ATR-FTIR spectroscopy of heparin: using multivariate analyses for next-generation quality control.

Thesis submitted in accordance with the requirements of Keele University for the degree of Doctor of Philosophy

By

Anthony James Devlin

June, 2023.

Abstract

Heparin is a member of the glycosaminoglycan (GAG) family of naturally occurring, polydisperse, linear polysaccharides. Heparin is a ubiquitous anticoagulant drug, a member of the WHO's list of vital medicines and the second used drug by weight after insulin. The structure of heparin is complex, extremely challenging to sequence - requiring large amounts of pure starting material, and there is currently no way of synthesising viable full-length heparin. Pharmaceutical heparin is therefore a natural product, extracted from the mucosa of pigs and cows. In 2007/2008, batches of heparin were found to have been adulterated with a semi-synthetic analogue (OSCS), which resulted in at least 100 deaths in the US alone. In 2011, Rudd *et al.* introduced the idea of multivariate analysis for heparin quality control and demonstrated that, through use of NMR and principal components analysis (PCA), pharmaceutical heparins could be defined in N-dimensions and that samples contaminated with OSCS could be discriminated to the level of 1% (w/w).

Here, infrared (IR) spectroscopy, coupled with multivariate analyses, is explored as a potential tool for the quality control of pharmaceutical heparin. Levels of OSCS as low as 0.25% (w/w) were detected and through expansion of the model, levels of 1 - 5% (w/w) of the other GAGs chondroitin sulphate (CS), dermatan sulphate (DS), hyaluronic acid (HA) and the sulphated semi-synthetic analogues thereof were detected. PC-regression (PCR) was applied to IR spectra of crude heparins – 2:1 (w/w) mixes of heparin:CS/DS which are believed to be where OSCS contamination first occurred – and was able to accurately detect and quantify heparin, CS and DS in the sample (RMSE = 2.2, 1.7 and 0.8 respectively).

PCR and partial least squares-regression were applied to the IR spectra of crude heparins, specifically aimed at the quantification of sulphate moieties on the carbon ring and was able to predict the levels of 6-O-, 2-O- and N- sulphate accurately ($RMSE_{average}$ = 0.91 and 0.76 respectively). The techniques were applied to heparins extracted from different animal sources and through cluster analysis, discrimination of heparins from pig, cow and sheep was achieved with 99% accuracy. Blends of porcine heparin with bovine and ovine heparin were also examined and detection was facilitated at the level of 10%.

Contained within this thesis is an exploration of the use of IR spectroscopy as both a tool for QC and a tool for GAG characterisation. IR, coupled with multivariate analyses is shown to detect low levels (sub 1%) of contaminants, accurately quantify and discriminate between different GAGs in a mixture and accurately quantify and discriminate between different sulphate moieties, all with accuracies that parallel those of NMR.

Acknowledgements

"If I have seen further than others, it is by standing on the shoulders of giants".

- Sir Isaac Newton

List of Publications

Publications:

Devlin, A., Mycroft-West, C., Turnbull, J., Lima, M., Guerrini, M., Yates, E., Skidmore, M. (2023) Analysis of heparin samples by attenuated total reflectance Fourier-transform infrared spectroscopy in the solid state. ACS Central Science, DOI: 10.1021/acscentsci.2c01176

Mycroft-West, C., **Devlin, A.,** Cooper, L, Guimond, S., Proctor, P., Miller, G., Guerrini, M., Fernig, D., Yates, E., Lima, M., Skidmore, M. (2023); A sulphated glycosaminoglycan extract from Placopecten magellanicus inhibits the Alzheimer's disease β-site amyloid precursor protein cleaving enzyme 1 (BACE-1). Carbohydrate Research, DOI: 10.1016/j.carres.2023.108747

Mycroft-West, C., **Devlin, A.**, Cooper, L, Guimond, S., Proctor, P., Guerrini, M., Miller, G., Fernig, D., Yates, E., Lima, M., Skidmore, M. (2020); Glycosaminoglycans from Litopenaeus vannamei Inhibit the Alzheimer's Disease β Secretase, BACE1. Marine Drugs., DOI: 10.3390/md19040203

Mycroft-West, C., **Devlin, A.,** Cooper, L, Proctor, P., Miller, G., Fernig, D., Guerrini, M., Yates, E., Guimond, S., Lima, M., Skidmore, M. (2020); Inhibition of BACE1, the Beta-Secretase Implicated in Alzheimer's Disease, by a Chondroitin Sulphate Extract from Sardina pilchardus. Neural Regen. Research., DOI: 10.4103/1673-5374.274341

Devlin, A., Mycroft-West, C., Proctor, P., Cooper, L., Guimond, S., Lima, M., Yates, E., Skidmore, M. (2019) Tools for the Quality Control of Pharmaceutical Heparin. Medicina., DOI: 10.3390/medicina55100636

Mycroft-West, C., Cooper, L, **Devlin, A.**, Proctor, P., Guimond, S., Guerrini, M., Fernig, D., Lima, M., Yates, E., Skidmore, M.. (2019) A Glycosaminoglycan Extract from Portnus pelagicus Inhibits BACE1, the Beta Secretase Implicated in Alzheimer's Disease. Marine Drugs., DOI: 10.3390/md17050293

Boyle, M., Skidmore, M., Dickerman, B., Cooper, L., **Devlin, A.,** Yates, E., Horrocks, P., Freeman, C., Chai, W., Beeson, J. (2017) Identification of Heparin Modifications and Polysaccharide Inhibitors of Plasmodium falciparum Merozoite Invasion That Have Potential for Novel Drug Development. Antimicrobial Agents and Chemotherapy, 61(11)., DOI: 10.1128/AAC.00709-17

Preprints:

Devlin, A., Mauri, L., Guerrini, M., Yates, E., Skidmore, M. (2019) (Preprint); The Use of ATR-FTIR Spectroscopy to Characterise Crude Heparin Samples by Composition and Structural Features. BioRxiv., DOI: 10.1101/744532

Devlin, A., Mycroft-West, C., Turnbull, J., Guerrini, M., Yates, E., Skidmore, M. (2019) (Preprint); Analysis of Solid-State Heparin Samples by ATR-FTIR Spectroscopy. BioRxiv., DOI: 10.1101/538074

List of Figures

CHAPTER 4: THE DETECTION OF, DISCRIMINATION OF AND QUANTIFICATION OF CONTAMINANTS OTHER THAN OSCS. 102

CHAPTER 5: IDENTIFICATION AND CHARACTERISATION OF CRUDE HEPARINS AND DISCRIMINATION BETWEEN HEPARINS FROM DIFFERENT ANIMAL SOURCES. 132

CHAPTER 6: FINAL DISCUSSION

CHAPTER 7: REFERENCES

CHAPTER 8: APPENDIX

[viii]

List of Tables

Glossary of Terms

2OH	hydroxyl at position 2 of GlcA/IdoA
2OST	2-O-sulfotransferase
2S	sulphate at position 2 of GlcA/IdoA
3OST	3-o-sulfotransferase
3S	sulphate at position 3 of GlcN
4S	sulphate at position 4 of GalN
6OH	Hydroxyl at position 6 of GlcN/GalN
6OST	6-O-sulfotransferase
6S	sulphate at position 6 of GlcN/GalN
6X	Position 6 can have a -OH or a -O-sulphate
ACT	activated clotting time
AMAC	2-aminoacridone
AP	2-aminopyridine
APTT	activated partial thromblastin time
AT	antithrombin
ATR	attenuated total reflectence
AUC	area under the curve
BHRA	Brazilian Health Regulation Authority
BK	bradykinin
BLH	bovine lung heparin
BMH	bovine mucosal heparin
BODIPY	4,4-difluoro-4-bora-3a,4a-diaza-*s*-indacene
BSE	bovine spongiform encepelopathy
C4ST-1:3	chondroitin-4-O-sulfotransferase-1:3
C6ST-1	chondroitin-6-O-sulfotransferase
CD	circular dichroism
CE	capillary electrophoresis
ChGn-1:2	Chondroitin GalNAc transferase 1:2
ChSy-1:3	Chondroitin Synthase 1:3
CMH	Chemically modified heparin
COS-Firs	correlation spectroscopy filtering with random iterative sampling
CS	chondroitin sulphate
D4ST-1	dermatan-4-O-sulfotransferase
DAP	diaminopropane
DBSCAN	density-based spatial clustering of applications with noise
DeS	Dextran sulphate
DMBB	Dimethylene blue
DNA	deoxyribose nucleic acid
DoS	degree of sulphation
DOSY	diffusion-order spectroscopy
DS	deramatan sulphate
ECM	extracellular matrix
EDTA	ethylenediaminetetraacetic acid
EM	expectation maximisation
EOF	electroosmotic flow

EP	european pharmacoepia
ESI	electrospray ionisation
EXTL1:3	Exostosin-like glycosyltransferase 1:3
FAB	fast atom bombardment
FDA	Food and drug administration
FGF	fibroblast growth factor
FGFR	Fibroblast growth factor receptor
FRET	Forster resonance energy transfer
FT	Fourier transform
Fuc	fucose
GAG	glycosaminoglycan
Gal	Galactose
GalN	Galactoseamine
GalNAc	Galactosamine-acetate
GalNac4,6S	Galactoseamine-4,6-O-sulphate
GalNAc4-6ST	GalNac4S 6-O-sulfotransferase
GalNAc4S	Galactosamine-4-O-sulphate
GalNAc6S	Galactosamine-6-O-Sulphate
GalNAcT-1:2	GalNac Transferase I:2
Glc	Glucose
GlcA	Glcuronic acid
GlcA2S	glucuronic acid -2-O-sulphate
GlcAT-II	GlcA transferase II
Glce	Glc C5 epimerase
GlcN	Glcusoamine
GlcN,6,3S	Glucosamine-N,6,3-O-sulphate
GlcN,6S	Glucosamine-N,6-O-sulphate
GlcNAc	Glucosamine-acetate
GlcNAc6ST	N-acetylglucosaminyl-6-sulfotransferase
GlcNS	Glucosamine-N-sulphate
GNR	Glucuronic acid at the non reducing end
GPC	gel permeation chromatography
HA	hyaluronic acid
HAS1:3	HA Synthase 1 through 3
HCA	hierarchical cluster analysis
HIT	Heparin induced thrombocytopenia
HILIC	hydrophobic interaction chromatography
HMQC	heteronuclear multiple-quantum correlation
HMWHA	high molecular weight hyaluronic acid
HS	heparan sulphate
HSQC	heteronuclear single-quantum correlation
IC	independent component
ICA	independent components analysis
ICP-AES	inductively coupled plasma-atom emission spectroscopy
ICP-MS	inductively coupled plasma-MS
IdoA	Iduronic acid
IdoA2S	Iduronic acid-2-O-sulphate
IR	infrared

IU	international units
KMC	K-means clustering
KMC-CC	K-means clustering with precomputed centroids
KNN	K-nearest neighbours
KS	keratan sulphate
KSGalST	KS galactosyl sulfotransferase
LALS	low angle light scattering
LC	liquid chromatography
LD	linear discriminant
LDA	Linear Discriminant Analysis
LMWH	Low molecular weight heparin
LMWHA	low molecular weight hyaluronic acid
LOD	limit of detection
LOQ	limit of quantification
LR	linkage region
LRM	logistic regression model
LV	Latent variable
MALDI	matrix-assisted laser desporption isolation
MALS	multi-angle light scattering
Man	Mannose
MHC	major histocompatibility complex
MLR	multiple linear regression
MS	mass spectrometry
MS/MS	tandem mass spectrometry
MurNAc	Muraminic acid acetate
MW	molecular weight
NDST	N-deacetylase/N-sulfotransferase
NIR	near-infrared
NMR	nuclear magnetic resonance
NOESY	nuclear Overhauser effect spectroscopy
NS	N-sulphate on GlcN
OMH	ovine mucosal heparin
OSAS	oversulphated-agarose
OSCS	oversulphated chondroitin sulphate
OSDS	oversulphated-dermatan sulphate
OSHA	oversulphated-hyaluronic acid
OSHp	oversulphated-heparin
OSHS	oversulphated-heparan sulphate
PAGE	polyacrylamide gel electrophoresis
PAPS	3'phosphoadenosine 5'-phosphosulphate
PC	principal component
PCA	principal component analysis
PCR	Principal Components regression
PF4	Platelet Factor 4
PG	proteoglycan
PLS-DA	Partial Least Squares discriminant analysis
PLS-R	Partial least squares regression
PMH	porcine mucosal heparin

PT	prothrombin time
pyr-GAG	pyridine salt of a GAG
QC	quality control
Q-PCR	Quantitative-polymerase chain reaction
RALS	right angle light scattering
RMSE	root mean squared error
RNA	ribose nucleid acid
ROC	response operator characteristic
RPIP	reversed phase ion pairing
SAX	strong-anion exchange
SEC	size exclusion chromatography
Sial	Sialic acid
SLS	static right angle scattering
TAE	tris-acetate-EDTA
TBE	Tris-borate-EDTA
TCA	trichloroacetic acid
TDA	triple detector
Thz	Terahertz
TOCSY	total correlation spectroscopy
t-SNE	t-distributed stochastic neighbourhood embedding
UDP-	Uridine disphosphate
UFH	unfractionated heparin
USP	united states pharmaceoepia
UST	uronyl-2-O-sulfotransferase
UV	ultraviolet
vCJD	variant Creutzfeldt Jacobs disease
Vis	visible
WAX	weak-anion exchange
WHO	world health organisation
Xyl	Xylose
β3GnT	β-1,3-N-acetyl-glucosaminyltransferase
β4GalT-1	β1,4-galactosyl-transferase-1
β4GalT-4	β1,4-galactosyl-transferase-4

Table of Contents

[xvi]

Chapter 1: Introduction

1.1 Carbohydrates

Carbohydrates are one of the four classes of biological macromolecule. Typically denoted with the suffix "-ose", carbohydrates are polymers consisting of repeating monosaccharides, the most basic of which comprise solely carbon, oxygen and hydrogen, with the empirical formula $C_n(H_2O)_n$. A carbohydrate polymer can also be referred to as a polysaccharide or a glycan, but the term glycan may infer that the carbohydrate is a moiety to a macromolecule, as in a proteoglycan. Carbohydrates possess no single unifying function; they may be readily lysed into manageable monosaccharides, making them an efficient energy source as in starch or can possess strong, chemically resistant structures that make them powerful structural components as in the cell wall of many eukaryotic species. Carbohydrates have a wide array of unique monosaccharides, linkages and may be naturally functionalised further, making them information rich molecules that are essential for many processes, including fertilization, immune response, pathogenesis and cell signalling (Dwek, 1996).

1.2 Monosaccharides.

Monosaccharides are hydrocarbon chains, containing between 3 and 9 carbons, each with a hydroxy functional group (-OH) and one with a ketone or aldehyde functional group (=O), forming tri- to non- aldose and ketose sugars. The presence of a =O and -OH allows the molecule to form a hemiacetal/hemiketal in equilibrium, resulting in a chiral alkoxy ring, usually with 5 or 6 members (furan- and pyran- respectively). Not all monosaccharides form rings and some will remain linear in aqueous solution. The directions of the hydroxyls in relation to each other, or more easily visualised against the plane of the ring, result in a series of different diastereomers, each of which are uniquely named. The relationship between the hydroxyls and the direction of the hydroxyl on the ring-forming carbon results in a pair of enantiomers and gives rise to D- L- isomerism, with almost all carbohydrates in nature being D- isomers. Different enantiomers of carbohydrates can infer different activities, for example, the presence of L-fucose, the deoxylated enantiomer of D-galactose represents the key difference between the Lewis[b] and Lewis[a] antigen, which also correlates with differing bacterial pathogenicity in some gastric illnesses (Becker and Lowe, 2003; Hooper and Gordon, 2001)

Cyclic carbohydrates possess a unique form of isomerism – anomerism. The location of the =O (C1 for aldose sugars) denotes the anomeric carbon and the direction of the hydroxyl of the anomeric carbon, relative to the direction of the remainder of the carbon chain at the bound anomeric reference (usually C4 or 5 for furan- and pyranose rings) results two isoforms: α

(functional groups on opposing sides) and β (functional groups on the same side) (**Fig. 1.1**). The anomer of the carbohydrate can greatly change its base properties and therefore its functions, for example α-glucose polymers, i.e starch, are soluble, can be heavily branched and may be enzymatically digested in humans, while β-glucose polymers form, amongst others, cellulose – an insoluble and important linear structural polymer. Cyclic forms of monosaccharides form different conformers in solution as the bond angles about each carbon alter into different 3D structures. Furanose rings typically form a "twist" or an "envelope" conformer, while pyranose rings typically form a "chair" conformer. Seven conformers are found in nature and monosaccharides can transfer freely between them but different conditions, including temperature, pH or pendant groups will alter the different ratios of certain conformers in solution (**Fig. 1.1**).
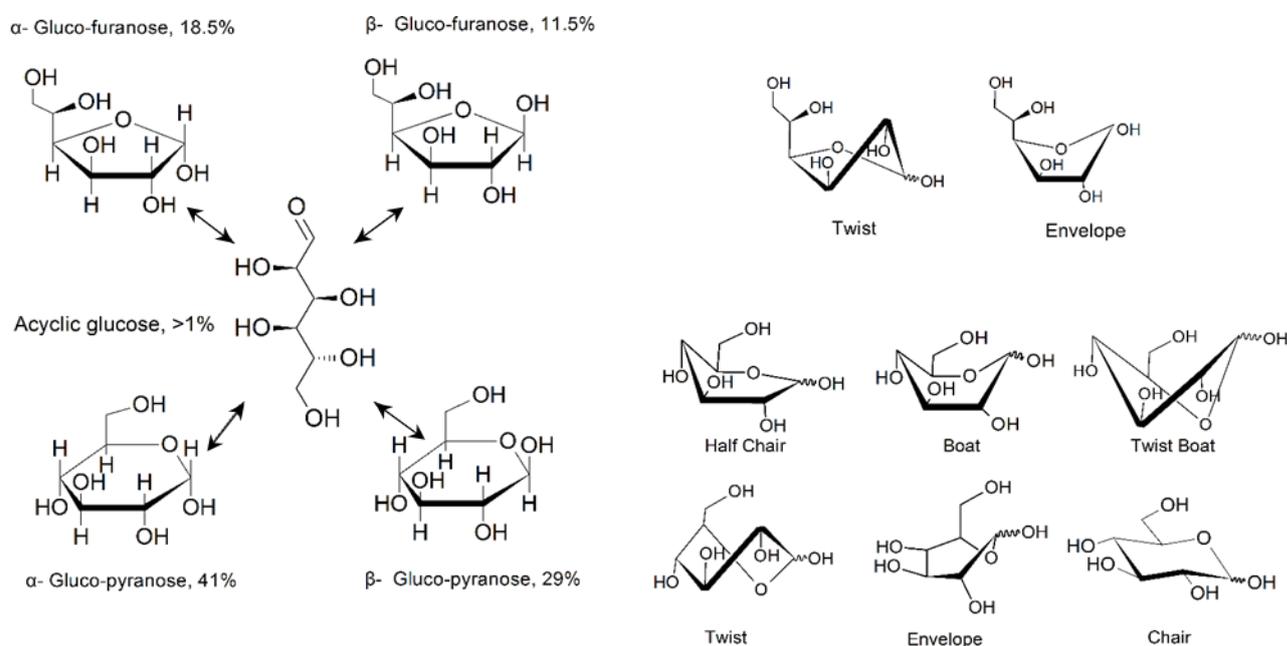
Figure 1.1: Anomer and conformer isomers of glucose. *Top row*) *Glucofuranose (5 carbon) rings forming α- and β- anomers the anomeric reference for furanose rings is at C4, whilst the anomeric carbon is at C1. **Bottom row**) Glucopyranose (6 carbon) rings forming α- and β- anomers. The anomeric reference for pyranose rings as it C5, whilst the anomeric carbon is C1. The different conformers that can form in 3D space for both furanose rings (top) and pyranose rings (bottom) are displayed to the right. Figure reproduced using (Dwek, 1996; Sinnott, 2007).*

1.3 Monosaccharide derivatives

Saccharides may be oxidised at one end to form an -aldonic acid, both ends to form an -aldaric acid or at a hydroxyl to form a -uronic acid. Further to this, the hydroxyls may form esters with inorganic molecules, including sulphates and phosphates, the locations of which can greatly alter function and 3D structure. The hydroxyls may also be methylated, facilitating increased propensity to bind to other methylated carbohydrates, as in pectins – an important structural carbohydrate found in the cell walls of most plants (Liang et al., 2012). Saccharides may also be aminated, forming amino-sugars and the resulting free amine may be further functionalised

with acetals, such as acetate or pyruvate, or with sulphates. Saccharides may also form deoxy-sugars, where a hydroxyl is replaced with a hydrogen (H). Oxygen bridges may also form between carbons, creating anhydro-sugars.

1.4 Oligosaccharides and Polysaccharides

Oligosaccharides are considered anywhere between three and 10 monosaccharides, and polysaccharides are anything above 10, but the two are often interchangeable. In theory, any free hydroxyl on a carbon may react with any free hydroxyl on another monosaccharide, allowing a vast number of glycosidic bonds to form, but generally, C1 binds to another monosaccharide. As monosaccharides may be acyclic in solution, one end of a polysaccharide may also be acyclic, dubbed the reducing end due to its ability to be oxidised, however not all polysaccharides, particularly those whose anomeric carbons are involved in glycosidic bonds, have a reducing end. Most polysaccharides usually consist of a repeat unit, be it a monosaccharide, disaccharide or higher.

Saccharides are structurally complex, a single hexose sugar can form 8 diasteromers of 2 anomers, each with unique characteristics and chemical properties. Furthermore, reduction, oxidation or functionalisation of the hydroxyl groups and the ability to form a glycosidic bond to any other hydroxyl on any other monosaccharide, make saccharides incredibly information dense molecules. (Buleon, et al., 1998), (Ghazarian, et al., 2011).

1.5 Glycoconjugates

Polysaccharides can covalently bind to many different molecules such as polyamino acids, resulting in a peptidoglycan and glycopeptides; proteins, resulting in glycoproteins; lipids, resulting in glycolipids/ lipopolysaccharides and other small molecules, resulting in glycosides. The non-glycan moieties can serve as an anchor for the glycan, as in glycolipids, attaching it to cell membranes and presenting the polysaccharide as a ligand, allowing for cell-cell recognition (Schnaar, 2004), or serving as a linker to a peptide (Paulick and Bertozzi, 2008), or be used to transport a glycan moiety for glycosylation to another molecule (Mule et al., 2021). The non-glycan moieties may also serve to strengthen the molecule, as in peptidoglycan – a part of the bacterial cell wall, where long polymers of N-acetylglucose (GlcNAc) and N-acetylmuramic acid (the ether of GlcNAc and lactic acid) are covalently bound to poly-amino acid moieties which then cross-link to provide structural rigidity and also retain permeability to many small molecules (Demchick and Koch, 1996; Scheffers and Pinho, 2005), or in lipopolysaccharide, a bacterial glycolipid with structural roles (Rietschel et al., 1994). Glycosides can be used to solubilise and deactivate chemicals found in an organism, serving as a method of detoxification (Easson et al., 2021) or as a defence mechanism: storing a toxin for release upon metabolism (Gleadow and Møller, 2014).

Glycopeptides are used medicinally as drugs, as in the antimicrobial vancomycin but biologically, they play important roles in immune response, allowing glycans which cannot usually interact with the major histocompatibility complex (MHC) to interact with it (Werdelin et al., 2002). Glycoproteins are universal to life; glycosylation representing the most complex post-translational modification (Spiro, 2002). Glycosylation allows for the non-genetic marking of proteins, indicating protein age or disease states, the glycan moieties can also alter the secondary structure of the bound protein, both activating or deactivating the protein and modulating the stability and solubility, aid in localisation, or act as a linker between proteins (Dennis et al., 1999; Mule et al., 2021; Spiro, 2002). The ABO blood groups are a well-known example of a glycoprotein, the study of which has been vital to modern medicine. Glycans bound to a protein can be broadly split into two categories (N-linked and O-linked), depending upon the linkage formed between the glycan and the protein but others do exist (**Fig. 1.2**).
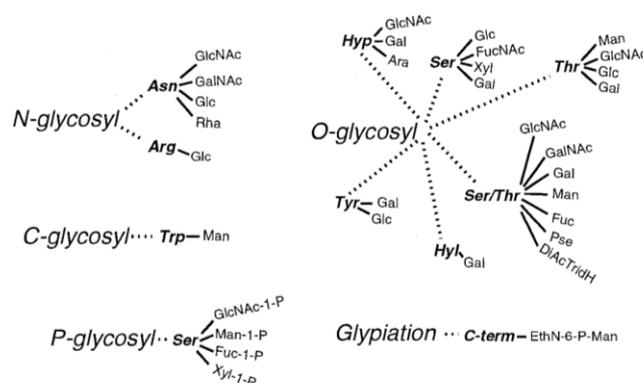
Figure 1.2: Types of glycosylation found in nature. Figure reproduced from (Spiro, 2002). Hyp: hydroxyproline, Hyl: hydroxylysine.

1.6 N-Linked glycans

N-glycosylation occurs primarily on asparagine residues that are part of the sequence NX(S/T) where X is any amino acid bar a proline. N-glycans are based off the same initial branched 14-mer, which is assembled and covalently linked via an amide bond (N-linked) to the target protein (**Fig. 1.3**). The core 14-mer aids in the action of chaperone proteins through interaction and the monitoring of the core 14-mer's structure during protein-folding is vital for correct protein folding, assembly and disluphide-bridge formation (Dennis et al., 1999; Ferris et al., 2014). The 14-mer may also be trimmed or expanded or both, to form one of three main N-glycan classes: an oligomannose, essentially a trimmed 14-mer or a Glc expanded 14-mer: a complex N-glycan, comprising many further antennae from the core: and a hybrid N-glycan, consisting of polymerised mannose and one or two antennae. The resultant N-glycans may also be capped with Sials, or fucosylated, and the levels of these modifications may be indicative of underlying processes in the cells, for example fucosylation is present in hepatocellular carcinoma but not in chronic liver disease (Taniguchi and Kizuka, 2015, p. 2).

The ratios of different N-glycans can modulate cell-cell interactions including cell-matrix adhesion, cell migration and modulate protein clearance (Schachter, 2000; Taniguchi and Kizuka, 2015).
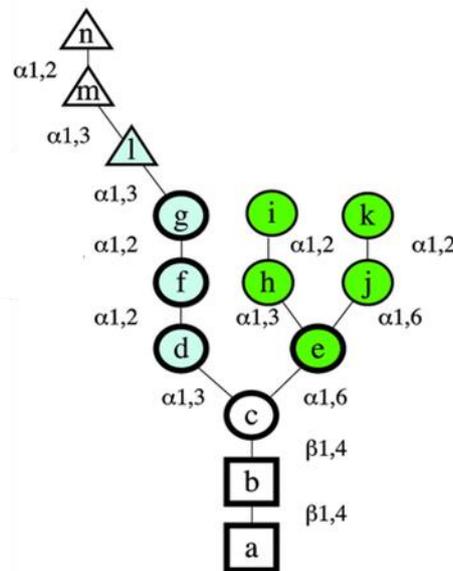
Figure 1.3: The core linker for all N-linked glycans. Glc: triangles, GlcNAc: squares, Man: circles. Figure reproduced from (Helenius and Aebi, 2004)

1.7 O-Linked glycans

O-linked glycans begin from the hydroxyl found on Ser/Thr residues and are split into six groups based on their primary linkage saccharide. O-Linked glucans encompass O-fucoses (O-Fuc), O-glucoses (O-Glc), O-mannoses (O-Man), O-aminogalactose (O-GalNAc), O-GlcNAcs and O-Xyloses (O-Xyl). O-fucosylation and O-glucosylation are responsible for correct protein folding and O-fucose may also serve as a marker for correctly folded epidermal growth factor-like repeats and therefore influence the binding of proteins to their ligands. O-fucosylation and O-glucosylation are both involved in the correct folding of the Notch protein, a vital protein in Notch signalling – an evolutionarily conserved pathway that regulates cell-fate (Holdener and Haltiwanger, 2019, p.; Yu and Takeuchi, 2019). O-glucosylation is rare and has been almost entirely studied in the context of Notch signalling, hence its true nature is unknown. O-mannosylation is vital for correct protein folding and cell wall integrity in fungi, where manno-proteins (with both N-glycans and O-mannans) are integral to its structure (Loibl and Strahl, 2013). Similarly, O-mannosylation is found in humans, where it is responsible for the correct structure of α-dystroglycan, a protein integral to the formation of muscle fibres (Dobson et al., 2013), and the correct structure of proteins that modulate cell-cell and cell-matrix interactions (Praissman and Wells, 2014), cancer metastasis and viral entry (Sheikh et al., 2017).

O-GlcNAcylation is found on over 80% of secretory and cell surface proteins but is particularly well known for its decoration of mucins, an essential component in mucous and the first line of defence in blocking pathogen invasion. O-GlcNAcylation begins also with a core, the simplest of which being GalNAc (also known as the Tn antigen). The GalNAc can be expanded, to form four common cores and four rare cores, which may all be matured with numerous epitopes, including the ABO blood group and Lewis antigens (Wang et al., 2021). Mucin dysregulation is common in cancers, where they have antiadhesive effects that benefit metastatic cells and can also disguise cancerous cells from the immune system (Hollingsworth and Swanson, 2004) O-GlcNAcylation is a unique modification, whereby expansion of it does not happen, but instead enzymes can add and cleave GlcNAc in a dynamic system analogous to phosphorylation (Spiro, 2002). O-GlcNAcylation is believed to respond in a dynamic manner to different nutrient concentrations, allowing the cell to fine-control its responses – in particular it modifies and is modified by phosphate cycling (Hart et al., 2011). O-GlcNAcylation is associated with cellular stress response, has roles in protein production and maintenance, is a component of the cytoskeleton and is also associated with many disease states, an extensive review of all can be found in (Hart et al., 2011).

O-Xylosylation is involved solely in the binding of glycosaminoglycans (GAGs) to a protein core, resulting in a proteoglycan (PG). GAGs are a group of polydisperse linear polysaccharides, consisting of repeat units of an amino-sugar (GlcNAc as in hyaluronic acid (HA), heparin and heparan sulphate (HS) and GalNAc as in chondroitin sulphate (CS)) bound to a glucuronic acid residue (GlcA) which may be epimerised at C5 to iduronic acid (IdoA).

1.8 Glycosaminoglycans

GAGs and their associated PGs are responsible for numerous biological functions, typically biological roles require the protein core, while pharmaceutical uses are found in just the GAG component. GAGs exist in an environment densely populated with various proteins, with which they play a vital role in assembly of protein-protein complexes and initiate and inhibit various signalling and biochemical pathways (Raman et al., 2005) and may also provide important architectural support to the surrounding tissues, especially in cartilaginous and corneal structures (Bastow et al., 2008; Roughley and Mort, 2014). There are six GAGs, some with subtypes (**Fig. 1.4**). HA is an entirely unsubstituted chain of GlcNAc-β-(1-4)-GlcA-β-(1-3)- (Vigetti et al., 2014), while CS is a chain of GalNAc-β-(1-4)-GlcA-β-(1-3)-, with O-sulphation at position 4 (4S) or 6 (6S) of the GalNAc denoting two of the different CS subtypes of -A and -C respectively (Sugahara et al., 2003). DS, originally identified as CSB, is a C5 epimer of CS and possesses a repeating disaccharide of GalNAc6S-β-(1-4)-L-IdoA2S-α(1-3)-(Sugahara et al., 2003). KS consists of GlcNAc-β-(1-3)-Gal-β-(1-4)- and may be O-sulphated at position 6 on one, both or neither monosaccharide (Caterson and Melrose, 2018). The final two GAGs,

heparin and HS are very similar in backbone structure and consist of a disaccharide repeat of GlcNAc-β-(1-4)-GlcA-β-(1-4)-, which may be sulphated at position 6 and – most commonly in heparin - the amine and position 3 of the GlcNAc, and at position 2 and 6 of the GlcA which may also undergo epimerisation to form IdoA, albeit mostly in heparin (Rabenstein, 2002a). GAGs have numerous roles and are involved in numerous structures and signalling pathways, ranging from development to immunity (Yamada et al., 2011).

All the GAGs are synthesised in specific compartments of the Golgi and are found naturally attached to a protein, with the exception of HA, which is synthesised at the plasma membrane from a segregated pool of precursors (Prydz, 2015). All GAGs that are linked to a protein core are synthesised from the same core tetrasaccharide – Xyl-(4-1)β-Gal-(3-1)β-Gal-(3-1)β-GlcA – which is attached at the *cis*-Golgi (Prydz, 2015) with the exception of KS, which may form multiple linkage regions (Caterson and Melrose, 2018). Attachment of the linker to a protein requires a specific GAG attachment site – Aaa- Aaa- Aaa- Aaa-Gly-Ser-Gly- Aaa-Bbb- Aaa, where Aaa = Glu or Asp and Bbb = Gly, Glu or Asp (Prydz, 2015).

Modifications to the linker will determine which GAG it matures into; capping of the linker with β-GlcNAc will result in a HS/Heparin chain and capping with β-GalNAc will result in a CS/DS chain (Prydz, 2015). Sulphation of the linker has also been found on CS/DS chains but not HS/heparin chains, suggesting that sulphation of the linker attenuates HS/heparin maturation. Sulphation of the linker Gals at position 4 or 6 also corresponds to the sulphation pattern of the matured GAG and may be integral to driving its synthesis (Pavão et al., 2006). The structure of the protein that the linker is bound too may also influence which GAG is synthesised, and removal of protein domains may alter the levels of HS vs CS while mutation of amino-acids can also alter the relative amount of HS chains present on the PG (Prydz, 2015). CS chains are functionally defined by the protein core to which they are bound, whereby levels and location of sulphation can alter significantly depending on the structure of the core. For example, the CS chains that decorate the PGs Syndecan-1 and Syndecan-4 have an average degree of sulphation (DoS) of 0.78 and 1.03 respectively, with an increased level of unsulphated disaccharides in Syndecan-1 and a ~10% increase in both 6S and 4S in Syndecan-4 (Deepa et al., 2004).
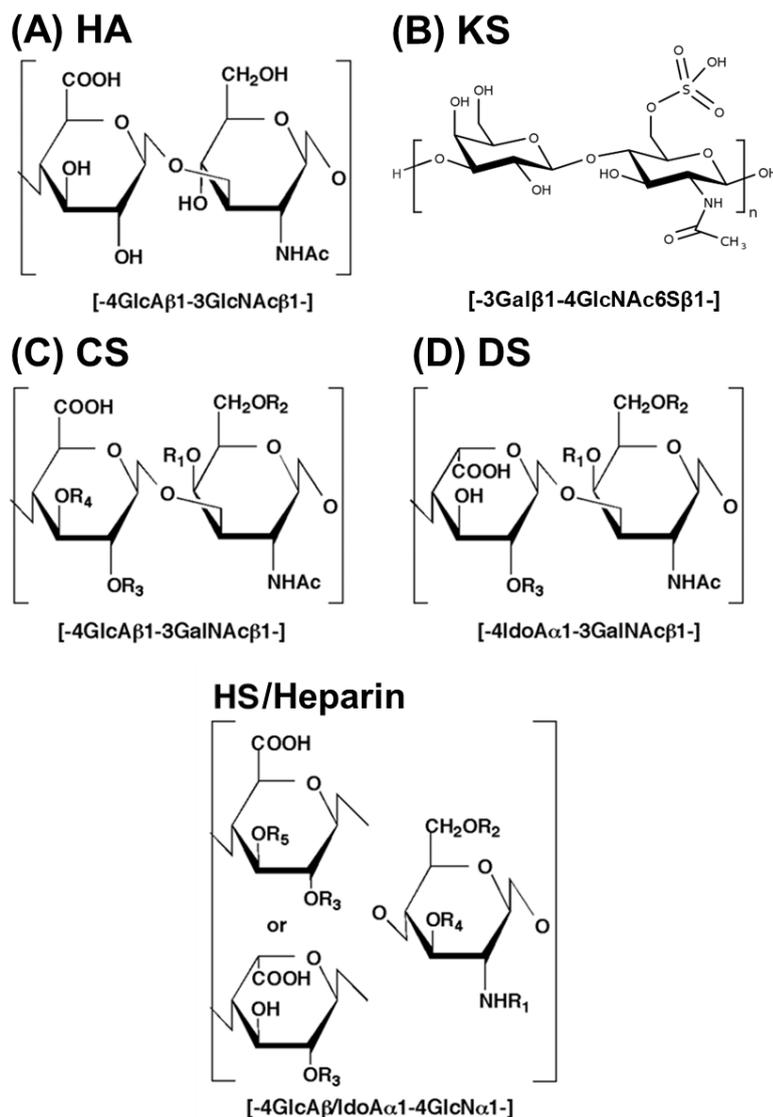
Figure 1.4: The major disaccharide repeats for each GAG. A) The GlcNAc-GlcA repeat of HA. **B)** The GlcNAc-Gal repeat of KS. **C)** The GalNAc-GlcA repeat of CS. **D)** The GalNAc-IdoA repeat of CS. **E)** The disaccharides available to HS/Heparin. R1= SO3 or CH3CO, R2,3,4,5 = H or SO3. Figure created using (Funderburgh, 2002; Trowbridge and Gallo, 2002; Yamada et al., 2011).

1.9 Hyaluronic Acid Biosynthesis

HA is synthesised in mammals by one of three HA-synthase enzymes (HAS 1 through 3 respectively) all of which share 55-71% sequence homology. The different enzymes create essentially the same product, but differ in their elongation rates and affinity for substrates. The HAS used may even influence the formation of HA matrices post-synthesis (Itano, 2008). HA has no core saccharide to start its synthesis, and instead the free nucleotide-sugars UDP-GlcNAc and UDP-GlcA enter the enzyme and elongate the growing HA chain. The elongation seems to simultaneously extrude the forming HA chain straight into the extracellular space through a pore formed by the HAS, alternatively HA may be secreted by an ABC system (Itano, 2008; Weigel and DeAngelis, 2007).

1.10 Hyaluronic acid Structure and Function

HA is structurally the simplest GAG, possessing no sulphation and no attachment to a protein core. It is the largest GAG, ranging from $2x10^5$ to $5x10^6$ kDa. HA is hygroscopic, has a large hydration volume and interacts with neighbouring molecules to form viscoelastic solutions, organising them while simultaneously filling extra space, making it a potent shock absorber (Lee and Spicer, 2000). HA also displays excluded volume effects, which restrict the access of larger molecules while allowing it to act as a permeable membrane that only allows components of low molecular weight to pass (Bastow et al., 2008). The ability of HA to act as a permeable barrier is important to some prokaryotes, which use HA as a shield from host-recognition (Vigetti et al., 2014). HA also stabilises water at increasing temperatures (Dong et al., 2020). Chondrocytes (cells which produce and maintain the cartilage) synthesise and secrete vast amounts of proteoglycans (primarily aggrecan – a CS and KS PG) and HA into the extracellular matrix (ECM). HA and aggrecan interact non-covalently with each-other and a small protein called link protein to form an essentially unbreakable interaction, allowing aggrecans to decorate the HA, driving the formation of an unusually high concentration of aggrecan without creating a gel. The aggrecan-HA units are dispersed throughout the collagen fibrils located in collagen, where they fill space and provide support. The CS/KS chains on aggrecan can swell with water due to their sulphate moieties creating an equilibrium of pressure throughout the collagen fibrils, providing a high compressive resilience (Bastow et al., 2008; Kiani et al., 2002; Roughley and Mort, 2014).

Unlike in other GAGs, where the levels and positions of sulphates correlate with function, the chain length of HA denotes biological activity. The larger the molecular weight (MW) of the HA, the more structural its role seems to be, with molecules larger than 1 MDa (HMW) having organisational and space filling roles, and supressing angiogenic, immune and inflammatory responses, and even inducing macrophage polarisation to an inflammatory phenotype (Lee et al., 2021). Smaller molecules (LMW) are involved in cell migration through matrix reorganisation, danger signalling as a ligand to multiple proteins, immune activation and wound repair (Lee-Sayer et al., 2015). Since all HAS enzymes create full length HMWHA (Weigel and DeAngelis, 2007), it is believed that LMWHA is produced through enzymatic, chemical or mechanical degradation and hence serves as a signalling molecule repair and stress in response to its degradation (Lee-Sayer et al., 2015).

1.11 Chondroitin sulphate and dermatan sulphate synthesis

CS/DS synthesis can be summarised in three steps: 1) synthesis of the core linkage region, 2) elongation of the polymer through addition of successive GlcA and GalNAc residues and 3) modification of the chain, through sulfation or epimerisation (Silbert and Sugumaran, 2002; Sugahara et al., 2003). After assembly of the linkage region and subsequent addition of the

initiating GalNAc by GalNAc transferase I (GalNAcT-I) (Prydz, 2015), the unsulphated backbone is polymerised by alternate addition of the activated sugars UDP-GlcA and UDP-GalNAc residues by action of GlcA transferase II (GlcAT-II) and GalNAc transferase II (GalNAcT-II) respectively (Silbert and Sugumaran, 2002). Three synthase enzymes, chondroitin synthase 1:3 (ChSy-1:3 respectively) possess dual transferase activity but require the protein chondroitin polymerizing factor to function (Sugahara et al., 2003). Two GlcNAc transferases that are specific to chondroitin also exist: chondroitin-GalNAc transferase 1 and 2 (ChGn-1 and 2) (Sugahara et al., 2003).

GlcA can be epimerised to IdoA through the action of 2 epimerases, DS-epi1 and DS-epi2 (Maccarana et al., 2009). Seven sulfotransferases are responsible for the sulfation of CS/DS through transfer of a sulphate group from 3'phosphoadenosine 5'-phosphosulphate (PAPS) to the desired hydroxyl. First, sulphation of position 4 or 6 occurs to generate CSA and CSC, after this further sulphation can occur to create highly sulphated variants, such as CSD or CSE (Mikami and Kitagawa, 2013). Four 4-sulfotransferase enzymes exist in *Homo sapiens*; chondroitin-4-o-sulfotransferase-1:3 (C4ST-1:3) and dermatan-4-O-sulfotransferase (D4ST-1) which are responsible for CSC and DS sulfation respectively (He et al., 2017). C4ST-2 and 3 and D4ST-1 are functionally redundant, but C4ST-1 seems to play a distinct regulatory role in both CSA synthesis and total CS synthesis. Chondroitin 6-O-sulfotransferase-1 (C6ST-1) is involved in CSC formation and no DS-6S sulfotransferase has been identified (Mikami and Kitagawa, 2013). Two final sulfotransferase enzymes remain, and they are responsible for IdoA2S formation (uronyl 2-O-sulfotransferase (UST) and formation of GalNAc4,6S from GalNAc4S (GalNAc4S 6-O-sulfotransferase (GalNAc42-6ST)) (Malmström et al., 2012; Mikami and Kitagawa, 2013; Pavão et al., 2006). Chain elongation may also be stimulated through action of (C4ST-1), which in combination with ChGn-2, polymerises the CS further meanwhile, ChGn-1 is thought add the first GlcNAc to the linkage region, which in combination with 4-sulphation from C4ST-2 can begin polymerisation. ChGn-1 is therefore thought to regulate the number of CS chains on the PG (Mikami and Kitagawa, 2013).

1.12 Chondroitin and dermatan sulphate Structure and Function.

CS is the most predominant GAG, with roles in signalling, nervous development, cellular adhesion and pathogenic invasion (Sugahara et al., 2003). Chondroitin sulphates and their derivatives also interact with proteins that bind other GAGs, such as antithrombin, albeit with differing affinities (Amarasekara et al., 2007). Chondroitin sulphates are normally split into subtypes, the four most common of which are CSA, and -C through -E. CSA and -C are GalNAc-4S and -6S respectively, while CSD and -E are "highly sulphated" variants, with CSD being a GlcA-2S form of CSC, and CSE containing both 4S and 6S on GalNAc (Sugahara et al., 2003) (**Fig 1.5**). Other forms of CS exist, including trisulphated and 3S variants, but these

are typically rare and species specific (Ueoka et al., 1999; Yamada et al., 2011). CS profiles which also include non-sulphated and highly sulphated disaccharides are found dependant on core protein they are bound to, the species of organism and even the tissue location (Alkhalil et al., 2000; Deepa et al., 2004). DS is a GlcA C5 epimer of CSD, where the GlcA-2S is epimerised to IdoA-2S (sometimes referred to as CSB). DS, while closely structurally related to CS is found in entirely different cellular locations – primarily the skin when compared with CSs, which are found primarily in the cartilage (Mishra and Ganguli, 2021). DS can be up to twice as massive as CS and is responsible for different biological functions, including wound repair and fibrosis. DS chains may contain GlcA and also there may be small amounts of IdoA found in normal CS, representing hybrid chain structures containing low levels (~15%) of DS, synthesis of which is influenced by its bound protein (Prydz, 2015). The naming conventions for CS and DS are unclear, typically CSA will contain small amounts of CSC and vice versa. The key distinguisher between the two is a higher abundance of the relevant type of sulphation. To compound this, some individuals will claim that a single IdoA residue is indicative of an entire DS chain. In reality, chondroitins and dermatans exist on a spectrum which contain amounts of each CS type and C5 epimer that are relevant to the biological function required.
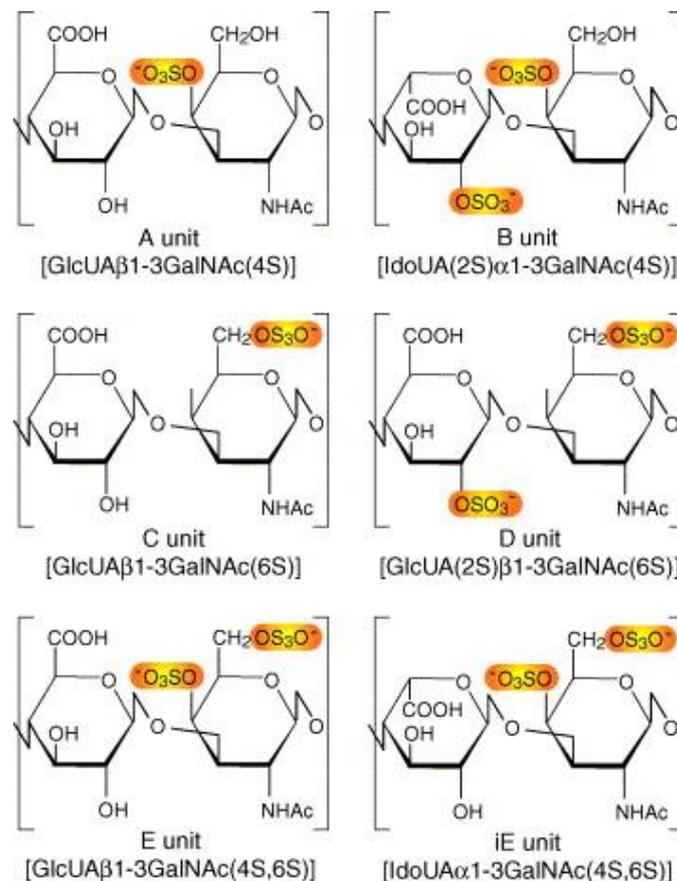
Figure 1.5: Common disaccharides found in different chondroitin sulphates. Figure reproduced from (Sugahara et al., 2003).

The activities of CS/DS are tied closely to their sulphation level and pattern (Alkhalil et al., 2000; Hitchcock et al., 2006; Kawashima et al., 2002; Mikami and Kitagawa, 2013; Sugahara et al., 2003). CSA, -C and -D do not bind to L- or P- selectin – cellular adhesion molecules involved in cell-cell interactions – while CSB and CSE do (Hashii et al., 2010; Kawashima et al., 2002). CSA can also impede neurite outgrowth, while CSC has no effect and CSE conversely stimulates neurite outgrowth (Mikami and Kitagawa, 2013). CSA is also required by the malarial parasite *Plasmodium falciparum* for binding and subsequent infection of red blood cells in the placenta. CSC and highly sulphated CS was found to inhibit parasite binding possibly due to steric hindrance or the importance of free hydroxyls. Some degree of unsulphated residues are required, ranging from 70—48% for maximal binding (Alkhalil et al., 2000; Gowda, 2006). CS may also sequester metal ions such as Cu^{2+} and Fe^{2+}, due to the high anionic charge of the sulphate groups, inhibiting free radical production and modulating oxidative stress. CSA in particular exerts this effect, and it is believed to be due to the spatial relationship between the C-4S and carboxylate group (Mishra and Ganguli, 2021). Highly sulphated CS and DS both bind to proteins that bind heparin/HS such as fibroblast growth factor (FGF), fibronectin, and proteins involved in the coagulation cascade, including thrombin and heparin cofactor II (Trowbridge and Gallo, 2002).

The sulphation pattern and length of CS bound to aggrecan has been shown to alter with age, decreasing in size and increasing in levels of 6-S from approximately 55:30:15 6S:4S:0S to 95:5 6-S:4S in adults (Plaas et al., 1997; Roughley et al., 1987). The increased chain length of the CS in young people, coupled with the decrease in overall sulphation suggests that the same osmotic force is maintained by the same amount of sulphate but over a larger distance. In patients suffering with osteoarthritis, a chronic disease characterised by dysregulation of the collagenous fibres and chondrocyte death, CS epitope 846 – a currently undefined epitope which is also found in foetal aggrecan, is increased, suggesting that the sulphation pattern changes (Rizkalla et al., 1992). This was later confirmed by (Lin et al., 2020), where both the total amount of CS and the level of CS-4S compared with CS-6S altered significantly. Administered CS has also been shown to stimulate HAS activity, particularly upregulating HAS1 and HAS2 and producing HMWHA, helping to treat the dysregulated collagen of osteoarthritis, suggesting tight regulation of GAGs by each other (David-Raoudi et al., 2009).

Versican is a CS/DS PG from the same family as aggrecan, that is found in the ECM and has roles in cell adhesion, migration and proliferation. Versican, like aggrecan also produces a viscoelastic environment and helps to stabilise the ECM but is instead found instead in various soft tissues. Unlike aggrecan, versican GAG chains consist of around 85% 4S residues, and also contains ~10% IdoA compared with aggrecans ~2%. The IdoA is thought to provide flexibility to the CS/DS chains, while the sulphation type seems required to prevent CS from interfering with versican binding to other proteins (Hitchcock et al., 2006). Interestingly, an

increased level of 6-S, particularly the formation of the highly sulphated 4-6S (E isoform), makes the CS/DS chains that decorate versican bind tightly to and compete with versican, retarding function (Kawashima et al., 2002). Alterations in sulphation pattern can also be indicative of cancer, whereby the expression of synthetic enzymes is altered, resulting in unusually oversulphated residues as seen in pancreatic cancer and the structural alterations can drive metastatic progression in aggressive breast cancer, by binding to P-selectin and encouraging secondary tumour growth (Cooney et al., 2011, 2011; Ren et al., 2021).

1.13 Keratan Sulphate Synthesis

KS synthesis begins post linkage-region synthesis. KS does not have the same linkage region as the other GAGs and is in turn split into three sub-types (KSI, KSII and KSIII) which are defined by the linkage regions (N- Linked, O-GalNAc-Linked and O-Man-Linked respectively) that they possess (Funderburgh, 2002). KSI is believed to begin from a similar process to other N-linked glycans, while KSII and KSIII are elongated from the cores of O-GalNAc and of O-mannosyls respectively. Extension of the chain from the core occurs from the sequential action of β-1,3-N-acetyl-glucosaminyltransferase (β3GnT), N-acetylglucosaminyl-6-sulfotransferase (GlcNAc6ST), β1,4-galactosyl-transferase-1 (β4GalT-1) and KS galactosyl sulfotransferase (KSGalST), which are responsible for addition of GalNac, sulphation of GalNAc to GalNAc6S, addition of Gal and sulphation of Gal to Gal6S respectively (Caterson and Melrose, 2018). Another enzyme, β1,4-galactosyl-transferase-4 (β4GalT-4) transfers Gal onto solely a GlcNAc-6S acceptor and is also capable of generating the initial branch points found in branched KS (Caterson and Melrose, 2018).

1.14 Keratan Sulphate structure and function

KS is located in corneal, cartilaginous, bone and brain tissues, where it helps to organise collagen fibrils and hydrate the microenvironment it is found in (Quantock et al., 2010). KS in the brain directs axonal growth by acting as a barrier for neuritogenesis and is involved in neuronal repair through association with glial scar formation (Caterson and Melrose, 2018; Zhang et al., 2006). The highly anionic KS chains can interact with water and ions, helping to hydrate nearby tissues and also sequestering Ca^{2+} for the generation of action potentials at the synapse (Caterson and Melrose, 2018). The organisation of collagen fibrils in the cornea is thought to be controlled by KS chains, regulation of which keeps the cornea transparent, hence dysregulation of KSPGs results in hazing of the cornea (Quantock et al., 2010). In aggrecan, the true function of KS is unknown, but the KS located near the N-terminus, is highly sulphated and is believed to supress aggrecan breakdown (Caterson and Melrose, 2018).

KSI is found throughout the body, while KSII is found exclusively bound to aggrecan and KSIII exclusively in the brain (Funderburgh, 2002). The linkage region of KSI possesses 5

saccharides in common with the standard N-glycan core, is bound to an asparagine and may possess biantennary expansions from C2 of the penultimate mannose which are similar in length to the primary expansion. KSII has a small, sialic acid (Sial) capped disaccharide expansion and, KSIII has the same base core as O-mannosyls and seems to be entirely linear (Funderburgh, 2002). Sulphation is vital to the activities of KS, and no-, low- and high- sulphate regions, consisting of GlcNAc-GalA, GlcNAc6S-GalA or GlcNAc6S-Gal6S respectively can be found throughout the KS chain. In corneal KS, the highly sulphated region is located at the end of the chain and the low-sulphated region, closer to the protein core (Funderburgh, 2002). This may be to allow the far end of the molecule to interact and bind with collagen while retaining some degree of flexibility to allow for collagen positioning. Other types of KS do not possess a domain structure, may be smaller in chain length (Funderburgh, 2002) and KSIII can contain non-sulphated disaccharides (Zhang et al., 2006).

1.15 Heparin and Heparan Sulphate Biosynthesis

Biosynthesis of HS and heparin happens in many overlapping steps and involves many different processes including elongation, N-deacetylation and N-sulphation, C5 epimerisation of GlcA to IdoA and O-sulphation at positions 2, 6 and more rarely 3. After formation of the linkage region, any of the exostosin-like (EXTL) family of glycosyltransferase enzymes initiates chain elongation. Of the three EXTL enzymes (EXTL1:3), EXTL3 is the main initiator of HS synthesis but all three may initiate chain elongation. EXTL2 may transfer either an α-GlcNAc or an α-GalNAc (Kitagawa et al., 1999), which may initiate HS chain elongation or block CS chain elongation respectively (Kreuger and Kjellén, 2012). EXTL2 may also regulate HS chain length, but the underlying mechanism is not clear (Katta et al., 2015). The chain is then elongated by action of two exostosin enzymes (EXT1 and 2) in a complex, which successively transfer GlcA and GlcNAc onto the chain. EXT1 possesses both GlcA and GlcNAc transferase activity and it is not clear if EXT2 possesses the same activity, however complexation of the EXT1 and EXT2 results in significantly higher polymerisation activity and both are required for HS synthesis *in vivo* (Busse-Wicher et al., 2014).

From here, post-elongation modification is thought to begin with deN-acetylation and subsequent N-sulphation, whereby N-deacetylase/N-sulfotransferase (NDST) enzymes move along the chain, removing acetyl groups from GlcNAc, to form GlcN and sulphating them to GlcNS (Deligny et al., 2016). Very rarely, the N-sulphation fails, possibly due to a lack of PAPS, leaving behind a free amine. NDST1 is primarily associated with HS N-sulphation, while NDST2 is primarily associated with heparin N-sulphation, however NDST2 can also sulphate HS, and may be involved in modulating chain length (Deligny et al., 2016). Both NDST1 and 2 are found in the same cells and may compete to sulphate the chain. They also N-sulphate in opposing directions, NDST1 from non-reducing end to protein and NDST2 from protein to

non-reducing end, which may explain why NDST2 is involved in elongation, and why the two compete (Deligny et al., 2016). EXT2 may also complex with NDST1, increasing its activity along HS chains, meaning that the relative concentration of EXT1, EXT2 and NDST1 could influence the sulphation pattern of HS (Presto et al., 2008).

The chain undergoes C5 epimerisation from GlcA to IdoA through action a Glc C5 epimerase (Glce). Glce binds specifically to a GlcA flanked by two GlcNS residues, and its activity is diminished by 2- and 6-OS, as they attenuate contact of C5 with the three reactive tyrosine residues T^{468}, T^{528} and T^{546} in the active site (Debarnot et al., 2019; Qin et al., 2015). The chain can then undergo O-sulphation, first at position 2 by action of a 2-O-sulphotransferase (2OST). 2OST also requires NS at the two flanking disaccharides, and can sulphate either an IdoA or a GlcA (albeit with lower activity (Rong et al., 2001)) and similarly to Glce, 2OST activity is also diminished by 6S (C. Liu et al., 2014). 2OST and Glce have also been shown to interact with each other and form a complex, with 2OST increasing Glce activity, and subsequently 2-O-sulphating the same acceptor and potentially also preventing re-epimerisation back to GlcA (Qin et al., 2015). 6-O-sulphotransferase (6OST) has 3 isoforms, which sulphate GlcNAc (and GlcNAc derivatives) to GlcNAc6S. None of the OSTs show particularly different substrate specificities and have been observed to sulphate all targets, but 6OST1 tends to preferentially sulphate IdoA-GlcNS, 6OST2 tends to preferentially sulphate 2S containing regions and 6OST3 has an intermediate specificity (Annaval et al., 2020).

The maturing chain may be further modified by 3-O-sulphation by action of one of seven 3OST isoforms. 3S is a very rare modification, encompassing around 0.5% of the final sulphation of the molecule and may be involved in the creation of more specific binding motifs (Gulberti et al., 2020). The seven 3OSTs, like the 6OSTs, exhibit preferential targeting, for instance 3OST-1 targets GlcN units with a uronic acid with no 2S at the reducing end, 3OST-2:4 and 6 require 2S and 3OST-5 sulphates GlcN irrespective of 2-O-sulphation (Annaval et al., 2020). 3OST-1 and 5 are involved in the creation 3S disaccharides which ultimately bind to AT, while 3OST 2:4 and 6s action has been linked to the facilitation of binding to *Herpes simplex* viral envelope protein but the actual physiological roles are not understood (Annaval et al., 2020). HS synthesis is usually thought of as happening in steps, in part due to the fairly strict substrate requirements of each enzyme but it is currently unclear what the true order is. The order is assumed to be based upon the enzyme and their substrate concentrations, whereby increased PAPs, activated sugars or even different enzyme isoform concentrations can result in differing chain lengths, sulphate levels and patterns, but the real underpinning reasons remain elusive (Prydz, 2015).

1.16 Heparan sulphate and Heparin Structure and Function

HS and heparin have closely related structures, albeit HS possessing a more complex and information rich one. HS and heparin consist of the same disaccharide repeats of either GlcA, which may be epimerised to IdoA, both of which may be 2-O-sulphated and GlcNAc, which may be 6-O-sulphated, N-sulphated or acetylated and more rarely 3-O-sulphated. HS consists mostly of β(1,4)-GlcNAc-β(1,4)-GlcA-, but other, more sulphated residues exist throughout the molecule (Rabenstein, 2002a) (**Fig. 1.6**). The varyingly sulphated disaccharides group together throughout the structure, creating a domain system, whereby unsulphated residues of β(1,4)-GlcNAc-β(1,4)-GlcA- comprise an NA-domain and a highly sulphated NS-domain containing IdoA-β(1,4)-GlcNS-. Between the NS and NA domains exists an NA/NS transition domain, containing both GlcNAc and GlcNS (Rabenstein, 2002a). HS has a DoS of 0.6-1.5 (Dreyfuss et al., 2009), while heparin has a DoS of 2.2 to 2.8 sulphates per disaccharide (Devlin et al., 2019b).
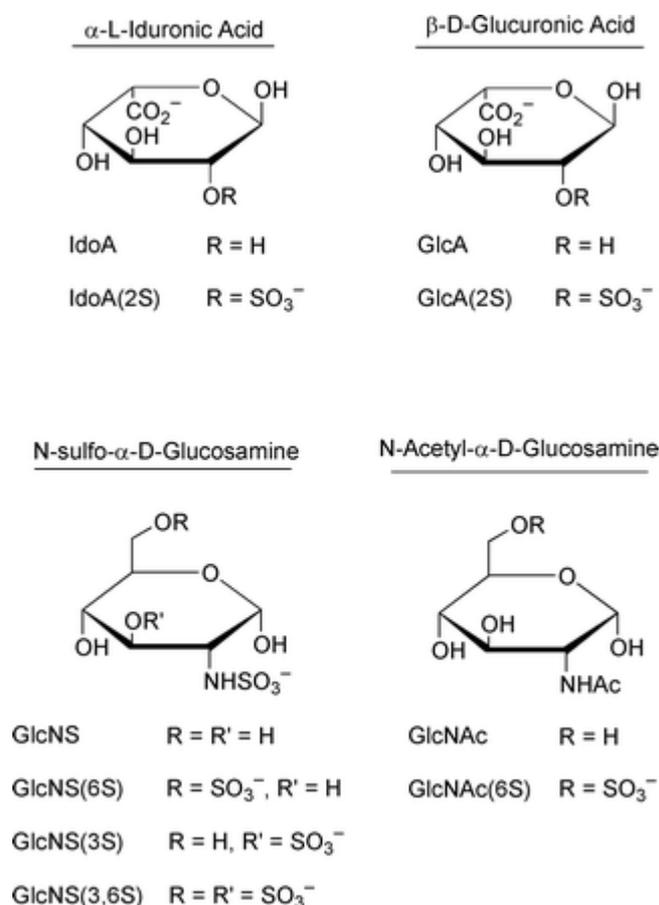
Figure 1.6: Different saccharides that comprise heparin and HS chains. Figure reproduced from (Rabenstein, 2002a).

The NS and the NA/NS domains afford HS with regions of high anionic charge which may bind to cations, including amines and proteins rich in lysine, arginine and sometimes histidine (Dreyfuss et al., 2009). The IdoA residues found in the NS region also facilitate great flexibility into the NS domain, allowing it to wrap more tightly around its target, increasing the binding

affinity (Mulloy and Forster, 2000). The NA domain has not been shown to be involved in protein binding, but may act as a flexible linker between NS domains, allowing the NS domain a degree of freedom to bind to its ligand, and also allowing multiple NS domains to interact with multiple ligands (Gallagher, 2001). The NS domains in HS allows binding to numerous cytokines, chemokines, growth factors and morphogens, both protecting the ligands from attack, and facilitating the formation of concentration gradients which are involved in the relative actions of the bound ligands (Sarrazin et al., 2011). Selective degradation of HS by heparanases can also release the sequestered molecules when needed (Sarrazin et al., 2011). HSPGs are also found in the ECM and basement membranes, where they have structural roles, mediate cell-cell and cell-membrane interactions and also endocytosis (Christianson and Belting, 2014; Sarrazin et al., 2011)

HS binding can exert three effects on ligands: 1) induction of a conformational change to the target, mediating activity: 2) facilitation of restricted diffusion of ligands whereby loose binding or disassociation/reassociation of the ligand controls ligand movement from high to low concentration across a defined path, either across the HS chain or to other nearby chains and: 3) stabilisation of more than one ligand into a complex (Sarrazin et al., 2011; Yan and Lin, 2009) The most studied heparin/HS protein interaction is that of heparin/HS with antithrombin (AT), an important protein involved in coagulation. The glycan binds to an allosteric site on AT, resulting in a conformational change that increases affinity for the active site of thrombin, blocking it and attenuating activity (Olson and Björk, 1994). HS/heparin binding also has the dual aspect of binding to the targets of AT and then presenting them to AT (Jin et al., 1997). Another example of protein:HS interaction is that of HS is and fibroblast growth factor (FGF) signalling. FGFs are a family of 23 signalling molecules some of which bind to FGF-receptors (FGFR), of which 5 exist (Ornitz and Itoh, 2001). HS acts as a skeleton, binding both the FGF isoform and the FGFR isoform into a ternary complex, facilitating subsequent auto-phosphorylation of FGFR results in signal transduction (Guimond and Turnbull, 1999). Both the FGF isoforms and FGFR isoforms exhibit unique structural requirements for HS binding, for example 2-O-S is integral the binding of FGF-2 (an isoform of FGF) to FGFR2c (an isoform of FGFR2c) (Guimond et al., 1993). Specific combinations of disaccharides have been shown to bind to various proteins however binding motifs on both the HS and the ligand are not necessarily unique, and numerous other motifs that facilitate the same activity may also exist (Meneghetti et al., 2015; Rudd et al., 2010; Sarrazin et al., 2011).

Heparin is structurally similar to HS but highly sulphated, comprising of 70-80% Ido2S-β(1,4)-GlcN,6S and 20-30% other with a DoS of 2.2-2.8 [(Devlin et al., 2019b; Dreyfuss et al., 2009). Heparin is almost exclusively found in the secretory vesicles of mast cells, attached to a serglycin PG (Korpetinou et al., 2014). The high anionic charge of heparin due to high levels of sulphate groups is thought to allow binding with cationic mast cell proteases and amines

(such as histamine), sequestering them in high concentrations in the secretory granule (Humphries et al., 1999; Wernersson and Pejler, 2014). The physiological role of heparin is still not fully understood (Devlin et al., 2019b), however upon secretory granule release, it is shed from the PG and has been linked to an increase in vascular permeability, promoting the immune inflammatory response (Oschatz et al., 2011). Shed heparin may also be transported to the lymph nodes, taking with it bound tumour necrosis factor and other proteins, aiding in communication between the peripheral and the secondary lymph tissues (Kunder et al., 2009).

1.17 Pharmaceutical Heparin

In 1918, heparin was identified as a potent anti-coagulant due to its ability to bind to and activate antithrombin (AT) (Howell and Holt, 1918), thus modulating coagulation, and resulting in the discovery of one of the most important pharmaceuticals in history. Heparin is the second most used drug by weight after insulin and is classified as an essential medicine by the World Health Organisation (WHO, 2022). Heparin is primarily used to treat atrial fibrillation, deep-vein thrombosis and pulmonary embolism, is important for some heart surgeries and is also utilised with medical devices including extracorporeal life support and blood dialysis systems, on indwelling catheters and sometimes during blood tests (Alquwaizani et al., 2013). The major side effects associated with heparin use are bleeding and heparin-induced thrombocytopenia (HIT) (Alquwaizani et al., 2013). HIT is caused by the formation of antibodies to heparin bound to platelet factor 4 (PF4), the antibody-heparin-PF4 complex can then bind to and activate platelets, resulting in the formation of micro clots which leads to thrombocytopenia (Warkentin, 2006).

Heparins may undergo depolymerisation using chemical or enzymatic means, creating low molecular weight heparins (LMWHs) which exhibit the same pharmaceutical properties as full length heparin (Alquwaizani et al., 2013; Ingle and Agarwal, 2014). LMWHs have reduced side effects, increased bioavailability and a longer half-life when compared with full length but do not respond to the antidote protamine as well (Alquwaizani et al., 2013; Mulloy et al., 2014a). There are multiple methods of depolymerisation and depending on the reagents and methods used, different structural motifs can be found at the ends of the molecule where cleavage occurs **(Table 1.1)**.

Table 1.1: The production methods of different LMWHs produced from porcine material and associated structural changes. Data were acquired from (Ingle and Agarwal, 2014; Yan et al., 2017).

Method	With	LMWH	Average MW	Non-reducing end	Reducing end
Deaminative cleavage	Nitrous acid	Dalteparin	4000-6000	IdoA2S	2,5-Anhydromannitol-6S
	Nitrous acid	Nadroparin	4500		
	Isoamyl nitrite	Certoparin	5400		
Oxidative depolymerisation	Cu+ and H2O2	Parnaparin	4500-5000	IdoA2S/GlcNS6S	GlcNS6S/IdoA2S
Alkaline depolymerisation	Performed on benzyl-ester of heparin	Enoxaparin	4500	4-enopyranose-uronate-2S/ enopyranose-uronate	GlcNS6S; characterized by the presence of 1,6-anhydro-gluco/ manno residues and mannosamine
Enzymatic depolymerisation	Heparinase	Tinzaparin	4900	4-enopyranose-uronate-2S	GlcNS6S

1.18 Coagulation

Haemostasis is the process of maintaining blood in a damaged vessel. Coagulation, which heparin inhibits, is one of the main steps involved in haemostasis, whereby the blood is thickened and creates clots which can seal holes in the blood vessels. After initial vessel damage, primary haemostasis occurs whereby a series of interactions between platelets and the site of damage activates them, forming an initial, activated platelet plug. This plug is strengthened, by fibrils of fibrin, which are in turn cross-linked by factor thirteen (FXIII) to form a strong and supportive mesh that keeps the now formed clot together (Palta et al., 2014).

Production of fibrin from its zymogen fibrinogen is essentially mediated entirely by the protein thrombin (also known as FIIa), which is directly responsible for the production of fibrin (FIa) from fibrinogen (FI) as well as forming part of a positive feedback loop that activates more thrombin from prothrombin (FII) and forms part of negative feedback loops that prevent further thrombin formation (Grover and Mackman, 2019; Palta et al., 2014). The coagulation pathways are described as a cascade, whereby a chain reaction leads to a great surplus of fibrin formation. There are two coagulation pathways: the intrinsic and extrinsic, both of which meet at FX, converting it to its active form FXa to form the common pathway, whereby thrombin is formed from prothrombin (Grover and Mackman, 2019; Palta et al., 2014). The

extrinsic pathway can be thought of as the initial start of a clot, whereby tissue factor (TF, FIII) is released from damaged cells and activates factor seven (FVII), which in turn activates FX. The intrinsic pathway serves as a means of amplification of coagulation and can be triggered in two ways: first through positive feedback from thrombin activated by the extrinsic pathway or second through activation of FXII, which occurs through contact with a charged surface – usually created from secretions of activated platelets, or through reciprocal activation from/of the contact system through bradykinin (BK) (Grover and Mackman, 2019). FXII activates FXI which in turn activates FIX and FVIII (with FVIIIa acting as a cofactor for FIXa) which then activates FX which, with the help of its cofactor, FV, activates thrombin (**Fig. 1.7**) (Grover and Mackman, 2019).

Antithrombin and plasmin are two mediators of negative autoregulation that the coagulation cascade exhibits, with plasmin breaking down FXIII and thus loosening fibrin fibrils, and AT which blocks the active sites of many of the coagulation factors, preventing them from activating more of the pathway, including its namesake, thrombin (Palta et al., 2014). AT is a serpin (serine protease inhibitor) that works by presenting an active serine, which, upon attack by the target serine protease, induces a conformational change that blocks the action of the attacker, creating an inactive, equimolar and stable serine-protease-serpin complex which can ultimately take days to dissociate (Olson and Björk, 1994) Heparin possess the ability to bind to an allosteric site on AT, inducing a conformational change that expels the active serine further, encouraging serine-protease binding (Olson and Björk, 1994). A distinct pentasaccharide sequence has been identified as the primary binding sequence for heparin to AT, GlcNAc6S-GlcA-GlcNAc(N,3S)-IdoA2S-GlcNAc2,6S, which is sometimes written as AGA*IA, with A* representing the rare 3S that is required for binding (**Fig. 1.7C**). It is important to note however, that while the 3OS of saccharide 3 is heavily implicated in AT binding, only it and three other sulphates (a 6-OS at saccharide 1, and an NS at saccharides 3 and 5) are required for AT binding, and while sulphation or lack thereof of the other positions may increase or decrease its binding capability, there is no "unique" pentasaccharide required for AT binding (Casu et al., 1981; Choay, 1989; Choay et al., 1983, 1981; Meneghetti et al., 2015). Furthermore, maximal anticoagulation occurs during the forming of a ternary complex with AT, thrombin and heparin. This requires an extended chain length, which can both allosterically activate AT, and bind to thrombin, bringing the two together (**Fig. 1.7B**) (Petitou et al., 1999).

(A)

(B)

(C)

R = SO₃⁻ or COCH₃

Minimal binding site for antithrombin
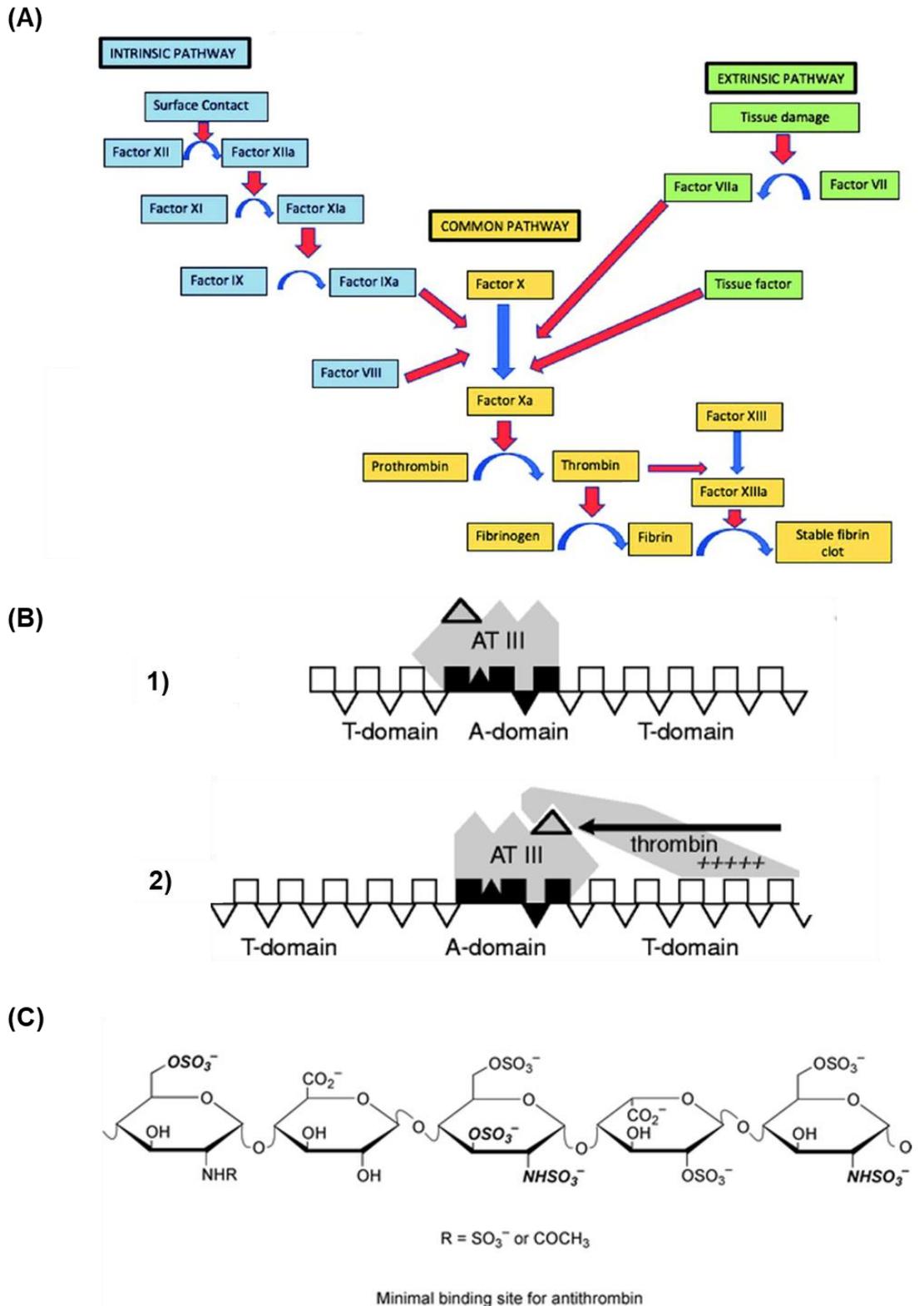
Figure 1.7: The coagulation cascade and the effect of heparin upon it. A) Schematic of the coagulation cascade, including the intrinsic (blue), extrinsic (green) and common pathways (yellow) (Dammann et al., 2020). **C)** The binding of heparin to AT and subsequent attraction of thrombin. *1)* Heparin binds to AT allosterically, altering the structure of AT, making it more likely to bind thrombin. *2)* The free residues of heparin bind to thrombin, guiding it to AT (Petitou et al., 1999). **C)** A pentasaccharide repeat found in heparin that binds to AT, potentiating the anticoagulant effect of heparin (Rabenstein, 2002a).

1.19 The extraction of glycosaminoglycans.

Glycosaminoglycan extraction generally requires 5 steps, 1) homogenisation of the starting material, 2) removal of unwanted macromolecules, 3) extraction of the GAGs from solution and any remaining small molecule impurities, 4) separation/removal of unwanted GAG species and 5) any extra clean-up (Guimond et al., 2009; Taylor et al., 2019). Unwanted macromolecules include bound proteins and lipids or DNA and may be removed by enzymatic digestion and solvent fractionation. After breakdown or removal of undesired macromolecules, the GAGs can be separated from the mixture by precipitation of the GAG using an alcohol (Charles and Scott, 1936). Step 4 proves to be one of the most difficult steps, as GAG structures can be very similar to each other, possessing different structures that have the same charge and mass. If a single GAG is desired, this step can be performed with an enzymatic digestion of the unwanted GAGs (Guimond et al., 2009; Zappe et al., 2021), but if the entire GAG profile is to be studied or GAGs are required at the large scales, the samples are often separated by SAX-HPLC (Nicola Volpi, 1994) .

Homogenisation is normally performed through use of blending, pulping and other means (Brito et al., 2018; Mycroft-West et al., 2021). Chemicals such as Trizol may also be used to homogenise samples on much smaller scales (on an organ by organ in animal basis) directly (Guimond et al., 2009), and guanidium hydrochloride (Chandrasekhar et al., 1987) may also be used to lyse cells and denature proteins that may bind to or degrade the extractant. Large extractions often involve solvents such as di and tri-chloromethane (chloroform), acetone or alcohols to remove or reduce lipid and other unwanted molecular content, both to reduce the working volume dramatically and to resist microbial growth during extraction (Brito et al., 2018; Freeman et al., 1957; Guimond et al., 2009; LaRiviere et al., 2021; Mycroft-West et al., 2021; N. Volpi, 1996). DNA, RNA and proteins are normally removed through enzymatic digestion or solvent fractionation (Hitchcock et al., 2007; Zappe et al., 2021), and sometimes the enzyme/protein remnants will be removed from the extract by precipitation with trichloroacetic acid (TCA) which may also be used to extract whole PGs prior to any protease steps (Aikawa et al., 1985). The GAG itself may also be enzymatically cleaved from its conjugates and sometimes more aggressive means may be required to free the GAGs, including chemical hydrazinolysis or beta-elimination through action of a strong base such as NaOH (Hitchcock et al., 2007). These steps can be tailored to the sample that is undergoing extraction, particularly lean animals or tissues may not need a defatting step for example.

Extracting the GAGs from solution can be done in a number of ways and normally revolve around selective solubilisation or precipitation of the target compound. Sometimes extractors opt to wash the extractant with a high (3M) electrolyte wash to remove any bound fragments such as proteins or undesirable cationic salts and precipitate the GAGs in a solvent, such as chloroform (normally at an acidic pH to drive GAG dissolution) (Gemst et al., 2016). Others

may perform a strategy in the opposing manner by placing an ion exchange resin into the sample, binding to the highly anionic GAGs and then removing and washing the resin with a high electrolyte concentration to elute the GAGs (Brito et al., 2014; Taylor et al., 2019). Some may perform a simple precipitation with an alcohol (Mycroft-West et al., 2021), or with quaternary salts such as cetylpyridinium chloride (Bigler and Brenneisen, 2009).

Further fractionation of the extracted GAGs may be performed with solvent washes, through stepwise elutions of increasing solvent concentration, with solvents such as acetone or ethanol (N. Volpi, 1994). Extractants may also be separated further or solely with chromatography particularly by anion-exchange (Guimond et al., 1993; Mycroft-West et al., 2021), affinity (Brito et al., 2014) and size exclusion (Iozzo et al., 1990). CsCl density gradient centrifugation has also been used in place of size exclusion and may also be used during the extraction steps (Razin et al., 1982).

Final clean-up steps may be required due to high electrolyte contamination from fractionation or extraction steps. Usually, these steps revolve around removal of small or insoluble particulates that have not been removed in previous steps. Filtration may be used to remove insoluble matter and desalting steps may be undertaken, such as dialysis to an applicable Da cut off (usually 1-8 kDa) or desalting with a size exclusion column (Mycroft-West et al., 2021; Taylor et al., 2019). Further steps can include cation exchange, often to ensure that the GAGs are in the generally accepted standard Na^+ or Ca^{2+} form, and bleaching to remove discoloration and some organic contaminants – this is particularly prevalent in pharmaceutical preparations (Al-Hakim, 2021; Fu et al., 2016; Taylor et al., 2019). Care must also be taken however, as some methods such as guanidine hydrochloride solubilisation may alter the structure of the extracted GAG, however if the aim is only to assess quantities or display the mere presence of GAGs, denaturation of the structure may not be an issue. Chemical excision of GAGs can also depolymerise them and high/low pHs, high temperatures (introduced during homogenisation or volume reduction) or both may also cause desulphation, deacetylation, epoxide formation and depolymerisation (Beccati et al., 2010; Guimond et al., 2009; E A Yates et al., 1996).

1.20 Analysis of GAGs.

Structural analysis of GAGs involves separation of GAG species or fragments thereof and subsequent detection or analysis of the structures present. Separation relies on the underlying structures of the compounds possessing different properties which may be exploited, allowing differentiation of them. Separation is generally crucial to carbohydrate analyses, as many structural features are present in all carbohydrates and it is generally their relative positions and amounts that confer structural complexity. Furthermore, features such as charge or mass can be equivalent for different disaccharide structures, making separation difficult (Powell et

al., 2004; Skidmore et al., 2010). Detection and identification revolve around the elucidation of different features of a molecule, often in the form of bond types, interactions and the presence and location of functional groups, all of which can be determined through spectroscopic or chemical means.

1.20.1 Solvent and ionic fractionation

The simplest separatory techniques for GAGs comprise ion pairing and subsequent separation. Different GAGs can pair with different ions, which can modulate solubility in different solvents and facilitate sulphation. Heparin can be targeted solely with potassium acetate (Toledo and Dietrich, 1977) and copper in alkaline solution can precipitate DS from a mixture of GAGs (Cifonelli et al., 1958). Sequential precipitation can also be used, whereby increasing amounts of solvents such as acetone and ethanol are added, separating GAGs that dissolve more readily into them with each step (Meyer et al., 1956; N. Volpi, 1994). Soluble barium has been shown to complex with highly sulphated regions of heparins (Bianchini et al., 1985). As an increase in chain length will lead to more sulphation and therefore more complexation, different Ba-salts of heparin can be extracted with varying temperatures, which select for highly sulphated and large MW (25 degrees), highly sulphated and medium MW (5 degrees) and lowly sulphated and low MW (soluble at 5 degrees) chains, each possessing differing activities (Bianchini et al., 1985; Viskov et al., 2009). Separations may also be performed enzymatically, whereby GAG lyases that target unwanted GAGs can be used to degrade and therefore remove them (Zappe et al., 2021).

1.20.2 The detection of GAGs.

It is difficult to detect GAGs due to a lack of natural chromophores. Ionic dyes such as azure A, toluidine blue, Alcian blue (Nicola Volpi, 1996), O-safranin (Sunwoo et al., 1998) and Stains-All (Volpi and Maccari, 2002) have been used, but they can be disassociated under separatory conditions and may not be specific to the analytes . GAGs absorb weakly at 200 nm, due to the presence of carboxylates in acetate groups and due to a small percentage of residues present in their acyclic form in solution at the reducing end (Skidmore et al., 2010). Borate buffers can be used, where boron forms a reversible complex with the acyclic site, to increase the relative amount of acyclic reducing ends and therefore the absorption by up to 20fold (Colon et al., 1993; Roy et al., 1984). Chemical and enzymatic depolymerisation results in unsaturated residues, which absorb strongly at 232nm, detection of approximately 1 nmol of derivatised material is possible, but molecules with larger chain lengths and therefore less reducing ends will need more material for detection (Skidmore et al., 2010). To overcome this, fluorescent labelling is also utilised, whereby the target molecule is covalently bound to a fluorophore. Primarily, GAGs are derivatised through reductive amination at the reducing end with a fluorophore such as 2-aminopyridine (AP) or 2-aminoacridone (AMAC) which greatly increase the sensitivity of detection to 3 pmol in ion exchange chromatography and 10 pmol

in gel electrophoresis for AP and AMAC respectively (Calabro et al., 2000; Plaas et al., 1996).The fluorescent tag BODIPY hydrazide has also been utilised and detection in the fmol range has been reported (Skidmore et al., 2010). Some GAGs may also be detected with antibodies, and antibodies targeted towards CSA, CSC, DS, HS and KS are available but are usually used for immunohistochemical staining in microscopy (Georgakopoulos et al., 2019; Shibutani et al., 1990).

Derivatization of other functional groups is also possible and functionalisation of the carboxylic acid and amine moieties with fluorinated groups has been achieved, allowing monitoring of AT-heparin interactions with NMR and microscopic visualisation of the labelled GAGs (Lima et al., 2017). Nitrous acid depolymerisation of heparins leaves behind a distinct anhydromanose residue, which can be functionalised with a tyramine moiety, which may then be fluorescently labelled with a fluorescein molecule or radio labelled with I[125] (Malsch et al., 1994). Aside from covalent radiolabelling with I[125], GAGs may also be radiolabelled with S[35], during biosynthesis, whereby the cells are fed radioactive substrates (Eiber and Danishefky, 1957).

1.20.3 Electrophoretic separation

The most accessible separatory techniques comprise electrophoresis with polyacrylamide and agarose gels or cellulose acetate strips (Nicola Volpi, 1996). Electrophoresis works by subjecting a charged molecule to an electric field of sufficient strength to move the molecule through a porous material. The distance the molecule has moved is a function of its charge, size, shape, strength of the electric field and the size of the pores in the medium it traverses. The anionic nature of GAGs due to sulphates, carboxylates or both, makes them susceptible both to movement through an electric field and subsequent staining with cationic dyes, such as azure A, toluidine blue and Alcian blue, the sensitivities of which may be expanded with silver staining (LaRiviere et al., 2021; Nicola Volpi, 1996). GAGs may also be detected with antibodies (Bartold, 1990) or with fluorescently labelled derivatives (Karousou et al., 2015, 2004; Volpi and Maccari, 2006). Gel formats are generally used instead of paper ones, due to the ability of the analyst to fine tune pore size as a function of polymer percentage, while cellulose acetate strip pore size is only tuneable at manufacture.

Acrylamide and agarose gels with standard nucleotide separation buffers, such as tris-acetate-EDTA (TAE) and Tris-borate-EDTA (TBE) can be used to separate GAGs by MW as a function of the pore size of the gel. TBE acrylamide gels have recently been able to separate samples as small as 6-mer oligosaccharides using a 22% acrylamide gel and also to separate HA samples ranging from 495 to 4 kDa with a 4-20% acrylamide gradient gel (Bhilocha et al., 2011) while TAE and TBE agarose gels have both been used to separate much larger HAs ranging from 1510 to 30 kDa and 6100 to 30 kDa (Bhilocha et al., 2011; Cowman et al., 2011).

Heparins ranging from 11.6 to 1.6 kDa have also been separated on agarose gels (Volpi and Maccari, 2006). Densitometric analysis can be applied to the gels and can be used to determine average MW of unknown samples, offering a cheap, facile and rapid method of size estimation (Bhilocha et al., 2011; Cowman et al., 2011; Edens et al., 1992; Mourão et al., 1979). GAG amounts as low as 0.1µg have been detected on agarose and acrylamide gels through use of subsequent silver staining, but generally 1-5µg is easily visualised with standard dyes (Volpi and Maccari, 2006).

While size and overall charge are important factors to be able to separate by, important structures in GAGs are defined by charge densities and the patterns therein. Electrophoretic separation can therefore be enhanced through the use of different buffer systems which interact specifically with the GAGs. Ba acetate may be added to the running buffer to separate GAGs, with a system that is essentially a Ba-fractionation during an electrophoretic migration, whereby the samples movement is expressed as a function of the complexation of Ba^{2+} (i.e molecular weight and sulphate density) with the sample (Funderburgh and Chandler, 1978; Oreste and Torri, 1980; Viskov et al., 2009). Cationic solvents, such as diaminopropane (DAP) can also be used, whereby separation is achieved by the interactions of the GAGs with DAP (Dietrich et al., 1977). DAP has also been coupled with Ba for marginally increased separation, particularly of heparin sub-populations (Dietrich et al., 1985; Volpi and Maccari, 2002).

Capillary electrophoresis (CE) is another form of electrophoresis, whereby the analyte is electrophoretically moved through a fused silica capillary rather than a gel pore. The silica itself is ionised by the buffer running through it, and a double layer of cations forms on the capillary that facilitates flow of the liquid, dubbed electroosmotic flow (EOF), through the capillary from anode to cathode. EOF allows analytes with opposing or low charges to move through the capillary, albeit at significantly reduced rates, facilitating separation. The polarity may also be switched through use of different buffering systems and capillary coatings which can afford better separations of some GAG species due to their anionic nature (Volpi et al., 2008). The samples inside the capillary is separated based on charge, hydrodynamic volume, pH and any interactions with the buffer (Mantovani et al., 2018) CE offers inline detection of the sample, through UV or derivatisation, which is much more reliable than the densitometric readings required to quantify a gel. CE also offers a lower limit of detection than gel electrophoresis, with amounts as low as 0.03 µg detected with UV absorbance at 200nm (Volpi and Maccari, 2006) and increased band resolution, separating 4- to 150mers without the smearing that is often observed on gels (Mao et al., 2002) CE has also been used for disaccharide analyses - one of the most used approaches for GAG characterisation, whereby the levels of constituent disaccharides within the GAG are quantified. Differences between disaccharide structures affect the relative pK_a values for the carboxylate and amine functional

groups and hence grant individual disaccharides with similar properties (overall charge, MW etc) unique migrations (Eldridge et al., 2009).

1.20.4 Chromatographic separation

Chromatography moves analytes in a mobile phased other a stationary phase and separates them based on their interaction with the stationary phase. The simplest chromatography used with GAG analysis is size exclusion chromatography (SEC). SEC closely resembles the principal of size-separatory gel electrophoresis whereby the analyte is passed through a column that is packed tightly with polymer beads that may or may not contain pores. Smaller molecules will be able to pass through more regions and hence experience a larger volume to travel through, while larger molecules will experience the opposite. The analyte is moved through the matrix by flow. Molecules of different sizes will elute at different volumes as a function of the volume that they experience in the column (**Fig.1.8A**). SEC is primarily used to differentiate GAGs, oligosaccharides of various lengths, to remove differently sized contaminants such as electrolytes, or to calculate MW (Guimond and Turnbull, 1999; Mulloy et al., 2014a; Staples and Zaia, 2011; Ziegler and Zaia, 2006).

SEC is an inherently low-resolution technique and is most often used during preparatory work to pool oligosaccharides of varying lengths for later analyses. For GAGs, an acrylamide matrix such as P6/P10 is normally used for separation (Bisio et al., 2009; Mulloy et al., 1997). SEC has separated 2- to 12mer fractions (Bisio et al., 2009; Langeslay et al., 2012b) with a P10 matrix and larger oligosaccharides up to 24mers can also be separated with longer columns (Goger et al., 2002). The oligosaccharides may be further separated afterwards to select for isomeric variants within the fraction as in (Guimond and Turnbull, 1999; Turnbull et al., 1999) where pooled decasaccharides were separated into 27 distinct species using anion exchange chromatography. SEC can also be used to remove unwanted molecules introduced during sample extraction such as salts and small organic molecules and is often achieved with a Sephadex matrix such as G25 or G50 usually due to cost, but acrylamide matrices can also be used (Guimond and Turnbull, 1999; Palhares et al., 2019). Sephadex does not offer the separation observed with acrylamide matrices but can be implemented more easily. Analytes flowed past SEC can be detected with standard UV or fluorescent detection of appropriately labelled saccharides, but can also be measured with conductometers, refractive index (RI) detectors, viscometers or light scattering (LS) detectors, the latter three of which may be combined, to form a "triple detector" which, through combination of RI and LS measurements can determine the average MW of a sample and provide information on the spread of MW (Bertini et al., 2005a).

Ion-exchange chromatography is a technique which separates analytes based on the relative affinities of the analyte for the stationary phase. Analytes will elute based on competition with

counter ions in the mobile phase (**Fig. 1.8D**). Strong anion-exchange (SAX) chromatography is primarily used for GAG separations whereby analytes with the smallest charge/mass ratio will elute first, based on competition with an inorganic ion, most commonly, a quaternary ammonium group in the stationary phase (Fasciano and Danielson, 2016). SAX is vital for disaccharide analysis and oligosaccharide differentiation however it is also used for intact GAG separation and is used to quality control pharmaceutical heparin by separating heparin from other GAGs that may be present (Fasciano and Danielson, 2016; Mauri et al., 2017b; Skidmore et al., 2010, 2010). Weak anion-exchange (WAX) chromatography is less common than SAX and follows essentially the same methodology, albeit with the use of a mobile phase that has a more specific working pH range, alteration of which can be used to alter the selectivity in binding (Fasciano and Danielson, 2016). WAX chromatography has been used to separate intact GAGs (Huang et al., 1995).

Reversed phase ion pairing (RPIP) chromatography is similar to ion exchange, but the stationary phase is highly hydrophobic and hence more polar compounds are eluted first. A somewhat non-polar mobile phase, usually acetonitrile or an alcohol is flown over the stationary phase at increasing concentration, until hydrophobic enough to elute the analyte (**Fig. 1.8B**). As GAGs are highly polar molecules, an ion pairing agent – usually a polyamine is added to the mobile phase, which binds to the GAG, neutralising its polarity. Optimisation of the concentration of the ion-pairing agent can also be used to adjust the retention of the analytes, facilitating better separation (**Fig. 1.8C**). Some of these ion-pairing agents can create a fluorescent complex, which removes the need for pre-column derivatisation, for example, GAGs fully hydrolysed to GlcN and GalN, can be coupled with o-phthaldialdehyde and 3-mercaptopropionic acid to form a fluorescent isoindole, and allows GlcN/GalN and therefore different GAG species identification and quantification (Studelska et al., 2006). RPIP-HPLC has been shown to allow rapid disaccharide analysis of heparin and HS, with run times of ~15minutes (Toyoda et al., 1999), and with post column fluorescent derivatisation, has been used to perform disaccharide analysis of both CS and heparin/HS, at nanogram amounts of starting material quantities (Sinnis et al., 2007). RPIP has also differentiated some LMWH species (Patel et al., 2009, p.).

Hydrophilic interaction chromatography (HILIC) is similar to RPIP, using a similar set up but in reverse, whereby the analyte interacts with the stationary phase in a hydrophobic mobile phase and is eluted primarily in water. HILIC also separates by polarity without the use of inorganic salts for elution (Pepi et al., 2021; Zappe et al., 2021) (**Fig. 1.8E**). The lack of inorganic salts and use of volatile solvents makes HILIC (and RPIP) chromatography suitable for separations with in-line mass spectrometry (MS) detection, however volatile salts can be applied to other HPLC methods which facilitate hyphenation to MS (Pepi et al., 2021). HILIC coupled with MS has been used for HS (Antia et al., 2018), CS (Tóth et al., 2020) and DS (Gill

et al., 2013) disaccharide analysis and LMWH (Li et al., 2012) and HS (Wu et al., 2019) oligosaccharide analysis.

Affinity chromatography is a type of chromatography that exploits the affinity of one molecule for a ligand. Affinity chromatography is achieved by derivatising a ligand in the stationary phase. Bound molecules are subsequently eluted with increasing concentrations of a competitor for the bound ligand, usually increasing concentrations of NaCl. Molecules that bind more strongly to the ligand require higher NaCl concentrations to be eluted (Powell et al., 2004) (**Fig. 1.8F**). Affinity chromatography is used to purify GAGs based on their affinity for a ligand and is how the pentasaccharide sequence responsible for AT binding in heparin was first discovered, whereby fragments that were undigestible with lyase enzymes were found to bind with high affinity to an AT derivatised column (Andersson et al., 1976; Choay, 1989). Affinity chromatography has been used to identify heparin and HS fragments that bind to FGF-1 and -2 (Kreuger et al., 2001) and FGFR4 (Loo et al., 2001), show that HS and DS can bind to hepatocyte growth factor/scattering factor (Lyon et al., 1998) and extract a HS from shrimp with binding affinity for AT but negligible anticoagulant activity (Chavante et al., 2014).

1.20.5 Mass Spectrometry

MS is yet another separatory technique, but for fragments that make up the sample, whereby the analytes are separated by their mass to absolute charge ratio (m/z). A sample is vaporised and ionised, and the gaseous ions are accelerated through a magnet. The magnetic field deflects the ions as a function of the m/z ratio. The abundance of each ion at each m/z is recorded. Ionisation of the analytes often results in fragmentation and various ions of various sizes are formed, which in turn can be analysed, allowing the analyst to develop an image of the various functional groups and fragments that make up the entire molecule. Some of the fragments can be subjected to further MS separation, as in tandem-MS (MS/MS), where ions of interest are entered into a second, tandem MS, fragmented further (through a host of further ionisation methods) and separated again, allowing further information to be gleaned about the parent ion, including differentiating between different structural isomers (Pepi et al., 2021; Staples and Zaia, 2011).
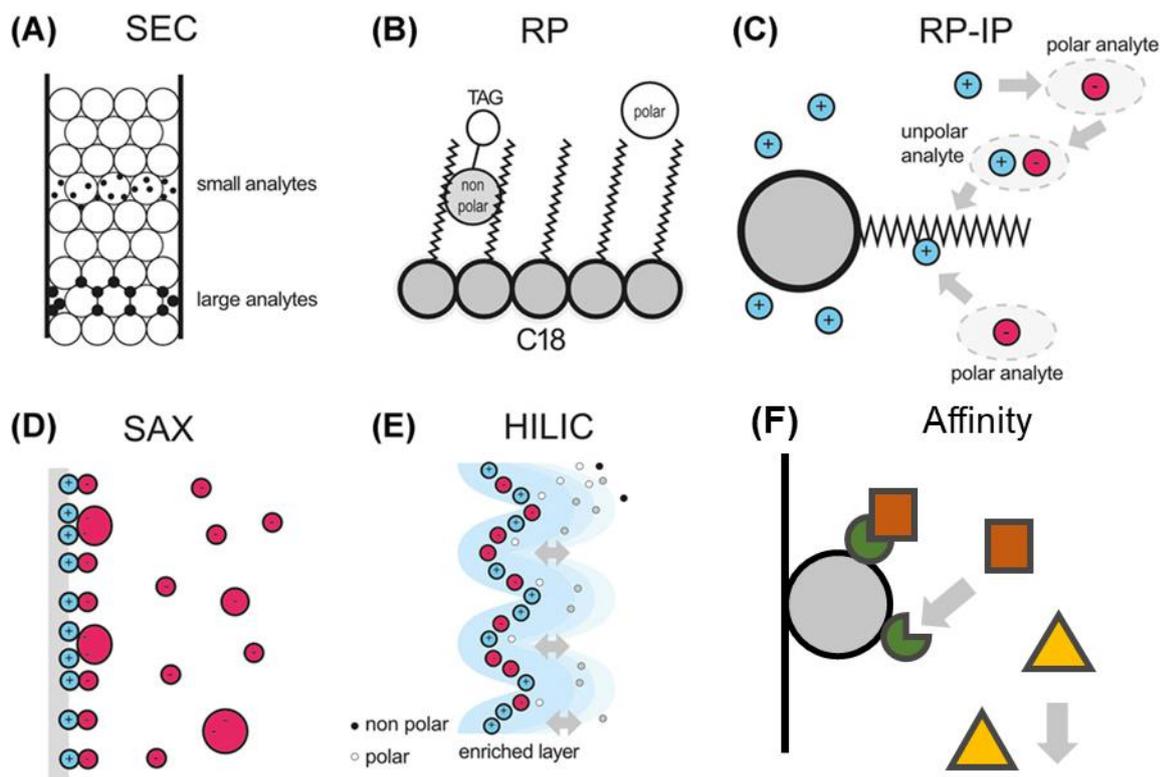
Figure 1.8: Different chromatography modes. Figure reproduced from (Zappe et al., 2021).

GAG MS comes with many difficulties as they are polar, non-volatile and thermally labile. Ionisation of whole or depolymerised saccharides can lead to liberation of sulphate groups and hence incorrect analysis of their positions and the large number of ionisable sites are subject to attack from alkali metal ions (Pepi et al., 2021; Zaia, 2005; Zappe et al., 2021). For GAGs, negative ion mode is used whereby deprotonated molecules are measured. Initially, fast atom bombardment (FAB) was used to ionise GAGs, whereby a high energy atom (usually Xe or Ar) is fired at the sample, fragmenting and ionising it (Yang et al., 2011). Using FAB, heparin oligosaccharides (Mallis et al., 1989) and tetrasaccharides (Khoo et al., 1993) could be identified. Currently, softer ionisation techniques are used, including electrospray ionisation (ESI) and matrix-assisted laser desorption ionisation (MALDI) (Pepi et al., 2021). ESI involves applying a high voltage to a liquid, creating an ionised aerosol of the analyte. Analytes ionised by ESI are in the liquid phase, and hence ESI-MS can be readily hyphenated to various HPLC techniques. In MALDI, the analyte is mixed with a matrix and irradiated irradiates by a laser, causing ablation and subsequence ionisation. Both techniques have been used to perform disaccharide analysis, define molecular mass of oligosaccharides and profile GAGs in different tissues (Hitchcock et al., 2007), due in part to the incredibly small amount of starting material required for detection (Pepi et al., 2021). Other ionisation techniques may be used with GAGs and are reviewed in (Pepi et al., 2021; Zappe et al., 2021). SEC, SAX, RPIP and HILIC have all been hyphenated to MS and used to separate complex mixtures of GAGs and LMWHs (Pepi et al., 2021; Wang and Chi, 2018; Zappe et al., 2021). The power of MS analysis

lies in the ability to partially sequence oligosaccharide fragments. m/z profiles for 36 distinct oligosaccharides, ranging from di- to decasaccharides have been published, allowing generation and characterisation of a HS microarray for structure/function studies (Miller et al., 2020)

1.20.6 Identification of GAG structures

Identification of GAG structures can be performed both before or after separation and primarily revolves around the detection of specific functional groups or bonds through either chemical or spectroscopic means. The simplest techniques involve the introduction of cationic dyes, which bind to the highly charged GAGs, eliciting a colour change that can be recorded and quantified. Dimethylmethylene-blue (DMBB) (Dey et al., 1992), Alcian blue (Tuckett and Morriss-Kay, 1988) and Azure A (Klein et al., 1982) are commonly used and can provide the analyst with details on the total quantity of anionic substrate. As the dyes primarily bind to the anionically charged sulphates, they can also be used as a read out for sulphate levels in known GAGs, monitoring derivatisation reactions or overall conformations (Torode et al., 2015).

More specific chemical methods have been developed and most widely used method is the carbazole assay for quantification of uronic acid levels. In the carbazole reaction, uronic acids, which have been liberated by acid hydrolysis of the compound, are reacted with carbazole creating a coloured complex that may be quantified at 550 nm, serving as a readout of total GAG content (Bitter and Muir, 1962). Aside from dye-binding assays, sulphate content assays also exist, whereby the GAG is hydrolysed entirely in acid, and the liberated sulphate groups are reacted with barium chloride and the subsequent solid barium sulphate produced is quantified through gravimetric or turbidimetric means (Dodgson and Price, 1962). The acid-lysed GAGs may also undergo elemental analysis with inductively coupled plasma -mass spectrometry or -atomic emission spectroscopy (ICP-MS, ICP-AES) whereby the individual elements are detected by their m/z ratio or by their characteristic absorption bands respectively (Boyraz et al., 2022; Mittelstaedt et al., 2016) or the entire GAG could undergo elemental analysis with CHN(O)S analysis, whereby the sample is burned at ~1600°C and the released gasses are quantified with gas chromatography. Methods to monitor the positions at sulphates are found also exist, primarily for NS, whereby free amine, produced from the liberation of sulphate (and not acetate) from the amine under acidic conditions can be marked with indole, and quantified at 490 nm (Lagunoff and Warren, 1962). The indole reaction can also be used to detect aminosugars (Dische and Borenfreund, 1950) These readouts can be combined to form a measure of degree of sulphation, by comparing uronate levels with sulphate levels (Lee et al., 1973). In modern times, these assays can be performed in a 96 well plate format, facilitating high throughput screening (Cesaretti et al., 2003; Frazier et al., 2008).

1.20.7 Spectroscopic Methods of identification

While chemical methods have their place for GAG structural studies, they are cumbersome, destroy the assayed sample and require high amounts of starting material. Furthermore, the total level of different moieties is not necessarily as important for GAG activity as their positions on the carbon ring or dispersal across the chain. Chemical methods do not provide this information, but instead provide an average across the sample. More refined, higher resolution methods are required and for this, spectroscopy is often used.
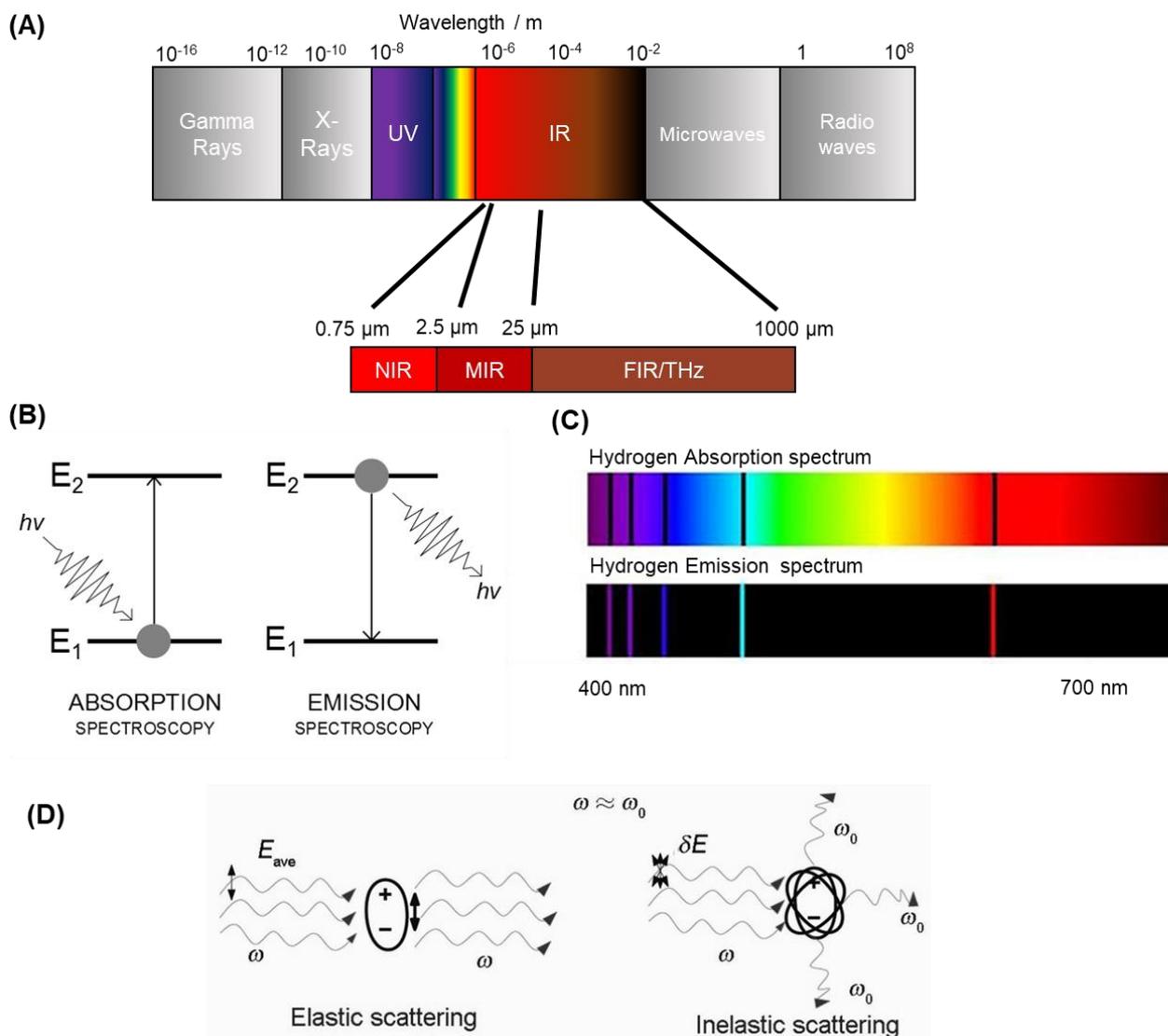
Figure 1.9: The electromagnetic spectrum and schematics for absorption, emission, elastic scattering and inelastic scattering spectroscopy. A) Electromagentic spectrum with zoom in of the IR region. **B)** Schematics for absorption and emission spectroscopy. **C)** Examples of visible light absorbance and emission spectra for hydrogen. **D)** Schematics for elastic and inelastic scattering. Where ω is the energy of the incident and elastically scattered light and ω₀ is the energy of inelastically scattered light. Images recreated from (Gauglitz and Vo-Dinh, 2006).

1.20.8 Electromagnetic spectroscopy

Spectroscopy is the study of the interaction between matter and electromagnetic radiation, or perhaps more colloquially, the study of the colours found in a compound. All structural

features, including the interactions between molecules, the functional groups present within each molecule, the bonds between them and even the elements that the molecule consists of are made of quantised energies which can be probed by an analyst. In theory, the entire electromagnetic spectrum can be used, but more specifically, set portions of it are studied in relation to GAGs, specifically the radio waves from nuclear resonance, terahertz (THz), infrared (IR), visible (Vis) and ultraviolet (UV) regions (**Fig. 1.9A**).

Spectroscopy as applied to GAGs is broadly split into absorbance, emission, scattering and resonance spectra. Emission spectra very rarely applied to GAGs and is usually brought about through modification of the polysaccharide through tagging of the molecules with an emitting moiety, as in chromatography coupled with fluorescent detection. Monitoring of the emission of other molecules which may interact with GAGs can be used however (Doi et al., 1983; Toda et al., 1981; Uniewicz et al., 2010).

1.20.9 Absorption spectroscopy

Absorption spectroscopy is the measure of the amount of a photon at a particular frequency being absorbed by a material (**Fig. 1.9B**). Absorption spectra are usually measured by transmittance, whereby the electromagnetic radiation leaving the sample is detected, and hence the missing radiation has been absorbed. UV/vis absorbance spectra provide little information on GAG structure, possessing a band at 190-200nm which is attributed to carboxylate and acetate chromophores in uronate and N-acetyl (Lima et al., 2011). UV spectroscopy can be used to monitor depolymerisation, through detection of the unsaturated carbon ring formed (Alkrad et al., 2003), and has been used to detect small molecule contaminants in pharmaceutical preparations (Lima et al., 2011). THz spectra provide information on the available conformational states in biomolecules (Hand and Yates, 2017) and is a relatively novel form of spectroscopy which has only recently become viable for GAG study due to technological progress (Ghann and Uddin, 2017). THz radiation lies at the energy range of hydrogen bonds, charge transfer and Van der Waals interactions (Romanenko et al., 2017) and thus can be used to probe the higher levels of structural complexity within GAGs (Holder et al., 2012). THz spectroscopy has shown that heparin may have a large number of possible conformations in space which it can assume in order to facilitate target-binding, whilst another carbohydrate (dextran) has relatively less possible conformations (Holder et al., 2012).

Infrared spectroscopy is a commonly used technique and is universal to most labs and facilities. Bonds in a molecule vibrate and the frequency of the vibration is dependent on the type of vibration, the type of bond, the length, the atoms that lie at either end of it and the interactions felt by these components by any nearby interactions, both through space and through bond. The energy in these structures can be probed by absorption of resonant (i.e of

equivalent energy) photons and are very specific. IR spectroscopy represents a facile, simple and accessible method of GAG structural elucidation, whereby the structure and any physical changes applied to it can be monitored. Characteristic bands for different sulphates have been located and distinct spectra can even be found for different cation forms. Mostly, IR spectra are used for GAG identification (Donghui et al., 2006; Duan et al., 2013; Foot and Mulholland, 2005; Mainreck et al., 2011) and structural monitoring of GAG derivatives (Grant et al., 1987a, 1989; Grant et al., 1991). IR spectroscopy has previously been used to study heparin primary and secondary structures and cation forms (Grant et al., 1987a, 1989; Neely, 1957). IR spectra are unfortunately very broad, and many similar vibrational modes will overlap in frequency, producing unresolved, information dense bands.

Near-IR spectroscopy (NIR) focuses on the electromagnetic region between UV and IR (1000-2500 nm), a region in which various complex overtones (a transition above the first excitation state that is detected in conventional, sometimes dubbed "mid" IR) are detected. NIR is used commonly in the food industry to detect the quantities of components in a complex mixture, but has been applied to GAGs, facilitating detection of the animal origin of CS (Zang et al., 2012) and, quantification of CS in both medications (Liu et al., 2014) and cartilage (Palukuru et al., 2014) as well as the quantification of HA in broths (Dong et al., 2014; Puvendran et al., 2018). NIR has also been used to predict the properties of some GAGs, including the MW of HA (Dong et al., 2010), the effects of HA on water structure (Dong et al., 2020) and even the potency of heparin for its anticoagulant activity (Sun et al., 2010).

1.20.10 Scattering spectroscopy

Two forms of scattering spectroscopy are known, namely elastic and inelastic scattering (**Fig. 1.9D**). In elastic scattering, the direction of the scattered light changes, a phenomenon exploited in the LS detectors used for MW determination. Inelastic scattering involves a changing of energy and direction of the scattered light, often through absorption by the studied sample and is dubbed the Raman effect after its discoverer. Raman spectroscopy, like IR, examines the vibrational modes of molecules. The corresponding shift of the scattered light from the wavelength of the light source is indicative of the vibrational modes and bonds present in a sample. Raman spectroscopy has been used to identify the different GAGs, (Donghui et al., 2006; Mainreck et al., 2011), identify the CS and heparin in small quantities (Liu et al., 2014; Monfared et al., 2013) and to identify the GAGs found on a single cell's membrane (Brézillon et al., 2017). It has also been used to detect contaminants in pharmaceutical preparations of CS (da Cunha et al., 2015), characterise GAGs by sulphation level and tissue location (Mainreck et al., 2011; Monfared et al., 2013) and to monitor protein-GAG interactions, through observation of band shifts (Ishwar et al., 2009). The Raman effect has also been applied to microscopy, and has been used to probe the absorption of HA of varying MWs into the skin (Essendoubi et al., 2016).

1.20.11 Nuclear magnetic resonance

The gold standard for GAG spectroscopy is nuclear magnetic resonance spectroscopy (NMR). In NMR, nuclei with a spin are aligned in a strong magnetic field. Once aligned, the nuclei will precess with a force characteristic to the chemical environment it is found in. A variable oscillating magnetic field is introduced, and if this field is of equal strength to that of the precession i.e in resonance, the alignment of some of the nuclei will invert, reducing the applied magnetic moment to 0 and making all the nuclei precess together, creating a frequency pulse. The radio frequency can be detected, as a function of the resonant frequency required. The measured frequency is dubbed a chemical shift, as it is fundamentally a measurement of the resonant shift of the nuclei from the original, applied magnetic field and is measured in Hz (**Fig. 1.10A**). The environment of each measured nucleus and hence its resonant frequency of precession is formed based on the interactions of the nuclei through-bond, through-space and from local fluctuation during the experiment, creating distinct resonances – dubbed chemical shielding. Some chemical shifts are dependent on the orientation of the spin of other nuclei, resulting in spin-spin coupling, which can split signals into multiplets, depending on the spin of the other nuclei ("What is NMR?," 2022) (**Fig.1.10B**). Multiple nuclei can be probed for GAG NMR, H^1, C^{13}, N^{15} and occasionally F^{19} are used, which detect the named nuclei and the environment in which they exist.
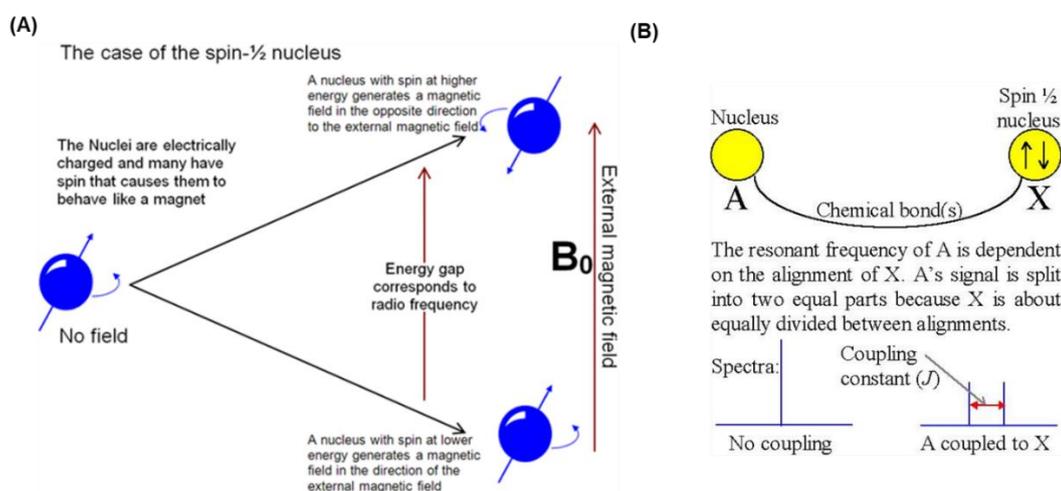
Figure 1.10: Schematic of NMR spectroscopy. A) Generating the radio frequencies for detection. **B)** Spin-Spin coupling. Acquired from ("What is NMR?," 2022)

Complete chemical shift assignments for heparin, HS, CS, HA and KS have been made in the most widely used H^1 and C^{13} dimensions (Pomin, 2014). The peak volumes of each signal can be used to succinctly quantify underlying structures, including uronate epimerisation, α/β anomers and N- and O- sulphation both in terms of relative amounts and positions (Mauri et al., 2017a). Identification of GAGs is also possible, including GAGs found in complex mixtures and different CS subfamilies through characteristic NMR profiles (Mauri et al., 2017b). NMR can also identify and study the conformational states of GAGs where it has been used to

monitor ring conformer populations and measure the overall 3D structure of the molecule (Hricovíni et al., 2002).

The true power of NMR lies in the ability to perform 2D experiments. Homonuclear experiments such as total correlation spectroscopy (TOCSY), whereby a particular frequency is excited, and the energy transfer through bonds is measured, has been used to map the rings in GAGs allowing the analyst to "trace" the signals for each carbon, while heteronuclear experiments such as heteronuclear single-quantum correlation (HSQC) and heteronuclear multiple-quantum correlation (HMQC), whereby the correlations between different nuclear dimensions, can be used to separate overlapping data and hence increase the spectral resolution, allowing minute structural components to be detected and quantified (Mauri et al., 2017a). Through-space correlations, such as nuclear-Overhauser effect spectroscopy (NOESY), where energy is transferred through space to nearby chemical environments and the corresponding effects detected, may also be performed, and is vital for the elucidation of GAG structure in 3D space (Bigler and Brenneisen, 2009; Guerrini et al., 2008) (Guerrini et al., 2008),(Bigler and Brenneisen, 2009).

Recently, other NMR dimensions have been studied, namely N^{15}, which has become more popular with the advent of more readily available high field strength (≥900 MHz) spectrometers. Probing the N^{15} dimension has allowed novel elucidation of differing chemical properties in HA end residues compared with core residues (Blundell et al., 2004), and perturbations in the amines, caused by the properties of neighbouring residues can also be detected in heparins (Langeslay et al., 2013).

1.20.12 Circular dichroism

Circular dichroism (CD) is a spectroscopic technique whereby the difference of left- and right-handed circularly polarised light is detected at different wavelengths across the electromagnetic spectrum. CD is often used for the study of protein structure and interaction with ligands but with the advent of more affordable CD spectrometers, has become a powerful tool for the study of carbohydrate structure and carbohydrate/protein interaction. GAGs have many CD chromophores (chiral compounds) in the UV region, including ring, glycosidic and hydroxyl oxygens, C5 epimers, carboxylate resonance and NS and NAc transitions (Rudd et al., 2008). The relative contributions of these chromophores to the spectrum are influenced heavily by the chemical environment of the carbohydrate, including sulphate substitution and bound cations (Rudd et al., 2007). CD produces succinct spectra for the different GAGs, as well as various heparin derivatives, including low molecular weight heparins and chemically modified heparins (Rudd et al., 2009). CD has been used to identify GAGs (Mycroft-West et al., 2020, p. 1) and is even sensitive to the direction of the various hydroxyls on the pyranose rings (Morris et al., 1975).

1.21 Techniques applied to the quality control of pharmaceutical heparin.

Full length heparin cannot currently be synthesised chemically and hence remains to this day a natural product, extracted from animals – primarily pigs and cows. This raises a series of issues, firstly heparin extracted from different animals, herds of the same species and even tissues within the same animal exhibit different structures (Al-Hakim, 2021; Bianchini et al., 1985; Casu et al., 1996; Devlin et al., 2019b; Shi and Zaia, 2009). Secondly, certain animals are not fit for use for certain demographics due to religious or ethical reasons and thirdly, heparin is currently unsequencable (Wu et al., 2019) and possesses a complex structure which cannot easily be defined and is liable to further, significant changes post-synthesis (Beccati et al., 2010).

Early heparin monographs were designed to prove that the sample to be controlled mirrors the action of what it is claimed to be, not whether it could be, or contain, anything else (Liu et al., 2009). A simple clotting test, the activated clotting time (ACT) test was utilised which entirely circumvented the need for structural definition and confirmation of heparin. Screening in this manner possessed a fatal flaw; it assumed that the tested substance was in fact heparin or that at the very least, a contaminant would alter the activity in such a manner that could be detected in the screen. The limits of this assumption were tested in 2007-2008, when batches of heparin, contaminated with an unknown agent began to cause severe adverse effects resembling anaphylaxia (Blossom et al., 2008; Guerrini et al., 2008; Kishimoto et al., 2008; Liu et al., 2009; Ramacciotti et al., 2011). At least 100 deaths occurred in north America and contaminated heparins were found in 10 other countries (FDA, 2013).The FDA believe that the contamination was intentional, owing to a low supply of heparin that year due to blue-eared-pig disease thinning the pig herds used to extract heparin (Colombo et al., 2022; FDA, 2013).

Through the concerted effort of numerous laboratories, the contaminant oversulphated-CS (OSCS) was identified (Guerrini et al., 2008; Liu et al., 2009). OSCS is a derivative of normal CS which has been chemically sulphated artificially to unnaturally high levels. The high charge density of OSCS makes it act as a charged surface which activates FXII and in turn stimulates coagulation and the contact system, which mediates vascular permeability and vasodilation, resulting in edema and hypotension (Adam et al., 2010; Devlin et al., 2019b).

Following the contamination, a slew of new techniques were added to the heparin monograph, with the express goal of confirming pharmaceutical heparin's identity, before screening for its activity. Two techniques, namely 300 MHz ^{1}H NMR spectroscopy (Beyer et al., 2008) and CE (Volpi et al., 2009), were added to the heparin USP monograph and to the monographs of many other pharmacopoeias (Alban et al., 2011b; Fernandes, 2018a). ^{1}H NMR offers a high-resolution technique that can demonstrate the presence of contaminants. Most, if not all, of

the major heparin ^{1}H NMR signals have been assigned with 4-5 characteristic signals depending on the monograph (**Table 1.2**) routinely used to confirm the presence of heparin in the sample. (Guerrini et al., 2001; E A Yates et al., 1996). Changes to, or a lack of, any of these signals may indicate a contaminant or a suboptimal heparin sample.

Separation of OSCS with CE relied on the high sulphate levels of OSCS compared with NMR. CE also allows for very small sample quantities to be separated. NMR and CE are orthogonal techniques that elucidate different components within the structure of heparin, hence, features of heparin that may be overlooked in one technique will be observed within the other (Guerrini et al., 2009; Keire et al., 2010a). For example, unusual sequences that resist enzymatic digestion will be poorly resolved or ignored entirely in separation techniques but may be detected by NMR (Spencer et al., 2009), thereby providing a more robust form of quality control. NMR and CE can also detect reasonable levels of OSCS contamination to the level of 0.5% and 1% (w/w) respectively (Edens et al., 1992; Volpi et al., 2009; Zhang et al., 2009).

The monograph has since been updated (**Fig.1.11**) (Fernandes, 2018b) and now requires a 500 MHz NMR spectrometer in order to increase the spectral resolution of the spectrum and therefore its discriminatory power. SAX-HPLC has replaced CE as it affords improved separation of potential contaminants, is more universally available, requires less-specialised training, and does not suffer from reproducibility issues that have been shown to plague CE (Trehy et al., 2009a). The ACT was also revised, opting for chromogenic Factor Xa and Factor IIa assays which more directly assay the effects of heparin on the coagulation cascade. The limits for each test and thus the definition of an acceptable heparin was defined by pharmaceutical heparins that have already been available on the market.
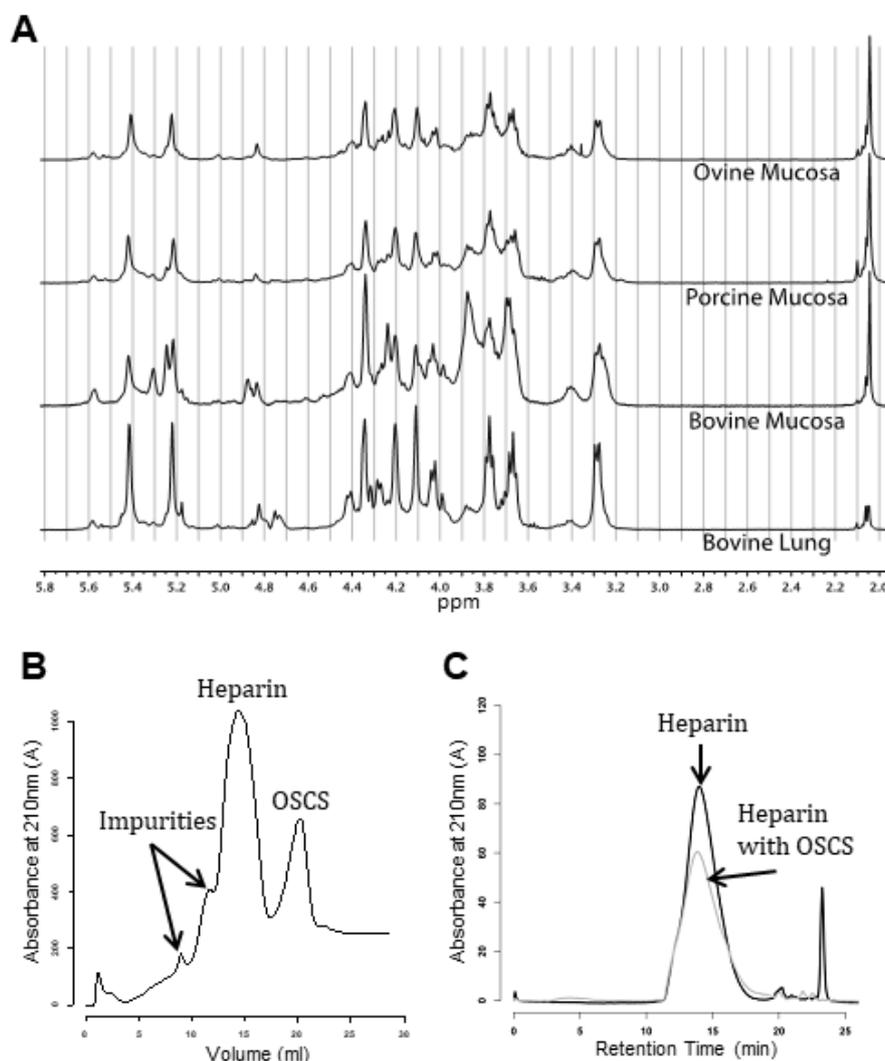
Figure 1.11: Examples of techniques that are currently or have been suggested to be used in the heparin monograph. **A)** NMR spectroscopy. **B)** SAX chromatography. **C)** SEC, displaying that OSCS MW can be manipulated to hide beneath the chromatogram. Figure reproduced from (Devlin et al., 2019b).

Currently, the monographs for the European union (EU) (EP, 2010), the US (USP, 2009) and Brazil (BHRA, 2019) require 300, 500 and 500MHz H-1 NMR respectively, and the presence of 5 major signals at 5.42, 5.22, 4.34, 3.27 and 2.04 ppm; 4 major heparin signals at 5.42, 5.21, 3.28 and 2.05 ppm; and 5 major signals at 5.40, 5.31, 5.22, 3.28 and 2.05 ppm respectively. In the EU monograph, the presence of DS at 2.08 ppm is accepted. In all three monographs, the chemical shifts can have a range of +/- 0.03 ppm and no extraneous signal in defined regions at more than 4% the relative height of the signals at 5.42, 5.42 or 5.21, or 5.40 for the EU, US and Brazilian monographs respectively. SAX is required, and the test sample is to have a retention time of ~1 compared with a heparin standard. Expected retention times for heparin, DS, CS and OSCS contaminants are described in each monograph and vary depending on the column set-up required. All three require a chromatogram of a standard DS, heparin and OSCS for comparison with the test sample. There must be no OSCS present

in all three and DS/CS cannot be more than 4% in the Brazilian monograph. Full details of the three monographs, their requirements and extra tests for purity can be found in **Table 1.2.**

Table 1.2: Comparison of the monographs for the EU, US and Brazil. Full monographs can be found in (BHRA, 2019; EP, 2010; USP, 2009).

Monograph Technique/Test	Details	EU	US	Brazil
Potency	Anti Factor IIa	180 IU/mg	180 IU/mg	180 IU/mg
	Anti-factor Xa:Anti-factor IIa	0.9-1.11	0.9-1.1	0.9-1.1
H-1 NMR	Minimum Field Strength (MHz)	300	500	500
	Signals required (ppm +/- 0.03)	5.42, 5.21, 4.34, 3.28, 2.05	5.42, 5.21, 3.28, 2.05	5.40, 5.31, 5.22, 3.28, 2.05
	Unidentified Signals	no unidentified signals larger than 4% compared to the signal at 5.42 in 0.10-2.00 ppm, 2,10-3.10 ppm and 5.70-8.00 ppm.	No signals greater than 4% the average height of the average of the signals at 5.42 and 5.21 can be present in the ranges of 0.10-2.00, 2.10-3.20 and 5.70-8.00 ppm	no unidentified signals larger than 4% compared to the signal at 5.40 in 0.10-2.00 ppm, 2,10-3.20 ppm and 5.70-8.00 ppm.
	Variation in Signal Intensities	Variation in the intensities of the peaks between 3.35 and 4.55 is allowed.	Variable signal intensities may be observed between 0.10 and 3.00 ppm.	
	Notes	DS may be observed at 2.08 +/- 0.02 ppm	Shift for OSCS at 2.16 +/- 0.03 ppm is present in standard and must not be present in test sample.	Shift for OSCS at 2.16 +/- 0.03 ppm is present in standard and must not be present in test sample. Ratio between signal at 5.31 and 5.40 must be > 20.
SAX-Chromatography	RT compared to Standard RT — Heparin	1 (~26min)	1 (~30min)	1 (~19min)
	RT compared to Standard RT — DS/CS	0.9 (~23min)	0.67 (~20min)	0.8 (~15min)
	RT compared to Standard RT — OSCS	1.3 (~34min)	1.7 (~50min)	1.1 (~21min)
	Precolumn	size: l = 0.05 m, Ø = 2 mm, Stationary Phase: Dionex AG11-HC (recommended) (13 um)	size: l = 0.05 m, Ø = 2 mm, Stationary Phase: Dionex Ion Pac AG-11, Dionex Ion Pac AG11-HC, Dionex IonPac AS11 (13um)	size: l = 0.05 m, Ø = 2 mm, Stationary Phase: Ionpac AG11-HC (9 um)
	Column	size: l = 0.25 m, Ø = 2 mm, Stationary Phase: Dionex AS11-HC (recommended) (9 um)	size: l = 0.25 m, Ø = 2 mm, Stationary Phase: Dionex Ion Pac AG-11, Dionex Ion Pac AG11-HC, Dionex IonPac AS11 (13um)	size: l = 0.25 m, Ø = 2 mm, Stationary Phase: Ionpac AS11-HC (recommended) (9 um)
	Notes	Sum of DS/CS peak area cannot exceed the sum of the DS/CS peak from the standard. No unidentified peaks.	RT of the major peaks in the sample correspond with the RT of the standard.	Amount of DS > 4% PMH. There can be no other peaks aside from PMH and DS.
Other Tests	Appearance	Clear solution	n/a	white/nearly white powder
	pH	5.5 - 8.0	5.0 - 7.5	5.0 - 8.0
	Residual Protein (Folin	>0.5%(w/w)	>1.0%(w/w)	>0.5%(w/w)
	Nucleotide Impurities (UV Spectrometry at 260nm)	Absorbance for 4mg/ml < 0.15	Absorbance for 4mg/ml < 0.20	Absorbance for 4mg/ml < 0.15
	Nitrogen	1.5% < x < 2.5% (w/w)	1.3% < x < 2.5% (w/w)	1.5% < x < 2.5% (w/w)
	Sodium	9.5% < x < 12.5% (w/w)	n/a	10.5% < x < 13.5% (w/w)
	Heavy Metals	<30ppm	<30ppm	<30ppm
	Loss on drying (dry under	<8.0%	<5.0%	<8.0%
	Bacterial endotoxins	<0.01/IU heparin	<0.03/IU heparin	<0.03/IU heparin
	GalA:Hexosamine (HPLC Amperometric test)	n/a	<1% GalA	n/a

1.22 Additional techniques for heparin quality control

The current heparin monographs are set up to monitor heparin identity based on two orthogonal approaches: species separation (SAX-HPLC) and structural elucidation (NMR). Numerous groups have explored many alternative and some novel techniques, which may complement or entirely replace these methodologies to better test the quality of heparin.

1.22.1 Species Separation:

Electrophoretic separation has already been used in the heparin monograph in the form of CE, where detection of OSCS and DS (a common impurity in heparin) at the levels of 0.05% and 0.1%, respectively, have been achieved (KAKOI et al., 2009; Somsen et al., 2009; Volpi et al., 2009; Wielgos et al., 2009). Other electrophoretic formats such as PAGE and agarose gel electrophoresis have also been used to distinguish OSCS from heparin whereby the test sample is degraded with nitrous acid, a procedure that primarily affects the NS groups found in heparin but not OSCS, allowing separation (Zhang et al., 2009). PAGE separation using this method relies on the contaminant to not be N-sulphated, and thus is targeted primarily at OSCS, which it has detected to the level of 0.1% (w/w), whilst being low-cost and facile for performance in most labs (Beni et al., 2011). Agarose gel electrophoresis has also been applied, with 1% (w/w) OSCS detection reported (Volpi and Buzzega, 2012; Volpi and Maccari, 2006).

Chromatographic methods are currently in use in the form of SAX, but there are numerous alternatives that may be of use, as reviewed in (Fasciano and Danielson, 2016). SAX can detect 0.02%–0.1% (w/w) OSCS (Fasciano and Danielson, 2016; Trehy et al., 2009a), while WAX can detect OSCS at the level of 0.25% (w/w) (Hashii et al., 2010). Unlike SAX, WAX-HPLC provides the ability to trap OSCS oligosaccharides on the column, redirecting heparin to waste-allowing multiple injections of the same sample to reveal low OSCS contents (Beni et al., 2011). RPIP has also been used and is particularly suited to constituent saccharide analyses of samples, where individual di-, tri-, and tetrasaccharide levels can be determined. The synthetic nature of OSCS means it resists digestion by conventional enzymes and therefore does not depolymerise into its saccharide constituents, allowing easy detection and resulting in reduced peak integrals on RPIP chromatograms. HILIC has also been used and can separate 1% (w/w) OSCS contamination, with an estimated lower level at 0.2% (w/w) (Li et al., 2015; Yang et al., 2011). Chromatographic methods are useful in quality control as they can separate many possible contaminants, assuming that they are susceptible to separation by the desired method and unusual chromatograms may be indicative of novel contaminants and hence they can be used proactively. These techniques suffer however when separating contaminants or impurities with similar primary structures, as they may undergo similar separation to the drug they contaminate.

Methods of separation lend themselves to hyphenation in the form of downstream detection, such as MS (Beni et al., 2011; Ruiz-Calero et al., 2001), fluorescent detection (Skidmore et al., 2010; Volpi and Maccari, 2006) and additional size definition with isoelectric focusing (Holman et al., 2010). MS methods can be employed post-depolymerisation and separation by other means (Volpi et al., 2008; Zamfir and Peter-Katalinić, 2004) but they may also be employed to separate whole heparin molecules and other GAGs (Nemes et al., 2013). Oligosaccharide fragments created during ionisation are relatively well documented, hence unknown or unusual fragments, corresponding presumably to contaminants can be identified and quantified accordingly, whilst also facilitating a higher throughput (Nemes et al., 2013; Wang and Chi, 2018). Performing hyphenated separations using established methods allows an increase to the sensitivity of the individual methodologies, for example HILIC- and RPIP-HPLC separation coupled to MS has detected 0.1% and 0.5% (w/w) OSCS contamination (Brustkern et al., 2010; G. et al., 2014). Recently however, time of flight secondary ion MS (ToF-SIMS) has been used without pre-MS separation and, when coupled with chemometric analysis and monitoring of specific signals, has detected OSCS at the level of 0.001% (w/w) from GAG films (Hook et al., 2021).

Further separation of depolymerised species is another avenue of considerable interest; impurities such as oversulphated heparan sulphate (OSHS) may respond very similarly to heparin during separation as they possess similar charge densities and disaccharides, separation of which is very difficult due to their similar charge densities, polarities, and hydrophobicity. Isoelectric focusing is one method that has been used to further separate these as described in (Holman et al., 2010).

1.22.2 Size Definition

The size definition of heparin is yet another challenging task as heparin exists as a heterogeneous population of varying MWs. Larger heparin molecules possess increased activity, but provide a larger surface area for potential side effects to occur (Mulloy et al., 2014b). The size of heparin can be defined in many ways, including legacy methods such as ultracentrifugation and viscosity measurements (Barlow, 1985; Stivala et al., 2004) and more modern methods such as calibration through use of standards, utilising PAGE (Edens et al., 1992; Ly et al., 2011; Malsch and Harenberg, 1994; Sommers et al., 2011b), and SEC (Ahsan. et al., 1995; Guo et al., 2003; Rodriguez and Vanderwielen, 1979; Sommers et al., 2011b) or through use triple detectors (Barlow et al., 1961; Beirne et al., 2011; Bigler and Brenneisen, 2009; Knobloch and Shaklee, 1997; Sommers et al., 2011b; Szajek et al., 2016). MW has also been defined using ^{13}C NMR (Desai and Linhardt, 1995) and MS (Chi et al., 2005; Juhasz and Biemann, 1994). Size definition when compared to standards is difficult, as carbohydrate standards are hard to define or acquire, meaning that they are usually created in-house, leading to variation between labs, however a USP MW standard has been made available in

an attempt to reduce this (Bertini et al., 2017; Mulloy et al., 2014b; Sommers et al., 2011b). A triple detector (TDA) approach has also been utilised, where a RALS detector, a differential refractive index detector (SLS detection), and a viscometer are used in concert to accurately define the MW of heparin, and may also be used to create uniform standards (Bertini et al., 2005b). MW determination can also be used to screen for contaminants due to the resistance of OSCS to enzymatic depolymerisation, resulting in easy detection (Viskov et al., 2009).

Size definition of the heparin chains was suggested for use in the monograph. It was suggested that acceptable heparin be defined as possessing no more than 20% exceeding 24,000 Da; an average MW between 15,000 and 19,000 Da and a ratio of populations between 8000:16,000 and 16,000:24,000 Da at no less than 1.0 (Mulloy et al., 2014b),(Szajek et al., 2016). The upper limit exists to diminish potential side effects associated with higher MW larger chains (Walenga et al., 2005), while the average exists to ensure good potency and the ratio serves to prevent blending of failed heparin lots with accepted ones (Mulloy et al., 2014b). However, owing to the nefarious nature of heparin:OSCS contamination, the OSCS that was blended with the heparins was of a comparable MW, avoiding detection (**Fig. 1.11C**)

1.22.3 Spectroscopic methods

While ^{1}H NMR spectroscopy is a part of the heparin monograph due to its high resolution and complete signal assignment, there are many other ways that NMR spectroscopy could be utilised, for example employing the ^{13}C nucleus, which is able to distinguish identical chemical features located in distinct saccharide sequences (Yates et al., 1996). ^{15}N NMR has also been studied, with some nitrogen microenvironments being assigned, producing fingerprints of intact and modified heparins (Langeslay et al., 2013, 2012a). However, some contaminants such as sulphated agarose or OSHS may lurk beneath the spectrum of heparin, avoiding detection (Devlin et al., 2019a; Rudd et al., 2011a). The true potential for NMR in quality control lies in powerful higher-dimensional experiments, including COSY (Guerrini et al., 2008; Rudd et al., 2011a, 2012a), TOCSY (Guerrini et al., 2008; Langeslay et al., 2013, 2012a), HSQC (Guerrini et al., 2009, 2008), HMQC (Guerrini et al., 2009)**,** NOESY (Guerrini et al., 2008),(Bigler and Brenneisen, 2009), and diffusion ordered spectroscopy (DOSY) (Sitkowski et al., 2008)**.** These 2D NMR methods have been used in the study of the higher-level structure of heparins and were pivotal in first identifying OSCS as a contaminant (Liu et al., 2009),(Guerrini et al., 2008). Furthermore, HSQC and HMQC using the ^{1}H and ^{13}C dimensions have been used to distinguish contaminants other than OSCS, such as OSHS, while the ^{15}N dimension has been used to fingerprint heparins (Langeslay et al., 2013). DOSY, which measures the differences between diffusion coefficients in molecules has also been applied to detecting OSCS contamination, showing clearly OSCS contamination and also distinguishing it from other types of contamination (Sitkowski et al., 2008). HSQC with SAX-HPLC have also been validated as complementary methods that may study heparin structure

and therefore its quality (Spelta et al., 2019). 2D NMR was however deemed too complex, time-consuming, and expensive for use in the heparin monograph.

CD could be used to detect contamination, due to the alteration of normal heparin bands appearing in contaminated samples, with down to 3.5% (w/w) OSCS contamination detected when isolating wavelengths correlated to known contaminants (Stanley and Stalcup, 2011a). The vibrational spectroscopies NIR, IR and Raman may also be studied for use in heparin quality control. When applied to OSCS screening, 38/41 samples could be correctly identified with IR and 36/38 samples could be correctly identified with Raman spectra, down to the level of 1.3% (w/w) OSCS (Spencer et al., 2009).

1.22.4 Other Methods

Other, less conventional methods have also been applied to heparin screening. Full-length heparin and OSCS competitively inhibit many DNA polymerases, including Taq polymerase, the enzyme used in the polymerase chain reaction. Therefore, the digestion of the heparin component should leave any OSCS, and subsequently prevent the polymerase chain reaction from occurring. This rationale has been applied to a quick heparin screen, and as low as 0.16% (w/w) OSCS contamination has been detected (Tami et al., 2008). A microplate based array has also been applied to heparin screening, taking the form of a colorimetric assay, which changes colour depending on the molecule (full length or depolymerised heparin and OSCS for example) bound to the cationic polythiophene polymer (3-(2-(N-(N0-methylimidazole))ethoxy)-4-methylthiophene (LPTP)). The colorimetric shift can be recorded, and perturbations in the shift correlating to OSCS at the level of 0.001% (w/w) have been reported (Sommers et al., 2011a). Fluorescent methods have also been developed, typically taking the form of a Förster resonance energy transfer (FRET) system, whereby some fluorescent molecule that interacts with both heparin and OSCS is introduced to the test sample and, through manipulation of the heparin molecule, either through enzymatic depolymerisation (Alban et al., 2011a; Ding et al., 2017, 2015; Hu et al., 2016; Kalita et al., 2014; Lühn et al., 2011), or cation modification (Lee and Tseng, 2015), only OSCS is left to cause fluorescence. Levels of OSCS between 0.5% (w/w) and 0.0001% (w/w) have been detected with these methods. The differing charge densities between heparin and OSCS have also been exploited to help detect OSCS, whereby the electric potential across an anionic-sensitive membrane can be altered depending on the charges of the GAGs present, and hence 0.5% (w/w) and theoretically lower OSCS and DS can be detected (Wang et al., 2008). These techniques, while powerful, are strictly applicable to OSCS and in some examples OSHS and DS, as they rely on the unique interplay between OSCS and heparin and cannot be applied to quality control of any other or novel foreign body.

1.23 Novel Issues Facing Heparin Quality Control

Numerous techniques exist that are used or may be applied to the quality control of heparin but none of them directly combat one of the major issues in heparin analysis: its high levels of heterogeneity. As heparin is an inherently polydisperse molecule, it is very difficult to confirm the fidelity of each technique to the true structure of the sample. Unfortunately, most heparin QC methods are targeted specifically to the detection of OSCS and very rarely attempt to detect other contaminants. In 2011, Rudd et al. put forward the idea of screening for a proposed heparin against a library of previously accepted heparins and through use of the chemometric method principal component analysis (PCA), the authors were able to create a region in the N-dimensional space where accepted heparins should fall, and thus detect "unusual" heparin samples (i.e., those contaminated with OSCS, down to the level of 3%) with NMR (Rudd et al., 2011a).

While NMR is powerful, it is expensive, technically challenging and requires trained staff to run and maintain. NMR of pharmacopeia-grade field strength in the US and Brazil is further expensive and technically challenging and also relatively rare. Entire universities or commercial laboratories may contain a single NMR machine of sufficient strength, the use of which may be highly competitive. Ideally, a different spectroscopic technique could be applied to heparin quality control, one which is cheap, facile, readily available and has a similar sensitivity. Other spectroscopic techniques have been applied to heparin quality control such as CD (Stanley and Stalcup, 2011a) and Raman (Spencer et al., 2009) but they too are difficult to find and expensive. Furthermore, there are fairly few GAG chromophores in CD and the ones that can be found are broad and overlap, while relatively little work has been put into Raman spectral band assignment. 6-O-, 2-O- and N-sulphate bands have been assigned, but quantification of each sulphate type is difficult as the band seems to signify the presence of, rather than a relative amount of each sulphate (Atha et al., 1996; Cabassi et al., 1978). The bands in Raman spectra instead seem relative to the quantity of a specific GAG species, rather than the quantity of a functional group found within, due in part to the highly complex interplay between these chromophores and their surroundings. IR has also been applied, particularly NIR but it also suffers with poor band assignment and broad, overlapping signals (Spencer et al., 2009).

Chemometric analyses (the mathematical breakdown of a data set into a series of more manageable trends) have become a major staple in spectroscopy, with broad but complex spectra as they afford rapid analysis. This is seen particularly with NIR, where most contemporaneous applications are based on chemometric models, which allow correlation between the spectrum and the samples physical activities (Dong et al., 2014; X. Liu et al., 2014; Sun et al., 2010). Chemometric analysis creates a unique opportunity to rapidly screen large spectroscopic data sets for interesting and previously uncovered trends, and is now

commonly applied to many spectroscopic techniques, including retrospectively to heparin NMR (Monakhova and Diehl, 2019).

IR-microscopy has been shown to distinguish different GAGs as dried films (Mainreck et al., 2011) and liquid and pellet IR has been used to identify the locations of different sulphate moieties, ring conformations and interactions across monosaccharides (Cabassi et al., 1978; Casu et al., 1978; Grant et al., 1987a; Grant et al., 1987; Grant et al., 1989; Grant et al., 1989b; Grant et al., 1991; Mathlouthi and Koenig, 1986; Vasko et al., 1971). Reflectance IR has rarely been used for heparin structural studies, primarily being used to detect heparin when attached to medical devices (Aksoy et al., 2008; Harada et al., 2005). Reflectance IR has also been briefly applied to the detection of OSCS and was shown to be comparable with NMR for the detection of OSCS (Alban et al., 2011b; Burmistrova et al., 2020; Norwig et al., 2009). Solid state IR represents a unique chance to rapidly perform quality control of heparins, due to its minimal sample preparation, fast spectral acquisition time and comparable OSCS detection to NMR.

1.24 Infrared Spectroscopy.

IR, sometimes referred to as vibrational spectroscopy is the study of infrared absorption by molecules. A covalent bond between two molecules in a diatomic system can vibrate in 2 ways: it may stretch, or it may bend. The frequency at which this vibration occurs is variable, based on the masses of the atoms at each end of the bond and the type of vibration. The vibration can be thought conceptually as a massless spring between two atoms and thus its frequency of vibration can be defined using Hooke's law (**eq. 1.1**).

$$f = \frac{1}{2\pi}\sqrt{\frac{k}{\mu}}$$

(**Eq. 1.1**)

Where f is the frequency, k is the spring constant, a constant that is characteristic of the type of vibration and μ is the reduced mass for the two atoms (**eq. 1.2**).

$$\mu = \frac{m_1 m_2}{m_1 + m_2}$$

(**Eq. 1.2**)

Where m_1 is the mass of the first atom and m_2 is the mass of the second atom.

The derivation of **eq. 1.1** is above the scope of this introduction and can be found in (Siebert and Hildebrandt, 2007). As shown in **eq.1.1** an increase or decrease in the mass of the atoms will result in a lower or higher frequency respectively and a change in the type of vibration (i.e

[46]

a change to the spring constant) will result in a different frequency. In real terms however, various other factors including the orientation of the molecules in 3D space and subsequent forces felt by it will affect the final frequency, and the **eq. 1.1** is a gross simplification. An electromagnetic wave of the equivalent frequency, i.e, a resonant frequency can be absorbed by the bond, exciting it to a higher energy level. The frequencies of these bands are found in the mid-IR region, more specifically between $3.8 \times 10^{14} - 1.2 \times 10^{14}$ Hz ($2.5 - 25$ µm wavelength).

The positions of the atoms in 3-dimensional space plays a role in the available vibrations, for example displacement of 2 atoms in the x direction by n and -n respectively results in a stretch. Specific combinations of movement in the three directions correspond to spatial and rotational translation that result in no stretching or bending of the bonds, and hence do not produce new vibrations, for example, if all the atoms move by n in the x direction, the entire molecule has essentially moved by n in the x direction, and no new modes are introduced. Due to this there are 3n-6 vibrational states for non-linear molecules and 3n-5 vibrational states for linear molecules (rotation about the axis of the linear bond results in no spatial change), where n is equal to the number of atoms.

As the number of atoms in the system increases, the complexity of the interactions between them becomes more difficult to predict and calculate. Stretches and bends may happen in all 3 cardinal directions and occur in a symmetric and an asymmetric manner, producing different types of vibration, namely symmetric and asymmetric stretching, scissoring, rocking, wagging and twisting (**Fig. 1.12**). Importantly, these vibrations can only be investigated by IR radiation if they undergo a change in dipole moment, as the change creates a field that may interact with the probing IR radiation. Hence symmetric molecules, such as N_2 and symmetric vibrations, as in symmetric stretching of the C=O bonds in CO_2 do not absorb IR frequencies.

In a traditional IR experiment, IR light of a specific frequency is shone through the sample and then to a detector. If the frequency of the IR light matches the frequency of a bond present in the sample, it is absorbed and the detector registers less IR light of that frequency. The process is then performed over a range of IR frequencies, generating a transmission spectrum. Currently however, a polychromatic IR source is instead shone through the sample and subsequently detected with a detector that can only detect the intensity of the absorbed light. The polychromatic source goes through a beam splitter and part of it is reflected back by a movable mirror and another part by a stationary mirror. The two beams, one now reflected at a different phase to the other combine and create interference, resulting in a different intensity absorbed at the detector. As the mirror moves, the interference pattern changes in a manner indicative of the phase difference between the two waves, generating a graph of the interference as a function of the distance moved by the secondary mirror. Should any of the

polychromatic wavelengths be absorbed by the sample, the resultant intensity at the detector will be decreased at certain distances of the mirror. This information can be converted into a spectrum of the wavelengths detected, with the Fourier transform (FT) and thus produce a transmission spectrum (**Fig.1.13**). The resultant wavelength is often measured in wavenumbers, which is equivalent to $\frac{1}{\lambda}$ where λ is equal to the wavelength. Wavenumber is often measured in cm^{-1} and is historically used due to its ease to process by the human brain but can also easily be converted into energy for other analyses.
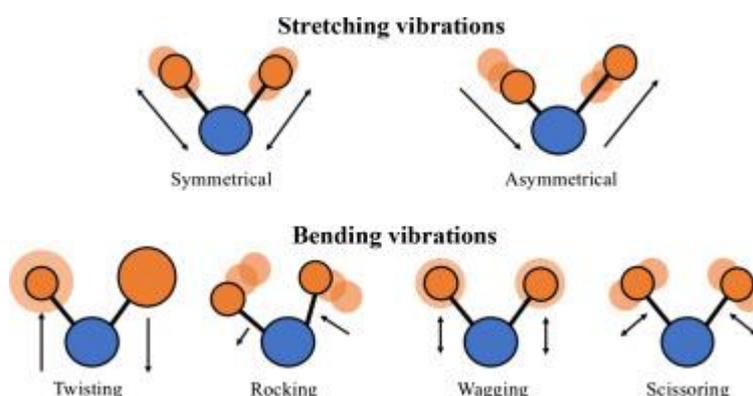
Figure.1.12: Types of stretching and bending modes. Figure reproduced from (Cameron et al., 2020, p. 10).

The absorption of IR radiation increases, based on Beer's law, both due to the amount of a single mode of the absorbed frequency found within the sample and the total concentration of the sample. The absorption band may also broaden, as the available frequencies for a single vibration increase dramatically due to intramolecular interactions such as hydrogen bonding and steric effects. The broadening is more obvious for more prevalent bonds, such as the O-H band found in water and alcohols. Certain bonds, such as the C=O and O-H have strong, characteristic bands, whose frequencies shift by a minimal amount and as such can be detected readily. These bands appear in a region of the IR spectrum known as the functional group region ($\geq$ 1500 cm^{-1}). Other bands, which are usually much sharper and arise due to unique interactions in the sample are found below 1500 cm^{-1} and represent the "fingerprint region". Hence, the functional group region can be used as an overview to identify some of the functional groups present in the sample and the fingerprint region can be used to identify unique molecules.
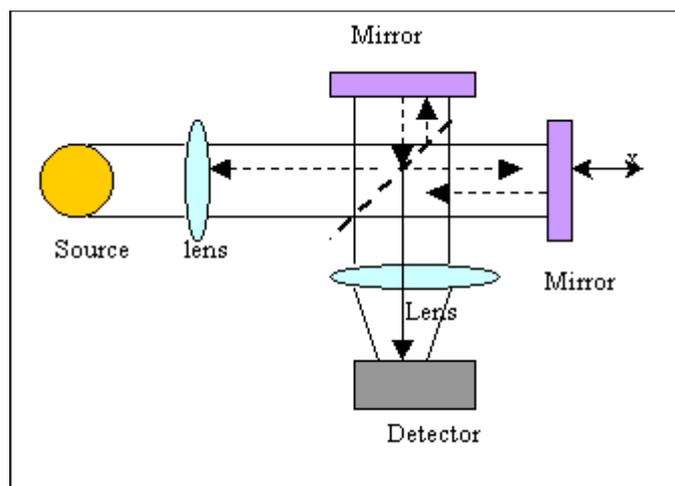
Figure 1.13: Schematic of an FTIR instrument, using a Michelson interferometer and movable mirror. Figure acquired from (physics.mq.edu.au, n.d.)

1.24.1 Types of Infrared Spectroscopy

Standard transmission IR spectroscopy can be performed on all three states of matter, through utilisation of IR invisible components as sample holders. Gaseous and liquid samples can be held between two salt plates (ionic interactions are IR invisible) while solid samples can either be diluted in a mulling agent, such as Nujol (with a very simple and noticeable IR spectrum) and applied to a salt plate, or milled with a salt, usually KBr and pressed into a thin pellet which can be analysed. If applicable, the sample may also be microtomed and a thin slice applied directly to the instrument. Spectra may also be acquired through reflectance spectroscopy, whereby the IR beam is reflected from the sample, and the absorption properties can be determined from the reflected light. There are two types of reflectance spectroscopy: external and internal reflectance. External reflectance involves shining the IR beam either directly onto a reflective sample (specular or diffuse reflection) or by shining the light through a thin layer of the sample, onto a reflective surface (transflectance) and measuring the reflected light. External reflectance is used most commonly for the evaluation of surfaces, surface coatings and impurities (Aksoy et al., 2008; Harada et al., 2005).

1.24.2 The attenuated total reflectance variant.

Internal reflectance spectroscopy involves internally reflecting an IR beam in a crystal that is contacting the sample and is known as Attenuated total reflectance (ATR). ATR requires the sample to come into contact with a crystal that has a higher refractive index than the sample. The IR light is shone through the crystal at an angle below the critical angle (**eq.1.4**), resulting in total internal reflection within the crystal. Whenever the light beam hits the interface (sometimes multiple times throughout the crystal), it creates an evanescent wave perpendicular to the crystal which permeates the sample. The evanescent wave interacts with the sample and this information is transferred back into the reflected wave, which is then

[49]

detected (**Fig. 1.13**). The permeation occurs to a distance that is a function of the wavelength of light and the refractive indices of the sample and crystal (**Eq. 1.5**).

$$\Theta_{\text{critical}} = \sin^{-1}\left(\frac{n_1}{n_2}\right)$$

<div align="right">(Eq. 1.4)</div>

Where n_1 and n_2 are the refractive indices for the crystal and the sample respectively.

$$d = \frac{\lambda}{2\pi\left(\sin^2\theta - \left(\frac{n_2}{n_1}\right)^2\right)}$$

<div align="right">(Eq. 1.5)</div>

Where d is the penetration depth and λ is the wavelength of the incident beam.
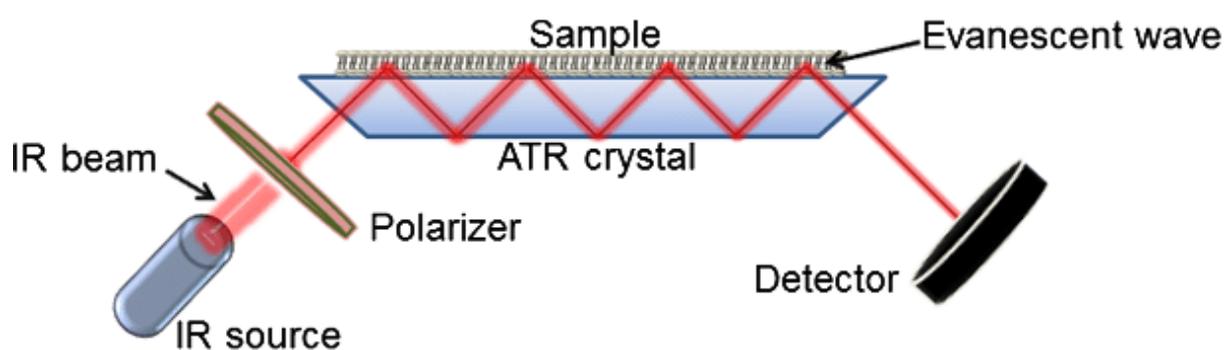
Figure 1.14: Schematic of the ATR technique. Figure acquired from (Ausili et al., 2015).

The ATR method offers a rapid technique for spectral acquisition, requiring almost no sample preparation as seen with other acquisition techniques and being incredibly simple to use. ATR does however have a drawback, as the penetration depth of the evanescent wave alters with wavelength (**Eq. 1.5**), thus the path length for each wavelength is different, and hence more absorption will be seen at greater wavelengths (lower wavenumbers) due to Beer's law. This amplification of path length is increased further with each reflection into the sample and subsequent evanescent waves formed hence, ATR spectra can have slightly different spectral intensities to standard transmission spectra, making quantification difficult.

1.24.3 The application of FTIR to GAGs.

Due to the complex nature of GAGs, the bands found in IR are generally very broad, consisting of multiple similar modes for the same chemical bonds. These modes can be visualised by differentiation of the spectrum, highlighting some of the different bands present. Regardless, the numerous bands found in the fingerprint region overlap considerably and are very difficult

to assign. Some assignments have been made and attributed to certain sulphate moieties in heparin and therefore HS, and CS and therefore DS (Alberto-Silva et al., 2020; Amarasekara et al., 2007; Cabassi et al., 1978; Casu et al., 1978; D Grant et al., 1987a; D. Grant et al., 1987; D Grant et al., 1989; D. Grant et al., 1989b; Grant et al., 1991; Mathlouthi and Koenig, 1986; Myron et al., 2017; Vasconcelos Oliveira et al., 2017; Vasko et al., 1971).

1.25 Chemometric Analysis

The broad, complex, and dense nature of IR spectra opens them up readily to examination with chemometric analysis. Chemometric analysis is the extraction of information from large data sets detailing chemical systems. This ultimately takes the form of multivariate analysis of spectroscopic data sets whereby a large and relatively impenetrable data set is broken down a series of underlying trends which an analyst can more easily examine. Spectroscopy is inherently multivariate, revolving around the detection of data at a series of variables, in the case of IR, wavenumbers. Chemometric techniques can be used to check the quality of medicines qualitatively or quantitatively (Biancolillo and Marini, 2018) and have, until recently, been used primarily for NIR spectra – spectra that, even for simple compounds are very broad and indistinct (Chen et al., 2018; Dong et al., 2014; X. Liu et al., 2014; Sarraguça and Lopes, 2009; Sun et al., 2010).

1.25.1 Principal components analysis.

The simplest form of chemometric analysis is PCA. PCA is an exploratory analysis, which aims to reduce the complexity of a data set by projecting its components onto a series of new axes which removes the dimensions in the data set but still retains the more important trends found within the data, compressing large, complex data sets into simple scatter plots (**Fig. 1.15**). PCA performs a rotation and a stretch, changing the point of view from which the analyst can observe the data. PCA is a transformation, hence the original data can always be recreated from it.
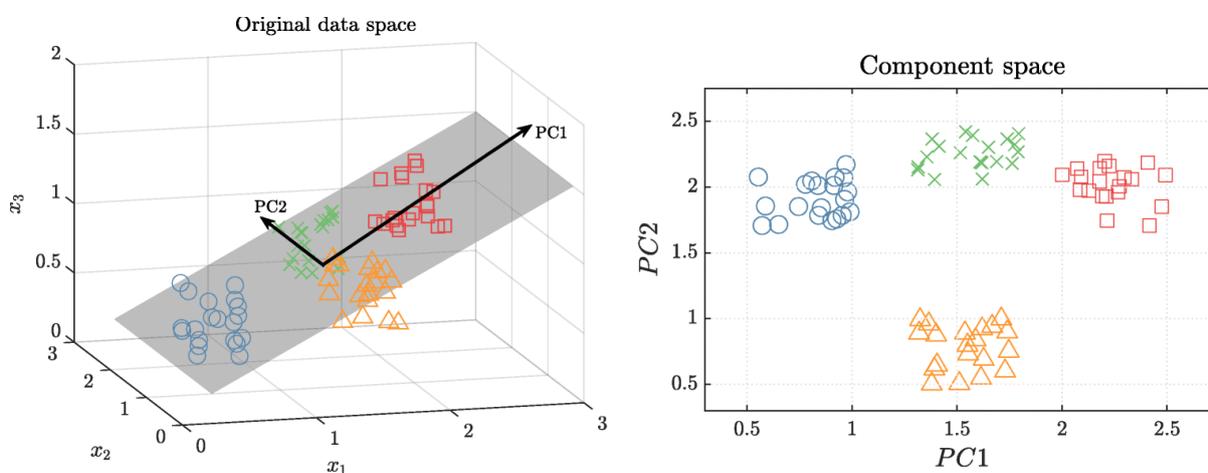
Figure 1.15: Schematic representation of PCA. Figure acquired from (Ghasemi et al., 2019)

PCA achieves this by computing a series of scores, which are the linear combinations of the data and computed scores (**Eq. 1.6**)

$$T = XW$$

Where T is a matrix of the principal component scores, X is a matrix of the original data and W is a matrix of the loadings.

The first component will explain the highest variance and thus be considered the most important, while the second vector will contain the second highest variance and thus be considered the second most important and so on.

Construction of the principal components begins with mean centring the data. If the data is not mean centred the first component will simply reflect the different means of the variables. There are two ways to compute the PC's, the first through construction of a covariance/variance matrix. For a data matrix of y samples with d dimensions, a covariance/variance matrix of d x d will be computed using **eq. 1.7**.

$$\mathrm{cov}(X, Y) = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{x})(Y_i - \bar{y})$$

The eigenvalues and eigenvectors are then computed for the covariance variance matrix using the characteristic eigenvector equation (**eq. 1.8**)

$$|X - \lambda I| = 0$$

Where λ are the eigenvalues, X is the covariance/variance matrix and I is the identity matrix.

First, the equation is solved for λ, generating a characteristic polynomial based on the value of d, the roots of which are the eigenvalues. If d = 2, then a quadratic polynomial is formed and so on. Using the eigenvalues, the eigenvectors can be calculated. The eigenvectors represent the direction from which the analyst will view the data, and the eigenvalues represent the magnitude of the effect that this projected view has, hence eigenvectors can be ordered by importance through ordering of the associated eigenvalues. From here, the original data matrix is transformed by the dot product of each eigenvector with the original data to form the principal components (PCs). The top x PCs can then be examined to understand the major underlying trends in the data set.

The PCs can also be computed through a method called singular value decomposition (SVD). SVD states that any matrix X can be factorised as in **eq. 1.9**.

$$X = USV^T$$

Where U and V are eigenvectors calculated from XX^T and X^TX respectively and S represents the square roots of the eigenvalues, dubbed the singular values.

Using SVD, the scores (T) can be written as eq. **1.10.**

$$T = US$$

In SVD, the eigenvalues and eigenvectors are computed as above with Eq 1.8 but for XX^T matrix, instead of the variance/covariance matrix. SVD is used more commonly, owing to faster computational times, but both methods can and are used for PCA (Jolliffe and Cadima, 2016).

1.25.2 The Uses of PCA

PCA highlights internal clusters and trends within the data and can also highlight which portions of the data are important for these trends. It is used in numerous fields of research including but not limited to genetics where it has been used to monitor genetic variation across geographical location (Lawson et al., 2012), environmental science where it has been used to identify factors which affect water quality in rivers (Mohd Nasir et al., 2011) and in forensic science to discriminate between the origins of different facial creams (Asri et al., 2021). Applied to glycosaminoglycans, PCA has been used to identify different low molecular weight heparins and heparinoids with CD (Rudd et al., 2009), identify different GAGs with IR microspectroscopy (Mainreck et al., 2011) and H-1 NMR (Rudd et al., 2011a), monitor the structures of crude heparins (GAG mixtures which are eventually converted into pharmaceutical heparin) (Mauri et al., 2017). Applied specifically to heparin quality control, PCA has been coupled with NMR and MS, where it has been used to identify product contamination with OSCS down to the level of 5% w/w and 0.0001% w/w respectively (Hook et al., 2021; Rudd et al., 2011a), to identify the source of heparin from different animals and blends thereof (Colombo et al., 2022; Hook et al., 2021; Ouyang et al., 2019; Rudd et al., 2019) and, to predict the anticoagulant activity of heparins with H-1 NMR (Monakhova et al., 2019).

Chapter 2.0: Materials and Methods

2.1 Materials

Details of all polysaccharides can be found in the appendix, **Tables A.1:4**. All other chemicals and reagents were procured from Fisher Scientific, UK.

2.2.0 Infrared spectroscopy sample preparation

Prior to the acquisition of spectra, brief sample preparation was performed as follows; 1-10 mg of dry sample was solubilized with 1 ml of deionised water, frozen at -80°C and lyophilized overnight.

2.2.1 Attenuated total reflectance Fourier-transform infrared spectroscopy

Samples were recorded using a Bruker Alpha I or Bruker Alpha II spectrometer (Bruker, UK) in the region of 4000 to 400 cm^{-1}, with 3 scans at a resolution of 2 cm^{-1} (approx. 70 seconds acquisition time), n=5. A background spectrum was obtained prior to recording the spectrum of each sample, using the same settings as for sample acquisition. 1-10 mg of each dried sample was placed on the crystal stage, ensuring that the entirety of the crystal was covered. A sufficient amount of sample was employed to ensure that at least 5 μm thickness was obtained, as this is the extent to which the evanescent ATR wave penetrates. The instrument stage was cleaned with water and acetone and dried between acquisitions. Spectra were acquired using OPUS 8.1 software (Bruker, UK) and exported using a CSV format.

2.2.2 Fourier-Transform Infrared spectra processing

2.2.2.1 Spectral Import and preliminary smoothing

All data processing and subsequent analyses were performed using a custom-built PC, equipped with an Intel core i7- 7700k. Spectra were imported into R studio v1.1.463 before preliminary smoothing, employing a Savitzky-Golay algorithm (**signal** package, *sgolayfilter)*, with a 21 neighbor, 2nd-degree polynomial smooth. Preliminary smoothing reduced the effects of noisy outliers failing to be baseline corrected effectively.

2.2.2.2 Baseline correction

Each individual smoothed spectrum received a baseline correction using a 7th order polynomial. Baseline correction was used to removed the effects of routine variations in sample film thickness and stage compression. Briefly, the spectra were divided into 6 equally spaced regions (buckets). The minimum absorbance value for each of these buckets, and the corresponding relevant wavenumber (x-axis) values were computed. The start and end values

for the spectrum were added to these values and from the resultant 8 x-y pairs, the coefficients for a 7th order polynomial were calculated using the ***base*** R *lm* function. The baseline was calculated utilizing the calculated coefficients and the x-values of the original x-axis, before subtraction from the smoothed spectrum.

2.2.2.3 Spectral Preparation for Principal Component Analysis

In order to remove the effects of inconsistent sample loading or densities before the recording of spectra, the corrected spectra were normalized (0-1) using the equation:

$$x_c = \frac{x - x_{min}}{x_{max} - x_{min}}$$

(**Eq. 2.1**)

Where x is the value to be corrected, x_c is the resultant corrected value, x_{max} the maximum x value for the spectrum and x_{min} the minimum x value for the spectrum.

The normalized spectra had variable regions which occur due to fluctuating CO_2 and H_2O levels in the environment, deleted (< 700 cm^{-1}, between 2000 and 2500 cm^{-1}, and >3600 cm^{-1}).

For second derivative analyses, the second derivative was taken using the Savitzky-Golay algorithm with 41 neighbors and 2nd order polynomial. The preliminary smooth is not always required, but if this step is omitted, more neighbors are required for optimum output during this later step. It was observed that less aggressive smoothing at the beginning of the process, removed anomalous baseline corrections entirely.

For some analyses, the differentiated data were autoscaled (**eq. 2.2**) to both reduce the ATR intensity effect and level the importance of weaker IR bands.

$$x = \frac{x - \mu}{\sigma}$$

(**Eq. 2.2**)

2.2.2.4 Defining optimum smoothing and correction parameters

To ascertain the optimum smoothing parameters, different degrees of smoothing (both in terms of neighbors and polynomial) were applied to the spectra and the resultant PC scores compared qualitatively. If spectra were not smoothed sufficiently, their individual 5 repeats spread across the score plots as they were dissimilar, while if the samples are over-smoothed, all samples from diverse polysaccharide types overlapped, yielding no meaningful separation. The plots with the least smoothing and the tightest sample grouping were taken forward, and these comprised 21 neighbors using a 2nd order polynomial for the preliminary smooth, and

21 neighbors and a 2nd order polynomial for the pre-differentiation smooth. The optimum baseline polynomial was also defined in a similar manner, using distinct polynomials in the range of 2nd to 9th order. For an nth order polynomial, the spectra were divided into n-1 buckets, and the same script run as in **Section 2.2.2.2**. Second and 3rd order polynomials generate poor baselines, often resulting in early or late baseline anomalies, in which alien bands are introduced as a consequence of rigidity; for 4th order polynomials and higher, the baselines are sufficient. A 7th order polynomial was chosen because it yielded the fewest unusable corrections, i.e., spectra in which the baseline became more curved.

2.2.3 Preparation of spectral libraries

Prior to sample comparison using PCA, individual sample libraries were created for each polysaccharide class (PMH, OMH, BMH, BLH, CS-A, CS-C, DS, HA, HS and OSCS). Each polysaccharide library was compared to itself with PCA, and through the use of the first 10 principal components, any outstanding samples (i.e., any that appeared distinctly removed from the main data-cluster of the library) were removed from the spectral library. Samples that were removed underwent qualitative spectral inspection, to ensure that unusual spectra were present (i.e spectra that were not smoothed properly, or contained unusual salt or contaminant peaks)

2.2.4 Preparation of samples at different concentrations.

Samples were prepared in deionised water at a stock concentration of 200mg.ml^{-1} and subsequently diluted with water to the concentrations of 100, 75, 50, 37.5, 25, 20, 15, 10, 7.5, 5, 2, and 1 mg.ml^{-1}. 2ml of each sample was decanted into an Eppendorf tube before freezing at -80 and lyophilisation overnight.

2.2.5 Calculating and plotting difference spectra

The spectra to be compared have either the average spectrum of all spectra to be compared or, in the case of a series, the spectrum corresponding to the lowest value in the series subtracted from them. Resultant spectra are plotted against each other either in 2D or 3D for easier viewing.

2.2.6 2D Correlation spectroscopy filtering of iterative random samples

Analysis was carried out as in (Rudd et al., 2012a), in brief a covariance matrix was created for a known library and a known library and a test sample. The covariance matrix of the known library was subtracted from the covariance matrix of the sample to be filtered. This was repeated 50 times, and the subsequent average matrix represented the maximal variance in

the spectrum that related to the test sample against the library. The power spectrum of this matrix was plotted as needed.

2.2.7 Regression Analyses

2.2.7.1 Linear Regression

Independent variables (usually the intensity of a selected IR band) and dependent variables underwent linear regression using regression analysis in Excel 2016 (Microsoft). The regression model was fit based on the lowest sum of squared residuals.

2.2.7.2 Multiple Linear Regression (MLR)

Multiple linear regression is essentially linear regression with many independent variables. MLR was performed using the *lm* function in **base** R. Two or more independent variables were input against a dependent variable. A series of coefficients were generated (analogous to the gradient generated in a simple linear regression) for each independent variable and the size of the coefficient was correlated with the importance of that particular independent variable and the dependent variable.

2.2.7.3 Logistic Regression Model (LRM)

Logistic regression was performed with the *glm* function in **base** R. One or more independent variables were used depending on the analysis, with the dependent variable (likelihood) set to 1 or 0 for a contaminated or not sample respectively. The *nagelkerke* function from the **rCompanion** package was used to gather descriptive statistics including R^2 values about the generated model. Likelihood thresholds were generated based on the analysis being performed. Generally, the likelihoods at which there were no false-negatives were taken forward, and were found through use of a response operator characteristic (ROC) curve, created with the *roc* function from the **pROC** package.

2.2.7.4 Predictions using a Logistic Regression Model

Likelihood values for a sample to be predicted were calculated using the *predict* function in **base** R. Relevant independent variables for the sample to be predicted were input and the computed likelihoods were compared to previously established likelihood thresholds, if they were below the threshold they are part of the group whose y=0 and if they are equal to or above they are part of the group whose y=1.

2.2.8 Chemometric Analyses

2.2.8.1 Principal Component Analysis (PCA)

The data were subject to PCA using singular value decomposition with the *prcomp* function in **base** R. During this process, the matrix was mean-centered. Through comparison of the scree

and loading plots, suitable PC scores were chosen to plot against each other as x-y scatter graphs.

2.2.8.2 Principal Components Regression (PCR)

PCR was performed when regression for prediction, or correlation of PCs to different features were required. In brief, an MLR is performed across a series of relevant PCs. The *pcr* function from the **pls** library was used. Into this function, an **x** and **y** matrix are input, whereby the **x** matrix is a PCA scores (independent variables) matrix and the **y** matrix is a vector or matrix of the dependent variables. For a matrix of **y**, a MLR is performed for each column.

2.2.8.3 Selecting ideal PCs with PCR

For continuous data, a percentage variance is calculated for each PC for each different dependent variable. The magnitude of the variance correlates with how closely related the PC is to the dependent variable and through interrogation of these values, relevant PCs can be taken forward for prediction (Massy, 1965).

2.2.8.4 Predicting with PCR.

Samples to have their dependent variables predicted were prepared in the same way as the samples which were used to perform the original PCA. The *predict* function from **base** R was used to generate predicted PC scores, whereby the sum of the loadings multiplied by the mean centered original spectra are taken (**eq. 2.3**).

$$X_{Pred} = \sum \omega x$$

(**Eq. 2.3**)

Where X_{Pred} are the predicted PC scores, ω are the loadings and *x* is the input spectrum.

The predicted scores underwent dependent variable prediction using the *predict* function in the **pls** package. The top **n** components were entered into the function or a selection of components, based on the percentage variance covered for relevant dependent variables.

2.2.8.5 Selecting ideal PCs for discrete analyses.

For discriminant and clustering analyses, ideal components were selected either through visual inspection of the components or monitoring of the frequency density for each group in each component. To do this, the frequency density for each discrete group was calculated at each score value with the *density* function in **base** R. The densities for each discrete group were plotted as a curve against the PC score value. PCs which showed separations in their densities between one or more groups from the other group(s) were selected as appropriate.

2.2.8.6 Partial Least Squares Analysis (PLS)

The data were subject to PLS analysis using the *pls* function from the **pls** library in R. Into this function, an **x** and **y** matrix are input, whereby the **x** matrix is a matrix of IR spectra and the **y** matrix is a vector or matrix of the dependent variables. The PLS loadings are then calculated based on correlations between the **x** and **y** matrices and subsequent scores for the **x** matrix are calculated from them. For each latent variable (LV) – analogous to the PCs of PCA – a percentage variance is calculated for each dependent variable which correlates with how closely related the LV is to the dependent variable and through interrogation of these values, relevant LVs can be taken forward for prediction.

2.2.8.7 Prediction with Partial Least Squares Analysis Regression (PLS-R)

Samples to have their dependent variables predicted were prepared in the same way as the samples which were used to perform the original PLS. The *predict* function from the **pls** package was used to generate both LV scores and subsequent dependent variables. The top **n** components based on the calculated percentage variance values were used for predictions.

2.2.8.8 Partial Least Squares Discriminant Analysis (PLS-DA)

PLS-DA is undertaken in essentially the same way as PLS but with the *plsda* function from the **mixOmics** package. The **x** and **y** inputs are handled in essentially the same way but the **y** inputs are categorical.

2.2.8.9 Partial Least Squares Discriminant Analysis Prediction

Samples to have the category into which they belong predicted were prepared in the same way as the samples which were used to generate the original PLS-DA. Using the *predict* function from the **mixOmics** package, the LVs were calculated for the new samples and Mahanalobis distance between each sample to be predicted and the population for each category was calculated. Samples with low Mahanalobis distances are more likely to be a part of the relevant distribution and hence the distribution which shows the lowest Mahanalobis distance was the most likely category of the sample to be predicted.

2.2.8.10 Linear Discriminant Analysis (LDA)

The data were subject to LDA analysis using the *lda* function from the **MASS** library in R. Into this function, an **x** and **y** matrix are input, whereby the **x** is a matrix of IR spectra and **y** is a vector of qualitative data labels. LDA generates a number of linear discriminants (LDs) – analogous to PCs and LVs – based on the number of discrete labels in **y,** which attempt to separate **x** into separate distributions.

2.2.8.11 Linear Discriminant Analysis Prediction

Samples to have the category into which they belong predicted were prepared in the same way as the samples which were used to generate the original LDA. Using the *predict* function

from the **MASS** package, the LD scores are calculated for the new samples and the likelihood that each sample falls into the distribution which describes each input sample label is calculated using maximum a posteriori estimation. The distribution in which the sample has the highest likelihood of being in is taken as the predicted distribution.

2.2.8.11 Independent Components Analysis (ICA)

The data were subject to ICA using the *fastICA* function from the **fastICA** package in R. The algorithm used was set to "deflation" to ensure that more than 2 ICs could be generated. 5 Ics were generated.

2.2.8.12 t-Distributed Stochastic Neighbor Embedding (tSNE)

The data were subject to tSNE using the *tsne* function from the **Rtsne** package. Perplexity was set to 30 or 50.

2.2.9 Cluster Analyses

2.2.9.1 Hierarchical Cluster Analysis (HCA)

HCA was performed using the *heatmap* function as part of **base** R on normalised and differentiated IR spectra and on PCA scores of IR spectra. *Heatmap* calls the *dendrogram* function which performs the HCA and orders samples accordingly. *Heatmap* also plots the calculated dendrogram.

2.2.9.2 K-means clustering (KMC)

Dimensionality reduced data (i.e through PCA or tSNE) had potential clusters analysed by KMC using the *kmeans* function in **base** R. Calculated means were plotted on x/y plots where appropriate. The value for K varied between 3 and 12 depending on the analysis.

2.2.9.3 K-means clustering with computed centroids (KMC-CC)

In brief, the means (known *a priori*) for each cluster in each selected component were calculated and used as the centroids for the clusters. Samples were classified based on their Euclidian distance (**eq. 2.4**) from each mean with the lowest distance meaning the highest probability of being a part of that cluster.

$$d(p, \bar{x}) = \sqrt{(p_1 - \bar{x}_1)^2 + (p_2 - \bar{x}_2)^2 + \cdots + (p_i - \bar{x}_i)^2 + \cdots + (p_n - \bar{x}_n)^2}$$

(Eq. 2.4)

Where d is the Euclidean distance between a point (*p*) and the relevant centroid mean ($\bar{x}$) in *n* dimensions.

The Euclidian distances for each cluster were subjected to an inverse SoftMax transformation (**eq. 2.5**), computing the subsequent probability that each sample is in the given cluster (Bishop, 2006).

$$softmax(x) = 1 \Big/ \frac{e^x}{\sum_{j-1}^{n} x_j}$$

Where *x* are the computed Euclidean distances for *n* clusters.

The highest probability, regardless of its closeness to other clusters, was assumed to be the cluster the sample belonged to.

2.2.9.4 Density based spatial clustering of applications with noise (DBSCAN)

Dimensionally reduced data (i.e through PCA or tSNE) had potential clusters analysed by DBSCAN using the *dbscan* function from the **dbscan** package.

2.2.9.5 Expectation maximisation (EM)

Dimensionally reduced data (i.e through PCA or tSNE) had potential clusters analysed by EM using the *mclust* function from the **mclust** package.

2.2.9.6 K-nearest neighbours (KNN)

Dimensionally reduced data (i.e through PCA, PLS or tSNE) had potential clusters analysed by KNN using the *knn3Train* function from the **caret** package. The value for K was set to 4 unless otherwise stated.

2.2.10 Testing the power of discriminant/clustering techniques

To test each discriminant technique, 1/3 of the data, picked at random without replacement was removed to become the test data. The remaining 2/3s were taken, and the discriminant technique trained on them. The test data is then predicted with the model and the entire process from test-data generation to completion was repeated 3 times. The outcomes were collected and pooled together. Confusion matrices for results comparison were produced using the *ConfusionMatrix* function from the **caret** package on relevant outputs. Where appropriate, Chi squared analysis was also performed using the *chisq.test* function in **base** R.

2.2.11 Modification of GAG samples

2.2.11.1 Conversion of GAGs to the pyridine salt:

Preparation was followed as in (Ricketts, 1952; E A Yates et al., 1996), in brief, approximately 1g of sample is dissolved into 5-10ml of deionised water and added to DOWEX marathon ion exchange beads until they have reached pH 1 (around 1 hour) the beads are removed via filtration and the resultant solution is neutralised with pyridine to yield pyr-GAG.

2.2.11.2 Collection of modified samples:

Work-up post reaction was followed as described in (E A Yates et al., 1996), in brief, 1-5 volumes of methanol saturated with Na-Acetate were added and the sample left to precipitate for at least 1hour. The samples were centrifuged and the precipitate collected. The precipitate underwent extensive dialysis in 8 kDa cut-off dialysis membrane for 3 days. The resultant samples were freeze dried and may undergo a further desalting step using a PD10 if necessary (**Section 2.2.11.3**).

2.2.11.3 PD10 Desalting

A PD10 is a commercially available SEC column, containing Sephadex G-25 medium and was used as per the manufacturer's instructions. Briefly, a PD10 column was washed with 10 column volumes of water (25 ml total). 1 column volume (2.5ml) of sample was added and allowed to drip through. 3.5 ml of water was added to the column and the resultant 3.5ml collected. The column was then washed with 10 column volumes of water, before being re-used.

2.2.11.4 Over-sulphation

Reaction was carried out as first described in (Ricketts, 1952), in brief 1g of pyr-GAG was dissolved in 10 ml of pyridine. On ice, 1 ml of chlorosulphonic acid was added. The mixture was heated to 100°C and allowed to react for 1 hour. The reaction was quenched by the addition of saturated Na carbonate solution until the mixture reached pH 7.

Sulphation was monitored by the presence or increase in intensity of the sulphate band in IR at ~1250 cm^{-1} (Amarasekara et al., 2007).

2.2.12 Determining limit of quantification (LOQ) and limit of detection (LOD).

2.2.12.1 Limit of quantification

The LOQ was defined as the maximum amount of a contaminant that was falsely predicted to be in a sample that should have been pure. For example, if a 100% heparin sample was predicted to have 10% OSCS, the LOQ was defined as 10% (or rounded up to the nearest w/w that was used in the model, for example, if 6% OSCS was detected and the model was generated using 5%, 7.5% and 10% OSCS, 7.5% was said to be the LOQ.)

2.2.12.2 Limit of detection

For PCA transformed data, LODs were defined as the point at which a contaminated sample was no longer able to be discriminated from pure samples using logistic regression. The threshold for this was determined qualitatively, based on the LRM used. Once determined, the same threshold is used for all analyses applied to that LRM unless stated otherwise. For MRA data, the LOQ as described above was used as a pseudo-LOD.

Chapter 3: The detection of OSCS with ATR-FTIR

3.1 Introduction

Heparin, a member of the GAG family, is a polydisperse, linear polysaccharide, which pharmaceutically is a potent anticoagulant drug (Rabenstein, 2002a). Heparin is the second most widely used drug by weight after insulin, and is on the World Health Organisation's list of vital medicines (WHO, 2022). The anticoagulant activity of heparin is afforded by its ability to bind to antithrombin and, depending upon its molecular weight, facilitate inactivation of factors X and II (thrombin), attenuating coagulation (Olson and Björk, 1994). Heparin is comprised of a disaccharide repeat unit of GlcNAc and IdoA, and more rarely the C5 epimer GlcA. IdoA/GlcA may be 2-O-sulphated and GlcNAc may be 6-O-sulphated and N-Sulphated, or more rarely 3-O-sulphated or a free-amine (GlcN) (Rabenstein, 2002b). Primarily, pharmaceutical heparin consists of trisulphated disaccharide repeats (IdoA2S-GlcNS,6S) which are interspersed with low amounts of disulphated, monosulphated and unsulphated saccharides, and those containing 3OS, the presence of these rarer residues are required to convey heparins pharmaceutical activity (Fu et al., 2013a).

The quality control (QC) of pharmaceutical heparin, is challenging because of the physiochemical properties of the extracted polysaccharide (Devlin et al., 2019b). Due to an inability to synthesise viable pharmaceutical heparin, it is extracted as a natural product from animals. Currently, the ability to obtain pure heparin from animal extracts is unfeasible and small amounts of contaminant DS (< 5%), and potentially HS, will be found within the final product (Baytas and Linhardt, 2020; Taylor et al., 2019). The complete sequence of heparin is also currently unknown, most likely changes on a case-by-case basis and cannot be determined, as there are currently no techniques available to sequence full length heparin chains due to a lack of both sensitivity and resolution, this is further compounded by the lack of an apparent, template-driven synthesis (Baytas and Linhardt, 2020).

In late 2007 - early 2008, batches of pharmaceutical heparin were contaminated by an unknown source with over-sulphated, chondroitin sulphate (OSCS), a non-natural, chemically sulphated GAG, which has similar anticoagulant potency to heparin, but proved toxic upon human administration. The contamination resulted in at least 100 deaths in the US alone and was found 10 other countries (FDA, 2013). Following this, the QC, or lack thereof, of heparin was called into question and numerous new techniques were proposed, with a particular emphasis on the detection of low levels of OSCS (Devlin et al., 2019b). New methods included standard spectral screening; searching for alien or shifted signals or use of the unique interplay of OSCS with current screening, such as the ability to inhibit DNA polymerase, and the ability to resist depolymerisation with current lyase enzymes (**Section 1.21**). During the

contamination, heparin lots with as high as 37% (w/w) OSCS were found (Mendes et al., 2019). OSCS has been shown in rat (Corbier et al., 2011) and pig (McKee et al., 2010) models to display no adverse effects at approximately 3% (w/w), hence an ideal screen is able to detect from ~3% w/w or below OSCS.

IR spectroscopy is a routinely used tool, used for structural study. IR has been briefly studied applied to GAGs; in particular, it is used to confirm the identity of GAG species and monitor underlying structures such as 4 and 6 sulphation in CS (Alberto-Silva et al., 2020; Amarasekara et al., 2007; Myron et al., 2017). The IR spectra of carbohydrates, particularly polysaccharides, are amongst the broadest due to the overlap of numerous functional groups within the spectra (Wiercigroch et al., 2017). Carbohydrate spectra are characterised by an intense, broad band at 3000 cm^{-1}, correlated to O-H stretching, and an intense, complex band between 900 – 1100 cm^{-1}, attributed to C-O-H bending and stretching, referred to from here as the main carbohydrate band (Wiercigroch et al., 2017). The spectra of GAGs contain extra bands at 1250 cm^{-1}, 1450 and 1600 cm^{-1}, which can be attributed primarily to sulphate, carboxylate and amine groups, respectively (Grant et al., 1989). These bands, while associated primarily with the ascribed functional regions, show complex interplay with each other, demonstrated by frequency shifts, which correspond with N-sulphate and carboxylate interactions within heparin (Grant et al., 1989b). Subtle shifts and shoulders in the sulphate band at 1250 cm^{-1} can be indicative of the underlying sulphate pattern, allowing distinction between CSA, CSC, and DS for example (Cabassi et al., 1978). Alteration of the C-O-H band can also be seen for GAGs, with additional, sharp bands appearing – most likely due to the conversion of C-O-H to C-O-S, and S=O stretching absorbing at 1050-1020 cm^{-1}. Some assignments have also been made to the directionality of sulphate moieties on the iduronate residues, indicating spectral differences that correlate with ring conformer structures (Grant et al., 1991) and the directionality of sulphate moieties (Amarasekara et al., 2007). The underlying subtleties between these bands have not been studied to any great length and instead, broad comparisons are often employed (Amarasekara et al., 2007; Burson et al., 1956; Myron et al., 2017).

3.2 Chapter aims

1. Establish whether ATR-FTIR can distinguish between heparins and OSCS and, heparins and heparins contaminated with OSCS.
2. If distinguishable, establish the limit of detection (LOD) for OSCS in heparins. The target level of OSCS is 3% (w/w).

3.3 Results

3.3.1 Exploring the viability of FTIR for heparin QC.

In order to establish FTIR's ability to quality control heparin, spectra of heparin and spectra of the primary contaminant OSCS were acquired and compared (**Fig. 3.1 A**). The two have distinct spectra, with OSCS possessing, amongst other differences, a band from the main carbohydrate band at ~1000 cm^{-1} and a more intense and slightly shifted sulphate band at 1223 cm^{-1}. The two major differences can be associated with an increase in total sulphation and of sulphation at new positions. Bands at 818 cm^{-1} and 781 cm^{-1} can also be associated with sulphate substitution. OSCS is, when compared with the natural CSs (**Fig. 3.1 B**) the most spectrally similar to heparin. The barycentre of the main carbohydrate band is located between 1029 and 1041 for the CSA through E, while it is located at ~1010 cm^{-1} for heparin and OSCS. The band at 781 cm^{-1} is unique to OSCS. The similarity of OSCS to heparin spectrally presents a potential difficulty in the screening for this molecule, but the resolution of it from heparin is still possible.
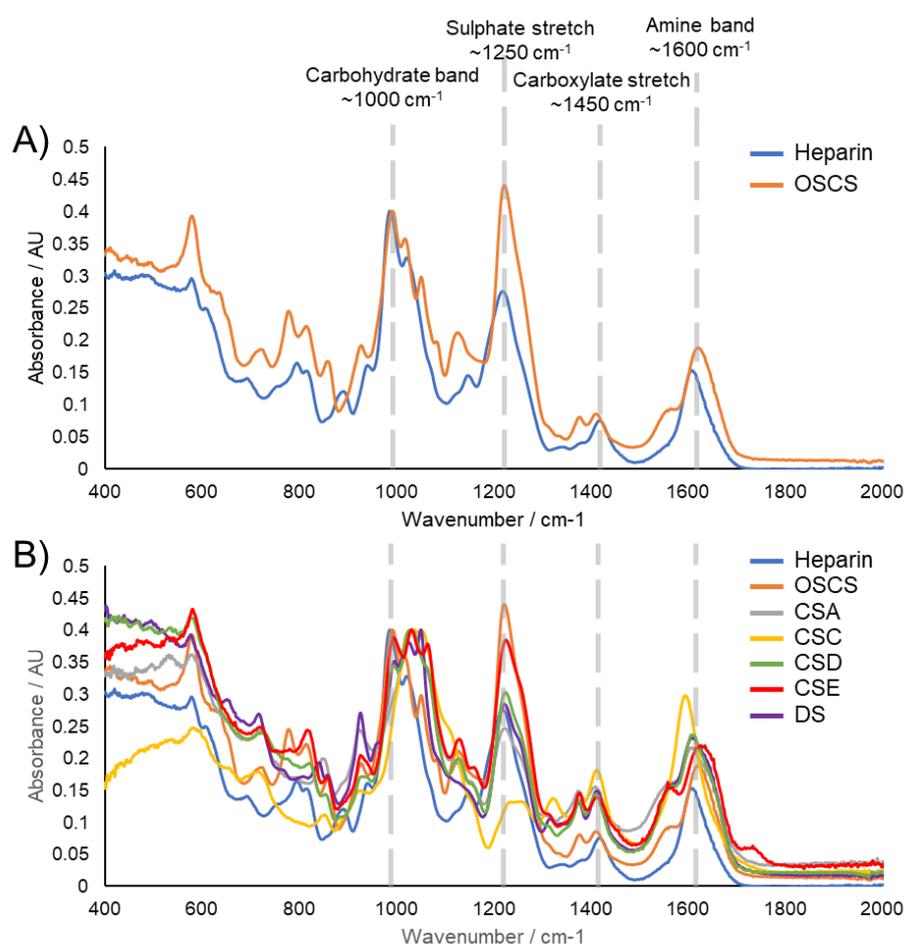
Figure 3.1: ATR-FTIR spectra of heparin vs different CSs. A) Heparin vs OSCS, B) Heparin vs other CSs. Heparin: blue, OSCS: Orange, CSA: grey, CSC: yellow, DS: purple, CSD: green, CSE: red. Spectra are between 400 and 2000 cm^{-1} and are averages of 5 repeats of 3 scans and were normalised to the absorbance at 1050 cm^{-1}.
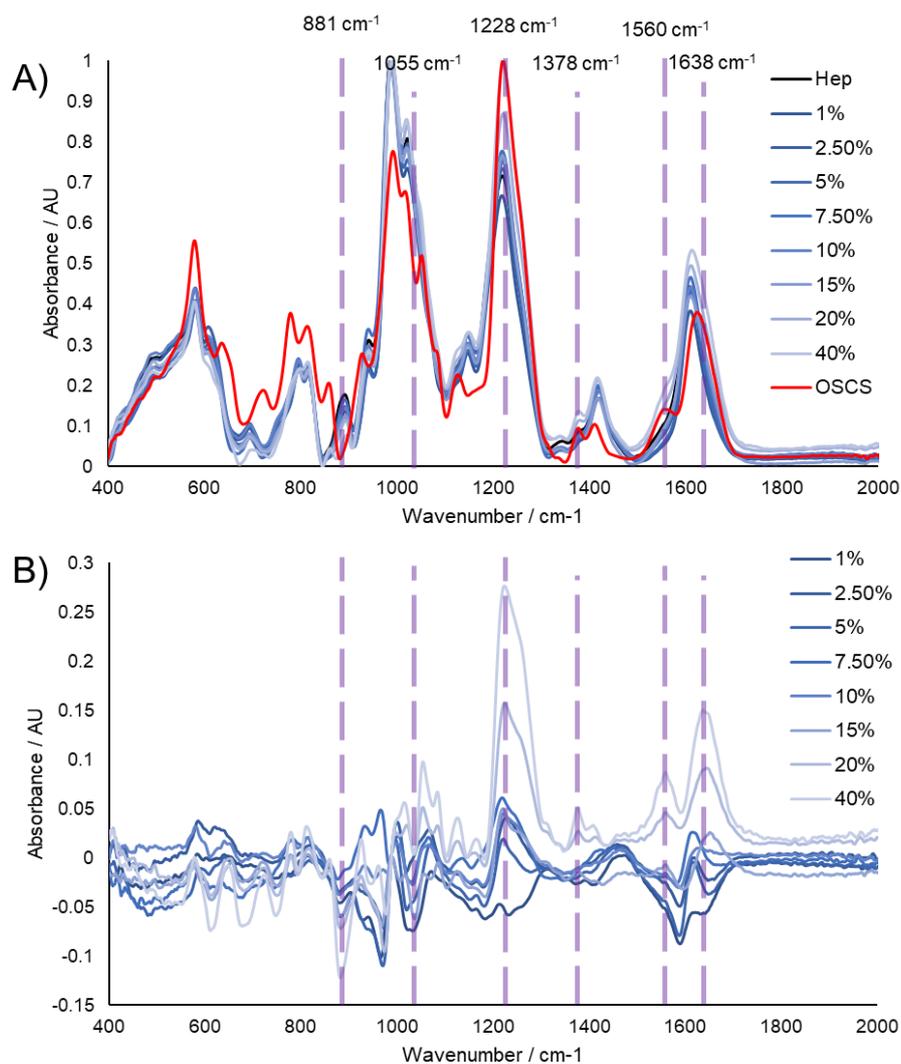
Figure 3.2: ATR-FTIR spectra and difference spectra of heparin and heparin contaminated with OSCS. A) ATR-FTIR spectra of heparin, heparin contaminated with OSCS to the levels of 1, 2.5, 5, 7.5, 10, 15, 20 and 40% w/w. **B)** Difference spectra of heparin vs heparin contaminated with OSCS to the same levels as panel A. Heparin: black, OSCS: red, percentage of OSCS in heparin: intensity of blue. Spectra are between 400 and 2000 cm^{-1} averages of 5 repeats of 3 scans and were smoothed, baseline corrected with a 7th order polynomial and normalised (0-1). Difference spectra were calculated based on the differences of each sample vs the heparin spectrum. Purple lines indicate the bands taken forward for MRA.

Based on the spectral differences observed, the bands which most strongly correlated with OSCS level by visual inspection were selected and the ability of them to predict the level of OSCS in a heparin examined. To do this, linear regression models were constructed based on band intensity and used as a predictive tool. An additional two *bona fide* heparins were blended with another two OSCSs at the same levels as the original series and 6 pharmaceutical heparins were randomly selected to serve as a 0% w/w point, their spectra acquired and their bands intensities at 881, 1055, 1228, 1378, 1560 and 1639 cm^{-1} recorded (**Fig. 3.2**, purple lines). The intensities of the bands at 1228 cm^{-1}, 1378 cm^{-1}, 1560 cm^{-1} and 1638 cm^{-1} increase linearly as a function of the level of OSCS present in the sample, with R^2 values of 0.98, 0.92, 0.85 and 0.95 respectively (**Fig. 3.3**). Interestingly, the band at 881 cm^{-1} decreases linearly as a function of OSCS contamination, however with a lower R^2 value of 0.87. At 881 cm^{-1}, the linear range is at OSCS levels of 7.5% to 40%, (w/w) with an R^2 value of 0.97 (**Fig. 3.3 E**, Orange line). The differences observed in the carbohydrate band at **~1000 cm^{-1}** correlated with the presence of OSCS but weakly compared to the other bands with the band at 1055 having an R^2 value of 0.77.
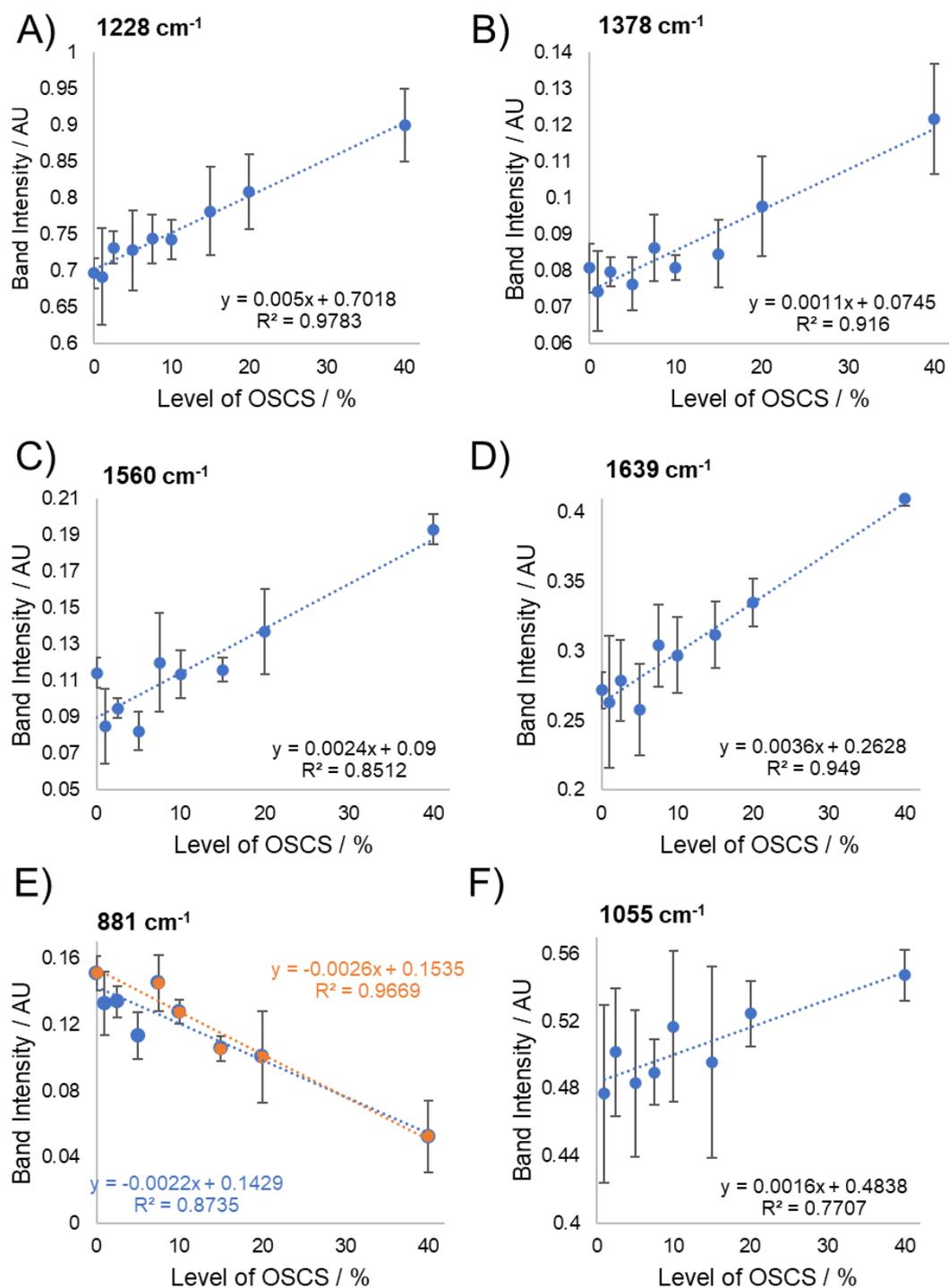
Figure 3.3: Linear regression models for the band intensities vs level of OSCS. A) Linear regression of band intensity at 1228 cm^{-1}, **B)** Linear regression of band intensity at 1378 cm-1, **C)** Linear regression of band intensities at 1560 cm^{-1}, **D)** Linear regression of band intensities at 1639 cm-1, **E)** Linear regression of band intensities at 881 cm^{-1}. Two lines were fitted to this plot, one including all samples (blue) and the other containing pharmaceutical heparin (0%) and heparins contaminated with OSCS from 7.5% up (orange), **F)** Linear regression of band intensities at 1055 cm^{-1}. Linear regression models were constructed from signal intensities at the relevant wavenumber against the level of OSCS expected in the sample.

These data suggest that ATR-FTIR can be used to detect OSCS contamination, and the level of OSCS contamination corresponds to a colinear change in a series of bands across the spectrum, particularly those that correlate with sulphation pattern and level across the sample.

[68]

However, when a library of pharmaceutical heparins are compared, the natural heterogeneity of heparin is often reflected in the intensities and positions of the aforementioned bands (**Fig. 3.4**), and thus a perfectly acceptable pharmaceutical heparin can show an apparent amount of OSCS that it doesn't contain. Using each of the above models, most pharmaceutical heparins are predicted to contain OSCS to a level as high as 25%. On average, a pharmaceutical heparin contains 1.06%, -0.81%, 7.27%, 2.81% and 0.78% (w/w) using each model (**Table 3.1**). An ideal model would correctly predict 0% of OSCS in pharmaceutical heparins, but the "best" model at 1228 cm^{-1} still detects up to 11.4% OSCS in heparins, suggesting a limit of detection (LOD) of ~15% (w/w) for a 100% true positive rate or around 7.5% (w/w) to be within one standard deviation of the mean.
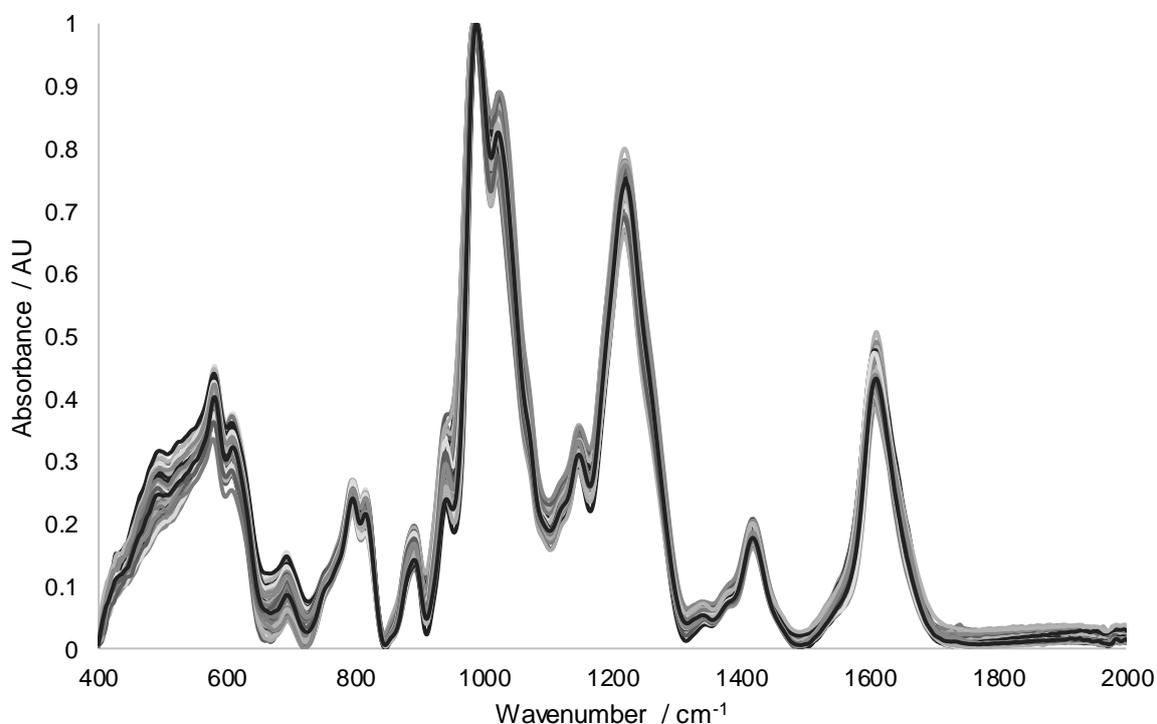
Fig 3.4: ATR-FTIR spectra of multiple pharmaceutical heparins. Spectra are between 400 and 2000 cm-1 averages of 5 repeats of 3 scans and were smoothed, baseline corrected with a 7th order polynomial and normalised (0-1).

Table 3.1: The prediction of OSCS in pharmaceutical heparin, using linear regression and MRA models.

Heparin	Predicted level of OSCS / %						
	Band Wavenumber / cm^{-1}					MRA	
	881.75	1228.79	1379.64	1560.95	1639.56	MRA (All)	MRA (881,1639)
G9850	-5.58	-5.54	12.30	11.06	0.67	-1.69	-2.72
P1	-4.68	-3.96	12.93	11.10	1.23	-1.19	-1.93
P3	4.76	5.99	7.05	4.04	-0.64	1.44	2.27
P7	-3.39	-2.80	3.15	8.43	-0.33	-0.23	-2.04
P8	-5.54	0.59	-0.13	14.03	7.29	4.04	0.81
P9	-4.47	-0.19	-0.95	11.96	7.07	4.36	1.28
P10	-2.08	1.69	2.75	10.86	8.51	5.36	3.37
P11	-11.86	-5.37	11.65	7.62	-0.25	-7.11	-6.66
P12	4.45	11.43	13.94	-1.27	5.57	0.86	5.39
P13	6.66	8.39	10.71	-2.30	3.50	2.07	5.51
P14	4.09	6.74	7.79	-0.97	2.66	1.15	3.65
P15	13.56	-5.82	-2.24	-5.94	-4.56	6.22	5.03
P16	13.33	-7.57	-1.24	-8.91	-6.69	4.24	3.78
P17	2.64	-6.80	11.84	-5.92	-4.96	-3.65	-1.18
P19	-2.68	1.35	12.36	11.95	6.91	2.60	2.19
P20	-3.25	-1.74	0.03	11.94	7.07	5.23	1.96
P21	-12.26	-8.59	7.21	15.59	1.96	-2.60	-5.71
P23	-1.28	-7.31	15.62	1.55	-0.02	-2.16	-0.72
P24	-3.87	-1.52	3.75	9.30	5.78	2.65	0.93
P25	1.25	1.72	3.80	2.66	0.87	0.97	1.14
P26	7.03	-9.18	2.46	-3.08	-2.97	3.29	2.29
P27	-10.84	-3.57	-0.77	9.99	4.87	-0.96	-3.38
P28	2.72	8.34	14.81	-1.28	2.41	-1.60	2.77
P29	2.58	-7.77	17.57	-8.14	-5.07	-5.29	-1.27
P30	8.27	-10.08	-2.41	-6.98	-7.15	1.88	0.76
P31	5.27	-1.79	3.95	0.86	-2.33	1.96	1.66
P32	4.64	5.97	6.47	0.32	0.13	0.88	2.61
P33	3.00	-4.22	-3.46	0.97	-2.81	2.11	0.16
P34a	10.22	5.56	4.26	-3.62	1.64	4.77	6.48
P34b	-7.42	-5.46	5.63	8.38	3.28	-0.93	-2.35
P35	5.86	1.96	2.33	-1.94	-0.63	2.16	2.89
P36	1.84	0.15	7.14	0.36	-1.22	-0.77	0.36
P37	-1.66	-0.31	5.45	5.91	3.05	1.17	0.70
P38	-10.54	-1.61	19.39	7.56	1.51	-7.48	-5.00
P39	3.20	-7.61	-0.03	1.26	-2.56	2.39	0.40
P40	5.33	-2.72	1.41	2.43	-0.10	4.28	2.87
P41	0.53	-2.45	19.88	-0.35	-0.67	-3.52	-0.06
P42	14.92	-15.02	-4.22	-9.96	-11.84	4.09	1.92
P43	-1.55	4.97	7.77	13.73	13.06	7.39	6.07
P44	12.08	-3.62	6.52	-9.44	-3.12	3.01	4.98
P48	2.57	9.00	16.16	0.36	2.10	-1.80	2.52
P49	-0.39	7.50	13.69	0.06	2.37	-2.92	1.04
P50	2.04	8.37	18.41	1.89	0.08	-3.14	1.16
P51	-2.00	-2.41	7.48	6.71	-2.16	-1.68	-2.25
P52	0.36	8.80	24.86	3.89	1.75	-4.12	1.12
Avg	1.06	-0.81	7.27	2.81	0.78	0.62	1.00
St Dev	6.59	6.20	7.14	6.96	4.59	3.48	3.01
Max	14.92	11.43	24.86	15.59	13.06	7.39	6.48
Min	-12.26	-15.02	-4.22	-9.96	-11.84	-7.48	-6.66

Generally, a heparin with an incorrect percentage of OSCS in one model had a more reasonable or accurate percentage in another, hence it was hypothesised that a combination of the models may cancel out some of the extreme values and therefore be sufficient to improve OSCS detection, hence a multiple linear regression (MLR) analysis was performed, as MLR is a linear regression in n dimensions, instead of 2. The size of the gradient in the direction of each factor (dubbed the coefficient) can be used to determine the importance of each factor and, with associated calculated P-values, can be used to determine if the measured models are significant or not. Multiple linear regression was performed with the dependant variable set to level of OSCS and the values in each model input as independent variables. The MLR had an R^2 of 0.95, and when converted to a calibration curve (a linear combination of the values vs OSCS % (w/w)), had an R^2 value of 0.99 (**Fig. 3.5**). The MLR model showed an average level of OSCS of 0.62% (w/w) +/- 3.5%. The highest false positive was 7.4%, suggesting a LOD of 7.5% (w/w) for a 100% true positive rate or ~ 5% to be within 1 standard deviation of the mean.

$$y = 0.9569x + 0.33$$
$$R^2 = 0.9911$$
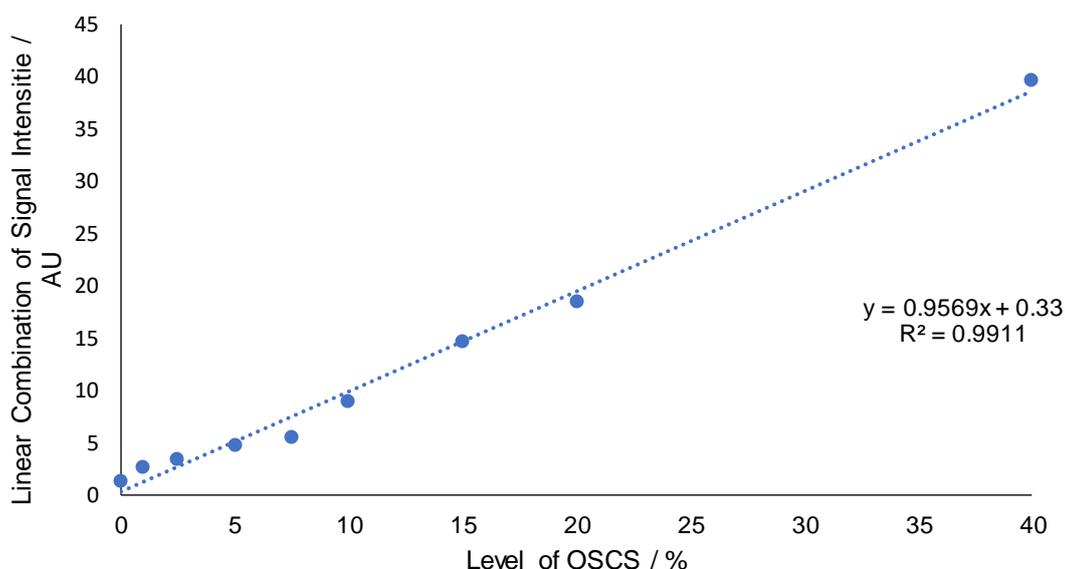
Figure 3.5: MLR of band intensities at bands 881, 1055, 1228, 1378, 1560 and 1639 cm^{-1}. The linear combination is the sum of intensities at the relevant band multiplied by the coefficient for that band calculated in the MLR.

Of note, the values at 881 and 1640 cm^{-1} were shown to be the most important in determining the percentage contamination of OSCS, with coefficients of -239.98 and 143.58, and P-values of 1.54×10^{-8} and 0.014 respectively; with the other values being insignificant to the model. Hence a new MLR was constructed, using only the values at 881 and 1639 cm^{-1} (**Fig. 3.6**). This model had a lower R^2 of 0.92, and when the linear combination, calculated from the computed coefficients and intensity values is used, it had an R^2 of 0.99 (**Fig. 3.6 B**). Both models had P-values of 2.07×10^{-8} and 5.69×10^{-9} each. The iterated MLR gave the closest for predicting OSCS%, with an average level of OSCS of 1% +/- 3.0% (w/w) and the highest

false positive of 6.5%, suggesting a LOD of 7.5% (w/w) for 100% true positive and ~ 3% (w/w) to be within 1 standard deviation of the mean.
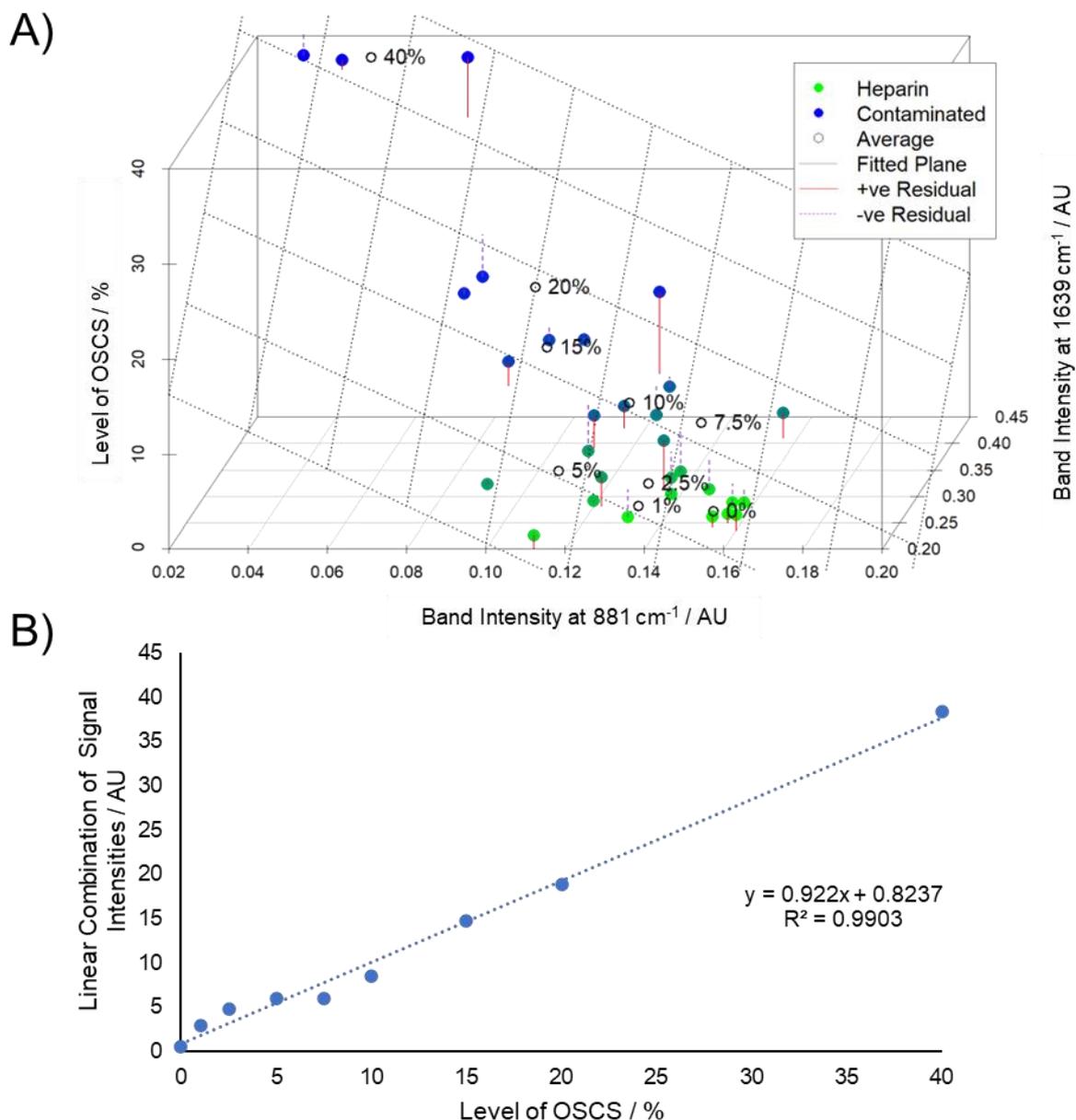
Figure 3.6: MLR of band intensities at bands 881 and 1639 cm^{-1}. A) 3D plot of the level of OSCS vs both the band intensities at 881 cm-1 and 1639 cm-1. **B)** Linear combination of the intensities and coefficients vs the level of OSCS. The linear combination is the sum of intensities at the relevant band multiplied by the coefficient for that band calculated in the MLR.

3.3.2 Applying logistic regression to heparin QC with FTIR.

To better define the LOD, a logistic regression model (LRM) was created from the studied wavelengths, whereby a sigmoidal response is plotted against the probability that a sample is contaminated (all contaminated samples P = 1) and the probability that the sample is not (P = 0). The analyst can then determine a probability threshold based on the fit curve, below which the sample is likely not contaminated, and above which it is. The threshold can be decided

based on how important detection is deemed to be, and is often decided via a response operator characteristic (ROC) curve. The LRM was calculated iteratively, using calculated P-values and comparison of AIC values to remove insignificant bands. Ultimately all wavelengths bar 1640 cm^{-1} were required (P values 3.85×10^{-8}, 3.06×10^{-11}, 1.91×10^{-5} and 7.92×10^{-7} for 881, 1229, 1379 and 1560 cm^{-1} respectively, AIC = 263.06) for optimal separation (**Fig 3.7**). The R^2 value was low, at 0.53 but the model was still trialled. In order to have 100% detection of contaminated samples, 80% of the pharmaceutical heparins would be misclassified and for a ~ 90% detection rate, around 20% would be misclassified (**Fig. 3.7 B**). A 100% true positive rate is essentially unobtainable with the current fit, as it requires a threshold set in the asymptotic region. Samples of 5% (w/w) contamination also fall into the asymptotic region, while samples at the level of 7.5% (w/w) OSCS fall to the right side of the linear portion, and therefore can be classified correctly, hence the LOD of OSCS contamination using the selected wavelengths for a simple yes/no classifier is 7.5% (w/w).
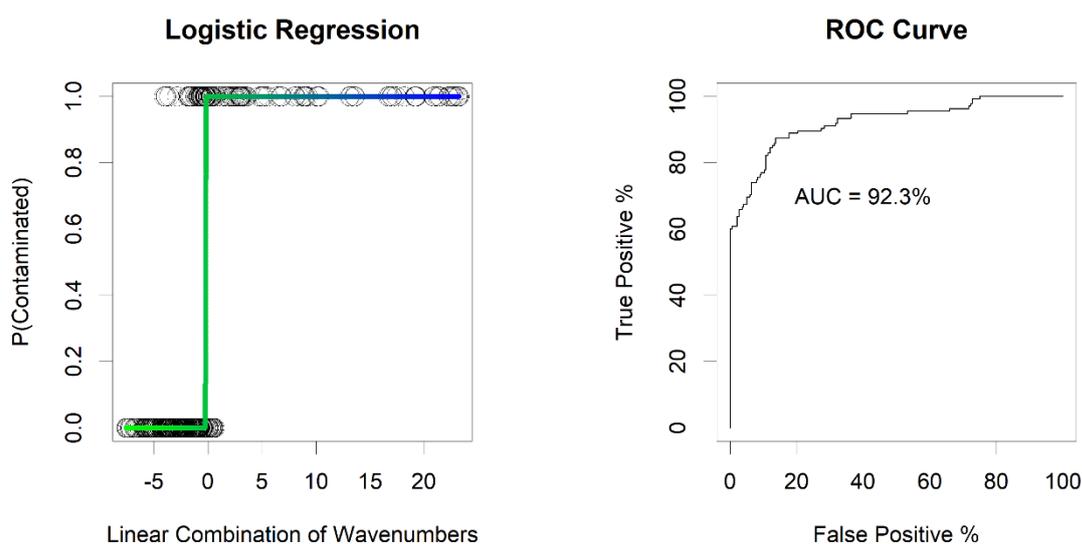
Figure 3.7: Logistic regression of the linear combination of the band intensities at bands 881, 1055, 1228, 1378 and 1639 cm^{-1}. Left) Logistic response curve for a sample that contains OSCS (y = 1) or a pharmaceutical heparin (y = 0). **Right)** ROC curve of the responses.

A LOD of 7.5% (w/w) is a marked increase in the sensitivity of detection from the LOD of the original model at 15% (w/w) but is still not at the ideal value of 3% (w/w) for pharmaceutical heparin QC. Improvements to the LOD have been shown above to correspond to the models used. The initial models were selected based on their perceived importance (the magnitude of the change) in calculated difference spectra, but it is not necessarily the largest change that corresponds to the best fit. To further improve the LOD, new models could be built and trialled. In theory, an MLR could be performed across the entire spectrum, containing 3391 models - one for every measured wavenumber. In reality however, generating the model and analysing it would be too laborious, and it is likely that most wavenumbers will not correlate with OSCS%,

and the large number of independent variables will make analysis difficult – typically, 10 variables are the maximum for modern computer algorithms.

3.3.3 The use of dimensionality reduction to improve the selection of model wavenumbers.

To better select model wavenumbers, a reduction in dimensionality is required, whereby the number of wavenumbers is reduced for further modelling. Spectral region deletion and bucketing could be used, and subsequent difference spectra, linear and MLR models constructed, but this is similarly laborious, and forces the analyst to select subjectively "important" regions which may not correlate with OSCS%. Ideally, the dimensions would be reduced to a set that corresponds to an underlying trend in the data – preferably one that correlates with the percentage of OSCS. The simplest and most routinely used method of dimension reduction is principal component analysis (PCA). PCA has also been applied to NMR spectra, as a means of distinguishing OSCS and OSCS contaminated samples from heparin (Rudd et al., 2011a).

A library of 45 heparins, all the heparins that had been contaminated with OSCS to the various levels here, and raw OSCS were taken and prepared in the same way as above. In brief, the samples were baseline corrected with an 7th order polynomial and normalised between 0 and 1. The variable air regions, where bands that can be attributed to CO_2 and H_2O found in air (< 700 cm^{-1}, between 2000 and 2500 cm^{-1}, and >3600 cm^{-1}) were removed and the data were mean centred. PCA was performed on this data matrix, with the variables set to the wavenumbers, using the *prcomp* function in R. PCA reduces the data set to a series of linear combinations, calculated based on the computed eigenvectors (the directions in which the linear combination reflects) and eigenvalues (the magnitude and therefore perceived importance of the direction) of the covariance matrix of the input data matrix, which group and separate samples based on the largest variance throughout. The eigenvectors are new directions from which the analyst can observe the data and explain underlying trends within the data. The relative importance, i.e the amount of variance covered in each component, is correlated with the magnitude of the eigenvalue. The % (w/w) of the variance covered is calculated here as the percentage that each eigenvalue covers given the sum of all the eigenvalues calculated. Generally, the first *n* components whose variances sum to cover the majority of the variance (usually >85%) are examined and the remaining components discarded.

Following PCA, PC1 through 3 accounted for 54.39%, 19.09% and 13.78% of the variance in the data respectively (**Fig. 3.8**). The three components together accounted for 87.26% of the total variance, which is considered to cover most of the variance in the data set (~85%) and should therefore account for the majority of the major trends within. Through interrogation of

the score plots, pure OSCS was seen to strongly separate from the main heparin library in PCs 2 and 3, with PC3 providing the most distinction (**Fig. 3.8 B**). PC2 also separates contaminated samples from pharmaceutical heparin, with heparins contaminated with 40% and 20% (w/w) OSCS seen most distinct from the main heparin cloud (**Fig. 3.8 A**). Little to no separation is observed for 15-1% (w/w) however.

The score plot for PC3 (**Fig. 3.8 H**) when compared with OSCS and the parent heparin (**Fig. 3.8 J**) clearly shows the major differences between OSCS and heparin that were observed in the difference spectra in **Figure. 3.2 B**, namely the extra eruption at 1050 cm^{-1} (or rather, the trough at 1036 cm^{-1} formed between this eruption and the main carbohydrate band, reduced band breadth and increased intensity in the sulphate band at 1200 cm^{-1}, particularly at 1190 cm^{-1} and the change in the C=O band at 1600 cm^{-1}, accounting for the strong separation observed. A strong separation is also observed due to the band at 973 cm^{-1} - a band which was not highlighted within the difference spectra. Small contributions from the bands at 726, 780, 859, 881, 1008, 1152 and 1427 cm^{-1} are also observed.

The loading plot for PC2 (**Fig. 3.8 G,I**), which separates OSCS, but also separates highly contaminated samples, separates by the same bands, but at slightly different locations; namely the eruption at 1050 cm^{-1}, the sulphate band at 1228 cm^{-1} and the C=O band at 1640 cm^{-1}. Other minor contributors include the bands at 783, 861, 881, 927, 959, 986 and 1087 cm^{-1} amongst others. The bands at 881 and 1640 cm^{-1} are highlighted here, as expected from the MLR model constructed earlier (**Fig. 3.6**) . The sulphate band at 1228 cm^{-1} is the most important predictor of OSCS contamination in PC2, despite being determined as less significant in the MLR model. This is most likely due to the prominence of the sulphate band in highly contaminated samples, which are separated here, but not in less contaminated samples.

The PCA highlighted new bands of interest which could be used to create a new model, but it is likely that these bands separate solely for > 20% (w/w) contaminated samples, as the less contaminated samples are not separated, and hence may not provide the increased LOD sought. In an attempt to increase the LOD a PCA was performed with the same sample preparation and the same library, minus the pure OSCS samples (**Fig 3.9**). Essentially the same score plot was acquired, but PCs 1 and 2 covered 84.36% of the total variance. PC2 acted similarly to PC2 in the original PCA, separating > 20% (w/w) contaminated heparin samples from the heparin library at virtually the same bands. No increase in separation was observed.
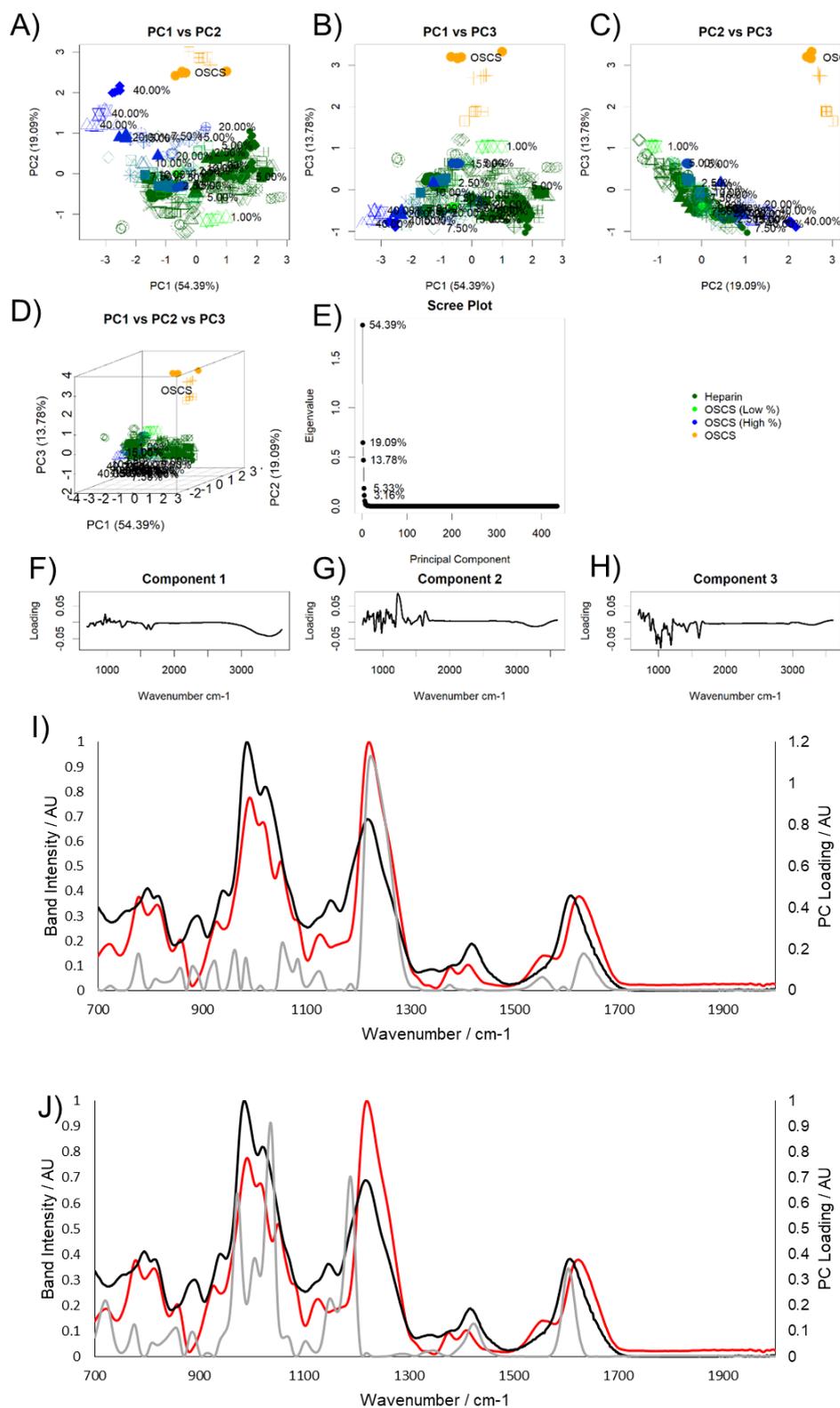
Figure 3.8: PCA of a heparin library against heparins contaminated with OSCS and OSCS alone. A) Scatter plot of PC1 vs PC2 scores. **B)** Scatter plot of PC1 vs PC3 scores. **C)** Scatter plot of PC2 vs PC3 scores. **D)** 3D scatter plot of PC1 vs PC2 vs PC3 scores. **E)** Scree plot. **F)** Loading plot of PC1. **G)** Loading plot of PC2. **H)** Loading plot of PC3. **I)** Comparison of PC2 loadings (grey) with ATR-FTIR spectra of heparin (black) and OSCS (red). **J)** Comparison of PC3 loadings (grey) with ATR-FTIR spectra of heparin (black) and OSCS (red). For PCA scatter plots, samples contaminated with OSCS are indicated with a scale of green to blue – the more blue the sample the higher the level of OSCS present. Pharmaceutical heparins are dark green and OSCS is orange. All spectra involved in PCA are in the regions of 700 and 2000 and, 2500 and 3600 cm^{-1} and are averages of 3 scans. The spectra were smoothed, baseline corrected with a 7th order polynomial and normalised (0-1). Spectra in I and J are averages of 5 repeats.
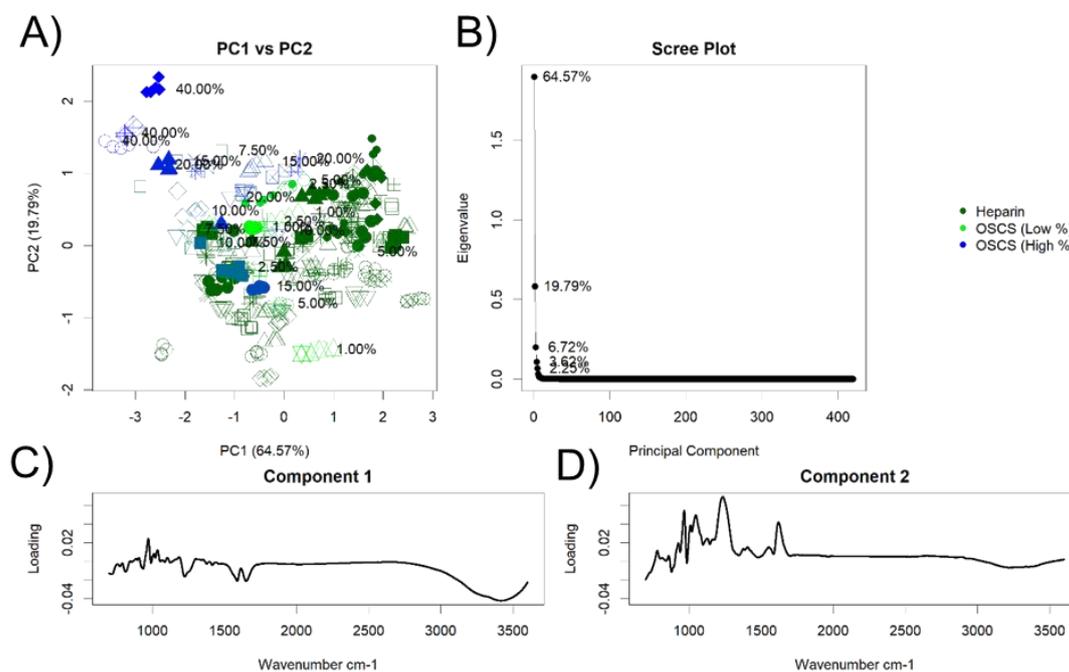
Figure 3.9: PCA of a heparin library against heparins contaminated with OSCS. A) Scatter plot of PC1 vs PC2 scores. **B)** Scree plot. **C)** Loadings for PC1. **D)** Loadings for PC2. Heparin: dark green, OSCS: orange, samples contaminated with OSCS are indicated with a scale of green to blue – the more blue the sample the higher the level of OSCS present. All spectra involved in PCA are in the regions of 700 and 2000 and, 2500 and 3600 cm-1 and are averages of 3 scans. The spectra were smoothed, baseline corrected with a 7th order polynomial and normalised (0-1).

Upon closer observation of the PC score plots, it was noted that some of the repeats of the same sample were separating from each other. Separations such as these are normally observed when the variable regions, which are susceptible to vibrations from atmospheric H_2O and CO_2 are not removed, and the lack of clustering here suggested that noise may be interfering with the PCA scores, hence the effects of smoothing on the PCA was assessed. At the start of the process, the spectra undergo a Savitzky Golay smooth, with a 2nd order polynomial to the nearest 21 neighbours, as part of a pre-treatment process to prevent artefacts introduced during the baseline fitting process and all smoothing performed from here are performed on already smoothed data. Care was taken to not smooth the data in such a manner that bands merge or lose their shape through inspection of the smoothed spectra.

3.3.4 Optimising the preparation of IR spectra for the PCA approach.

There are no hard and fast rules about how much and when to smooth, and the amount of smoothing is often at the analyst's discretion (van den Berg et al., 2006). In order to retain consistency in the matrix, all spectra were smoothed to exactly the same degree, custom smoothing of noisy samples may interfere with final results, and an overly noisy spectrum is more likely due to an error in acquisition. Ideally, as little smoothing as possible will be used to retain as much information in the spectra as possible. Smoothing between **21** and **41** neighbours with a 2nd or 4th order polynomial was deemed ideal based on the raw spectra as the general shape of the original spectrum is not lost (**Fig. 3.10**).
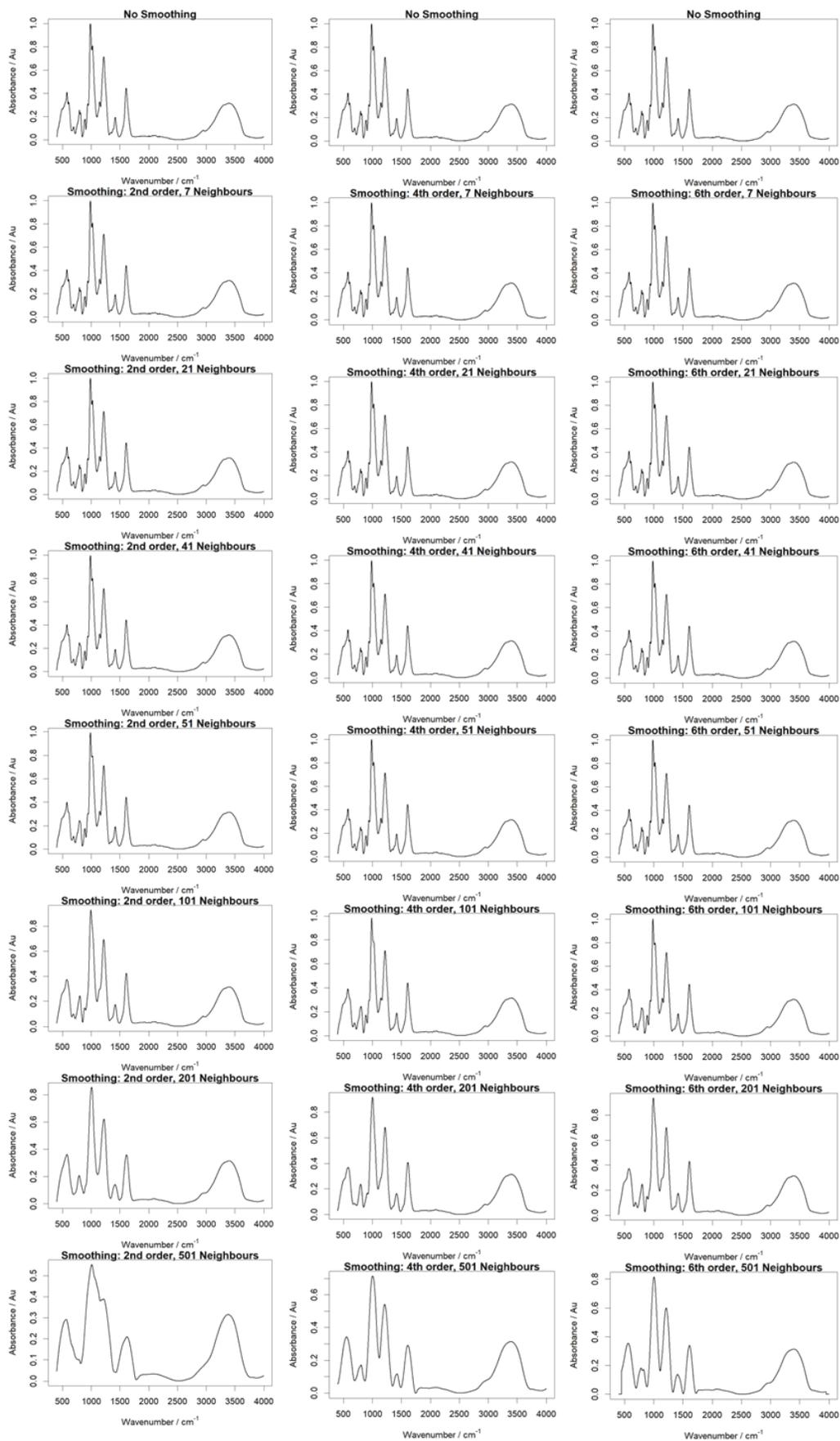
Figure 3.10: The effects of smoothing with different parameters on ATR-FTIR spectra of heparin. Spectra are first baseline corrected with a 7th order polynomial and then normalised (0-1) before being subjected to smoothing with the Savitzky-Golay algorithm to different parameters indicated on each plot.

PCA of data that had been smoothed to a 2nd and 4th order polynomial and at 21 and 41 neighbours were performed (**Figs. 3.11 to 3.14**) and while there was a very slight increase in separation with samples of 20% and 15% (w/w) contamination moving slightly further from the main heparin cloud, overall, there was little to no difference in the produced plots. Small variations in the eigenvalues for each component were observed (0.02 – 0.37% in PC1) and the loadings were essentially the same, suggesting that smoothing was not the issue observed.
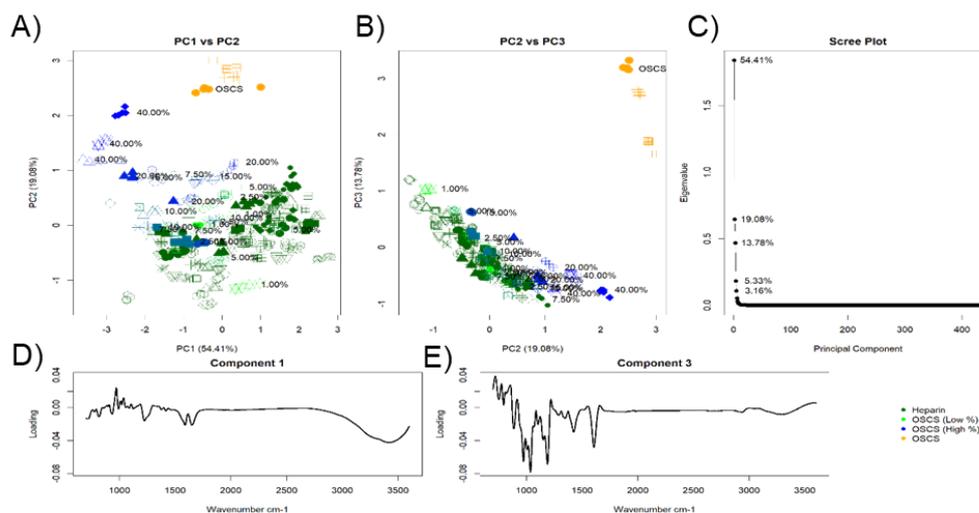
Figure 3.11: PCA of a heparin library and heparins contaminated with OSCS, after subjecting the spectra to Savitzky Golay smoothing using a 2nd order polynomial and 21 neighbours. A) Scatter plot of PC1 vs PC2 scores. **B)** Scatter plot of PC2 vs PC3 scores. **C)** Scree plot. **D)** Plot of the loadings for PC1. **E)** Plot for the loadings for PC2. Heparin: darkgreen, OSCS: orange, samples contaminated with OSCS are indicated with a scale of green to blue – the more blue the sample the higher the level of OSCS present. All spectra involved in PCA are in the regions of 700 and 2000 and, 2500 and 3600 cm^{-1} and are averages of 3 scans. The spectra were smoothed, baseline corrected with a 7th order polynomial and then smoothed with a Savitzky-Golay algorithm using a 2nd order polynomial and 21 neighbours and then normalised (0-1).

Figure 3.12: PCA of a heparin library and heparins contaminated with OSCS, after subjecting the spectra to Savitzky Golay smoothing using a 2nd order polynomial and 41 neighbours. A) Scatter plot of PC1 vs PC2 scores. **B)** Scatter plot of PC2 vs PC3 scores. **C)** Scree plot. **D)** Plot of the loadings for PC1. **E)** Plot for the loadings for PC2. Heparin: darkgreen, OSCS: orange, samples contaminated with OSCS are indicated with a scale of green to blue – the more blue the sample the higher the level of OSCS present. All spectra involved in PCA are in the regions of 700 and 2000 and, 2500 and 3600 cm^{-1} and are averages of 3 scans. The spectra were smoothed, baseline corrected with a 7th order polynomial and then smoothed with a Savitzky-Golay algorithm using a 2nd order polynomial and 41 neighbours and then normalised (0-1).

Figure 3.13: PCA of a heparin library and heparins contaminated with OSCS, after subjecting the spectra to Savitzky Golay smoothing using a 4th order polynomial and 21 neighbours. A) Scatter plot of PC1 vs PC2 scores. B) Scatter plot of PC2 vs PC3 scores. C) Scree plot. D) Plot of the loadings for PC1. E) Plot for the loadings for PC2. Heparin: darkgreen, OSCS: orange, samples contaminated with OSCS are indicated with a scale of green to blue – the more blue the sample the higher the level of OSCS present. All spectra involved in PCA are in the regions of 700 and 2000 and, 2500 and 3600 cm^{-1} and are averages of 3 scans. The spectra were smoothed, baseline corrected with a 7th order polynomial and then smoothed with a Savitzky-Golay algorithm using a 4th order polynomial and 21 neighbours and then normalised (0-1).

Figure 3.14: PCA of a heparin library and heparins contaminated with OSCS, after subjecting the spectra to Savitzky Golay smoothing using a 4th order polynomial and 41 neighbours. **A)** Scatter plot of PC1 vs PC2 scores. **B)** Scatter plot of PC2 vs PC3 scores. **C)** Scree plot. **D)** Plot of the loadings for PC1. **E)** Plot for the loadings for PC2. Heparin: darkgreen, OSCS: orange, samples contaminated with OSCS are indicated with a scale of green to blue – the more blue the sample the higher the level of OSCS present. All spectra involved in PCA are in the regions of 700 and 2000 and, 2500 and 3600 cm^{-1} and are averages of 3 scans. The spectra were smoothed, baseline corrected with a 7th order polynomial and then smoothed with a Savitzky-Golay algorithm using a 4th order polynomial and 41 neighbours and then normalised (0-1).

As PCA is a projection onto a new multivariate space, spectra can be reconstructed from the PCs. If all PCs are taken, the original spectrum is recreated perfectly, but through removal of PCs that do not correspond to the trends which relate to the desired outcome, the analyst can remove "unimportant" data, and reconstruct a spectrum that highlights the differences observed between the samples, based on the selected components. Techniques, such as principal component regression (PCR) – a MLR performed on PCs, exist to determine which

PCs are important and those that are not (Massy, 1965). But here, since it is clear that PCs 2 and 3 separate the desired groups, the spectra were reconstructed from only these two components. Reconstruction involves calculating the product of the scores matrix and the transverse of the loadings matrix for the components to be used and addition of the mean that was subtracted during mean centring (**eq. 3.3.1**).

$$Y = SL^T + \mu$$

Reconstruction of the spectra highlights the major signals for contaminated samples (**Fig. 3.15**). The band at wavenumber 1228 cm^{-1} is the most apparent, but is only observed in samples with very high contamination (>15%) The other, less contaminated samples are shifted to a lower wavenumber at ~1190 cm^{-1}. The fact that only highly contaminated samples, and not lowly contaminated samples, are significantly varied at 1228 cm^{-1} may explain why this band was shown to not correlate with the level of OSCS in the MLR analysis (**Fig. 3.5**) and explain why samples below 15% (w/w) are not separated in the PCA. The heparins shown in this plot are generally featureless, showing only shallow troughs around wavenumbers 900-1000, 1200 and 1600 cm^{-1}. The bands at 881 and 1640 cm^{-1} have a good spread across the different levels of OSCS and the heparins. An observation made due to the PCA and the reconstructed spectra, is the dense amount of information in the band(s) between 900 and 1100 cm^{-1}. These bands were essentially left out after inspection of the difference spectra (**Fig. 3.2 B**), due to incoherent band intensities and locations between different levels of OSCS%.

The loading plot for PC2 (**Fig. 3.14E**), the component that separates the contaminated samples from heparins the best, shows strong signals at wavenumbers 880, 987, 1228 and 1640 cm^{-1}, with the signals at 880 and 987 cm^{-1} correlating with amount of pharmaceutical heparin while the signals at 1228 and 1640 cm^{-1} have strong positive values and hence correlate with the amount of OSCS. Despite the perceived lack of separation for less contaminated samples, the PCA highlighted that more information can be gleaned from the complex carbohydrate band at ~ 1000 cm^{-1}.

3.3.5 The optimisation of 2nd derivative IR spectra for PCA QC of heparin.

A common technique used in IR spectroscopy is second derivative spectroscopy, whereby complex, broad and overlapping bands undergo derivatisation, highlighting the points of inflection through the spectrum, and therefore revealing the locations of bands lying beneath the major band. Second derivative spectroscopy has been used to calculate the macromolecular content of cartilage (Rieppo et al., 2012; Yin and Xia, 2010), the purity (Byler et al., 1995) and secondary structure (Susi and Michael Byler, 1983) of proteins, and to

separate overlapping carbonyl signals found in polyphenols (Zumbühl et al., 2017). Second derivative spectra have also been applied to Raman spectroscopy of GAGs, aiding detection of GAGs in cell lines (Mohamed et al., 2017) and has been used to monitor the different sulphation patterns of fucosylated sea cucumber chondroitins (Myron et al., 2017). Derivatisation has also improved the groupings and separations in FTIR-PCA of chondroitins (Foot and Mulholland, 2005). Put simply, the second derivative of the spectrum is calculated and these spectra are compared instead of the original. It was hypothesised that the second derivative may reveal some of the underlying features of the complex band at 1000 cm^{-1}. As derivatives highlight changes in gradient, random noise, which essentially amounts to many, informationless bands can greatly affect the outcome. Generally, derivatisation requires smoothing, as it was already demonstrated that smoothing to a 2nd order polynomial and 21 neighbours using the Savitsky-Golay algorithm resulted in little change to the PCA in **Fig. 3.11** and did not result in reduced noise but retained signal shape (**Fig. 3.16**), this level of smoothing was taken forward. The second derivative may also be calculated as a by-product of the Savitsky-Golay algorithm, hence this smoothing method and algorithm were used.
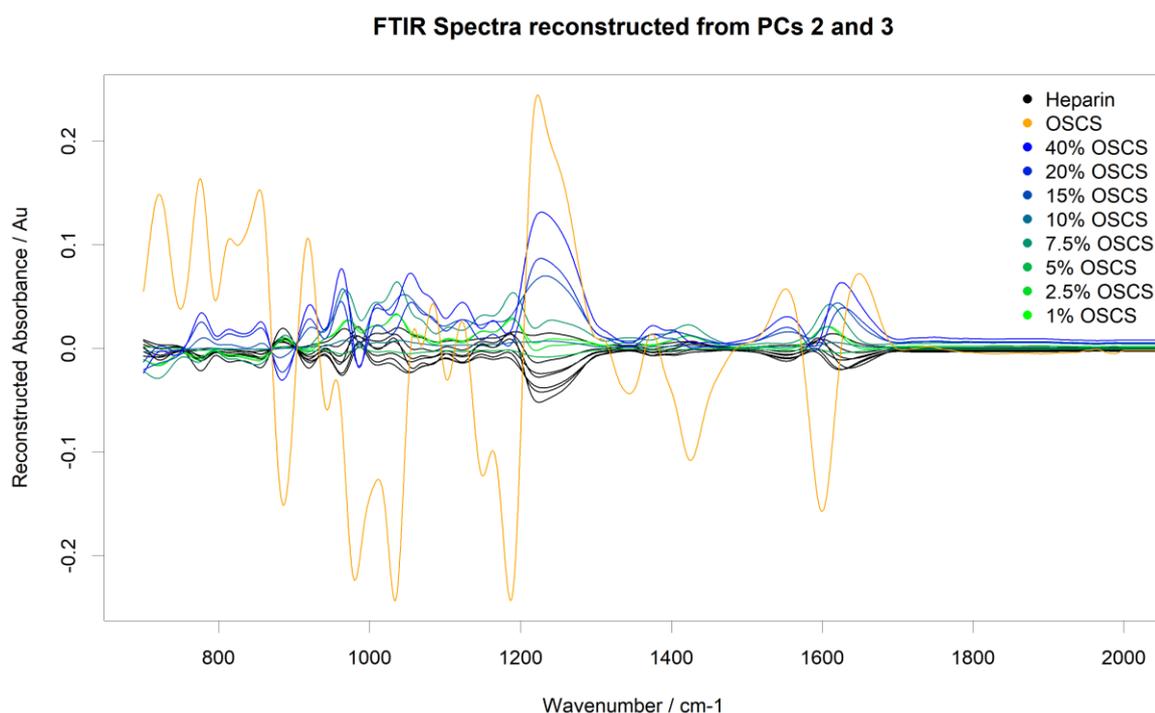
Figure 3.15: ATR-FTIR spectra of heparins that have been reconstructed from PCs 2 and 3. Heparin: black, OSCS: orange, samples contaminated with OSCS are indicated with a scale of green to blue – the more blue the sample the higher the level of OSCS present.
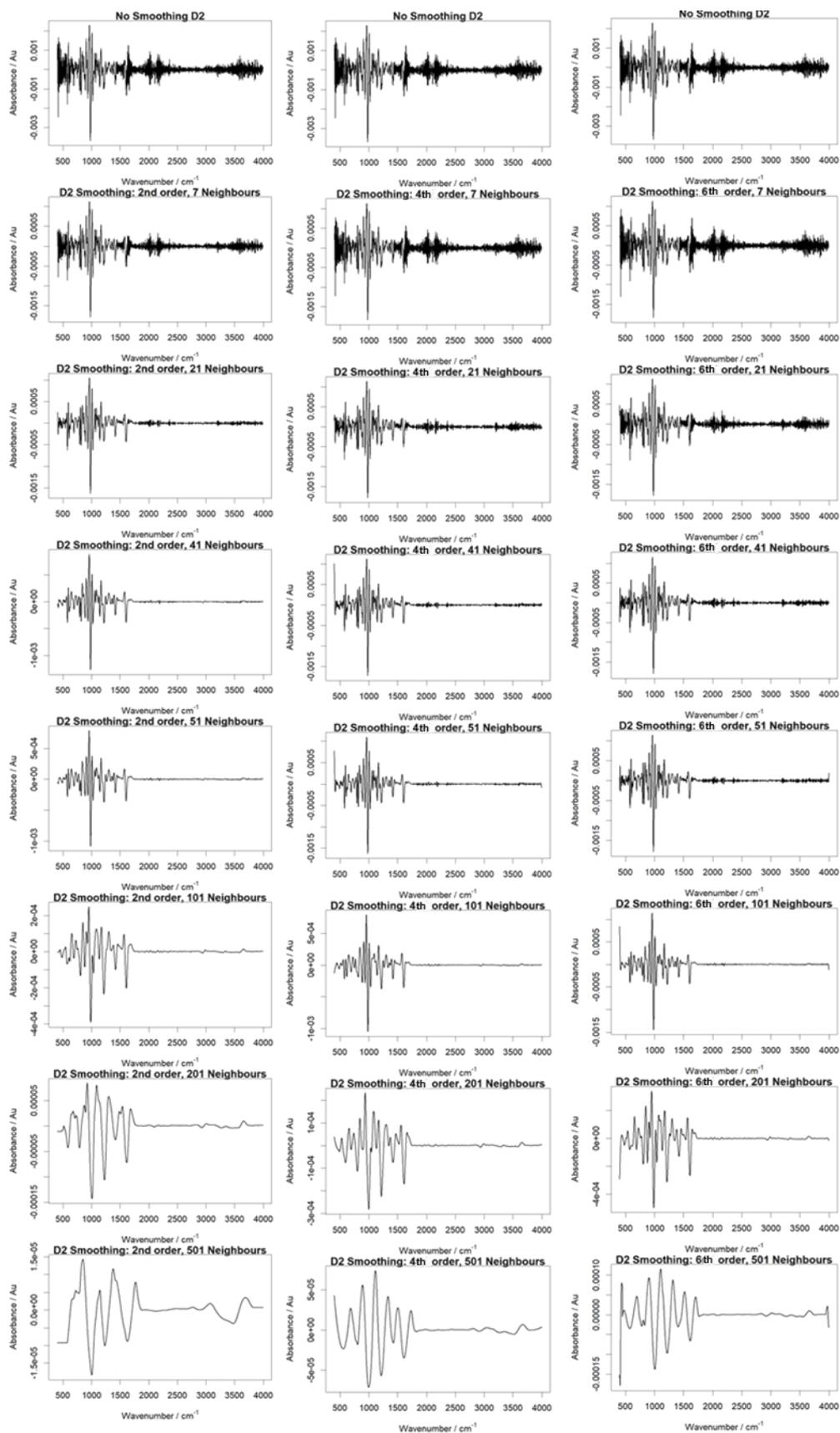
Figure 3.16: The effects of smoothing with different parameters on second derivative ATR-FTIR spectra of heparin. The spectra were first baseline corrected with a 7th order polynomial, normalised (0-1) before being subjected to smoothing with the Savitzky-Golay algorithm to different parameters indicated on each plot. The Savitsky-Golay algorithm was also used to calculate the second derivative of the spectrum post smoothing.

Upon calculation and comparison of the second derivative spectra (**Fig. 3.17 top**) and their difference spectra (**Fig. 3.17 bottom**), the complexity of the band at a wavenumber of 1000 cm^{-1} was highlighted and immediately, numerous clear and defined regions for exploration of the percentage of OSCS became apparent. Signals at wavenumbers 737, 767, 787, 802, 872, 902, 932, 972, 991, 1026, 1046, 1065, 1077, 1092, 1112, 1139, 1157, 1179, 1203, 1585, and 1615 cm^{-1} all appear visually to correlate with the percentage of OSCS. An MLR could be performed for these signals to determine which are relevant to the percentage of OSCS, but MLR tends to struggle with >10 variables, hence PCA was used instead.
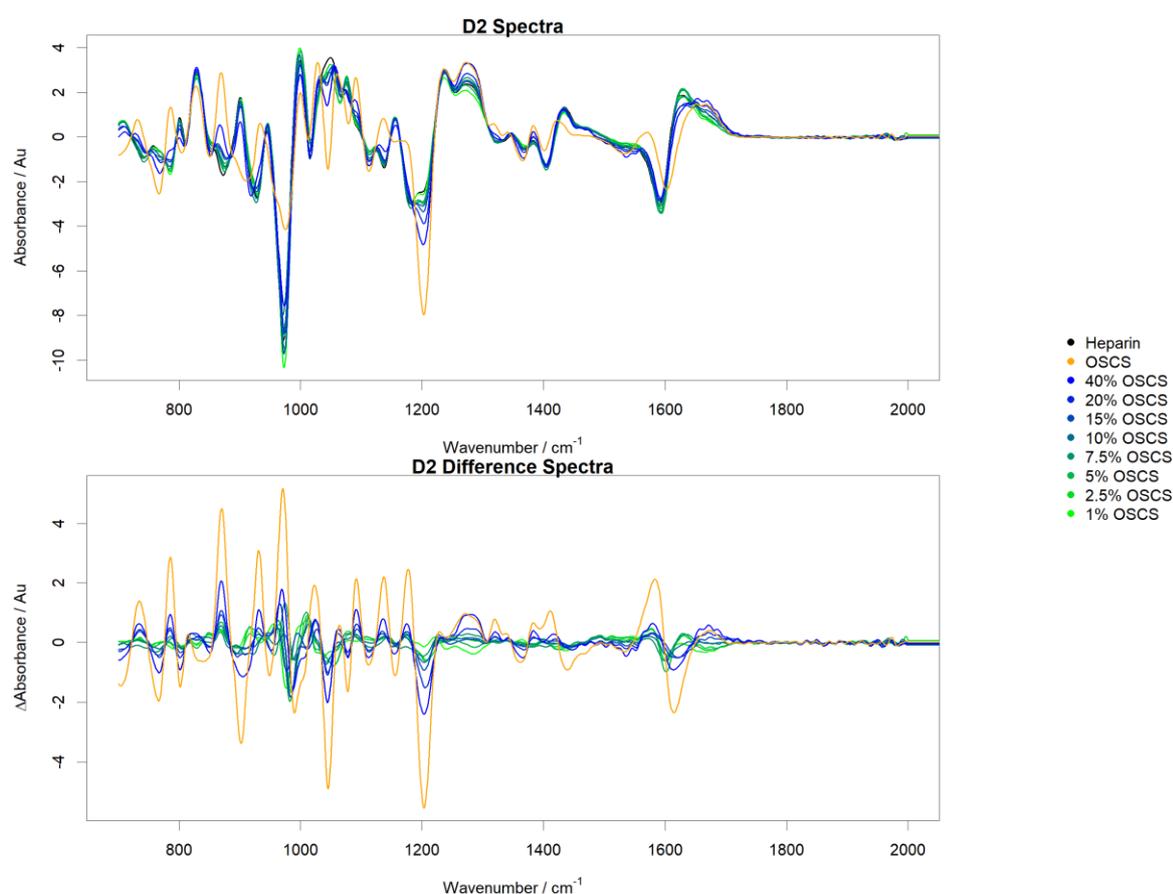
Figure 3.17: Second derivative ATR-FTIR spectra and difference spectra of heparin, OSCS and heparin contaminated with OSCS. Top) Second derivative ATR-FTIR spectra of heparin, heparin contaminated with OSCS to the levels of 1, 2.5, 5, 7.5, 10, 15, 20 and 40% w/w. **Bottom)** Difference spectra of second derivative ATR-FTIR spectra of heparin vs heparin contaminated with OSCS to the same levels as panel A. Heparin: black, OSCS: red, percentage of OSCS in hepain: intensity of blue. Spectra are between 400 and 2000 cm^{-1} averages of 5 repeats of 3 scans and were smoothed, baseline corrected with a 7th order polynomial and normalised (0-1), before smoothing and derivatisation with the Savitsky-Golay algorithm. Difference spectra were calculated based on the differences of each sample vs the heparin spectrum.

The PCA of the differentiated spectra displayed strong separation of the contaminated samples and OSCS (**Fig. 3.18**). A PCR was performed on the top 7 components (the number of components before the eigenvalue was ≈ to all lower values) and the top 6 components were shown to be statistically relevant to the percentage of OSCS. PC1, 2, 3 and 5 were the most relevant with P values of 9.69 $x10^{-11}$, 2.00$x10^{-10}$, 0.139$x10^{-3}$ and 0. 338$x10^{-3}$ respectively.

Through visual inspection, PC1, 3 and 5 correlate with the percentage of OSCS while PC2 appears to account for some natural variation within the heparin library. Iterative MLR of these components revealed that PC1 and 3 are the two significant components, however PC1 is by far the most significant, with a P-value of $< 2 \times 10^{-16}$, compared with PC3 at 0.0152. A plot of PC1 (**Fig. 3.18 H**) and the linear combination of PC1 and 3 (**Fig. 3.18 I**) scores against the percentage of OSCS reveal excellent correlations between the two (R^2 = 0.99 and 0.99 respectively), with the best R^2 value coming from PC1 alone at 0.9921.

In an effort to further increase the separation seen in the PCA plots, autoscaling was applied to the differentiated matrix prior to PCA (**Fig. 3.19**). Autoscaling attempts to make all data points in the spectrum equally important, preventing the PCA from discounting subtle but important perturbations in the spectra (van den Berg et al., 2006). As with smoothing, there are no hard and fast rules about when to use scaling and the use and type are at the discretion of the analyst (Rudd et al., 2011a). Autoscaling was used to distinguish between contaminated heparin and normal heparin with NMR (Rudd et al., 2011a), hence it was trialled here. The separation was further increased and immediately more apparent in PC1. PCR revealed that PC1, 4, 5 and 6 were correlated strongly with the percentage of OSCS (R^2 0.99). Upon visual inspection, PCs 5 and 6 separated the entirety of the first OSCS contamination series, irrespective of the other two series. PC 5 and 6 had comparatively low P-values of 0.553×10^{-3} and 4.369×10^{-3}, compared with that of PC1 and PC4 with P-values of 4.22×10^{-6} and 5.94×10^{-5}, respectively; hence PCs 5 and 6 were not considered further. Further PCR of PC1 and PC4 confirmed significance, with P-values of $<2 \times 10^{-16}$ and 0.573×10^{-3} respectively. Using PC1 and PC4, separation was observed down to the level of 2.5% (w/w) and samples at 1% (w/w) moved to the edge of the heparin cloud (**Fig. 3.19 A**). Separation of 2.5% (w/w) is lower than the threshold of 3% (w/w) required for heparin QC.
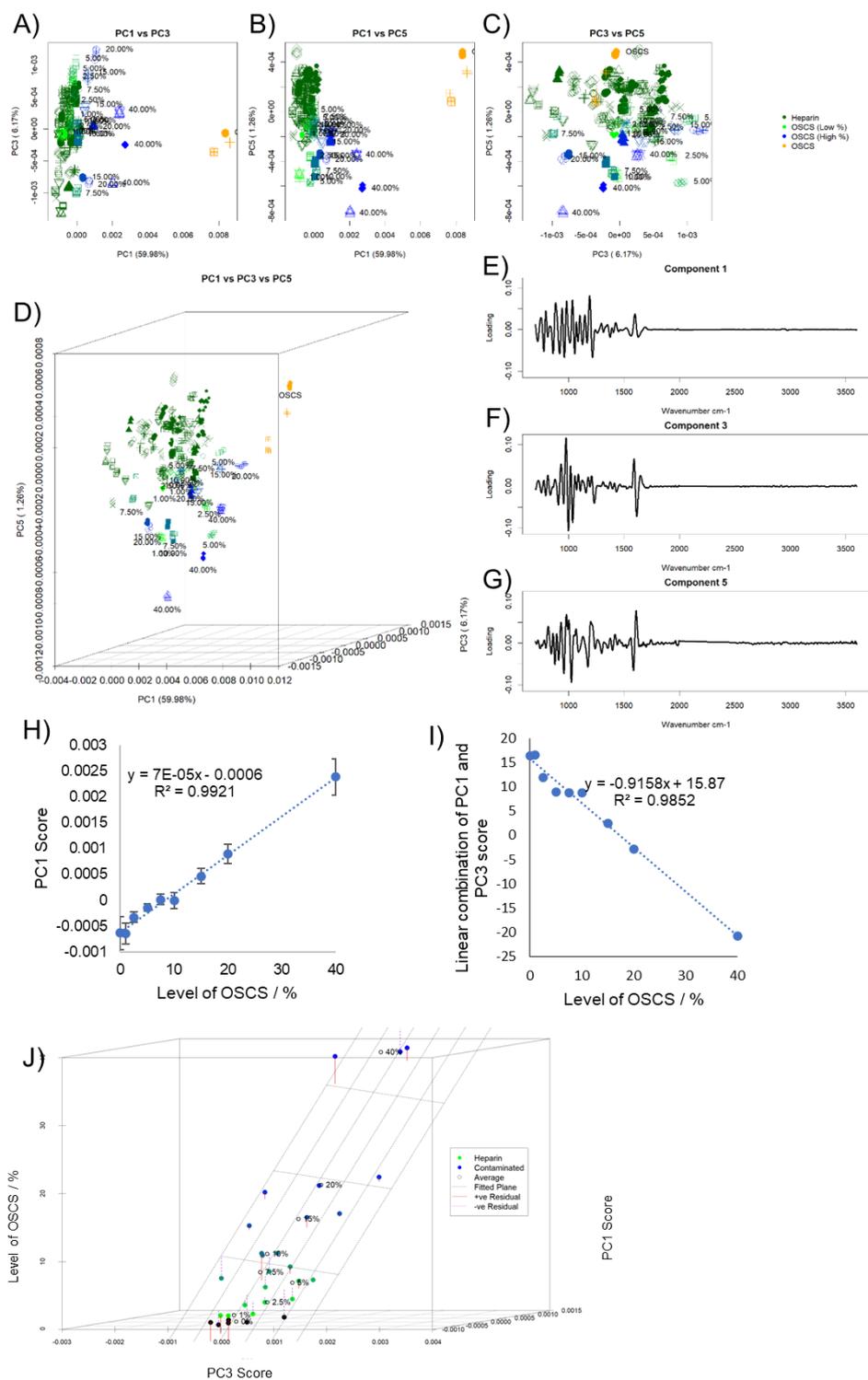
Figure 3.18: PCA of a heparin library and heparins contaminated with OSCS, after taking the second derivative of the spectra and subsequent linear regression analysis. A) Scatter plot of PC1 vs PC3 scores. **B)** Scatter plot of PC1 vs PC5 scores. **C)** Scatter plot of PC3 vs PC5 scores. **D)** 3D scatter plot of PC1 vs PC3 vs PC5 scores. **E)** Plot of the loadings for PC1. **F)** Plot for the loadings for PC3. **G)** Plot for the loadings for PC5. H) Linear regression analysis of PC1 scores vs level of OSCS present. I) Linear regression analysis of a linear combination of PC1 and PC3 scores vs level of OSCS present. J) 3D plot of PC1 score vs PC3 score vs level of OSCS present. Heparin: darkgreen, OSCS: orange, samples contaminated with OSCS are indicated with a scale of green to blue – the more blue the sample the higher the level of OSCS present. All spectra involved in PCA are in the regions of 700 and 2000 and, 2500 and 3600 cm-1 and are averages of 3 scans. The spectra were smoothed, baseline corrected with a 7th order polynomial, normalised (0-1) before smoothing and derivatisation with the Savitzky-Golay algorithm. Coefficients used to calculate the linear combinations in panel I are computed as part of a multiple linear regression analysis.
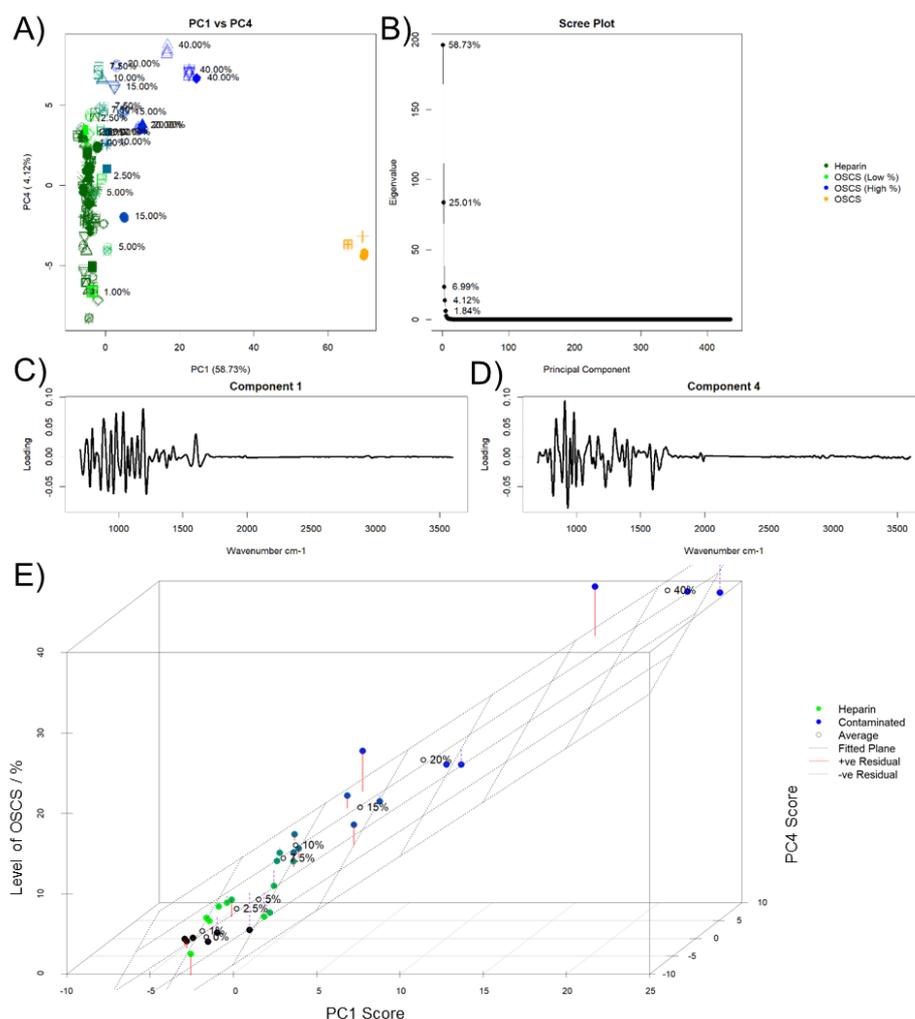
Figure 3.19: PCA of a heparin library and heparins contaminated with OSCS, after autoscaling the second derivative of the spectra and subsequent multiple linear regression analysis. A) Scatter plot of PC1 vs PC4 scores. **B)** Scree Plot. **C)** Plot of the loadings for PC1. **D)** Plot for the loadings for PC4. **E)** 3D plot and multiple linear regression analysis of PC1 score vs PC4 score vs level of OSCS present. Heparin: dark green, OSCS: orange, samples contaminated with OSCS are indicated with a scale of green to blue – the more blue the sample the higher the level of OSCS present. All spectra involved in PCA are in the regions of 700 and 2000 and, 2500 and 3600 cm⁻¹ and are averages of 3 scans. The spectra were smoothed, baseline corrected with a 7th order polynomial, normalised (0-1), smoothed and derivatised with a Savitzky-Golay algorithm and finally autoscaled.

After establishing that PCA can distinguish heparins contaminated with OSCS from pharmaceutical heparin (**Fig. 3.19 A**), individual heparins which had been blended with OSCS were probed against the entire library (**Fig. 3.20**), in a manner similar to how a real sample would be. Samples contaminated with 1% (w/w) OSCS were visually discriminated from the heparin cloud (**Fig. 3.20 D**). Hence, samples with less than 1% (w/w) contamination were generated at the levels of 0.5, 0.25, 0.125 and 0.625% (w/w). The samples were generated by diluting the sample blended with 1% OSCS with the same parent heparin in a 1:2 dilution series. Samples contaminated with 0.5 (**Fig. 3.20 C**) and 0.25% (w/w) (**Fig. 3.20 B**) were seen to move to the edge of the heparin library.
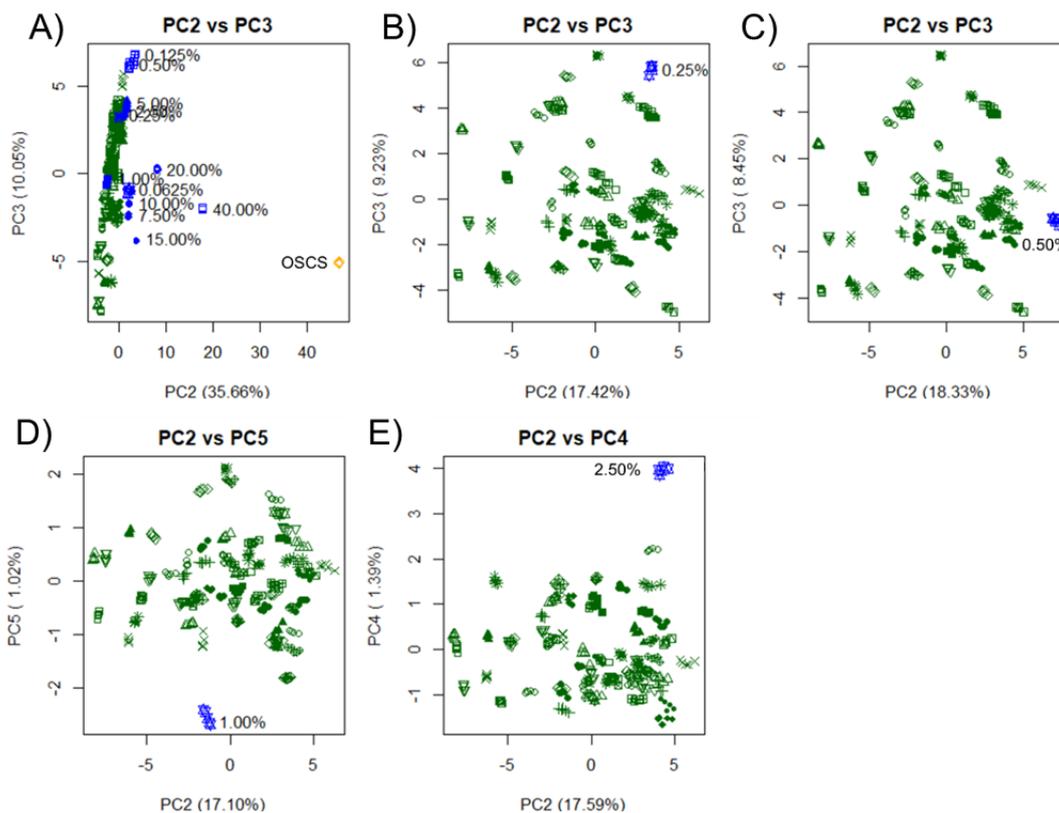
Figure 3.20: Lone samples, contaminated with OSCS at varying levels are compared to a pharmaceutical heparin library using PCA. A) An entire contamination series of heparin with OSCS is compared to a heparin library and OSCS. **B)** A heparin contaminated with 0.25% OSCS (w/w) is compared to a heparin library using PC2 and PC3. **C)** A heparin contaminated with 0.50% OSCS (w/w) is compared to a heparin library using PC2 and PC3. **D)** A heparin contaminated with 1.0% OSCS (w/w) is compared to a heparin library using PC2 and PC5. **E)** A heparin contaminated with 2.50% OSCS (w/w) is compared to a heparin library using PC2 and PC4. The contamination series: blue, consists of heparin blended with OSCS at levels of 0.625, 0.125, 0.25, 0.5, 1, 2.5, 5, 7.5, 10, 15, 20, 40 (w/w); heparin: green, OSCS: orange. All spectra involved in PCA are in the regions of 700 and 2000 and, 2500 and 3600 cm^{-1} and are averages of 3 scans. The spectra were smoothed, baseline corrected with a 7th order polynomial, normalised (0-1), smoothed and derivatised with a Savitzky-Golay algorithm and finally autoscaled.

Attempts to predict the level of OSCS were undertaken using the PCA in **Figure 3.19**. PCR was performed, and the majority of variance covered by OSCS was contained within the first two components (79.6% total variance, 97.8% of the variance covered by OSCS). From this PCR, samples, blended with OSCS, that were not in the original library had PC scores predicted and from the scores, the levels of OSCS calculated. The RMSE for these predictions was 4.2%, with an R^2 value of 0.998. However, a T-test performed between the expected and predicted means revealed a significant difference between them (P-value = 1.2 x 10^{-16}), suggesting that the values are not reliable.

While the PCA demonstrated an ability to differentiate contaminated heparin from pharmaceutical heparin, with visual discrimination down to the level of 1% (w/w) in some cases, it could not predict the level of contamination reliably. PCA of a test sample versus a heparin library cannot be easily applied to a QC situation due to difficulty in predicting single contaminated samples (such as those that would be screened) the same, defined PCs. Different PCs are required to separate samples each time (**Fig. 3.20**), and subsequent visual

discrimination is at the discretion of the analyst. To overcome this, a pre-made OSCS contamination series could be placed into the library to allow calibration and therefore PCR to determine relevant components with which to discriminate potentially contaminated samples.

Nevertheless, through examination of lone heparins contaminated with OSCS, heparin contaminated with as little as 1% (w/w) OSCS can be visually detected. However, there is no ability to classify contaminated samples reliably each time. Ideally, the analyst would be able to take a raw spectrum of the sample and determine if the sample is of pharmaceutical quality or not, with a simple yes/no classifier. PCA cannot be trained to determine if a sample is contaminated or not, but PCA can be used to determine the bands that strongly discriminate OSCS contaminated samples, and this can be probed further.

In agreement with the results in **Table 3.1** a single signal is not satisfactory for OSCS detection, and some features are more significant than others, as highlighted in MLR of signals selected from difference spectra. Through MLR, linear combinations of relevant signals can be determined to improve the detection of OSCS. PCs are themselves linear combinations of the measured variables which highlight trends in the data that correlate with the largest change in variance, hence new test spectra can be mapped into the component space which most strongly correlates with the percentage of OSCS and the resultant component scores used to determine if the sample is contaminated or not.

In the PCAs used for Figures **3.18** and **3.19**, PC1 separated the most strongly for OSCS contamination across all OSCS contamination series, hence the loadings of PC1 were taken forward. The wavenumbers specifically selected by PC1 align with many features in the original spectra, including peaks and troughs, points of inflection and band sections. Each positive and negative feature in the loading plot corresponds to a portion of the spectrum that corresponds to heparin in the positive portion and OSCS in the negative portion (**Fig. 3.21**). The intensity of each band indicates its relative importance and essentially all the features between wavenumbers 700 and 1200 cm^{-1} are of similar importance while the features between wavenumbers 1200 and 1800 cm^{-1} are less important.
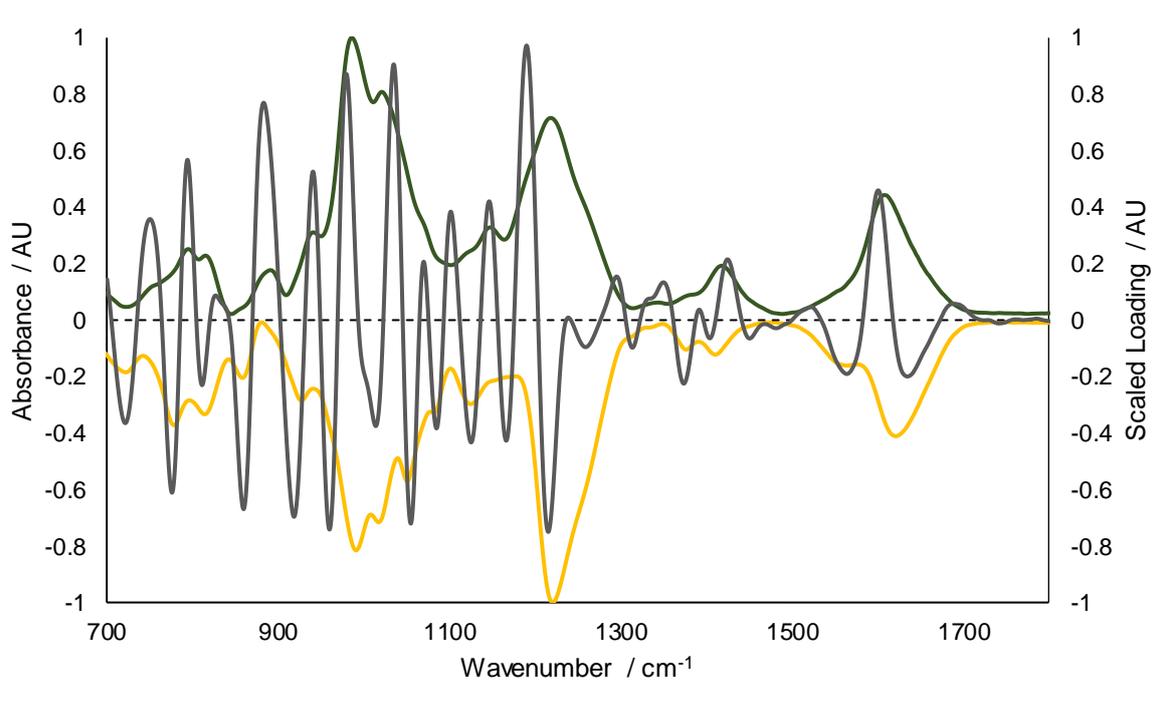
Figure 3.21: ATR-FTIR spectra of PMH and OSCS compared to the loading plot of PC1. Heparin: green, OSCS: orange, PC1 loading: grey. ATR-FTIR spectra were smoothed, baseline corrected with a 7th order polynomial and normalised (0-1). The inverse of the OSCS spectrum is plotted here, to display which signals of OSCS are correlated with the negative loadings.

The scores of PC1 from the PCA in Figure **3.19** were used to create a LRM, so that an analyst having the ability to use a simple yes/no binary operator for if a sample is contaminated or not. During preparation of the logistic data, it was observed that there was much crossover between uncontaminated heparins and heparins contaminated to the level of 1 - 10% (w/w) OSCS, while heparins contaminated at the level of >15% (w/w) were separated more reliably (**Fig. 3.22**). As observed in the difference spectra (**Fig. 3.2B**), heparins contaminated at the level of >20% (w/w) have much more distinct spectral differences, and it was hypothesised that these larger differences may be being separated for in earlier components, causing a loss in resolution at the lower end. Since heparin contaminated at the higher end can be detected easily, the logistic model should be targeted towards heparins contaminated at the low end, with an ideal target of ~3% (w/w) OSCS. A PCA was therefore performed on the same heparin library vs the contamination series at the levels of 1%, 2.5%, 5%, 7.5% and 10% (w/w) OSCS contamination (**Fig. 3.23**).
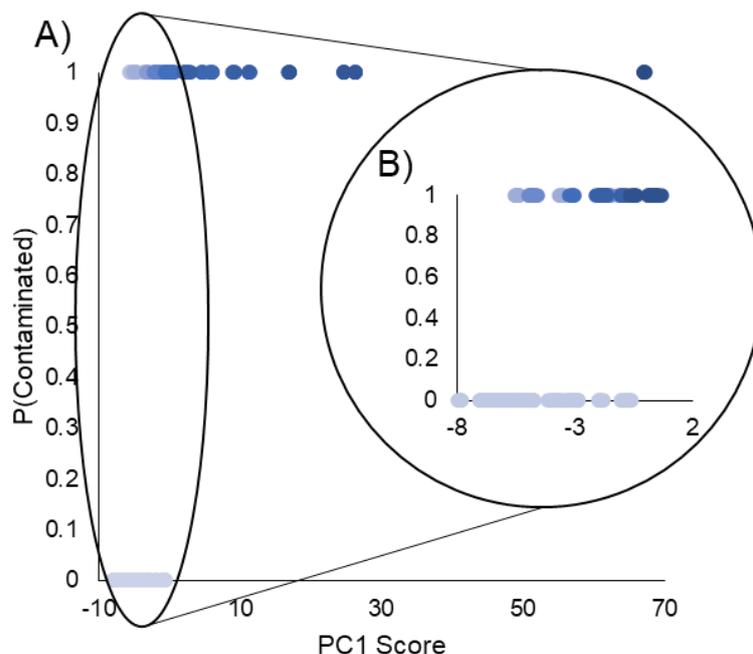
Figure 3.22: Preparation of LRM reveals that samples contaminated with 10% or less OSCS are obscured within the heparin library. A) logistic regression input data for the entire contamination series from 1 to 40%. **B)** Zoom in of panel A, showing the samples contaminated with 1 to 10%. Samples contaminated with OSCS are indicated with a scale of blue, the more contaminated the sampler the bluer it is. Heparin: blue/grey. Heparin samples have y = 0, whilst samples contaminated with OSCS have y = 1.
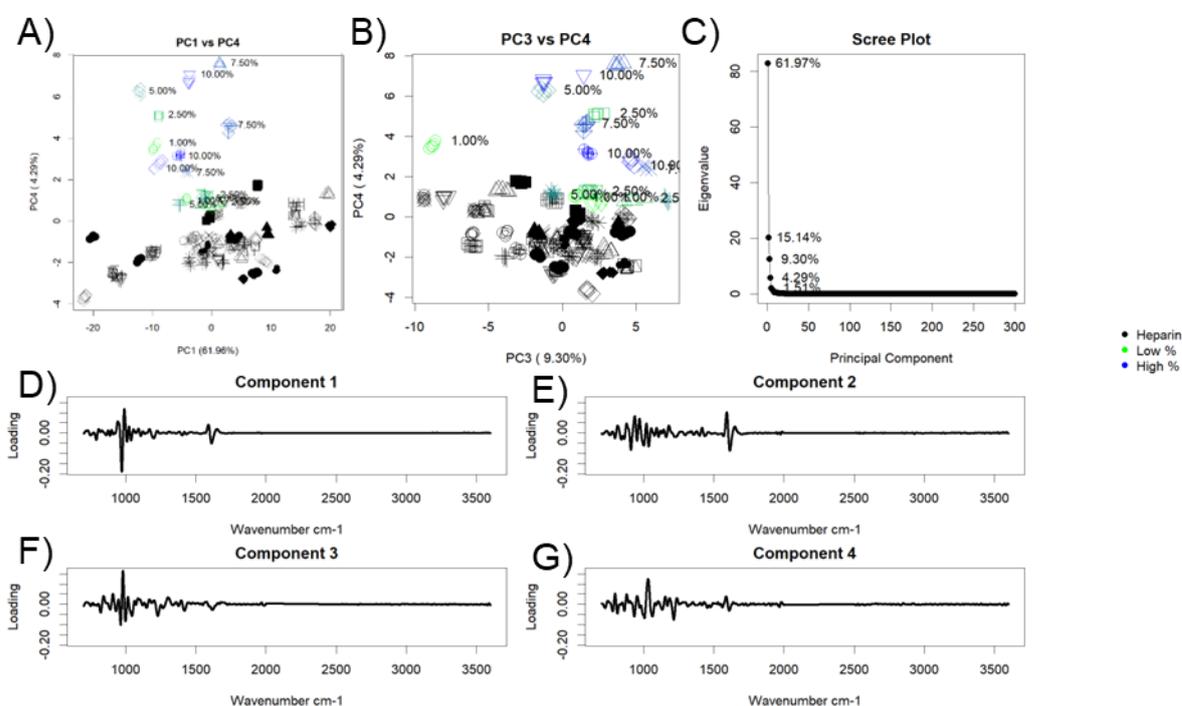
Figure 3.23: PCA of an OSCS contamination series between 1 and 10% against a heparin library. **A)** Scatter plot of PC1 vs PC4 scores. **B)** Scatter plot of PC3 vs PC4 scores. **C)** Scree Plot. **D)** Loading plot for PC1. **E)** Loading plot for PC2. **F)** Loading plot for PC3. **G)** Loading plot for PC4. Heparin: black, samples contaminated with OSCS are indicated with a scale of green to blue – the more blue the sample the higher the level of OSCS present. All spectra used in PCA are in the regions of 700 and 2000 and, 2500 and 3600 cm^{-1} and are averages of 3 scans. The spectra were smoothed, baseline corrected with a 7th order polynomial, normalised (0-1), smoothed and derivatised with a Savitzky-Golay algorithm and finally autoscaled.

[91]

PCR of the first five components revealed that PCs 1 through PC4 where important for the percentage of OSCS ($R^2 = 0.7746$). PCs 3 and 4 had the lowest P-values ($<2x10^{-16}$, compared with $6.01x10^{-13}$ and 0.0147 for PC1 and 2) and hence were taken forward. PCR of PC3 and PC4 confirmed their importance ($R^2 = 0.7274$) and both components retained their low P-values. Through visual inspection, PC4 separates OSCS contaminated samples more than PC3, hence PC4 was taken forward. Comparison of the loading plots (**Fig. 3.24**) between the PC1 from the PCA that contained contamination series of OSCS from 1 to 40% (w/w) (**Fig. 3.19**) and PC4 from the PCA tgat contained the contamination series of OSCS from 1 to 10% (w/w) (**Fig.3.23**) shows the similarities between the two, corroborating that PC4 is the relevant PC while also highlighting some of the differences required to detect low the percentage of OSCS.
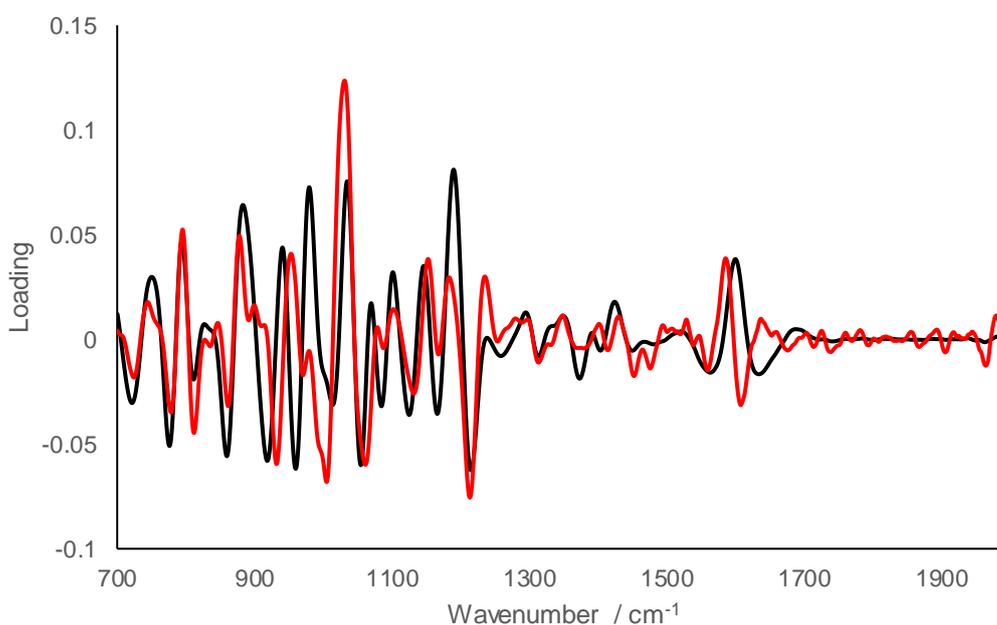
Figure 3.24: Comparison of loadings plots which separate heparin from heparins contaminated with OSCS at levels of 1-40% and, 1-10%. Loadings of PC1 for heparin contaminated with OSCS to the levels of 1-40%: black, Loadings of PC4 for heparin contaminated with OSCS to the levels of 1-10%: red.

The LRM constructed from PC4 of the PCA that covered the contamination series of OSCS from 1 to 10% (w/w) (PCA from **Fig. 3.23**) showed much less overlap between contaminated and uncontaminated samples, increasing the resolution of the model (**Fig. 3.25**). The LRM still had overlap of the OSCS samples contaminated below 10% (w/w) OSCS however, with little separation between them. An LRM was successfully generated ($R^2 = 0.69$). A ROC curve was constructed (AUC = 97.2%) and from it, a probability threshold range of 0.233 to 0.648, correlating with a PC4 score of 0.54 and 1.33 respectively was selected. Depending on how serious a false negative is, the analyst can opt for different thresholds – a PC4 score of 0.54 (**Fig. 3.25,** red line) or lower detects all contaminated samples, with a relatively low false

positive rate of 11%, while a PC4 score of 1.3 (**Fig 3.25,** yellow line) as a threshold correctly identifies all heparins but misses 40% of the contaminated samples. A threshold between these values could also be taken, where 92% of samples are correctly identified, but only 10% of samples are falsely identified (**Fig. 3.25**, blue line).
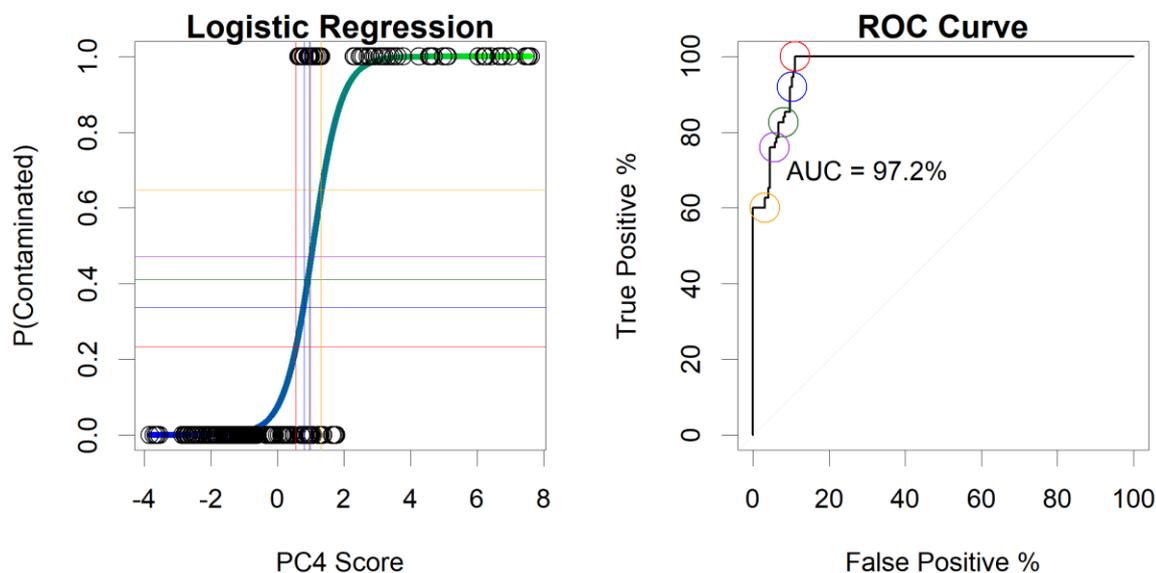
Figure 3.25: LRM of PC4 scores for heparins vs heparins blended with OSCS to the levels of 1-10%. Left: LRM plot for a heparin library (y = 0) vs heparins blended with OSCS (y = 1). Right: ROC curve of responses. Thresholds which were investigated are indicated with coloured lines in the left panel and correspond to the same coloured circles in the right panel.

Cross validation of the LRM was performed, using heparins and contaminated heparins that had not been used in the original PCA (**Fig. 3.26**). The PC4 scores of the samples were calculated using the *predict* function in **base** R and subsequently compared against the established LRM thresholds. At the lowest threshold (**Fig. 3.26**, green line), 37% of the samples were classified correctly, with the majority false negative and at the highest threshold (**Fig. 3.26**, red line), 32% of the samples were classified correctly, with the majority also false negatives.

3.3.6 The optimisation of input IR spectral regions to account for differences associated with sample concentration.

The cross validation failed as the new input samples were not correctly predicted. A LRM could be created that fit the "new" heparins and their contaminated forms, but may not predict the original samples and vice versa. Both of these data sets were taken in different locations and under different circumstances. The contamination series in the original models were acquired at 1mg.mL^{-1} (final volume 1 mL) and the new data set at 10 mg.ml^{-1} (final volume 100 µL) while the entire heparin library was taken at an essentially random concentration. As the samples were recorded in the solid state, it was assumed that the starting concentration before drying would not affect the spectra. It was assumed that the intensity across the spectra

was a function of the amount of sample compressed into the crystal, but it appeared that a change in concentration prior to drying would manifest within the spectra in a noticeable manner. It was therefore hypothesised that the two different data sets were separating from each other due to concentration of the starting material and that correction of this phenomenon would facilitate cross-validation.
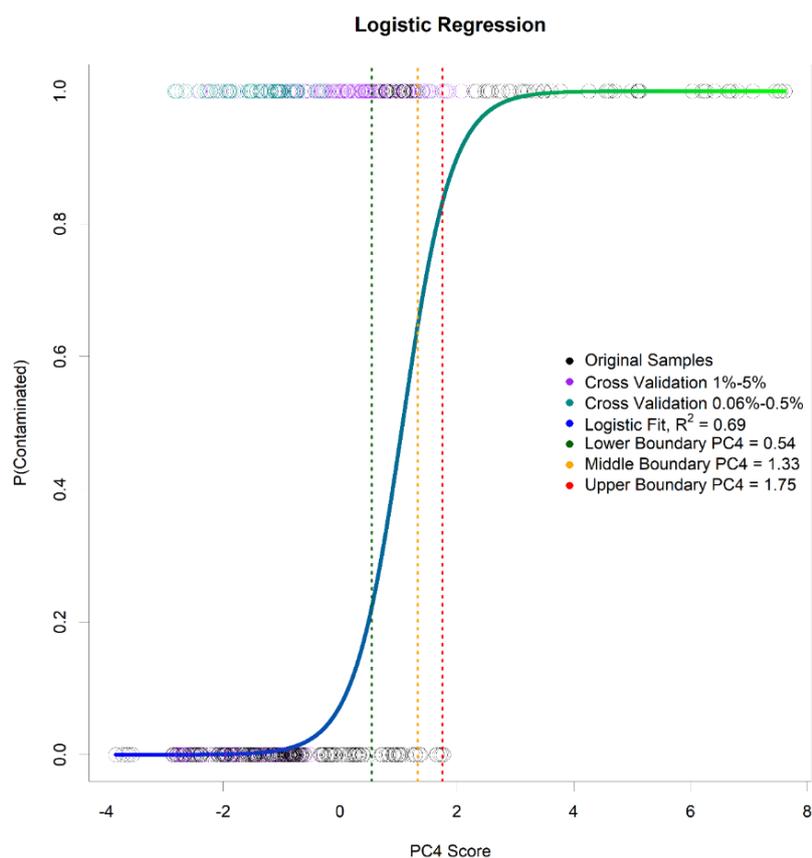
Figure 3.26: Cross validated LRM based on PC4 scores computed from a heparin library vs heparins contaminated with OSCS to the levels of 1-10%. Values that were used to compute the LRM: black, values that were input for cross validation: purple, samples that were input for cross validation with levels of OSCS < 1%: teal. Proposed thresholds for discrimination are indicated with three dashed lines.

To explore the effect of concentration pre-drying on ATR-FTIR spectra, a concentration series was constructed using a random heparin. Heparin was dissolved at concentrations of 0.125, 0.25, 0.5, 1, 2, 5, 7.5, 10, 15, 20, 25, 37.5, 50, 75, 100 and 200 mg.mL^{-1} and after lyophilisation, their spectra were collected. Difference spectra were calculated and plotted (**Fig. 3.27 A and B**) and showed a difference between samples dried from different concentrations.
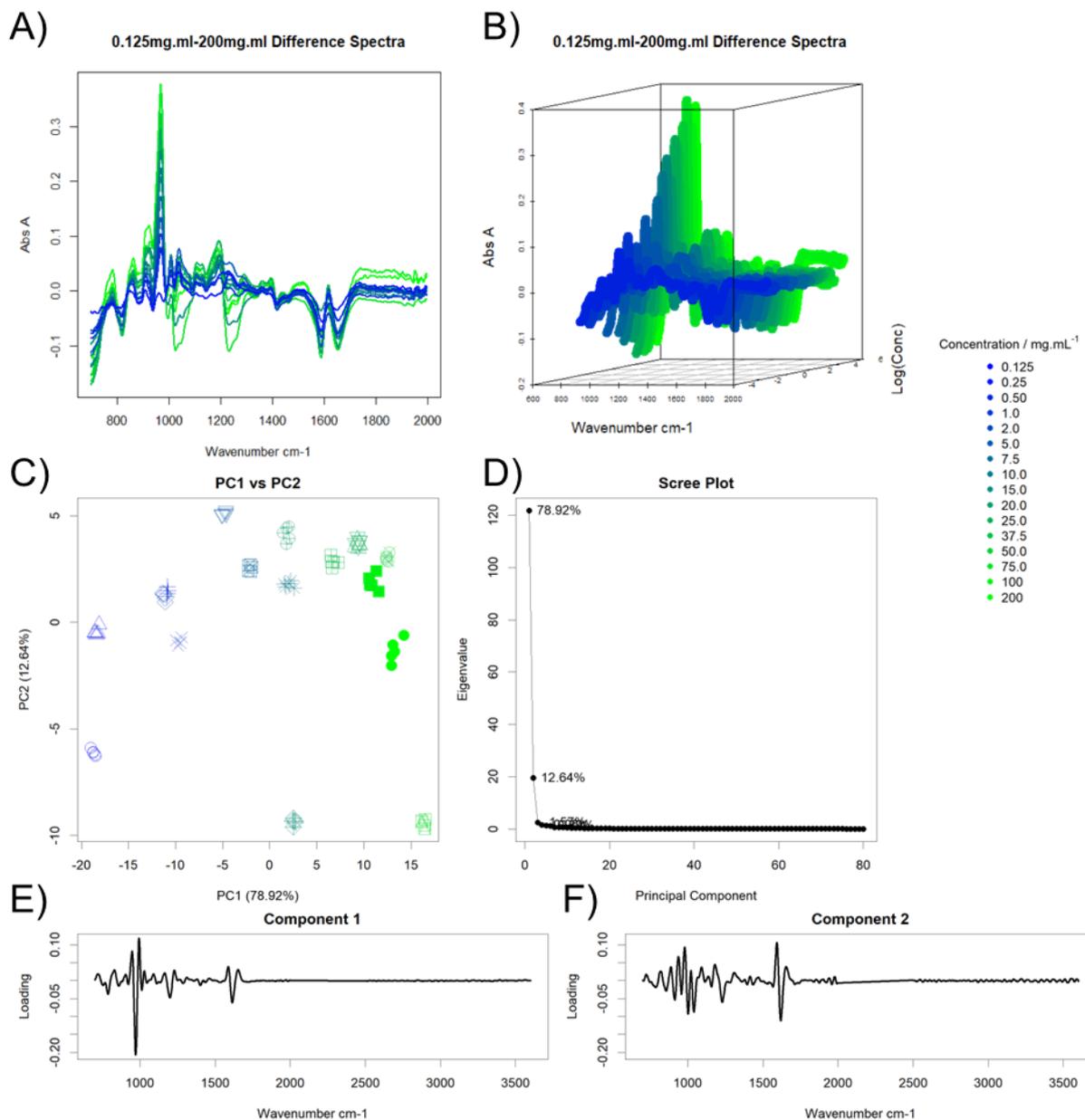
Figure 3.27: Analysis of the effects of sample concentration post-lyophilisation on ATR-FTIR spectra. A) Difference spectra for samples at different concentrations. **B)** 3D difference spectra. **C)** Scatter plot of PC1 vs PC2 scores. **D)** Scree plot. **E)** Loadings plot for PC1. **F)** Loadings plot for PC2. Samples are coloured according to their concentration using a blue to green scale, whereby the greener the sample the more concentrated it is. To create the concentration series, a stock solution of heparin at 200 mg.ml-1 was created and subsequently diluted to make solutions at concentrations of 100, 75, 50, 37.5, 25, 20, 15, 10, 7.5, 5, 2, 1, 0.5, 0.25 and 0.125 mg.ml-1 of heparin. 2 ml aliquots of each concentration were transferred into 2 ml Eppendorf tubes and subsequently lyophilised and the resultant powders had their ATR-FTIR spectra recorded. Difference spectra were calculated by subtraction of the spectrum sample at the lowest concentration (0.125 mg.ml-1). All spectra involved in PCA are in the regions of 700 and 2000 and, 2500 and 3600 cm^{-1} and are averages of 3 scans. The spectra were smoothed, baseline corrected with a 7th order polynomial, normalised (0-1), smoothed and derivatised with a Savitzky-Golay algorithm and finally autoscaled.

PCA of the acquired spectra was performed (**Fig. 3.27**) and a very clear separation was observed (comprising 79% of the total variance in the data set) that separated the samples from very low to very high concentration across PC1 (**Fig. 3.27 C**). This variation was observed almost entirely in the carbohydrate band (**Fig. 3.27 A, E**), with the variable region centring at

wavenumber ~ 985 cm^{-1}. Fluctuations in this region were observed in the PC3 of the PCA in **Figure 3.23**, hence it was not used in conjunction with PC4 for the LRM in **Figure 3.25**. This variation was also observed in the early linear regression models after investigation of the feature at wavenumber 1005 cm^{-1} – the feature with the poorest R^2 value (**Figure 3.3, Table 3.1**). There are small effects noted in the bands at wavenumber ~ 1200 cm^{-1} and ~ 1600 cm^{-1}; the bands which also show great variance for the prediction of the level of OSCS within the heparin library (**Table 3.1**).

To account for this, the region between wavenumber 960 – 1007 cm^{-1}, accounting for the entirety of the spike in PC1, was deleted from the heparin libraries, and the same analyses performed (**Fig 3.28**).
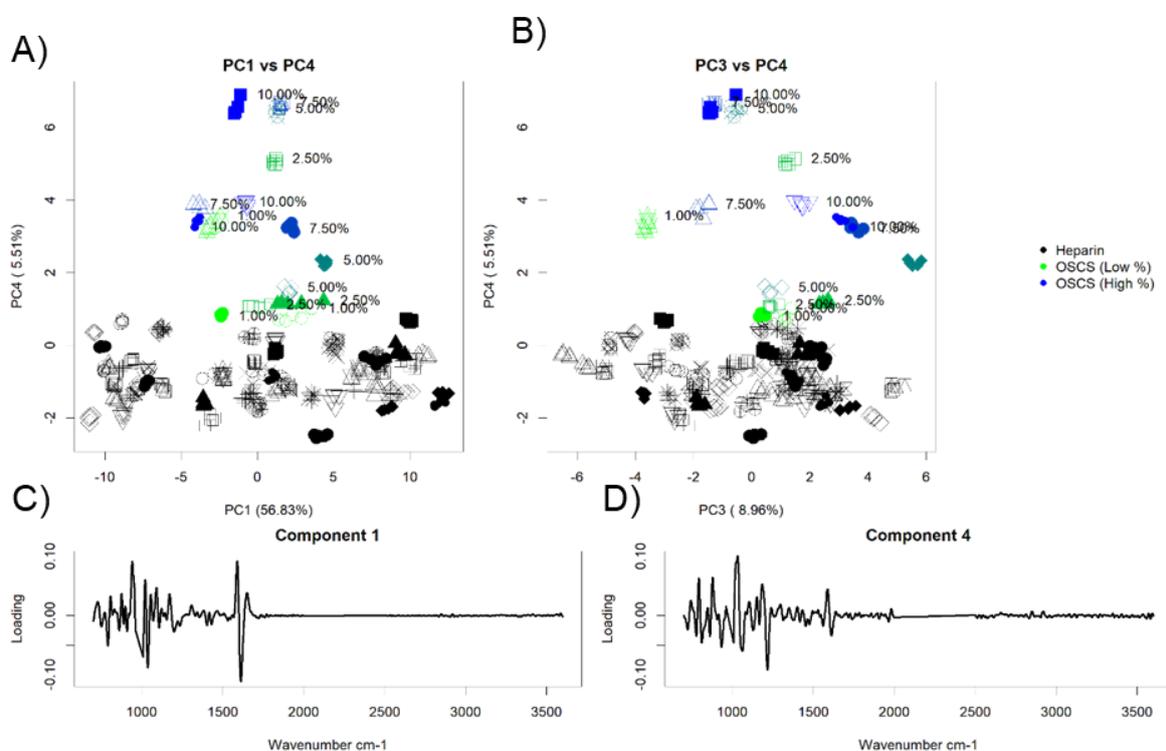
Figure 3.28: PCA of a heparin library vs a contamination series of heparin with OSCS in which the ATR-FTIR spectra have had the signals at 960 – 1007 cm^{-1} removed. A) Scatter plot of PC1 vs PC4 scores. **B)** Scatter plot of PC3 vs PC4 scores. **C)** Loading plot from PC1. **D)** Loading plot from PC4. Heparin: black, samples contaminated with OSCS are indicated with a scale of green to blue – the more blue the sample the higher the level of OSCS present. All spectra involved in PCA are in the regions of 700 and 960, 1007 and 2000 and, 2500 and 3600 cm^{-1} and are averages of 3 scans. The spectra were smoothed, baseline corrected with a 7th order polynomial, normalised (0-1), smoothed and derivatised with a Savitzky-Golay algorithm and finally autoscaled.

Improved separation in PC4 was seen, with the heparin library flattening out across PC4, most likely due to the decreased variability across the library (**Fig. 3.28 A, B**). The LRM was vastly improved with an R^2 of 0.97, a ROC curve with an improved area of 99.9% and almost complete separation between contaminated and pharmaceutical heparin was observed (**Fig. 3.29**).
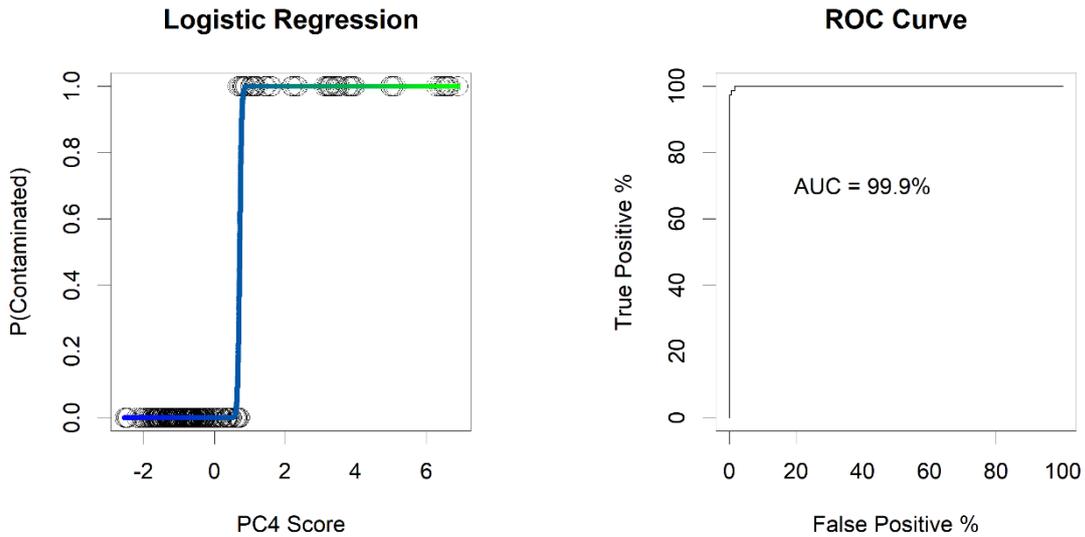
Figure 3.29: LRM of PC4 scores for samples blended with OSCS to the levels of 1-10% or not with the band at 970-1007 cm-1 removed. Left)LRM plot for a heparin library (y = 0) vs heparins blended with OSCS (y = 1). **Right)** ROC curve of responses.

Cross validation was performed (**Fig. 3.30**) as before (**Fig. 3.26**) and also improved, with 97% of samples from the new library correctly identified. Importantly, all samples contaminated to the level of 2.5% (w/w) or higher were correctly identified, which is lower than the required 3% (w/w) OSCS threshold.
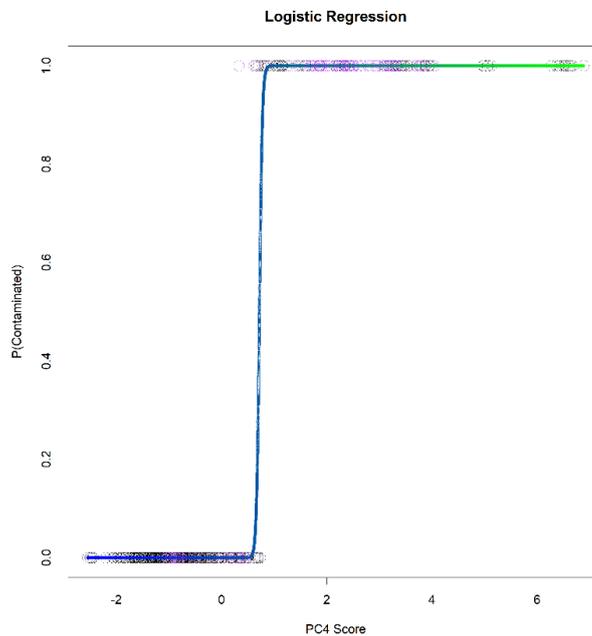
Figure 3.30: Cross validated LRM based on PC4 scores computed from a heparin library vs heparins contaminated with OSCS to the levels of 1-10%. Values that were used to compute the LRM: black, values that were input for cross validation: purple.

[97]

As so many samples were correctly predicted, new samples(n=24) which consisted of blends below 1% at the levels of 0.5%, 0.25%, 0.125% and 0.0625% (w/w) OSCS were then produced introduced into the LRM. The samples were not distinguished from the heparin library in PCA (**Fig.3.31**) in components that correlate with OSCS contamination but some separation was observed with PC1, the loadings for which (**Fig 3.31 C**) resemble the concentration effects observed in (**Fig. 3.27**). In the LRM, 64% of the samples were correctly identified as contaminated, with the remainder falling below the LOD of the curve (i.e. below the minimum threshold of PC4 score = 0.54).
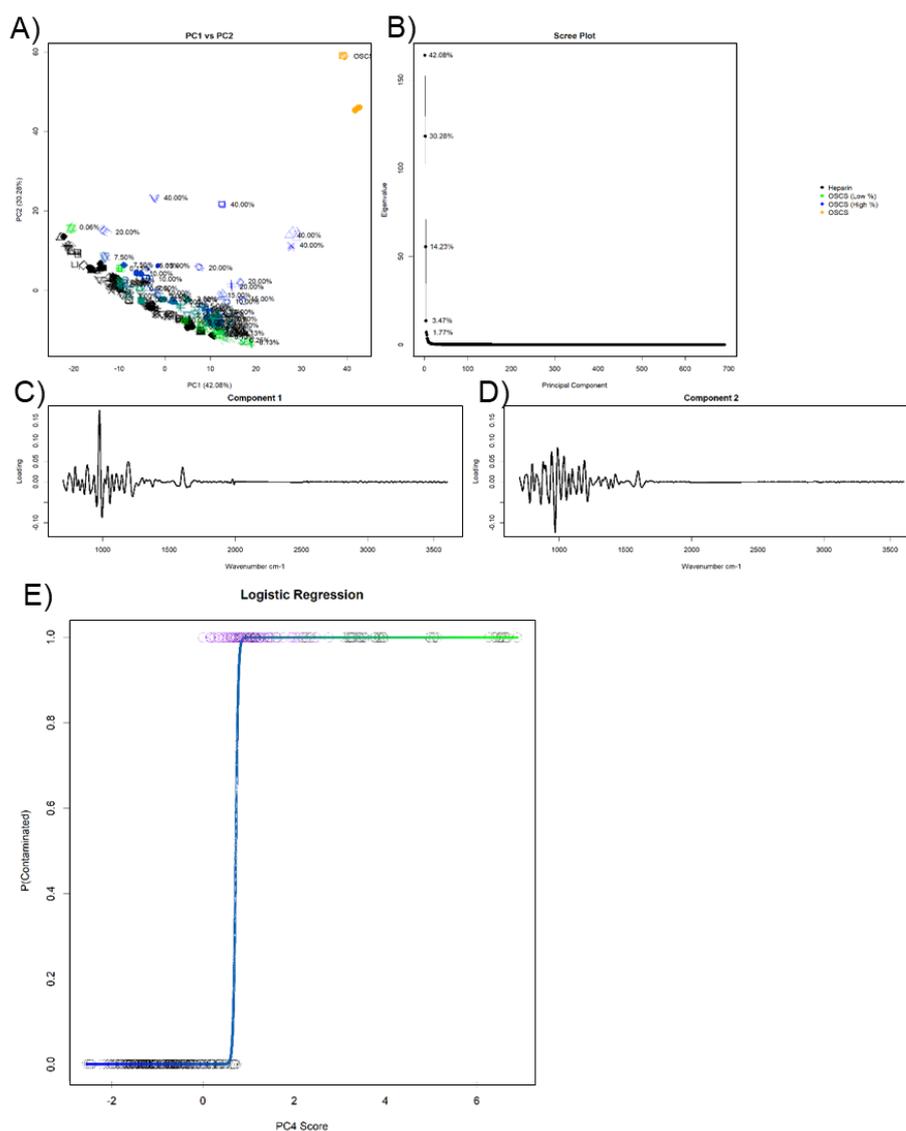
Figure 3.31: PCA of a heparin library vs a contamination series of heparin with OSCS, including samples that are below 1% (w/w) contamination and cross validation of the LRM model with the same sub 1% (w/w) samples. A) Scatter plot of PC1 vs PC2 scores. **B)** Scree plot. **C)** Loading plot for PC1. **D)** Loading plot for PC2. **E)** Cross validation of LRM from Figure 3.29, including sub 1% OSCS samples. For PCA; Heparin: black, samples contaminated with OSCS are indicated with a scale of green to blue – the more blue the sample the higher the level of OSCS present. For LRM; Values that were used to compute the LRM: black, values that were input for cross validation: purple. All spectra involved in PCA are in the regions of 700 an d 960, 1007 and 2000 and, 2500 and 3600 cm^{-1} and are averages of 3 scans. The spectra were smoothed, baseline corrected with a 7th order polynomial, normalised (0-1), smoothed and derivatised with a Savitzky-Golay algorithm and finally autoscaled.

3.4 Discussion

Initially, simple spectroscopic techniques were utilised to locate spectral features of interest that may correlate with levels of OSCS contamination. Difference spectra of heparin contaminated with different levels of OSCS (**Fig. 3.2B**) revealed 5 signals of interest, the intensities of which correlated linearly with the level of OSCS, with a loss of resolution at or below 5% (w/w) OSCS (**Fig 3.3, Table 3.1**). Unfortunately, not all heparins are the same compositionally, and each heparin has a slightly different IR spectrum (**Fig. 3.4**) which overlap heparin samples blended with <10% OSCS, masking the latter from detection. The linear models became difficult to use, misclassifying many heparins as contaminated at high percentages (**Table 3.1**).

To account for the spectral effect created by the natural variation of heparin, an MLR was performed on the identified features (**Fig. 3.5**). The MLR correlated with the percentage of OSCS ($R^2 = 0.99$) but also highlighted the lack of importance of three of the features that were identified. The three features were removed, and an MLR of just 881 and 1639 cm^{-1} was constructed (**Fig. 3.6**), with the same R^2 value of 0.99. The MLR was able to improve the LOD to 7.5% (w/w) (**Table 3.1**). Ideally, <3% (w/w) OSCS would be detected, as 3% was shown to be the limit of medical issues in rats and pigs (Corbier et al., 2011; McKee et al., 2010). A LRM of the linear models was also created, which corroborated the LOD of 7.5% (w/w) (**Fig. 3.7**). Through use of the LRM and MLR models together, an analyst could classify contaminated samples and simultaneously confidently quantify the level of contaminant from 7.5% up to a level of 40% (w/w).

Manually selecting signals based on difference spectra introduced bias on behalf of the analyst and subtle but important changes may not be identified or overlooked. FTIR spectra are naturally broad, and identification of features of interest is difficult for the human eye. PCA was performed on the contaminated and pharmaceutical spectra. PCA showed improved separation, immediately shifting samples contaminated to the level of 5% (w/w) to the edge of the heparin cloud (**Fig. 3.8**). PCR was performed to identify relevant components for study and analysis of the relevant PC loading plots revealed new features of interest. Efforts were then undertaken to improve the separation of the PC plots, separation was observed between repeats of the same sample, which suggested that random noise and spectral artefacts were causing an issue.

Through iteration, derivatised spectra with a small level of smoothing and autoscaling were shown to have markedly increased separation in PCA, with samples contaminated to the level of 2.5%, i.e. below the 3% (w/w) threshold detected visually (**Fig. 3.19**). PCA was then performed on a series of lone contaminated heparin samples against the entire library to probe whether PCA could be used as a viable QC tool (**Fig. 3.20**). Discrimination was achieved down

to the level of 1% (w/w) in some cases (**Fig. 3.20 D**), with samples contaminated at the level of 0.5% and 0.25% (w/w) OSCS w/w moving to the edge of the heparin cloud (**Fig. 3.20 B, C**). This analysis highlighted a major issue with PCA as a quality control tool – as the level of contamination decreases, the importance of the component that separates it decreases, making the correct component harder to detect and subsequently to be used. While it is well documented that higher PCs are not necessarily the most important when applied to a specific task (Massy, 1965), knowing which PCs correlate with the desired outcome becomes difficult, especially at lower levels of OSCS.

While it is entirely possible for the analyst to compare the loading plots of their PCs to loading plots of accepted PCs that discriminate OSCS (as in **Fig. 3.24**), subtle changes can still be observed in these loadings, and it is again at the discretion of the analyst to decide if the PC is important or not. These issues could relate to a number of factors which can alter ATR-FTIR spectra, most likely sample concentration before drying (**Fig. 3.27**), room/sample temperature, air composition (level of CO_2, humidity etc) and the amount of sample placed onto the ATR crystal. ATR-FTIR is sensitive enough to these factors to induce variations that are detected in some PCs and the analyst must take care in deciding what is important or not. To overcome these issues, PCR could be performed on the target sample, the heparin library and an OSCS contamination series, to facilitate better identification of relevant components. An attempt to quantify the level of OSCS was made using PCR but the predicted values were significantly different to the real values (P value = 1.2×10^{-16}) even at the level of 10% (w/w) OSCS, quantification failed but importantly the samples were still distinguishable from the heparin library by eye.

In an ideal QC situation, a simple yes/no binary operator would be used which removes all elements of bias from the analysts. PCs are formed from a linear combination of coefficients which have been calculated in a way that most aptly demonstrates the natural variation in the data set, similarly to MLR but over many more variables. A PC or a series of PCs that can separate heparins contaminated with OSCS from pharmaceutical heparins, could be identified and the corresponding loadings used to map new samples into this PC-space. This removes the requirement of the analyst to decide which components are necessary in a singular QC situation.

A LRM (**Fig. 3.25**) was constructed based on PC1 scores from the PCA in **Figure 3.23**. Importantly, the threshold of detection of the LRM could be tuned to discriminate all heparins contaminated at or above the level of 3% (w/w) - i.e. the level that is considered to cause medical harm (Corbier et al., 2011; McKee et al., 2010). After tuning of the input data (**Fig. 3.29**) and corresponding PCs, all samples at the level of 2.5% (w/w) were discriminated. The LRM also separated, in some cases, samples to the level of 0.25% (w/w) OSCS (**Fig. 3.31 E**).

This model has potential for use in heparin QC; only requiring an IR instrument and a computer to perform. The separations observed in PC4 correlate with the presence of OSCS, but the level of OSCS cannot be quantified, hence if quantification is required, this approach is not viable. Regardless, the drastic increase in sensitivity from 7.5% down to 0.25% w/w OSCS in some cases makes the inability to quantify less important.

One thing noted during the preparation of the PCA-LRM was the sensitivity of the IR instrument to underlying interactions within the sample. The concentration of the input sample changed the relative intensities across the spectrum and represented a major source of variation throughout the dataset (**Fig.3.27**). Removal of signals which correspond with these effects, while not ideal, vastly improved the sensitivity of the approach (**Fig. 3.25 and 3.26 vs Fig 3.29, 3.30 and 3.31 E**).

Chapter 4: The detection of, discrimination of and quantification of contaminants other than OSCS.

4.1 Introduction

One major issue with the PCA-LRM model described in **Figure. 3.29** is that it has been specifically tailored to detect OSCS. Most publications based on heparin QC are also targeted solely at the discrimination of OSCS and occasionally DS, only one is targeted at multiple alien contaminants (Rudd et al., 2011a). This is a pertinent problem with OSCS contamination as the unusually high level of side effects displayed by the contaminated heparin samples have led some to posit that other, highly sulphated contaminants were present, including perhaps OSDS and OSHS, molecules that may be hidden due to the natural presence of under-sulphated variants in accepted pharmaceutical heparin (Ramacciotti et al., 2011). Such contaminants, if present, are likely sulphated forms of polysaccharides separated from crude heparin (Taylor et al., 2019) - crude heparin being the mixture of GAGs extracted prior to purification and clean-up to form the final pharmaceutical product. The LRM (**Fig. 3.29**) may not be able to effectively categorise samples contaminated with such different molecules and hence the robustness should be tested.

NMR with PCA was shown to be able to discriminate heparins contaminated with alien molecules other than OSCS, some of which hide behind the H[1] NMR spectra due to a lack of acetyl signals – the primary signal for which OSCS is detected with H[1] NMR (BHRA, 2019; EP, 2010; Rudd et al., 2011a; USP, 2009), with similar discriminatory power to that of OSCS (Rudd et al., 2011a). While PCA-LRM is powerful, it may be unable to discriminate samples contaminated with a novel contaminant, particularly a contaminant whose structure is unknown to the analyst, hence, a contamination series and subsequent relevant LRM could not be constructed. Therefore, exploration of a technique that detect anything "alien" is vital. It should be noted that, unlike in NMR, the signals of OSCS, particularly at low percentages, lie beneath the IR spectra but can still be detected.

Another potential source of contamination in the future could be that of novel, semi-synthetic polysaccharides, which are usually easy to source polysaccharides that have been chemically modified to have heparin-like structures and activities (Kareem et al., 2018; Skidmore et al., 2017). While semi-synthetic polysaccharides are currently unlikely to enter the heparin supply chain naturally, they represent a possible adulterant, and contain other unique structural features, the detection of which may be necessary in future contaminant detection. IR is regularly used to identify semi-synthetic modifications made to carbohydrates (Boyle et al., 2017; Huang et al., 2020; Suflet et al., 2010) and hence it is likely that it will possess the ability to detect underlying levels of these carbohydrates.

Another facet of heparin QC is that of species identification. In the monograph, SAX chromatography is used to discriminate and quantify whole GAGs within the sample. NMR can also be used to discriminate different CS/DS from heparin/HS within the sample, specifically by detection of sulphates associated with CS/DS and levels of GalNAc and IdoA-GalNAc with H^1, C^{13} HSQC (Mauri et al., 2017b). Solution-based FTIR coupled with PCA has been employed previously to discriminate different GAG species, particularly by the type and level of sulphation (Mainreck et al., 2011), and has been used to detect structural changes in GAGs in chondrocytes and in cartilage (Brézillon et al., 2017; Mohamed et al., 2017; Sanden et al., 2019). It is also entirely possible that the detection of small levels of these molecules may be possible by FTIR, as observed for OSCS contamination (**Fig. 3.29**).

Quantification of OSCS was undertaken in **Chapter.3,** but was significantly different to the expected values (P value = 1.2×10^{-16}). A RMSE of 4.2 was obtained and, as the majority of the samples that had the level of OSCS quantified were at or below 5% (w/w) OSCS, this is likely the reason the quantification failed. ~5% (w/w) is a reasonable margin of error for the prediction of GAG levels in samples that are approximately 2 parts heparin and 1 part "other" and hence, better quantification may be observed here.

Detection and quantification of the levels of CS and DS is paramount for heparin QC, for example, pharmaceutical heparin is extracted first as a crude mixture containing approximately 2 parts heparin and 1 part "other", usually CS, DS and some residual DNA/RNA. There are currently no strict rules for the quality of crude heparin, but there is a push to define a "standard" crude heparin, as it is posited that the crude stage is where initial contamination with OSCS began (FDA, 2013; Mendes et al., 2019).

Crude heparins comprise the product which will be later purified and to pharmaceutical quality heparin. During heparin production, the intestines of the relevant animal are collected and processed in slaughterhouses approved by regulatory authorities. The mucosa is first preserved with sodium metabisulphite (Al-Hakim, 2021). The mucosal epithelium is extracted and boiled to solubilise the tissue, to which protease is added, following digestion the mixture is centrifuged and the solid matter filtered off (Al-Hakim, 2021; Taylor et al., 2019). The solution is precipitated with an alkonium halide or ethanol and the dissolved precipitate is loaded onto an anion exchange resin and then transported to crude heparin processing facilities (Mauri et al., 2017b; Taylor et al., 2019). The loaded resin is washed with water and a solution of low-ionic strength, removing unwanted molecules and enriching the heparin content (Mauri et al., 2017b). The enriched fraction is separated from the resin with a high electrolyte (NaCl) concentration, filtered, precipitated, and vacuum dried to form the substance formally known as crude heparin. The mixture also contains residual proteins, a range of bound small

molecules and ions and nucleic acids, on which the polymerase chain reaction may be used identify the species of origin (Al-Hakim, 2021; WHO, 2022).

The crude heparin then undergoes oxidative and alkaline treatments to remove residual colour, and to inactivate endotoxins and viruses that remain following extraction. The mixture is precipitated with increasing volumes of methanol or ethanol to fractionate heparin from CS and DS (Van der Meer et al., 2017). The product is filtered at low pH to remove any residual proteins (Liu et al., 2009) and cation-exchange is performed which both converts the heparin to its relevant cation form (usually Na^+ or Ca^{2+}) and removes extraneous metal ions, resulting in pharmaceutical heparin (Taylor et al., 2019). Separation of heparin from CS and DS may also be achieved with spray drying (Van der Meer et al., 2017). The exact extraction and purification process is a closely guarded secret for pharmaceutical heparin suppliers, hence exact details as to specific processes cannot be described here (Al-Hakim, 2021).

Aside from heparin and crude heparin QC, GAGs extracted from novel sources, such as marine animals (dubbed marine GAGs from hereon) are becoming studied more frequently (Brito et al., 2014; Chavante et al., 2014; Mycroft-West et al., 2021, 2020; Palhares et al., 2019). Marine GAGs possess unique or rare structures compared with mammalian GAGs and therefore have interesting pharmaceutical applications, for example, a heparin/HS hybrid (i.e, a highly sulphated HS) from the heads of the shrimp _Litopenaeus vannamei_, which possesses high levels of GlcN,6S and GlcN,6S-IdoA2S compared to normal HS but comparable anticoagulant ability to heparin in both APTT and Anti-XA assays has been described in (Brito et al., 2014), whilst another, distinct glycosaminoglycan mixture from _L. vannamei_ has been shown to Inhibit BACE-1, a currently target for Alzheimer's disease therapies (Mycroft-West et al., 2021). Similarly, a GAG mixture from _Portunus pelagicus_, a species of crab, has also been shown to inhbit BACE-1 (Mycroft-West et al., 2019). Rapid compositional analysis of these and the many potential future marine GAGs could be of use for future GAG research.

4.2 Chapter Aims

1. Establish the specificity of the PCA-LRM model (**Fig. 3.29**)
2. Define LODs for other possible contaminants.
3. Determine whether the GAGs in a mixture, such as crude heparin or in marine GAGs can be detected, identified and quantified.

4.3 Results

4.3.1 The expansion of the PCA-LRM model to other contaminants

To begin, a series of OS-GAGs, including OSDS, OS-heparin (OSHp), OSHS, OSHA and a new OSCS were synthesised through the chlorosulphonic acid-pyridine route that is most likely to have been used in the synthesis of OSCS in 2008 (personal communication, M. Guerrini). Further to this strategy, a semi-synthetic, oversulphated agarose (OSAS) was synthesised and an already naturally sulphated carbohydrate, dextran (DeS) was also selected – both samples that have been tested with NMR-PCA (Rudd et al., 2011a). A contamination series akin to before of all these samples was formulated, to the levels of 1%, 2.5%, 5%, 7.5%, 10%, 15%, 20% and 40% (w/w). The parent compounds, (DS, HA and CSA), were also used to create contamination series, alongside a CSC. The resultant ATR-FTIR spectra were prepared, as in (**Fig. 3.28**), and the linear combinations, based on the PCA-LRM PC4 loadings (**Fig. 3.29**), were calculated using the *predict* function in **base** R. For the PCA-LRM, a calculated PC4 score of < 0.60 was characterised as a pharmaceutical quality heparin, a score of 0.60 < x < 0.87 was defined as possibly contaminated (i.e, it falls within the linear portion of the logistic model and may be classified as contaminated, or not, depending on the chosen threshold) and a value of > 0.87 was defined as contaminated.

Table 4.1: LODs for the detection of different natural and semi-synthetic GAGs blended with pharmaceutical heparin. *: n=1

	CSA	CSC	DS	OSDS	HA	OSHA	DeS	OSAS
LOD	≤1% - 2.5%	2.50%	≤1% - 5%	≤1%*	NA	5%*	5%*	2.5%*
Average PC4 score at boundary value	0.76 +/- 0.31	-0.25 +/- 1.18	0.32 +/- 0.2	1.41	-17 : 0.4	0.63	0.72	0.51

	Generic Polysaccharide	OS-Polysaccharide
Average PC4 score	-4.4	10.0
Min : Max	-25.3 : 20.4	-22.2 : 37.0
Lower Quartile	-10.1	-0.23
Upper Quartile	1.7	22.2
Median	-4.8	14.1

From these values, LODs could be determined, based on the PCA-LRM created earlier (**Fig. 3.29**). The model was robust, given that it was created specifically to detect OSCS, detecting comparable levels of most of the contaminants. CSA and OSDS were detected at the level of 1% (w/w) (a level below this was never tested) while OSAS and DS were detected to the level of 2.5%. OSHA and DeS were detected to 5% and CSC to 7.5% (w/w). HA was not detected, as all of its PC4 values were greatly below the threshold. All heparins tested had PC4 scores between -2.0 and 0.60, while the HA contaminated samples had PC4 values down to -17. The

data in **Table 4.1** suggests that the model has become inadvertently tuned to sulphated polysaccharides. Saccharides that have sulphations at positions not found in heparin, such as CSA (4S) and the OS-GAGs have lower LODs of 1%, 1% and 0.5% (w/w) for CSA, OSDS and OSCS respectively, while compounds with a sulphation pattern more similar to heparin such as CSC with 6S and DS with a 2S are detected less readily with LODs of ~2.5%.

To further test this hypothesis, a series of chemically modified, semi-synthetic polysaccharides and the original parent compounds had PC4 scores predicted and tested using the LRM. Generally, the parent compounds were classified as uncontaminated with very low PC4 scores ($\bar{x}$ = -4.41), similar to that of HA ($\bar{x}$ = -8.34), while the OS-derivatives had very large PC4 scores ($\bar{x}$ = 10.02), similar to that of the OS-GAGs. Some polysaccharides were classified as "contaminated", but not all of them were naturally sulphated. Some were 2- sulphated (iota and lambda carrageenan) and 4-sulphated (kappa-carrageenan) while others were methylated, ethylated and acetylated (methyl-, ethyl-cellulose and karaya gum respectively (BeMiller, 2018; Campo et al., 2009)) and others possess 1-6 branches or amino-sugars (levan and locust bean gum and chitin derivatives respectively (Finney and Siegel, 2008; Glicksman, 1963; Meyer, 2015, p. 2)). Sulphation of compounds that were not discriminated with LRM led to detection. This suggests that the LRM-PCA model is either specific to detecting heparin, and not to its contaminant, or it is separating samples based on broad and overlapping features – particularly sulphation in locations (2-S and 4-S) or at levels (DOS > 2.8) alien to heparin or of features (branches and alkane derivatives) alien to heparin.

The strong negative values observed in unsubstituted saccharides suggests that the LRM may be further extended to better define heparin and its potential contaminants resulting in a region across PC4, which describes heparin. To explore this further, another LRM was created between heparin and heparin contaminated with HA. The PC4 score values for heparin contaminated with HA were predicted and an LRM constructed (**Fig. 4.1**), possessing a low R^2 value of 0.30, a ROC AUC of 77.4%. Considerable overlap between the lower end (≤ 15%) of the HAs was observed and this made differentiations not possible; only at infinity was a false positive rate of 0% found, and at 100% true positive, 44.6% of samples were misclassified.

This is most likely because PC4 is targeted at sulphated polysaccharides (**Table 4.1**) and hence HA fails to be detected. A new PCA was performed (**Fig.4.2**), using both OSCS and HA, which revealed that contaminated heparins with both OSCS and HA could be readily separated from the main heparin library, but in an orthogonal manner. Orthogonal separation of HA and OSCS from heparin may provide an explanation as to why the previous LRM did not work for HA contamination.
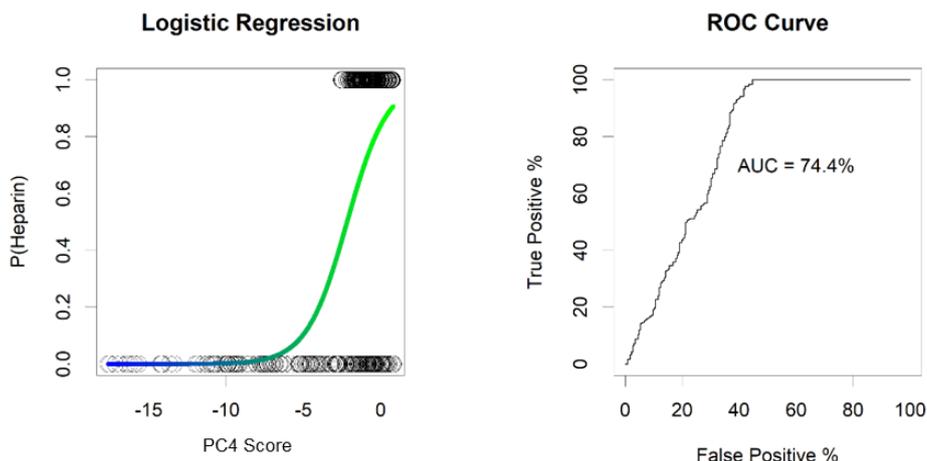
Figure. 4.1: LRM of PC4 scores for samples blended HA vs heparin. Left: LRM plot for a heparin library (y = 1) vs heparins blended with HA (y = 0). Right: ROC curve of responses. PC4 scores were calculated for the heparins blended with HA using the predict function in base R using the PCA from Figure 3.28 All spectra involved in PCA are in the regions of 700 and 960, 1007 and 2000 and, 2500 and 3600 cm-1 and are averages of 3 scans. The spectra were smoothed, baseline corrected with a 7th order polynomial, normalised (0-1), smoothed and derivatised with a Savitzky-Golay algorithm and finally autoscaled.
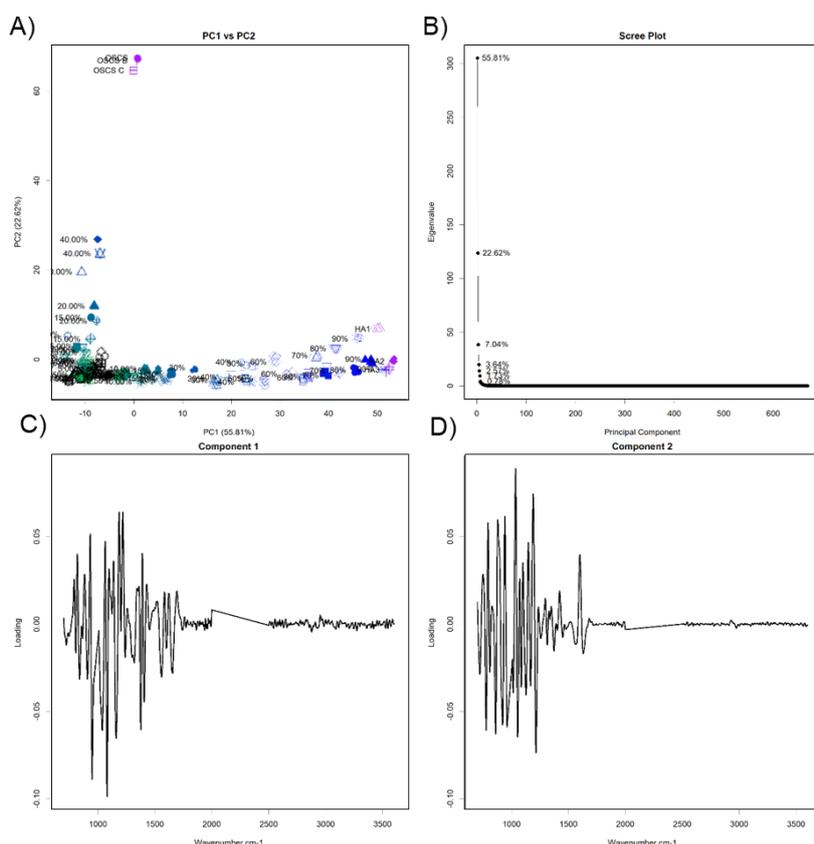
Fig.4.2: PCA of heparins blended with HA vs heparins blended with OSCS vs pharmaceutical heparins. A) Scatter plot of PC1 vs PC2 scores. **B)** Scree plot. **C)** Loadings plot for PC1. **D)** Loadings plot for PC2. Heparin: black, OSCS/HA: purple samples blended with OSCS/HA are indicated with a scale of green to blue – the more blue the sample the higher the level of OSCS/HA present. All spectra involved in PCA are in the regions of 700 and 2000 and, 2500 and 3600 cm^{-1} and are averages of 3 scans. The spectra were smoothed, baseline corrected with a 7th order polynomial, normalised (0-1), smoothed and derivatised with a savitzky-golay algorithm and finally autoscaled.

The heparins contaminated at the levels of 1, 2.5, 5, 7.5 and 10% (w/w) with HA were taken forward and a subsequent PCA performed (**Fig. 4.3 A,B**) and from PC1 a LRM constructed

[107]

(**Fig. 4.3 C,D**). The LRM had an R^2 of 0.71, the ROC had an AUC of 99.9% and at a 100% true positive, only 3.5% of samples were falsely identified as contaminated.
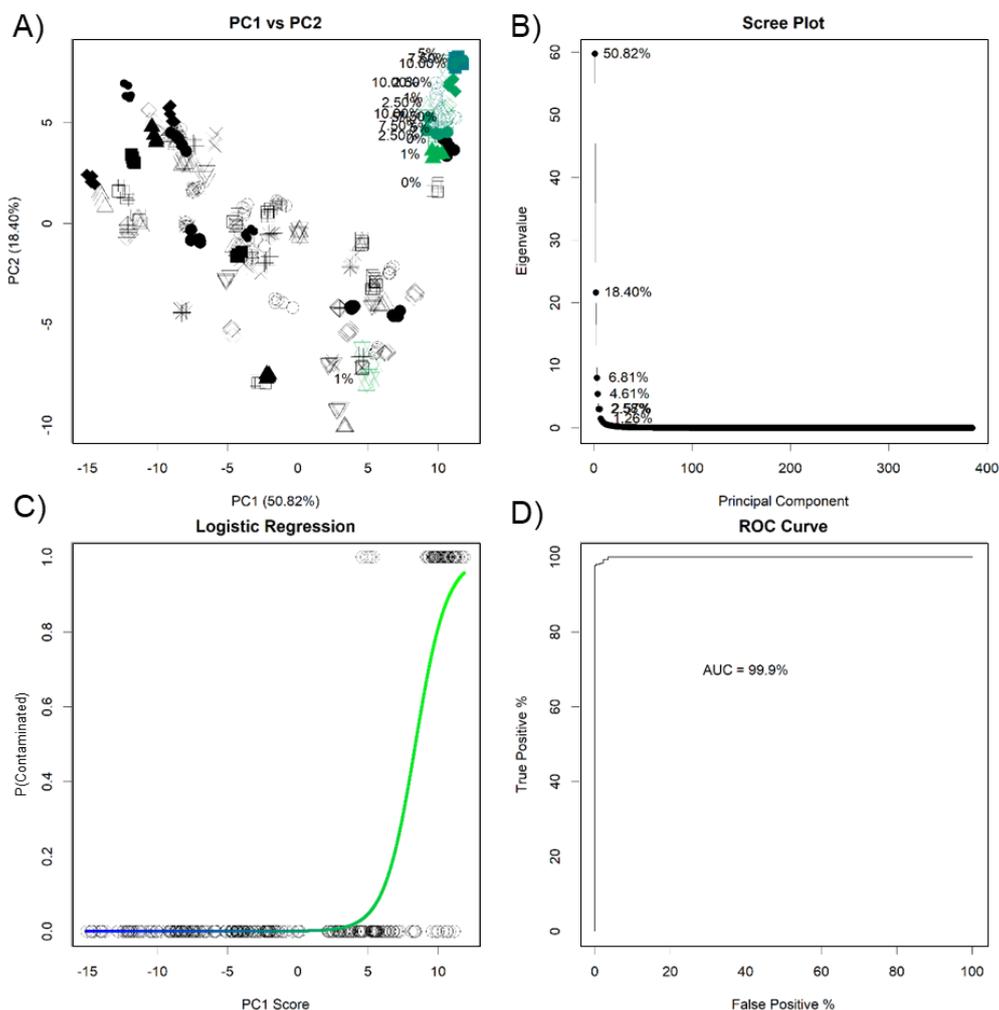
Figure 4.3: PCA to identify features associated with heparin blended with HA and subsequent logistic regression. A) Scatter plot of PC1 vs PC2 scores. **B)** Scree plot. **C)** LRM of pharmaceutical heparins (y = 0) vs heparins blended with HA (y = 1). **D)** ROC curve of responses. Heparin: black, OSCS/HA: purple samples blended with OSCS/HA are indicated with a scale of green to blue – the more blue the sample the higher the level of OSCS/HA present. All spectra involved in PCA are in the regions of 700 and 2000 and, 2500 and 3600 cm-1 and are averages of 3 scans. The spectra were smoothed, baseline corrected with a 7th order polynomial, normalised (0-1), smoothed and derivatised with a Savitzky-Golay algorithm and finally autoscaled.

4.3.2 Using PCA for contaminant identification and subsequent PCR for quantification.

The discovery that different components separate different contaminants suggests that FTIR-PCA may be able to discriminate between different types of contamination. While PCA-LRM is very powerful for its use-case; allowing the user to quickly and confidently identify a heparin to a certain standard, it is very poor at discriminating how or why the sample is of poor quality. Use of the LRMs in **Figures 3.29** and **4.3** would allow an analyst to determine if the failure is due to HA, or a sulphated polysaccharide, but is otherwise unable to discriminate between different contaminants. For example, current pharmaceutical heparin may have DS at the level of 5% (w/w) and, as the PCA-LRM can distinguish DS in some cases at the level of 2.5% (w/w)
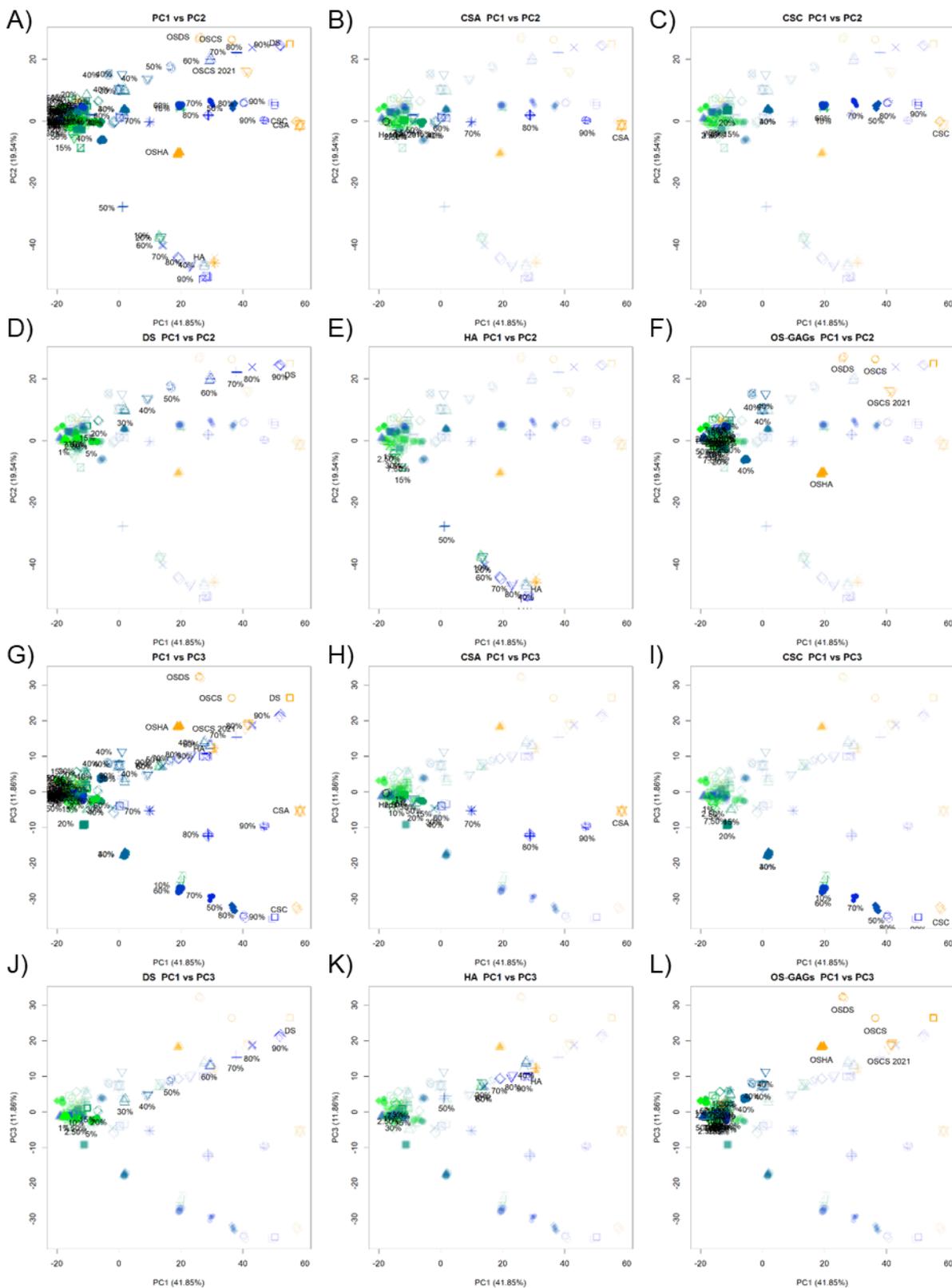
Figure 4.4.1: PCA score plots (PCs 1 vs 2 and 1 vs 3) for heparins and heparins blended with CSA, CSC, DS, HA and OS-GAGs. A through **D)** Scatter plots for PC1 vs PC2 scores. **G** through **L)** Scatter plots for PC1 vs PC3 scores. **A,G)** Original scores plots with no recolouring. **B,H)** Scores plots recoloured to visualise CSA. **C,I)** Scores plots recoloured to visualise CSC. **D,J)** Scores plots recoloured to visualise DS. **E,K)** Scores plots recoloured to visualise HA. **F,L)** Scores plots recoloured to visualise OS-GAGs. Details as to the analyses can be found in the legend for **Figure 4.4.3.**
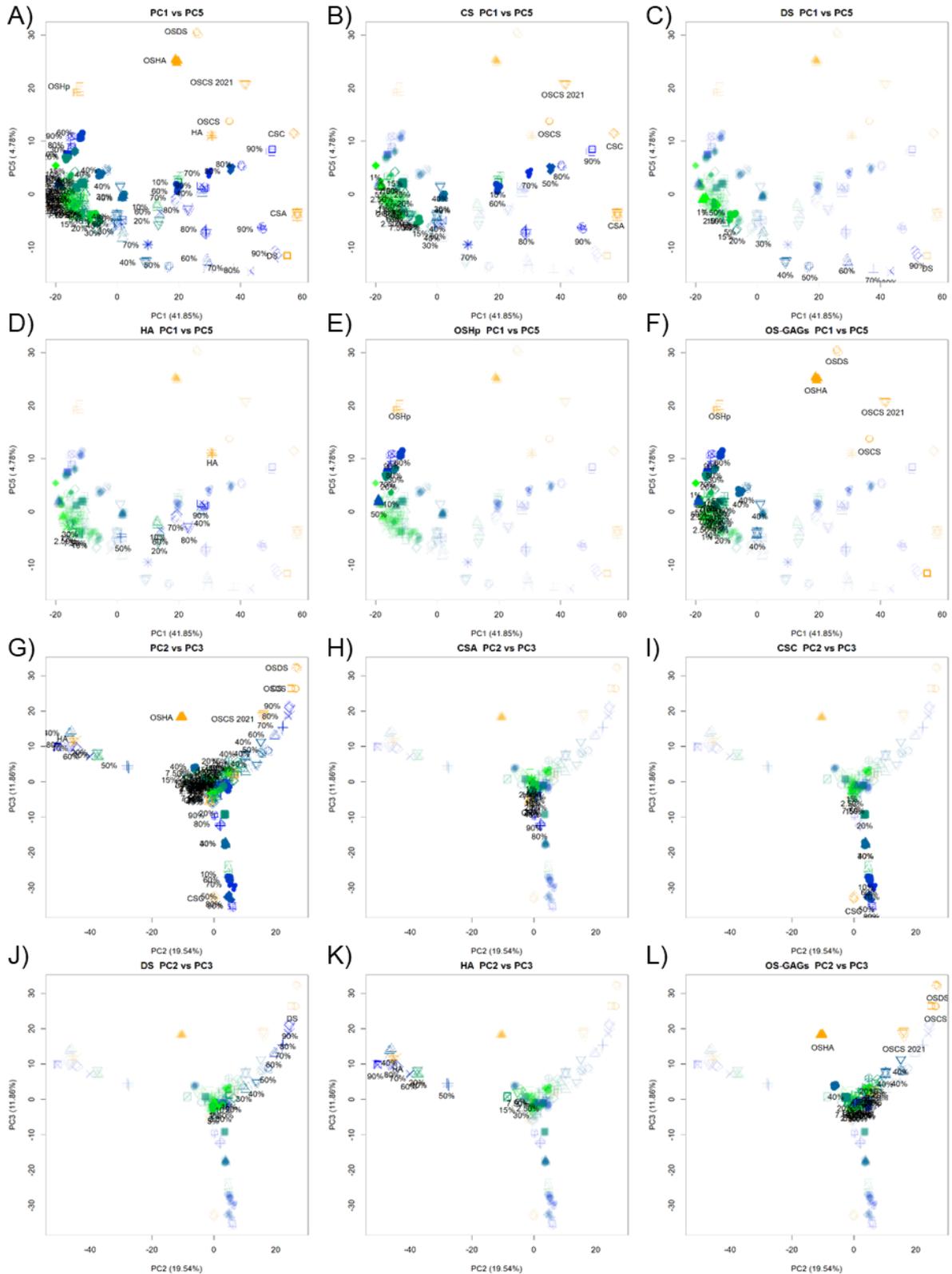
Figure 4.4.2: PCA score plots (PCs 1 vs 5 and 2 vs 3) for heparins and heparins blended with CSA, CSC, DS, HA and OS-GAGs. A through **D)** Scatter plots for PC1 vs PC5 scores. **G** through **L)** Scatter plots for PC2 vs PC3 scores. **A,G)** Original scores plots with no recolouring. **B)** Scores plots recoloured to visualise CSA and CSC, **C,J)** Scores plots recoloured to visualise DS. **D,K)** Scores plots recoloured to visualise HA. **E)** Scores plots recoloured to visualise OS-Hp. **F,L)** Scores plots recoloured to visualise OS-GAGs. **H)** Scores plots recoloured to visualise CSA. Details as to the analyses can be found in the legend for **Figure 4.4.3.**
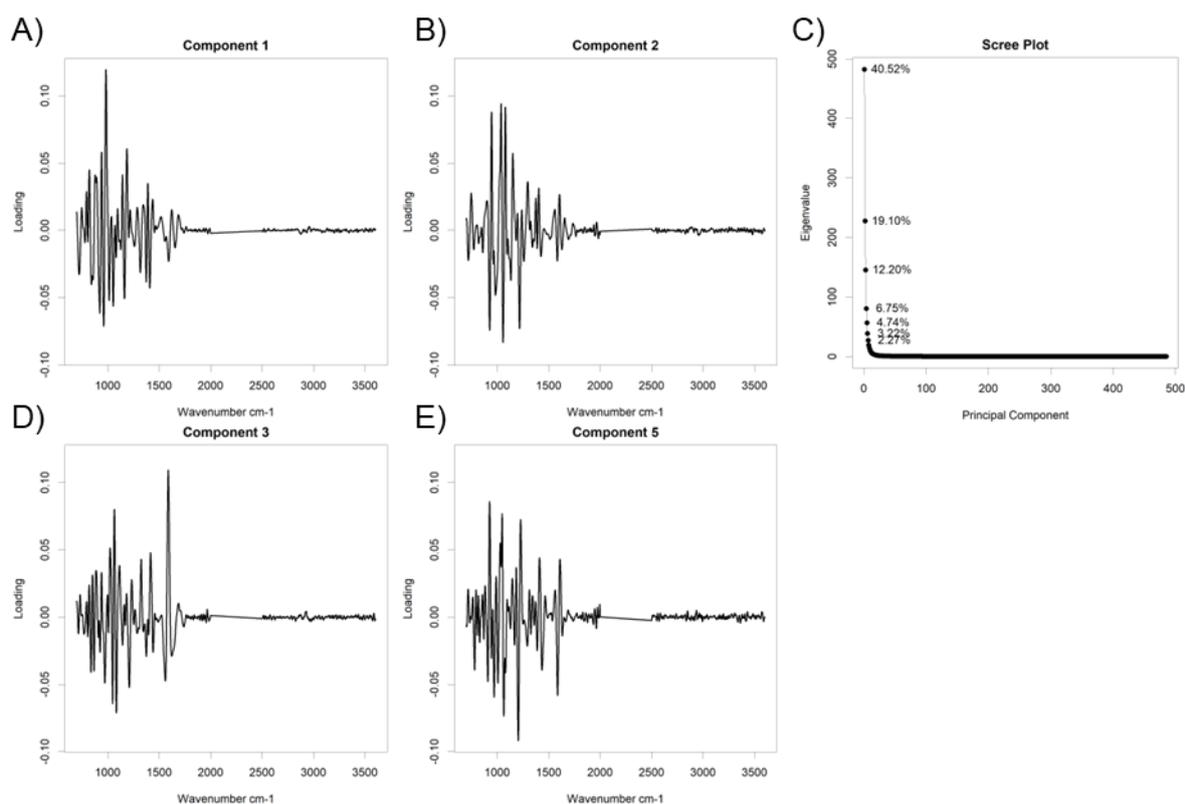
Figure 4.4.3: PCA Loading plots and scree plot for heparins and heparins blended with CSA, CSC, DS, HA and OS-GAGs.
A) Loadings plot for PC1. **B)** Loadings plot for PC2. **C)** Scree plot. **D)** Loadings plot for PC3. **E)** Loadings plot for PC5. In Figures 4.4.1 and 4.4.2, colours are: Heparin: black, pure CSA, CSC, DS, HA and OS-GAGs: orange, heparin samples blended with different GAGs are indicated with a scale of green to blue – the more blue the sample the higher the level of GAG present. All spectra involved in PCA are in the regions of 700 and 2000 and, 2500 and 3600 cm^{-1} and are averages of 3 scans. The spectra were smoothed, baseline corrected with a 7th order polynomial, normalised (0-1), smoothed and derivatised with a Savitzky-Golay algorithm and finally autoscaled.

the sample may be of pharmaceutical quality, but provide a false positive result. Furthermore, using multiple models is cumbersome and an ideal model would detect many contaminants at once. Hence, the contamination series that were used in the PCA-LRM model were probed with PCA (**Figs. 4.4.1-3**). For unmodified GAGs (i.e not OS-GAGs) the contaminated range was extended to the level of 100% (w/w) at 10% intervals.

In a similar manner to contamination with OSCS, a clear, almost linear separation is observed between heparin and heparins blended with the other GAGs. The separation is improved for different contaminants in different components. PC1 separates CSA, CSC and DS from heparin (**Figs. 4.4.1A,G** and **4.4.2A**) with some overlap of their OS-counterparts, and correlates with the level of heparin, accounting for 54.4% of the variance associated with the level of heparin (**Table 4.2**). PC2 separates strongly by the level of HA (**Fig. 4.4.1E**) (54.0% of the variance associated with HA) and weakly by the level of DS (**Fig. 4.4.1D**) (12.0% of the variance associated with DS) OSDS is also separated, but has little variance associated with PC2, so it is likely separated due to underlying DS signatures. PC3 separates CSC strongly (**Fig. 4.4.1I**) (45.1% of the variance associated with CSC is covered) and again DS weakly (**Fig. 4.4.1J**, variance covered = 8.2%). The OS-GAGs also separate upwards with DS (**Fig.**

4.4.1L). PC4 separates CSA (9.1%) and OSCS (23.0%) from DS (8.3%) and by extension OSDS (0.51%). PC5 separates the OSGAGs (**Fig. 4.4.2F**) and offers separation of OSHp (8.3%) from the heparins and other GAGs (**Fig. 4.4.2E**).

Table 4.2: Table of variance covered for each PC for heparin contaminated with each GAG between the levels of 1 and 100% (w/w). Values are computed with PCR, using the *pcr* function from the **pls** package.

PCR 1:100% Composition	*Variation Covered / %*							
	PC1	*PC2*	*PC3*	*PC4*	*PC5*	*PC6*	*PC7*	*PC8*
Total Variation	41.85	19.54	11.86	6.39	4.78	3.04	2.28	1.49
Heparin	54.36	0.01	1.14	0.99	6.74	4.47	0.02	0.66
CSA	1.38	0.00	2.21	9.14	7.07	13.85	17.99	1.91
CSC	7.36	0.77	45.09	0.68	1.72	1.43	7.97	1.04
OSCS	2.65	3.03	2.82	22.95	0.55	0.02	2.11	1.53
DS	8.77	12.03	8.16	8.28	21.98	4.03	0.08	0.08
OSDS	1.13	1.37	0.51	3.57	1.19	0.61	2.11	1.41
HA	1.91	53.97	4.05	0.39	0.96	0.39	0.16	0.04
OSHA	1.99	0.28	0.06	0.30	0.75	0.89	3.77	1.58
OSHp	3.14	0.30	0.04	16.29	8.25	7.54	1.34	16.44

As different GAG types and modified GAGs are separated in different components, different contaminants can be selected for based on the component(s) selected. In theory, a LRM could be made for each of these components to help determine which contaminant is present, but in practice this is laborious and it is likely that a combination of LRMs and detective work would be required to truly determine the identity of the alien species. A plot such as the one above (**Fig. 4.4**) could be used post PCA-LRM of contaminated samples however to help ascertain the contaminants present, but the separation is not as clear and samples at the levels of 5 - 20%(w/w) can become indistinguishable in some cases. It is likely that, as observed with OSCS blends (**Figs. 3.19** and **3.23**), contamination series that cover smaller percentages will provide better separation at the lower end (here, <40%), hence contamination series that covered the lower end were compared to the heparin library (**Fig. 4.5, Table 4.3**). The distinction between the identity of different blended GAGs was lost, and two components separated the majority of the blended GAGs with PC1 possessing the highest variation that covered the level of heparin (29.6%). PC2 separated blends of the un-sulphated GAG HA (19.9%) from heparin (**Fig. 4.5E**) while PC3 separated blends of the sulphated contaminants CSA, CSC, DS and the OS-GAGs (**Fig.4.5A**) from heparin however, the variation covered is low in the PCR (<7% for each feature) – most likely due to the encroachment of the heparin library over the lower end in PC2. Separation of all GAG blends is obtained through use of both components 2 and 3 in a manner analogous to **Figure 4.4**.

Samples contaminated to the levels of 1 to 5% (w/w) are close to the edge of the heparin cloud, but are often separated, in agreeance with the data generated from the PCA-LRM

model (**Table 4.1**) and with previous prediction attempts, where prediction of OSCS level led to a RMSE of 4.2%. Interestingly, the loadings observed here do not match up with the original PCA-LRM PC4 loading, but still separated OSCS, suggesting that there are multiple ways to detect OSCS (**Fig. 4.6**).
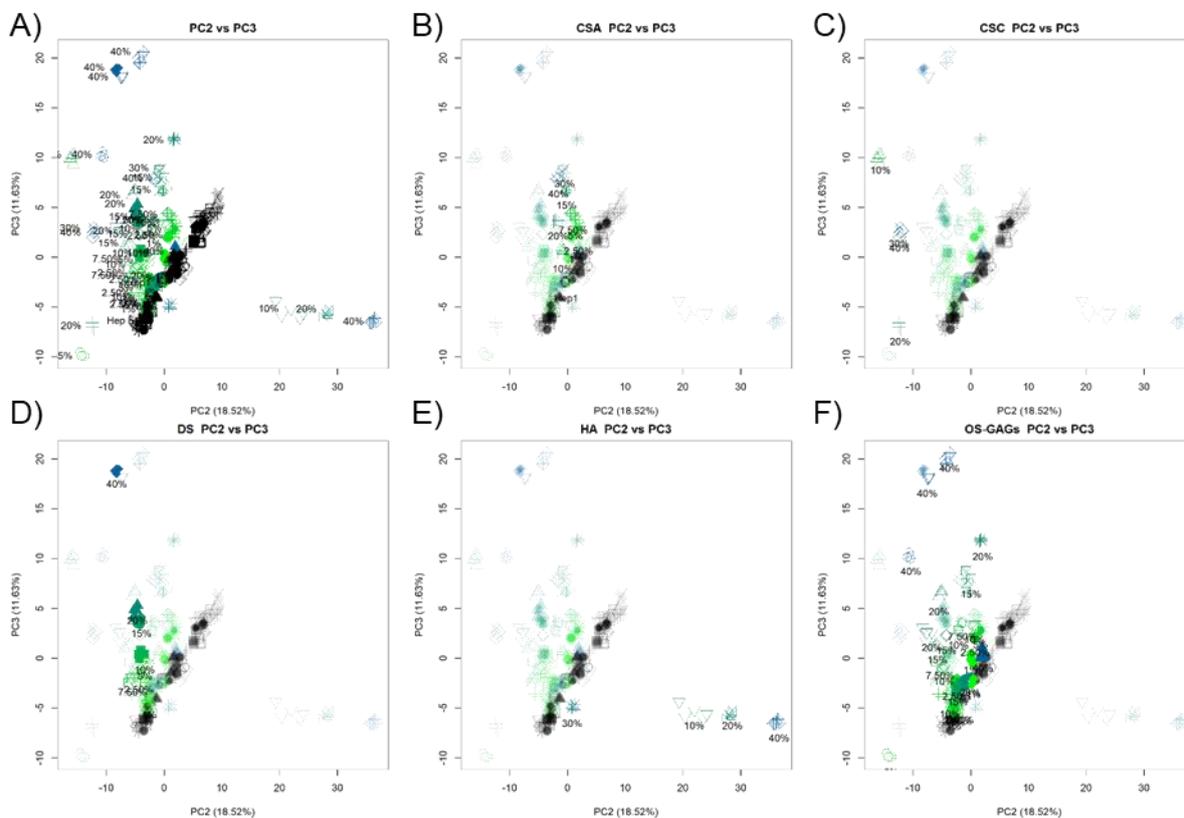
Figure 4.5: PCA score plots of PC2 vs PC3 for heparins and heparins blended with CSA, CSC, DS, HA and OS-GAGs to the maximum level of 40%. A) Original scores plots with no recolouring. **B)** Scores plots recoloured to visualise CSA. **C)** Scores plots recoloured to visualise CSC. **D)** Scores plots recoloured to visualise DS. **E)** Scores plots recoloured to visualise HA. **F)** Scores plots recoloured to visualise OS-GAGs. Heparin: black, heparin samples blended with different GAGs are indicated with a scale of green to blue – the more blue the sample the higher the level of GAG present. All spectra involved in PCA are in the regions of 700 and 2000 and, 2500 and 3600 cm^{-1} and are averages of 3 scans. The spectra were smoothed, baseline corrected with a 7th order polynomial, normalised (0-1), smoothed and derivatised with a Savitzky-Golay algorithm and finally autoscaled.

The linear nature of the PCA plots in **Figure 4.4** suggests that linear regression models may be able to predict levels of GAG in a heparin blend, and the orthogonal nature of the separation between different GAGs suggests that the GAGs present in the blend can be identified too. To test this, a library of multiple heparin:GAG titrations ($n \geq 3$) were created to account for some natural variation in the underlying GAG structures (as observed for heparins in **Figure 3.4**). For clarity in the plots the entire heparin library is excluded, but the heparins used to construct the blends are still present in the data processing.
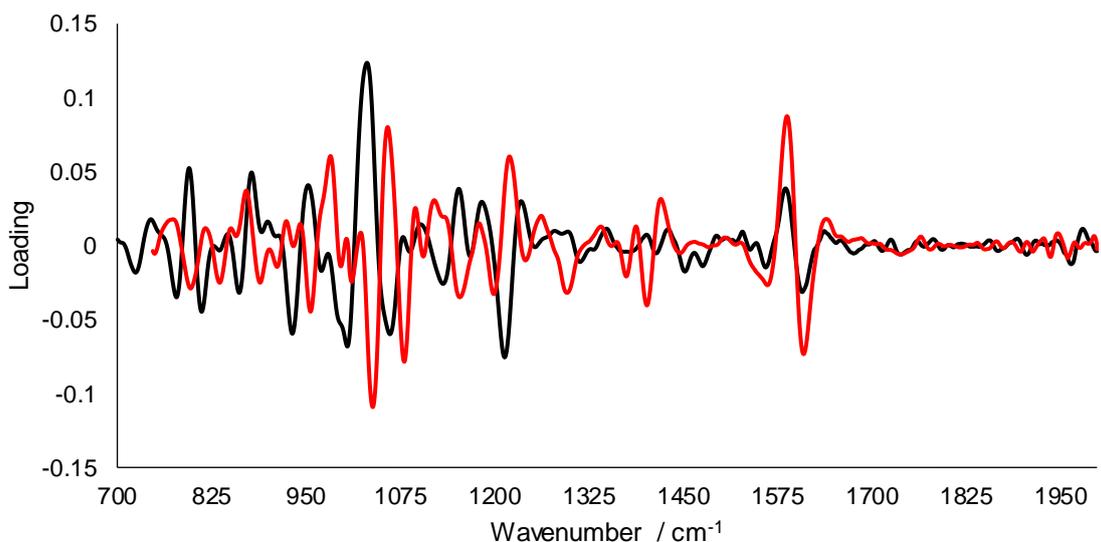
Figure 4.6: Loading plot of the PC used in PCA-LRM (Fig. 3.29; PC4), which is targeted at OSCS, compared with the loadings required to separate OSCS and other sulphated GAGs (Fig 4.5; PC2). Loading for PC4 from the PCA-LRM, used to detect OSCS: black, Loading from PC2 used to detect sulphated GAGs: red.

Table 4.3: Table of variance covered for each PC for heparin contaminated with each GAG between the levels of 1 and 20% (w/w). Values are computed with PCR, using the *pcr* function from the **pls** package.

PCR 1:20% Composition	*Variation Covered / %*							
	PC1	*PC2*	*PC3*	*PC4*	*PC5*	*PC6*	*PC7*	*PC8*
Total Variation	49.00	16.83	9.09	6.79	4.46	2.65	1.22	1.02
Hp	29.59	1.60	3.77	11.64	3.20	0.62	0.04	2.82
CSA	3.88	0.28	4.47	0.09	0.23	8.52	3.25	0.07
CSC	6.87	1.79	0.08	5.24	8.35	5.29	1.72	0.83
OSCS	0.28	1.02	4.19	1.29	21.25	0.14	5.55	8.23
DS	5.73	0.40	3.58	0.57	0.32	6.17	0.49	33.91
OSDS	4.38	0.71	6.33	0.09	1.14	22.99	7.96	2.28
HA	7.38	19.86	2.91	5.85	0.33	0.04	0.01	0.22
OSHA	5.04	0.52	0.70	1.89	0.00	1.92	0.02	8.21

PCA and subsequent PCR of the new library (**Fig 4.5, Table 4.4**) showed the same trends as **Figure 4.4**, PC1 separated by level of heparin, PC2 by level of DS and HA, PC3 by level of CSC and PC5 by level of CSA (**Table. 4.4**). To test if the PCA can predict GAG levels and discriminate between the constituents in the mixture, 15 test samples (T1 through 9 and T11 through 16) were made by blending a series of GAGs together, w/w. Care was taken to ensure that the GAGs were blended with partners and at levels that were not already in the library to synthesise unique combination as a test of robustness. In brief, T1 to 3 and 9 were single GAG in heparin blends at the level of 25% to 75% (w/w), T4 to 9 were double GAG to heparin blends, with the GAG mix totalling 25% (w/w) and heparin equalling 75% (w/w), T13:15 were

50:50 GAG:GAG blends and finally, T12 and T16 were triple GAG blends, at levels similar to those found in crude samples.

Through visual inspection of the PCR variance covered for each GAG, the first 5 components were taken forward as the components which would be used for prediction. The levels for each different GAG were computed from the PCR and averaged across the 5 individual spectral repeats for each sample. The root mean squared error (RMSE) was calculated between the real values and predicted values and from this, an R^2 value was calculated. The original and predicted means for each GAG level underwent a T-test to determine if the observed and predicted values were significantly different or not.

The PCR predicted the values for the samples correctly with P values of 0.57, 0.87, 0.91, 0.70, 0.67 for the level of heparin, CSA, CSC, DS, HA and OSCS respectively. As P > 0.01, the predictions are not significantly different from the expected values (**Table. 4.5**). Levels of CSC and HA has RMSE of 4.74 and 4.91 and hence were within ~5% of their real values, while levels of heparin, CSA and DS had RMSE of 7.26, 8.70 and 7.08 respectively. All predictions except for predictions of HA had high correlations between their predicted and expected values ($R^2 > 0.82$). It is likely that the poor R^2 value for HA is due to only one measurement.

The level of heparin is predicted correctly in almost all cases where heparin is present, but in GAG:GAG mixtures that do not contain heparin, it is overpredicted by up to 17.88%, taking away from CSA, CSC and DS predictions in T13, T14 and T15. OSCS is also overpredicted in T13 and T14, suggesting that signals in heparin and OSCS overlap with signals in CS and DS. In T12, a sample which contains heparin, the levels of both CSC and DS are underpredicted and the difference is reflected in the amount of heparin predicted, this is likely due to overlap of the 6S and 2S signals respectively. The spectral similarities between heparin, DS, CSA and CSC, can be observed in the PCR variances (**Table 4.4**). The largest portions of the variance covered by DS are primarily in PC1, 2, 3 and 5 by 23.4%, 30.6%, 9.3% and 19.6% respectively, while the largest portions of the variance covered by CSA are primarily in PC1 and 5 at 26.9% and 58.3% respectively, the largest portion of the variance covered by heparin is primarily in PC1 at 83.4% and the largest portion of the variance covered by CSC is primarily in PC3 at 62.7%. The components which corelate with DS overlap with the components which correlate with heparin, CSA and CSC and the component which correlates with heparin overlaps with a component which correlates with CSA. This may explain the difficulties observed in quantifying these samples.

Figure 4.7: PCA scores plots of heparins blended with CSA, CSC, DS, HA and OSCS (n ≥ 3 for each series). A) Score plot of PC1 vs PC2. **B)** Recoloured score plot of PC1 vs PC2 to highlight DS. **C)** Score plot of PC1 vs PC3. **D)** Recoloured score plot

of PC1 vs PC3 to highlight CSC. **E)** Score plot of PC1 vs PC5. **F)** Recoloured score plot of PC1 vs PC5 to highlight CSA. Heparin: black, CSA, CSC, DS, HA and OSCS: orange, heparin samples blended with different GAGs are indicated with a scale of green to blue – the more blue the sample the higher the level of GAG present. All spectra involved in PCA are in the regions of 700 and 2000 and, 2500 and 3600 cm-1 and are averages of 3 scans. The spectra were smoothed, baseline corrected with a 7th order polynomial, normalised (0-1), smoothed and derivatised with a Savitzky-Golay algorithm and finally autoscaled. Spectra used for comparison are between 400 and 4000 cm^{-1}, are averages of 5 repeats of 3 scans that were first smoothed and baseline corrected with a 7th order polynomial before normalisation (0-1).

Table 4.4: Table of variance covered for each PC from the PCA in Figure 4.12. Values are computed with PCR, using the pcr function from the pls package.

Feature	Variance Covered / %							
	PC1	**PC2**	**PC3**	**PC4**	**PC5**	**PC6**	**PC7**	**PC8**
Total Variance	39.82	16.53	11.26	9.58	5.72	2.24	2.08	1.56
CSA	26.86	0.08	0.29	1.49	56.32	0.76	3.33	2.89
DS	23.38	30.60	9.30	1.14	19.60	3.17	0.36	2.24
CSC	2.84	5.77	62.65	1.07	9.21	1.13	0.01	0.07
HA	0.69	57.21	25.45	0.08	3.27	0.69	0.36	1.32
OSCS	1.47	4.00	0.08	0.71	0.76	4.51	1.78	1.80
Hp	83.43	3.87	0.00	0.74	1.38	1.47	0.81	0.67

Table. 4.5: PCR predictions of GAG composition for 15 simulated GAGs. PCA scores from the PCA shown in **Figure 4.7** were used to build the PCR model. Samples to undergo prediction had their PC scores predicted for the relevant PCA using the *predict* function in **base** R. The levels of GAG were predicted using the *predict* function from the **pls** package on the relevant PCR and predicted PC scores. The top 5 components were used for prediction.

Sample	Percentage of Component / %											
	Heparin		**CSA**		**CSC**		**DS**		**HA**		**OSCS**	
	Original	Predicted	Original	Predicted	Original	Predicted	Original	Predicted	Original	Predicted	Original	Predicted
T1	75.00	72.79 +/- 1.95	0.00	-1.65 +/- 2.78	25.00	19.15 +/- 1.15	0.00	4.39 +/- 2.68	0.00	3.7 +/- 1.21	NA	1.62 +/- 0.2
T2	75.00	79.68 +/- 1.81	0.00	-0.49 +/- 2.17	0.00	5.28 +/- 0.75	25.00	7.11 +/- 1.39	0.00	6.95 +/- 1.31	NA	1.47 +/- 0.05
T3	75.00	79.82 +/- 1.39	25.00	14.02 +/- 1.55	0.00	3.14 +/- 1.28	0.00	0.98 +/- 1.92	0.00	0.9 +/- 0.53	NA	1.13 +/- 0.17
T4	75.00	78.04 +/- 0.43	12.50	8.99 +/- 1.22	12.50	16.11 +/- 0.32	0.00	-3.82 +/- 0.72	0.00	0.36 +/- 0.57	NA	0.32 +/- 0.06
T5	75.00	80.93 +/- 0.59	18.75	11.43 +/- 0.9	6.25	6.89 +/- 0.46	0.00	0.4 +/- 0.56	0.00	-1.3 +/- 0.05	NA	1.65 +/- 0.05
T6	75.00	77.96 +/- 0.73	6.25	9.04 +/- 0.73	18.75	16.17 +/- 0.56	0.00	-3.62 +/- 0.28	0.00	-0.62 +/- 0.27	NA	1.07 +/- 0.06
T7	75.00	77.51 +/- 0.99	6.25	3.16 +/- 0.85	0.00	1.85 +/- 0.52	18.75	13.09 +/- 0.82	0.00	0.99 +/- 0.39	NA	3.4 +/- 0.06
T8	75.00	81.34 +/- 0.53	0.00	0.91 +/- 0.57	6.25	4.07 +/- 0.79	18.75	11.12 +/- 0.56	0.00	-0.83 +/- 0.32	NA	3.4 +/- 0.03
T9	75.00	82.13 +/- 0.78	18.75	9.29 +/- 0.58	0.00	1.74 +/- 0.67	6.25	4.96 +/- 0.9	0.00	-0.06 +/- 0.79	NA	1.95 +/- 0.08
T11	75.00	79.96 +/- 1.04	0.00	-2.83 +/- 0.76	0.00	6.33 +/- 0.91	0.00	7.59 +/- 0.76	25.00	7.68 +/- 0.37	NA	1.27 +/- 0.1
T12	55.00	64.99 +/- 0.47	15.00	13.08 +/- 1.1	20.00	16.52 +/- 0.6	10.00	5.51 +/- 0.9	0.00	-2.83 +/- 0.47	NA	2.74 +/- 0.07
T13	0.00	17.88 +/- 0.62	0.00	0.42 +/- 0.37	50.00	36.17 +/- 0.23	50.00	39.03 +/- 0.58	0.00	-2.37 +/- 0.36	NA	8.88 +/- 0.06
T14	0.00	9.45 +/- 0.41	50.00	21.33 +/- 0.66	0.00	-3.03 +/- 0.34	50.00	60.38 +/- 0.67	0.00	0.34 +/- 0.5	NA	11.53 +/- 0.09
T15	0.00	8.82 +/- 0.49	50.00	41 +/- 1.41	50.00	50.08 +/- 1.08	0.00	0.07 +/- 1.12	0.00	-1.7 +/- 0.41	NA	1.73 +/- 0.09
T16	65.00	70.54 +/- 1.03	10.00	9.38 +/- 1.08	20.00	18.26 +/- 1.1	5.00	-0.96 +/- 1.45	0.00	1.61 +/- 0.91	NA	1.17 +/- 0.12
RMSE		7.26		8.70		4.74		7.08		4.91		NA
R2		0.99		0.82		0.95		0.88		0.38		NA
P-value		0.57		0.87		0.91		0.70		0.67		NA

It is possible that the failure of the PCR from Table 4.5 to discern CSA, CSC and DS is due to a lack of samples in the library that represent these mixtures. Some amount of heparin is present in almost every sample in the library, likely explaining why heparin is overpredicted in samples T13, 14 and 15. To improve the model, mixtures of the GAG species (CSA:CSC, DS:CSA and DS:CSC) were produced at the levels of 20:80, 40:60, 60:40 and 20:80 w/w. Two DS:CSA and DS:CSC blends were formulated, as DS blends were noted to separate from each other in some components, hence DS samples that cover these separations were

[117]

included. A new PCA was performed (**Fig. 4.8**) and a subsequent PCR was calculated, including the GAG:GAG mixtures (termed series M1 through 5) and the composition of the T-samples predicted (**Table. 4.7**). All RMSE values were reduced from 7.26, 8.70, 4.74, 7.08 and 4.91 to 5.60, 6.98, 4.11, 6.96 and 4.96 for heparin, CSA, CSC, DS and HA respectively. All R^2 values stayed the same or increased from 0.99, 0.82, 0.95 and 0.88 to 0.99, 0.89, 0.96 and 0.88 for heparin, CSA , CSC and DS but decreased from 0.38 to 0.37 for HA. All P-values remained insignificant at 0.67, 0.98, 0.99, 0.74 and 0.59 for heparin, CSA, CS, DS and HA respectively. The GAG:GAG mixtures (T13:15) improved the most, with overpredicted heparin reducing from 17.88% 9.45% and 8.82% to 12.50%, 4.98% and 5.98% (w/w) for samples T13, 14 and 15 respectively. There was still some overlap between heparin, CSA and DS as these three still share their main PCs (1, 3 and 5) however, the percentage of the variance described by DS in each of these components becomes more homogenous (**Table 4.6**).
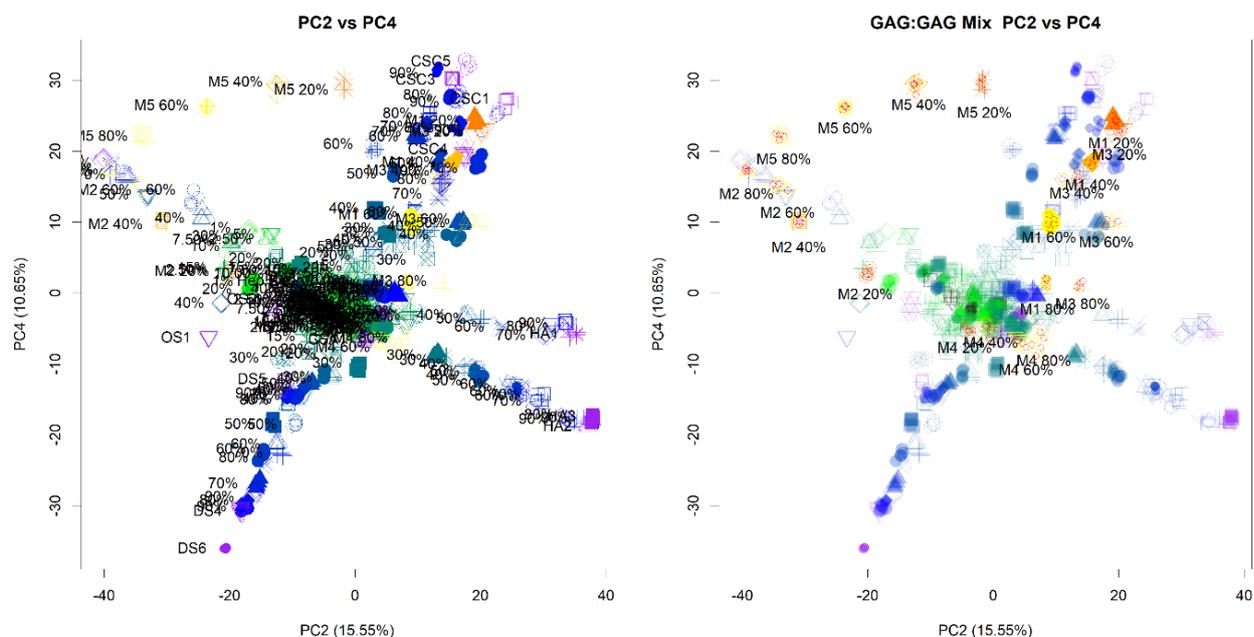
Figure 4.8: PCA of heparin:GAG blends and GAG:GAG mixtures. Left) Scores plot of PC2 vs PC4. **Right)** Recoloured scores plot of PC2 vs PC4 to show the newly included GAG:GAG mixtures. Heparin: black, CSA, CSC, DS, HA and OSCS: purple, heparin samples blended with different GAGs are indicated with a scale of green to blue – the more blue the sample the higher the level of GAG present. For the new GAG:GAG mixtures, a scale from yellow to orange is used to indicate the levels of each GAG. All spectra involved in PCA are in the regions of 700 and 2000 and, 2500 and 3600 cm-1 and are averages of 3 scans. The spectra were smoothed, baseline corrected with a 7th order polynomial, normalised (0-1), smoothed and derivatised with a Savitzky-Golay algorithm and finally autoscaled. Spectra used for comparison are between 400 and 4000 cm^{-1}, are averages of 5 repeats of 3 scans that were first smoothed and baseline corrected with a 7th order polynomial before normalisation (0-1).

Table 4.6: Table of variance covered for each PC from the PCA in Figure 4.8. Values are computed with PCR, using the *pcr* function from the ***pls*** package.

	Variance Covered / %							
Feature	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
Total Variance	40.76	15.55	10.92	10.66	5.31	2.16	1.93	1.49
CSA	29.59	0.01	0.04	0.03	50.18	1.13	0.55	5.19
DS	27.71	25.17	0.05	5.64	19.78	0.04	1.84	3.44
CSC	5.37	12.02	12.13	48.26	6.32	0.08	0.99	0.01
HA	0.04	34.36	37.42	9.47	3.98	0.00	0.90	1.51
OSCS	0.93	3.33	0.15	0.02	0.31	3.40	3.26	0.41
Hp	85.80	2.47	1.10	0.74	1.24	0.29	1.46	0.29

Table. 4.7: PCR predictions of GAG composition for 15 simulated GAGs using GAG:GAG mixtures in the PCR. PCA scores from the PCA shown in **Figure 4.13** were used to create the PCR model. Here, a set of GAG:GAG mixtures are also introduced to facilitate better separation. Samples to undergo prediction had their PC scores predicted for the relevant PCA using the *predict* function in ***base*** R. The levels of GAG were predicted using the *predict* function from the ***pls*** package on the relevant PCR and predicted PC scores. The top 5 components were used for prediction.

Sample	Percentage of Component / %											
	Heparin		CSA		CSC		DS		HA		OSCS	
	Original	Predicted	Original	Predicted	Original	Predicted	Original	Predicted	Original	Predicted	Original	Predicted
T1	75.00	71.84 +/- 1.98	0.00	-0.95 +/- 2.63	25.00	19.69 +/- 1.06	0.00	4.79 +/- 2.73	0.00	3.53 +/- 1.15	NA	1.12 +/- 0.14
T2	75.00	79.33 +/- 1.89	0.00	-0.05 +/- 2.23	0.00	5.4 +/- 0.69	25.00	7.11 +/- 1.5	0.00	6.99 +/- 1.27	NA	1.24 +/- 0.04
T3	75.00	79.78 +/- 1.44	25.00	13.87 +/- 1.38	0.00	3.27 +/- 1.25	0.00	0.97 +/- 1.91	0.00	1 +/- 0.49	NA	1.11 +/- 0.12
T4	75.00	77.71 +/- 0.5	12.50	7.92 +/- 1.2	12.50	16.31 +/- 0.35	0.00	-2.94 +/- 0.76	0.00	0.37 +/- 0.56	NA	0.6 +/- 0.03
T5	75.00	80.56 +/- 0.57	18.75	12.06 +/- 0.95	6.25	7.38 +/- 0.53	0.00	-0.01 +/- 0.58	0.00	-1.17 +/- 0.26	NA	1.21 +/- 0.03
T6	75.00	77.28 +/- 0.84	6.25	8.99 +/- 0.77	18.75	16.96 +/- 0.64	0.00	-3.39 +/- 0.31	0.00	-0.61 +/- 0.27	NA	0.77 +/- 0.02
T7	75.00	76.34 +/- 1.08	6.25	4.95 +/- 0.88	0.00	2.67 +/- 0.53	18.75	12.81 +/- 0.86	0.00	1.05 +/- 0.38	NA	2.29 +/- 0.05
T8	75.00	80.24 +/- 0.49	0.00	2.84 +/- 0.5	6.25	5.06 +/- 0.74	18.75	10.5 +/- 0.53	0.00	-0.75 +/- 0.3	NA	2.21 +/- 0.04
T9	75.00	81.88 +/- 0.77	18.75	9.86 +/- 0.61	0.00	2.1 +/- 0.71	6.25	4.58 +/- 0.93	0.00	0.06 +/- 0.77	NA	1.54 +/- 0.06
T11	75.00	79.7 +/- 1.1	0.00	-2.51 +/- 0.75	0.00	6.16 +/- 0.88	0.00	7.72 +/- 0.76	25.00	7.76 +/- 0.38	NA	1.17 +/- 0.06
T12	55.00	63.48 +/- 0.45	15.00	14.1 +/- 1.07	20.00	17.69 +/- 0.64	10.00	6.02 +/- 0.88	0.00	-2.96 +/- 0.47	NA	1.73 +/- 0.05
T13	0.00	12.5 +/- 0.65	0.00	7.3 +/- 0.4	50.00	38.86 +/- 0.25	50.00	40.25 +/- 0.63	0.00	-3.32 +/- 0.35	NA	4.7 +/- 0.05
T14	0.00	4.98 +/- 0.41	50.00	30.87 +/- 0.59	0.00	-1.14 +/- 0.3	50.00	59.77 +/- 0.64	0.00	-0.74 +/- 0.49	NA	6.65 +/- 0.06
T15	0.00	5.98 +/- 0.52	50.00	41.89 +/- 1.35	50.00	50.86 +/- 1.08	0.00	3.33 +/- 1.11	0.00	-2.94 +/- 0.41	NA	0.87 +/- 0.07
T16	65.00	69.4 +/- 1.08	10.00	9.31 +/- 1.08	20.00	19.08 +/- 1.14	5.00	-0.18 +/- 1.42	0.00	1.55 +/- 0.89	NA	0.85 +/- 0.1
RMSE		5.60		6.98		4.11		6.96		4.96		NA
R2		0.99		0.89		0.96		0.88		0.37		NA
P-value		0.67		0.98		0.99		0.74		0.59		NA

4.3.3 Testing the robustness of PCR prediction with novel marine GAG samples.

To test the robustness of the method, 7 crude GAG samples, which had been extracted from marine sources and characterised in terms of level of heparin, CS and DS using NMR (**Table.4.8)** by (Mycroft-West, 2021) underwent ATR-FTIR spectral acquisition in order to undergo GAG quantification. To quantify the GAGs in each of the crude samples, PC scores were predicted using the PCA that contained the MGAGs (**Fig. 4.8, Table 4.7**) and input into the PCR model.

Table. 4.8: PCR predictions of GAG composition for 7 crude marine GAG mixtures, using GAG:GAG mixes in the PCR. PCA scores from the PCA shown in **Figure 4.8** were used to create the PCR model. Samples to undergo prediction had their PC scores predicted for the relevant PCA using the *predict* function in **base** R. The levels of GAG were predicted using the *predict* function from the **pls** package on the relevant PCR and predicted PC scores. The top 5 components were used for prediction

Sample		% of composition by NMR			% of composition by PCR						
Scientific Name	Colloquial Species	HS/heparin	CS	DS	Heparin	CS Total	CSA	CSC	OSCS	DS	HA
P.pelagicus	Crab	70	30	1	71.31 +/- 0.11	23.87 +/- 0.56	2.23 +/- 0.36	19.93 +/- 0.19	1.71 +/- 0.01	7.7 +/- 0.22	-2.85 +/- 0.15
L.vannamei	Prawn	60	40	5	62.21 +/- 1.1	20.47 +/- 1.69	-0.67 +/- 0.64	19.91 +/- 0.97	1.24 +/- 0.09	9.99 +/- 1.25	7.34 +/- 0.7
S.pilchardus	Pilchard	10	78	12	15.86 +/- 0.32	56.94 +/- 1.15	37.08 +/- 0.55	16.73 +/- 0.57	3.13 +/- 0.03	30.51 +/- 0.33	-3.3 +/- 0.29
M.aeglefinus	Haddock	2	58	42	-5.88 +/- 0.26	50.78 +/- 0.61	63.75 +/- 0.4	-18.07 +/- 0.2	5.09 +/- 0.01	53.63 +/- 0.25	1.58 +/- 0.18
M.merluccius	Hake	25	63	12	-9.36 +/- 0.16	57.9 +/- 0.49	68.28 +/- 0.22	-15.09 +/- 0.26	4.7 +/- 0.02	49.16 +/- 0.24	2.41 +/- 0.21
C.batrachus	Catfish	5	90	5	4.41 +/- 0.55	70.87 +/- 1.62	57.26 +/- 0.87	9.89 +/- 0.72	3.73 +/- 0.03	32.54 +/- 0.21	-7.77 +/- 0.72
O.gorbuscha	Salmon	10	60	30	-11.64 +/- 0.15	37.84 +/- 0.44	54.19 +/- 0.24	-22.22 +/- 0.18	5.87 +/- 0.02	67.93 +/- 0.21	5.97 +/- 0.02

The predictions of heparin were accurate in some samples but incorrect in others. For example, the level of heparin in crab, prawn, pilchard and catfish was predicted at 71.09%, 62.05%, 15.51% and 4.12% (w/w) compared with the expected values of 70%, 60%, 10% and 5% (w/w) respectively. For the other samples, the level was vastly underpredicted at -5.96%, -9.44% and -11.57% (w/w) compared with expected values of 2%, 25% and 10% (w/w) for haddock, hake and salmon respectively. The level of CS and DS was difficult to distinguish between, with DS overpredicted and CS under predicted in all cases.

As an improvement in the predictions was observed after the addition of GAG:GAG mixtures, it was hypothesised that the library did not contain enough samples that represented the crude GAGs, hence the T-samples, which were originally produced to test the library were input into the library and the composition of the crude GAGs predicted against the new library (Table.4.9).

Table. 4.9: PCR predictions of GAG composition for 7 crude marine GAG mixtures, using more GAG:GAG mixes in the PCR. A new PCA was performed, using the samples from the PCA shown in **Figure 4.8** and the T-GAGs that were formulated as test samples, that used to create the PCR model. Samples to undergo prediction had their PC scores predicted for the relevant PCA using the *predict* function in **base** R. The levels of GAG were predicted using the *predict* function from the **pls** package on the relevant PCR and predicted PC scores. The top 5 components were used for prediction.

Sample		% of composition by NMR			% of composition by PCR						
Scientific Name	Colloquial Species	HS/heparin	CS	DS	Heparin	CS Total	CSA	CSC	OSCS	DS	HA
P.pelagicus	Crab	70	30	1	71.09 +/- 0.11	24.08 +/- 0.56	2.67 +/- 0.37	19.89 +/- 0.19	1.52 +/- 0.01	7.69 +/- 0.22	-2.83 +/- 0.14
L.vannamei	Prawn	60	40	5	62.05 +/- 1.1	20.54 +/- 1.67	-0.47 +/- 0.63	19.87 +/- 0.96	1.14 +/- 0.08	10.08 +/- 1.24	7.35 +/- 0.7
S.pilchardus	Pilchard	10	78	12	15.51 +/- 0.33	57.17 +/- 1.15	37.11 +/- 0.56	16.93 +/- 0.57	3.12 +/- 0.02	30.58 +/- 0.34	-3.24 +/- 0.29
M.aeglefinus	Haddock	2	58	42	-5.96 +/- 0.26	51.11 +/- 0.62	64.19 +/- 0.4	-18.11 +/- 0.2	5.03 +/- 0.01	53.33 +/- 0.25	1.62 +/- 0.17
M.merluccius	Hake	25	63	12	-9.44 +/- 0.15	58.2 +/- 0.49	68.71 +/- 0.22	-15.14 +/- 0.25	4.63 +/- 0.02	48.87 +/- 0.24	2.47 +/- 0.21
C.batrachus	Catfish	5	90	5	4.12 +/- 0.55	71.11 +/- 1.61	57.53 +/- 0.87	9.94 +/- 0.71	3.64 +/- 0.03	32.49 +/- 0.21	-7.66 +/- 0.72
O.gorbuscha	Salmon	10	60	30	-11.57 +/- 0.16	38.08 +/- 0.44	54.57 +/- 0.24	-22.33 +/- 0.19	5.84 +/- 0.02	67.57 +/- 0.21	6.02 +/- 0.02

A very small improvement in the level of heparin in crab, prawn and pilchard GAGs was observed, with the error decreasing from 1.31, 2.2 and 5.86 to 1.09, 2.05 and 5.51 respectively. The error in the catfish prediction increased from 0.59 to 0.88. The levels of CS and DS were still overlapping with no improvement of the quantification observed.

It was hypothesised that the unique structures found in marine GAGs may be attenuating the ability of the model to predict the levels of CS and DS accurately. Five crude heparins, the compositions of which had been characterised with SAX chromatography underwent compositional prediction with ATR-FTIR-PCR using the same PCR model that was used for the crude GAGs, specifically the model which contained the MGAGs and T-samples described earlier (**Table 4.9**). The RMSE of heparin and CS was 5.26 and 4.42, which are in line with the RMSE of 5.60 and 5.54 observed in the predictions of the T-samples in **Table 4.5** respectively. DS had the lowest RMSE at 2.88, which is drastically reduced from the RMSE of 6.96 observed earlier. This suggests that CS and DS can be separated from each other and quantified.

Table 4.10: PCR predictions of GAG composition for 5 crude heparin samples. PCA scores from the PCA used in **Table 4.9** were used to create the PCR model. Samples to undergo prediction had their PC scores predicted for the relevant PCA using the *predict* function in ***base*** R. The levels of GAG were predicted using the *predict* function from the ***pls*** package on the relevant PCR and predicted PC scores. The top 5 components were used for prediction.

	Sample	Composition / %					
		Hep		DS		CS	
		Expected	Predicted	Expected	Predicted	Expected	Predicted
Crude Heparin	G12131	84.60	77.8 +/- 0.43	13.50	14.38 +/- 0.12	1.80	6.16 +/- 0.49
	G12138	91.70	86.3 +/- 0.56	7.90	7.02 +/- 0.29	0.00	5.35 +/- 0.53
	G12161	86.70	77.5 +/- 0.72	12.00	14 +/- 0.22	1.30	7.59 +/- 0.74
	G12179	100.00	93.27 +/- 0.65	0.00	4.05 +/- 0.26	0.00	-2.24 +/- 0.5
	G9913	90.20	83.03 +/- 0.3	7.60	7.78 +/- 0.23	2.20	6.98 +/- 0.57
	Average	89.34 +/- 4.12	84.87 +/- 3.76	8.30 +/- 3.43	10.08 +/- 3.36	2.29 +/- 2.00	6.13 +/- 2.92
	RMSE	5.26		2.88		4.42	

From the quantifications achieved on crude heparins (**Table 4.10**), it was clear that DS and CS in a GAG mixture could be identified and quantified to within 5% of the expected level. It is possible that the improvement in separation compared with marine GAGs was achieved either due to SAX quantification instead of NMR, a lack of samples in the library that adequately reflect the marine GAG structures or that the IR signals overlap too much and cannot be distinguished. NMR analysis of constituent GAGs uses peak volumes associated with IdoA-GalN to quantify DS – since DS may not necessarily contain 100% IdoA, the levels of DS may be calculated incorrectly using NMR whilst SAX chromatography separates the whole molecules by charge, separating for molecules that contain some IdoA, which is canonically labelled DS.

Evidence of overlapping signals can be seen in **Table 4.4,** where components that describe variance associated with DS overlap components that describe variance associated with heparin, CSA and, CSC and a component that describes a portion of variance associated with heparin overlaps with a component that a portion of the variance associated with CSA. To

investigate whether the signals overlap too much, another technique – independent components analysis (ICA) was used. ICA is similar to PCA, but instead of seeking principal components that are orthogonal and calculated based on maximum variance within the data, ICA seeks directions in the data that correlate with independence; mathematically described as a reduction in gaussianity. ICA is most commonly explained with the "cocktail party" problem, whereby an analyst listens to a single conversation and, despite the presence of many other overlapping conversations, understands the target conversation by unmixing the input conversations into their constituents. ICA can be illustrated as (**eq. 4.1)**

$$X = CS$$

(**Eq. 4.1**)

whereby, X denotes the input matrix, C denotes the "mixing" matrix and S the underlying signals that constitute X.

C describes the amount of each constituent signal, S that is required to make up the original, but it is unitless and often has an arbitrary sign. Unlike PCA, latent signals do not have to be orthogonal to each other and they can correlate. ICA has been applied to FTIR spectra previously, where it has amongst many other uses, separated constituent atmospheric gasses (Ju et al., 2020), characterised plant rhizome extracts (Wang et al., 2009) and has allowed for detection of monosaccharides in solution (Hahn and Yoon, 2006). Theoretically, ICA should detect component spectra within the sample and therefore reveal their contaminants.

Before the component spectra could be extracted, it is important to first establish whether ATR-FTIR can identify individual GAG species from the spectra that may be extracted . Mainreck et al., 2011 showed that GAGs can be discriminated with FTIR microscopy of solid films and PCA hence it is likely that ATR-FTIR can do the same. A library of 243 GAGs was assembled, comprising 176 heparins, 13 CSA samples, 16 CSC samples, 21 DS samples, 11 HA samples and 6 OSCS samples, against which the identities of the major GAG components of the crude marine samples (given in **Table 4.8**) could be analysed using PCA (**Fig.4.9**). Each GAG fell into its own distinct group, forming heparin, HS, HA, CSA, CSC, DS and OSCS regions (**Fig 4.9A**), analogous to the trends observed by (Mainreck et al., 2011). Samples with more than 50% (w/w) heparin/HS appear in the heparin/HS region of the scores plot PC1 vs PC2 and samples with less than 50% (w/w) appear in the CS/DS region (**Fig. 4.9A**).
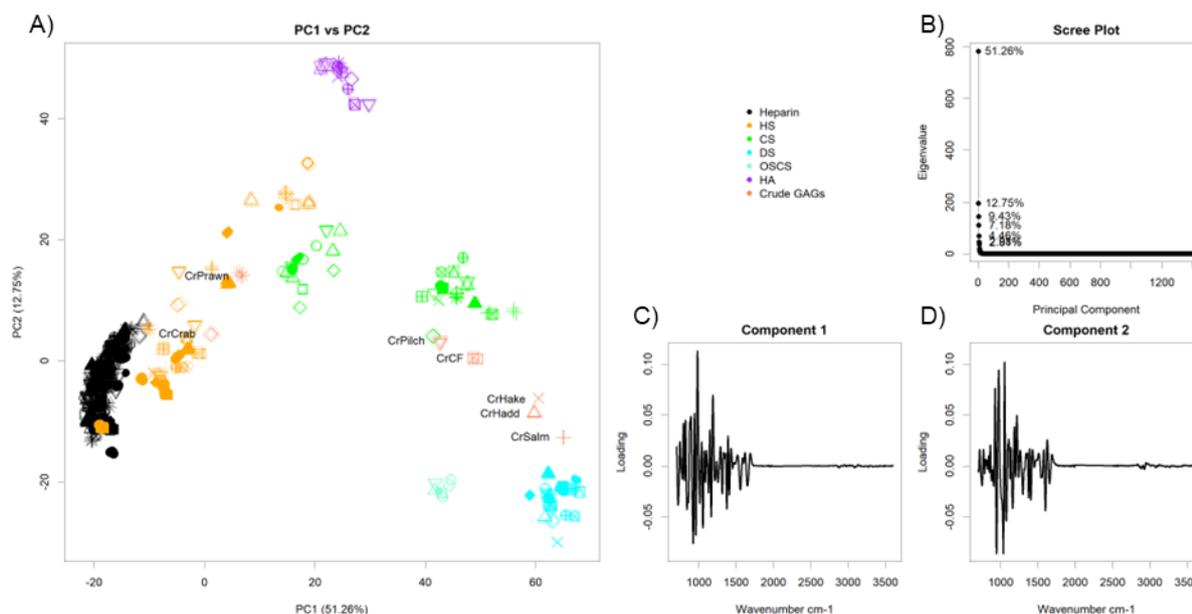
Figure 4.9: PCA of a GAG library vs crude samples extracted from marine sources. A) Scores plot of PC1 vs PC2. **B)** Scree plot. **C)** Loadings for PC1. **D)** Loadings for PC2. Heparin: black, HS: orange, CS: green, DS: teal, OSCS: aquamarine, HA: purple, crude samples: pink. Abbreviations: Cr: crude, Pilch: pilchard, CF: catfish, Hadd: haddock, Salm: salmon. All spectra involved in PCA are in the regions of 700 and 2000 and, 2500 and 3600 cm^{-1} and are averages of 3 scans. The spectra were smoothed, baseline corrected with a 7th order polynomial, normalised (0-1), smoothed and derivatised with a Savitzky-Golay algorithm and finally autoscaled.

After demonstrating that ATR-FTIR-PCA can identify GAGs, ICA was performed on each of the crude marine GAGs and a total of 5 (selected by the analyst) signals were generated for each sample using the *fastICA* function from the **fastICA** package in R. The algorithm used was set to "deflation" to ensure that more than two ICs (as there are mixtures of 3 GAGs) could be generated. The resulting signals underwent PC score prediction using the PCA in **Figure 4.9** and were plotted, along with their parent compound in PCs 1 and 2. It is important to note, that not all 5 component signals are necessarily meaningful and may simply resemble residual noise.

The ICA failed to separate the spectra into any constituents (**Fig. 4.10** and **4.11**). The only constituent extracted consisted of the input spectrum and the remaining components was random noise. This suggests that the constituents within the spectrum are too similar to each other to discern them (**Fig. 4.11**).

[123]
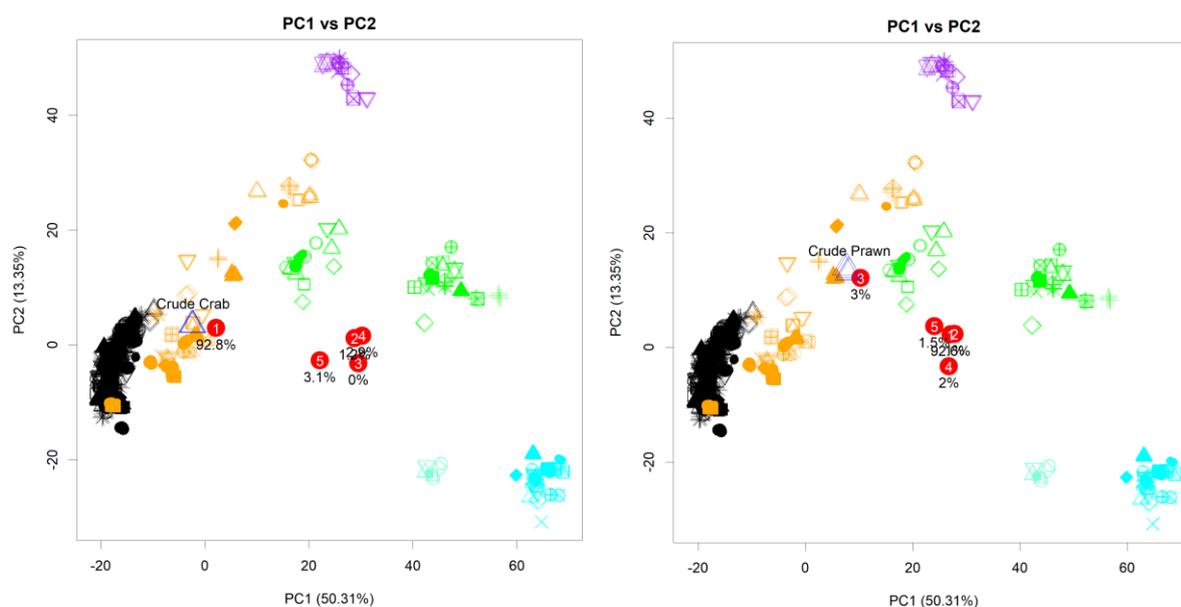
Figure 4.10: ICA output signals plotted over PCA of a GAG library. The PC scores were predicted for the signals extracted from ICA analysis using the predict function in base R and plotted over the original PCA (Fig.4.7) score plots. **Left)** Score plot for PC1 vs PC2, with signals extracted from the crude crab sample. **Right)** Score plot for PC1 vs PC2, with signals extracted from the crude prawn sample. Heparin: black, HS: orange, CS: green, DS: teal, OSCS: aquamarine, HA: purple, original spectrum of crude sample: blue, extracted signals from ICA: red. All spectra involved in PCA and ICA were prepared in the same way; they are the average of 3 repeats and the spectral regions from 700 to 2000 and from 2500 to 3600 cm^{-1} were used. The spectra were smoothed, baseline corrected with a 7th order polynomial, normalised (0-1), smoothed and derivatised with a Savitzky-Golay algorithm and finally autoscaled.
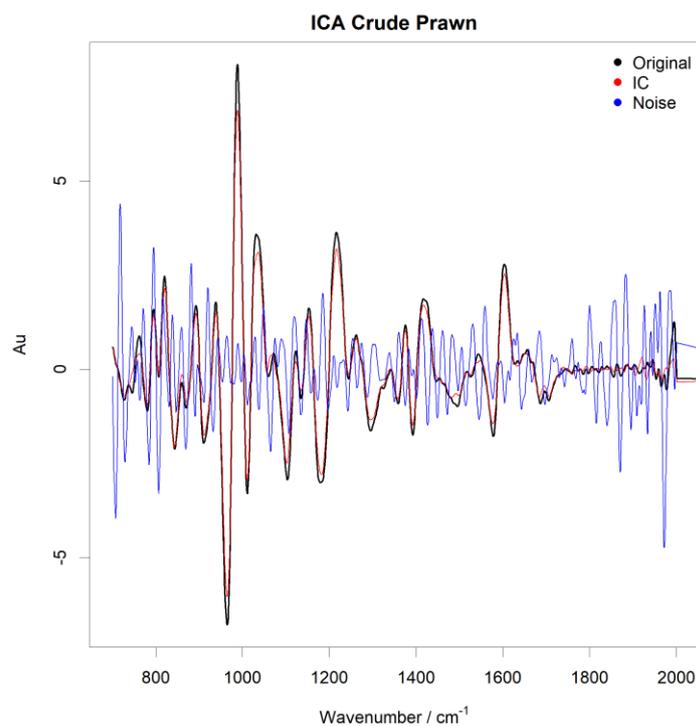
Figure 4.11: Extracted signals from ICA compared with the original spectrum for the crude prawn sample. Original spectrum: black, extracted signal: red, extracted signal that equates to noise: blue.

4.3.4 The use of spectral filtering to investigate signal masking.

The inability to resolve the spectra into its constituents may help to explain resolution issues observed with marine GAG compositional analysis. In order to study the signals relevant to

each GAG in a heparin-GAG mix, a new technique was used. Correlation spectroscopy filtering with random iterative sampling (COS-Firs) is a technique that has been utilised by Rudd et al. (2011) for the extraction and identification of NMR signals which correlate to alien features in heparins. The technique effectively computes a more resolute difference spectrum, the contents of which are solely correlated to signals alien to a test matrix. In brief, a COS-Firs difference spectrum is calculated through the subtraction of two correlation matrices: one matrix containing the natural variation expected within a population and the other matrix containing natural samples and a test sample. The power spectrum of the resultant correlation matrix can be used to identify signals unique to the test sample that do not fall within the natural variation expected. Each matrix is constructed from a library of representative spectra and the test sample randomly replaces one of these spectra, a series of iterations of the calculated power spectra are averaged to create the output. A small library of heparin (n = 12) was also used to create an average heparin signal, to which the extracted signals could be compared.

At high GAG levels, each distinct GAG has a unique set of signals that correlate to it (**Fig. 4.12A, B**), but as the level of GAG reduces in the sample, many of these unique signals diminish until only some at wavenumbers 980 and 1000 cm^{-1} remain (**Fig. 4.12**). At the level of 5% (w/w), only CSC has unique signals at wavenumbers 1000 and 1010 cm^{-1}, while the other GAGs overlap significantly (**Fig. 4.12E**). At the level of 1% (w/w), all the GAGs overlap and are indistinguishable (**Fig. 4.12F**). CSC was predicted the best in all above models and consistently had the lowest RMSE of 4.11, this is corroborated by COS-Firs, where at 5% (w/w) CSC, CSC has unique signals but at 1% (w/w) it does not. At 10% (w/w), which is higher than the RMSE of the other GAGs (~7), heparin, CSA, DS and HA retain unique signals. It is likely that the reason the model cannot separate the other GAGs from each other at these levels is due to the overlapping signals at low levels. This corroborates the detection observed at low levels with PCA-LRM (**Table 4.1**) and the subsequent inability to separate and therefore quantify the GAGs to below a RMSE of ~5 (**Table 4.7**).

As the COS-Firs spectra correspond to signals that are solely related to contaminants in heparin, it was hypothesised that they may provide better quantification in a PCR model. The entire library, including the T-samples and the MGAGs underwent COS-Firs, and using a PCA of the output spectra of heparin:GAG blends and the MGAGs (**Fig. 4.13**), the compositions of the T-GAGs were predicted (**Table 4.11**). The predictions were worse in all cases and all RMSEs increased from 5.60, 6.98, 4.11, 6.96 and 4.96 to 8.74, 13.57, 8.88, 10.73 and 6.51 for heparin, CSA, CSC, DS and HA respectively. No attempts to predict the levels of the crude marine GAGs were made due to the worse predictions with the COS-Firs PCR model.
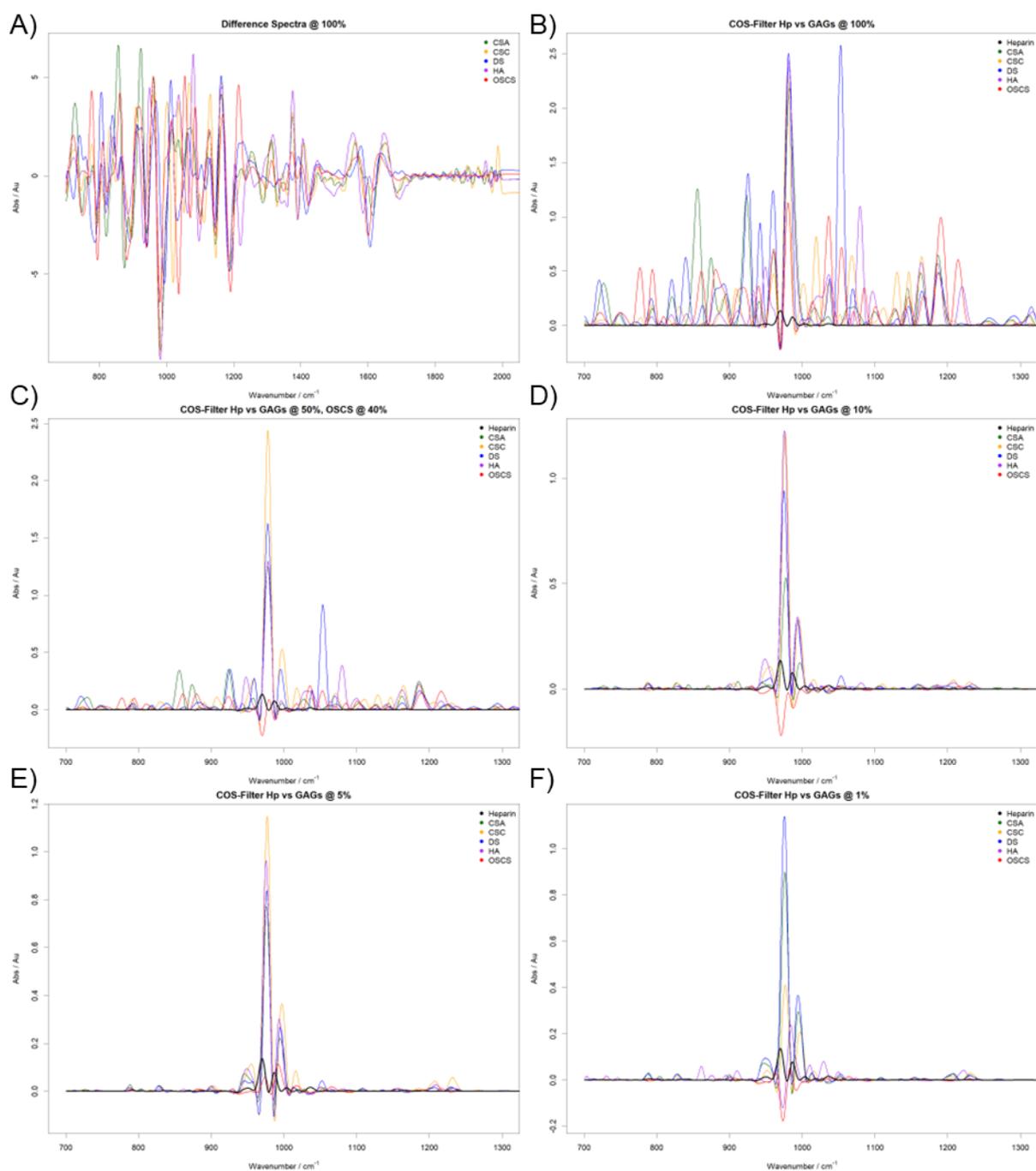
Figure 4.12: Difference spectra and COS-firs difference spectra of heparin blended with CSA, CSC, DS, HA and OSCs.
A) difference spectra of heparin and CSA, CSC, DS, HA and OSCS. B) COS-firs difference spectra of heparin vs CSA, CSC, DS, HA and OSCS. C) COS-firs difference spectra of heparin vs heparins blended with CSA, CSC, DS, HA and OSCS at the level of 40% (w/w). D) COS-firs difference spectra of heparin vs heparins blended with CSA, CSC, DS, HA and OSCS at the level of 10% (w/w). E) COS-firs difference spectra of heparin vs heparins blended with CSA, CSC, DS, HA and OSCS at the level of 5% (w/w). F) COS-firs difference spectra of heparin vs heparins blended with CSA, CSC, DS, HA and OSCS at the level of 1% (w/w). COS-Firs difference spectra were calculated using the second derivative of spectra that had been smoothed, base-line corrected to a 7th order polynomial, normalised (0-1) and then smoothed again. Difference spectra were calculated by subtracting the heparin spectrum from the relevant GAG spectrum. COS-Firs difference spectra were calculated by subtracting a covariance matrix of a heparin library from a covariance matrix of a heparin library that contains a test sample. The multiple matrices are constructed, using different combinations of heparins each time – the test sample never changes. After 50 iterations, the average of the resultant covariance matrices is calculated and the power spectrum recorded and plotted.
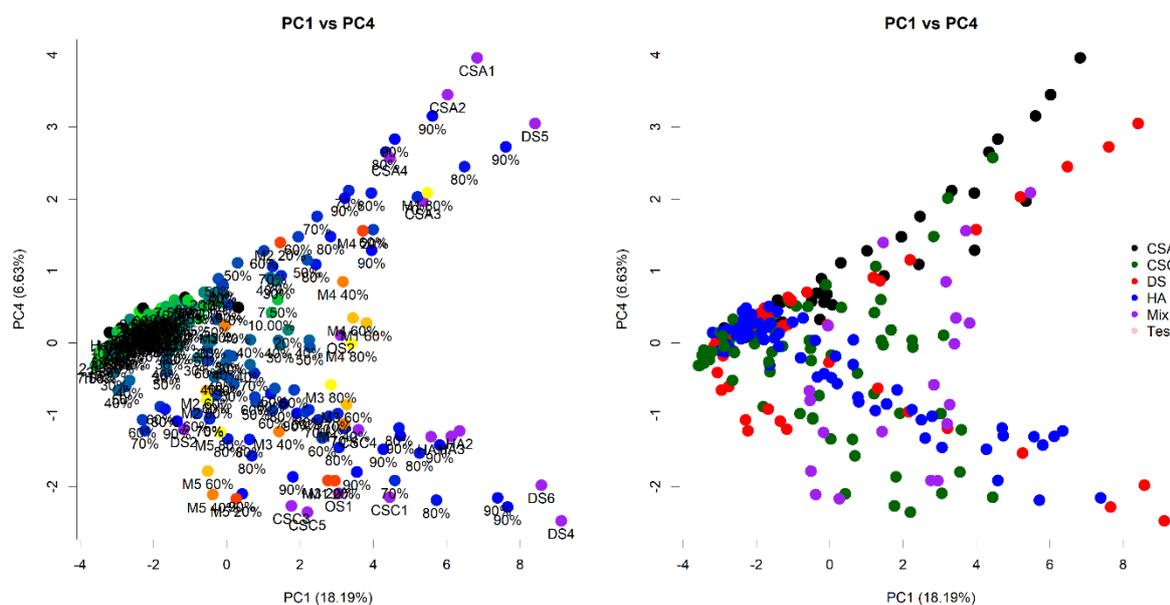
Figure 4.13: PCA of difference spectra that have undergone COS-Firs against a heparin library. Left) Scores plot of PC1 vs PC4. Heparin: black, CSA, CSC, DS, HA: purple, heparin samples blended with different GAGs are indicated with a scale of green to blue – the more blue the sample the higher the level of GAG present. For the new GAG:GAG mixtures, a scale from yellow to orange is used to indicate the levels of each GAG. **Right)** Recoloured scores plot of PC1 vs PC4, whereby samples are coloured based on which GAGs are blended with which, not the relative levels. Heparin:CSA blends: dark green, heparin:CSC blends: green, heparin:DS blends: red, heparin:HA blends: blue, GAG:GAG mixtures (MGAGs): purple, T-GAGs: pink. COS-Firs difference spectra were calculated using the second derivative of spectra that had been smoothed, base-line corrected to a 7th order polynomial, normalised (0-1) and then smoothed again in the regions of 700 to 2000 and 2500 to 3700 cm^{-1}. COS-Firs difference spectra were calculated by subtracting a covariance matrix of a heparin library from a covariance matrix of a heparin library that contains a test sample. The multiple matrices are constructed, using different combinations of heparins each time – the test sample never changes. After 50 iterations, the average of the resultant covariance matrices is calculated and the power spectrum recorded and plotted. The COS-Firs output spectra were used for PCA with no further modifications.

Table. 4.11: PCR predictions of GAG composition for 15 simulated GAGs using the PCA of COS-FIRs difference spectra. PCA scores from the PCA shown in **Figure 4.13** were used to create the PCR model. Samples to undergo prediction had their PC scores predicted for the relevant PCA using the *predict* function in ***base*** R. The levels of GAG were predicted using the *predict* function from the ***pls*** package on the relevant PCR and predicted PC scores. The top 5 components were used for prediction.

Sample	Percentage of Component / %											
	Heparin		**CSA**		**CSC**		**DS**		**HA**		**OSCS**	
	Original	Predicted	Original	Predicted	Original	Predicted	Original	Predicted	Original	Predicted	Original	Predicted
T1	75.00	78.52 +/- 0	0.00	7.88 +/- 0	25.00	4.75 +/- 0	0.00	1.73 +/- 0	0.00	4.37 +/- 0	NA	2.74 +/- 0
T2	75.00	80.41 +/- 0	0.00	9.55 +/- 0	0.00	3.32 +/- 0	25.00	-1.78 +/- 0	0.00	5.21 +/- 0	NA	3.29 +/- 0
T3	75.00	81.66 +/- 0	25.00	8.68 +/- 0	0.00	2.23 +/- 0	0.00	1.01 +/- 0	0.00	4.11 +/- 0	NA	2.3 +/- 0
T4	75.00	81.72 +/- 0	12.50	4.71 +/- 0	12.50	8.46 +/- 0	0.00	2.92 +/- 0	0.00	1.89 +/- 0	NA	0.3 +/- 0
T5	75.00	82.31 +/- 0	18.75	9.26 +/- 0	6.25	0.29 +/- 0	0.00	0.74 +/- 0	0.00	3.65 +/- 0	NA	3.75 +/- 0
T6	75.00	81.25 +/- 0	6.25	6.83 +/- 0	18.75	3.78 +/- 0	0.00	3.01 +/- 0	0.00	2.39 +/- 0	NA	2.74 +/- 0
T7	75.00	78.81 +/- 0	6.25	8.48 +/- 0	0.00	1.41 +/- 0	18.75	3.31 +/- 0	0.00	3.73 +/- 0	NA	4.26 +/- 0
T8	75.00	81.72 +/- 0	0.00	7.83 +/- 0	6.25	-0.82 +/- 0	18.75	4.08 +/- 0	0.00	2.74 +/- 0	NA	4.45 +/- 0
T9	75.00	81.72 +/- 0	18.75	9.05 +/- 0	0.00	0.37 +/- 0	6.25	1.22 +/- 0	0.00	4.04 +/- 0	NA	3.59 +/- 0
T11	75.00	78 +/- 0	0.00	9 +/- 0	0.00	3.17 +/- 0	0.00	0.44 +/- 0	25.00	6.47 +/- 0	NA	2.91 +/- 0
T12	55.00	71.3 +/- 0	15.00	9.63 +/- 0	20.00	8.89 +/- 0	10.00	4.47 +/- 0	0.00	1.87 +/- 0	NA	3.84 +/- 0
T13	0.00	13.13 +/- 0	0.00	2.93 +/- 0	50.00	44.96 +/- 0	50.00	37.8 +/- 0	0.00	-5.72 +/- 0	NA	6.9 +/- 0
T14	0.00	17.77 +/- 0	50.00	5.69 +/- 0	0.00	13.19 +/- 0	50.00	70.16 +/- 0	0.00	-9.79 +/- 0	NA	2.98 +/- 0
T15	0.00	-4.07 +/- 0	50.00	38.65 +/- 0	50.00	48.65 +/- 0	0.00	5.22 +/- 0	0.00	7.65 +/- 0	NA	3.9 +/- 0
T16	65.00	75.7 +/- 0	10.00	7.93 +/- 0	20.00	6.9 +/- 0	5.00	2.06 +/- 0	0.00	4.61 +/- 0	NA	2.81 +/- 0
RMSE	8.74		13.57		8.88		10.73		6.51		NA	
R2	0.98		0.57		0.82		0.72		0.00		NA	
P-value	0.51		0.48		0.51		0.64		0.69		NA	

[127]

4.4 Discussion

The aims of this chapter were three-fold: 1) to determine the ability of PCA-LRM to quality control heparin against many distinct contaminants, 2) determine the LODs of these contamiannts and 3) to determine if ATR-FTIR could identify the contaminant(s), if any, present in a sample. To satisfy objective 1, a series of relevant saccharides were gathered (the other GAGs) and some were synthesised (the OS-GAGs). Heparin was then contaminated at various levels with these saccharides, and the PCA-LRMs ability to distinguish these samples investigated. All sulphated saccharides were detected at meaningful levels (≤5%) (**Table 4.1**), suggesting that the model detects levels of heparin, rather than levels of OSCS. The GAG HA was not detected at all however, despite its OS-variant being detected at 5%, which suggests that the model is detecting sulphation that is not native to heparin.

To test this, a series of natural saccharides, many of which are unsulphated, and their chemically over-sulphated counterparts were taken and input into the library. No contamination series of these were made, as many of the parent compounds are insoluble and strict LODs for them are likely not meaningful for heparin QC. The generic, generally unsulphated saccharides had low PC4 scores (mean = -4.4, median = -4.8) (**Table. 4.1**), which fell below the heparin libraries PC4 scores, while the over-sulphated saccharides had high PC4 scores (mean = 10.0, median = 14.1) that were readily classified as "contaminated" heparins (threshold PC4 score for a heparin = 0.87). Attempts were then made to improve the LRM below the PC4 score of accepted heparins to account for potential under-sulphated contaminants, but this failed (**Fig. 4.1**). This suggested that PC4 was primarily separating for sulphation. Curiously, generic naturally occurring, unsulphated polysaccharides were classified as "contaminated" in some cases. The "contaminated" samples contained methyl-, ethyl- and acetyl- esters and also branching which suggests that, more broadly, the PCA-LRM model separates by alien moieties about the carbon ring. The broad range of potential contaminants detected by the PCA-LRM makes it more robust in general use than previously expected, but its inability to detect under-sulphated contaminants represented a weakness.

A new PCA, of both OSCS contaminated and HA contaminated heparin was performed and separation for both contaminants was shown to be orthogonal across 2 different components. This suggested that another LRM would be needed to detect under-sulphated contamination. An LRM could be constructed with scores from the new PCA to detect both HA and OSCS, or another could be produced based on just HA separation from heparin. The model based on just HA was calculated and was significantly better than the PC4-LRM model used previously (AIC score of 115.51 vs 511.64 respectively), allowing discrimination of all contaminated samples at a low 3% false positive rate (**Fig. 4.3**). Using both these models, a pharmaceutical

heparin can be defined clearly and easily for both over- sulphated constituents and HA, which may account for under-sulphated molecules.

The generalised nature of contaminant detection, while powerful in many aspects is an issue in others. Some pharmaceutical heparins can contain up to 5% (w/w) DS which, as less than 5% (w/w) DS has been detected in some cases, may falsely detect "contaminated" heparins. The PCA-LRM model has no ability to identify contaminants, but PCA has demonstrated an ability to not only detect levels of contaminants, but also to separate the type of contamination orthogonally (**Fig.4.2A**). To probe this, the contamination series constructed to test the robustness of the PCA-LRM became the basis of a new library, on which PCA was performed. The 8 contaminants: CSA, CSC, OSCS, DS, OSDS, HA, OSHA and OSHp were all separated easily across 6 components (**Figs. 4.4.1 and 4.4.2**). The discovery that 6 components were required to separate the 8 contaminants revealed a potential weakness however, as not all samples are orthogonal to each other and may overlap (**Table 4.4**).

To test the ability of FTIR-PCA to identify different contaminants, a series of test samples were produced and the composition of each predicted with a PCR model, generated from the PCA in (**Fig. 4.7**). The levels of heparin, CSA, CSC, DS and HA were predicted without being significantly different to the expected values (P-values > 0.01) and had RMSE of 7.26, 8.70, 4.74, 7.08 and 4.91 respectively (**Table 4.5**). Through the addition of samples that encompass GAG:GAG mixtures, not just heparin:GAG mixtures, the predictions were improved, all P-values remained > 0.01 and the RMSE were decreased to 5.60, 6.98, 4.11, 6.96 and 4.96 for heparin, CSA, CSC, DS and HA respectively (**Table 4.7**). DS was difficult to separate from other GAGs however, and this was likely due to separation being obtained across multiple PCs (**Table 4.4**) and subsequent overlapping signals with other GAGs, which also had portions of the variance associated with them in these components. The level of heparin was generally correct in all cases, despite some overlap with CSA, meaning that a simple glucosaminoglycan vs galactoseaminoglycan ratio can be achieved rapidly with simple analysis.

Application of PCR to marine GAGs provided mixed results. Some samples had the level of heparin predicted with great accuracy (RMSE = 2.49 for crude prawn, crab, pilachard and catfish GAGs) whilst others were incorrectly predicted (RMSE = 21.29 for crude haddock, hake and salmon GAGs) (**Table 4.8**). In an attempt to rectify this, the samples originally used to test the PCR models (the T-samples), some of which contain triple GAG mixtures and hence more closely resemble the crude marine GAG mixtures, were input into the library and used for PCR prediction. The levels of heparin prediction improved mildly for the crude prawn, crab, pilchard and catfish GAGs (RMSE = 2.38) and decreased somewhat for the remaining crude GAGs (RMSE = 21.32). The levels of DS vs CS were incorrectly predicted in both cases (RMSE =

16.03 and 24.23 for CS and DS respectively when PCR model used for **Table 4.7** is used and RMSE = 14.11 and 20.52 for CS and DS respectively when the PCR model used for **Table 4.9** is used) however large improvements were seen with the addition of the new samples – suggesting that the library used is not sufficient to discriminate between the GAG structures observed in the crude marine GAGs.

To ascertain whether the library did not contain enough relevant samples, 5 crude heparins, which are comprised of mammalian GAGs, such as those used to construct the library, underwent compositional analysis. The model which provided the best predictions for the crude marine GAGs (that which was used for **Table 4.9** and contains the T-samples) was used. The prediction of heparin, CS and DS was improved (**Table 4.10**) when compared with the T-samples (**Table 4.9**). Interestingly, the RMSE of heparin remained steady (RMSE = 5.28 compared with RMSE = 5.60) but the RMSE for CS and DS was drastically improved (From 14.11 and 20.52 to 4.42 and 2.88 respectively). This suggests that the underlying DS and CS signals in the marine GAGs are too different to the modelled mammalian GAGs. A decrease in RMSE for heparin, CS and DS has been observed with every iteration of the GAG blend library that has been used to predict the crude marine GAG compositions. As each iteration of the library included samples that more closely resembled the tested GAGs compositionally, it is possible that the library is incomplete for the compositional analysis of marine GAGs.

Nevertheless, as overlaps between signals that correspond to DS and signals that correspond to heparin, CSA, CSC and HA and between signals that correspond with heparin with signals that correspond with CSA have been observed (**Table 4.4**), it is also possible that the signals associated with these GAGs are simply too similar to discriminate between at low levels. To explore whether the signals corresponding to DS, CSA, CSC and heparin could be distinguished, attempts were made to separate and identify which signalds correspond to which GAG. ICA was performed, which attempts to extract latent signals from an input spectrum and quantify how much of each are present. ICA failed to extract any new signals from the crude marine GAGs instead produced the original input signal with a slightly reduced intensity and some signals that appear to be random noise (**Fig. 4.11**). This suggested that the signals associated with the different GAGs may not be separable with FTIR. To test this further, COS-Firs, a technique which has been used to show signals which correspond to very low levels of heparins from different animal sources with NMR, was utilised. COS-Firs calculates what effectively amounts to a high-resolution difference spectrum, containing signals that are not correlated with natural fluctuations in the filtered sample library. Here, samples contaminated with different GAGs were filtered against a heparin and the computed signals compared (**Fig. 4.12**). At the level of 10% (w/w), all GAGs had distinct signals which may correspond to each unique GAG (**Fig. 4.12D**), corresponding with observed RMSE values of ~7 for heparin, CSA and DS (**Table 4.7**). CSC and OSCS showed unique signals at the

level of 5% (w/w) (**Fig. 4.12E**), while the rest of the GAGs overlapped each other significantly and OSCS showed unique signals down to the level of 1% (w/w) (**Fig. 4.12F**). The unique signals of OSCS at 1% (w/w) corroborate with observed LODs in the PCA-LRM (**Fig. 3.29**).

FTIR has, therefore, been shown to be able to quality control heparin against a broad range of contaminants, some relevant and some potential. It has also, coupled with PCA, been able to predict the levels of heparin, CSA, CSC, DS and HA in mixtures of the 5. There are two potential issues with the current approach: 1) a lack of resolution with which to distinguish CSA, CSC and DS and 2) a lack of relevant model compounds with which to model all the structural variation observed. Both of these problems are likely contributing to the present analysis; as the inability to observe differences between heparin, CSA and DS at the level of 5% (w/w) is shown with COS-Firs spectra and as the addition of new blends which better describe the test samples reduces the RMSE in all cases. Another potential issue is that of the characterisation method for the levels of different GAGs used for the marine GAGs and the crude heparins. The marine GAGs underwent compositional analysis with NMR spectroscopy whilst the crude heparins underwent compositional analysis with SAX chromatography. The level of DS was ascribed based on the relative level of IdoA in the sample for the marine GAGs, whilst the level of DS was ascribed to whole chains (which may not contain explicitly IdoA) that had increased overall charge for the crude GAGs. Since the library is constructed from DS samples which may not contain explicitly IdoA, this may explain the failed predictions with the marine GAGs. Expansion of the library either through incorporation of previously extracted and characterised marine (or other) samples, a wider range of parent compound entries, or chemical or chemoenzymatic modification to introduce new structures may provide better separation and facilitate separation of multiple GAG species in future work.

Chapter 5: Identification and characterisation of crude heparins and discrimination between heparins from different animal sources.

5.1 Introduction

FTIR-ATR-PCA has been shown to be able to quantify the levels of heparin and to some extent remaining GAGs in simulated blends, crude heparins and crude GAG extracts (**Tables 4.7, 4.9 and 4.10**). Crude heparins represent an important part of pharmaceutical heparin QC. It is thought that the introduction of OSCS during the contamination of pharmaceutical heparin in 2007-8 happened at the crude heparin stage due to the identification of OSCS in crude heparins (FDA, 2013). The traceability of the suppliers of the initial intestines from which pharmaceutical heparin will be extracted is poor and weakly regulated. Further to this, many sites from many locations are involved and many crude heparin lots are blended into one before being sent for final product preparation, hence the risk of contamination from an unknown source is greatly increased and difficult to monitor(Al-Hakim, 2021; Liu et al., 2009; Mauri et al., 2017b). In 2013, the FDA published guidance on what should constitute a crude heparin in an effort to reduce the risk of contamination with OSCS and recommended testing for OSCS in crude heparins with SAX (FDA, 2013).

In 2017, Mauri *et al.* utilised NMR and chemometrics to monitor the structural features of 88 crude heparins that have been acquired over the course of 5 years, from 13 different manufacturers. This paper highlighted two things: 1) that crude heparins could be contained within their own region of n-dimensional PC space and 2) that chemical and structural composition of crude heparins could be undertaken with NMR. The crude heparins form lobes which describe key compositional differences such as higher levels of DS and CS than other crude heparins and key structural differences such as high levels of NAc and higher overall sulphation. The paper shows that PCA of just the acetyl region in H[1] NMR creates a homogenous lobe into which all the previously accepted crude samples fall. This suggests that previously accepted crude samples can be used to create a library, into which new crude samples can fall, facilitating crude heparin QC.

Heparin QC will soon face another hurdle, following the potential introduction of heparin extracted from different animal sources. Heparin is extracted primarily from pig mucosa, leading to porcine mucosal heparin (PMH) which is currently the only FDA- and EU approved pharmaceutical heparin, and is used solely in America, England and Europe (Van der Meer et al., 2017) despite the historical use of heparin sourced from dogs and cows and (liver and lungs [bovine lung heparin; BLH]) (Linhardt, 2003; Van der Meer et al., 2017). Various other

heparins have been extracted and probed for pharmaceutical applications, including those from poultry, seafood and other domesticated animals (Van der Meer et al., 2017) but despite the viability of these heparins it is only heparin from sheep mucosa (ovine mucosal heparin (OMH)) that, in modern times, has been considered alongside porcine and bovine heparins for pharmaceutical use amongst some pharmacopoeias (Colombo et al., 2022; Fareed et al., 2019; Ouyang et al., 2019; Vilanova et al., 2019b).

The sole use of PMH in Europe and north America is due to a host of ethical and economical reasons, which are unique to the country in which it will be used. In England, PMH is primarily used over BLH due in part to fears associated with bovine spongiform encephalopathy (BSE), known colloquially as mad cow disease; a prion-borne illness found in cows (Bett et al., 2020). Upon consumption of infected tissues, BSE can spread to humans, causing variant Creutzfeldt Jakob disease (vCJD), a lethal neurodegenerative disease which led to the deaths of some 174 British people between the years of 1980 and 1996, and resulted in strict controls on the feeding and age of cows and changed the cuts of beef that could be sold (Casalone and Hope, 2018, p. 7; Nathanson et al., 1997). Ovine heparins are often not sourced due to a similar prion disease, Scrapie, that affects sheep but has not been known to transfer to humans (Hunter, 2007). Heparin and HS have been shown to potentiate prion protein aggregate formation in murine models (Imamura et al., 2016) and hence may be a potential risk for prion borne illnesses. Other countries, such as India, and Islamic countries, such as Malaysia, would prefer to use bovine heparins, due, either to belief of the sacredness of the cow, dirtiness of the pig or a combination of both (Vilanova et al., 2019b).

80% of all PMH is manufactured in, and exported from, China due in part to the huge output of pig products from China and also comes from a single species of pig, representing a huge potential threat to the heparin supply chain (Vilanova et al., 2019a)). Illness may threaten the pigs and subsequently the heparin supply; in 2007 it is believed that blue-ear pig disease, endemic in Asia resulted in decreased supply of heparin and ultimately resulted in the contamination of heparin with OSCS – a similar situation is currently unfolding with African swine fever afflicting Chinese pigs (Fareed et al., 2019). Due to this, some countries such as Brazil have opted to re-introduce bovine heparins into the pharmaceutical supply (Vilanova et al., 2019b) and the US is currently examining whether bovine or ovine heparins could be introduced before a possible shortage becomes an issue (Fareed et al., 2019).

The gross structure of heparin changes depending on the animal from which it is extracted, and this is reflected in the sulphate substitution pattern, MW and therefore the anticoagulant activity (Devlin et al., 2019b; Mauri et al., 2019). PMH consists primarily of 60-85% trisulphated disaccharides and ~ 20% disulphated saccharides, with an average DoS of 2.45 and MW of 17 kDa (Devlin et al., 2019b; Fu et al., 2013a; St Ange et al., 2016a). BLH and OMH are

structurally very similar to PMH, possessing 70 - 85% and 60 - 90% trisulphated disaccharides and both possessing 10 - 20% disulphated disaccharides and an average MW of ~ 15 kDa respectively (Chen et al., 2019b; Colombo et al., 2022; Devlin et al., 2019b). Importantly, and unlike heparins sourced from other organisms (Van der Meer et al., 2017), the similar structures found in PMH, BLH and OMH confer a similar anticoagulant activity (Devlin et al., 2019b; Fareed et al., 2015).

Heparin from cow mucosa (bovine mucosal heparin (BMH)) was used in Brazil up to 2009, and after a brief removal from the Brazilian pharmacopoeia, was reintroduced in 2016 (Vilanova et al., 2019b). BMH possesses, for pharmaceutical heparin, an unusually low level of 6-sulphation, particularly of disaccharides containing NS2S residues (St Ange et al., 2016a; Vilanova et al., 2019b). BMH contains 50 - 55% trisulphated, 16-32% disulphated residues and has an average MW of ~15 kDa and also possesses the lowest anticoagulant activity (possibly due to the different levels of sulphation) of the four heparins considered here (Devlin et al., 2019b; St Ange et al., 2016a). Despite their chemical and pharmaceutical similarity, the different heparins are still considered as separate pharmaceutical entities, with bovine derived and porcine derived heparins possessing unique monographs in Brazil (BHRA, 2019; Vilanova et al., 2019b).

Table 5.1 Structural and biological activity differences of heparins extracted from porcine bovine and ovine sources. Data based on published results (Bianchini et al., 1997; Casu, 1985; Chen et al., 2019a; Fasciano and Danielson, 2016; Fu et al., 2013b; Guan et al., 2016; Jeske et al., 2019; Mulloy et al., 2000; St Ange et al., 2016b; Tovar et al., 2012; Watt et al., 1997; Zhang et al., 2011), table reproduced from (Devlin et al., 2019b). Data are expressed as "minimum–maximum (mean)".

Heparin	Level of Ido2S-GlcNS,6S / %	Specific Activity / IU.mg^{-1}			Mw/Da
		Anti-Xa	Anti-IIa	APTT	
PMH	51.5–85 (68.3)	145–220 (194)	172–230 (197)	145-277 (196)	12,000–27,090 (19,002)
BLH	70–87 (79.8)	105–156 (133)	130.6–180 (153)	89–167 (139)	12,000–15,240 (14,230)
BMH	47.4–64.2 (54.5)	113.6–159 (134)	92.2–160.7 (126)	88.1–181 (136)	14,900–16,417 (15,439)
OMH	60–89.4 (75.2)	196–205 (201)	191–201 (195)	165–165 (165)	12,200–20,023 (14,773)

The Brazilian monograph for bovine heparin is similar to that of porcine heparin, requiring the same techniques explained in **Table 1.2**. H^1 NMR is required and detection of the same 5 characteristic signals at 5.40, 5.31, 3.28 and 2.05 ppm +/- 0.03 is used, however, the ratio between the intensities of the signals at 5.40 and 5.31 have different requirements and are calculated using the formula below (**Eq 5.1**). PMH must be ≤20 while BMH must be between 42 and 58.

$$\frac{\delta_{5.31} \times 100}{\delta_{5.41}} = ratio$$

<div align="right">**Eq. 5.1**</div>

Ideally, QC of pharmaceutical heparin would include the ability to detect the source of the heparin, and blends thereof, however high structural similarities make this a difficult task to undertake. Currently, the only FDA approved method for identifying the source animal of heparin is the polymerase chain reaction on crude heparin products (FDA, 2013). Furthermore, this method is not essential as it is only recommended as part of a guidance on crude heparins. Efforts to discriminate the animal source of crude heparins with NMR (Mauri et al., 2017b), the animal source of pharmaceutical heparins and blends thereof have been studied previously with NMR and MS (Colombo et al., 2022; Hook et al., 2021).

5.2 Chapter aims.

1. Establish whether ATR-FTIR can be used to identify crude heparins as a separate chemical entity, facilitating QC.
2. Establish whether ATR-FTIR can monitor the structural and compositional features of crude heparins, such as sulphate level and composition.
3. Establish whether ATR-FTIR can discriminate between heparins extracted from porcine, bovine and ovine sources – the key differences of which are sulphate level and composition.
4. Define LODs and LOQs for blends of porcine heparin with other heparins.

5.3 Results

5.3.1 The application of FTIR to crude heparin samples.

To begin, analysis of crude heparins in comparison with pharmaceutical heparins was undertaken in order to ascertain whether crude heparins are chemically distinct from pharmaceutical heparins with ATR-FTIR. 73 of the 83 crude heparins explored in (Mauri et al., 2017b) underwent ATR-FTIR spectral acquisition. The paper provides complete structural and compositional characterisation of the crude heparins and comprised of crude heparins with a mean DS and CS content of 8.3 ± 3.38% (w/w) and 2.3 ± 1.95% (w/w) of which, 15.5 ± 16.62% (w/w) was CSC respectively, but essentially pure heparin samples and samples as low as 80% (w/w) heparin were present. The samples comprised on average 57.9 ± 2.09% trisulphated residues with 85.0 ± 1.94% NS, 73.4 ± 3.36% 6S and 71.5 ± 2.49% 2S.

Through visual inspection of the acquired IR spectra, the crude heparins represent a standard heparin (**Fig. 5.1**), and no clear trends pertaining to the intensities of the bands are observed as to the underlying composition. The spectra of samples with high CS, high DS and both high CS and DS compared with the other samples in the library were compared against a heparin sample from a random pharmaceutical heparin from **Figure 3.4** (i.e., a sample with <LOD for CS and DS content). There are noticeable differences between the spectra (**Fig. 5.1A**), particularly in the intensities of the bands at wavenumbers ~1450 cm^{-1} and ~1600 cm^{-1} and an extra shoulder appears at ~1430 cm^{-1}. The main carbohydrate band at ~1000 cm^{-1} has a more intense signal at the secondary maxima at ~1050 cm^{-1}. The intensity at ~1150 cm^{-1} is lower in these samples. At the second maxima (~1050 cm^{-1}), the sample containing both high CS and DS displays, paradoxically an intensity between those of the samples with high CS and high DS levels, potentially an artefact of sample concentration or of the underlying structures in the heparin samples themselves. The samples were compared in a similar manner for sulphation type, with the overall degree of sulphation (DoS) (**Fig. 5.1B**), levels of NS vs NAc (**Fig. 5.1C**) and levels of I2S vs I2O H (**Fig. 5.1D**). These spectra display a high level of similarity and the bands do not alter in an obvious manner, with any major differences noted in the main carbohydrate band at ~1000cm^{-1} and the sulphate band at ~1250 cm^{-1}.
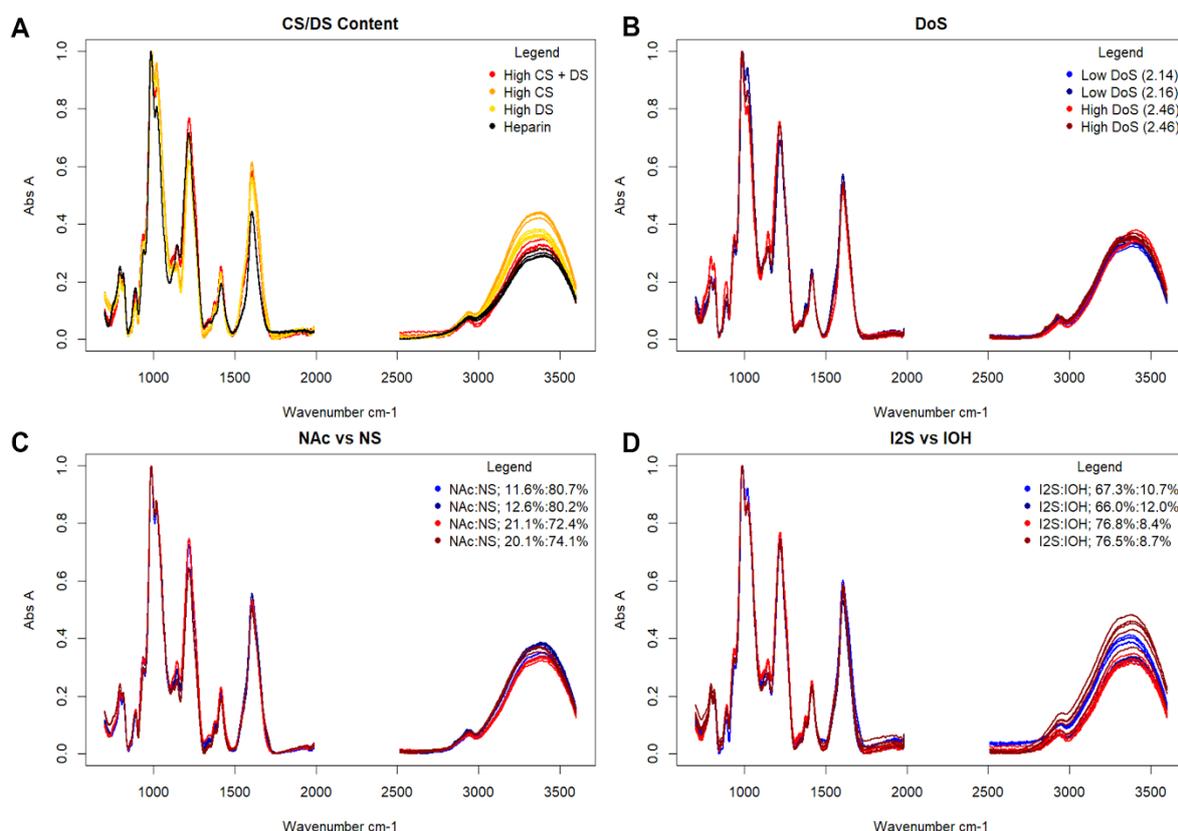
Figure 5.1: ATR-FTIR spectra of crude heparins vs pharmaceutical heparins. A) A pharmaceutical heparin (black) is compared with crude heparins that contain either relatively high CS (orange), high DS (yellow) or a combination of both high CS and DS (red). **B)** Two crude heparins that contain a relatively low DoS (blue) are compared with two crude heparins that contain a relatively high DoS (red). **C)** Two crude heparins that contain a relatively low level of NAc (blue) are compared with two crude heparins that contain a relatively high level of NAc (red). **D)** Two crude heparins that contain a relatively low level of I2S (blue) are compared against two crude heparins that contain a relatively high level of I2S (red). The "relative" level of each feature is given based on the values contained within the crude heparin library, for example a "relatively high" level indicates a high level within the 73 crude samples studied here. On these plots, all 5 repeats of each spectrum are shown. Spectra are between the regions of 700 and 2000 and, 2500 and 3700 cm^{-1}. All spectra were smoothed, baseline corrected with a 7th order polynomial and normalised (0-1) before visual comparison.

The crude samples were compared with PCA to both a GAG library and a library of pharmaceutical heparins (**Fig. 5.2).** The crude heparins locate between the heparin and CS region, verging on the edge of the heparins in PC1 and PC2 (**Fig. 5.2A**). When compared to heparin alone, the crude heparins appear distinct in PCs 1 and 2 (**Fig. 5.2B**). In both plots, crude heparins which are 100% heparin (coloured dark red) appear in the heparin library. This suggests that currently accepted crude heparins can be distinguished and defined using ATR-FTIR.
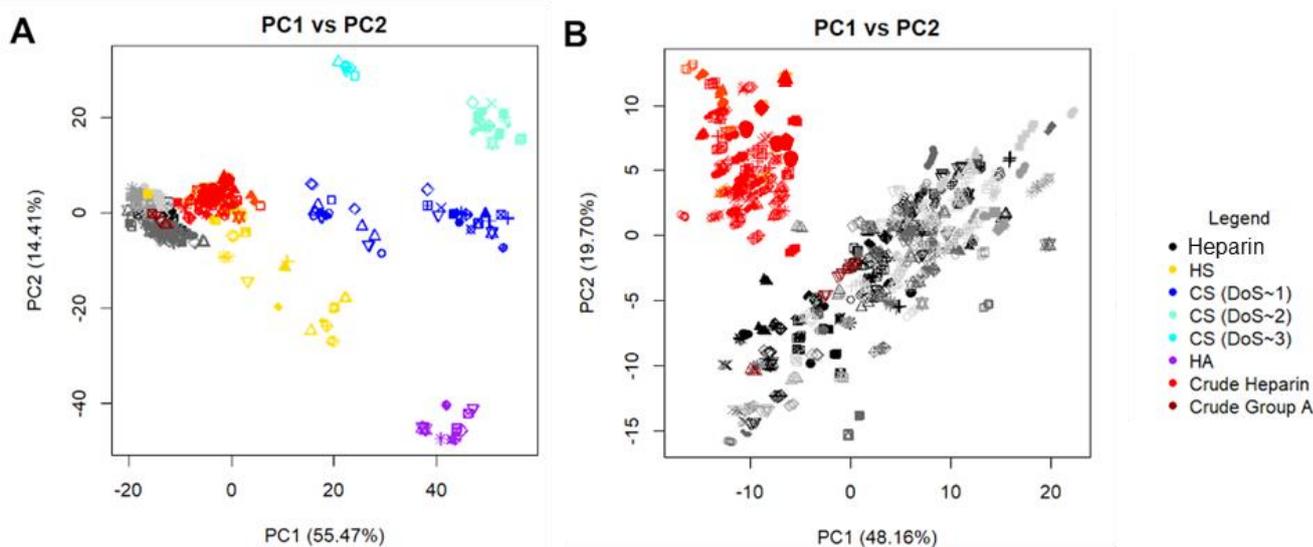
Figure 5.2: PCA of crude heparins compared with a GAG and a heparin library. A) Scores plots of PC1 and PC2 for crude heparins compared to a GAG library. **B)** Scores plots of PC1 and 2 for crude heparins compared to a heparin library. Crude heparins: red, heparin: shades of grey/black, HS: yellow, CS: blue, DS: aquamarine, HA: purple, OSCS: teal. All spectra involved in PCA are in the regions of 700 and 2000 and, 2500 and 3700 cm^{-1} and are averages of 3 scans. The spectra were smoothed, baseline corrected with a 7th order polynomial, normalised (0-1), smoothed and derivatised with a Savitzky-Golay algorithm and finally autoscaled.

5.3.2 The prediction of the composition of crude heparin samples.

The crude heparins alone underwent PCA and through visual inspection and PCR, components 1, 2, 3 and 5 were shown to correlate with the levels of CS and DS present in the samples (**Fig. 5.3 Table 5.2**). PC1 and 5 separated by the level of CS (**Fig 5.3C**) (24.4% and 25.5% of the variance associated with CS was associated with these components respectively) while PC2 and 3 which cover 30.6% and 33.9% of the variance associated with the level of DS, separated by the level of DS (**Fig 5.3A**) most likely through detection of Ido-2S associated with DS (13.4% and 28.2% of the variance associated with DS-2S respectively). A smooth transition for the level of IdoA-2S in DS was not observed in any of the components, instead, a node of no DS-2S and DS-2S was observed (**Fig. 5.3B**). The level of CS-6S had a similar relationship with components 1 and 5 (5.9 and 27.0% of the variance which accounts for CS-6S respectively); samples with no CS-6S formed a node to the bottom right, but samples containing different levels of CS-6S otherwise did not follow a clear gradient (**Fig. 5.3D**).

PCs 1, 5 and 6 correlated strongly with the levels of sulphation found in the samples (**Fig.5.4**). Specifically, PC1 correlates most strongly with the DoS of heparin, containing 27.52% of the variance associated with DoS. This also encapsulates levels of 6S, 2S and NS (25.9%, 20.9% and 11.36% of the variance associated with 6S, 2S and NS respectively). PC5 correlated with NAc levels and 2S levels (46.2% and 35.1% of the associated variance respectively), suggesting it may correlate with moieties bound to C2 of the carbon ring. PC5 also covers a large amount of the variation that encapsulates 6S (40.3%) and is also correlated with the level of CS (25.40%) and with CS-6S (26.9%). PC3 correlated with the level of DS (33.9%),

the presence of NH$_2$ (16.1%) and the percentage of saccharide chains that still possess a GAG linkage region (17.8%) while PC4 correlated weakly with the level of both I2S and NS (3.6%) and also with the presence of the linkage region (11.0%).
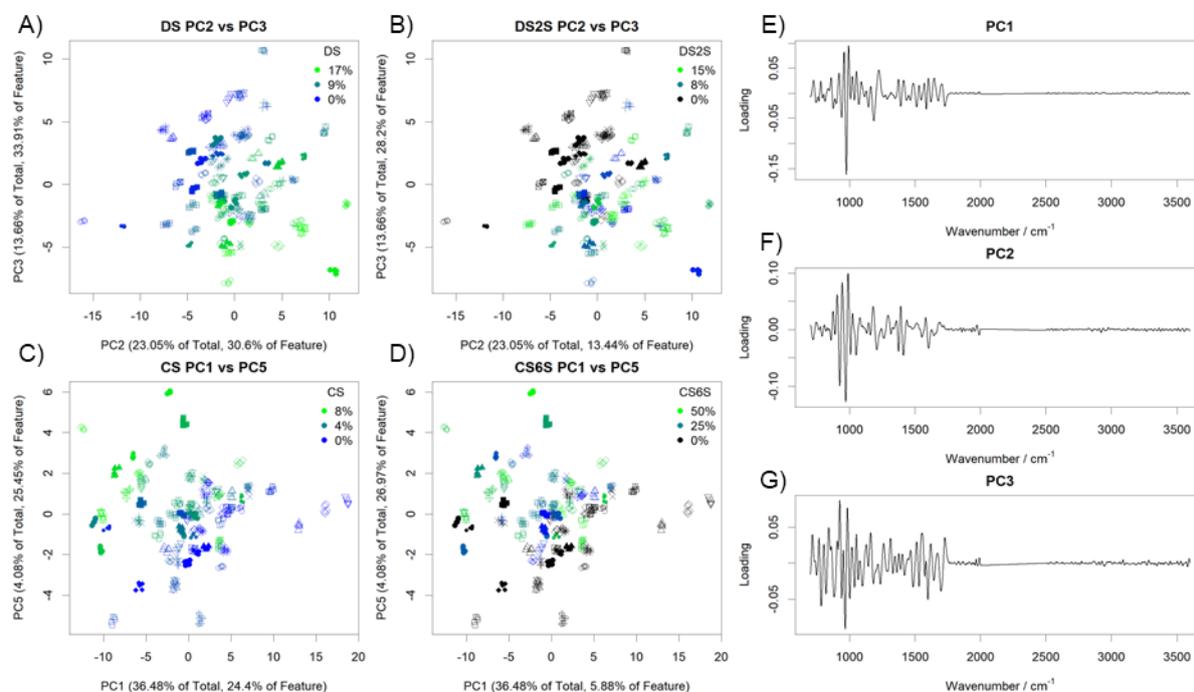
Figure 5.3: PCA analysis of crude heparins, coloured to indicate the levels of CS and DS. A) Scores plot of PC2 vs PC3, coloured to indicate level of DS. **B)** Scores plot of PC2 vs PC3, coloured to indicate the level of IdoA-2S present in the DS moieties. **C)** Scores plot of PC1 vs PC 5, coloured to indicate the level of CS. **D)** Scores plot of PC1 vs PC5, coloured to indicate the level of GalNAc-6S present in the CS moieties. **E)** Loading plot for PC1. **F)** Loading plot for PC2. **G)** Loading plot for PC3. All samples are coloured depending on the amount of each structural feature on a green to blue scale – the greener the sample the higher the level. PCs were selected based on the portion of variance which they cover, given the structural feature they are examining (Table 5.2). All spectra involved in PCA are in the regions of 700 and 2000 and, 2500 and 3700 cm^{-1} and are averages of 3 scans. The spectra were smoothed, baseline corrected with a 7th order polynomial, normalised (0-1), smoothed and derivatised with a Savitzky-Golay algorithm and finally autoscaled.

PC6 has small contributions to the levels of each sulphate type, correlating with the DoS (4.48%), 6S (3.23%), 2S (2.91%) and most strongly with the level of NS and Nac (8.7% and 6.7% respectively). Structural features, when compared with the levels of features which may also occupy the same carbon (for example NS vs Nac) separate orthogonally across components that describe them, as would be expected. PC7 correlated strongly with the level of GlcN,6,3-S with 12.9% of the variation for this feature explained, every other feature studied had a very low correlation with this component, and it covers a minute 1.31% of the total variance observed in the data set. Combination of PC1 and 7 allowed for meaningful separation of GlcN,6,3-S levels, the rare sulphate responsible for heparins pharmaceutical activity. Components 1, 5, 6 and 7 all have major loadings that correspond to different sections of the main carbohydrate band at ~1000 cm^{-1} with other small contributions across the spectrum [1] (**Figs. 5.3E** and **5.4I,J and K**). PC2, 3 and 4 possess broader contributions across the whole spectrum – most likely due to their correlations with alien saccharide constituents,

including DS, CS and saccharides found in the still attached linkage region (**Figs. 5.3F and G and 5.4H**).
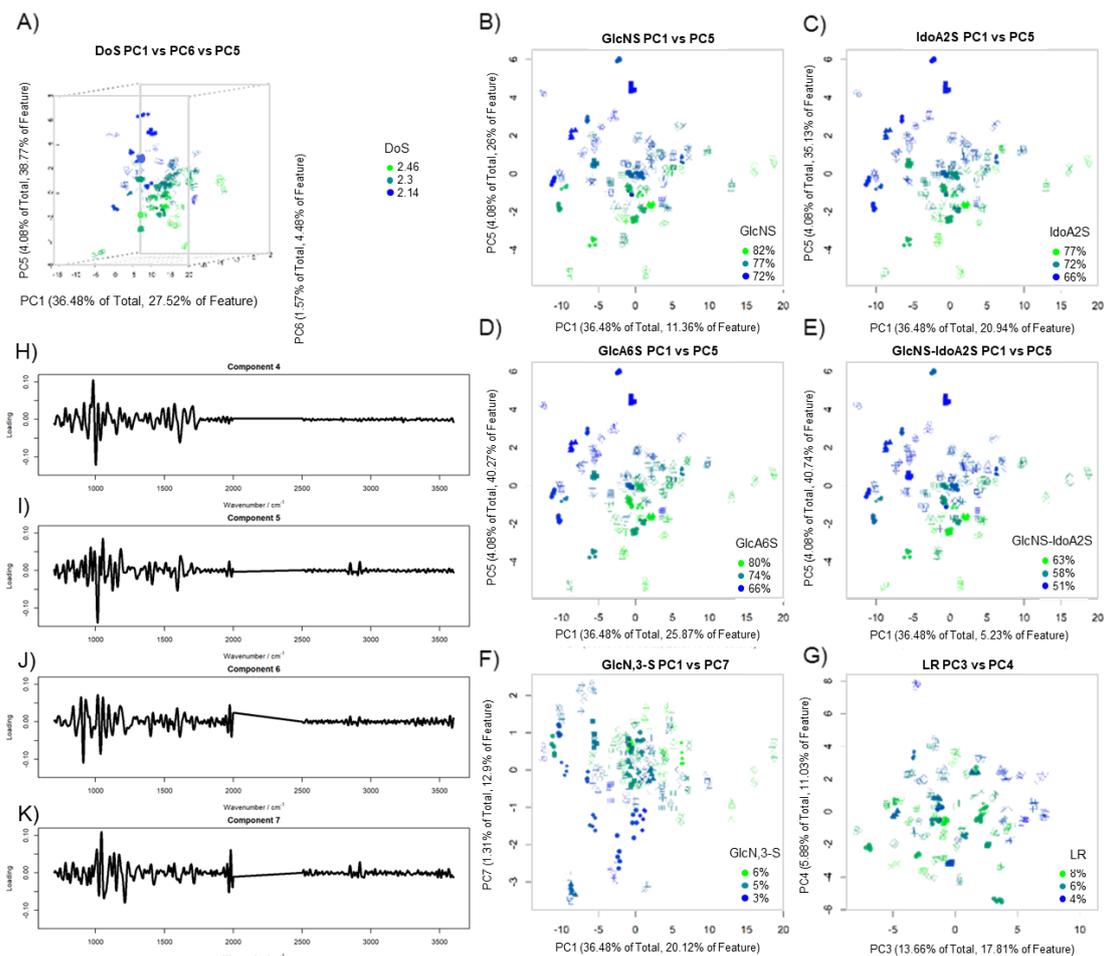
Figure 5.4: PCA analysis of crude heparins, coloured to indicate the levels of different structural features. A) 3D scores plot of PC1 vs PC5 vs PC6, coloured to indicate the DoS of heparin in the sample. **B)** Scores plot of PC1 vs PC5, coloured to indicate the levels of GlcNS. **C)** Scores plot of PC1 vs PC5, coloured to indicate the levels of IdoA-2S. **D)** Scores plot of PC1 vs PC5, coloured to indicate the levels of GlcN-6S. **E)** Scores plot for PC1 vs PC5, coloured to indicate the levels of GlcNS-IdoA2S. **F)** PC scores plot of PC1 vs PC7, coloured to indicate the levels of GlcN,3-S. **G)** Scores plot of PC3 vs PC4, coloured to indicate the levels of heparin which still have some of their protein linkage region remaining. **H)** Loadings plot for PC4. **I)** Loadings plot for PC5. **J)** Loadings plot for PC6. K) Loadings plot for PC7. PCs were selected based on the portion of variance which they cover, given the structural feature they are examining (**Table 5.2**). All samples are coloured depending on the amount of each structural feature on a green to blue scale – the greener the sample the higher the level. All spectra involved in PCA are in the regions of 700 and 2000 and, 2500 and 3700 cm-1 and are averages of 3 scans. The spectra were smoothed, baseline corrected with a 7th order polynomial, normalised (0-1), smoothed and derivatised with a Savitzky-Golay algorithm and finally autoscaled.

Table 5.2: Table of variance covered for each PC for many structural features of heparin, CS and DS. Values are computed with PCR, using the *pcr* function from the ***pls*** package.

PCR Structural Feature	*Variance Covered / %*							
	PC1	**PC2**	**PC3**	**PC4**	**PC5**	**PC6**	**PC7**	**PC8**
Total Variance	36.48	23.05	13.66	5.88	4.08	1.57	1.31	1.03
DS Level	10.54	30.60	33.91	0.06	5.50	0.01	0.47	0.47
CS Level	24.40	11.01	0.00	1.45	25.45	0.46	0.97	2.16
IdoA2S-GalNac (DS-2S)	0.12	13.44	28.20	0.40	2.31	0.71	2.01	0.27
GalNac-6S (CS-6S)	5.88	2.26	0.52	0.09	26.97	1.94	1.76	2.02
GlcN,3S,6X	20.12	0.17	0.03	0.24	4.63	0.97	12.90	0.23
GlcA-GlcNS,3S,6X	0.83	5.19	6.07	0.12	0.70	0.29	6.00	0.38
IdoA2S-GlcN	25.87	0.75	0.72	0.08	25.99	2.86	0.01	0.82
DoS	27.52	0.04	1.65	0.16	38.77	4.48	1.96	0.09
GlcNAc,6X-G	15.41	0.11	0.66	4.81	40.74	4.15	2.65	0.72
GlcNS,6X-GlcA	0.02	0.00	0.00	1.02	7.10	1.03	0.01	0.88
GlcNS,6X-IdoA	2.99	0.56	2.24	0.00	1.15	0.12	0.58	6.96
GlcNS,6X-IdoA2S	5.23	0.00	2.10	3.55	40.74	3.12	0.04	2.43
IdoA2S-GlcNH2,6X	0.13	0.25	16.10	1.48	19.79	0.12	2.02	0.01
GlcNAc,6X	16.11	0.16	0.89	1.64	46.18	6.66	0.64	0.77
GlcNS,6X	11.36	1.94	0.33	0.34	28.00	8.74	0.34	1.01
GlcNx,6S	25.87	0.01	2.77	0.00	40.27	3.23	3.05	0.25
GlcA-GlcNAc,6X	13.61	0.00	6.62	0.61	40.23	0.38	0.75	0.73
GlcA-GlcNS,6X	9.29	0.13	0.03	0.49	0.80	4.89	0.93	0.37
IdoA-GlcNx	15.13	2.41	0.37	1.82	6.99	1.57	0.00	0.80
IdoA-GlcNx,6S	3.86	0.68	1.96	0.02	4.97	1.40	0.19	6.23
GlcNH2,6X	0.83	1.51	1.20	2.68	35.33	1.10	0.06	0.25
GlcA2OH	20.37	0.72	1.68	0.65	37.04	2.16	0.07	0.08
GlcA2S	1.56	16.88	0.95	0.09	0.03	0.31	1.99	5.78
IdoA2OH	11.43	1.90	1.98	0.66	9.00	2.04	0.08	5.28
IdoA2S	20.94	0.00	2.35	0.28	35.13	2.91	0.26	0.28
LR	3.38	0.36	17.81	11.03	0.53	4.40	0.92	0.00
GNR	0.69	8.06	0.67	4.92	0.18	0.00	4.13	0.04

The PCR suggests that the levels of sulphate in crude heparins can be identified, hence the ability to predict each of the different structural features was tested. A random subset of 13 crude heparins were removed from the library, PCA and subsequent PCR analysis was performed on the crude samples, not including the selected 13, and from the PCR model(s) of the first 5 PCs, the 27 structural features for the test 13 crude heparins were predicted, and the predicted values compared to the actual values. For each feature, an R^2, P-value and root-mean squared error (RMSE) were calculated. All features bar one (the level of I2S-GlcNH2) were predicted accurately with PCR (**Table 5.3**). Aside from those of I2S-GlcNH2, all predicted values correlated with the expected values ($R^2 > 0.7$, $R^2_{average} = 0.895$). T-tests between the predicted and real values were performed, and all features had P-values > 0.01 i.e., they were not significantly different to the real values. For the structural features in

heparin, the RMSE was on average 0.91%. The LODs observed for CS and DS (**Table 4.1**) with PCA-LRM were also confirmed here, with a RMSE of 1.71% (i.e., an observed LOD of 2.5%) for DS and a RMSE of 0.77 (an observed LOD of <1.0%) for CSs. The RMSE for heparin, CS and DS were lower than those that were calculated with the mammalian GAG library constructed in **Chapter 4**, **Table 4.10.** The RMSEs observed in **Table 4.10** were 5.26, 4.42 and 2.6 for heparin, CS and DS respectively compared with 2.48, 1.71 and 0.77 for heparin, CS and DS respectively seen here, corroborating the idea that an improved library that is more specific to the test samples leads to improved separation.

Table 5.3: Descriptive statistics for the PCR prediction of various compositional and structural features of heparin, CS and DS in crude heparins. 13 randomly selected samples were chosen from the crude library. A PCR was performed on the remainder of the library. The PC scores for the 13 test samples were predicted and underwent compositional and structural analysis with the resulting PCR model. R2, P- and RMSE values are calculated based on the observed and expected levels of each feature. The predicted levels of the samples should not be significantly different to the expected, hence P- values should be greater than the critical value of 0.001.

	Feature	DS	CS	DS2S	CS6S						PCR
Composition	R²	0.96	0.93	0.77	0.75						
	P-value	0.72	0.85	0.13	0.84						
	RMSE	1.71	0.77	4.64	11.11						
Structural Features	Feature	DoS	GlcNS	IdoA2S	GlcA6S	GlcN,3-S	GlcNAc	IdoA2OH	GlcA2OH	GlcNH2	
	R²	0.95	0.93	0.92	0.96	0.81	0.94	0.93	0.88	0.96	
	P-value	0.82	0.12	0.86	0.64	0.04	0.58	0.31	0.57	0.63	
	RMSE	0.04	1.38	1.85	1.92	0.44	1.14	0.70	1.32	0.39	
	Feature	GlcA-GlcNAc	GlcA-GlcNS	IdoA-GlcN	idoA-GlcNG6S	GlcUa-GlcN,3-S	IdoA2S-GlcN	GlcNAc-GlcA	GlcNS-GlcA	GlcNS-IdoA	GlcNS-IdoA2S
	R²	1.00	0.80	0.90	0.87	0.70	0.92	0.90	1.00	0.74	0.94
	P-value	1.00	0.18	0.27	0.70	0.01	0.92	0.35	1.00	0.61	0.15
	RMSE	0.00	0.77	0.47	0.60	0.59	1.77	1.18	0.00	1.84	1.84
Linkage Region	Feature	LR	GNR								
	R²	0.93	0.97								
	P-value	0.49	0.15								
	RMSE	0.40	0.23								

	Average RMSE				Average R²	
	All	Composition	GAG structural features	Heparin Structural features		
	1.37%	1.24%	7.88%	0.91%	0.89	

5.3.3 The use of different chemometric models to improve composition prediction.

As the predictions observed improve with the addition of new samples to each new model, it was hypothesised that an entirely new method of generating the model might facilitate increased detection. It has already been observed that different linear combinations of variables can detect the same contaminants, as with OSCS in **Figure 4.6**. Partial least squares (PLS) is a trained, iterative technique that generates latent variables (LVs – analogous to PCs), based on known dependant variables (Wold et al., 2001, p. 0). PLS is performed on 2 matrices, the Y matrix: a data frame of the values to be predicted, and the X matrix, a data frame of the variables to be predicted from – the same data frame entered into a PCA. PLS will create LVs which explain the maximum covariance between matrices X and Y, not the maximum variance in matrix X as in PCA. The LVs can be explored in essentially the same way as PCs post PCR. In principle, PLS will generate LVs that contain trends which explain

the features that the analyst is observing and hence should provide improved separation compared to PCR and hence, to the library of ATR-FTIR spectra of crude heparins, PLS-regression (PLS-R) was performed.

Table 5.4: Table of variance covered for each LV for many structural features of heparin, CS and DS. Values are computed with PLS-R, using the *pls* function from the ***pls*** package.

PLS Structural Feature	*Variation Covered / %*							
	LV1	**LV2**	**LV3**	**LV4**	**LV5**	**LV6**	**LV7**	**LV8**
Total Variance	34.36	21.05	7.41	13.80	5.58	1.58	1.16	0.84
DS Level	32.07	41.47	5.15	0.54	1.35	0.85	0.06	0.03
CS Level	33.43	0.01	29.85	1.42	0.48	0.14	3.69	1.23
IdoA2S-GalNac (DS-2S)	2.74	37.32	2.40	2.76	0.06	1.88	2.87	1.02
GalNac-6S (CS-6S)	8.88	0.04	24.27	8.68	1.63	2.60	4.50	10.98
GlcN,3S,6X	15.71	5.79	1.99	1.43	1.02	1.43	6.02	2.56
GlcA-GlcNS,3S,6X	4.07	8.09	0.33	1.80	0.94	7.41	0.17	2.85
IdoA2S-GlcN	25.04	4.63	20.90	3.63	1.28	4.42	1.08	0.70
DoS	23.48	10.04	30.79	5.69	0.31	0.37	1.20	2.78
GlcNAc,6X-G	13.34	7.02	35.72	5.33	1.82	0.56	0.41	1.03
GlcNS,6X-GlcA	0.03	0.17	6.15	1.09	1.71	1.08	0.13	0.03
GlcNS,6X-IdoA	1.43	4.04	0.57	0.01	1.35	2.39	2.17	2.38
GlcNS,6X-IdoA2S	3.86	6.47	36.68	3.96	0.04	4.67	0.57	0.00
IdoA2S-GlcNH2,6X	0.98	10.54	26.67	0.78	0.13	1.04	0.65	1.55
GlcNAc,6X	14.45	6.52	39.70	6.88	0.00	1.66	0.14	1.27
GlcNS,6X	13.23	1.12	27.01	3.64	1.02	6.27	0.10	1.02
GlcNx,6S	20.82	12.39	32.18	5.27	0.38	0.15	0.81	3.72
GlcA-GlcNAc,6X	10.03	12.78	38.24	1.92	0.09	0.00	0.17	3.25
GlcA-GlcNS,6X	8.66	0.97	0.58	0.33	0.65	0.01	0.05	0.54
IdoA-GlcNx	9.50	8.04	2.27	5.36	4.99	1.24	0.06	2.52
IdoA-GlcNx,6S	2.03	4.74	2.93	0.38	2.41	5.21	5.56	0.09
GlcNH2,6X	2.04	4.63	29.60	4.34	2.64	2.68	0.14	0.97
GlcA2OH	19.27	6.11	33.46	3.62	0.09	0.83	0.11	1.51
GlcA2S	6.49	4.29	1.46	7.56	1.21	9.12	3.23	0.50
IdoA2OH	6.64	9.30	4.26	2.50	5.42	4.98	3.28	0.28
IdoA2S	17.07	9.91	28.55	4.47	0.65	1.77	0.83	1.44
LR	6.73	7.52	0.71	16.80	9.11	1.65	1.04	2.03
GNR	0.00	1.91	3.93	5.32	5.35	0.15	0.33	0.17

The generated LVs were first inspected for the variance which they cover (**Table 5.4**), and then graphed for a better visual representation (**Figs. 5.5** and **5.6**). The total variance covered by the LVs was equivalent to the total variance covered by each PC in the PCR model (**Table 5.2**), but the portion of the variance covered for each response variable in each LV was increased. The majority of the variance for each response variable (>85%) was covered by LVs 1 to 4, as opposed to PCs 1 to 6 in PCR.
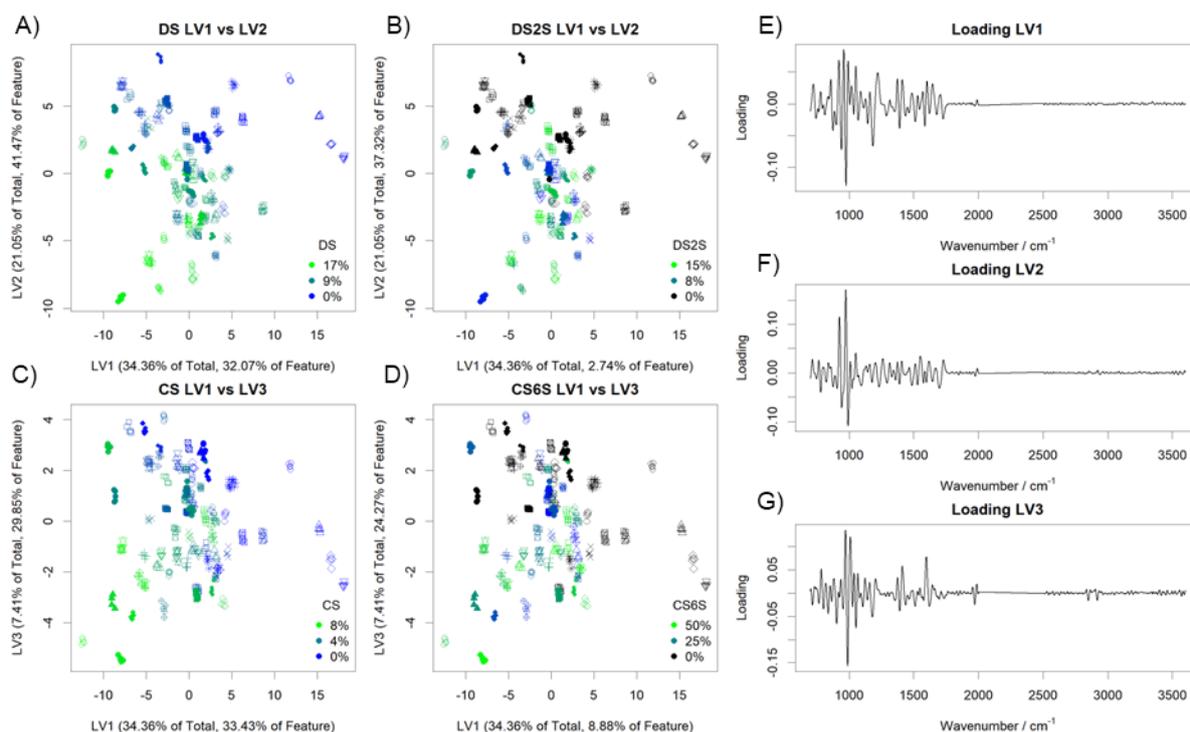
Figure 5.5: PLS-R analysis of crude heparins, coloured to indicate the levels of CS and DS. A) Scores plot of LV1 vs LV2, coloured to indicate level of DS. **B)** Scores plot LV1 vs LV2, coloured to indicate the level of IdoA-2S present in the DS moieties. **C)** Scores plot of LV1 vs LV3, coloured to indicate the level of CS. **D)** Scores plot of LV1 vs LV3, coloured to indicate the level of GalNAc-6S present in the CS moieties. **E)** Loading plot for LV1. **F)** Loading plot for LV2. **G)** Loading plot for LV3. All samples are coloured depending on the amount of each structural feature on a green to blue scale – the greener the sample the higher the level. LVs were selected based on the portion of variance which they cover, given the structural feature they are examining (**Table 5.4**). All spectra involved in PLS-R are in the regions of 700 and 2000 and, 2500 and 3700 cm^{-1} and are averages of 3 scans. The spectra were smoothed, baseline corrected with a 7th order polynomial, normalised (0-1), smoothed and derivatised with a Savitzky-Golay algorithm and finally autoscaled.

The trends observed in PCR were observed in PLS, but with higher components. LV1 to 3 contained most of the variance for both sample composition and the structural features, with the sums of the variances associated with the level of DS and CS, and the DoS and the levels of 6S, 2S and NAc being 78.69%, 63.29, 64.31%, 65.39%, 55.21% and 61.05% compared with 75.05%, 35.41%, 38.77%, 40.27%, 23.29% and 17.16% for the sums of the variance covered for each figure in PCs 1 to 3. LV4 and 5 correlated to the level of linkage region (16.80% and 9.11% respectively), and LV7 correlated with GlcN,3,6-S (6.02%). Similar scores plots were observed with PLS as were observed with PCA (**Figs. 5.5** and **5.6**).
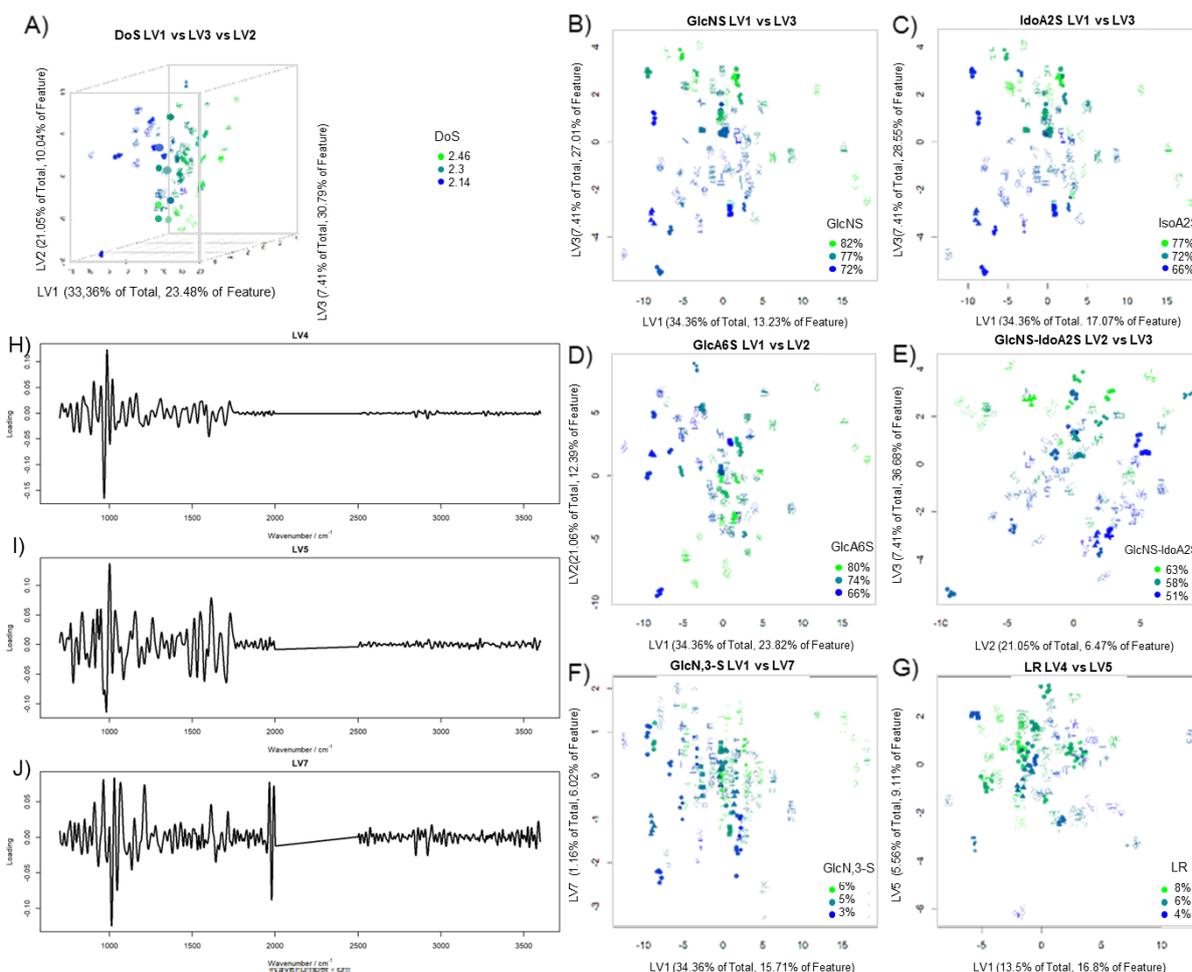
Figure 5.6: PLS-R analysis of crude heparins, coloured to indicate the levels of different structural features. A) 3D scores plot of LV1 vs LV2 vs LV3, coloured to indicate the DoS of heparin in the sample. **B)** Scores plot of LV1 vs LV3, coloured to indicate the levels of GlcNS. **C)** Scores plot of LV1 vs LV3, coloured to indicate the levels of IdoA-2S. **D)** Scores plot of LV1 vs LV2, coloured to indicate the levels of GlcN-6S. **E)** Scores plot for LV2 vs LV3, coloured to indicate the levels of GlcNS-IdoA2S. **F)** Scores plot of LV1 vs LV7, coloured to indicate the levels of GlcN,3-S. **G)** Scores plot of LV4 vs LV5, coloured to indicate the levels of heparin which still have some of their protein linkage region remaining. **H)** Loadings plot for LV4. **I)** Loadings plot for LV5. J) Loadings plot for LV7. All samples are coloured depending on the amount of each structural feature on a green to blue scale – the greener the sample the higher the level. LVs were selected based on the portion of variance which they cover, given the structural feature they are examining (**Table 5.4**). All spectra involved in PLS-R are in the regions of 700 and 2000 and, 2500 and 3700 cm^{-1} and are averages of 3 scans. The spectra were smoothed, baseline corrected with a 7th order polynomial, normalised (0-1), smoothed and derivatised with a Savitzky-Golay algorithm and finally autoscaled.

Following this, prediction of the structural features for the same 12 crude samples that were used in **Table 5.5** was undertaken. As with PCR, all variables had P-values > 0.01 showing that the predicted levels were not significantly different to the expected. All predictions retained high high R^2 values that were higher than those observed with PCR (all R^2 > 0.80 compared with 0.70 in PCR and R$^2_{average}$ = 0.920, compared with 0.890 in PCR). The average RMSE decreased too, and for the structural features in heparin, an improvement of from 0.91 with PCR to 0.76 was achieved. This shows that ATR-FTIR can detect levels of sulphate at specific positions under 1%.

[145]

Table 5.5: Descriptive statistics for the PLS-R prediction of various compositional and structural features of heparin, CS and DS in crude heparins. 13 randomly selected samples were chosen from the crude library. A PLS-R was performed on the remainder of the library. The spectra for the 13 test samples underwent compositional and structural analysis with the resulting PLS-R model. R2, P- and RMSE values are calculated based on the observed and expected levels of each feature. The predicted levels of the samples should not be significantly different to the expected, hence P- values should be greater than the critical value of 0.001.

Composition

Feature	DS	CS	DS2S	CS6S		PLS
R²	0.98	0.95	0.81	0.76		
P-value	0.74	0.89	0.28	0.92		
RMSE	1.41	0.66	4.20	10.93		

Structural Features

Feature	DOS	GlcNS	IdoA2S	GlcA6S	GlcN,3-S	GlcNAc	IdoA2OH	GlcA2OH	GlcNH2
R²	0.96	0.97	0.93	0.96	0.80	0.97	0.93	0.91	0.98
P-value	0.85	0.31	0.89	0.72	0.03	0.73	0.37	0.64	0.69
RMSE	0.04	0.98	1.68	1.87	0.46	0.80	0.68	1.18	0.31

Feature	GlcA-GlcNAc	GlcA-GlcNS	IdoA-GlcN	idoA-GlcN6S	GlcUa-GlcN,3-S	IdoA2S-GlcN	GlcNAc-GlcA	GlcNS-GlcA	GlcNS-IdoA	GlcNS-IdoA2S
R²	1.00	0.88	0.90	0.90	0.86	0.94	0.96	1.00	0.81	0.96
P-value	1.00	0.38	0.37	0.68	0.07	0.94	0.50	1.00	0.72	0.30
RMSE	0.00	0.61	0.47	0.52	0.41	1.54	0.78	0.00	0.73	1.48

Linkage Region

Feature	LR	GNR
R²	0.94	0.97
P-value	0.41	0.23
RMSE	0.51	0.45

Average RMSE				Average R2
All	Composition	GAG structural features	Heparin Structural features	
1.31%	1.04%	7.57%	0.76%	0.92

To test the robustness of the model, in a manner similar to that undertaken in Chapter 4 (**Tables 4.8** and **4.9**), marine GAGs also underwent sulphate level prediction with the PCR and PLS models. Of the crude GAGs examined in **Tables 4.8** and **4.9**, the two with high levels of heparin (prawn and crab) were taken forward and enriched with SAX-chromatography. From the crude prawn, two fractions (F4 and F5) were taken and from the crude crab another (F4) and the levels of each of the sulphates were assessed with NMR. Compositionally, these samples resemble crude heparin (2 parts heparin, 1 part CS/DS), and hence underwent prediction using the PCR and PLS-R models that were trained on crude-heparin ATR-FTIR spectra (**Table 5.5**). With PCR, the levels of NS and 6S were predicted well, all P-values > 0.01 and RMSE of 4.29 and 3.94 respectively. The values predicted with PCR were better than those observed with PLS, which had RMSE of 6.81 and 11.93 for the level of NS and 6S respectively. The DoS was close to the expected values for both prawn F4 and crab F5 (residuals of 0.15, and 0.01) but was incorrect for prawn F5 (residual of 0.50). Both models failed to predict the level of 2S, with RMSE of 27.31 and 30.05 for PCR and PLS-R respectively. The level of 2S was overpredicted for both prawn fractions but was predicted

correctly for the crab GAG. This is most likely due to low levels of 2S in both prawn GAGs at 40% and 30% for F4 and F5 respectively. These levels are not covered by the levels found in the model (min = 63.3%, mean = 73.3 +/- 3.4%). Nevertheless, strong predictions, some of which fell outside of the bounds of the model, were achieved with these techniques, suggesting that ATR-FTIR can be used as a predictive tool for the structures of heparin and heparinoids in heparin containing mixtures.

Table 5.5: PCR and PLS-R prediction of the DOS, and levels of NS, 6S and 2S for three marine GAGs.

GAG Sample	Expected from NMR / %				PCR Prediction / %				PLS Prediction / %			
	DOS	NS	6S	2S	DOS	NS	6S	2S	DOS	NS	6S	2S
L. Vannamei F4	1.71	70	50	40	1.86	68.48	54.83	59.68	2.04	73.12	61.58	63.7
L. Vannamei F5	1.82	70	70	30	2.32	77.18	74.35	72.9	2.45	80.54	79.24	76.24
P. Pelagicus F4	2.05	75	60	70	2.04	73.88	62.05	66.93	2.34	79.26	74.41	72.99
RMSE					0.30	4.29	3.94	27.31	0.44	6.81	11.93	30.05
P-value					0.27	0.64	0.67	0.19	0.06	0.10	0.21	0.13

5.3.4 The application of FTIR to heparins from different animal species.

The final facet of heparin quality control is that of heparin extracted from different animal sources. Currently, there are 3 sources of heparin that are being considered for pharmaceutical use: PMH (the standard in many countries) BMH, reintroduced in Brazil, BLH (standard for Islamic countries) and finally OMH, a type not currently in pharmaceutical use, but under consideration (Fareed et al., 2019; Vilanova et al., 2019b). Discretion between these compounds is paramount as, like in Brazil, they may possess unique monographs and, as in the UK and Islamic countries, they may be banned for use due to medical concerns and religious issues respectively. The key chemical differences between each heparin from each source amount to underlying levels of sulphation, overall pattern and chain lengths (**Table 5.1**). As ATR-FTIR-PCA has been shown to detect varying levels of different sulphates in crude heparins (**Table 5.3**), it is reasonable to assume that it can discriminate between the different heparin types and hence a library of each heparin type was constructed, including 69 PMHs, 33 OMHs, 57 BMHs and 19 BLHs before subjection to PCA. As the sample groups are discrete, a regression cannot be performed to identify relevant components hence, to decide which PCs to observe, the frequency density for each heparin type in each component was calculated and graphed. Using the calculated densities, it is easy to observe which components separate for which heparin (**Fig. 5.7**) for example, in (**Fig. 5.7H**) the OMH samples are seen to separate from the other three heparins..
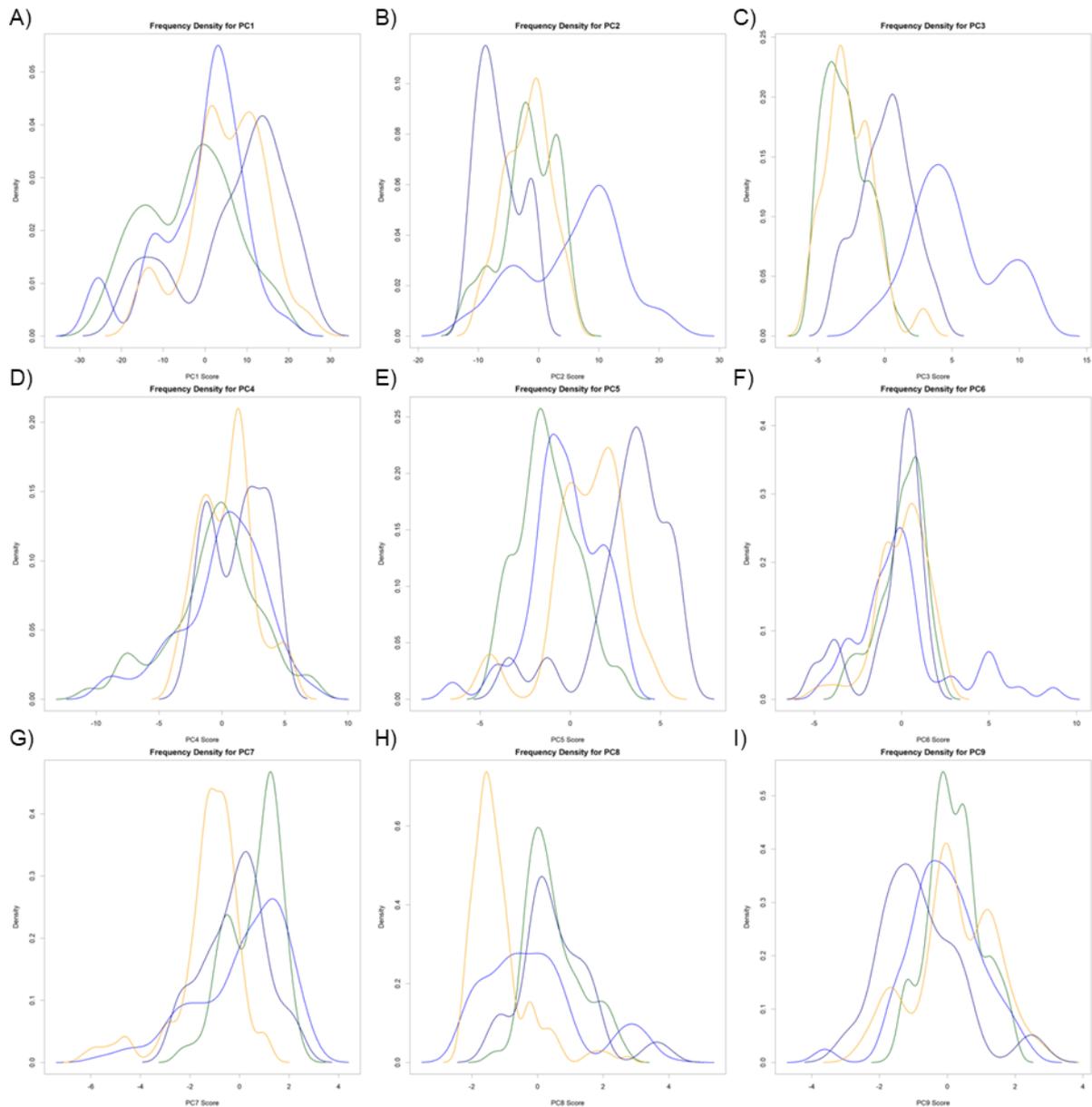
Figure 5.7: Frequency densities of PMH, BMH, BLH and OMH in different PCs. A through **I)** correspond to the frequency densities for each heparin in PCs 1 through 9 respectively. PMH: green, BMH: blue, BLH: dark blue, OMH: orange.

Clear separation of species or the formation of distinct lobes was observed (**Fig. 5.8**). BMH separated strongly from the rest of the heparins with PCs 2 and 3 (**Fig. 5.8A**). Through PC 5, all heparin species form one large cluster, however, 4 lobes within this cluster are evident - each populated with a single source of heparin, with BLH at the top, BMH to the right, PMH at the bottom left and OMH at the top left (**Figs. 5.7E** and **5.8B**). PMH, OMH and BLH do not separate entirely, but do form distinct nodules (**5.8C and E**). Separation between BLH and the other samples can be achieved with PC2 and 5 (**Fig. 5.8D**), where it forms a separate node at the top left and separation between PMH and OMH can be observed in PC8 (**Fig. 5.8E**).
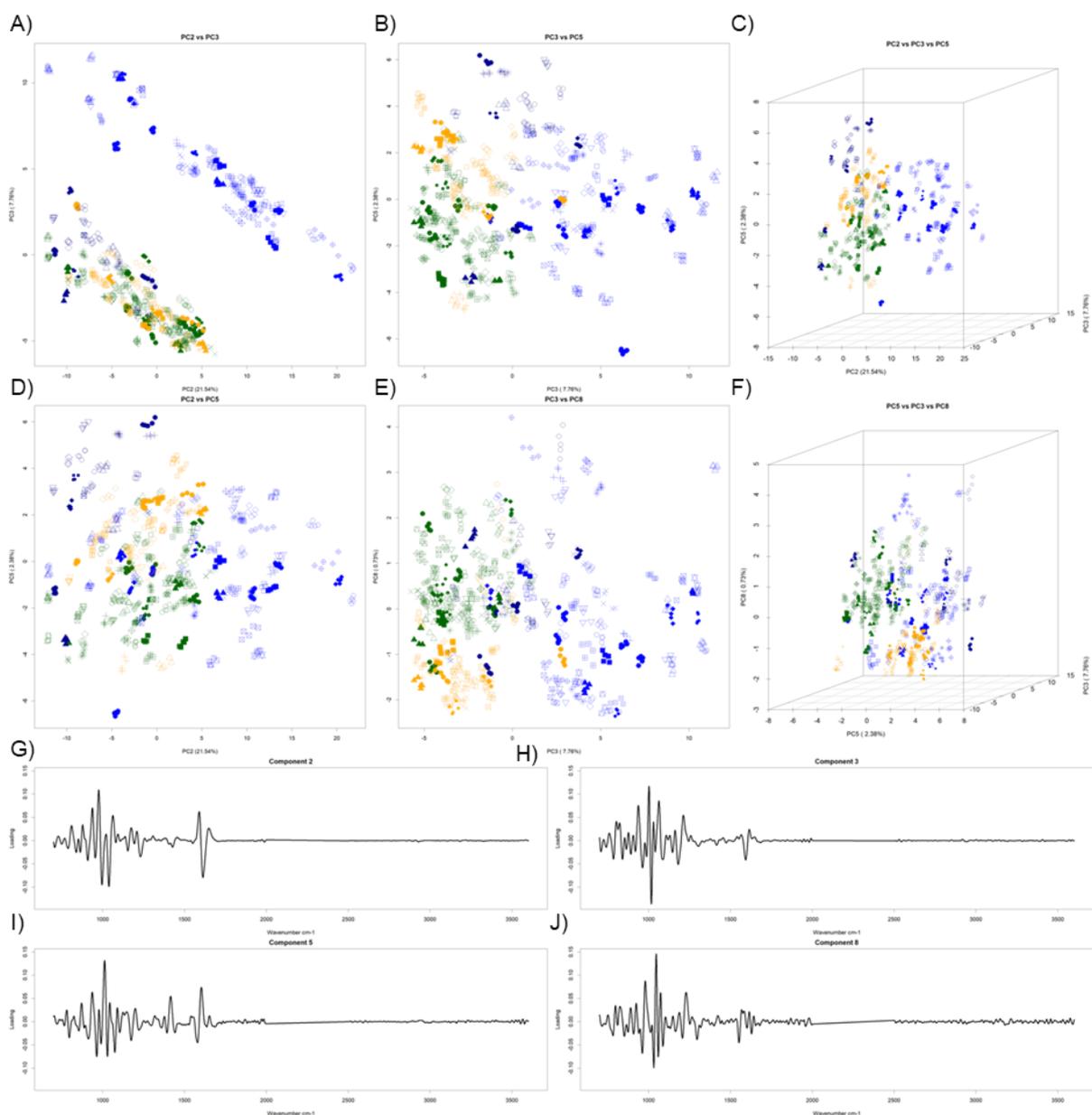
Figure 5.8: PCA of ATR-FTIR spectra of heparins extracted from different animal sources. A) Scores plot of PC2 vs PC3. **B)** Scores plot of PC3 vs PC5. **C)** 3D scores plot of PC2 vs PC3 vs PC5. **D)** Scores plot of PC2 vs PC5. **E)** Scores plot of PC3 vs PC8. **F)** 3D Scores plot of PC3 vs PC5 vs PC8. **G)** Loadings plot for PC2. **H)** Loadings plot for PC3. **I)** Loadings plot for PC5. **J)** Loadings plot for PC8. PMH: green, OMH: orange, BMH: blue, BLH: dark blue. PCs were selected for graphing through visual inspection of frequency densities in each component **Figure 5.7**. All spectra involved in PCA are in the regions of 700 and 2000 and, 2500 and 3700 cm-1 and are averages of 3 scans. The spectra were smoothed, baseline corrected with a 7th order polynomial, normalised (0-1), smoothed and derivatised with a Savitzky-Golay algorithm and finally autoscaled.

Prediction of the animal source of a heparin requires the identification of a relevant lobe. This will be difficult due to the close proximity of each lobe and the requirement for multiple components to separate each heparin type. Subtle outliers of one group may be misclassified as part of the other and a more rigorous method of identification is required. As there are numerous heparin types, a logistic model is impractical as for $n_{species}$, $n\frac{(n-1)}{2}$ logistic models must be made.

5.3.5 The application of untrained cluster analyses to detect and identify heparins from different animal species.

Cluster analysis is one option that would allow mathematical identification of unknown samples in compact clusters (Aggarwal and Reddy, 2013). Cluster analysis involves computation of a cluster which accounts for one sample type, and subsequently testing for which cluster new samples belong to. Clustering can be split into 5 main types, including connectivity-based, centroid-based, distribution-based, density-based and grid-based clustering. Connectivity-based clustering involves moving a sample closest to its most similar neighbor and then moving the two samples to their most similar neighbor and so on, forming a dendrogram. Centroid-clustering involves creating a central point which is not necessarily an input data point and moving it to minimise the distances between each data point that belongs to the cluster. New samples can then have their distances between each point calculated and be subsequently mapped to new clusters. This style of clustering often requires input of the number of centroids, hence some prior knowledge of the number of expected clusters is required. Distribution-based clustering relates samples by some mathematical distribution, often gaussian, and seeks to solve for a distribution that covers different clusters. This style of clustering can require input of the number of clusters and can have reproducibility issues as different runs can produce slightly different clusters, due to random initialisation. Density-based clustering creates clusters based on the density of points in an area. Areas with few or distant points are counted as noise or border points. Density-based clusters can also create "ghost" areas, in which new points may fall, however it struggles greatly with overlapping lobes and with areas that are uniform in density. Grid-based clustering is an iterative density cluster applied to a multivariate data set.

When applied to heparin from different sources, it is generally accepted what types there are and how many clusters there will be, hence any of the above methods can be applied. To begin, a simple HCA clustering algorithm (Nielsen, 2016) was applied to the normalised data (**Fig. 5.9**), the differentiated and the auto scaled data (**Fig. 5.10**) and then to the scores calculated after PCA (**Figs. 5.11 and 5.12**). For visual clarity of the signals separated for, a heatmap was also plotted with the data that was classified by HCA. The normalised data itself did not have the ability to form clusters of heparin from different animal sources. Some clusters of BMH, OMH and PMH appeared, but the samples were not resolved when applied to the D2 data and all the PCs, the majority of the BMHs were clustered together, but the rest of the samples are mixed, as observed with PCA (**Fig. 5.8A**). Selective application of HCA to the PCs which correspond best with heparin species separation (PCs 2, 3 and 5) increased PMH clustering but produced 2 separate BMH clusters. The BLHs and OMHs were not clustered with HCA.

[150]

Heatmap of Normalised Data
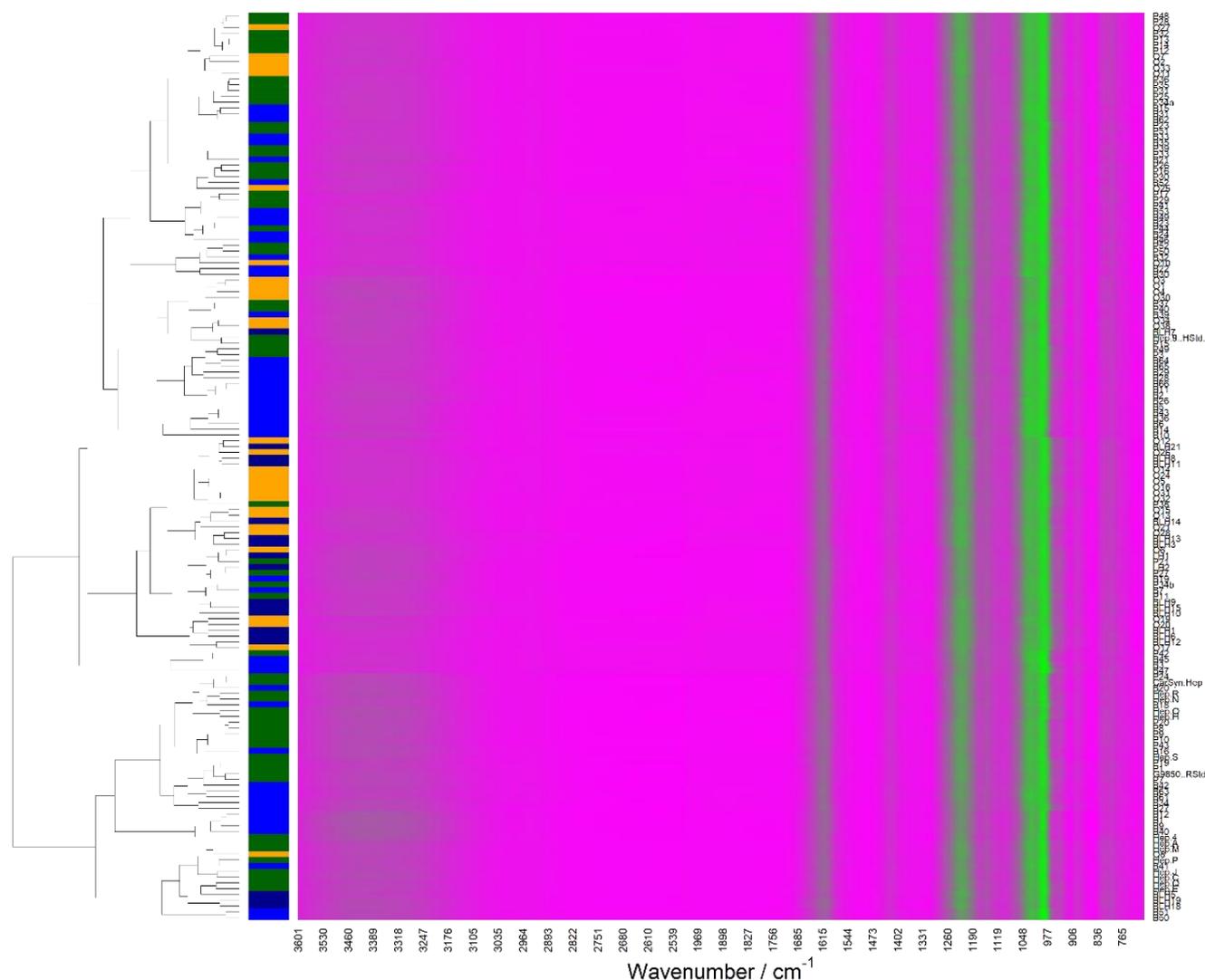
Figure 5.9: Heatmap and subsequent HCA of ATR-FTIR spectra of heparins extracted from different animal species. The dendrogram is located to the left of the plot. Between the plot and dendrogram are colour codes for which animal source the sample is from. PMH: green, BMH: blue, BLH: dark blue, OMH: orange. Spectra to undergo HCA were the averages of 3 scans in the regions of 700 to 2000 cm-1 and, 2500 to 3700 cm**-1**. They were smoothed, baseline corrected with a 7th order polynomial and normalised (0-1). Averages of 5 spectra for each sample were used.

Three more clustering techniques comprising the remaining categories of clustering type; centroid-based K-means clustering (KMC) (Hamerly and Elkan, 2002); density-based: density based spatial clustering of applications with noise (DBSCAN) (Kriegel et al., 2011) and; distribution-based, expectation maximisation (EM) (Dempster et al., 1977) were trialled. Generally, these clustering techniques are applied to data sets that have already had their dimensions reduced (i.e., through PCA) hence they were trialled on the scores of PCs 2, 3 and 5 from the PCA in **Figure 5.8**.

Heatmap of Autoscaled D2 Data
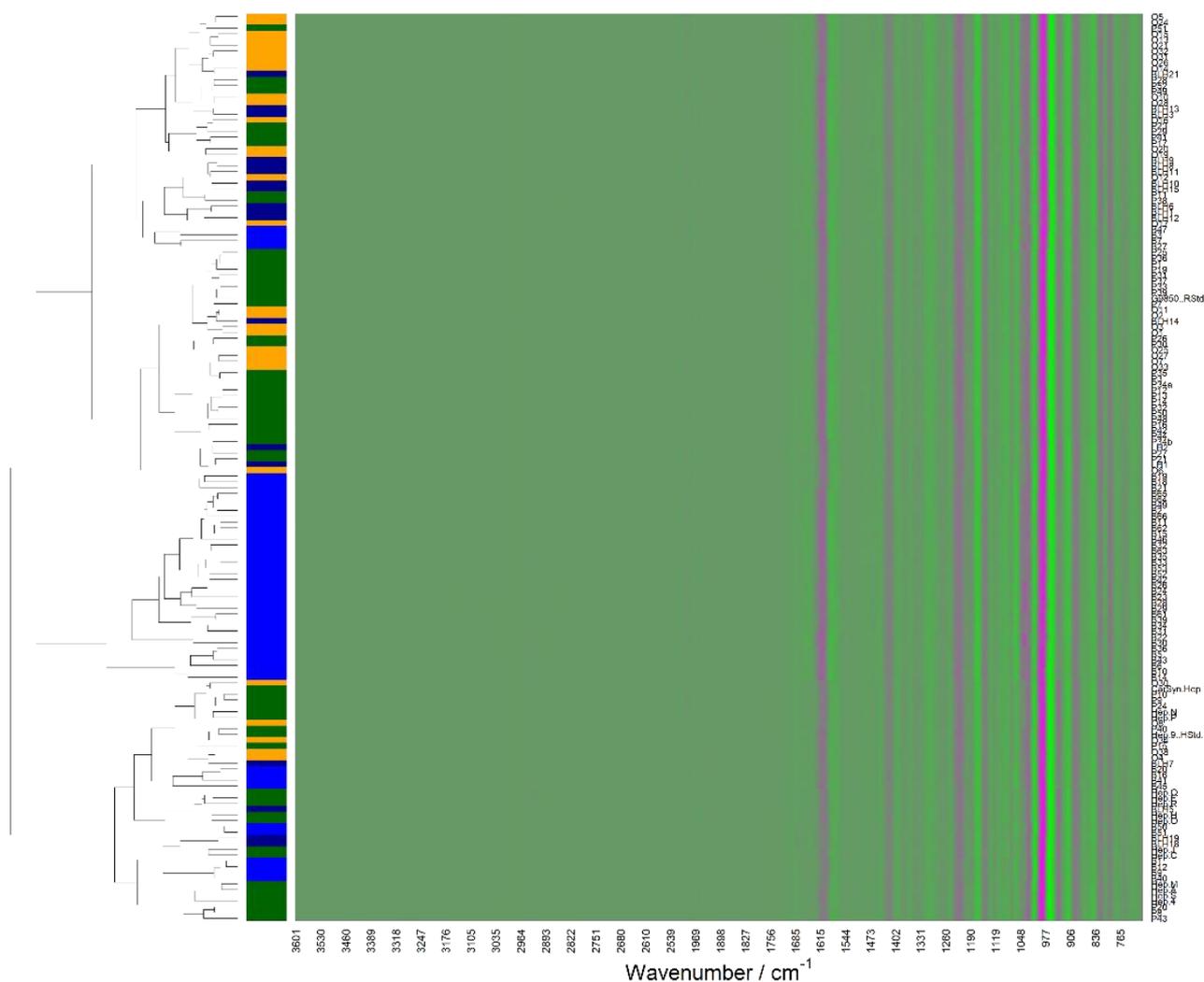
Figure 5.10: Heatmap and subsequent HCA of second derivative ATR-FTIR spectra of heparins extracted from different animal species. The dendrogram is located to the left of the plot. Between the plot and dendrogram are colour codes for which animal source the sample is from. PMH: green, BMH: blue, BLH: dark blue, OMH: orange. All spectra involved in HCA are in the regions of 700 and 2000 and, 2500 and 3700 cm**-1** and are averages of 3 scans. The spectra were smoothed, baseline corrected with a 7th order polynomial, normalised (0-1), smoothed and derivatised with a Savitzky-Golay algorithm and finally autoscaled.

As with HCA, BMH was clustered from the others in all circumstances, albeit into 2 clusters for KMC and DBSCAN (**Figs. 5.13, 5.14, 5.15 and 5.16**). KMC and DBSCAN had no separation between PMH, OMH and BLH, either separating the PMH-OMH-BLH lobe perpendicular to the expected splits or not at all (**Figs. 5.13B and F** and, **5.14D**). Partial separation of BLH was observed with KMC in PCs 3 and 5, but this separation also clustered some OMH samples, incorrectly (**Fig. 5.13D**). EM offered clustering which separated the heparins from different sources, creating 2 BMH clusters, a BLH cluster and within the remaining samples, 3 more clusters (**Fig. 5.15**). Arguably, the ability to differentiate bovine heparins from others, as with EM clustering, is the most important due to religious issues surrounding bovine use, and the contemporaneous use of bovine heparins in countries such as Brazil, hence the ability to cluster BLH and BMH from the other heparins is of great

importance. Furthermore, EM is an example of "soft" clustering, where samples are not strictly within the cluster, but are instead most likely to exist in their cluster but may exist in others. Hence, EM also offers the analyst the ability to flag potential outliers and edge cases for further study. With PC8, some clustering could be achieved which discriminated some PMH and OMH with KMC and EM clustering, albeit into multiple clusters.
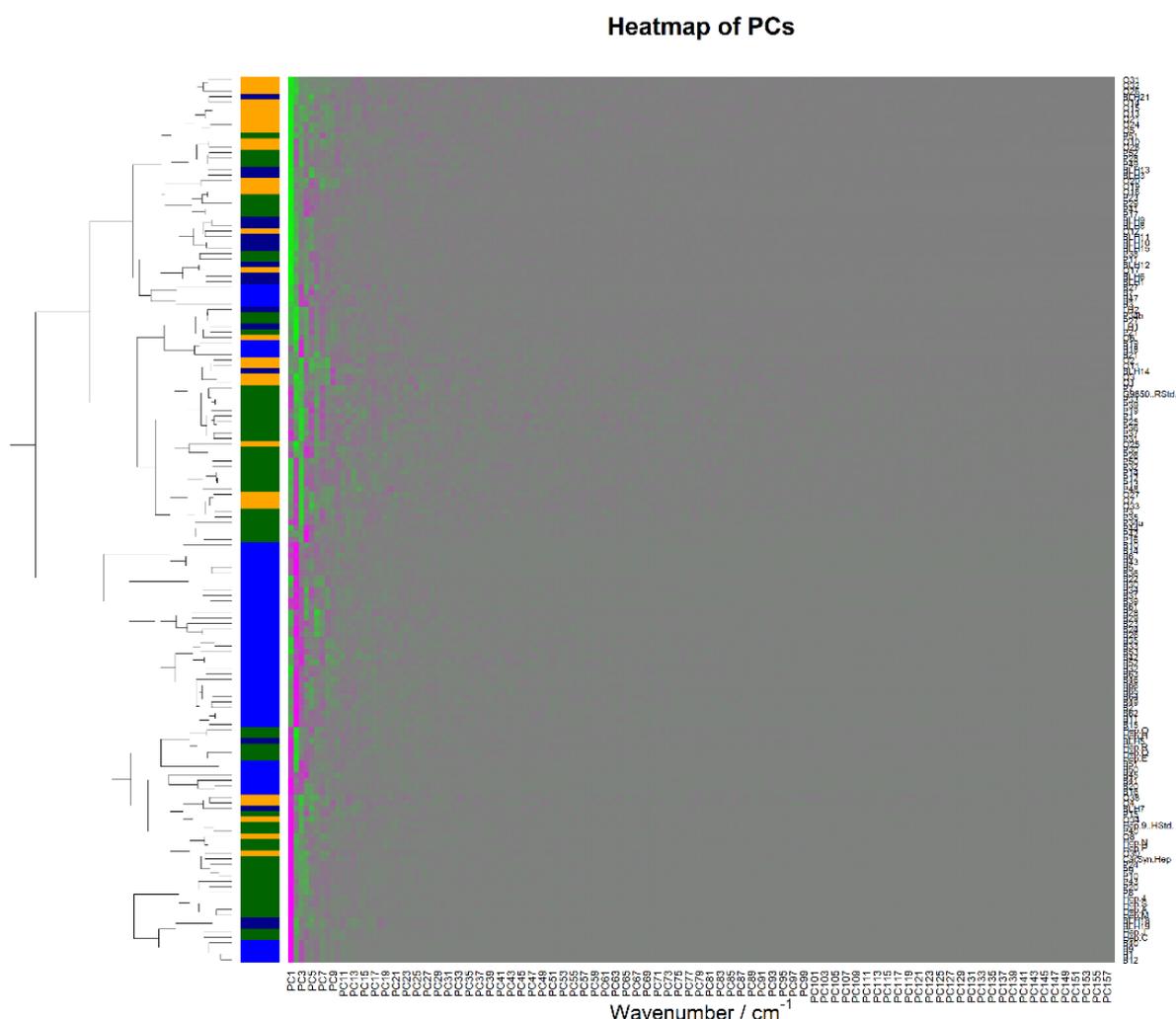
Figure 5.11: Heatmap and subsequent HCA of PCs, generated from second derivative ATR-FTIR spectra of heparins extracted from different animal species. The dendrogram is located to the left of the plot. Between the plot and dendrogram are colour codes for which animal source the sample is from. PMH: green, BMH: blue, BLH: dark blue, OMH: orange. All spectra involved in PCA are in the regions of 700 and 2000 and, 2500 and 3700 cm^{-1} and are averages of 3 scans. The spectra were smoothed, baseline corrected with a 7th order polynomial, normalised (0-1), smoothed and derivatised with a Savitzky-Golay algorithm and finally autoscaled. The PC scores were used as an input for the HCA.
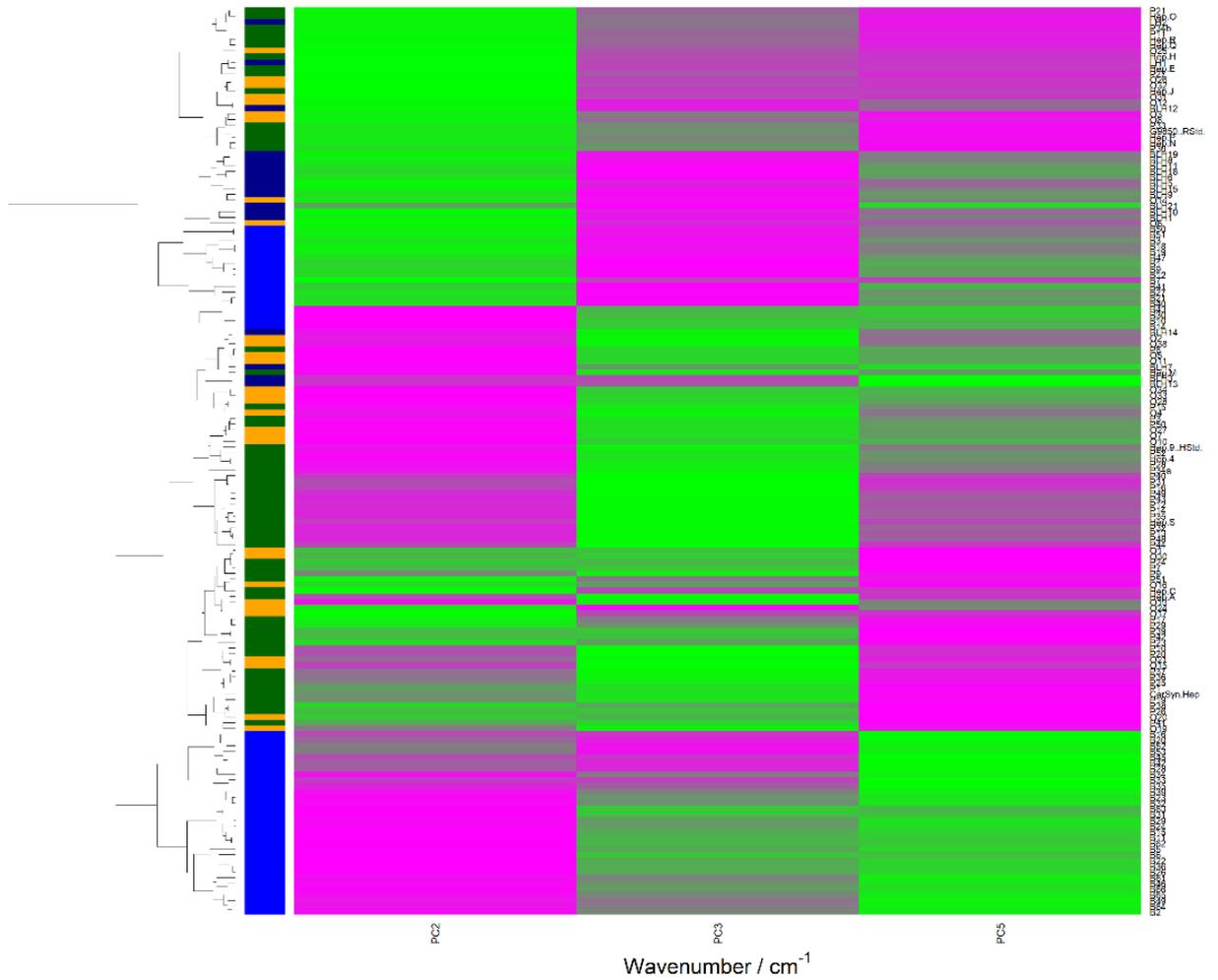
Figure 5.12: Heatmap and subsequent HCA of PCs 2, 3 and 5, generated from second derivative ATR-FTIR spectra of heparins extracted from different animal species. The dendrogram is located to the left of the plot. Between the plot and dendrogram are colour codes for which animal source the sample is from. PMH: green, BMH: blue, BLH: dark blue, OMH: orange. All spectra involved in PCA are in the regions of 700 and 2000 and, 2500 and 3700 cm-1 and are averages of 3 scans. The spectra were smoothed, baseline corrected with a 7th order polynomial, normalised (0-1), smoothed and derivatised with a Savitzky-Golay algorithm and finally autoscaled. The PC scores from PCs 2, 3 and 5 were used as an input for the HCA.
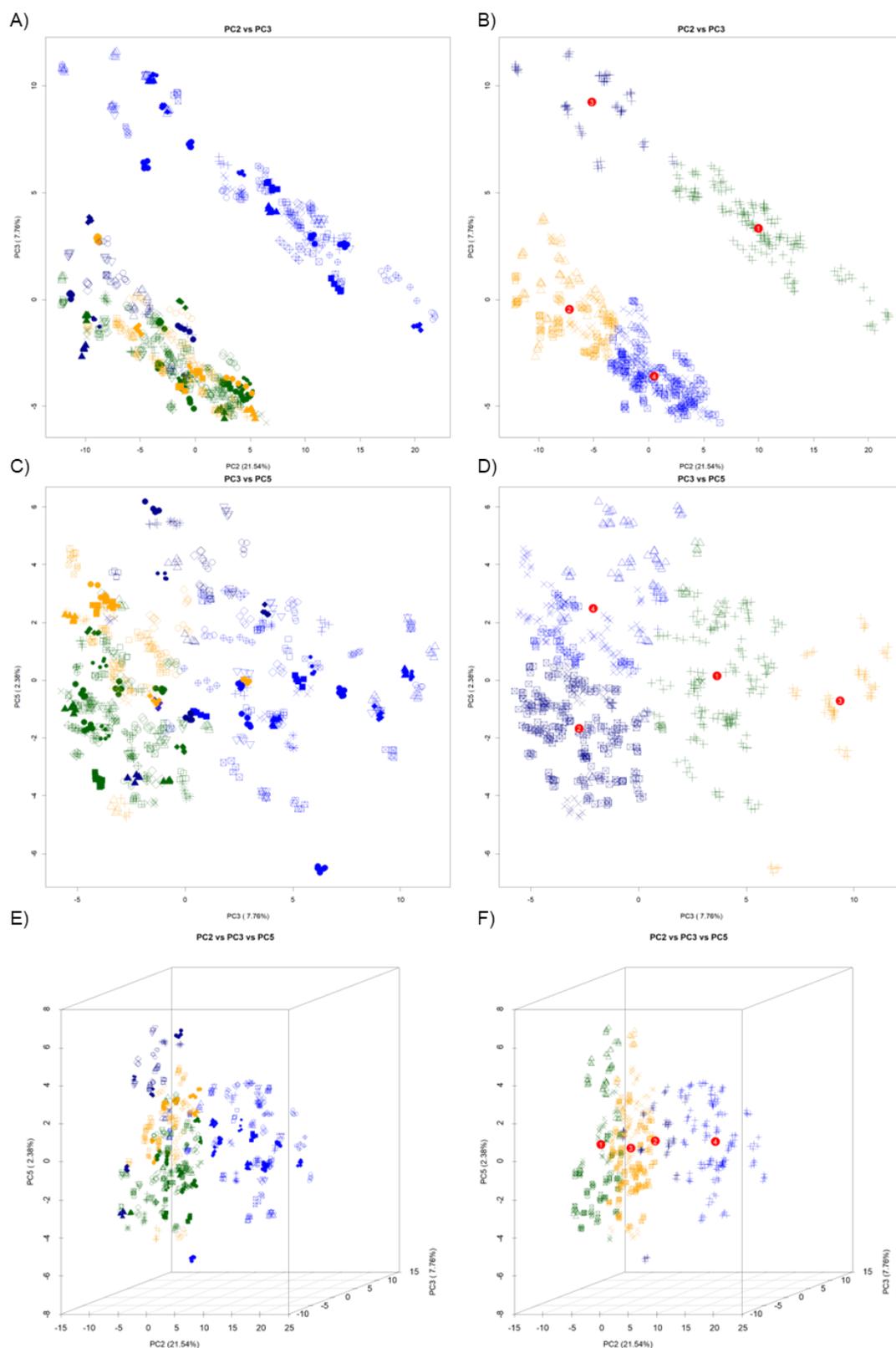
Figure 5.13: PCA scores plots of heparins extracted from different animal sources that have been subjected to cluster analysis with KMC. A,C,D) PC scores plots coloured to indicate the animal of origin of for PCs 2 and 3, PCs 3 and 5 and, PCs 2, 3 and 5 respectively **B,D,F)** Recoloured PC scores plots, representing the clusters computed with KMC for PCs 2 and 3, PCs 3 and 5 and, PCs 2, 3 and 5 respectively. PMH: green, BMH: blue, BLH: dark blue, OMH: orange, KMC cluster means: red. All spectra involved in PCA are in the regions of 700 and 2000 and, 2500 and 3700 cm^{-1} and are averages of 3 scans. The spectra were smoothed, baseline corrected with a 7th order polynomial, normalised (0-1), smoothed and derivatised with a Savitzky-Golay algorithm and finally autoscaled.
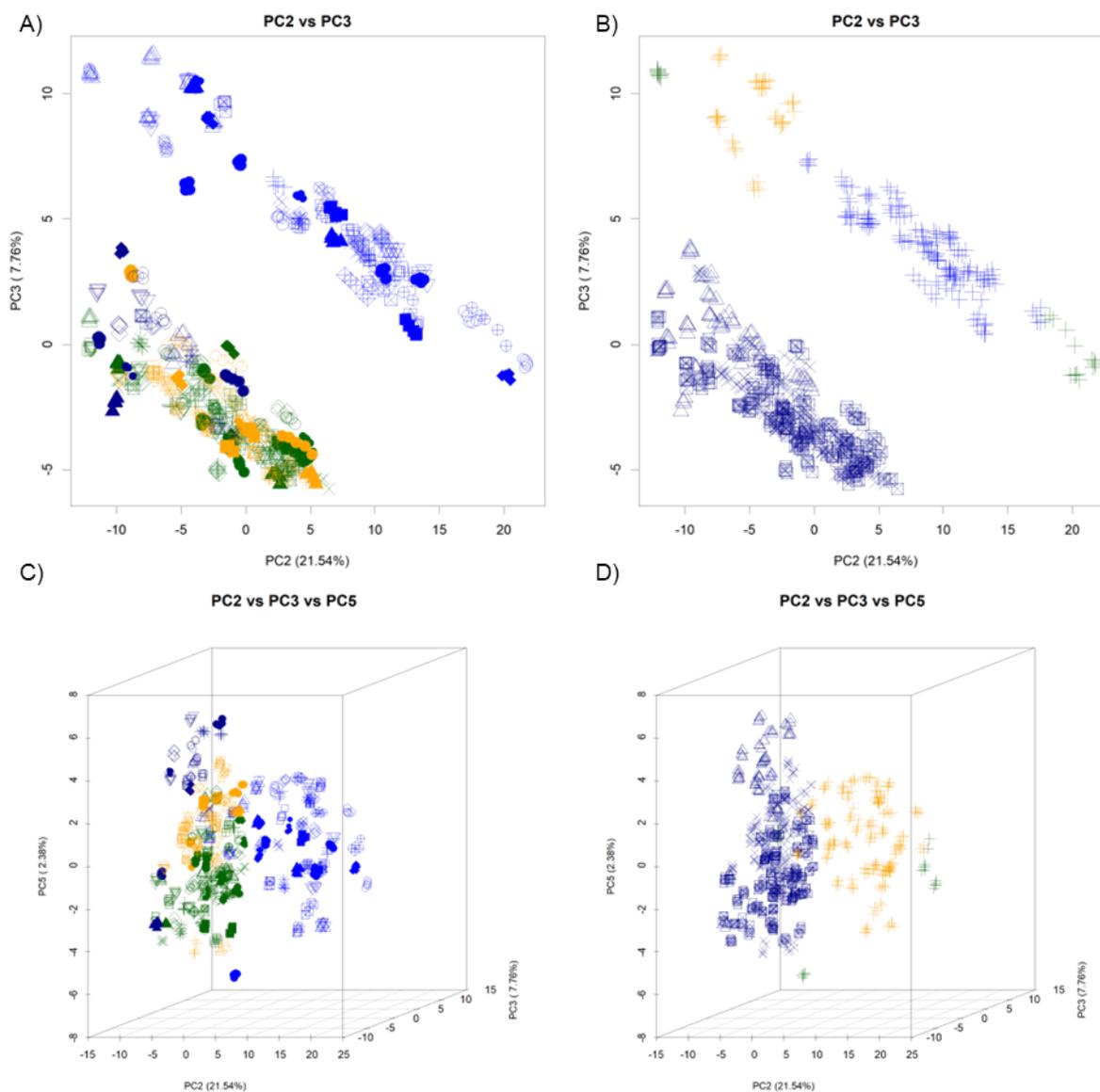
Figure 5.14: PCA scores plots of heparins extracted from different animal sources that have been subjected to cluster analysis with DBSCAN. A,C) PC scores plots coloured to indicate the animal origin of heparins for PCs 2 and 3 and, PCs 2, 3 and 5 respectively **B,D)** Recoloured PC scores plots, representing the clusters computed with DBSCAN for PCs 2 and 3 and, PCs 2, 3 and 5 respectively. PMH: green, BMH: blue, BLH: dark blue, OMH: orange. All spectra involved in PCA are in the regions of 700 and 2000 and, 2500 and 3700 cm^{-1} and are averages of 3 scans. The spectra were smoothed, baseline corrected with a 7th order polynomial, normalised (0-1), smoothed and derivatised with a Savitzky-Golay algorithm and finally autoscaled.
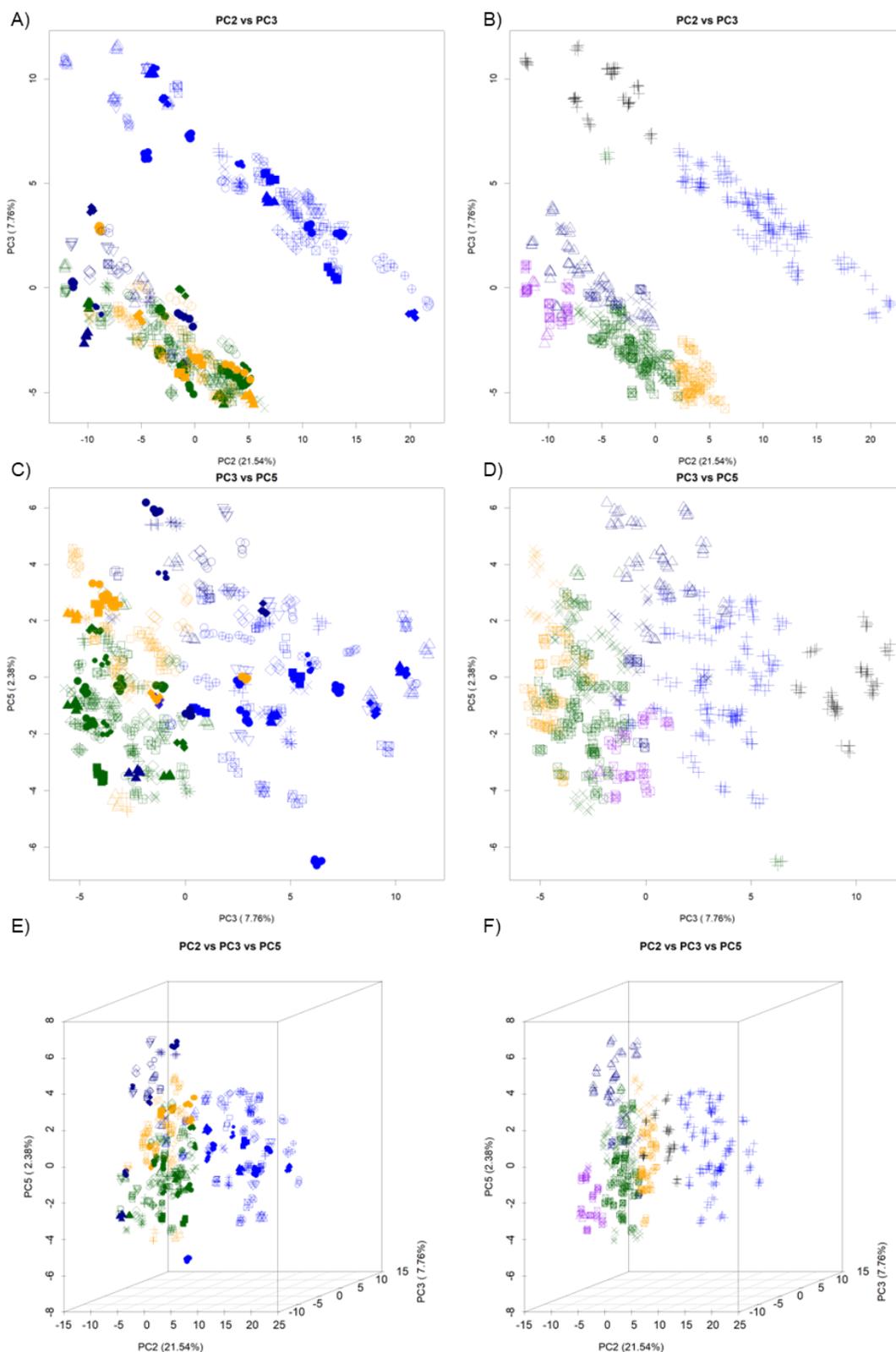
Figure 5.15: PCA scores plots of heparins extracted from different animal sources that have been subjected to cluster analysis with EM. A,C,D) Correctly coloured PC scores plots for heparins extracted from different animal sources for PCs 2 and 3, PCs 3 and 5 and, PCs 2, 3 and 5 respectively **B,D,F)** Recoloured PC scores plots, representing the clusters computed with EM for PCs 2 and 3, PCs 3 and 5 and, PCs 2, 3 and 5 respectively. PMH: green, BMH: blue, BLH: dark blue, OMH: orange. All spectra involved in PCA are in the regions of 700 and 2000 and, 2500 and 3700 cm^{-1} and are averages of 3 scans. The spectra were smoothed, baseline corrected with a 7th order polynomial, normalised (0-1), smoothed and derivatised with a Savitzky-Golay algorithm and finally autoscaled.

[157]
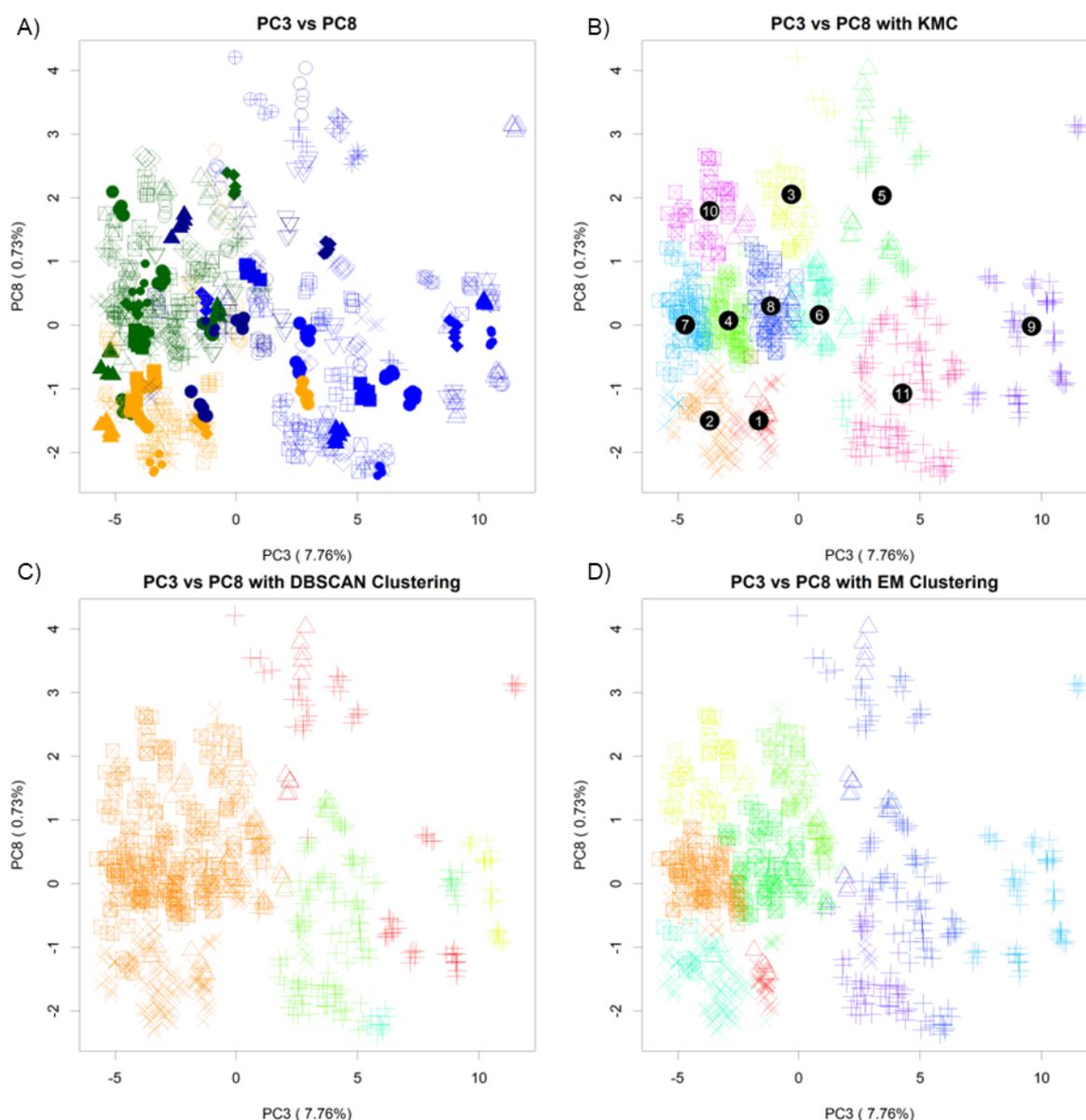
Figure 5.16: Clusters, calculated with KMC, DBSCAN and EM displayed over PC3 and PC8 scores plots compared with the correct clusters for heparins extracted from different animal sources. A) Correctly coloured score plot of PC3 vs PC8, PMH: green, BMH: blue, BLH: dark blue, OMH: orange. **B)** 10 clusters computed with KMC. KMC means: black. **C)** 7 clusters computed with DBSCAN. **D)** 10 clusters computed with EM. All spectra involved in PCA are in the regions of 700 and 2000 and, 2500 and 3700 cm^{-1} and are averages of 3 scans. The spectra were smoothed, baseline corrected with a 7th order polynomial, normalised (0-1), smoothed and derivatised with a Savitzky-Golay algorithm and finally autoscaled.

The lack of clusters which resemble the animal sources of each heparin may have been due to a lack of separation in the PC plots, hence attempts were made to improve the separation and therefore the clustering. t-SNE, another common method of dimensionality reduction, which seeks to move like data close together and unlike data far away, while compressing the entire data set into 2 or 3 dimensions was applied to the autoscaled data second derivative ATR-FTIR spectra in place of PCA using the *tsne* function from the package **Rtsne** (Hinton and Roweis, 2002; Maaten and Hinton, 2008). t-SNE is sometimes applied to PCA when numerous PCs are required to explain the data, to further reduce dimensionality and improve readability. Strong separation between BMH and the other three heparins was observed (**Fig.**

5.17A). It was theorised that the signals associated with concentration were again an issue, as PC1 had a loading plot that corresponded with the concentration band at 960-1007 cm^{-1} hence this band was removed and the t-SNE performed again (**Fig. 5.17B**). BMH formed a band at the bottom, while PMH formed a band across the middle, OMH a lobe atop this band and BLH a lobe atop that lobe. There was still some overlap between the OMH, BLH and PMH however but separation was greatly improved.

The perplexity, a measure of how the t-SNE forms the groupings across a local and global scale, was changed from the standard 30 to 50 as a robust t-SNE should not change significantly between these perplexity values. A rotation between t-SNE1 and t-SNE2 was observed, but the major groups were preserved (**Fig. 5.17C**). The same three clustering algorithms that were used for PCA were performed on the new t-SNE plot (**Fig. 5.17D,E and F**). KMC and EM were both able to separate a BLH cluster, but struggled to differentiate between OMH and PMH. All clustering techniques separated BMH from the rest.

t-SNE was then applied to PCs 2, 3, 5 and 8, the PCs which appeared to separate the different sources of heparin (**Figs. 5.7** and **5.8**) and to the subsequent t-SNE plot, the same clustering techniques applied. The spectra input into the PCA did not have the band associated with concentration removed, but instead PC1 which appeared to correlate with concentration, was removed from the analyses. BMH was again separated from the rest of the heparins, and the separation between OMH and PMH improved (**Fig. 5.18**). BLH moved mostly into its own cluster, close to BMH but still showed overlap with PMH and OMH. The same three clustering techniques as before were applied across the t-SNE plot and, while all the clustering algorithms located a dense BLH region, there was some separation between OMH and PMH with KMC and EM, most likely due to the improved separation in the t-SNE plot (**Fig. 5.18B,C and D**).

With t-SNE, new data is unable to be predicted onto the 2-dimensional space without use of an algorithm that is unique to each t-SNE plot (Hinton and Roweis, 2002). Since the separation observed with t-SNE was similar to that of PCA, but would be computationally challenging to predict from, t-SNE was not taken forward.

[159]
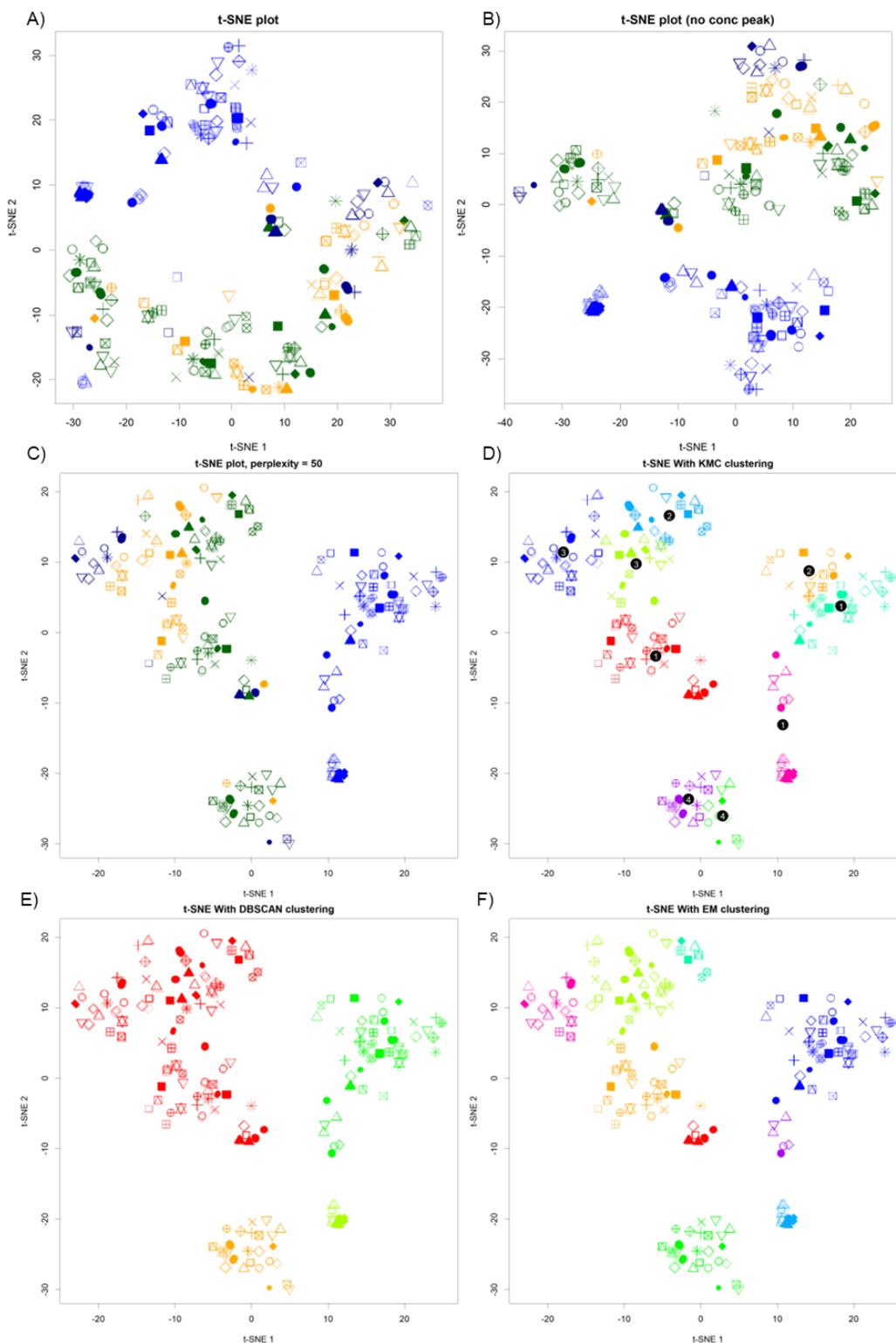
Figure 5.17: t-SNE analysis of ATR-FTIR spectra of heparins from different animal sources and subsequent cluster analysis. A) t-SNE plot of ATR-FTIR spectra in the regions of 700 to 2000 cm-1 and, 2500 to 3700 cm-1. Perplexity = 30 **B)** t-SNE plot of ATR-FTIR spectra in the regions of 700 to 960, 1007 to 2000 and, 2500 to 3700cm-1 i.e, spectra with the band associated with concentration removed. Perplexity = 30. **C)** t-SNE analysis with no band associated with concentration, and a perplexity of 50. **D)** Cluster analysis of the t-SNE plot from panel C using KMC. black: the computed KMC means. **E)** Cluster analysis of the t-SNE plot from panel C using DBSCAN. **F)** Cluster analysis of the t-SNE plot from panel C using EM. PMH: green, BMH: blue, BLH: dark blue, OMH: orange. On panels in which cluster analysis has been undertaken, the clusters are indicated with colours of the rainbow. Unless otherwise stated, the perplexity is set to 50. All spectra involved in t-SNE are averages of 3 scans. The spectra were smoothed, baseline corrected with a 7th order polynomial, normalised (0-1), smoothed and derivatised with a Savitzky-Golay algorithm and finally autoscaled.

[160]
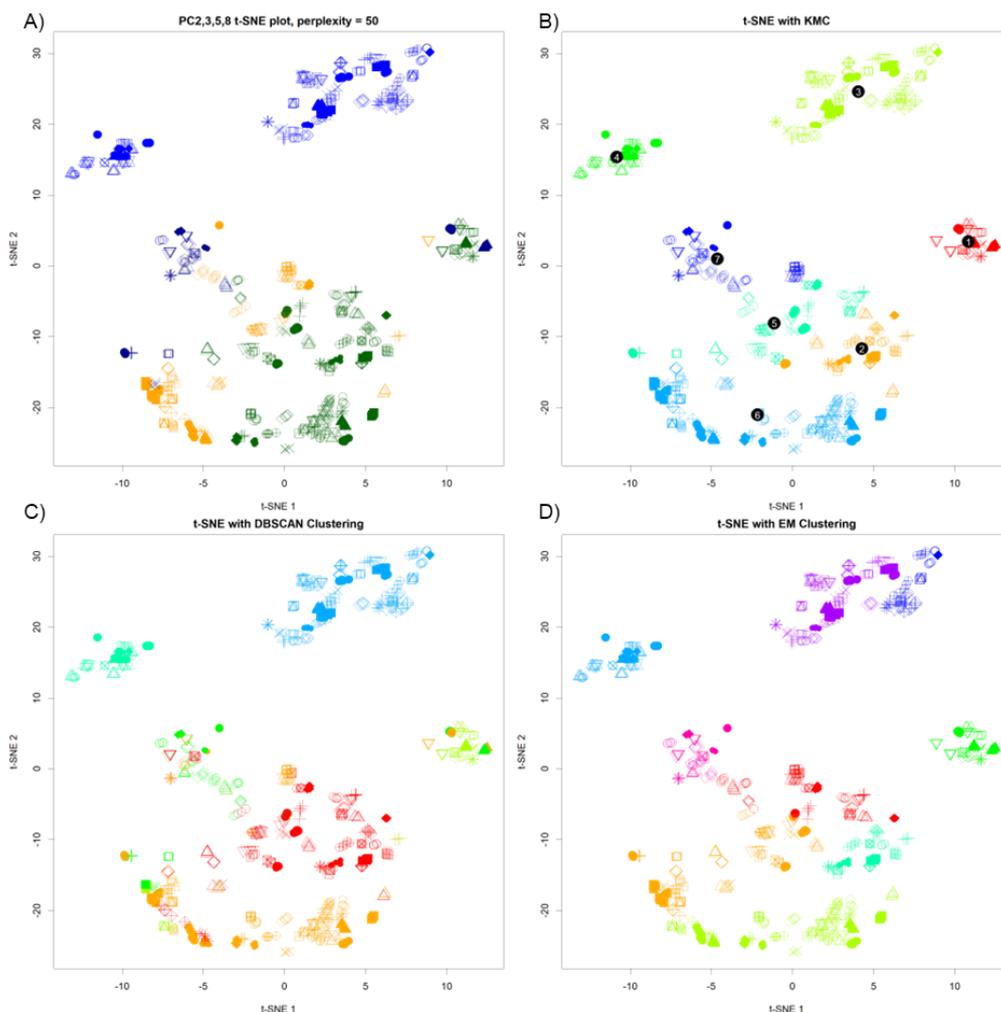
Figure 5.18: t-SNE analysis of the scores of PCs 2, 3, 5 and 8, calculated through PCA of ATR-FTIR spectra of heparins from different animal sources and subsequent cluster analysis. A) Correctly coloured t-SNE plot for heparins from different animal sources, PMH: green, BMH: blue, BLH: dark blue, OMH: orange. **B)** Cluster analysis of the t-SNE plot from panel C using KMC. black: the computed KMC means. **C)** Cluster analysis of the t-SNE plot from panel C using DBSCAN. **D)** Cluster analysis of the t-SNE plot from panel C using EM. Plots which have undergone cluster analysis have computed clusters coloured with colours of the rainbow. All spectra involved in PCA and subsequently in t-SNE are in the regions of 700 to 960, 1007 to 2000 and, 2500 to 3700cm⁻¹ and are averages of 3 scans. The spectra were smoothed, baseline corrected with a 7th order polynomial, normalised (0-1), smoothed and derivatised with a Savitzky-Golay algorithm and finally autoscaled.

5.3.6. The application of trained cluster analyses to detect and identify heparins from different animal species.

The lack of ability to cluster the data by animal source was probably because the clustering techniques used are primarily used for the prediction of *possible* clusters within the data. Since the sources of heparins present here are known *a priori,* a slightly different clustering technique - KMC, with precomputed centroids - was employed. In brief, the means for each heparin type in each selected component were calculated and used as the centroids for the clusters. Samples were classified based on their Euclidian distance from each mean with the lowest distance meaning the highest probability of being a part of that cluster. The Euclidian distances were subjected to a SoftMax transformation, computing the subsequent probability that each sample is in the given cluster. For quick classification, the highest probability, regardless of its closeness to other clusters, was assumed to be the cluster the sample

belonged to, but incorrect classifications and edge cases can be examined with these probabilities. Like EM, this algorithm is an example of a soft classification technique.

Table 5.6: Confusion matrix for the predicted identities of heparins from different animal sources, using PCA-KMC-CC.

		Confusion Matrix					Statistics			
		Reference					Accuracy	Precision	Specificity	Chi squared P-value
		BLH	BMH	OMH	PMH	Sum				
Predicted	BLH	77	0	21	17	115	0.86	0.67	0.94	< 2.2x10-16
	BMH	0	244	15	15	274	0.97	0.89	0.94	
	OMH	13	0	73	36	122	0.72	0.60	0.92	
	PMH	9	1	34	234	278	0.84	0.84	0.91	
	Sum	99	245	143	302	789	0.85	0.75	0.93	

To test the ability of PCA-KMC-CC's to predict heparin source, the dataset used to formulate the original PCA (**Fig. 5.8**) was taken, and 1/3rd of the data, picked at random without replacement, was removed to become the test data. The remaining 2/3rds were taken, and a PCA performed on them. To this PCA model, the test samples were predicted using KMC-CC of PCs 2, 3, 5, 7 and 8 – the components which appear to discriminate the sources of each heparin. The entire process was repeated 3 times and the predicted values collected. The predicted values were compared to the expected values with Chi^2 and a confusion matrix was created (**Table 5.6**). The observed identities were not statistically different from the expected identities (P-value $< 2.2x10^{-16}$) and were predicted with reasonable accuracy (average accuracy = 85%). Prediction of OMH and BLH was not precise with precisions of 67% and 60% respectively. OMH and BLH were predicted falsely as either each other or PMH. Prediction of PMH and BMH was more precise however at 84% and 89% respectively. As observed in the PC plots, BLH, OMH and PMH overlap significantly, but are each distinct from BMH, this may explain the lack of separation.

The overall accuracy of these predictions can be increased through selected use of relevant components (**Table 5.7**). Using components 1 through 5 provides high accuracy for BMH (99%), but poor accuracy and precision for BLH and OMH (80% and 59%). Using components 1 through 9, the average accuracy is increased to 81% from 80%, but the discrimination of BMH is weakened from 99% accuracy to 88%. When PCs 2, 3, 5, 7 and 8 are used, the accuracy was increased markedly in all cases to an average of 87%. The precision of BLH remains low at 64%, but increases greatly for BMH to 84% from 70%, taking the average precision for all four heparins to 82% (from 67% at its lowest). The 5 chosen components also remain consistent across all iterations of PCA (tested for n = 9) (**Fig.5.19**).

Table 5.7: Confusion matrices for the predicted identities of heparins from different animal sources, using PCA-KMC-CC with specific PCs. First matrix: PCs 1 through 5 are used for classification. **Second Matrix:** PCs 1 through 9 are used for classification. **Third Matrix:** PCs 2, 3, 5, 7 and 8 are used for classification.

	Confusion Matrix						Statistics			
PC1:5		**Reference**					**Accuracy**	**Precision**	**Specificity**	**Chi squared P-value**
		BLH	**BMH**	**OMH**	**PMH**	**Sum**				
Predicted	**BLH**	22	0	13	8	43	0.80	0.51	0.91	< 2.2x10-16
	BMH	0	93	3	2	98	0.99	0.95	0.97	
	OMH	6	0	11	8	25	0.59	0.44	0.94	
	PMH	4	0	17	76	97	0.84	0.78	0.88	
	Sum	32	93	44	94	263	0.80	0.67	0.92	

	Confusion Matrix						Statistics			
PC1:9		**Reference**					**Accuracy**	**Precision**	**Specificity**	**Chi squared P-value**
		BLH	**BMH**	**OMH**	**PMH**	**Sum**				
Predicted	**BLH**	21	0	9	3	33	0.80	0.64	0.95	< 2.2x10-16
	BMH	1	93	9	30	133	0.88	0.70	0.77	
	OMH	4	0	24	0	28	0.76	0.86	0.98	
	PMH	6	0	2	61	69	0.80	0.88	0.95	
	Sum	32	93	44	94	263	0.81	0.77	0.91	

	Confusion Matrix						Statistics			
PC2,3,5,7,8		**Reference**					**Accuracy**	**Precision**	**Specificity**	**Chi squared P-value**
		BLH	**BMH**	**OMH**	**PMH**	**Sum**				
Predicted	**BLH**	21	0	6	6	33	0.80	0.64	0.95	< 2.2x10-16
	BMH	1	93	5	12	111	0.95	0.84	0.89	
	OMH	4	0	32	0	36	0.85	0.89	0.98	
	PMH	6	0	1	76	83	0.88	0.92	0.96	
	Sum	32	93	44	94	263	0.87	0.82	0.95	

With the success of the KMC-CC method, another, widely used trained technique was applied to the PCA plots. K-nearest neighbours (KNN) is a machine learning algorithm with close ties to KMC. Unlike KMC, KNN is trained on a matrix of data with known clusters. New inputs are classified whereby the most common class of the K data points (K=4 in these examples) with the smallest Euclidian distance is ascribed to the test data (Altman, 1992). The KMC-CC and KNN methods were applied, using PCs 2, 3, 5, 7 and 8 to another three iterations of sample testing and the outputs examined in **Table. 5.8.** KMC-CC with PCs 2, 3, 5 and 8 had an accuracy of 89% but the precision for BLH was still low however, falsely identifying BLHs and PMHs. KNN improved the average accuracy and precision to 98% and 97% respectively from 89% and 80% with KMC-CC. The precision of BLH was improved to 96% from 54%, with only 3 samples falsely identified as PMH, down from 41 with KMC-CC.
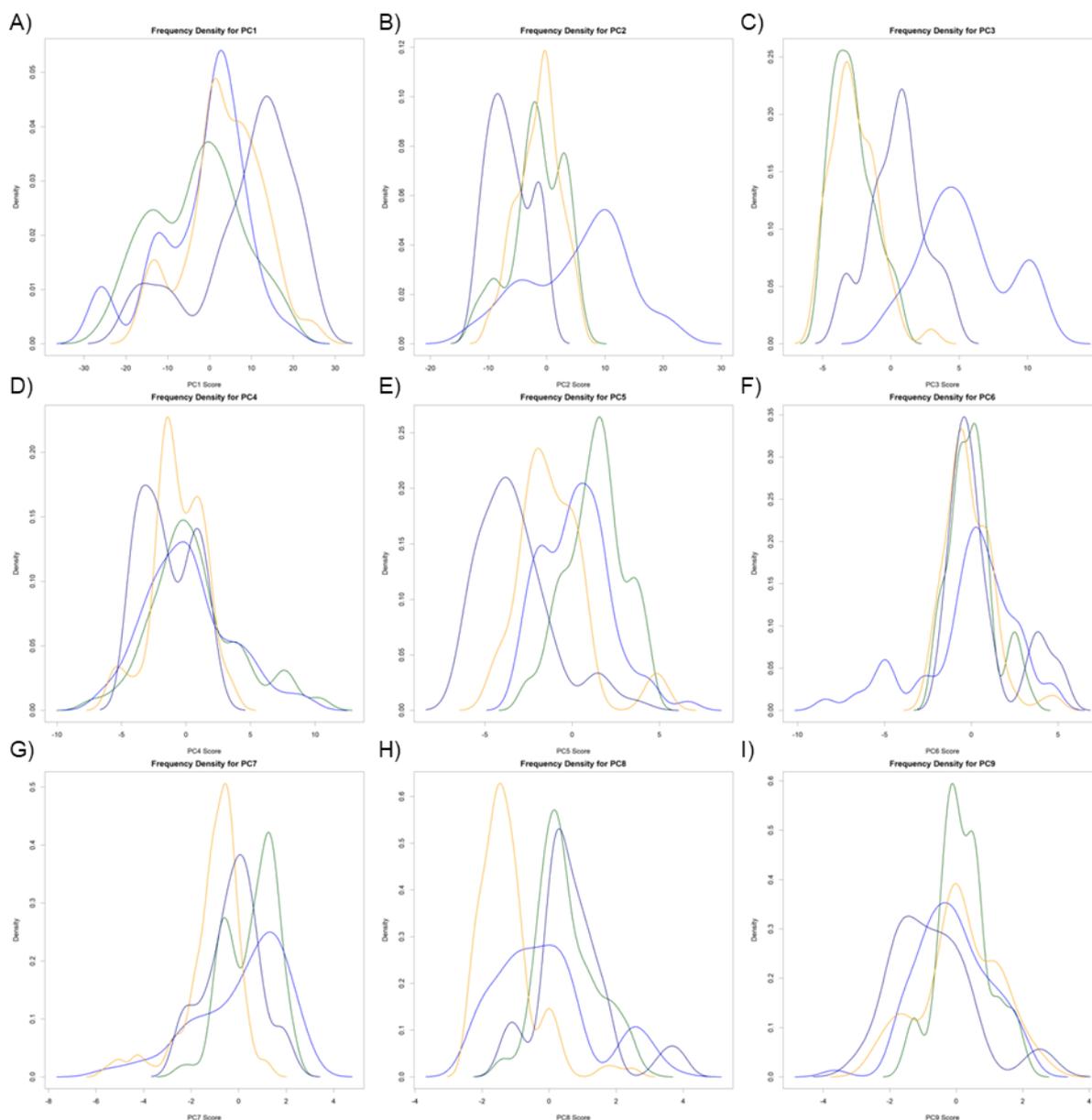
Figure 5.19: Frequency densities of PMH, BMH, BLH and OMH in different PCs, after a test library has been removed, indicating the stability of the PCs in each iteration. A through **I)** correspond to the frequency densities for each heparin in PCs 1 through 9 respectively. PMH: green, BMH: blue, BLH: dark blue, OMH: orange.

Two other, trained techniques were applied to the spectra. The first was PLS-discriminant analysis (PLS-DA); essentially PLS as outlined above but for the creation of linear combinations that facilitate the discrimination of different sample types. The PLS-DA produced score plots with clear and distinct lobes with minimal overlap across all heparin types (**Fig.5.20**). The PLS loading plots show different features selected for when compared to PCA (**Fig. 5.20D,E and F)**. PCA has higher loadings in the main carbohydrate band at ~1000 cm^{-1} (**Fig. 5.8G,H,I and J**) with some loadings in other regions, but the PLS loading plots are disperse across the entire spectrum.

Table 5.8: Confusion matrices for the predicted identities of heparins from different animal sources, using KMC-CC and KNN on PCs 2,3,5,7 and 8. First matrix: classification using KMC-CC. **Second matrix:** classification using KNN.

		Confusion Matrix					Statistics			
KMC-CC		**Reference**					**Accuracy**	**Precision**	**Specificity**	**Chi squared P-value**
		BLH	**BMH**	**OMH**	**PMH**	**Sum**				
Predicted	**BLH**	65	0	15	41	121	0.84	0.54	0.72	< 2.2x10-16
	BMH	4	246	4	25	279	0.97	0.88	1.00	
	OMH	11	0	140	16	167	0.90	0.84	0.85	
	PMH	6	1	4	211	222	0.85	0.95	0.72	
	Sum	86	247	163	293	789	0.89	0.80	0.82	

		Confusion Matrix					Statistics			
KNN		**Reference**					**Accuracy**	**Precision**	**Specificity**	**Chi squared P-value**
		BLH	**BMH**	**OMH**	**PMH**	**Sum**				
Predicted	**BLH**	80	0	0	3	83	0.96	0.96	0.98	< 2.2x10-16
	BMH	0	246	0	0	246	1.00	1.00	1.00	
	OMH	6	0	159	4	169	0.98	0.94	0.97	
	PMH	0	1	4	286	291	0.98	0.98	0.98	
	Sum	86	247	163	293	789	0.98	0.97	0.98	
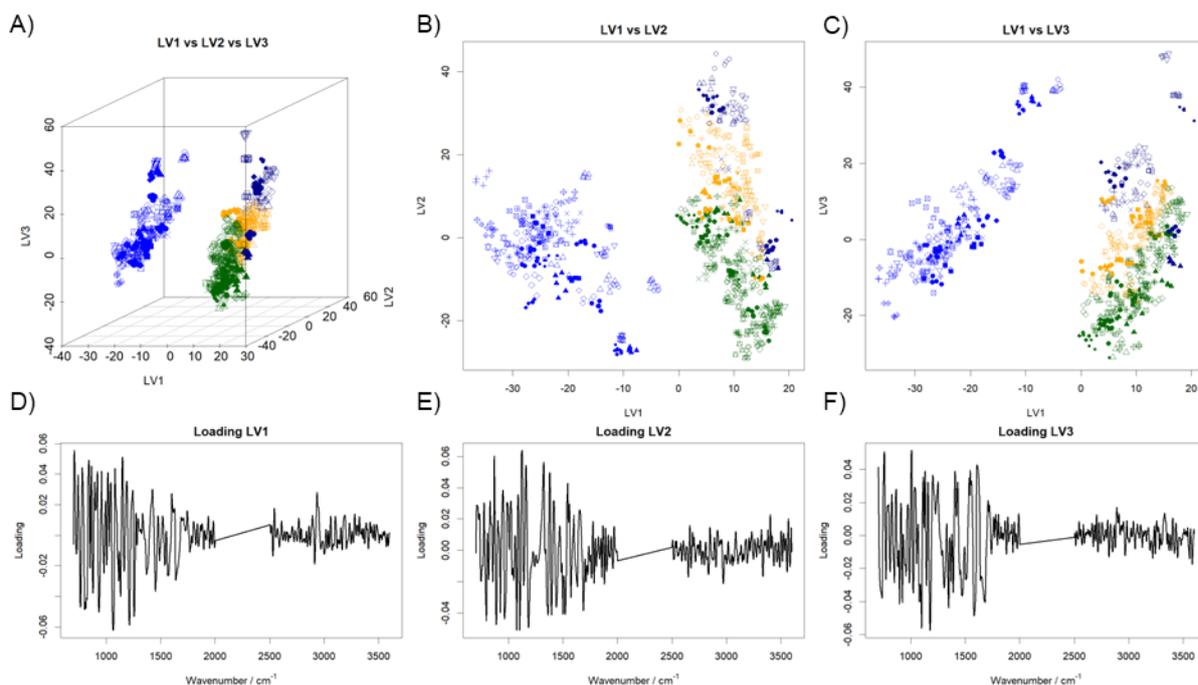
Figure 5.20: PLS-DA of ATR-FTIR spectra of heparins extracted from different animal sources. **A)** 3D scores plots for LV1 vs LV2 vs LV3. **B)** Scores plots for LV1 vs LV2. **C)** Scores plots for LV1 vs LV3. **D)** Loading plot for LV1. **E)** Loading plot for LV2. **F)** Loading plot for LV3. All spectra involved in PLS-DA are in the regions of 700 and 2000 and, 2500 and 3700 cm-1 and are the averages of 3 scans. The spectra were smoothed, baseline corrected with a 7th order polynomial, normalised (0-1), smoothed and derivatised with a Savitzky-Golay algorithm and finally autoscaled.

The ability of PLS-DA to discriminate heparin species was validated as outlined previously for PCA-KMC-CC, using both KMC-CC and KNN of the generated scores for the first 4 LVs (**Table**

5.9). The predicted examples were not statistically different to the expected values (P-value < 2.2×10^{-16}) and were predicted with high accuracy (average accuracy = 98% for CC clustering and 99% for KNN). The predictions were also more precise and more specific than the PCA clustered samples. Both KMC-CC clustering and KNN approached 100% correct prediction with very few false positives (13 and 4 for KMC-CC and KNN respectively) and false negatives (3 for both techniques).

Table 5.9: Confusion matrices for the predicted identities of heparins from different animal sources, using PLS-DA and PLS-DA-KNN. First matrix: classification using PLS-DA. **Second matrix:** classification using PLS-DA-KNN.

		Confusion Matrix					Statistics			
PLS-DA		Reference					Accuracy	Precision	Specificity	Chi squared P-value
		BLH	BMH	OMH	PMH	Sum				
Predicted	**BLH**	70	0	2	0	72	0.95	0.97	1.00	< 2.2x10-16
	BMH	0	251	1	0	252	1.00	1.00	1.00	
	OMH	1	0	145	0	146	0.98	0.99	1.00	
	PMH	7	2	3	307	319	0.99	0.96	0.98	
	Sum	78	253	151	307	789	0.98	0.98	0.99	
		Confusion Matrix					Statistics			
PLS-DA-KNN		Reference					Accuracy	Precision	Specificity	Chi squared P-value
		BLH	BMH	OMH	PMH	Sum				
Predicted	**BLH**	77	0	1	1	79	0.99	0.97	1.00	< 2.2x10-16
	BMH	0	251	1	0	252	1.00	1.00	1.00	
	OMH	1	0	148	0	149	0.99	0.99	1.00	
	PMH	0	2	1	306	309	1.00	0.99	0.99	
	Sum	78	253	151	307	789	0.99	0.99	1.00	

A second technique, linear discriminant analysis (LDA) was also employed. LDA, like PCA and PLS, seeks to create a linear combination of variables that best describes the given data (Martinez and Kak, 2001). Unlike PCA and PLS, LDA calculates linear combinations in a manner which maximises the distance between gaussian distributions which best describe each group present within the data (Cohen et al., 2002). As with PCA and PLS, scores plots and loadings can be graphed with LDA (**Fig. 5.21**). The scores plots showed very tight groupings for each source of heparin. LDA provided unparalleled identification of the different sources of heparin; of the 789 samples tested, a total of 6 were misclassified with LDA, providing an accuracy of 99%. For the first time, all BLH samples were correctly discriminated (**Table 5.10**).

Table 5.10: Confusion matrices for the predicted identities of heparins from different animal sources, using LDA.

		Confusion Matrix					Statistics			
LDA		**Reference**					**Accuracy**	**Precision**	**Specificity**	**Chi squared P-value**
		BLH	**BMH**	**OMH**	**PMH**	**Sum**				
Predicted	**BLH**	86	0	0	0	86	1.00	1.00	1.00	< 2.2x10-16
	BMH	0	242	3	1	246	1.00	0.98	0.99	
	OMH	0	0	147	2	149	0.99	0.99	1.00	
	PMH	0	0	0	308	308	1.00	1.00	1.00	
	Sum	86	242	150	311	789	0.99	0.99	1.00	
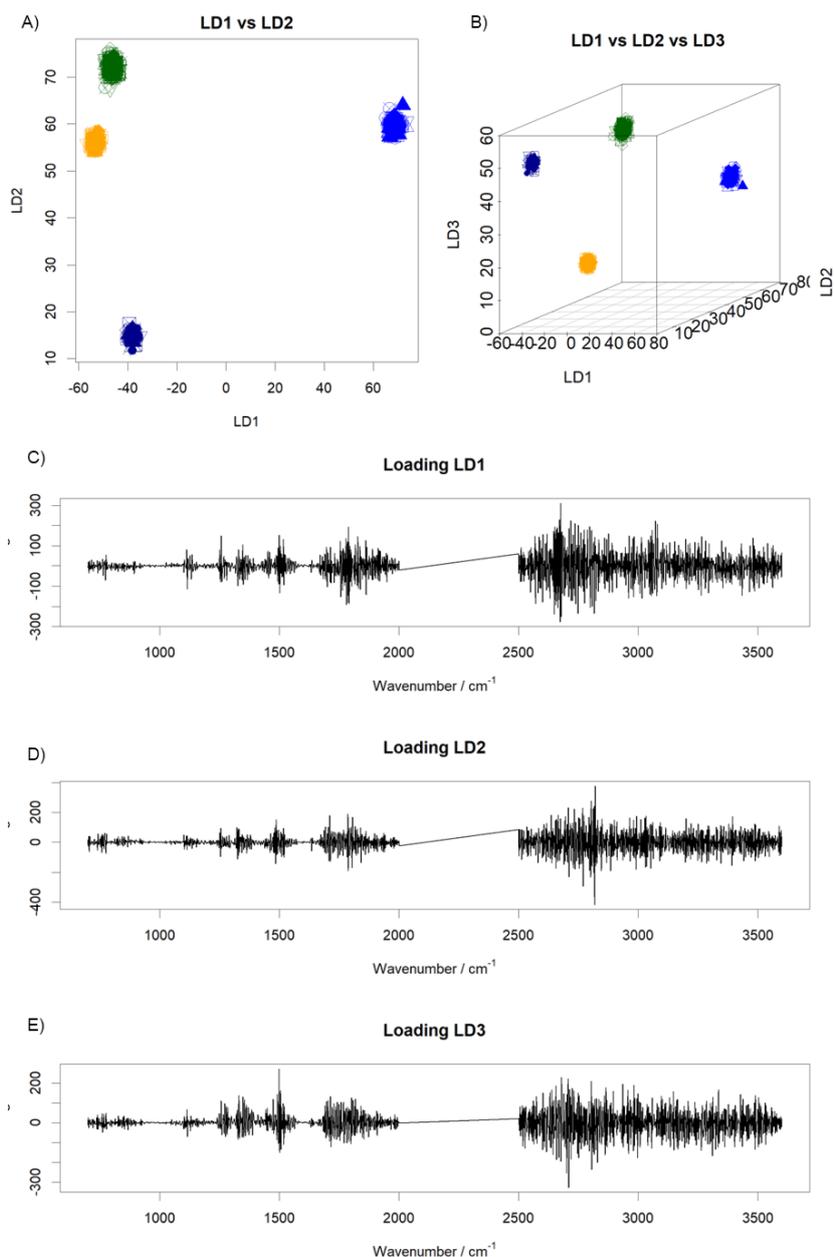
Figure 5.21: LDA of ATR-FTIR spectra of heparins extracted from different animal sources. A) Scores plot for LD1 vs LD2. **B)** 3D score plots for LD1 vs LD2 vs LD3. **C)** Loading plot for LD1. **D)** Loading plot for LD2. **E)** Loading plot for LD3. All spectra involved in LDA are in the regions of 700 and 2000 and, 2500 and 3700 cm^{-1} and are the averages of 3 scans. The spectra were smoothed, baseline corrected with a 7th order polynomial, normalised (0-1), smoothed and derivatised with a Savitzky-Golay algorithm and finally autoscaled.

FTIR-ATR has been shown to be able to detect heparin contaminated with OSCS, heparin blended with other GAGs, characterise the levels of different sulphates in crude heparin and by extension, discriminate heparin from different sources. One potential future area of contamination is that of contamination of heparin with heparins from other animal sources. This may be a particularly serious issue in Muslim and Jewish countries, where contamination with pork products is of serious concern. Such issues reciprocate to the contamination of pork products with beef products, potentiating fears of spreading BSE in the UK. To explore this, three randomly selected PMHs was contaminated with three randomly selected BMHs (and by extension, BMH with PMH) at the levels of 99, 90, 80, 70, 60, 50, 40, 30, 20, 10, 1 % (w/w). The resulting series of samples had their spectra recorded and were subjected to PCA and subsequent PCR and PLS-R analysis. The PCR and PLS regression models were constructed using the levels of PMH and BMH in all samples (including zeroes and hundreds for the PMH and BMH libraries respectively). A final contamination series, which had not been used to create the library, was then screened for the levels of PMH and subsequently BMH present.

Table 5.11: PCR and PLS-R predictions for the levels of PMH and BMH and PMH:BMH blends.

Expected PMH Level / %	Predicted PMH Level / %			PMH Level in PMH library / %	
	PCR 4 PCs	PLS - 5 LVs		PCR 4 PCs	PLS - 5 LVs
100.00	96.74 +/- 1.8	94.93 +/- 0.82	*Average*	97.4 +/- 8.06	97.51 +/- 6.99
99.00	99.45 +/- 0.38	95.43 +/- 1.07	*Min*	73.40	73.95
90.00	82.8 +/- 1.47	82.64 +/- 1.18	*Max*	118.82	115.05
80.00	66.28 +/- 0.91	68.61 +/- 1.2			
70.00	57.34 +/- 1.48	58.86 +/- 1.49		**BMH Level in BMH library / %**	
60.00	52.42 +/- 3.91	50.65 +/- 4.37		PCR 4 PCs	PLS - 5 LVs
50.00	59.57 +/- 3.25	47.05 +/- 3.5	*Average*	96.88 +/- 9.3	96.5 +/- 9.41
40.00	50.55 +/- 2.54	41.45 +/- 2.84	*Min*	69.76	72.84
30.00	25.84 +/- 2.83	24.67 +/- 3.02	*Max*	115.74	120.33
20.00	5.77 +/- 2.4	-1.72 +/- 2.52			
10.00	1.47 +/- 1.46	-0.07 +/- 2.08			
1.00	-10.44 +/- 4.23	-6.81 +/- 3.39			
0.00	13.18 +/- 3.83	12.98 +/- 4.16			
RMSE	9.63	9.65			
R2	0.98	0.97			
P-value	0.78	0.66			

Table 5.12.1: PCR and PLS-R predictions for the levels of PMH and BMH in PMH:BMH, PMH:BLH and PMH:OMH blends.

PMH level / %			BMH level / %		
Expected	Predicted		Expected	Predicted	
	PCR	PLS		PCR	PLS
100.00	90.05 +/- 8.47	102.7 +/- 6.89	0.00	-6.71 +/- 5.08	-5.1 +/- 6.35
99.00	92.2 +/- 4.48	102.64 +/- 3.31	1.00	1.57 +/- 4.33	1.82 +/- 4.12
90.00	84.57 +/- 3.55	93.54 +/- 3.71	10.00	8.2 +/- 1.6	8.82 +/- 1.1
80.00	69.88 +/- 9.46	89.04 +/- 13.13	20.00	13.08 +/- 5.13	13.85 +/- 5.05
70.00	67.49 +/- 3.1	76.49 +/- 2.64	30.00	29.01 +/- 2.49	29.31 +/- 2.53
60.00	46.4 +/- 5.32	52.57 +/- 3.59	40.00	35.37 +/- 2.76	35.79 +/- 2.49
50.00	34.68 +/- 7.58	46.59 +/- 9.21	50.00	43.9 +/- 4.4	45.41 +/- 4.18
40.00	35.21 +/- 8.24	40.23 +/- 7.71	60.00	52.28 +/- 2.61	51.87 +/- 2.42
30.00	23.8 +/- 5.99	28.44 +/- 5.21	70.00	63.04 +/- 2.78	63.35 +/- 2.93
20.00	9.4 +/- 11.51	12.09 +/- 11.3	80.00	62.12 +/- 3.58	62.24 +/- 3.81
10.00	13.8 +/- 7.15	15.21 +/- 6.78	90.00	76.57 +/- 3.28	76.74 +/- 3.3
1.00	11.47 +/- 1.63	10.01 +/- 2.76	99.00	83.11 +/- 3.61	83.04 +/- 2.91
0.00	7.19 +/- 1.86	4.15 +/- 1.88	100.00	87.9 +/- 2.79	87.48 +/- 2.85
99.00	96.29 +/- 2.37	105.88 +/- 2.49	0.00	-12.55 +/- 1.91	-10.79 +/- 1.99
90.00	83.72 +/- 3.53	87.58 +/- 1.29	0.00	-3.77 +/- 1.07	-2.91 +/- 1.44
80.00	66.92 +/- 10.4	70.64 +/- 7.6	0.00	-7.47 +/- 1.61	-7.19 +/- 1.86
70.00	60.97 +/- 7.83	68.27 +/- 9.53	0.00	-3.4 +/- 1.99	-5.01 +/- 2.38
60.00	53.67 +/- 5.48	56.24 +/- 3.47	0.00	1.14 +/- 2.14	-0.2 +/- 1.8
50.00	54.63 +/- 3.77	57.23 +/- 3.62	0.00	0.35 +/- 2.84	-0.7 +/- 3.04
40.00	47.07 +/- 2.99	46.52 +/- 4.78	0.00	-0.45 +/- 4.23	-1.49 +/- 3.89
30.00	37.65 +/- 7.31	38.39 +/- 5.38	0.00	-3.01 +/- 3.63	-4.73 +/- 3.69
20.00	28.05 +/- 2.51	29.29 +/- 2.53	0.00	3.95 +/- 1.94	4.11 +/- 1.93
10.00	25.16 +/- 3.63	24.77 +/- 3.73	0.00	-2.44 +/- 2	-2.78 +/- 2.18
1.00	22.62 +/- 5.45	19.78 +/- 5.4	0.00	-2.98 +/- 2.44	-3.3 +/- 2.56
0.00	13.93 +/- 1.37	13.43 +/- 2.05	0.00	-1.07 +/- 0.96	-1.16 +/- 0.98
100.00	84.91 +/- 3.63	91.21 +/- 2.28	0.00	2.96 +/- 0.78	3.49 +/- 0.7
99.00	88.93 +/- 3.17	95.64 +/- 2.67	0.00	-2.98 +/- 1.23	-3.74 +/- 1.4
90.00	78.18 +/- 10.34	89.51 +/- 11.97	0.00	-0.07 +/- 5.31	-1.25 +/- 4.84
80.00	70.97 +/- 6.13	75.1 +/- 3.29	0.00	1.72 +/- 3.59	1.78 +/- 3.91
70.00	59.23 +/- 9.06	61.82 +/- 7.57	0.00	3.44 +/- 2.05	2.86 +/- 2.11
60.00	58.11 +/- 6.63	60.6 +/- 9.8	0.00	6.02 +/- 2.75	6.66 +/- 3.4
50.00	59.73 +/- 8.14	56.53 +/- 7.7	0.00	4.53 +/- 3.97	3.18 +/- 3.57
40.00	45.63 +/- 3.99	40.84 +/- 3.22	0.00	7.96 +/- 1.56	7.62 +/- 1.64
30.00	41.47 +/- 7.67	34.16 +/- 5.95	0.00	7.18 +/- 1.93	6.17 +/- 1.94
20.00	16.37 +/- 10.45	11.94 +/- 10.13	0.00	10.79 +/- 3.29	9.81 +/- 2.57
10.00	33.43 +/- 3.66	20.73 +/- 3.14	0.00	3.6 +/- 2.22	0.57 +/- 2.16
1.00	23.45 +/- 5.49	11.18 +/- 4.25	0.00	6.85 +/- 5.23	5.59 +/- 5.41
0.00	29.17 +/- 4.79	17.62 +/- 6.12	0.00	-0.12 +/- 3.23	-1.28 +/- 2.76
RMSE Average	11.84	6.93		24.88	25.37
RMSE BMH	9.00	9.46		14.36	4.59
RMSE BLH	10.90	4.84		38.36	32.20
RMSE OMH	14.77	5.56		16.17	31.24

Table 5.12.2: PCR and PLS-R predictions for the levels of BLH and OMH in PMH:BMH, PMH:BLH and PMH:OMH blends.

BLH level / %			OMH level / %		
Expected	Predicted		Expected	Predicted	
	PCR	PLS		PCR	PLS
0.00	20.77 +/- 4.69	14.44 +/- 9.45	0.00	-12.04 +/- 15.67	-4.11 +/- 8.47
0.00	3.61 +/- 5.03	7.05 +/- 6.75	0.00	-11.51 +/- 2.34	2.62 +/- 2.78
0.00	9.39 +/- 3.73	14.17 +/- 5.66	0.00	-16.53 +/- 5.52	-2.16 +/- 3.31
0.00	23.74 +/- 9.03	16.39 +/- 17.65	0.00	-19.29 +/- 21.32	-6.7 +/- 9.46
0.00	5.74 +/- 3.83	9.88 +/- 8.66	0.00	-15.68 +/- 5.61	-2.24 +/- 4.46
0.00	13.71 +/- 2.9	21.34 +/- 3.42	0.00	-9.7 +/- 2.45	4.51 +/- 2.28
0.00	19.64 +/- 7.88	25.75 +/- 13.77	0.00	-17.75 +/- 14.72	1.77 +/- 4.34
0.00	2.94 +/- 0.74	13.74 +/- 3.67	0.00	-5.84 +/- 9.93	9.58 +/- 8.12
0.00	8.6 +/- 2.86	14.74 +/- 3.71	0.00	-6.53 +/- 5.87	4.57 +/- 5.2
0.00	23.85 +/- 4.47	23.11 +/- 11	0.00	2.56 +/- 8.28	4.63 +/- 9.44
0.00	14.42 +/- 3.63	21.4 +/- 8.6	0.00	-13.35 +/- 11.88	-4.8 +/- 7.45
0.00	6.85 +/- 4.94	15.82 +/- 8.99	0.00	-8.86 +/- 6.52	-1.44 +/- 3.93
0.00	8.5 +/- 2	14.05 +/- 4.74	0.00	-5.69 +/- 4.26	-3.6 +/- 4.61
1.00	7.56 +/- 3.62	5.85 +/- 7.89	0.00	-0.94 +/- 9.64	8.7 +/- 3.54
10.00	12.28 +/- 3.04	10.83 +/- 8.26	0.00	4.5 +/- 8.69	7.77 +/- 3.57
20.00	18.35 +/- 7.51	13.96 +/- 11.6	0.00	22.58 +/- 15.3	22.2 +/- 16.65
30.00	12.45 +/- 5.51	2.02 +/- 5.9	0.00	34.73 +/- 10.92	29.98 +/- 10.28
40.00	20.44 +/- 3.5	16.61 +/- 8.66	0.00	27.35 +/- 8.37	24.75 +/- 5.97
50.00	17.9 +/- 2.91	18.06 +/- 2.7	0.00	25.4 +/- 3.39	27.12 +/- 3.92
60.00	22.15 +/- 3.22	22.83 +/- 8.65	0.00	32.14 +/- 8.2	31.24 +/- 3.12
70.00	23.04 +/- 1.86	25.95 +/- 8.52	0.00	40.39 +/- 8.53	42.33 +/- 2.91
80.00	33.88 +/- 2.43	41.95 +/- 3.28	0.00	24.65 +/- 3.67	34.12 +/- 3.97
90.00	37.3 +/- 0.77	48.49 +/- 3.04	0.00	29.52 +/- 2.83	39.98 +/- 4.66
99.00	38.33 +/- 1.63	50.77 +/- 4.7	0.00	32.76 +/- 7.93	42.03 +/- 5.91
100.00	39.48 +/- 1.65	45.42 +/- 1.31	0.00	42.31 +/- 3.16	47.65 +/- 1.71
0.00	6.11 +/- 2.94	10.02 +/- 5.31	0.00	-4.71 +/- 6.99	6.03 +/- 5.53
0.00	0.23 +/- 5.08	-0.24 +/- 10.06	1.00	8.35 +/- 8.68	13.82 +/- 4.86
0.00	-1.1 +/- 3.71	1.39 +/- 8.56	10.00	10.35 +/- 12.87	23 +/- 10.23
0.00	10.73 +/- 3.37	9.92 +/- 6.72	20.00	13.2 +/- 6.26	16.58 +/- 2.61
0.00	14.34 +/- 4.01	13.7 +/- 4.7	30.00	21.63 +/- 12.72	22.99 +/- 12.02
0.00	13.64 +/- 1.51	7.47 +/- 7.19	40.00	25.27 +/- 11.79	22.24 +/- 5.68
0.00	9.71 +/- 5.69	3.44 +/- 7.67	50.00	36.85 +/- 11.67	26.03 +/- 10.63
0.00	18.48 +/- 3.81	13.07 +/- 2.52	60.00	38.47 +/- 1.37	27.93 +/- 3.74
0.00	21.92 +/- 2.9	20.04 +/- 4.11	70.00	39.63 +/- 8.8	29.44 +/- 10.33
0.00	24.19 +/- 6.59	13 +/- 10.05	80.00	65.25 +/- 14.7	48.65 +/- 12.4
0.00	11.81 +/- 4.09	7.48 +/- 5.62	90.00	71.22 +/- 2.27	51.16 +/- 1.32
0.00	25.25 +/- 7.43	16.66 +/- 14.26	99.00	66.57 +/- 13.01	44.45 +/- 7.44
0.00	18.75 +/- 3.54	13.79 +/- 4.16	100.00	69.87 +/- 4.23	52.2 +/- 6.23
RMSE Average	7.96	6.69		22.78	20.89
RMSE BMH	5.64	9.27		17.07	12.25
RMSE BLH	9.83	4.69		34.41	29.06
RMSE OMH	7.95	5.06		11.62	19.31

Table 5.13: The prediction of the levels of PMH, BMH, BLH and OMH for libraries of PMH, BMH ,BLH and OMH, using PCR and PLS-R models.

Heparin Species		Predicted level of each heparin species / %							
		PCR				PLS			
		PMH	BMH	BLH	OMH	PMH	BMH	BLH	OMH
PMH	Avg	82.3 +/- 13.18	3.65 +/- 5.99	1.6 +/- 8.91	12.45 +/- 10.47	87.01 +/- 13.61	3.47 +/- 5.78	0.68 +/- 8.11	8.84 +/- 12.34
	RMSE	22.00	6.98	8.98	16.21	18.73	6.71	8.07	15.10
	MIN	45.11	-10.20	-17.20	-11.18	48.30	-8.71	-15.07	-15.77
	MAX	109.13	19.84	19.00	37.44	117.95	20.16	19.21	48.18
BMH	Avg	2.96 +/- 23.52	94.81 +/- 10.34	1.82 +/- 16.5	0.42 +/- 13.01	3.09 +/- 18.06	95.49 +/- 10.39	3.53 +/- 13.7	-2.11 +/- 16.41
	RMSE	21.42	10.48	15.01	11.76	16.56	10.25	12.79	14.95
	MIN	-52.99	68.47	-36.11	-19.47	-38.35	67.56	-19.85	-37.16
	MAX	53.10	113.34	36.24	35.15	49.83	113.01	43.70	42.53
BLH	Avg	11.09 +/- 34.75	0.28 +/- 9.37	48.69 +/- 16.85	39.95 +/- 16.57	10.31 +/- 36.57	1.78 +/- 9.56	60.7 +/- 19.04	27.21 +/- 20.34
	RMSE	19.47	4.99	29.50	23.59	20.27	5.18	23.79	18.42
	MIN	-23.50	-18.50	13.60	-0.35	-30.01	-15.90	17.77	-24.62
	MAX	105.24	17.70	77.83	67.62	114.30	19.82	93.96	68.65
OMH	Avg	26.81 +/- 13.25	1.01 +/- 7.41	27.55 +/- 9.57	44.63 +/- 9.77	16.74 +/- 13.61	0.32 +/- 8	16.84 +/- 9.94	66.1 +/- 15.44
	RMSE	21.08	5.20	20.58	39.74	15.15	5.57	13.77	26.27
	MIN	-1.73	-10.37	13.97	9.75	-8.18	-12.56	-12.03	20.03
	MAX	46.90	25.80	45.36	61.11	36.91	27.10	32.89	104.47

Table 5.14: LODs for the classification of PMH samples that have been blended with BMH, BLH and OMH using PCA-KNN, PLS-DA, PLS-KNN and LDA. A sample is classified as contaminated when it is no longer classified as PMH. NA indicates that no separation was attained.

Contamination series			Level at which sample is no longer PMH / %
PCA-KNN	PMH:BMH	Lowest	20.00
		Average	30.00
	PMH:BLH	Lowest	NA
		Average	NA
	PMH:OMH	Lowest	10.00
		Average	10.00
PLS-DA	PMH:BMH	Lowest	10.00
		Average	20.00
	PMH:BLH	Lowest	10.00
		Average	20.00
	PMH:OMH	Lowest	40.00
		Average	60.00
PLS-KNN	PMH:BMH	Lowest	20.00
		Average	60.00
	PMH:BLH	Lowest	NA
		Average	NA
	PMH:OMH	Lowest	NA
		Average	NA
LDA	PMH:BMH	Lowest	50.00
		Average	60.00
	PMH:BLH	Lowest	10.00
		Average	20.00
	PMH:OMH	Lowest	10.00
		Average	10.00

The PCA and PLS score plots showed limited separation of PMH contaminated with BMH and PMH samples contaminated at the level of 20% (w/w) were seen to move into the space between PMH and BMH but samples contaminated below this level were not (**Fig. 5.22**).

[171]

Regression of both models had RMSE of 9.63 and 9.65 for PCR and PLS-R respectively (**Table 5.11**). The predicted values correlated with the expected values with R^2 values of 0.98 and 0.97 respectively and were not significantly different with P-values of 0.78 and 0.66 hence levels of BMH in PMH and vice versa can be accurately quantified. The levels of PMH and BMH in the PMH and BMH libraries were also predicted to determine LOQs. The PMH library showed on average 97.4 +/- 9.06% (w/w) and 97.51 +/- 6.99% (w/w) PMH for PCR and PLS-R models respectively whilst the BMH library showed levels of BMH of 96.88 +/- 9.3% and 96.5 +/- 9.41% (w/w) for PCR and PLS-R models respectively (**Table 5.11**). This suggests a LOQ of ~10% (w/w) for BMH in PMH and PMH in BMH and LODs of ~10% (w/w) for PMH in BMH and BMH in PMH.

The same procedure was then undertaken for PMH contaminated with the two remaining heparins; BLH and OMH whereby 6 randomly selected PMHs were contaminated with 3 randomly selected BLHs and 3 randomly selected OMHs at the same levels as above. Two other contamination series were formulated for testing purposes. All four libraries of different sources of heparin and the contamination series then underwent PCR and PLS-R (**Tables 5.12.1** and **5.12.2**). The libraries of PMH, BMH, BLH and OMH also underwent compositional analysis to determine LOQs (**Table 5.13**). The addition of the other heparins from different animal sources did not change the LOD of the models to detect PMH (RMSE 11.84 and 6.93 compared with 9.63 and 9.65 for PCR and PLS-R respectively) but it the LOD for the levels of BMH (RMSE = 24.88 and 25.37 for PCR and PLS-R). The LOD of BLH was at a similar level to PMH with RMSE of 7.96 and 6.69 for PCR and PLS-R respectively and the level of OMH had a LOD at a similar level as BMH with RMSE of 22.78 and 20.89 for PCR and PLS-R respectively. When only two heparins and blends thereof are compared the LOD improves; for the level of BMH in PMH, RMSEs of 14.36 and 4.59; for BLH in PMH, RMSEs of 9.83 and 4.69; and for OMH in PMH RMSE's of 11.62 and 19.31 for PCR and PLS-R respectively. This suggests that the models are able to differentiate between PMH and PMH contaminated with another heparin but struggles to distinguish between which type of heparin the PMH is contaminated with.

The decreased ability to predict is further shown following prediction of the four different heparins in each of the parent heparins (**Table 5.13**). The LOQ (i.e 100-RMSE of the heparin from an animal source in a *bona fide* sample of heparin from the same animal source) of PMH using all four heparin libraries increases from ~10% (w/w) for both models to ~20% and ~15% (w/w), for PCR and PLS-R respectively whilst the LOQ for BMH remains the same at ~10% (w/w). The LOQ for BLH and OMH is ~30% and ~25% (w/w) and, ~40% and 25% (w/w) for PCR and PLS-R respectively.
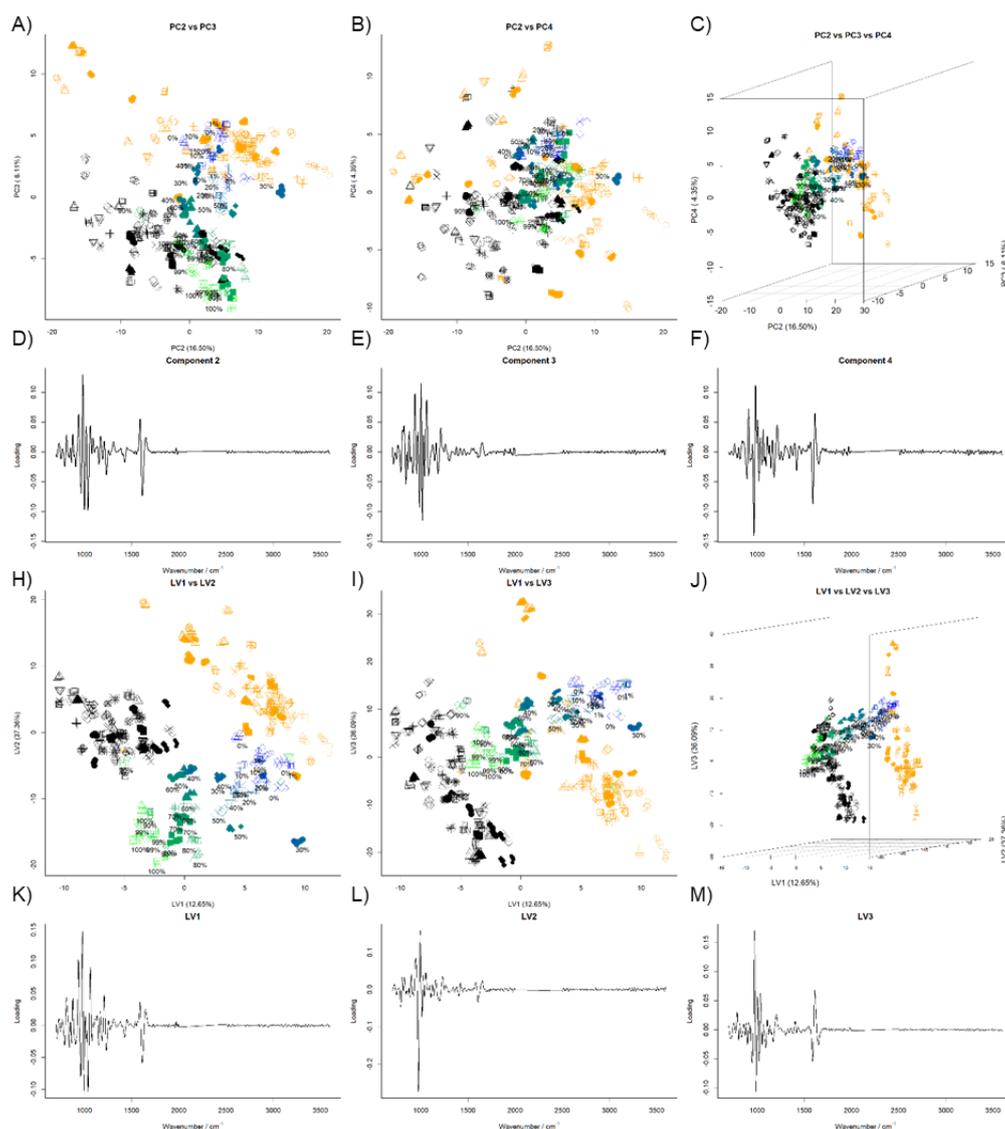
Figure 5.22: PCA and PLS-R models of PMH, BMH and PMH:BMH blends. A through F correspond to the PCA models and H through M correspond to PLS-R models. **A)** Scores plot for PC2 vs PC3. **B)** Scores plot for PC2 vs PC4. **C)** 3D score plot for PC2 vs PC3 vs PC4. **D)** Loadings plot for PC2. **E)** Loadings plot for PC2. **F)** Loadings plot for PC4. **H)** Scores plot for LV1 vs LV2. **I)** Scores plot for LV1 vs LV3. **J)** 3D scores plot for LV1 vs LV2 vs LV3. **K)** Loadings plot for LV1. **L)** Loadings plot for LV2. **M)** Loadings plot for LV3. PMH: black, BMH: orange, samples of PMH blended with BMH are indicated with a scale from green to blue – the bluer the sample, the higher the level of BMH. All spectra involved in PCA and PLS-R are in the regions of 700 and 2000 and, 2500 and 3700 cm-1 and are the averages of 3 scans. The spectra were smoothed, baseline corrected with a 7th order polynomial, normalised (0-1), smoothed and derivatised with a Savitzky-Golay algorithm and finally autoscaled.

The LODs were confirmed using cluster analysis, as the level at which the PMH blended with another heparin no longer belongs to the PMH cluster represents the LOD. To perform this analysis, samples of PMH which contain either BMH, BLH and OMH had the clusters to which they belong predicted. This clustering resembles the LRM used earlier in **Fig. 3.29**, but the linear portion of sigmoidal curve where thresholds are located is instead represented by a likelihood of belonging to a cluster or not. Four different clustering techniques: PCA-KNN, PLS-DA, PLS-KNN and LDA were used and, depending on the technique used, the lowest LODs for each heparin were 10%, which corroborate the RMSEs acquired during the predictions of the level of each heparin (**Table 5.14**). PLS-DA provided the best LODs across

[173]

all heparin sources at 10% (w/w) for BMH and BLH and 40% (w/w) for OMH but the LOD of OMH could be improved to 10% (w/w) using either LDA or PCA-KNN.

The PMHs contaminated with BMH underwent 2D-COS-Firs and were compared against a library of *bona fide* heparins to visualise if discrimination was possible at lower levels (**Fig. 5.23**). The variable region due to signals associated with concentration (highlighted in violet) was the region with the largest changes, but even down to the level of 1% (w/w) (averaged over 4 repeats) subtle shifts in regions not associated with concentration that are not found in pharmaceutical quality PMH can be located with ATR-FTIR spectra.
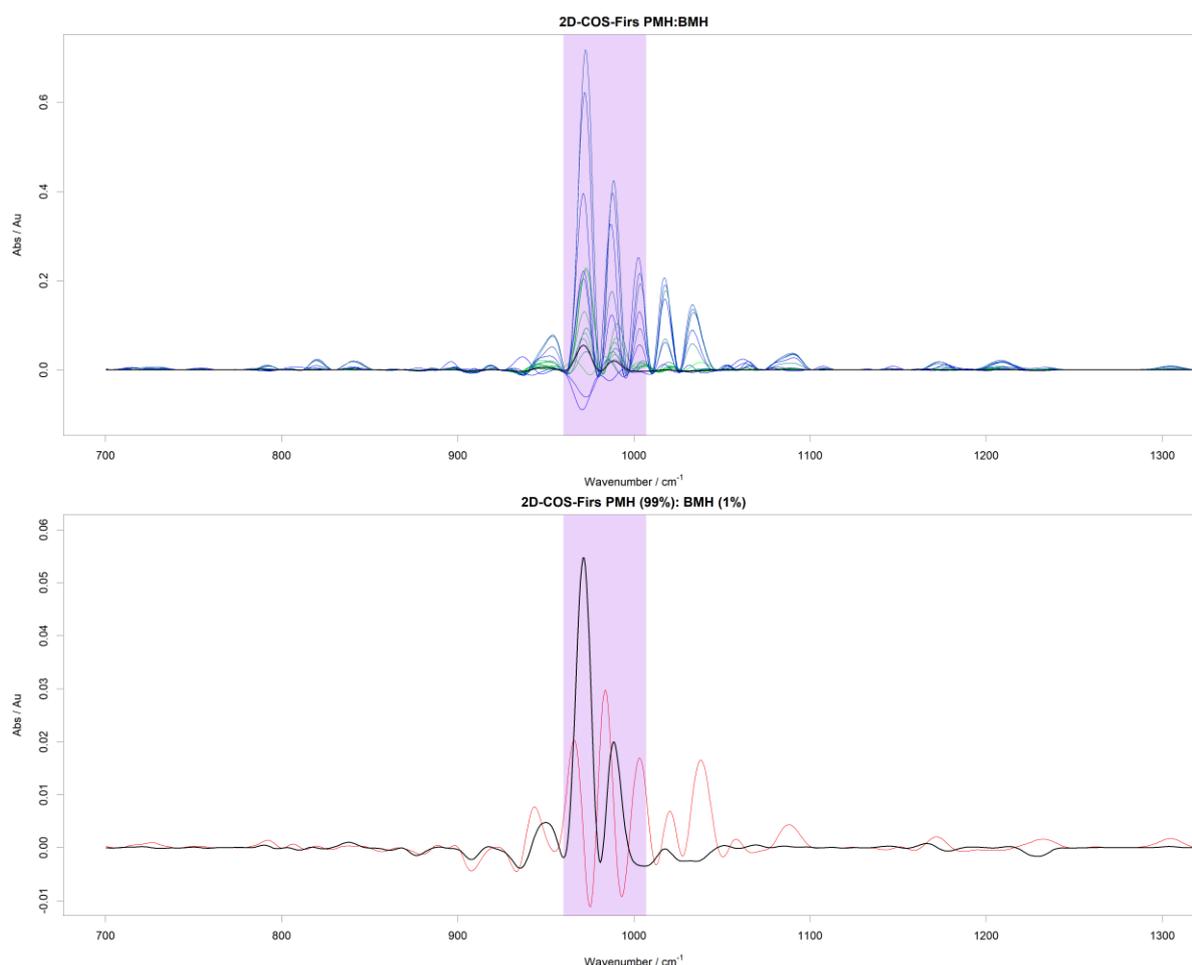
Figure 5.23: COS-Firs difference spectra of PMHs blended with BMHs and of a PMH:BMH 99:1 blend. Top) COS-Firs difference spectra for the PMH:BMH blend series at the levels of 1, 10, 20, 30, 40, 50, 60, 70, 80, 90, 99% (w/w) BMH. Colours are a scale from green to blue – the bluer the sample the higher the level of BMH. **Bottom)** COS-Firs difference spectra for a PMH (black) and a PMH blended with BMH to the level of 1% (w/w) (red). The purple region indicates the variable region associated with heparin concentration (960 – 1007 cm⁻¹). COS-Firs difference spectra were calculated using the second derivative of spectra that had been smoothed, base-line corrected to a 7th order polynomial, normalised (0-1) and then smoothed again. COS-Firs difference spectra were calculated by subtracting a covariance matrix of a heparin library from a covariance matrix of a heparin library that contains a test sample. The multiple matrices are constructed, using different combinations of heparins each time – the test sample never changes. After 50 iterations, the average of the resultant covariance matrices is calculated and the average power spectrum recorded and plotted.

[174]

5.4 Discussion

The key aims of this chapter were twofold, 1) to establish whether ATR-FTIR could determine the levels and types of sulphation in crude heparins, facilitating potential QC and 2) if so, apply these findings to the discrimination of heparins from different animal sources – the key structural differences being the level and types of sulphation.

On the subject of crude heparin QC, there is currently no strict definition or requirement as to what constitutes a crude heparin (Al-Hakim, 2021; FDA, 2013). The FDA has released a guideline on what it recommends to be "good" crude heparins, including monitoring of OSCS levels and testing for the animal source with the polymerase chain reaction but they are just suggestions to manufacturers (FDA, 2013). Mauri et al., 2017 demonstrated that a series of crude heparins from the previous 13 years and from 5 different suppliers are structurally similar and hence the potential to define a crude heparin exists. To establish this, Mauri et al. demonstrated that, using PCA of H-1 NMR spectra, crude heparins do not form any particular groupings or sub-types and suggest that they are confined to their own space in n-dimensions.

Similar results are observed with ATR-FTIR spectra, where crude heparins form their own region in GAG-PC space, creating a lobe near to the heparins when compared to both a GAG library and a heparin library. The crude heparins were fully compositionally and structurally characterised with SAX and NMR in Mauri et al., and these values were used to create a regression model with the intent of characterising crude heparin samples with PCA of ATR-FTIR spectra. The predictions were successful with many structural features predicted well (average RMSE = 0.91 for each feature) (**Table 5.3**). Prediction of the levels of DS and CS were also achieved and in keeping with the LODs observed in **Table 4.1**, LOD = ~1% and ~2.5% whilst the RMSE = 1.71 and 2.48 for DS and CS respectively. The RMSE were lower than those calculated with the mammalian GAG blend library 2.6 and 4.42 for DS and CS respectively (**Table 4.10**). The trend of improvement with relevant libraries is corroborated here – through use of a library specifically designed to characterise crude heparins, the ability to characterise crude heparins is improved. This lends credence to the idea that crude marine GAG compositional analysis will be improved with better libraries. Prediction of the level of sulphation of DS and CS was not obtained as the levels of IdoA-2S attributed to DS and GalN-6S attributed to CSC were not predicted, but rather the presence or lack thereof. This is similar again to the results observed in where the levels of CS and DS were confused with crude marine GAGs (**Tables 4.8 and 4.9**). However, the levels of DS and CS are low in the crude library (max = 16.7% and 7.7% (w/w), mean = 8.3 +/- 3.4%, 2.3 +/- 2.0% (w/w) for DS and CS respectively, as are the levels of IdoA-2S from DS and GalN-6S from CSC. On average, 22% of the CS moieties contain GalN-6S, hence IR would be required to detect 22% of 2.3% (0.5% of the total crude mixture), a value which is below any observed RMSE of prediction, including for those of heparin sulphation (average RMSE = 0.91 with PCR).

The level of 3S could be predicted, however the P-value was the lowest (P = 0.0011) albeit not significant, so care should be taken in interpreting the results. The low P-value may be ascribed to the correlation of 3S with PC7 – a low component, which accounts for only 1.31% of the total variance in the data, but 12.90% of the variance described by GlcN,3-S,6-X (**Table 5.2)**. It would make sense that the presence of 3S would appear in such a low component as 3S accounts for between 3 and 6% of the total sulphation in PMH and signals attributed to it would most likely be heavily masked by the other sulphate groups. Inspection of the loadings for PC7 shows bands near wavenumbers 1000 cm^{-1}, 1200 cm^{-1} and 1500 cm^{-1}, all regions which are associated with sulphation (**Fig. 5.4K**). These selected bands are unique to this component and are not observed to such a degree in the other components (**Figs. 5.3 and 5.4**), suggesting that this component is attributed to the level of 3S. The inclusion of loadings near 1500cm^{-1} suggest that some level of NS observed – as confirmed through the small amount of variance (2.65%) for ANAc6X-G being covered in this component. Correlations with DOS and 6S (1.96% and 3.05% of their respective variances) are also observed here. As 3OST1, the key 3OST involved in the biosynthesis of heparin, requires 6S (Moon et al., 2012), which in turn requires NS in the active site (Annaval et al., 2020), these correlations are indicative of the detection of 3S.

Similar observations as above were made with PLS-R albeit across less components. The predictions were improved in all cases with PLS-R, increasing from an average R^2 of 0.89 to 0.92. The RMSE for heparin structural features decreased from 0.91 to 0.76. Interestingly, the correlation in LV7 with 3S is still present, albeit at a lower level of 6.02%. The PCR model and the PLS-R model were able to structurally characterise crude heparins.

To test the robustness of the techniques, other molecules which structurally resemble crude heparins – three marine GAGs extracted from prawn and crab (Mycroft-West, 2021) – underwent structural prediction. PCR predicted the DoS and levels of NS and 6S in novel prawn and crab heparins (RMSE = 0.30, 4.29 and 3.94 respectively), while the PLS overpredicted in every circumstance (RMSE = 0.44, 6.81 and 11.93 respectively). PCR predicted 2S to a reasonable level when the 2S level fell within 10% of the levels that were found within the sample library it was tested against (error = 3.07 for crab F4) but was vastly overpredicted when the level was 40% or lower (errors = 19.68 and 42.9 for prawn F4 and F5) (**Table 5.5**). This is most likely because the model was not set up to account for low levels of 2S, and so, as observed in the improvement of RMSE for levels of CS and DS (**Tables 4.5** to **4.7**, **4.8** to **4.9** ,**4.10** to **5.3**), addition of samples that are more similar to the marine GAGs will facilitate improved detection. Hence, the PCR model proved robust in the characterisation of heparins and heparinoids in general, while the PLS-R model proved robust for the characterisation of crude heparins alone. It is possible that the trained nature of the PLS-R model makes it weaker in the detection of novel samples, whilst PCR, which looks at every

underlying trend, not trends that the analyst expects, is more flexible and hence can predict novel samples better.

While the PCR and PLS-R models predict every feature calculated from NMR and SAX, it does not necessarily mean that FTIR-ATR can directly observe these features. While the presence of different sulphation patterns should result in slightly altered chemical environments and hence different FTIR band patterns, dependent on the disaccharide composition, it is not clear if FTIR-ATR can simply observe certain levels of sulphate and these levels preclude certain disaccharides, or if FTIR-ATR can observe the subtle interactions that constitute individual disaccharides. Evidence already exists for the detection of intermonosaccharide interactions (D. Grant et al., 1989b) and simulated heparin disaccharide IR spectra show differences attributed to the substitution pattern (Monakhova et al., 2022). Nevertheless, care should be taken in the interpretation of the predictions of disaccharide moieties until more results are acquired. The numbers here are also predicted at the liberty of NMRs ability to discriminate them, hence errors accrued through NMR compositional analysis may be further amplified here.

ATR-FTIR was shown to accurately determine the sulphate levels of heparin in crude heparins. Since the main structural difference between heparins extracted from different animals is the level and pattern of sulphation (**Table 5.1**), it was deemed likely that ATR-FTIR could discriminate heparins from different animal sources. This is particularly important with the upcoming considerations about whether to diversify the heparin supply using bovine and ovine heparins, a decision which has already been made in Brazil with the re-addition of BMH to the heparin supply in 2016.

With PCA, separation was observed between BMH and the other heparins (**Fig. 5.8A**) and, within the remaining heparins, lobes could be seen to form through use of other PCs (**Fig 5.8**). As PMH, BLH and OMH clustered together, with little separation, attempts to calculate clusters which describe these areas were undertaken. Generally, the computed clusters failed to describe the real groupings of PMH, BLH and OMH. EM and KMC clustered BMH and BLH away from PMH and OMH (**Fig. 5.15F**) (**Fig 5.16B**). Weak separation of OMH from PMH was achieved with EM and KMC through use of PC8 (**Fig 5.16B and D**). In order to create clusters that explain the different heparins, multiple clusters were created into which the same heparin types would fall, resulting in a cluster of clusters which may describe each heparin (**Fig 5.16**). BMH was regularly 2 or 3 clusters, and if OMH and PMH were separated, it was into 3 or 5 clusters which represented PMH and 2 that represented OMH. While it is not difficult to merge these clusters together for the purposes of prediction, it shows that the clustering algorithms were not ideal for heparin species discrimination.

An attempt was made to increase the separations of the samples through the use of t-SNE (**Fig 5.17**). Again, EM and KMC produced clusters which contained the majority of the BMH and BLH samples but struggled with the separation of OMH and PMH (**Fig. 5.17D and E**). A visible separation between OMH and PMH was achieved with the application of t-SNE to PCs 2, 3, 5 and 8, but the clustering algorithms could not discriminate these (**Fig. 5.18**). PCs 2, 3, 5 and 8 were chosen here as they were stable as the main PCs for discriminating heparin species across numerous different iterations of PCA (**Figs. 5.7** and **5.19**). Visual distinction was improved with t-SNE compared with PCA, but t-SNE cannot be remapped with new data. To add new data, i.e test samples, either a new t-SNE and new clustering needs to be calculated, or a machine learning algorithm needs to be trained to predict the x-y values for a new sample onto the t-SNE. The difficult associated with testing new samples, weighed against the minimal improvement in separation meant that t-SNE was not taken forward.

As the clustering techniques used here are designed to cluster data sets with no known clusters, it was theorised that the clustering was failing due to the algorithms attempt to detect *potential* clusters. As the groupings are known *a priori*, it was hypothesised that trained methods may improve clustering and separation. An extension of KMC, whereby the centroids it bases its prediction on are pre-computed by the user, was used. The centroids were determined as the means for the scores of each different heparin source in each component, and the Euclidian distance between them and predicted samples was checked. While KMC traditionally uses the nearest neighbours of a sample to identity to classify its sample, it was decided that the Euclidian distances would be used. The Euclidian distances could be computed into a probability that a sample belongs to each group through use of the softmax function, generating soft classification. The new method, named KMC-CC was applied to the species across a range of different components. It was determined that components 2, 3, 5, 7 and 8 were the best for distinguishing different heparin species with this method (**Fig. 5.7**). Through use of component selection, up to 89% of samples could be correctly identified (**Table 5.7**). KNN, which is essentially a KMC-CC, except the centroid point that will best describe each cluster, not the mean of the expected values, is first computed through iteration. KNN of the selected components boosted the accuracy of discrimination from 89% to 98% and increased the ability of the model to predict BLH – BLH being the most difficult to predict in KMC-CC (**Table 5.8**). BLH was most likely the hardest to predict due to a small group of samples appearing as strongly PMH in the PCA plots (Fig 5.8B). Separate PCA of just the BLH samples does not reveal these samples to be outliers to BLH however, hence they cannot be discarded as outliers as described in **Section 2.2.3**.

Despite the improvement observed with KNN, KNN is a hard classification technique, i.e., there is no probability computed that a sample is or is not in that cluster as with KMC-CC. Hence PLS was applied again to the samples. Here however, PLS-DA rather than PLS-R was

used, which is similar to PLS-R but is instead aimed at grouping discrete samples. As observed with PLS-R, PLS-DA showed an improvement over PCA. When KMC-CC was applied over PLS plots a 98% accuracy was achieved. KNN boosted this to 99%, with only 7 samples being misclassified, compared with KMC-CCs 16, PCA-KNNs 18 and PCA-KMC-CCs 127 misclassified samples (**Table 5.9**). A further learned technique, LDA, was applied also. LDA separates samples based on underlying distributions which best explain each sample type by creating a linear combination that best separates them. LDA provided the best separation (**Fig. 5.21**) and prediction (**Table 5.10**) overall, with only 6 failed predictions. LDA seems to select more heavily for regions after wavenumber 2500 cm^{-1}, a region which is rarely involved in PCA and PLS loadings. This region is mostly ascribed to O-H bonds and hence, very subtle shifts here may be indicative of the underlying sulphate patterns (**Fig 5.21C,D and E**).

As with the PLS-R model before, there is always a risk that trained algorithms will only be applicable to the test samples they are trained on. As there are no "novel" heparin types to test here as with the marine GAGs for PLS-R, it is not clear if overprediction will occur when faced with a new sample. Hence, despite its lower discriminatory ability, PCA-KMC-CC may be the best approach for potentially novel samples, while PCA-KNN, PLS-DA-KNN and LDA would be the best approach for discrimination of PMH, BMH, OMH and BLH.

After demonstrating the ability of ATR-FTIR to discriminate between heparins from different animal sources, the ability to detect and quantify blends of the two was investigated. Blends of porcine heparins with bovine heparins represents a serious ethical issue in many countries, which can, currently, only be detected at the crude heparin stage with the polymerase chain reaction – the only technique recommended by the FDA. While it is entirely possible that PCR and PLS-R are able to discriminate the subtle changes in structural composition that would occur with the blending of heparins with differing sulphate patterns, the blend will remain undetected until the features themselves make the average composition structurally distinct from the parent heparin. For example, addition of 10% (w/w) of an entirely desulphated heparin, would result in an average decrease of 5% for all sulphate types, which could, for some heparins, still fall within the usual homogenous bounds of a normal heparin (but still be detected with ATR-FTIR; RMSE = 0.76 for sulphate levels with PLS-R). This issue is further amplified when the starting parental differences are only ~20% at most for a specific position that is sulphated for example, the level of 6S in BMH compared with PMH. This means that detection of 10% (w/w) contamination of PMH with BMH requires discrimination of a 2% change in the level of 6S (should only 6S be modelled), which is in line with the current RMSE of 1.92 and 1.87 for PCA and PLS-R respectively, but means that blends of less than 10% may not be detected. Nevertheless, PCR and PLS-R models were formulated, with the goal of quantifying the levels of PMH blended with varying levels of BMH, BLH and OMH.

First, just PMH blended with BLH was investigated as these are the two contemporaneous heparins in pharmaceutical use and hence are the two that are most likely to be blended together in a real-world situation but, blends of PMH:BLH and PMH:OMH were also analysed after. A LOQ of 8.06 and 9.3% (w/w) was observed for PMH and BMH respectively (**Table 5.11**) but when other heparins from different sources were added, the LOQs increased to 22.0% (w/w) for PMH and 10.5% (w/w) for BMH (**Table 5.13**). This is most likely because the underlying structures of PMH, BLH and OMH are similar and hence they overlap with the PMH samples, increasing the LOQ for each. LOQs for BLH and OMH were also obtained and, when all four sources of heparin are examined, they are 29.5% (w/w) and 39.7% (w/w) respectively when PCR is used. The LOQs could be improved to 11.8, 14.3, 9.8 and 11.6 for PMH, BMH, BLH and OMH respectively when just the libraries that contain the constituents of the blend are compared. The LODs for each heparin from a different species in PMH were also examined using clustering techniques and were similar to calculated RMSE values for quantification at 10% (w/w) for PMH, BMH, BLH and OMH depending on the clustering technique used (**Table 5.14**). Efforts to separate and quantify the levels of BMH and OMH in PMH have been studied with H^1, C^{13} HSQC NMR and chemometric analyses (Colombo et al., 2022). Using HSQC, the authors were able to discriminate OMH and BMH from PMH to the level of 8%, which is similar to that of FTIR, suggesting an equivalence in potential use.

In an attempt to observe subtle changes that may be underlying in the data, COS-Firs was utilised. Even at the level of 1% (w/w) BMH, unique spectral features were observed (**Fig. 5.23**). This suggests that there may be a way to detect low levels of heparin blends from sources with ATR-FTIR, but the method to do so has not yet been identified. This suggestion is corroborated by the differing LODs for the level of different heparins in PMH observed with different dimensionality reduction and clustering techniques, suggesting that the spectra contain the desired information, and that the different techniques may but do not always highlight these differences (**Table 5.14**).

Chapter 6: Final Discussion

6.1 FTIR-ATR as a quality control technique for pharmaceutical heparin

The primary aim of the thesis was to ascertain if FTIR could be used as a quality control technique for pharmaceutical, heparin. Current pharmacopeia-approved methodologies require both ^{1}H NMR, which is both expensive and technically challenging, and SAX-chromatography, which is comparatively slow and requires the regular production and subsequent use of standards. (**Table 1.2**)

NMR has been shown to have a LOD of 0.1% (w/w) OSCS (the major contaminant at study here) in crude heparins (Keire et al., 2010b) and has been observed down to the level of 1%

(w/w) in pharmaceutical heparins using NMR and PCA (Rudd et al., 2011b), while SAX has demonstrated a LOD of 0.03% (w/w) OSCS and 0.1% (w/w) for DS (Trehy et al., 2009b). Other methods which are not currently included in the pharmacopoeia and are technically comparable to FTIR include MS-ToF-SIMS which, combined with PCA and PLS has detected OSCS down to the level of 0.001% (w/w) (Hook et al., 2021) and CD which, through monitoring of selected signals has estimated LODs of 2.77% (w/w) for OSCS, 0.95% (w/w) for DS and 4.6% (w/w) for CSA (Stanley and Stalcup, 2011b). Here, FTIR has been shown to detect OSCS down to the level of 0.25%(w/w) in some instances (**Fig. 3.29**), and between 1 - 5% (w/w) for CSA, CSC and DS (**Table 4.1**). These values are confirmed with 2D-COS-firs data, which shows that heparins contaminated with 1% (w/w) CSA, CSC, DS or OSCS have alien signals present within the spectrum (**Fig. 4.12**).

NMR, SAX and MS all require some level of sample preparation. NMR for example, requires D_2O exchange and can require EDTA treatment to reduce line broadening from excess bound cations and internal standards where necessary (Mauri et al., 2017c). SAX requires preparation and characterisation of standards to compare retention times to, as OSCS is not a natural GAG, it must be made synthetically, the characterisation of which most likely requires use of NMR. Both NMR and SAX can take up to an hour per sample to run and shimming/column washes can add additional time to these analyses. MS-ToF-SIMS has a fast scan time of ~ 20 s, but preparation of samples is technically challenging, requiring novel inkjet fabrication this is coupled with the particularly large upfront cost of a SIMS-MS system and chemometric analysis must be applied to the signals at specially selected m/z ratios (Hook et al., 2021). As with MS, NMR and CD have a large upfront cost (~£200k, ~£800k and ~£80k respectively) and operation often requires skilled employees, also like MS, the LOD of NMR-PCA for OSCS can alter when different portions of the spectra are analysed, where it increases from 1 to 5% (w/w) when the whole spectrum instead of the anomeric region is analysed. Unlike the other methods discussed, FTIR is rapid, facile, and cheap (**Table 6.1**). FTIR is also routinely taught in many science degrees and is therefore easily accessible to many individuals with a low barrier to entry. Spectral acquisition requires ~ 10 s per repeat, which is considerably lower than NMR and SAX and still lower than MS and CD. The output data is relatively simple to handle and can be rapidly output as a simple CSV file with no prior manipulation. Furthermore, the entire spectrum (with the exclusion of variable regions owing to air composition) is used in every circumstance. No specific region or bands are required for the approaches outlined here and this improves the robustness of the technique, but it is important to note that differences observed due to sample concentration and other factors can be greatly reduced by excision of a single band. Sample preparation for FTIR is minimal, with only a dry powder used here, making the technique further applicable to other potentially insoluble samples. It is also entirely possible that multiple samples could be screened rapidly

using an IR microscope, analogously to the method used for MS-ToF-SIMS as used in (Mainreck et. al 2010).

Table 6.1: Comparison of different, monograph approved or similar approaches to ATR-FTIR for the detection of OSCS.

Method	Upfront cost / £	Acquisition time	LOD OSCS / %	Portability	Notes	Refs
NMR	~800 k	1h	0.1	-	Applied to multiple contaminants, to the level of ~1% with PCA	(Keire et al., 2010b)
SAX	~30 k	1h	0.03	+	Limited ability to identify contaminants	(Trehy et al., 2009b)
CD	~80 k	5 min	2.77	-		(Stanley and Stalcup, 2011b)
MS	~200 k	20 s/scan	0.001	-	Requires PCA of specific signals to achieve detection.	(Hook et al., 2021)
FTIR	~20 k	10 s/scan	0.5	+++	Requires PCA to detect contaminants.	The present work

The LOD for OSCS present within heparin using FTIR (0.5% (w/w) in all cases) is higher than MS (0.001% (w/w), (Hook et al., 2021)), NMR (0.1% (w/w) in crude heparins (Keire et al., 2010b)) and SAX (0.03%(w/w) (Trehy et al., 2009b)) but is lower than the 3% (w/w) threshold at which symptoms appear in mice and pigs (Corbier et al., 2011; McKee et al., 2010) and is much improved compared to NMR-PCA methods, which discriminate levels of OSCS of 1% (w/w) (Rudd et al., 2011b), however it is important to note that no LOD is quoted for this technique. The entire system is significantly faster and requires little set-up or maintenance. The technique does not require regular synthesis of standards as in SAX, nor specialist technical support as with NMR, MS and CD. Another added benefit to FTIR is the portability of the system – an IR instrument can be taken as carry-on luggage on a plane, and thus could be taken to the dispersed and difficult to manage crude heparin production facilities (Mauri et al., 2017c).

This work also represents the first time that a large panel of relevant and potential contaminants in pharmaceutical heparins has been analysed. Previously, (Rudd et al., 2011b) focused on two alien saccharides that would be difficult for NMR specifically to detect (OSAS and DeS), here, all main GAG species bar KS, commercially available polysaccharides and OS-variants of each were screened. Many works focus solely on detection of low levels of OSCS and occasionally of the detection of DS. Here, all GAGs and OS-variants are examined,

providing reasonable LODs for each (**Table 4.1**). Detection of HA with the LRM model was not achieved due, presumably, to the unsulphated nature of the contaminant. This was highlighted by addition of numerous "other" polysaccharide samples which had been artificially sulphated – all of which had relatively high PC4 scores, akin to those of OSCS, while the unsulphated parent compounds had, akin to HA, negative PC4 scores. High scores were also observed for unsulphated carbohydrates that instead had other ring modifications, including alkylations and branches, suggesting that the changes observed with this technique are based on stresses place upon the carbohydrate backbone by extra moieties. A separate PCA and subsequent LRM was performed on HA samples, providing a LOD of ~2.5% (w/w), which is in line with the other GAGs and also backed up by 2D-COS-firs **(Fig. 4.12)**. Hence, for potentially unsulphated or lowly-sulphated contaminants, additional care must be taken when using this approach and unusually low PC4 scores should be monitored. Multiple repeats using many heparins and contaminant GAGs of the same type were also studied as biological repeats; a methodology which is often not followed with other techniques, potentially due to the high acquisition times associated with them.

This method offers no way to quantify the amount of contaminant, but for a simple QC check this is no issue. The LOQ for OSCS appears to begin at ~5% (w/w), which is in line with CS and DS LOQ of ~5% (w/w). Also, the broadness of the LRM approach may be of detriment as there is no way to identify which contaminant is present. For example, small amounts of DS are accepted in most pharmaceutical heparins (BHRA, 2019; EP, 2010) but may be detected as a contaminant. In theory, the model created here contains only pharmaceutical heparins and hence previously accepted amounts of DS should permeate the library, meaning that any differences observed will account for natural DS presence, but the matter should be taken into consideration. The natural levels of DS in the library may account for the higher LOD for DS at up to 5% (w/w) as the underlying DS in the library may mask the contaminant.

6.2 Identification of the parental origin of heparin.

Another important aspect of heparin QC is that of discerning heparins from distinct parental origins – particularly heparins from porcine, bovine and ovine sources. ATR-FTIR coupled with PCA was shown to cluster each individual heparin species through use of multiple components. A series of additional dimensionality reduction techniques, including PLS-DA, LDA and tSNE were also applied and to the outputs of these, a series of clustering techniques were trialled. Discrimination of the four heparins had a 99% accuracy with both LDA coupled with likelihood estimation and PLS-KNN and, a 98% accuracy for PCA-KNN (**Tables 5.9 and 5.10**). Only trained techniques (LDA, PLS-DA, KNN and to some extent KMC-CC) were capable of grouping and predicting sample identities. PCA alone had much overlap between BLH, OMH and PMH, making visual discrimination difficult and untrained clustering techniques

applied over tSNE and PCA could neither cluster the different heparin sources nor predict them. This suggests that a degree of *a priori* knowledge is required, and any novel heparin species will require the production of a new library and subsequent model for detection.

One potential threat to modern pharmaceutical heparin regulation is the introduction of blends of heparins from different animal sources, which constitutes a major ethical and legal issue in many countries. Currently, the US and European pharmacopoeia approved method for species identification is through quantitative polymerase chain reaction (Q-PCR), which can only be performed at crude stages (as nucleic acids are considered contaminants in pharmaceutical heparin and hence a DNAse and an oxidation step are used to remove any such contaminants)(Auguste et al., 2012). Furthermore, low levels of potential contaminants such as OSCS can perturb the action of DNAse (Auguste et al., 2011). Efforts to detect both bovine and ovine heparin in porcine heparin have been undertaken with NMR, achieving detection at 8% (w/w) for both (Colombo et al., 2022). Here, with discriminant techniques, FTIR was shown to detect ~10% (w/w) BMH with PLS-DA and ~10% (w/w) OMH with PCA-KNN. Detection of BLH and OMH in PMH at good levels was observed with LDA, but detection of BMH was poor. Attempts to quantify levels of contaminants were also undertaken, with RMSE of 11.85 and 6.94 with PCR and PLS respectively for level of PMH contaminated with either BLH, BMH or OMH. Quantification of BLH and OMH was worse, but as PMH is currently at most risk of contamination this isn't such an issue. The LOQs for other heparins in PMH was as high as 18%, but should detection alone be required, a LOD of ~10% (w/w) for BMH is slightly more than NMR whilst OMH is in-line with current NMR-PLS techniques. 2D-COS-firs was utilised to show that signals (albeit fewer signalss) attributed to as low as 1% (w/w) contamination of PMH with BMH could be observed, which is in-line with filtered NMR (Rudd et al., 2012b). Making these specific signals more important in the dimensionality reduction process through scaling may make the LOD and LOQ lower but care must be taken as this may make the technique specific to BMH contamination. However, attempts to use the filtered signals as inputs for PCA were tested here, but for heparin:other GAG blends (**Figure 4.13**, **Table 4.13**) and this provided worse results than conventional PCR (**Table 4.7**). The filtered signals were not considered as loadings in this approach, instead the filtered spectra were used as inputs. A filtered signal could be used as a loading to create COS-Firs scores which should, in theory, select solely for the alien signals present in the target(s). Further work should be performed to investigate this, which may introduce a new, targeted dimensionality reduction technique.

MS-TOF-SIMs has shown detection of BMH and BLH down to the level of 0.001 and 0.01% (w/w) respectively in PMH with PLS. Similar to the detection of OSCS, this requires measurement of specific, pre-determined m/z signals to achieve. It is possible that through similar manipulation of the IR spectra, using signals determined from 2D-COS-firs for example,

that the LOD for ATR-FTIR and indeed NMR could be improved greatly. Furthermore, the signals used differ between techniques applied to detection of the same sample – for example, 18 ions are used for PCA of PMH:OSCS whilst 40 are required for PLS of PMH:OSCS. Between the two sets of ions required, only 1 is found in both sets. This deep understanding of the MS system and its outputs is the antithesis of the rapid and facile technique explored here with ATR-FTIR.

6.3 The application of FTIR-ATR to crude heparin samples.

Crude heparin is a 2 parts heparin 1 part "other" mixture that forms the raw product, which is eventually refined to make pharmaceutical heparin. It is believed that OSCS may have been introduced to the heparin supply chain at the crude stage (FDA, 2013; Mendes et al., 2019) and that this may be the entry point for heparins of different parental origins (Al-Hakim, 2021). Moreover, the only stage in heparin production at which different heparins can be located by pharmacopeia approved methods, is at the crude stage with Q-PCR. Recently there has been a push to control crude heparins and the FDA has published guidelines (FDA, 2013) for what a crude heparin should contain prior to any requirements for the final API quality monograph. Here, it is demonstrated that ATR-FTIR can discriminate between crude and pharmaceutical quality heparins (**Fig. 5.2**) and hence, may be used to create chemometric models that can define "good quality" crude heparin, akin to the methods shown with OSCS and other GAG contamination (**Fig. 3.29**).

Unfortunately, crude samples from different parental origin were not available for analysis during this project. Future experiments should be performed to see if crude ovine and bovine heparins can be discerned from porcine heparins, in an approach similar to that outlined with PMH, BMH, BLH and OMH, as in **chapter 5**, including for blends. It is possible that the differing CS/DS profiles both structurally and compositionally from different animals may alter the separability between types and thus change the LOD/LOQ. Detection of OSCS and other OS-GAGs was also not investigated for crude-heparins, due to a lack of available starting materials. Attempts to define a LOD of OS-GAGs should be attempted in crude heparins, as again, the presence of other GAGs in the mixture may alter the LOD.

ATR-FTIR lends itself particularly well to crude analysis due to its high portability. A QC agent could easily travel to a crude manufacturing facility and quickly analyse any batch that they desire on location. This allows robust spot checks to be undertaken and could also be used to detect which facilities, if any, that contamination is beginning at before the blending of many lots occurs. Should ATR-FTIR show similar predictive power for animal sources of and OSCS contamination in crude heparins as it does with pharmaceutical heparins, it would be of high value for crude heparin QC.

6.4 FTIR as a tool for the structural analysis of GAGs.

Since the major difference between pharmaceutical heparin and crude heparin is the presence of other GAGs in a mixture, and that the crude heparins separated from the pharmaceutical heparins towards the relevant GAGs in PCA, it was hypothesised that FTIR could be used for the compositional analysis of GAGs. Quantification of different GAG species was successful and through use of a large library of GAGs from multiple mammalian sources, the quantification of each GAG was achieved with a RMSE of ~5%. Through use of a more specific library, i.e. a crude heparin library to analyse crude heparin, RMSEs of 2.48 and 2.07 for PCR and PLS were achieved for the level of heparin, with RMSE as low as 0.66 for CS.

The rapidity coupled with the accuracy of this technique is unparalleled and complete GAG identification of GAG mixtures can be achieved in under a minute for a dry, powdered sample. Care must be taken with this approach however, as the library used for quantification can greatly alter the outcomes. This is observed with crude heparins, where a crude heparin specific library greatly increases the accuracy. Marine GAGs are GAGs extracted from marine sources and often contain interesting and unusual structural motifs which may provide novel bioactivities (Mycroft-West, 2021). ATR-FTIR can and has been used to pinpoint the major structural components of marine GAGs (Mycroft-West, 2021; Mycroft-West et al., 2021, 2020, 2020, 2019) and here, it has been used to quantify some of the GAG constituents with some level of success (**Table 4.9**). Samples which resemble crude heparins (i.e ~2/3 heparin) such as prawn and crab extracts had RMSE ~2.5 for the level of heparin and there was relative success at distinguishing between CS and DS albeit with a regular overprediction of DS. The overprediction of DS is likely due to detection of DS in NMR being determined by level of IdoA-GlcA, not necessarily of chains which contain some IdoA, hence more work is required to more accurately quantify the levels of CS/DS in these compounds. Some compounds that were not similar to crude heparins such as pilchard and catfish extracts also underwent compositional prediction of heparin to a good level of accuracy (RMSE= ~5) but still, the CS:DS ratio was not adequate. Other marine samples were incorrectly predicted however, this may be due to levels of HS, not heparin skewing the prediction to a negative value. Further work, using HS as a blending agent instead of heparin could rectify this.

One important observation was that of regular discrimination between CSA and CSC subtypes, as well as discrimination of OSHp from heparin and the heparins from different animal sources. Since the only differences between these molecules are sulphate type, levels or both respectively, it was hypothesised that FTIR could be used to structurally characterise GAGs by their sulphation levels and patterns. The crude heparins, which have been fully structurally characterised in (Mauri et al., 2017c) and contain varying levels of sulphation prior to product blending were used. Using a crude heparin library to predict crude heparins also gives the technique the best chance at success, and this should be taken into account for

other applications. Using PCR and PLS, RMSE of 0.91 and 0.76 were achieved respectively (**Tables 5.3** and **5.5**) for the levels of different sulphates in crude heparins, suggesting that structural analysis of sulphate positions is achievable with FTIR. Similar to compositional analysis, the robustness of the technique was tested through characterisation of marine GAGs and produced promising results with a RMSE of 3.9 and 4.3 for 6O- and NS respectively. Prediction of 2S, which was unusually low in two of the samples (at 40 and 30%, compared to the 65.4% to 77.7% of the crude heparin library) and hence failed to be predicted. As with compositional analysis, this highlights the importance of the library used for prediction and that the limits of the library cannot be exceeded. For *P. pelagicus* F4, for which the levels of sulphation fell within the limits of the crude library, the RMSE$_{average}$ for 6O-, 2O- and NS was 2.08 which is slightly worse than the RMSE for crude heparins at 0.91 but still very good, suggesting that the model can be used for crude-heparin-like samples.

Predictions of the level of 3OS were the worst, possessing the lowest P values and therefore the most chance of being predicted by random chance. Analysis of PCA and PLS loading plots shows that bands which would most likely correlate with sulphation (in this case, assumed to be 3OS) were selected for in components which strongly correlated with 3OS level but the naturally low levels of 3OS in heparin may obscure the true amounts to ATR-FTIR. Levels of different disaccharides were also quantified with good accuracy, but it is not clear if the presence of different sulphates precludes different disaccharides. It is established that interactions between monosaccharides in a polymer can be observed with FTIR (Grant et al., 1989b) and that disaccharide substitution patterns in simulated spectra can be discriminated (Monakhova et al., 2022), hence it is entirely possible that different disaccharides produce different signals which can be detected. If this is true, then ATR-FTIR may be able to quickly perform disaccharide analyses, akin to recent methods developed by (Mauri et al., 2017) through use of NMR and peak volume analysis but further work is required to support this idea.

To demonstrate that FTIR can differentiate between levels of different sulphates, a panel of chemically modified heparins (CMH) which encompass the many different sulphate positions in heparin could be produced as described in (Yates et al., 1996) and undergo spectral interrogation. In theory, the unique combinations of each sulphate should produce unique spectra, confirming that the sulphate positions can be detected. A similar approach could be used with disaccharides, whereby unique disaccharides should produce unique spectra – this may help to confirm that FTIR can quantify these structural features. Evidence that IR can detect different sulphate positions already exists and is demonstrated by the clear discrimination between CSA and CSC samples (**Fig. 4.9**), (Cabassi et al., 1978; Grant et al., 1989a; Mainreck et al., 2011)

Mathematical blending, whereby spectra of a parent and contaminant are added together accordingly to simulate a contaminated spectrum, has recently been published as a technique for LOD detection of different heparin species with NMR (Colombo et al., 2022). Application of such a technique to the IR libraries could make improving the accuracy of quantification much faster and simpler. Through use of CMHs and other chemically modified GAGs, techniques for the syntheses of which are well documented and/or under contemporaneous study (Palhares et al., 2021), one could mathematically construct large libraries which encompass *all* possible structural variation. Interactions between disaccharides may not be accounted for in such a model and may produce significant errors. To overcomet this, oligosaccharides containing distinct disaccharide sequences could instead be used, but this is expensive, time consuming and would require a large amount of work and collaboration. Another important aspect to consider is the cation form of the studied GAGs. Burmistrova et al., (2021) and Grant et al., (1987b) have shown that IR spectra can alter in response to different cations and IR has been used in conjunction with CD to display the binding sites of cations (Rudd et al., 2008). These spectral alterations may reduce the accuracy of predictions for heparins that are not in the sodium form as examined here, and must be considered moving forward, especially as sodium heparin and calcium heparin have different monographs. Such spectral alterations are also observed with NMR, where EDTA is often required to reduce line broadening associated with Ca^{2+} ions.

Therefore, the technique as it stands, can be used to confidently quantify the composition of mammalian (porcine, ovine and bovine) GAGs to the level of 5% (w/w) (2.5% (w/w) with more optimal libraries) and, with modification, may be used to quantify the composition of novel marine (and potentially more) GAGs rapidly. There is also evidence that ATR-FTIR can detect sulphates at various positions and quantify them to the level of 0.71% in the best case. This is highlighted by prediction of marine GAG structures, structures which are not included in the original library, but are predicted to the level of 2.3% (w/w). With both compositional and sulphate analyses, care must be taken to use an adequate library for the compound being studied. This requires some level of *a priori* knowledge which is not ideal with rapid identification of entirely novel GAGs. NMR and SAX (given that adequate standards and derivatisation methods exist) can easily identify and quantify novel structures through methods which are not based on prepared libraries of known compounds and will remain the port of call for truly novel GAGs. Nevertheless, the ability to rapidly analyse the structure of crude heparins and some crude marine GAGs in under a minute is currently unparalleled by any technique and represents a significant advance in the field.

6.5 Conclusions and future prospects

To conclude, FTIR has been shown to detect low (< 1% (w/w) in some cases) amounts of GAGs, semisynthetic GAGs (such as OSCS) and other, natural and semisynthetic polysaccharides in pharmaceutical heparins and has therefore shown potential to be used for heparin QC. This is comparable with other techniques in use such as NMR and SAX and is importantly below the 3% (w/w) threshold at which side effects are seen with OSCS. MS-ToF-SIMS is the only technique with a major increase in sensitivity at 0.001% (w/w) for OSCS but MS-ToF-SIMS is a very specialist piece of equipment, expensive to run and maintain and the model used can detect *only* OSCS at that level – in order to detect other contaminants different, bespoke models are required for each contaminant and dimensionality reduction technique used, whilst FTIR is performed on the same spectral space for every contaminant detected.

Crude heparins have been shown to be uniquely represented in PC space and thus may also be QCed effectively. Importantly, crude heparins had underwent GAG composition and sulphate pattern analysis and were quantified accurately (±2.3% (w/w) and ±0.71% respectively). Identification of heparins from four different origins has also been achieved with high (99%) accuracy and blends between them have been detected and to some degree quantified at the levels of 10% (w/w) or higher. Further work which studies crude heparins from different parental origins should be undertaken and may provide interesting results for FTIR as a QC technique.

Arguably the most important results to come from these works are preliminary work that shows propensity for FTIR to be used as a technique to accurately characterise the composition and structure of GAGs. This technique has shown robustness when faced with novel marine GAGs, of which there are no similar examples found in the original library but also highlighted that care must be taken in which library is used for characterisation. Creation of more relevant libraries could be achieved through addition of more, fully characterised, samples which cover more structural spaces, either through novel GAG extraction, chemo-enzymatic synthesis, traditional synthesis or mathematical blending of pre-described samples.

Importantly, FTIR as a technique is rapid (~10 s per scan), facile (minimal sample preparation and easy data handling) and cheap (costing ~£20,000 for a high-end machine). It is highly portable and easily accessible to many users allowing QC to be taken to relevant facilities and possible sources of contamination. The speed at which FTIR spectra and subsequent analysis (~1 minute for structure prediction) can be performed to a high degree of accuracy (±2.3% (w/w) and ±0.71% for composition and sulphate positions respectfully) will hopefully allow FTIR to become routine technique used for the structural analysis of GAGs in general. These results come at an interesting time, as infrared microspectroscopy is a technique that is

currently receiving much interest due to huge increases in spatial resolution (Gierlinger, 2018; Gieroba et al., 2020; Kumar et al., 2016; Mohamed et al., 2020). Such a technique, coupled with the models developed here may be very powerful, allowing non-destructive GAG analysis on live cells (Doherty et al., 2019) over time periods and in response to various cellular challenges (Chan et al., 2020; Elsheikha et al., 2019) whilst monitoring substrate turnover . Such an application is however in its infancy, due to the added spectral load of proteins, lipids, nucleotides, other carbohydrates and other small molecules that will no doubt also be detected (Lima et al., 2021).

Chapter 7: References

Ahsan, A., Jeske, W., Hoppensteadt, D., Lormeau, J.C., Wolf, H., Fareed, J., 1995. Molecular profiling and weight determination of heparins and depolymerized heparins. J. Pharm. Sci. 84, 724–727. http://dx.doi.org/10.1002/jps.2600840612

Adam, A., Montpas, N., Keire, D., Désormeaux, A., Brown, N.J., Marceau, F., Westenberger, B., 2010. Bradykinin forming capacity of oversulfated chondroitin sulfate contaminated heparin in vitro. Biomaterials 31, 5741–5748. https://doi.org/10.1016/j.biomaterials.2010.03.074

Aggarwal, C.C., Reddy, C.K. (Eds.), 2013. Data Clustering: Algorithms and Applications, 1st edition. ed. Chapman and Hall/CRC, Boca Raton.

Aikawa, J., Munakata, H., Isemura, M., Yosizawa, Z., Tada, K., Suzuki, M., Sakurai, T., 1985. Sulfated Glycopeptides from Middle Ear Effusions of Secretory Otitis Media. Tohoku J. Exp. Med. 146, 461–467. https://doi.org/10.1620/tjem.146.461

Aksoy, A.E., Hasirci, V., Hasirci, N., 2008. Surface Modification of Polyurethanes with Covalent Immobilization of Heparin. Macromol. Symp. 269, 145–153. https://doi.org/10.1002/masy.200850918

Alban, S., Lühn, S., Schiemann, S., 2011a. Combination of a two-step fluorescence assay and a two-step anti-Factor Xa assay for detection of heparin falsifications and protein in heparins. Anal. Bioanal. Chem. 399, 681–690. https://doi.org/10.1007/s00216-010-4252-0

Alban, S., Lühn, S., Schiemann, S., Beyer, T., Norwig, J., Schilling, C., Rädler, O., Wolf, B., Matz, M., Baumann, K., Holzgrabe, U., 2011b. Comparison of established and novel purity tests for the quality control of heparin by means of a set of 177 heparin samples. Anal. Bioanal. Chem. 399, 605–620. https://doi.org/10.1007/s00216-010-4169-7

Alberto-Silva, C., Malheiros, F.B.M., Querobino, S.M., 2020. Fourier-transformed infrared spectroscopy, physicochemical and biochemical properties of chondroitin sulfate and glucosamine as supporting information on quality control of raw materials. Future J. Pharm. Sci. 6, 98. https://doi.org/10.1186/s43094-020-00120-3

Al-Hakim, A., 2021. General Considerations for Diversifying Heparin Drug Products by Improving the Current Heparin Manufacturing Process and Reintroducing Bovine Sourced Heparin to the US Market. Clin. Appl. Thromb. 27, 10760296211052292. https://doi.org/10.1177/10760296211052293

Alkhalil, A., Achur, R.N., Valiyaveettil, M., Ockenhouse, C.F., Gowda, D.C., 2000. Structural Requirements for the Adherence of Plasmodium falciparum-infected Erythrocytes to Chondroitin Sulfate Proteoglycans of Human Placenta *. J. Biol. Chem. 275, 40357–40364. https://doi.org/10.1074/jbc.M006399200

Alkrad, J.A., Mrestani, Y., Stroehl, D., Wartewig, S., Neubert, R., 2003. Characterization of enzymatically digested hyaluronic acid using NMR, Raman, IR, and UV–Vis spectroscopies. J. Pharm. Biomed. Anal. 31, 545–550. https://doi.org/10.1016/S0731-7085(02)00682-9

Alquwaizani, M., Buckley, L., Adams, C., Fanikos, J., 2013. Anticoagulants: A Review of the Pharmacology, Dosing, and Complications. Curr. Emerg. Hosp. Med. Rep. 1, 83–97. https://doi.org/10.1007/s40138-013-0014-6

Altman, N.S., 1992. An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression. Am. Stat. 46, 175–185. https://doi.org/10.1080/00031305.1992.10475879

Amarasekara, A.S., Opoku, G., Qiu, X., Doctor, V., 2007. Effect of oversulfation on the chemical and biological properties of chondroitin-6-sulfate. Carbohydr. Polym. 68, 116–121. https://doi.org/10.1016/j.carbpol.2006.07.019

Andersson, L.-O., Barrowcliffe, T.W., Holmer, E., Johnson, E.A., Sims, G.E.C., 1976. Anticoagulant properties of heparin fractionated by affinity chromatography on matrix-bound antithrombin III and by gel filtration. Thromb. Res. 9, 575–583. https://doi.org/10.1016/0049-3848(76)90105-5

Annaval, T., Wild, R., Crétinon, Y., Sadir, R., Vivès, R.R., Lortat-Jacob, H., 2020. Heparan Sulfate Proteoglycans Biosynthesis and Post Synthesis Mechanisms Combine Few Enzymes and Few Core Proteins to Generate Extensive Structural and Functional Diversity. Molecules 25, 4215. https://doi.org/10.3390/molecules25184215

Antia, I.U., Mathew, K., Yagnik, D.R., Hills, F.A., Shah, A.J., 2018. Analysis of procainamide-derivatised heparan sulphate disaccharides in biological samples using hydrophilic interaction liquid chromatography mass spectrometry. Anal. Bioanal. Chem. 410, 131–143. https://doi.org/10.1007/s00216-017-0703-1

Asri, M.N.M., Verma, R., Ibrahim, M.H., Nor, N.A.M., Sharma, V., Ismail, D., 2021. On the discrimination between facial creams of different brands using Raman Spectroscopy and partial least squares discriminant analysis for forensic application. Sci. Justice J. Forensic Sci. Soc. 61, 687–696. https://doi.org/10.1016/j.scijus.2021.08.006

Atha, D.H., Gaigalas, A.K., Reipa, V., 1996. Structural analysis of heparin by Raman spectroscopy. J. Pharm. Sci. 85, 52–56. https://doi.org/10.1021/js950216c

Auguste, C., Dereux, S., Martinez, C., Anger, P., 2011. New developments in quantitative polymerase chain reaction applied to control the quality of heparins. Anal. Bioanal. Chem. 399, 747–755. https://doi.org/10.1007/s00216-010-4232-4

Auguste, C., Dereux, S., Rousset, M., Anger, P., 2012. Validation of quantitative polymerase chain reaction methodology for monitoring DNA as a surrogate marker for species material contamination in porcine heparin. Anal. Bioanal. Chem. 404, 43–50. https://doi.org/10.1007/s00216-012-6085-5

Ausili, A., Sánchez, M., Gómez-Fernández, J.C., 2015. Attenuated total reflectance infrared spectroscopy: A powerful method for the simultaneous study of structure and spatial orientation of lipids and membrane proteins. Biomed. Spectrosc. Imaging 4, 159–170. https://doi.org/10.3233/BSI-150104

Barlow, G.H., 1985. The determination of molecular weight distributions on heparin samples. Semin. Thromb. Hemost. 11, 26–28. https://doi.org/10.1055/s-2007-1004354

Barlow, G.H., Sanderson, N.D., McNeill, P.D., 1961. Macromolecular properties and biological activity of heparin. Arch. Biochem. Biophys. 94, 518–525. https://doi.org/10.1016/0003-9861(61)90080-7

Bartold, P.M., 1990. A biochemical and immunohistochemical study of the proteoglycans of alveolar bone. J. Dent. Res. 69, 7–19. https://doi.org/10.1177/00220345900690011201

Bastow, E.R., Byers, S., Golub, S.B., Clarkin, C.E., Pitsillides, A.A., Fosang, A.J., 2008. Hyaluronan synthesis and degradation in cartilage and bone. Cell. Mol. Life Sci. 65, 395–413. https://doi.org/10.1007/s00018-007-7360-z

Baytas, S.N., Linhardt, R.J., 2020. Advances in the preparation and synthesis of heparin and related products. Drug Discov. Today 25, 2095–2109. https://doi.org/10.1016/j.drudis.2020.09.011

Beccati, D., Roy, S., Yu, F., Gunay, N.S., Capila, I., Lech, M., Linhardt, R.J., Venkataraman, G., 2010. Identification of a novel structure in heparin generated by potassium permanganate oxidation. Carbohydr. Polym. 82, 699–705. https://doi.org/10.1016/j.carbpol.2010.05.038

Becker, D.J., Lowe, J.B., 2003. Fucose: biosynthesis and biological function in mammals. Glycobiology 13, 41R-53R. https://doi.org/10.1093/glycob/cwg054

Beirne, J., Truchan, H., Rao, L., 2011. Development and qualification of a size exclusion chromatography coupled with multiangle light scattering method for molecular weight determination of unfractionated heparin. Anal. Bioanal. Chem. 399, 717–725. https://doi.org/10.1007/s00216-010-4187-5

BeMiller, J.N., 2018. Carbohydrate Chemistry for Food Scientists. Elsevier.

Beni, S., Limtiaco, J.F.K., Larive, C.K., 2011. Analysis and characterization of heparin impurities. Anal. Bioanal. Chem. 399, 527–539. https://doi.org/10.1007/s00216-010-4121-x

Bertini, S., Bisio, A., Torri, G., Bensi, D., Terbojevich, M., 2005a. Molecular Weight Determination of Heparin and Dermatan Sulfate by Size Exclusion Chromatography with a Triple Detector Array. Biomacromolecules 6, 168–173. https://doi.org/10.1021/bm049693s

Bertini, S., Risi, G., Guerrini, M., Carrick, K., Szajek, A.Y., Mulloy, B., 2017. Molecular Weights of Bovine and Porcine Heparin Samples: Comparison of Chromatographic Methods and Results of a Collaborative Survey. Mol. Basel Switz. 22, 1–10. https://doi.org/10.3390/molecules22071214

Bett, C., Andrews, O., Asher, D.M., Pilant, T., Keire, D., Gregori, L., 2020. Eliminating Spiked Bovine Spongiform Encephalopathy Agent Activity from Heparin. Emerg. Infect. Dis. 26, 2478–2480. https://doi.org/10.3201/eid2610.200142

Beyer, T., Diehl, B., Randel, G., Humpfer, E., Schäfer, H., Spraul, M., Schollmayer, C., Holzgrabe, U., 2008. Quality assessment of unfractionated heparin using 1H nuclear magnetic resonance spectroscopy. J. Pharm. Biomed. Anal. 48, 13–19. https://doi.org/10.1016/j.jpba.2008.06.014

Bhilocha, S., Amin, R., Pandya, M., Yuan, H., Tank, M., LoBello, J., Shytuhina, A., Wang, W., Wisniewski, H.-G., de la Motte, C., Cowman, M.K., 2011. Agarose and polyacrylamide gel electrophoresis methods for molecular mass analysis of 5- to 500-kDa hyaluronan. Anal. Biochem. 417, 41–49. https://doi.org/10.1016/j.ab.2011.05.026

BHRA, 2019. Brazillian Pharmacopoeia 6th edition.

Bianchini, P., Liverani, L., Mascellani, G., Parma, B., 1997. Heterogeneity of unfractionated heparins studied in connection with species, source, and production processes. Semin. Thromb. Hemost. 23, 3–10. https://doi.org/10.1055/s-2007-996063

Bianchini, P., Osima, B., Parma, B., Nader, H.B., Dietrich, C.P., Casu, B., Torri, G., 1985. Fractionation and structural features of two heparin families with high antithrombotic, antilipemic and anticoagulant activities. Arzneimittelforschung. 35, 1215–1219.

Biancolillo, A., Marini, F., 2018. Chemometric Methods for Spectroscopy-Based Pharmaceutical Analysis. Front. Chem. 6.

Bigler, P., Brenneisen, R., 2009. Improved impurity fingerprinting of heparin by high resolution 1H NMR spectroscopy. J. Pharm. Biomed. Anal. 49, 1060–1064. https://doi.org/10.1016/j.jpba.2009.01.017

Bishop, C.M., 2006. Pattern Recognition and Machine Learning. Springer.

Bisio, A., Vecchietti, D., Citterio, L., Guerrini, M., Raman, R., Bertini, S., Eisele, G., Naggi, A., Sasisekharan, R., Torri, G., 2009. Structural features of low-molecular-weight heparins affecting their affinity to antithrombin. Thromb. Haemost. 102, 865–873. https://doi.org/10.1160/TH09-02-0081

Bitter, T., Muir, H.M., 1962. A modified uronic acid carbazole reaction. Anal. Biochem. 4, 330–334. https://doi.org/10.1016/0003-2697(62)90095-7

Blossom, D.B., Kallen, A.J., Patel, P.R., Elward, A., Robinson, L., Gao, G., Langer, R., Perkins, K.M., Jaeger, J.L., Kurkjian, K.M., Jones, M., Schillie, S.F., Shehab, N., Ketterer, D., Venkataraman, G., Kishimoto, T.K., Shriver, Z., McMahon, A.W., Austen, K.F., Kozlowski, S., Srinivasan, A., Turabelidze, G., Gould, C.V., Arduino, M.J., Sasisekharan, R., 2008. Outbreak of Adverse Reactions Associated with Contaminated Heparin. N. Engl. J. Med. 359, 2674–2684. https://doi.org/10.1056/NEJMoa0806450

Blundell, C.D., DeAngelis, P.L., Day, A.J., Almond, A., 2004. Use of 15N-NMR to resolve molecular details in isotopically-enriched carbohydrates: sequence-specific observations in hyaluronan oligomers up to decasaccharides. Glycobiology 14, 999–1009. https://doi.org/10.1093/glycob/cwh117

Boyle, M.J., Skidmore, M., Dickerman, B., Cooper, L., Devlin, A., Yates, E., Horrocks, P., Freeman, C., Chai, W., Beeson, J.G., 2017. Identification of Heparin Modifications and Polysaccharide Inhibitors of Plasmodium falciparum Merozoite Invasion That Have Potential for Novel Drug Development. Antimicrob. Agents Chemother. 61, e00709-17. https://doi.org/10.1128/AAC.00709-17

Boyraz, B., Saatz, J., Pompös, I.-M., Gad, M., Dernedde, J., Maier, A.-K.B., Moscovitz, O., Seeberger, P.H., Traub, H., Tauber, R., 2022. Imaging Keratan Sulfate in Ocular Tissue Sections by Immunofluorescence Microscopy and LA-ICP-MS. ACS Appl. Bio Mater. 5, 853–861. https://doi.org/10.1021/acsabm.1c01240

Brézillon, S., Untereiner, V., Mohamed, H.T., Hodin, J., Chatron-Colliet, A., Maquart, F.-X., Sockalingum, G.D., 2017. Probing glycosaminoglycan spectral signatures in live cells and their conditioned media by Raman microspectroscopy. Analyst 142, 1333–1341. https://doi.org/10.1039/C6AN01951J

Brito, A.S., Cavalcante, R.S., Cavalheiro, R.P., Palhares, L.C.G.F., Nobre, L.T.D.B., Andrade, G.P.V., Nader, H.B., Lima, M.A., Chavante, S.F., 2018. Anti-IIa activity and antitumor properties of a hybrid heparin/heparan sulfate-like compound from Litopenaeus vannamei shrimp. Int. J. Biol. Macromol. 118, 1470–1478. https://doi.org/10.1016/j.ijbiomac.2018.06.143

Brito, A.S., Cavalcante, R.S., Palhares, L.C.G.F., Hughes, A.J., Andrade, G.P.V., Yates, E.A., Nader, H.B., Lima, M.A., Chavante, S.F., 2014. A non-hemorrhagic hybrid heparin/heparan sulfate with anticoagulant potential. Carbohydr. Polym. 99, 372–378. https://doi.org/10.1016/j.carbpol.2013.08.063

Brustkern, A.M., Buhse, L.F., Nasr, M., Al-Hakim, A., Keire, D.A., 2010. Characterization of currently marketed heparin products: Reversed-phase ion-pairing liquid chromatography mass spectrometry of heparin digests. Anal. Chem. 82, 9865–9870. https://doi.org/10.1021/ac102301j

[194]

Burmistrova, N.A., Diehl, B.W.K., Soboleva, P.M., Rubtsova, E., Legin, E.A., Legin, A.V., Kirsanov, D.O., Monakhova, Y.B., 2020. Quality Control of Heparin Injections: Comparison of Four Established Methods. Anal. Sci. 36, 1467–1472. https://doi.org/10.2116/analsci.20P214

Burmistrova, N.A., Soboleva, P.M., Monakhova, Y.B., 2021. Is infrared spectroscopy combined with multivariate analysis a promising tool for heparin authentication? J. Pharm. Biomed. Anal. 194, 113811. https://doi.org/10.1016/j.jpba.2020.113811

Burson, S.L., Fahrenbach, M.J., Frommhagen, L.H., Riccardi, B.A., Brown, R.A., Brockman, J.A., Lewry, H.V., Stokstad, E.L.R., 1956. Isolation and Purification of Mactins, Heparin-like Anticoagulants from Mollusca1. J. Am. Chem. Soc. 78, 5874–5878. https://doi.org/10.1021/ja01603a045

Busse-Wicher, M., Wicher, K.B., Kusche-Gullberg, M., 2014. The extostosin family: Proteins with many functions. Matrix Biol., Proteoglycan Biology 35, 25–33. https://doi.org/10.1016/j.matbio.2013.10.001

Byler, D.M., Wilson, R.M., Randall, C.S., Sokoloski, T.D., 1995. Second Derivative Infrared Spectroscopy as a Non-Destructive Tool to Assess the Purity and Structural Integrity of Proteins. Pharm. Res. 12, 446–450. https://doi.org/10.1023/A:1016273122944

Cabassi, F., Casu, B., Perlin, A.S., 1978. Infrared absorption and raman scattering of sulfate groups of heparin and related glycosaminoglycans in aqueous solution. Carbohydr. Res. 63, 1–11. https://doi.org/10.1016/S0008-6215(00)80924-6

Calabro, A., Benavides, M., Tammi, M., Hascall, V.C., Midura, R.J., 2000. Microanalysis of enzyme digests of hyaluronan and chondroitin/dermatan sulfate by fluorophore-assisted carbohydrate electrophoresis (FACE). Glycobiology 10, 273–281. https://doi.org/10.1093/glycob/10.3.273

Cameron, J.M., Bruno, C., Parachalil, D.R., Baker, M.J., Bonnier, F., Butler, H.J., Byrne, H.J., 2020. Chapter 10 - Vibrational spectroscopic analysis and quantification of proteins in human blood plasma and serum, in: Ozaki, Y., Baranska, M., Lednev, I.K., Wood, B.R. (Eds.), Vibrational Spectroscopy in Protein Research. Academic Press, pp. 269–314. https://doi.org/10.1016/B978-0-12-818610-7.00010-4

Campo, V.L., Kawano, D.F., Silva, D.B. da, Carvalho, I., 2009. Carrageenans: Biological properties, chemical modifications and structural analysis – A review. Carbohydr. Polym. 77, 167–180. https://doi.org/10.1016/j.carbpol.2009.01.020

Casalone, C., Hope, J., 2018. Chapter 7 - Atypical and classic bovine spongiform encephalopathy, in: Pocchiari, M., Manson, J. (Eds.), Handbook of Clinical Neurology, Human Prion Diseases. Elsevier, pp. 121–134. https://doi.org/10.1016/B978-0-444-63945-5.00007-6

Casu, B., 1985. Structure and Biological Activity of Heparin. Adv. Carbohydr. Chem. Biochem. 43, 51–134. https://doi.org/10.1016/S0065-2318(08)60067-0

Casu, B., Guerrini, M., Naggi, A., Torri, G., De-Ambrosi, L., Boveri, G., Gonella, S., Cedro, A., Ferró, L., Lanzarotti, E., Paterno, M., Attolini, M., Valle, M.G., 1996. Characterization of sulfation patterns of beef and pig mucosal heparins by nuclear magnetic resonance spectroscopy. Arzneimittelforschung. 46, 472–477.

Casu, B., Oreste, P., Torri, G., Zoppetti, G., Choay, J., Lormeau, J.C., Petitou, M., Sinaε, P., 1981. The structure of heparin oligosaccharide fragments with high anti-(factor Xa) activity

containing the minimal antithrombin III-binding sequence Chemical and13C nuclear-magnetic-resonance studies. Biochem. J. 197, 599–609. https://doi.org/10.1042/bj1970599

Casu, B., Scovenna, G., Cifonelli, A.J., Perlin, A.S., 1978. Infrared spectra of glycosaminoglycans in deuterium oxide and deuterium chloride solution: Quantitative evaluation of uronic acid and acetamidodeoxyhexose moieties. Carbohydr. Res. 63, 13–27. https://doi.org/10.1016/S0008-6215(00)80925-8

Caterson, B., Melrose, J., 2018. Keratan sulfate, a complex glycosaminoglycan with unique functional capability | Glycobiology | Oxford Academic 28, 182–206.

Cesaretti, M., Luppi, E., Maccari, F., Volpi, N., 2003. A 96-well assay for uronic acid carbazole reaction. Carbohydr. Polym. 54, 59–61. https://doi.org/10.1016/S0144-8617(03)00144-9

Chan, K.L.A., Lekkas, I., Frogley, M.D., Cinque, G., Altharawi, A., Bello, G., Dailey, L.A., 2020. Synchrotron Photothermal Infrared Nanospectroscopy of Drug-Induced Phospholipidosis in Macrophages. Anal. Chem. 92, 8097–8107. https://doi.org/10.1021/acs.analchem.9b05759

Chandrasekhar, S., Esterman, M.A., Hoffman, H.A., 1987. Microdetermination of proteoglycans and glycosaminoglycans in the presence of guanidine hydrochloride. Anal. Biochem. 161, 103–108. https://doi.org/10.1016/0003-2697(87)90658-0

Charles, A.F., Scott, D.A., 1936. Studies on heparin: Observations on the chemistry of heparin. Biochem. J. 30, 1927–1933. https://doi.org/10.1042/bj0301927

Chavante, S.F., Brito, A.S., Lima, M., Yates, E., Nader, H., Guerrini, M., Torri, G., Bisio, A., 2014. A heparin-like glycosaminoglycan from shrimp containing high levels of 3-O-sulfated d-glucosamine groups in an unusual trisaccharide sequence. Carbohydr. Res. 390, 59–66. https://doi.org/10.1016/j.carres.2014.03.002

Chen, H., Lin, Z., Tan, C., 2018. Nondestructive Discrimination of Pharmaceutical Preparations Using Near-Infrared Spectroscopy and Partial Least-Squares Discriminant Analysis. Anal. Lett. 51, 564–574. https://doi.org/10.1080/00032719.2017.1339070

Chen, J., Yu, Y., Fareed, J., Hoppensteadt, D., Jeske, W., Kouta, A., Caijuan, J., Jin, Y., Yao, Y., Xia, K., Zhang, F., Chen, S., Ye, X., Linhardt, R.J., 2019. Comparison of Low-Molecular-Weight Heparins Prepared From Ovine Heparins With Enoxaparin. Clin. Appl. Thromb. 25, 1–9. https://doi.org/10.1177/1076029616686422

Chi, L., Amster, J., Linhardt, R., 2005. Mass Spectrometry for the Analysis of Highly Charged Sulfated Carbohydrates. Curr. Anal. Chem. 1, 223–240. https://doi.org/10.2174/157341105774573929

Choay, J., 1989. Chemically Synthesized Heparin-Derived Oligosaccharides. Ann. N. Y. Acad. Sci. 556, 61–74. https://doi.org/10.1111/j.1749-6632.1989.tb22490.x

Choay, J., Lormeau, J.C., Petitou, M., Sinaÿ, P., Fareed, J., 1981. Structural studies on a biologically active hexasaccharide obtained from heparin. Ann. N. Y. Acad. Sci. 370, 644–649. https://doi.org/10.1111/j.1749-6632.1981.tb29770.x

Choay, J., Petitou, M., Lormeau, J.C., Sinaÿ, P., Casu, B., Gatti, G., 1983. Structure-activity relationship in heparin: A synthetic pentasaccharide with high affinity for antithrombin III and eliciting high anti-factor Xa activity. Biochem. Biophys. Res. Commun. 116, 492–499. https://doi.org/10.1016/0006-291X(83)90550-8

Christianson, H.C., Belting, M., 2014. Heparan sulfate proteoglycan as a cell-surface endocytosis receptor. Matrix Biol., Proteoglycan Biology 35, 51–55. https://doi.org/10.1016/j.matbio.2013.10.004

Cifonelli, J.A., Ludowieg, J., Dorfman, A., 1958. Chemistry of beta-heparin (chondroitinsulfuric acid-B). J. Biol. Chem. 233, 541–545.

Cohen, J., Cohen, P., West, S.G., Aiken, L.S., 2002. Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences, 3rd ed. Routledge, New York. https://doi.org/10.4324/9780203774441

Colombo, E., Mauri, L., Marinozzi, M., Rudd, T.R., Yates, E.A., Ballabio, D., Guerrini, M., 2022. NMR spectroscopy and chemometric models to detect a specific non-porcine ruminant contaminant in pharmaceutical heparin. J. Pharm. Biomed. Anal. 214, 114724. https://doi.org/10.1016/j.jpba.2022.114724

Colon, L.A., Dadoo, Rajeev., Zare, R.N., 1993. Determination of carbohydrates by capillary zone electrophoresis with amperometric detection at a copper microelectrode. Anal. Chem. 65, 476–481. https://doi.org/10.1021/ac00052a027

Cooney, C.A., Jousheghany, F., Yao-Borengasser, A., Phanavanh, B., Gomes, T., Kieber-Emmons, A.M., Siegel, E.R., Suva, L.J., Ferrone, S., Kieber-Emmons, T., Monzavi-Karbassi, B., 2011. Chondroitin sulfates play a major role in breast cancer metastasis: a role for CSPG4 and CHST11gene expression in forming surface P-selectin ligands in aggressive breast cancer cells. Breast Cancer Res. 13, R58. https://doi.org/10.1186/bcr2895

Corbier, A., Le Berre, N., Rampe, D., Meng, H., Lorenz, M., Vicat, P., Potdevin, S., Doubovetzky, M., 2011. Oversulfated Chondroitin Sulfate and OSCS-Contaminated Heparin Cause Dose- and Route-Dependent Hemodynamic Effects in the Rat. Toxicol. Sci. 121, 417–427. https://doi.org/10.1093/toxsci/kfr072

Cowman, M.K., Chen, C.C., Pandya, M., Yuan, H., Ramkishun, D., LoBello, J., Bhilocha, S., Russell-Puleri, S., Skendaj, E., Mijovic, J., Jing, W., 2011. Improved agarose gel electrophoresis method and molecular mass calculation for high molecular mass hyaluronan. Anal. Biochem. 417, 50–56. https://doi.org/10.1016/j.ab.2011.05.023

da Cunha, A.L., de Oliveira, L.G., Maia, L.F., de Oliveira, L.F.C., Michelacci, Y.M., de Aguiar, J.A.K., 2015. Pharmaceutical grade chondroitin sulfate: Structural analysis and identification of contaminants in different commercial preparations. Carbohydr. Polym. 134, 300–308. https://doi.org/10.1016/j.carbpol.2015.08.006

Dammann, K., Gifford, A., Kelley, K., Stawicki, S.P., 2020. Operative Hemostasis in Trauma and Acute Care Surgery: The Role of Biosurgical Agents. IntechOpen. https://doi.org/10.5772/intechopen.94100

David-Raoudi, M., Deschrevel, B., Leclerq, S., Galéra, P., Boumediene, K., Pujol, J.-P., 2009. Chondroitin sulfate increases hyaluronan production by human synoviocytes through differential regulation of hyaluronan synthases: Role of p38 and Akt - David-Raoudi - 2009 - Arthritis & Rheumatism - Wiley Online Library. Arthritis Rheum. 60, 760–770.

Debarnot, C., Monneau, Y.R., Roig-Zamboni, V., Delauzun, V., Le Narvor, C., Richard, E., Hénault, J., Goulet, A., Fadel, F., Vivès, R.R., Priem, B., Bonnaffé, D., Lortat-Jacob, H., Bourne, Y., 2019. Substrate binding mode and catalytic mechanism of human heparan sulfate d-glucuronyl C5 epimerase. Proc. Natl. Acad. Sci. 116, 6760–6765. https://doi.org/10.1073/pnas.1818333116

Deepa, S.S., Yamada, S., Zako, M., Goldberger, O., Sugahara, K., 2004. Chondroitin Sulfate Chains on Syndecan-1 and Syndecan-4 from Normal Murine Mammary Gland Epithelial Cells Are Structurally and Functionally Distinct and Cooperate with Heparan Sulfate Chains to Bind Growth Factors: A NOVEL FUNCTION TO CONTROL BINDING OF MIDKINE, PLEIOTROPHIN, AND BASIC FIBROBLAST GROWTH FACTOR *. J. Biol. Chem. 279, 37368–37376. https://doi.org/10.1074/jbc.M403031200

Deligny, A., Dierker, T., Dagälv, A., Lundequist, A., Eriksson, I., Nairn, A.V., Moremen, K.W., Merry, C.L.R., Kjellén, L., 2016. NDST2 (N-Deacetylase/N-Sulfotransferase-2) Enzyme Regulates Heparan Sulfate Chain Length *♦. J. Biol. Chem. 291, 18600–18607. https://doi.org/10.1074/jbc.M116.744433

Demchick, P., Koch, A.L., 1996. The permeability of the wall fabric of Escherichia coli and Bacillus subtilis. J. Bacteriol. 178, 768–773. https://doi.org/10.1128/jb.178.3.768-773.1996

Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum likelihood from incomplete data via the EM algorithm. J. R. Stat. Soc. Ser. B 39, 1–38.

Dennis, J.W., Granovsky, M., Warren, C.E., 1999. Protein glycosylation in development and disease. BioEssays 21, 412–421. https://doi.org/10.1002/(SICI)1521-1878(199905)21:5<412::AID-BIES8>3.0.CO;2-5

Desai, U.R., Linhardt, R.J., 1995. Molecular Weight of Heparin Using. J. Pharm. Sci. 84, 212–215.

Devlin, A., Mycroft-West, C., Guerrini, M., Yates, E., Skidmore, M., 2019a. Analysis of solid-state heparin samples by ATR-FTIR spectroscopy. bioRxiv 538074. https://doi.org/10.1101/538074

Devlin, A., Mycroft-West, C., Procter, P., Cooper, L., Guimond, S., Lima, M., Yates, E., Skidmore, M., 2019b. Tools for the Quality Control of Pharmaceutical Heparin. Medicina (Mex.) 55, 636. https://doi.org/10.3390/medicina55100636

Dey, P., Saphos, C.A., McDonnell, J., Moore, V.L., 1992. Studies on the quantification of proteoglycans by the dimethylmethylene blue dye-binding method. Specificity, quantitation in synovial lavage fluid, and automation. Connect. Tissue Res. 28, 317–324. https://doi.org/10.3109/03008209209016823

Dietrich, C.P., de Paiva, JoséF., Moraes, C.T., Takahashi, H.K., Porcionatto, M.A., Nader, H.B., 1985. Isolation and characterization of a heparin with high anticoagulant activity from Anomalocardia brasiliana. Biochim. Biophys. Acta BBA - Gen. Subj. 843, 1–7. https://doi.org/10.1016/0304-4165(85)90041-8

Dietrich, C.P., McDuffie, N.M., Sampaio, L.O., 1977. Identification of acidic mucopolysaccharides by agarose gel electrophoresis. J. Chromatogr. A 130, 299–304. https://doi.org/10.1016/S0021-9673(00)89809-X

Ding, Y., Shi, L., Wei, H., 2015. A "turn on" fluorescent probe for heparin and its oversulfated chondroitin sulfate contaminant. Chem. Sci. 6, 6361–6366. https://doi.org/10.1039/c5sc01675d

Ding, Y., Zhou, M., Wei, H., 2017. A supercharged fluorescent protein based FRET sensing platform for detection of heparin contamination. Anal. Methods 9, 5593–5597. https://doi.org/10.1039/c7ay01907f

Dische, Z., Borenfreund, E., 1950. A spectrophotometric method for the microdetermination of hexosamines. J. Biol. Chem. 184, 517–522.

Dobson, C.M., Hempel, S.J., Stalnaker, S.H., Stuart, R., Wells, L., 2013. O-Mannosylation and human disease. Cell. Mol. Life Sci. 70, 2849–2857. https://doi.org/10.1007/s00018-012-1193-0

Dodgson, K., Price, R., 1962. A note on the determination of the ester sulphate content of sulphated polysaccharides | Biochemical Journal | Portland Press. Biochem. J. 84, 106–110.

Doherty, J., Raoof, A., Hussain, A., Wolna, M., Cinque, G., Brown, M., Gardner, P., Denbigh, J., 2019. Live single cell analysis using synchrotron FTIR microspectroscopy: development of a simple dynamic flow system for prolonged sample viability. Analyst 144, 997–1007.

Doi, A., Matsumoto, I., Seno, N., 1983. Fluorescence Spectral Studies on the Specific Interaction between Sulfated Glycosaminoglycans and Potato Lectin. J. Biochem. (Tokyo) 93, 771–775. https://doi.org/10.1093/jb/93.3.771

Dong, Q., Guo, X., Li, L., Yu, C., Nie, L., Tian, W., Zhang, H., Huang, S., Zang, H., 2020. Understanding hyaluronic acid induced variation of water structure by near-infrared spectroscopy. Sci. Rep. 10, 1387. https://doi.org/10.1038/s41598-020-58417-5

Dong, Q., Zang, H., Liu, A., Yang, G., Sun, C., Sui, L., Wang, P., Li, L., 2010. Determination of molecular weight of hyaluronic acid by near-infrared spectroscopy. J. Pharm. Biomed. Anal. 53, 274–278. https://doi.org/10.1016/j.jpba.2010.05.031

Dong, Q., Zang, H., Zang, L., Liu, A., Shi, Y., Zhang, H., 2014. Rapid determination of hyaluronic acid concentration in fermentation broth with near-infrared spectroscopy. J. Innov. Opt. Health Sci. 07, 1450012. https://doi.org/10.1142/S1793545814500126

Donghui, F., Beibei, W., Zheng, X., Qisheng, G., 2006. Determination of hyaluronan by spectroscopic methods. J. Wuhan Univ. Technol.-Mater Sci Ed 21, 32–34. https://doi.org/10.1007/BF02840874

Dreyfuss, J.L., Regatieri, C.V., Jarrouge, T.R., Cavalheiro, R.P., Sampaio, L.O., Nader, H.B., 2009. Heparan sulfate proteoglycans: structure, protein interactions and cell signaling. An. Acad. Bras. Ciênc. 81, 409–429. https://doi.org/10.1590/S0001-37652009000300007

Duan, Z.H., Cheng, C.Y., Hai, Y., Wang, J.L., 2013. Determination and Identification of Chondroitin Sulfate from Tilapia Byproducts. Adv. Mater. Res. 690–693, 1318–1321. https://doi.org/10.4028/www.scientific.net/AMR.690-693.1318

Dwek, R.A., 1996. Glycobiology: Toward Understanding the Function of Sugars. Chem. Rev. 96, 683–720. https://doi.org/10.1021/cr940283b

Easson, M.L.A.E., Malka, O., Paetz, C., Hojná, A., Reichelt, M., Stein, B., van Brunschot, S., Feldmesser, E., Campbell, L., Colvin, J., Winter, S., Morin, S., Gershenzon, J., Vassão, D.G., 2021. Activation and detoxification of cassava cyanogenic glucosides by the whitefly Bemisia tabaci. Sci. Rep. 11, 13244. https://doi.org/10.1038/s41598-021-92553-w

Edens, R.E., Al-Hakim, A., Weiler, J.M., Rethwisch, D.G., Fareed, J., Linhardt, R.J., 1992. Gradient polyacrylamide gel electrophoresis for determination of molecular weights of heparin preparations and low-molecular-weight heparin derivatives. J. Pharm. Sci. 81, 823–827. https://doi.org/10.1002/jps.2600810821

Eiber, H.B., Danishefky, I., 1957. The incorporation of S35-sulfate into heparin. J. Biol. Chem. 226.

Eldridge, S.L., Higgins, L.A., Dickey, B.J., Larive, C.K., 2009. Insights into the Capillary Electrophoresis Separation of Heparin Disaccharides from Nuclear Magnetic Resonance,

pKa, and Electrophoretic Mobility Measurements. Anal. Chem. 81, 7406–7415. https://doi.org/10.1021/ac901218q

Elsheikha, H.M., Elsaied, N.A., Chan, K.L.A., Brignell, C., Harun, M.S.R., Wehbe, K., Cinquee, G., 2019. Label-free characterization of biochemical changes within human cells under parasite attack using synchrotron based micro-FTIR. Anal. Methods 11, 2518–2530. https://doi.org/10.1039/C8AY02777C

EP, 2010. European Pharmacopoeia.

Essendoubi, M., Gobinet, C., Reynaud, R., Angiboust, J.F., Manfait, M., Piot, O., 2016. Human skin penetration of hyaluronic acid of different molecular weights as probed by Raman spectroscopy. Skin Res. Technol. 22, 55–62. https://doi.org/10.1111/srt.12228

Fareed, J., Jeske, W., Ramacciotti, E., 2019. Porcine Mucosal Heparin Shortage Crisis! What Are the Options? Clin. Appl. Thromb. 25, 1076029619878786. https://doi.org/10.1177/1076029619878786

Fareed, J., Nader, H., Lima, M., Hoppensteadt, D., Walenga, J., Jeske, W., Kumar, E., Raake, W., Bakhos, M., 2015. Bovine and Porcine Mucosal Heparins Exhibit Similar Biologic Profiles. FASEB J. 29, 610.4. https://doi.org/10.1096/fasebj.29.1_supplement.610.4

Fasciano, J.M., Danielson, N.D., 2016. Ion chromatography for the separation of heparin and structurally related glycoaminoglycans: A review. J. Sep. Sci. 39, 1118–1129. https://doi.org/10.1002/jssc.201500664

FDA, 2013. Heparin for Drug and Medical Device Use: Monitoring Crude Heparin for Quality.

Fernandes, A., 2018a. Heparin Sodium.

Fernandes, A., 2018b. Heparin Sodium Monograph.

Ferris, S.P., Kodali, V.K., Kaufman, R.J., 2014. Glycoprotein folding and quality-control mechanisms in protein-folding diseases | Disease Models & Mechanisms | The Company of Biologists. Dis. Model. Mech. 7, 331–341.

Finney, N.S., Siegel, J.S., 2008. In Memoriam: Albert Hofmann (1906–2008). CHIMIA 62, 444–444. https://doi.org/10.2533/chimia.2008.444

Foot, M., Mulholland, M., 2005. Classification of chondroitin sulfate A, chondroitin sulfate C, glucosamine hydrochloride and glucosamine 6 sulfate using chemometric techniques. J. Pharm. Biomed. Anal. 38, 397–407. https://doi.org/10.1016/j.jpba.2005.01.026

Frazier, S.B., Roodhouse, K.A., Hourcade, D.E., Zhang, L., 2008. The Quantification of Glycosaminoglycans: A Comparison of HPLC, Carbazole, and Alcian Blue Methods/title>. Open Glycosci. 1.

Freeman, L., Posthuma, R., Gordon, L., Marx, W., 1957. Determination of tissue heparin. Arch. Biochem. Biophys. 70, 169–177. https://doi.org/10.1016/0003-9861(57)90090-5

Fu, L., Li, G., Yang, B., Onishi, A., Li, L., Sun, P., Zhang, F., Linhardt, R.J., 2013a. Structural Characterization of Pharmaceutical Heparins Prepared from Different Animal Tissues. J. Pharm. Sci. 102, 1447–1457. https://doi.org/10.1002/jps.23501

Fu, L., Li, G., Yang, B., Onishi, A., Li, L., Sun, P., Zhang, F., Linhardt, R.J., 2013b. Structural Characterization of Pharmaceutical Heparins Prepared from Different Animal Tissues. J. Pharm. Sci. 102, 1447–1457. https://doi.org/10.1002/jps

Fu, L., Suflita, M., Linhardt, R.J., 2016. Bioengineered heparins and heparan sulfates. Adv. Drug Deliv. Rev., Extracellular Matrix (ECM) and ECM-like materials: Therapeutic Tools and Targets in Cancer Treatment 97, 237–249. https://doi.org/10.1016/j.addr.2015.11.002

Funderburgh, J.L., 2002. Keratan Sulfate Biosynthesis. IUBMB Life 54, 187–194. https://doi.org/10.1080/15216540214932

Funderburgh, J.L., Chandler, J.W., 1978. An agarose gel electrophoretic method for analysis of sulfated glycosaminoglycans of cultured cells. Anal. Biochem. 91, 464–472. https://doi.org/10.1016/0003-2697(78)90532-8

Li, G., Cai, C., Li, L., Fu, L., Chang, Y., Zhang, F., Toida, T., Xue, C., Linhardt, R., 2014. Method to detect contaminants in heparin using radical depolymerization and liquid chromatography-mass spectrometry. Anal. Chem. 86, 326–330. http://dx.doi.org/10.1021/ac403625a

Gallagher, J.T., 2001. Heparan sulfate: growth control with a restricted sequence menu. J. Clin. Invest. 108, 357–361. https://doi.org/10.1172/JCI13713

Gauglitz, G., Vo-Dinh, T., 2006. Handbook of Spectroscopy. John Wiley & Sons.

Gemst, J.J. van, Loeven, M.A., Graaf, M.J.J. de, Berden, J.H.M., Rabelink, T.J., Smit, C.H., Vlag, J. van der, 2016. RNA Contaminates Glycosaminoglycans Extracted from Cells and Tissues. PLOS ONE 11, e0167336. https://doi.org/10.1371/journal.pone.0167336

Georgakopoulos, C.D., Makri, O.E., Pagoulatos, D., Karamanos, N.K., 2019. Expression and Localization of Glycosaminoglycans/Proteoglycan in Pterygium: An Immunohistochemical Study. Med. Hypothesis Discov. Innov. Ophthalmol. 8, 39–43.

Ghann, W., Uddin, J., 2017. Terahertz (THz) Spectroscopy: A Cutting-Edge Technology | IntechOpen, Terahertz Spectroscopy: A Cutting Edge Technology.

Ghasemi, P., Aslani, M., Rollins, D.K., Williams, R.C., 2019. Principal component analysis-based predictive modeling and optimization of permanent deformation in asphalt pavement: elimination of correlated inputs and extrapolation in modeling. Struct. Multidiscip. Optim. 59, 1335–1353. https://doi.org/10.1007/s00158-018-2133-x

Gierlinger, N., 2018. New insights into plant cell walls by vibrational microspectroscopy. Appl. Spectrosc. Rev. 53, 517–551. https://doi.org/10.1080/05704928.2017.1363052

Gieroba, B., Krysa, M., Wojtowicz, K., Wiater, A., Pleszczyńska, M., Tomczyk, M., Sroka-Bartnicka, A., 2020. The FT-IR and Raman Spectroscopies as Tools for Biofilm Characterization Created by Cariogenic Streptococci. Int. J. Mol. Sci. 21, 3811. https://doi.org/10.3390/ijms21113811

Gill, V.L., Aich, U., Rao, S., Pohl, C., Zaia, J., 2013. Disaccharide Analysis of Glycosaminoglycans Using Hydrophilic Interaction Chromatography and Mass Spectrometry. Anal. Chem. 85, 1138–1145. https://doi.org/10.1021/ac3030448

Gleadow, R.M., Møller, B.L., 2014. Cyanogenic Glycosides: Synthesis, Physiology, and Phenotypic Plasticity. Annu. Rev. Plant Biol. 65, 155–185. https://doi.org/10.1146/annurev-arplant-050213-040027

Glicksman, M., 1963. Utilization of Natural Polysaccharide Gums in the Food Industry, in: Chichester, C.O., Mrak, E.M., Stewart, G.F. (Eds.), Advances in Food Research. Academic Press, pp. 109–200. https://doi.org/10.1016/S0065-2628(08)60065-8

Goger, B., Halden, Y., Rek, A., Mösl, R., Pye, D., Gallagher, J., Kungl, A.J., 2002. Different Affinities of Glycosaminoglycan Oligosaccharides for Monomeric and Dimeric Interleukin-8: A Model for Chemokine Regulation at Inflammatory Sites. Biochemistry 41, 1640–1646. https://doi.org/10.1021/bi011944j

Gowda, D.C., 2006. Role of Chondroitin-4-Sulfate in Pregnancy-Associated Malaria, in: Advances in Pharmacology, Chondroitin Sulfate: Structure, Role and Pharmacological Activity. Academic Press, pp. 375–400. https://doi.org/10.1016/S1054-3589(05)53018-7

Grant, D., Long, W.F., Moffat, C.F., Williamson, F.B., 1991. Infrared spectroscopy of heparins suggests that the region 750-950 cm-1 is sensitive to changes in iduronate residue ring conformation. Biochem. J. 275, 193–197.

Grant, D, Long, W.F., Moffat, C.F., Williamson, F.B., 1989. Infrared spectroscopy of chemically modified heparins. Biochem. J. 261, 1035–1038. https://doi.org/10.1042/bj2611035

Grant, D, Long, W.F., Williamson, F.B., 1987a. Infrared spectroscopy of heparin-cation complexes. Biochem. J. 244, 143–149. https://doi.org/10.1042/bj2440143

Grant, D., Long, W.F., Williamson, F.B., 1987. A model of two conformational forms of heparins/heparans suggested by infrared spectroscopy. Med. Hypotheses 24, 131–136. https://doi.org/10.1016/0306-9877(87)90095-8

Grant, D, Long, W.F., Williamson, F.B., 1987b. Infrared spectroscopy of heparin-cation complexes. Biochem. J. 244, 143–149.

Grant, D., Moffat, C.F., Long, W.F., Williamson, F.B., 1989b. N-Acetyl amide-II and methyl C—H bending infrared absorbances of chemically modified heparins. Biochem. Soc. Trans. 17, 502–503. https://doi.org/10.1042/bst0170502

Grover, S.P., Mackman, N., 2019. Intrinsic Pathway of Coagulation and Thrombosis. Arterioscler. Thromb. Vasc. Biol. 39, 331–338. https://doi.org/10.1161/ATVBAHA.118.312130

Guan, Y., Xu, X., Liu, X., Sheng, A., Jin, L., Linhardt, R.J., Chi, L., 2016. Comparison of Low-Molecular-Weight Heparins Prepared From Bovine Lung Heparin and Porcine Intestine Heparin. J. Pharm. Sci. 105, 1843–1850. https://doi.org/10.1016/j.xphs.2016.03.037

Guerrini, M., Beccati, D., Shriver, Z., Naggi, A., Viswanathan, K., Bisio, A., Capila, I., Lansing, J.C., Guglieri, S., Fraser, B., Al-Hakim, A., Gunay, N.S., Zhang, Z., Robinson, L., Buhse, L., Nasr, M., Woodcock, J., Langer, R., Venkataraman, G., Linhardt, R.J., Casu, B., Torri, G., Sasisekharan, R., 2008. Oversulfated chondroitin sulfate is a contaminant in heparin associated with adverse clinical events. Nat. Biotechnol. 26, 669–675. https://doi.org/10.1038/nbt1407

Guerrini, M., Bisio, A., Torri, G., 2001. Combined Quantitative 1H and 13C Nuclear Magnetic Resonance Spectroscopy for Characterization of Heparin Preparationsy. Semin. Thromb. Hemost. 27, 473–482. https://doi.org/10.1016/b978-0-12-813608-9.00002-2

Guerrini, M., Zhang, Z., Shriver, Z., Naggi, A., Masuko, S., Langer, R., Casu, B., Linhardt, R.J., Torri, G., Sasisekharan, R., 2009. Orthogonal analytical approaches to detect potential contaminants in heparin. Proc. Natl. Acad. Sci. 106, 16956–16961. https://doi.org/10.1073/pnas.0906861106

Guimond, S., Maccarana, M., Olwin, B.B., Lindahl, U., Rapraeger, A.C., 1993. Activating and inhibitory heparin sequences for FGF-2 (basic FGF). Distinct requirements for FGF-1, FGF-

2, and FGF-4. J. Biol. Chem. 268, 23906–23914. https://doi.org/10.1016/S0021-9258(20)80471-2

Guimond, S.E., Puvirajesinghe, T.M., Skidmore, M.A., Kalus, I., Dierks, T., Yates, E.A., Turnbull, J.E., 2009. Rapid Purification and High Sensitivity Analysis of Heparan Sulfate from Cells and Tissues: TOWARD GLYCOMICS PROFILING *♦. J. Biol. Chem. 284, 25714–25722. https://doi.org/10.1074/jbc.M109.032755

Guimond, S.E., Turnbull, J.E., 1999. Fibroblast growth factor receptor signalling is dictated by specific heparan sulphate saccharides. Curr. Biol. 9, 1343–1346. https://doi.org/10.1016/S0960-9822(00)80060-3

Gulberti, S., Mao, X., Bui, C., Fournel-Gigleux, S., 2020. The role of heparan sulfate maturation in cancer: A focus on the 3O-sulfation and the enigmatic 3O-sulfotransferases (HS3STs). Semin. Cancer Biol., Translating Extracellular Matrix 62, 68–85. https://doi.org/10.1016/j.semcancer.2019.10.009

Guo, X., Condra, M., Kimura, K., Berth, G., Dautzenberg, H., Dubin, P.L., 2003. Determination of molecular weight of heparin by size exclusion chromatography with universal calibration. Anal. Biochem. 312, 33–39. https://doi.org/10.1016/S0003-2697(02)00428-1

Hahn, S., Yoon, G., 2006. Identification of pure component spectra by independent component analysis in glucose prediction based on mid-infrared spectroscopy. Appl. Opt. 45, 8374–8380. https://doi.org/10.1364/AO.45.008374

Hamerly, G., Elkan, C., 2002. Alternatives to the k-means algorithm that find better clusterings | Proceedings of the eleventh international conference on Information and knowledge management. Proc. 2002 ACM CIKM Int. Conf. Inf. Knowl. Manag.

Hand, K., Yates, E., 2017. Terahertz: dictating the frequency of life. Do macromolecular vibrational modes impose thermal limitations on terrestrial life? J. R. Soc. Interface 14, 20170673. https://doi.org/10.1098/rsif.2017.0673

Harada, N.S., Oyama, H.T., Bártoli, J.R., Gouvêa, D., Cestari, I.A., Hui Wang, S., 2005. Quantifying adsorption of heparin on a PVC substrate using ATR-FTIR. Polym. Int. 54, 209–214. https://doi.org/10.1002/pi.1685

Hart, G.W., Slawson, C., Ramirez-Correa, G., Lagerlof, O., 2011. Cross Talk Between O-GlcNAcylation and Phosphorylation: Roles in Signaling, Transcription, and Chronic Disease. Annu. Rev. Biochem. 80, 825–858. https://doi.org/10.1146/annurev-biochem-060608-102511

Hashii, N., Kawasaki, N., Itoh, S., Qin, Y., Fujita, N., Hattori, T., Miyata, K., Bando, A., Sekimoto, Y., Hama, T., Kashimura, M., Tatsumi, M., Mabuchi, K., Namekawa, H., Sakai, T., Hirose, M., Dobashi, S., Shimahashi, H., Koyama, S., Herr, S.O., Kawai, K., Yoden, H., Yamaguchi, T., 2010. Heparin identification test and purity test for OSCS in heparin sodium and heparin calcium by weak anion-exchange high-performance liquid chromatography. Biologicals 38, 539–543. https://doi.org/10.1016/j.biologicals.2010.04.002

He, W., Zhu, Y., Shirke, A., Sun, X., Liu, J., Gross, R.A., Koffas, M.A.G., Linhardt, R.J., Li, M., 2017. Expression of chondroitin-4-O-sulfotransferase in Escherichia coli and Pichia pastoris. Appl. Microbiol. Biotechnol. 101, 6919–6928. https://doi.org/10.1007/s00253-017-8411-5

Helenius, A., Aebi, M., 2004. Roles of N-linked glycans in the endoplasmic reticulum. Annu. Rev. Biochem. 73, 1019–1049. https://doi.org/10.1146/annurev.biochem.73.011303.073752

Hinton, G., Roweis, S., 2002. Stochastic neighbor embedding, in: Proceedings of the 15th International Conference on Neural Information Processing Systems, NIPS'02. MIT Press, Cambridge, MA, USA, pp. 857–864.

Hitchcock, A.M., Costello, C.E., Zaia, J., 2006. Glycoform Quantification of Chondroitin/Dermatan Sulfate Using a Liquid Chromatography−Tandem Mass Spectrometry Platform. Biochemistry 45, 2350–2361. https://doi.org/10.1021/bi052100t

Hitchcock, A.M., Yates, K.E., Shortkroff, S., Costello, C.E., Zaia, J., 2007. Optimized Extraction of Glycosaminoglycans from Normal and Osteoarthritic Cartilage for Glycomics Profiling. Glycobiology 17, 25–35. https://doi.org/10.1093/glycob/cwl046

Holdener, B.C., Haltiwanger, R.S., 2019. Protein O-fucosylation: structure and function. Curr. Opin. Struct. Biol., Sequences and Topology ● Carbohydrates 56, 78–86. https://doi.org/10.1016/j.sbi.2018.12.005

Holder, G.M., Bowfield, A., Surman, M., Suepfle, M., Moss, D., Tucker, C., Rudd, T.R., Fernig, D.G., Yates, E.A., Weightman, P., 2012. Fundamental differences in model cell-surface polysaccharides revealed by complementary optical and spectroscopic techniques. Soft Matter 8, 6521–6527. https://doi.org/10.1039/C2SM25239B

Hollingsworth, M.A., Swanson, B.J., 2004. Mucins in cancer: protection and control of the cell surface. Nat. Rev. Cancer 4, 45–60. https://doi.org/10.1038/nrc1251

Holman, J., Skidmore, M.A., Rudd, T.R., Yates, E.A., 2010. The latent ampholytic nature of glycosaminoglycan (GAG) oligosaccharides facilitates their separation by isoelectric focusing. Anal. Methods 2, 1550–1554. https://doi.org/10.1039/c0ay00340a

Hook, A.L., Hogwood, J., Gray, E., Mulloy, B., Merry, C.L.R., 2021. High sensitivity analysis of nanogram quantities of glycosaminoglycans using ToF-SIMS. Commun. Chem. 4, 1–8. https://doi.org/10.1038/s42004-021-00506-1

Hooper, L.V., Gordon, J.I., 2001. Glycans as legislators of host–microbial interactions: spanning the spectrum from symbiosis to pathogenicity. Glycobiology 11, 1R-10R. https://doi.org/10.1093/glycob/11.2.1R

Howell, W.H., Holt, E., 1918. Two new factors in blood coagulation—heparin and pro-antithrombin. Am. J. Physiol.-Leg. Content 47, 328–341. https://doi.org/10.1152/ajplegacy.1918.47.3.328

Hricovíni, M., Guerrini, M., Bisio, A., Torri, G., Naggi, A., Casu, B., 2002. Active conformations of glycosaminoglycans. NMR determination of the conformation of heparin sequences complexed with antithrombin and fibroblast growth factors in solution. Semin. Thromb. Hemost. 28, 325–334. https://doi.org/10.1055/s-2002-34301

Hu, Y., Guo, W., Ding, Y., Cheng, H., Wei, H., 2016. Modulating luminescence of Tb3+ with biomolecules for sensing heparin and its contaminant OSCS. Biosens. Bioelectron. 86, 858–863. https://doi.org/10.1016/j.bios.2016.07.085

Huang, S., Yang, W., Huang, G., 2020. Preparation and activities of selenium polysaccharide from plant such as Grifola frondosa. Carbohydr. Polym. 242, 116409. https://doi.org/10.1016/j.carbpol.2020.116409

Huang, Y., Toyoda, H., Koshiishi, I., Toida, T., Imanari, T., 1995. Determination of a Depolymerized Holothurian Glycosaminoglycan in Plasma after Intravenous Administration by Postcolumn HPLC. Chem. Pharm. Bull. (Tokyo) 43, 2182–2186. https://doi.org/10.1248/cpb.43.2182

Humphries, D.E., Wong, G.W., Friend, D.S., Gurish, M.F., Qiu, W.-T., Huang, C., Sharpe, A.H., Stevens, R.L., 1999. Heparin is essential for the storage of specific granule proteases in mast cells. Nature 400, 769–772. https://doi.org/10.1038/23481

Hunter, N., 2007. Scrapie—Uncertainties, biology and molecular approaches. Biochim. Biophys. Acta BBA - Mol. Basis Dis. 1772, 619–628. https://doi.org/10.1016/j.bbadis.2007.04.007

Imamura, M., Tabeta, N., Kato, N., Matsuura, Y., Iwamaru, Y., Yokoyama, T., Murayama, Y., 2016. Heparan Sulfate and Heparin Promote Faithful Prion Replication in Vitro by Binding to Normal and Abnormal Prion Proteins in Protein Misfolding Cyclic Amplification. J. Biol. Chem. 291, 26478–26486. https://doi.org/10.1074/jbc.M116.745851

Ingle, R.G., Agarwal, A.S., 2014. A world of low molecular weight heparins (LMWHs) enoxaparin as a promising moiety - A review. Carbohydr. Polym. 106, 148–153. https://doi.org/10.1016/j.carbpol.2014.01.100

Iozzo, R.V., Kovalszky, I., Hacobian, N., Schick, P.K., Ellingson, J.S., Dodge, G.R., 1990. Fatty acylation of heparan sulfate proteoglycan from human colon carcinoma cells. J. Biol. Chem. 265, 19980–19989. https://doi.org/10.1016/S0021-9258(17)45471-8

Ishwar, A.R., Jeong, K.J., Panitch, A., Akkus, O., 2009. Raman Spectroscopic Investigation of Peptide—Glycosaminoglycan Interactions. Appl. Spectrosc. 63, 636–641. https://doi.org/10.1366/000370209788559656

Itano, N., 2008. Simple Primary Structure, Complex Turnover Regulation and Multiple Roles of Hyaluronan. J. Biochem. (Tokyo) 144, 131–137. https://doi.org/10.1093/jb/mvn046

Jeske, W., Kouta, A., Farooqui, A., Siddiqui, F., Rangnekar, V., Niverthi, M., Laddu, R., Hoppensteadt, D., Iqbal, O., Walenga, J., Fareed, J., 2019. Bovine Mucosal Heparins Are Comparable to Porcine Mucosal Heparin at USP Potency Adjusted Levels. Front. Med. 5, 1–9. https://doi.org/10.3389/fmed.2018.00360

Jin, L., Abrahams, J.P., Skinner, R., Petitou, M., Pike, R.N., Carrell, R.W., 1997. The anticoagulant activation of antithrombin by heparin. Proc. Natl. Acad. Sci. 94, 14683–14688. https://doi.org/10.1073/pnas.94.26.14683

Jolliffe, I.T., Cadima, J., 2016. Principal component analysis: a review and recent developments. Philos. Trans. R. Soc. Math. Phys. Eng. Sci. 374, 20150202. https://doi.org/10.1098/rsta.2015.0202

Ju, W., Lu, C., Liu, C., Jiang, W., Zhang, Y., Hong, F., 2020. Rapid Identification of Atmospheric Gaseous Pollutants Using Fourier-Transform Infrared Spectroscopy Combined with Independent Component Analysis. J. Spectrosc. 2020, e8920732. https://doi.org/10.1155/2020/8920732

Juhasz, P., Biemann, K., 1994. Mass spectrometric molecular-weight determination of highly acidic compounds of biological significance via their complexes with basic polypeptides. Proc. Natl. Acad. Sci. U. S. A. 91, 4333–7.

KAKOI, N., KINOSHITA, M., KAWASAKI, N., YAMAGUCHI, T., HAYAKAWA, T., KAKEHI, K., 2009. Capillary Electrophoresis Analysis of Contaminants in Heparin Sodium for the Japanese Pharmacopoeia Purity Test日本薬局方医薬品各条ヘパリンナトリウム純度試験へのキャピラリー電気泳動法の適用について. Yakugaku Zasshi 129, 1255–1264. https://doi.org/10.1248/yakushi.129.1255

Kalita, M., Balivada, S., Swarup, V.P., Mencio, C., Raman, K., Desai, U.R., Troyer, D., Kuberan, B., 2014. A nanosensor for ultrasensitive detection of oversulfated chondroitin sulfate contaminant in heparin. J. Am. Chem. Soc. 136, 554–557. https://doi.org/10.1021/ja409170z

Kareem, N., Yates, E., Skidmore, M., Hoole, D., 2018. In vitro investigations on the effects of semi-synthetic, sulphated carbohydrates on the immune status of cultured common carp (Cyprinus carpio) leucocytes. Fish Shellfish Immunol. 74, 213–222. https://doi.org/10.1016/j.fsi.2017.12.047

Karousou, E., Asimakopoulou, A.P., Zafeiropoulou, V., Viola, M., Monti, L., Rossi, A., Passi, A., Karamanos, N., 2015. Fast screening of glycosaminoglycan disaccharides by fluorophore-assisted carbohydrate electrophoresis (FACE): applications to biologic samples and pharmaceutical formulations. Methods Mol. Biol. Clifton NJ 1229, 143–159. https://doi.org/10.1007/978-1-4939-1714-3_14

Karousou, E.G., Porta, G., De Luca, G., Passi, A., 2004. Analysis of fluorophore-labelled hyaluronan and chondroitin sulfate disaccharides in biological samples. J. Pharm. Biomed. Anal. 34, 791–795. https://doi.org/10.1016/S0731-7085(03)00568-5

Katta, K., Imran, T., Busse-Wicher, M., Grønning, M., Czajkowski, S., Kusche-Gullberg, M., 2015. Reduced Expression of EXTL2, a Member of the Exostosin (EXT) Family of Glycosyltransferases, in Human Embryonic Kidney 293 Cells Results in Longer Heparan Sulfate Chains *. J. Biol. Chem. 290, 13168–13177. https://doi.org/10.1074/jbc.M114.631754

Kawashima, H., Atarashi, K., Hirose, M., Hirose, J., Yamada, S., Sugahara, K., Miyasaka, M., 2002. Oversulfated Chondroitin/Dermatan Sulfates Containing GlcAβ1/IdoAα1–3GalNAc(4,6-O-disulfate) Interact with L- and P-selectin and Chemokines∗. J. Biol. Chem. 277, 12921–12930. https://doi.org/10.1074/jbc.M200396200

Keire, D.A., Mans, D.J., Ye, H., Kolinski, R.E., Buhse, L.F., 2010a. Assay of possible economically motivated additives or native impurities levels in heparin by 1H NMR, SAX-HPLC, and anticoagulation time approaches. J. Pharm. Biomed. Anal. 52, 656–664. https://doi.org/10.1016/j.jpba.2010.02.019

Keire, D.A., Trehy, M.L., Reepmeyer, J.C., Kolinski, R.E., Ye, W., Dunn, J., Westenberger, B.J., Buhse, L.F., 2010b. Analysis of crude heparin by 1H NMR, capillary electrophoresis, and strong-anion-exchange-HPLC for contamination by over sulfated chondroitin sulfate. J. Pharm. Biomed. Anal. 51, 921–926. https://doi.org/10.1016/j.jpba.2009.10.017

Khoo, K.H., Morris, H.R., McDowell, R.A., Dell, A., Maccarana, M., Lindahl, U., 1993. FABMS/derivatisation strategies for the analysis of heparin-derived oligosaccharides. Carbohydr. Res. 244, 205–223. https://doi.org/10.1016/0008-6215(83)85002-2

Kiani, C., Chen, L., Wu, Y.J., Yee, A.J., Yang, B.B., 2002. Structure and function of aggrecan. Cell Res. 12, 19–32. https://doi.org/10.1038/sj.cr.7290106

Kishimoto, T.K., Viswanathan, K., Ganguly, T., Elankumaran, S., Smith, S., Pelzer, K., Lansing, J.C., Sriranganathan, N., Zhao, G., Galcheva-Gargova, Z., Al-Hakim, A., Bailey, G.S., Fraser, B., Roy, S., Rogers-Cotrone, T., Buhse, L., Whary, M., Fox, J., Nasr, M., Dal Pan, G.J., Shriver, Z., Langer, R.S., Venkataraman, G., Austen, K.F., Woodcock, J., Sasisekharan, R., 2008. Contaminated Heparin Associated with Adverse Clinical Events and Activation of the Contact System. N. Engl. J. Med. 358, 2457–2467. https://doi.org/10.1056/NEJMoa0803200

Kitagawa, H., Shimakawa, H., Sugahara, K., 1999. The Tumor Suppressor EXT-like Gene EXTL2 Encodes an α1, 4-N-Acetylhexosaminyltransferase That TransfersN-Acetylgalactosamine and N-Acetylglucosamine to the Common Glycosaminoglycan-Protein Linkage Region: THE KEY ENZYME FOR THE CHAIN INITIATION OF HEPARAN SULFATE *. J. Biol. Chem. 274, 13933–13937. https://doi.org/10.1074/jbc.274.20.13933

Klein, M.D., Drongowski, R.A., Linhardt, R.J., Langer, R.S., 1982. A colorimetric assay for chemical heparin in plasma. Anal. Biochem. 124, 59–64. https://doi.org/10.1016/0003-2697(82)90219-6

Knobloch, J.E., Shaklee, P.N., 1997. Absolute molecular weight distribution of low-molecular-weight heparins by size-exclusion chromatography with multiangle laser light scattering detection. Anal. Biochem. 245, 231–241. https://doi.org/10.1006/abio.1996.9984

Korpetinou, A., Skandalis, S., Labropoulou, V., Smirlaki, G., Noulas, A., Karamanos, N., THEOCHARIS, A., 2014. Serglycin: At the Crossroad of Inflammation and Malignancy. Front. Oncol. 3.

Kreuger, J., Kjellén, L., 2012. Heparan Sulfate Biosynthesis: Regulation and Variability. J. Histochem. Cytochem. 60, 898–907. https://doi.org/10.1369/0022155412464972

Kreuger, J., Salmivirta, M., Sturiale, L., Giménez-Gallego, G., Lindahl, U., 2001. Sequence Analysis of Heparan Sulfate Epitopes with Graded Affinities for Fibroblast Growth Factors 1 and 2 *. J. Biol. Chem. 276, 30744–30752. https://doi.org/10.1074/jbc.M102628200

Kriegel, H.-P., Kröger, P., Sander, J., Zimek, A., 2011. Density-based clustering. WIREs Data Min. Knowl. Discov. 1, 231–240. https://doi.org/10.1002/widm.30

Kumar, S., Lahlali, R., Liu, X., Karunakaran, C., 2016. Infrared spectroscopy combined with imaging: A new developing analytical tool in health and plant science. Appl. Spectrosc. Rev. 51, 466–483. https://doi.org/10.1080/05704928.2016.1157808

Kunder, C.A., St. John, A.L., Li, G., Leong, K.W., Berwin, B., Staats, H.F., Abraham, S.N., 2009. Mast cell–derived particles deliver peripheral signals to remote lymph nodes. J. Exp. Med. 206, 2455–2467. https://doi.org/10.1084/jem.20090805

Lagunoff, D., Warren, G., 1962. Determination of 2-deoxy-2-sulfoaminohexose content of mucopolysaccharides. Arch. Biochem. Biophys. 99, 396–400. https://doi.org/10.1016/0003-9861(62)90285-0

Langeslay, D.J., Beecher, C.N., Naggi, A., Guerrini, M., Torri, G., Larive, C.K., 2013. Characterizing the microstructure of heparin and heparan sulfate using N -sulfoglucosamine 1H and 15N NMR chemical shift analysis. Anal. Chem. 85, 1247–1255. https://doi.org/10.1021/ac3032788

Langeslay, D.J., Beni, S., Larive, C.K., 2012a. A closer look at the nitrogen next door: 1H-15N NMR methods for glycosaminoglycan structural characterization. J. Magn. Reson. 216, 169–174. https://doi.org/10.1016/j.jmr.2012.01.018

Langeslay, D.J., Jones, C.J., Beni, S., Larive, C.K., 2012b. Glycosaminoglycans: Oligosaccharide Analysis by Liquid Chromatography, Capillary Electrophoresis, and Specific Labeling, in: Rédini, F. (Ed.), Proteoglycans: Methods and Protocols, Methods in Molecular Biology. Humana Press, Totowa, NJ, pp. 131–144. https://doi.org/10.1007/978-1-61779-498-8_9

LaRiviere, W.B., Han, X., Oshima, K., McMurtry, S.A., Linhardt, R.J., Schmidt, E.P., 2021. Detection of Glycosaminoglycans by Polyacrylamide Gel Electrophoresis and Silver Staining. JoVE J. Vis. Exp. e62319. https://doi.org/10.3791/62319

Lawson, D.J., Hellenthal, G., Myers, S., Falush, D., 2012. Inference of Population Structure using Dense Haplotype Data. PLOS Genet. 8, e1002453. https://doi.org/10.1371/journal.pgen.1002453

Lee, B.M., Park, S.J., Noh, I., Kim, C.-H., 2021. The effects of the molecular weights of hyaluronic acid on the immune responses. Biomater. Res. 25, 27. https://doi.org/10.1186/s40824-021-00228-4

Lee, C.Y., Tseng, W.L., 2015. Molecular beacon-based fluorescent assay for specific detection of oversulfated chondroitin sulfate contaminants in heparin without enzyme treatment. Anal. Chem. 87, 5031–5035. https://doi.org/10.1021/acs.analchem.5b00692

Lee, J.Y., Spicer, A.P., 2000. Hyaluronan: a multifunctional, megaDalton, stealth molecule. Curr. Opin. Cell Biol. 12, 581–586. https://doi.org/10.1016/S0955-0674(00)00135-6

Lee, T.-Y., Jamieson, A.M., Schafer, I.A., 1973. Changes in the Composition and Structure of Glycosaminoglycans in the Human Placenta during Development. Pediatr. Res. 7, 965–977. https://doi.org/10.1203/00006450-197312000-00005

Lee-Sayer, S.S.M., Dong, Y., Arif, A.A., Olsson, M., Brown, K.L., Johnson, P., 2015. The Where, When, How, and Why of Hyaluronan Binding by Immune Cells. Front. Immunol. 6.

Li, H., Wickramasekara, S., Nemes, P., 2015. One-Hour Screening of Adulterated Heparin by Simplified Peroxide Digestion and Fast RPIP-LC-MS^2. Anal. Chem. 87, 8424–8432. https://doi.org/10.1021/acs.analchem.5b01788

Li, L., Zhang, F., Zaia, J., Linhardt, R.J., 2012. Top-Down Approach for the Direct Characterization of Low Molecular Weight Heparins Using LC-FT-MS. Anal. Chem. 84, 8822–8829. https://doi.org/10.1021/ac302232c

Liang, R., Chen, J., Liu, W., Liu, C., Yu, W., Yuan, M., Zhou, X., 2012. Extraction, characterization and spontaneous gel-forming property of pectin from creeping fig (Ficus pumila Linn.) seeds. Carbohydr. Polym. 87, 76–83. https://doi.org/10.1016/j.carbpol.2011.07.013

Lima, C., Muhamadali, H., Xu, Y., Kansiz, M., Goodacre, R., 2021. Imaging Isotopically Labeled Bacteria at the Single-Cell Level Using High-Resolution Optical Infrared Photothermal Spectroscopy. Anal. Chem. https://doi.org/10.1021/acs.analchem.0c03967

Lima, M.A., Cavalheiro, R.P., M.Viana, G., Meneghetti, M.C.Z., Rudd, T.R., Skidmore, M.A., Powell, A.K., Yates, E.A., 2017. 19F labelled glycosaminoglycan probes for solution NMR and non-linear (CARS) microscopy. Glycoconj. J. 34, 405–410. https://doi.org/10.1007/s10719-016-9723-x

Lima, M.A., de Farias, E.H.C., Rudd, T.R., Ebner, L.F., Gesteira, T.F., Mendes, A., Bouças, R.I., Martins, J.R.M., Hoppensteadt, D., Fareed, J., Yates, E.A., Sassaki, G.L., Tersariol, I.L.S., Nader, H.B., 2011. Low molecular weight heparins: Structural differentiation by spectroscopic and multivariate approaches. Carbohydr. Polym. 85, 903–909. https://doi.org/10.1016/j.carbpol.2011.04.021

Lin, T.-S., Hsieh, C.-H., Kuo, C., Juang, Y.-P., Hsieh, Y.S.Y., Chiang, H., Hung, S.-C., Jiang, C.-C., Liang, P.-H., 2020. Sulfation pattern of chondroitin sulfate in human osteoarthritis

cartilages reveals a lower level of chondroitin-4-sulfate. Carbohydr. Polym. 229, 115496. https://doi.org/10.1016/j.carbpol.2019.115496

Linhardt, R.J., 2003. 2003 Claude S. Hudson Award Address in Carbohydrate Chemistry. Heparin: Structure and Activity. J. Med. Chem. 46, 2551–2564. https://doi.org/10.1021/jm030176m

Liu, C., Sheng, J., Krahn, J.M., Perera, L., Xu, Y., Hsieh, P.-H., Dou, W., Liu, J., Pedersen, L.C., 2014. Molecular Mechanism of Substrate Specificity for Heparan Sulfate 2-O-Sulfotransferase *. J. Biol. Chem. 289, 13407–13418. https://doi.org/10.1074/jbc.M113.530535

Liu, H., Zhang, Z., Linhardt, R.J., 2009. Lessons learned from the contamination of heparin. Nat. Prod. Rep. 26, 313–321. https://doi.org/10.1039/b819896a

Liu, X., Li, L., Zhao, T., Dong, H., 2014. Determination of chondroitin sulfate in tablets by Raman spectroscopy and near-infrared spectroscopy combined with chemometric methods. Anal. Methods 6, 4219–4227. https://doi.org/10.1039/C3AY41853G

Loibl, M., Strahl, S., 2013. Protein O-mannosylation: What we have learned from baker's yeast. Biochim. Biophys. Acta BBA - Mol. Cell Res., Functional and structural diversity of the endoplasmic reticulum 1833, 2438–2446. https://doi.org/10.1016/j.bbamcr.2013.02.008

Loo, B.-M., Kreuger, J., Jalkanen, M., Lindahl, U., Salmivirta, M., 2001. Binding of Heparin/Heparan Sulfate to Fibroblast Growth Factor Receptor 4 *. J. Biol. Chem. 276, 16868–16876. https://doi.org/10.1074/jbc.M011226200

Lühn, S., Schiemann, S., Alban, S., 2011. Simple fluorescence assay for quantification of OSCS in heparin. Anal. Bioanal. Chem. 399, 673–680. https://doi.org/10.1007/s00216-010-3867-5

Ly, M., Wang, Z., Laremore, T.N., Zhang, F., Zhong, W., Pu, D., Zagorevski, D. V., Dordick, J.S., Linhardt, R.J., 2011. Analysis of E. coli K5 capsular polysaccharide heparosan. Anal. Bioanal. Chem. 399, 737–745. https://doi.org/10.1007/s00216-010-3679-7

Lyon, M., Deakin, J.A., Rahmoune, H., Fernig, D.G., Nakamura, T., Gallagher, J.T., 1998. Hepatocyte Growth Factor/Scatter Factor Binds with High Affinity to Dermatan Sulfate *. J. Biol. Chem. 273, 271–278. https://doi.org/10.1074/jbc.273.1.271

Maaten, L. van der, Hinton, G., 2008. Visualizing Data using t-SNE. J. Mach. Learn. Res. 9, 2579–2605.

Maccarana, M., Kalamajski, S., Kongsgaard, M., Magnusson, S.P., Oldberg, A., Malmström, A., 2009. Dermatan sulfate epimerase 1-deficient mice have reduced content and changed distribution of iduronic acids in dermatan sulfate and an altered collagen structure in skin. Mol. Cell. Biol. 29, 5517–5528. https://doi.org/10.1128/MCB.00430-09

Mainreck, N., Brézillon, S., Sockalingum, G.D., Maquart, F.-X., Manfait, M., Wegrowski, Y., 2011. Rapid Characterization of Glycosaminoglycans Using a Combined Approach by Infrared and Raman Microspectroscopies. J. Pharm. Sci. 100, 441–450. https://doi.org/10.1002/jps.22288

Mallis, L.M., Wang, H.M., Loganathan, D., Linhardt, R.J., 1989. Sequence analysis of highly sulfated, heparin-derived oligosaccharides using fast atom bombardment mass spectrometry. Anal. Chem. 61, 1453–1458. https://doi.org/10.1021/ac00188a030

Malmström, A., Bartolini, B., Thelin, M.A., Pacheco, B., Maccarana, M., 2012. Iduronic Acid in Chondroitin/Dermatan Sulfate: Biosynthesis and Biological Function. J. Histochem. Cytochem. 60, 916–925. https://doi.org/10.1369/0022155412459857

Malsch, R., Guerrini, M., Torri, G., Lohr, G., Casu, B., Harenberg, J., 1994. Synthesis of a N′-Alkylamine Anticoagulant Active Low-Molecular-Mass Heparin for Radioactive and Fluorescent Labeling. Anal. Biochem. 217, 255–264. https://doi.org/10.1006/abio.1994.1117

Malsch, R., Harenberg, J., 1994. High-performance size exclusion chromatography and polyacrylamide gel electrophoresis for characterization of unfractionated and low molecular mass glycosaminoglycans. Semin. Thromb. Hemost. 20, 135–143. https://doi.org/10.1055/s-2007-1001896

Mantovani, V., Galeotti, F., Maccari, F., Volpi, N., 2018. Recent advances in capillary electrophoresis separation of monosaccharides, oligosaccharides, and polysaccharides. ELECTROPHORESIS 39, 179–189. https://doi.org/10.1002/elps.201700290

Mao, W., Thanawiroon, C., Linhardt, R.J., 2002. Capillary electrophoresis for the analysis of glycosaminoglycans and glycosaminoglycan-derived oligosaccharides. Biomed. Chromatogr. 16, 77–94. https://doi.org/10.1002/bmc.153

Martinez, A.M., Kak, A.C., 2001. PCA versus LDA. IEEE Trans. Pattern Anal. Mach. Intell. 23, 228–233. https://doi.org/10.1109/34.908974

Massy, W.F., 1965. Principal Components Regression in Exploratory Statistical Research. J. Am. Stat. Assoc. 60, 234–256. https://doi.org/10.1080/01621459.1965.10480787

Mathlouthi, M., Koenig, J.L., 1986. Vibrational spectra of carbohydrates. Adv. Carbohydr. Chem. Biochem. 44, 7–89. https://doi.org/10.1016/s0065-2318(08)60077-3

Mauri, L., Boccardi, G., Torri, G., Karfunkle, M., Macchi, E., Muzi, L., Keire, D., Guerrini, M., 2017a. Qualification of HSQC methods for quantitative composition of heparin and low molecular weight heparins. J. Pharm. Biomed. Anal. 136, 92–105. https://doi.org/10.1016/j.jpba.2016.12.031

Mauri, L., Marinozzi, M., Mazzini, G., Kolinski, R.E., Karfunkle, M., Keire, D.A., Guerrini, M., 2017b. Combining NMR Spectroscopy and Chemometrics to Monitor Structural Features of Crude Hep-arin. Mol. Basel Switz. 22. https://doi.org/10.3390/molecules22071146

Mauri, L., Marinozzi, M., Phatak, N., Karfunkle, M., St. Ange, K., Guerrini, M., Keire, D.A., Linhardt, R.J., 2019. 1D and 2D-HSQC NMR: Two Methods to Distinguish and Characterize Heparin From Different Animal and Tissue Sources. Front. Med. 6.

McKee, J., Bairstow, S., Szabo, C., Ray, J., Wielgos, T., Hu, P., Chess, E., Nordhaus, M., Hai, T., Campbell, J., Donovan, S., Viseux, N., Riedel, N., Cammack, J., Johnson, R., 2010. Structure Elucidation and Biological Activity of the Oversulfated Chondroitin Sulfate Contaminant in Baxter Heparin. J. Clin. Pharmacol. 50, 1159–1170. https://doi.org/10.1177/0091270009355158

Mendes, A., Meneghetti, M.C.Z., Palladino, M.V., Justo, G.Z., Sassaki, G.L., Fareed, J., Lima, M.A., Nader, H.B., 2019. Crude Heparin Preparations Unveil the Presence of Structurally Diverse Oversulfated Contaminants. Molecules 24, 2988. https://doi.org/10.3390/molecules24162988

Meneghetti, M.C.Z., Hughes, A.J., Rudd, T.R., Nader, H.B., Powell, A.K., Yates, E.A., Lima, M.A., 2015. Heparan sulfate and heparin interactions with proteins. J. R. Soc. Interface 12, 20150589. https://doi.org/10.1098/rsif.2015.0589

Meyer, D., 2015. Chapter Two - Health Benefits of Prebiotic Fibers, in: Henry, J. (Ed.), Advances in Food and Nutrition Research. Academic Press, pp. 47–91. https://doi.org/10.1016/bs.afnr.2014.11.002

Meyer, K., Davidson, E., Linker, A., Hoffman, P., 1956. The acid mucopolysaccharides of connective tissue. Biochim. Biophys. Acta 21, 506–518. https://doi.org/10.1016/0006-3002(56)90188-3

Mikami, T., Kitagawa, H., 2013. Biosynthesis and function of chondroitin sulfate. Biochim. Biophys. Acta BBA - Gen. Subj. 1830, 4719–4733. https://doi.org/10.1016/j.bbagen.2013.06.006

Miller, R.L., Guimond, S.E., Schwörer, R., Zubkova, O.V., Tyler, P.C., Xu, Y., Liu, J., Chopra, P., Boons, G.-J., Grabarics, M., Manz, C., Hofmann, J., Karlsson, N.G., Turnbull, J.E., Struwe, W.B., Pagel, K., 2020. Shotgun ion mobility mass spectrometry sequencing of heparan sulfate saccharides. Nat. Commun. 11, 1481. https://doi.org/10.1038/s41467-020-15284-y

Mishra, S., Ganguli, M., 2021. Functions of, and replenishment strategies for, chondroitin sulfate in the human body. Drug Discov. Today 26, 1185–1199. https://doi.org/10.1016/j.drudis.2021.01.029

Mittelstaedt, D., Kahn, D., Xia, Y., 2016. Topographical and depth-dependent glycosaminoglycan concentration in canine medial tibial cartilage 3 weeks after anterior cruciate ligament transection surgery—a microscopic imaging study - Mittelstaedt - Quantitative Imaging in Medicine and Surgery. Quant. Imaging Med. Surg. 6, 648–660.

Mohamed, H.T., Untereiner, V., Cinque, G., Ibrahim, S.A., Götte, M., Nguyen, N.Q., Rivet, R., Sockalingum, G.D., Brézillon, S., 2020. Infrared Microspectroscopy and Imaging Analysis of Inflammatory and Non-Inflammatory Breast Cancer Cells and Their GAG Secretome. Mol. Basel Switz. 25, E4300. https://doi.org/10.3390/molecules25184300

Mohamed, H.T., Untereiner, V., Sockalingum, G.D., Brézillon, S., 2017. Implementation of infrared and Raman modalities for glycosaminoglycan characterization in complex systems. Glycoconj. J. 34, 309–323. https://doi.org/10.1007/s10719-016-9743-6

Mohd Nasir, M.F., Samsudin, M.S., Mohamad, I., Awaluddin, M.R.A., Mansor, M.A., Juahir, H., Ramli, N., 2011. River water quality modeling using combined principle component analysis (PCA) and multiple linear regressions (MLR): a case study at Klang River, Malaysia. World Appl. Sci. J. 14, 73–82.

Monakhova, Y.B., Diehl, B.W.K., 2019. Retrospective multivariate analysis of pharmaceutical preparations using 1H nuclear magnetic resonance (NMR) spectroscopy: Example of 990 heparin samples. J. Pharm. Biomed. Anal. 173, 18–23. https://doi.org/10.1016/j.jpba.2019.05.024

Monakhova, Y.B., Fareed, J., Yao, Y., Diehl, B.W.K., 2019. Anticoagulant activity of porcine heparin: Structural-property relationship and semi-quantitative estimation by nuclear magnetic resonance (NMR) spectrometry. J. Pharm. Biomed. Anal. 174, 639–643. https://doi.org/10.1016/j.jpba.2019.06.036

Monakhova, Y.B., Soboleva, P.M., Fedotova, E.S., Musina, K.T., Burmistrova, N.A., 2022. Quantum Chemical Calculations of IR Spectra of Heparin Disaccharide Subunits. Comput. Theor. Chem. 113891. https://doi.org/10.1016/j.comptc.2022.113891

Monfared, A.M.T., Tiwari, V.S., Tripathi, M.M., Anis, H., 2013. Raman spectroscopy for clinical-level detection of heparin in serum by partial least-squares analysis. J. Biomed. Opt. 18, 027010. https://doi.org/10.1117/1.JBO.18.2.027010

Moon, A.F., Xu, Y., Woody, S.M., Krahn, J.M., Linhardt, R.J., Liu, J., Pedersen, L.C., 2012. Dissecting the substrate recognition of 3-O-sulfotransferase for the biosynthesis of anticoagulant heparin. Proc. Natl. Acad. Sci. 109, 5265–5270. https://doi.org/10.1073/pnas.1117923109

Morris, E.R., Rees, D.A., Sanderson, G.R., Thom, D., 1975. Conformation and circular dichroism of uronic acid residues in glycosides and polysaccharides. J. Chem. Soc. Perkin Trans. 2 1418–1425. https://doi.org/10.1039/P29750001418

Mourão, P.A., Michelacci, Y.M., Toledo, O.M., 1979. Glycosaminoglycans and proteoglycans of normal and tumoral cartilages of humans and rats. Cancer Res. 39, 2802–2806.

Mule, S.N., Rosa-Fernandes, L., Coutinho, J.V.P., Gomes, V.D.M., Macedo-da-Silva, J., Santiago, V.F., Quina, D., de Oliveira, G.S., Thaysen-Andersen, M., Larsen, M.R., Labriola, L., Palmisano, G., 2021. Systems-wide analysis of glycoprotein conformational changes by limited deglycosylation assay. J. Proteomics 248, 104355. https://doi.org/10.1016/j.jprot.2021.104355

Mulloy, B., Forster, M.J., 2000. Conformation and dynamics of heparin and heparan sulfate. Glycobiology 10, 1147–1156. https://doi.org/10.1093/glycob/10.11.1147

Mulloy, B., Gee, C., Wheeler, S.F., Wait, R., Gray, E., Barrowcliffe, T.W., 1997. Molecular weight measurements of low molecular weight heparins by gel permeation chromatography. Thromb. Haemost. 77, 668–674.

Mulloy, B., Gray, E., Barrowcliffe, T.W., 2000. Characterization of unfractionated heparin: Comparison of materials from the last 50 years. Thromb. Haemost. 84, 1052–1056.

Mulloy, B., Heath, A., Shriver, Z., Jameison, F., Al Hakim, A., Morris, T.S., Szajek, A.Y., 2014a. USP compendial methods for analysis of heparin: chromatographic determination of molecular weight distributions for heparin sodium. Anal. Bioanal. Chem. 406, 4815–4823. https://doi.org/10.1007/s00216-014-7940-3

Mulloy, B., Heath, A., Shriver, Z., Jameison, F., Al Hakim, A., Morris, T.S., Szajek, A.Y., 2014b. USP compendial methods for analysis of heparin: chromatographic determination of molecular weight distributions for heparin sodium. Anal. Bioanal. Chem. 406, 4815–4823. https://doi.org/10.1007/s00216-014-7940-3

Mycroft-West, C.J., 2021. Non-anticoagulant glycosaminoglycan extracts isolated from aquatic species inhibit BACE-1: a key drug target in Alzheimer's Disease (Doctor of Philosophy). Keele University.

Mycroft-West, C.J., Cooper, L.C., Devlin, A.J., Procter, P., Guimond, S.E., Guerrini, M., Fernig, D.G., Lima, M.A., Yates, E.A., Skidmore, M.A., 2019. A Glycosaminoglycan Extract from Portunus pelagicus Inhibits BACE1, the β Secretase Implicated in Alzheimer's Disease. Mar. Drugs 17, 293. https://doi.org/10.3390/md17050293

Mycroft-West, C.J., Devlin, A.J., Cooper, L.C., Guimond, S.E., Procter, P., Guerrini, M., Miller, G.J., Fernig, D.G., Yates, E.A., Lima, M.A., Skidmore, M.A., 2021. Glycosaminoglycans from Litopenaeus vannamei Inhibit the Alzheimer's Disease β Secretase, BACE1. Mar. Drugs 19, 203. https://doi.org/10.3390/md19040203

Mycroft-West, C.J., Devlin, A.J., Cooper, L.C., Procter, P., Miller, G.J., Fernig, D.G., Guerrini, M., Guimond, S.E., Lima, M.A., Yates, E.A., Skidmore, M.A., 2020. Inhibition of BACE1, the β-secretase implicated in Alzheimer's disease, by a chondroitin sulfate extract from Sardina pilchardus. Neural Regen. Res. 15, 1546–1553. https://doi.org/10.4103/1673-5374.274341

Myron, P., Siddiquee, S., Azad, S.A., 2017. Partial structural studies of fucosylated chondroitin sulfate (FuCS) using attenuated total reflection fourier transform infrared spectroscopy (ATR-FTIR) and chemometrics. Vib. Spectrosc. 89, 26–36. https://doi.org/10.1016/j.vibspec.2016.12.008

Nathanson, N., Wilesmith, J., Griot, C., 1997. Bovine Spongiform Encephalopathy (BSE): Causes and Consequences of a Common Source Epidemic. Am. J. Epidemiol. 145, 959–969. https://doi.org/10.1093/oxfordjournals.aje.a009064

Neely, Brock.W., 1957. Infrared Spectra of Carbohydrates. Adv Carbohyd Chem 12, 13–33.

Nemes, P., Hoover, W.J., Keire, D.A., 2013. High-throughput differentiation of heparin from other glycosaminoglycans by pyrolysis mass spectrometry. Anal. Chem. 85, 7405–7412. https://doi.org/10.1021/ac401318q

Nielsen, F., 2016. Hierarchical Clustering, in: Nielsen, F. (Ed.), Introduction to HPC with MPI for Data Science, Undergraduate Topics in Computer Science. Springer International Publishing, Cham, pp. 195–211. https://doi.org/10.1007/978-3-319-21903-5_8

Norwig, J., Beyer, T., Brinz, D., Holzgrabe, U., Diller, M., Manns, D., 2009. Prediction of the oversulphated chondroitin sulphate contamination of unfractionated heparin by ATR-IR spectrophotometry. Pharmeuropa Sci. Notes 2009, 17–24.

Olson, S.T., Björk, I., 1994. Regulation of thrombin activity by antithrombin and heparin. Semin. Thromb. Hemost. 20, 373–409. https://doi.org/10.1055/s-2007-1001928

Oreste, P., Torri, G., 1980. Fingerprinting of heparins by low-amperage electrophoresis in barium acetate. J. Chromatogr. A 195, 398–401. https://doi.org/10.1016/S0021-9673(00)81475-2

Ornitz, D.M., Itoh, N., 2001. Fibroblast growth factors. Genome Biol. 2, reviews3005.1. https://doi.org/10.1186/gb-2001-2-3-reviews3005

Oschatz, C., Maas, C., Lecher, B., Jansen, T., Björkqvist, J., Tradler, T., Sedlmeier, R., Burfeind, P., Cichon, S., Hammerschmidt, S., Müller-Esterl, W., Wuillemin, W.A., Nilsson, G., Renné, T., 2011. Mast Cells Increase Vascular Permeability by Heparin-Initiated Bradykinin Formation In Vivo. Immunity 34, 258–268. https://doi.org/10.1016/j.immuni.2011.02.008

Ouyang, Y., Han, X., Yu, Y., Chen, J., Fu, L., Zhang, F., Linhardt, R.J., Fareed, J., Hoppensteadt, D., Jeske, W., Kouta, A., Zhang, Z., Xia, K., 2019. Chemometric analysis of porcine, bovine and ovine heparins. J. Pharm. Biomed. Anal. 164, 345–352. https://doi.org/10.1016/j.jpba.2018.10.052

Palhares, L.C.G.F., Brito, A.S., de Lima, M.A., Nader, H.B., London, J.A., Barsukov, I.L., Andrade, G.P.V., Yates, E.A., Chavante, S.F., 2019. A further unique chondroitin sulfate from the shrimp Litopenaeus vannamei with antithrombin activity that modulates acute inflammation | Elsevier Enhanced Reader. Carbohydr. Polym. 222. https://doi.org/10.1016/j.carbpol.2019.115031

Palhares, L.C.G.F., London, J.A., Kozlowski, A.M., Esposito, E., Chavante, S.F., Ni, M., Yates, E.A., 2021. Chemical Modification of Glycosaminoglycan Polysaccharides. Molecules 26, 5211. https://doi.org/10.3390/molecules26175211

Palta, S., Saroa, R., Palta, A., 2014. Overview of the coagulation system. Indian J. Anaesth. 58, 515–523. https://doi.org/10.4103/0019-5049.144643

Palukuru, U.P., McGoverin, C.M., Pleshko, N., 2014. Assessment of hyaline cartilage matrix composition using near infrared spectroscopy. Matrix Biol. 38, 3–11. https://doi.org/10.1016/j.matbio.2014.07.007

Patel, R.P., Narkowicz, C., Jacobson, G.A., 2009. Effective reversed-phase ion pair high-performance liquid chromatography method for the separation and characterization of intact low-molecular-weight heparins. Anal. Biochem. 387, 113–121. https://doi.org/10.1016/j.ab.2009.01.007

Paulick, M.G., Bertozzi, C.R., 2008. The Glycosylphosphatidylinositol Anchor: A Complex Membrane-Anchoring Structure for Proteins. Biochemistry 47, 6991–7000. https://doi.org/10.1021/bi8006324

Pavão, M.S.G., Vilela-Silva, A.-C., Mourão, P.A.S., 2006. Biosynthesis of Chondroitin Sulfate: From the Early, Precursor Discoveries to Nowadays, Genetics Approaches, in: Advances in Pharmacology, Chondroitin Sulfate: Structure, Role and Pharmacological Activity. Academic Press, pp. 117–140. https://doi.org/10.1016/S1054-3589(05)53006-0

Pepi, L.E., Sanderson, P., Stickney, M., Amster, I.J., 2021. Developments in Mass Spectrometry for Glycosaminoglycan Analysis: A Review. Mol. Cell. Proteomics 20. https://doi.org/10.1074/mcp.R120.002267

Petitou, M., Hérault, J.-P., Bernat, A., Driguez, P.-A., Duchaussoy, P., Lormeau, J.-C., Herbert, J.-M., 1999. Synthesis of thrombin-inhibiting heparin mimetics without side effects. Nature 398, 417–422. https://doi.org/10.1038/18877

physics.mq.edu.au, n.d. Fourier Transform Infrared Spectroscopy System (FTIR) [WWW Document]. URL http://physics.mq.edu.au/~goldys/optmicroweb/ftir/FTIR2old.html (accessed 9.24.22).

Plaas, A.H.K., Hascall, V.C., Midura, R.J., 1996. Ion exchange HPLC microanalysis of chondroitin sulfate: quantitative derivatization of chondroitin lyase digestion products with 2-aminopyridine. Glycobiology 6, 823–829. https://doi.org/10.1093/glycob/6.8.823

Plaas, A.H.K., Wong-Palms, S., Roughley, P.J., Midura, R.J., Hascall, V.C., 1997. Chemical and Immunological Assay of the Nonreducing Terminal Residues of Chondroitin Sulfate from Human Aggrecan *. J. Biol. Chem. 272, 20603–20610. https://doi.org/10.1074/jbc.272.33.20603

Pomin, V.H., 2014. NMR Chemical Shifts in Structural Biology of Glycosaminoglycans. Anal. Chem. 86, 65–94. https://doi.org/10.1021/ac401791h

Powell, A.K., Yates, E.A., Fernig, D.G., Turnbull, J.E., 2004. Interactions of heparin/heparan sulfate with proteins: Appraisal of structural factors and experimental approaches. Glycobiology 14, 17R-30R. https://doi.org/10.1093/glycob/cwh051

Praissman, J.L., Wells, L., 2014. Mammalian O-Mannosylation Pathway: Glycan Structures, Enzymes, and Protein Substrates. Biochemistry 53, 3066–3078. https://doi.org/10.1021/bi500153y

Presto, J., Thuveson, M., Carlsson, P., Busse, M., Wilén, M., Eriksson, I., Kusche-Gullberg, M., Kjellén, L., 2008. Heparan sulfate biosynthesis enzymes EXT1 and EXT2 affect NDST1 expression and heparan sulfate sulfation. Proc. Natl. Acad. Sci. 105, 4751–4756. https://doi.org/10.1073/pnas.0705807105

Prydz, K., 2015. Determinants of Glycosaminoglycan (GAG) Structure. Biomolecules 5, 2003–2022. https://doi.org/10.3390/biom5032003

Puvendran, K., Anupama, K., Jayaraman, G., 2018. Real-time monitoring of hyaluronic acid fermentation by in situ transflectance spectroscopy. Appl. Microbiol. Biotechnol. 102, 2659–2669. https://doi.org/10.1007/s00253-018-8816-9

Qin, Y., Ke, J., Gu, X., Fang, J., Wang, W., Cong, Q., Li, Jie, Tan, J., Brunzelle, J.S., Zhang, C., Jiang, Y., Melcher, K., Li, Jin-ping, Xu, H.E., Ding, K., 2015. Structural and Functional Study of d-Glucuronyl C5-epimerase *. J. Biol. Chem. 290, 4620–4630. https://doi.org/10.1074/jbc.M114.602201

Quantock, A.J., Young, R.D., Akama, T.O., 2010. Structural and biochemical aspects of keratan sulphate in the cornea. Cell. Mol. Life Sci. CMLS 67, 891–906. https://doi.org/10.1007/s00018-009-0228-7

Rabenstein, D.L., 2002a. Heparin and heparan sulfate: structure and function. Nat. Prod. Rep. 19, 312–331. https://doi.org/10.1039/B100916H

Rabenstein, D.L., 2002b. Heparin and heparan sulfate: structure and function. Nat. Prod. Rep. 19, 312–331. https://doi.org/10.1039/B100916H

Ramacciotti, E., Clark, M., Sadeghi, N., Hoppensteadt, D., Thethi, I., Gomes, M., Fareed, J., 2011. Contaminants in heparin: Review of the literature, molecular profiling, and clinical implications. Clin. Appl. Thromb. 17, 126–135. https://doi.org/10.1177/1076029610392214

Raman, R., Sasisekharan, V., Sasisekharan, R., 2005. Structural insights into biological roles of protein-glycosaminoglycan interactions. Chem. Biol. 12, 267–277. https://doi.org/10.1016/j.chembiol.2004.11.020

Razin, E., Stevens, R.L., Akiyama, F., Schmid, K., Austen, K.F., 1982. Culture from mouse bone marrow of a subclass of mast cells possessing a distinct chondroitin sulfate proteoglycan with glycosaminoglycans rich in N-acetylgalactosamine-4,6-disulfate. J. Biol. Chem. 257, 7229–7236.

Ren, Q., Wang, J., Liu, C., Meng, L., Qian, R., Gao, H., Qin, W., Zhou, C., Qiao, S., Wang, H., Zhang, L., Zhang, Y., 2021. Exploring the sulfate patterns of chondroitin sulfate/dermatan sulfate and keratan sulfate in human pancreatic cancer. J. Pharm. Biomed. Anal. 205, 114339. https://doi.org/10.1016/j.jpba.2021.114339

Ricketts, C.R., 1952. Dextran sulphate—a synthetic analogue of heparin. Biochem. J. 51, 129–133.

Rieppo, L., Saarakkala, S., Närhi, T., Helminen, H.J., Jurvelin, J.S., Rieppo, J., 2012. Application of second derivative spectroscopy for increasing molecular specificity of fourier transform infrared spectroscopic imaging of articular cartilage. Osteoarthritis Cartilage 20, 451–459. https://doi.org/10.1016/j.joca.2012.01.010

Rietschel, E.T., Kirikae, T., Schade, F.U., Mamat, U., Schmidt, G., Loppnow, H., Ulmer, A.J., Zähringer, U., Seydel, U., Di Padova, F., Schreier, M., Brade, H., 1994. Bacterial endotoxin: molecular relationships of structure to activity and function. FASEB J. 8, 217–225. https://doi.org/10.1096/fasebj.8.2.8119492

[215]

Rizkalla, G., Reiner, A., Bogoch, E., Poole, A.R., 1992. Studies of the articular cartilage proteoglycan aggrecan in health and osteoarthritis. Evidence for molecular heterogeneity and extensive molecular changes in disease. J. Clin. Invest. 90, 2268–2277. https://doi.org/10.1172/JCI116113

Rodriguez, H.J., Vanderwielen, A.J., 1979. Molecular weight determination of commercial heparin sodium USP and its sterile solutions. J. Pharm. Sci. 68, 588–591. https://doi.org/10.1002/jps.2600680519

Romanenko, S., Begley, R., Harvey, A.R., Hool, L., Wallace, V.P., 2017. The interaction between electromagnetic fields at megahertz, gigahertz and terahertz frequencies with cells, tissues and organisms: risks and potential. J. R. Soc. Interface 14, 20170585. https://doi.org/10.1098/rsif.2017.0585

Rong, J., Habuchi, H., Kimata, K., Lindahl, U., Kusche-Gullberg, M., 2001. Substrate Specificity of the Heparan Sulfate Hexuronic Acid 2-O-Sulfotransferase. Biochemistry 40, 5548–5555. https://doi.org/10.1021/bi002926p

Roughley, P.J., Mort, J.S., 2014. The role of aggrecan in normal and osteoarthritic cartilage. J. Exp. Orthop. 1, 8. https://doi.org/10.1186/s40634-014-0008-7

Roughley, P.J., White, R.J., Glant, T.T., 1987. The Structure and Abundance of Cartilage Proteoglycans during Early Development of the Human Fetus. Pediatr. Res. 22, 409–413. https://doi.org/10.1203/00006450-198710000-00009

Roy, R., Williams, R.E., Jennings, H.J., 1984. Increase in the acyclic form of sugars in the presence of borate ions, as measured by circular dichroism. Carbohydr. Res. 127, 165–169. https://doi.org/10.1016/0008-6215(84)85119-8

Rudd, T.R., Gaudesi, D., Skidmore, M.A., Ferro, M., Guerrini, M., Mulloy, B., Torri, G., Yates, E.A., 2011a. Construction and use of a library of bona fide heparins employing 1H NMR and multivariate analysis. Analyst 136, 1380–1389. https://doi.org/10.1039/c0an00834f

Rudd, T.R., Guimond, S.E., Skidmore, M.A., Duchesne, L., Guerrini, M., Torri, G., Cosentino, C., Brown, A., Clarke, D.T., Turnbull, J.E., Fernig, D.G., Yates, E.A., 2007. Influence of substitution pattern and cation binding on conformation and activity in heparin derivatives | Glycobiology | Oxford Academic. Glycobiology 17, 983–993.

Rudd, T.R., Macchi, E., Gardini, C., Muzi, L., Guerrini, M., Yates, E.A., Torri, G., 2012b. How To Find a Needle (or Anything Else) in a Haystack: Two-Dimensional Correlation Spectroscopy-Filtering with Iterative Random Sampling Applied to Pharmaceutical Heparin. Anal. Chem. 84, 6841–6847. https://doi.org/10.1021/ac301428d

Rudd, T.R., Mauri, L., Marinozzi, M., Stancanelli, E., Yates, E.A., Naggi, A., Guerrini, M., 2019. Multivariate analysis applied to complex biological medicines. Faraday Discuss. 218, 303–316. https://doi.org/10.1039/C9FD00009G

Rudd, T.R., Skidmore, M.A., Guimond, S.E., Guerrini, M., Cosentino, C., Edge, R., Brown, A., Clarke, D.T., Torri, G., Turnbull, J.E., Nichols, R.J., Fernig, D.G., Yates, E.A., 2008. Site-specific interactions of copper(II) ions with heparin revealed with complementary (SRCD, NMR, FTIR and EPR) spectroscopic techniques. Carbohydr. Res., Selected papers from the 14th European Carbohydrate Symposium, Lübeck, Germany, Sept. 2007 343, 2184–2193. https://doi.org/10.1016/j.carres.2007.12.019

Rudd, T.R., Skidmore, M.A., Guimond, S.E., Holman, J., Turnbull, J.E., Lauder, R.M., Fernig, D.G., Yates, E.A., 2009. The potential for circular dichroism as an additional facile

and sensitive method of monitoring low-molecular-weight heparins and heparinoids. Thromb. Haemost. 102, 874–878. https://doi.org/10.1160/TH08-12-0797

Rudd, T.R., Uniewicz, K.A., Ori, A., Guimond, S.E., Skidmore, M.A., Gaudesi, D., Xu, R., Turnbull, J.E., Guerrini, M., Torri, G., Siligardi, G., Wilkinson, M.C., Fernig, D.G., Yates, E.A., 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. Org. Biomol. Chem. 8, 5390–5397. https://doi.org/10.1039/C0OB00246A

Ruiz-Calero, V., Moyano, E., Puignou, L., Galceran, M.T., 2001. Pressure-assisted capillary electrophoresis-electrospray ion trap mass spectrometry for the analysis of heparin depolymerised disaccharides. J. Chromatogr. A 914, 277–291. https://doi.org/10.1016/S0021-9673(00)01181-X

Sanden, K.W., Kohler, A., Afseth, N.K., Böcker, U., Rønning, S.B., Liland, K.H., Pedersen, M.E., 2019. The use of Fourier-transform infrared spectroscopy to characterize connective tissue components in skeletal muscle of Atlantic cod (Gadus morhua L.). J. Biophotonics 12, e201800436. https://doi.org/10.1002/jbio.201800436

Sarraguça, M.C., Lopes, J.A., 2009. Quality control of pharmaceuticals with NIR: From lab to process line. Vib. Spectrosc. 49, 204–210. https://doi.org/10.1016/j.vibspec.2008.07.013

Sarrazin, S., Lamanna, W.C., Esko, J.D., 2011. Heparan Sulfate Proteoglycans. Cold Spring Harb. Perspect. Biol. 3.

Schachter, H., 2000. The joys of HexNAc. The synthesis and function of N-andO-glycan branches. Glycoconj. J. 17, 465–483. https://doi.org/10.1023/A:1011010206774

Scheffers, D.-J., Pinho, M.G., 2005. Bacterial Cell Wall Synthesis: New Insights from Localization Studies. Microbiol. Mol. Biol. Rev. 69, 585–607. https://doi.org/10.1128/MMBR.69.4.585-607.2005

Schnaar, R.L., 2004. Glycolipid-mediated cell–cell recognition in inflammation and nerve regeneration. Arch. Biochem. Biophys., Highlight Issue on Glycobiology Dedicated to the Memory of Victor Ginsburg 426, 163–172. https://doi.org/10.1016/j.abb.2004.02.019

Sheikh, M.O., Halmo, S.M., Wells, L., 2017. Recent advancements in understanding mammalian O-mannosylation. Glycobiology 27, 806–819. https://doi.org/10.1093/glycob/cwx062

Shi, X., Zaia, J., 2009. Organ-specific Heparan Sulfate Structural Phenotypes *. J. Biol. Chem. 284, 11806–11814. https://doi.org/10.1074/jbc.M809637200

Shibutani, T., Iwayama, Y., Tsubone, M., Ando, C., Yamada, K., Mori, M., 1990. Immunohistochemical localization of glycosaminoglycans with the use of monoclonal antibodies in salivary pleomorphic adenomas. Anticancer Res. 10, 1533–1542.

Siebert, F., Hildebrandt, P., 2007. Theory of Infrared Absorption and Raman Spectroscopy, in: Vibrational Spectroscopy in Life Science. John Wiley & Sons, Ltd, pp. 11–61. https://doi.org/10.1002/9783527621347.ch2

Silbert, J.E., Sugumaran, G., 2002. Biosynthesis of chondroitin/dermatan sulfate. IUBMB Life 54, 177–186. https://doi.org/10.1080/15216540214923

Sinnis, P., Coppi, A., Toida, T., Toyoda, H., Kinoshita-Toyoda, A., Xie, J., Kemp, M.M., Linhardt, R.J., 2007. Mosquito Heparan Sulfate and Its Potential Role in Malaria Infection and Transmission *. J. Biol. Chem. 282, 25376–25384. https://doi.org/10.1074/jbc.M704698200

Sinnott, M., 2007. Carbohydrate Chemistry and Biochemistry. https://doi.org/10.1039/9781847558015

Sitkowski, J., Bednarek, E., Bocian, W., Kozerski, L., 2008. Assessment of Oversulfated Chondroitin Sulfate in Low Molecular Weight and. J Med Chem 51, 7663–7665.

Skidmore, M.A., Guimond, S.E., Dumax-Vorzet, A.F., Yates, E.A., Turnbull, J.E., 2010. Disaccharide compositional analysis of heparan sulfate and heparin polysaccharides using UV or high-sensitivity fluorescence (BODIPY) detection. Nat. Protoc. 5, 1983–1992. https://doi.org/10.1038/nprot.2010.145

Skidmore, M.A., Mustaffa, K.M.F., Cooper, L.C., Guimond, S.E., Yates, E.A., Craig, A.G., 2017. A semi-synthetic glycosaminoglycan analogue inhibits and reverses Plasmodium falciparum cytoadherence. PLOS ONE 12, e0186276. https://doi.org/10.1371/journal.pone.0186276

Sommers, C.D., Mans, D.J., Mecker, L.C., Keire, D.A., 2011a. Sensitive detection of oversulfated chondroitin sulfate in heparin sodium or crude heparin with a colorimetric microplate based assay. Anal. Chem. 83, 3422–3430. https://doi.org/10.1021/ac200011s

Sommers, C.D., Ye, H., Kolinski, R.E., Nasr, M., Buhse, L.F., Al-Hakim, A., Keire, D.A., 2011b. Characterization of currently marketed heparin products: Analysis of molecular weight and heparinase-I digest patterns. Anal. Bioanal. Chem. 401, 2445–2454. https://doi.org/10.1007/s00216-011-5362-z

Somsen, G.W., Tak, Y.H., Toraño, J.S., Jongen, P.M.J.M., de Jong, G.J., 2009. Determination of oversulfated chondroitin sulfate and dermatan sulfate impurities in heparin by capillary electrophoresis. J. Chromatogr. A 1216, 4107–4112. https://doi.org/10.1016/j.chroma.2009.02.063

Spelta, F., Liverani, L., Peluso, A., Marinozzi, M., Urso, E., Guerrini, M., Naggi, A., 2019. SAX-HPLC and HSQC NMR Spectroscopy: Orthogonal Methods for Characterizing Heparin Batches Composition. Front. Med. 6, 1–11. https://doi.org/10.3389/fmed.2019.00078

Spencer, J.A., Kauffman, J.F., Reepmeyer, J.C., Gryniewicz, C.M., Yi, W., Duckhee, T., Buhse, L.F., Westenberger, B.J., 2009. Screening of Heparin API by Near Infrared Reflectance and Raman Spectroscopy. J. Pharm. Sci. 98, 3540–3547. https://doi.org/10.1002/jps

Spiro, R.G., 2002. Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. Glycobiology 12, 43R-56R. https://doi.org/10.1093/glycob/12.4.43r

St Ange, K., Onishi, A., Fu, L., Sun, X., Lin, L., Mori, D., Zhang, F., Dordick, J.S., Fareed, J., Hoppensteadt, D., Jeske, W., Linhardt, R.J., 2016a. Analysis of Heparins Derived From Bovine Tissues and Comparison to Porcine Intestinal Heparins. Clin. Appl. Thromb. 22, 520–527. https://doi.org/10.1177/1076029616643822

Stanley, F.E., Stalcup, A.M., 2011. The use of circular dichroism as a simple heparin-screening strategy. Anal. Bioanal. Chem. 399, 701–706. https://doi.org/10.1007/s00216-010-4272-9

Staples, G.O., Zaia, J., 2011. Analysis of Glycosaminoglycans Using Mass Spectrometry. Curr. Proteomics 8, 325–336.

Stivala, S.S., Herbst, M., Kratky, O., Pilz, I., 2004. Physico-chemical studies of fractionated bovine heparin. Arch. Biochem. Biophys. 127, 795–802. https://doi.org/10.1016/0003-9861(68)90291-9

Studelska, D.R., Giljum, K., McDowell, L.M., Zhang, L., 2006. Quantification of glycosaminoglycans by reversed-phase HPLC separation of fluorescent isoindole derivatives. Glycobiology 16, 65–72. https://doi.org/10.1093/glycob/cwj037

Suflet, D.M., Chitanu, G.C., Desbrières, J., 2010. Phosphorylated polysaccharides. 2. Synthesis and properties of phosphorylated dextran. Carbohydr. Polym. 82, 1271–1277. https://doi.org/10.1016/j.carbpol.2010.07.007

Sugahara, K., Mikami, T., Uyama, T., Mizuguchi, S., Nomura, K., Kitagawa, H., 2003. Recent advances in the structural biology of chondroitin sulfate and dermatan sulfate. Curr. Opin. Struct. Biol. 13, 612–620. https://doi.org/10.1016/j.sbi.2003.09.011

Sun, C., Zang, H., Liu, X., Dong, Q., Li, L., Wang, F., Sui, L., 2010. Determination of potency of heparin active pharmaceutical ingredient by near infrared reflectance spectroscopy. J. Pharm. Biomed. Anal. 51, 1060–1063. https://doi.org/10.1016/j.jpba.2009.11.022

Sunwoo, H.H., Nakano, T., Hudson, R.J., Sim, J.S., 1998. Isolation, characterization and localization of glycosaminoglycans in growing antlers of wapiti (Cervus elaphus). Comp. Biochem. Physiol. B Biochem. Mol. Biol. 120, 273–283. https://doi.org/10.1016/s0305-0491(98)10017-2

Susi, H., Michael Byler, D., 1983. Protein structure by Fourier transform infrared spectroscopy: Second derivative spectra. Biochem. Biophys. Res. Commun. 115, 391–397. https://doi.org/10.1016/0006-291X(83)91016-1

Szajek, A.Y., Chess, E., Johansen, K., Gratzl, G., Gray, E., Keire, D., Linhardt, R.J., Liu, J., Morris, T., Mulloy, B., Nasr, M., Shriver, Z., Torralba, P., Viskov, C., Williams, R., Woodcock, J., Workman, W., Al-Hakim, A., 2016. The US regulatory and pharmacopeia response to the global heparin contamination crisis. Nat. Biotechnol. 34, 625–630. https://doi.org/10.1038/nbt.3606

Tami, C., Puig, M., Reepmeyer, J.C., Ye, H., D'Avignon, D.A., Buhse, L., Verthelyi, D., 2008. Inhibition of Taq polymerase as a method for screening heparin for oversulfated contaminants. Biomaterials 29, 4808–4814. https://doi.org/10.1016/j.biomaterials.2008.08.024

Taniguchi, N., Kizuka, Y., 2015. Chapter Two - Glycans and Cancer: Role of N-Glycans in Cancer Biomarker, Progression and Metastasis, and Therapeutics, in: Drake, R.R., Ball, L.E. (Eds.), Advances in Cancer Research, Glycosylation and Cancer. Academic Press, pp. 11–51. https://doi.org/10.1016/bs.acr.2014.11.001

Taylor, S.L., Hogwood, J., Guo, W., Yates, E.A., Turnbull, J.E., 2019. By-Products of Heparin Production Provide a Diverse Source of Heparin-like and Heparan Sulfate Glycosaminoglycans. Sci. Rep. 9, 2679. https://doi.org/10.1038/s41598-019-39093-6

Toda, N., Doi, A., Jimbo, A., Matsumoto, I., Seno, N., 1981. Interaction of sulfated glycosaminoglycans with lectins. J. Biol. Chem. 256, 5345–5349.

Toledo, O.M.S., Dietrich, C.P., 1977. Tissue specific distribution of sulfated mucopolysaccharides in mammals. Biochim. Biophys. Acta BBA - Gen. Subj. 498, 114–122. https://doi.org/10.1016/0304-4165(77)90092-7

Torode, T.A., Marcus, S.E., Jam, M., Tonon, T., Blackburn, R.S., Hervé, C., Knox, J.P., 2015. Monoclonal Antibodies Directed to Fucoidan Preparations from Brown Algae. PLOS ONE 10, e0118366. https://doi.org/10.1371/journal.pone.0118366

Tóth, G., Vékey, K., Sugár, S., Kovalszky, I., Drahos, L., Turiák, L., 2020. Salt gradient chromatographic separation of chondroitin sulfate disaccharides. J. Chromatogr. A 1619, 460979. https://doi.org/10.1016/j.chroma.2020.460979

Tovar, A.M.F., Capillé, N.V.M., Santos, G.R.C., Vairo, B.C., Oliveira, S.N.M.C.G., Fonseca, R.J.C., Mourão, P.A.S., 2012. Heparin from bovine intestinal mucosa: Glycans with multiple sulfation patterns and anticoagulant effects. Thromb. Haemost. 107, 903–915. https://doi.org/10.1160/TH-11-07-0518

Toyoda, H., Yamamoto, H., Ogino, N., Toida, T., Imanari, T., 1999. Rapid and sensitive analysis of disaccharide composition in heparin and heparan sulfate by reversed-phase ion-pair chromatography on a 2 µm porous silica gel column. J. Chromatogr. A 830, 197–201. https://doi.org/10.1016/S0021-9673(98)00854-1

Trehy, M.L., Reepmeyer, J.C., Kolinski, R.E., Westenberger, B.J., Buhse, L.F., 2009a. Analysis of heparin sodium by SAX/HPLC for contaminants and impurities. J. Pharm. Biomed. Anal. 49, 670–673. https://doi.org/10.1016/j.jpba.2008.12.013

Trehy, M.L., Reepmeyer, J.C., Kolinski, R.E., Westenberger, B.J., Buhse, L.F., 2009b. Analysis of heparin sodium by SAX/HPLC for contaminants and impurities. J. Pharm. Biomed. Anal. 49, 670–673. https://doi.org/10.1016/j.jpba.2008.12.013

Trowbridge, J.M., Gallo, R.L., 2002. Dermatan sulfate: new functions from an old glycosaminoglycan. Glycobiology 12, 117R–25R. https://doi.org/10.1093/glycob/cwf066

Tuckett, F., Morriss-Kay, G., 1988. Alcian Blue staining of glycosaminoglycans in embryonic material: Effect of different fixatives | SpringerLink. Histochem. J. 20, 174–182.

Turnbull, J.E., Hopwood, J.J., Gallagher, J.T., 1999. A strategy for rapid sequencing of heparan sulfate and heparin saccharides. Proc. Natl. Acad. Sci. 96, 2698–2703. https://doi.org/10.1073/pnas.96.6.2698

Ueoka, C., Nadanaka, S., Seno, N., Khoo, K.H., Sugahara, K., 1999. Structural determination of novel tetra- and hexasaccharide sequences isolated from chondroitin sulfate H (oversulfated dermatan sulfate) of hagfish notochord. Glycoconj. J. 16, 291–305. https://doi.org/10.1023/a:1007022229813

Uniewicz, K.A., Ori, A., Xu, R., Ahmed, Y., Wilkinson, M.C., Fernig, D.G., Yates, E.A., 2010. Differential scanning fluorimetry measurement of protein stability changes upon binding to glycosaminoglycans: a screening test for binding specificity. Anal. Chem. 82, 3796–3802. https://doi.org/10.1021/ac100188x

USP, 2009. USP Heparin Sodium Monograph.

van den Berg, R.A., Hoefsloot, H.C., Westerhuis, J.A., Smilde, A.K., van der Werf, M.J., 2006. Centering, scaling, and transformations: improving the biological information content of metabolomics data. BMC Genomics 7, 142. https://doi.org/10.1186/1471-2164-7-142

Van der Meer, J.-Y., Kellenbach, E., Van den Bos, L.J., 2017. From Farm to Pharma: An Overview of Industrial Heparin Manufacturing Methods. Molecules 22, 1025. https://doi.org/10.3390/molecules22061025

Vasconcelos Oliveira, A.P., de Abreu Feitosa, V., de Oliveira, J.M., Coelho, A.L., de Araújo P. Vieira, L., de Assis Rocha da Silva, F., de Assis Avelino Figueredo Sobrinho, F., Duarte,

E.B., de Souza, B.W., de sá Moreira de Souza Filho, M., 2017. Characteristics of Chondroitin Sulfate Extracted of Tilapia (Oreochromis niloticus) Processing. Procedia Eng., 3rd International Conference on Natural Fibers: Advanced Materials for a Greener World, ICNF 2017, 21-23 June 2017, Braga, Portugal 200, 193–199. https://doi.org/10.1016/j.proeng.2017.07.028

Vasko, P.D., Blackwell, J., Koenig, J.L., 1971. Infrared and raman spectroscopy of carbohydrates: Part I: Identification of O・H and C・H-related vibrational modes for D-glucose, maltose, cellobiose, and dextran by deuterium-substitution methods. Carbohydr. Res. 19, 297–310. https://doi.org/10.1016/S0008-6215(00)86160-1

Vigetti, D., Karousou, E., Viola, M., Deleonibus, S., De Luca, G., Passi, A., 2014. Hyaluronan: biosynthesis and signaling. Biochim. Biophys. Acta 1840, 2452–2459. https://doi.org/10.1016/j.bbagen.2014.02.001

Vilanova, E., Tovar, A.M.F., Mourão, P.A.S., 2019a. Imminent risk of a global shortage of heparin caused by the African Swine Fever afflicting the Chinese pig herd. J. Thromb. Haemost. 17, 254–256. https://doi.org/10.1111/jth.14372

Vilanova, E., Vairo, B.C., Oliveira, S.-N.M.C.G., Glauser, B.F., Capillé, N.V., Santos, G.R.C., Tovar, A.M.F., Pereira, M.S., Mourão, P.A.S., 2019b. Heparins Sourced From Bovine and Porcine Mucosa Gain Exclusive Monographs in the Brazilian Pharmacopeia. Front. Med. 6.

Viskov, C., Bouley, E., Hubert, P., Martinez, C., Herman, F., Jeske, W., Hoppensteadt, D., Walenga, J.M., Fareed, J., 2009. Isolation and characterization of contaminants in recalled unfractionated heparin and low-molecular-weight heparin. Clin. Appl. Thromb. 15, 395–401. https://doi.org/10.1177/1076029609338710

Volpi, N., 1996. Purification of heparin, dermatan sulfate and chondroitin sulfate from mixtures by sequential precipitation with various organic solvents. J. Chromatogr. B Biomed. Appl. 685, 27–34. https://doi.org/10.1016/0378-4347(96)00154-5

Volpi, Nicola, 1996. Electrophoresis Separation of Glycosaminoglycans on Nitrocellulose Membranes. Anal. Biochem. 240, 114–118. https://doi.org/10.1006/abio.1996.0337

Volpi, Nicola, 1994. Dermatan sulfate from beef mucosa: structure, physicochemical and biological properties of fractions prepared by chemical depolymerization and anion-exchange chromatography. Carbohydr. Res. 255, 133–144. https://doi.org/10.1016/S0008-6215(00)90975-3

Volpi, N., 1994. Fractionation of Heparin, Dermatan Sulfate, and Chondroitin Sulfate by Sequential Precipitation: A Method to Purify a Single Glycosaminoglycan Species from a Mixture. Anal. Biochem. 218, 382–391. https://doi.org/10.1006/abio.1994.1196

Volpi, N., Buzzega, D., 2012. Agarose-gel electrophoresis for the quality assurance and purity of heparin formulations. J. Pharm. Biomed. Anal. 67–68, 144–147. https://doi.org/10.1016/j.jpba.2012.04.002

Volpi, N., Maccari, F., 2006. Electrophoretic approaches to the analysis of complex polysaccharides. J. Chromatogr. B Analyt. Technol. Biomed. Life. Sci. 834, 1–13. https://doi.org/10.1016/j.jchromb.2006.02.049

Volpi, N., Maccari, F., 2002. Detection of submicrogram quantities of glycosaminoglycans on agarose gels by sequential staining with toluidine blue and Stains-All. ELECTROPHORESIS 23, 4060–4066. https://doi.org/10.1002/elps.200290021

Volpi, N., Maccari, F., Linhardt, R.J., 2009. Quantitative capillary electrophoresis determination of oversulfated chondroitin sulfate as a contaminant in heparin preparations. Anal. Biochem. 388, 140–145. https://doi.org/10.1016/j.ab.2009.02.012

Volpi, N., Maccari, F., Linhardt, R.J., 2008. Capillary electrophoresis of complex natural polysaccharides. Electrophoresis 29, 3095–3106. https://doi.org/10.1002/elps.200800109

Walenga, J.M., Prechel, M., Jeske, W.P., Bakhos, M., 2005. Unfractionated heparin compared with low-molecular-weight heparin as related to heparin-induced thrombocytopenia. Curr. Opin. Pulm. Med. 11, 385–391. https://doi.org/10.1097/01.mcp.0000174232.78219.03

Wang, G., Dong, C., Shang, Y., Sun, Y., Fu, D., Zhao, J., 2009. Characterization of radix rehmanniae processing procedure using FT-IR spectroscopy through nonnegative independent component analysis. Anal. Bioanal. Chem. 394, 827–833. https://doi.org/10.1007/s00216-009-2759-z

Wang, L., Buchanan, S., Meyerhoff, M.E., 2008. Detection of high-charge density polyanion contaminants in biomedical heparin preparations using potentiometric polyanion sensors. Anal. Chem. 80, 9845–9847. https://doi.org/10.1021/ac801879t

Wang, S., Chen, C., Gadi, M.R., Saikam, V., Liu, D., Zhu, H., Bollag, R., Liu, K., Chen, X., Wang, F., Wang, P.G., Ling, P., Guan, W., Li, L., 2021. Chemoenzymatic modular assembly of O-GalNAc glycans for functional glycomics. Nat. Commun. 12, 3573. https://doi.org/10.1038/s41467-021-23428-x

Wang, Z., Chi, L., 2018. Recent advances in mass spectrometry analysis of low molecular weight heparins. Chin. Chem. Lett. 29, 11–18. https://doi.org/10.1016/j.cclet.2017.08.050

Warkentin, T.E., 2006. Think of HIT. Hematology 2006, 408–414. https://doi.org/10.1182/asheducation-2006.1.408

Watt, D.K., Yorke, S.C., Slim, G.C., 1997. Comparison of ovine, bovine and porcine mucosal heparins and low molecular weight heparins by disaccharide analyses and 13C NMR. Carbohydr. Polym. 33, 5–11. https://doi.org/10.1016/S0144-8617(97)00042-8

Weigel, P.H., DeAngelis, P.L., 2007. Hyaluronan Synthases: A Decade-plus of Novel Glycosyltransferases *. J. Biol. Chem. 282, 36777–36781. https://doi.org/10.1074/jbc.R700036200

Werdelin, O., Meldal, M., Jensen, T., 2002. Processing of glycans on glycoprotein and glycopeptide antigens in antigen-presenting cells. Proc. Natl. Acad. Sci. 99, 9611–9613. https://doi.org/10.1073/pnas.152345899

Wernersson, S., Pejler, G., 2014. Mast cell secretory granules: armed for battle. Nat. Rev. Immunol. 14, 478–494. https://doi.org/10.1038/nri3690

What is NMR? [WWW Document], 2022. URL http://chem.ch.huji.ac.il/nmr/whatisnmr/whatisnmr.html (accessed 9.24.22).

WHO, 2022. WHO List of Essential medicines [WWW Document]. URL https://list.essentialmeds.org/ (accessed 9.24.22).

Wielgos, T., Havel, K., Ivanova, N., Weinberger, R., 2009. Determination of impurities in heparin by capillary electrophoresis using high molarity phosphate buffers. J. Pharm. Biomed. Anal. 49, 319–326. https://doi.org/10.1016/j.jpba.2008.11.034

Wiercigroch, E., Szafraniec, E., Czamara, K., Pacia, M.Z., Majzner, K., Kochan, K., Kaczor, A., Baranska, M., Malek, K., 2017. Raman and infrared spectroscopy of carbohydrates: A review. Spectrochim. Acta. A. Mol. Biomol. Spectrosc. 185, 317–335. https://doi.org/10.1016/j.saa.2017.05.045

Wold, S., Sjöström, M., Eriksson, L., 2001. PLS-regression: a basic tool of chemometrics. Chemom. Intell. Lab. Syst., PLS Methods 58, 109–130. https://doi.org/10.1016/S0169-7439(01)00155-1

Wu, J., Wei, J., Chopra, P., Boons, G.-J., Lin, C., Zaia, J., 2019. Sequencing Heparan Sulfate Using HILIC LC-NETD-MS/MS. Anal. Chem. 91, 11738–11746. https://doi.org/10.1021/acs.analchem.9b02313

Yamada, S., Sugahara, K., Özbek, S., 2011. Evolution of glycosaminoglycans. Commun. Integr. Biol. 4, 150–158. https://doi.org/10.4161/cib.4.2.14547

Yan, D., Lin, X., 2009. Shaping Morphogen Gradients by Proteoglycans. Cold Spring Harb. Perspect. Biol. 1, a002493. https://doi.org/10.1101/cshperspect.a002493

Yan, Y., Ji, Y., Su, N., Mei, X., Wang, Y., Du, S., Zhu, W., Zhang, C., Lu, Y., Xing, X.H., 2017. Non-anticoagulant effects of low molecular weight heparins in inflammatory disorders: A review. Carbohydr. Polym. 160, 71–81. https://doi.org/10.1016/j.carbpol.2016.12.037

Yang, B., Solakyildirim, K., Chang, Y., Linhardt, R.J., 2011. Hyphenated techniques for the analysis of heparin and heparan sulfate. Anal. Bioanal. Chem. 399, 541–557. https://doi.org/10.1007/s00216-010-4117-6

Yates, E A, Santini, F., Guerrini, M., Naggi, A., Torri, G., Casu, B., 1996. 1H and 13C NMR spectral assignments of the major sequences of twelve systematically modified heparin derivatives. Carbohydr. Res. 294, 15–27.

Yates, E. A., Santini, F., Guerrini, M., Naggi, A., Torri, G., Casu, B., 1996. 1H and 13C NMR spectral assignments of the major sequences of twelve systematically modified heparin derivatives. Carbohydr. Res. 294, 15–27. https://doi.org/10.1016/s0008-6215(96)90611-4

Yin, J., Xia, Y., 2010. Macromolecular Concentrations in Bovine Nasal Cartilage by Fourier Transform Infrared Imaging and Principal Component Regression. Appl. Spectrosc. 64, 1199–1208. https://doi.org/10.1366/000370210793335124

Yu, H., Takeuchi, H., 2019. Protein O-glucosylation: another essential role of glucose in biology. Curr. Opin. Struct. Biol., Sequences and Topology ● Carbohydrates 56, 64–71. https://doi.org/10.1016/j.sbi.2018.12.001

Zaia, J., 2005. Mass Spectrometry Resource. J. Biomacromolecular Mass Spectrom. 1, 3–36.

Zamfir, A., Peter-Katalinić, J., 2004. Capillary electrophoresis-mass spectrometry for glycoscreening in biomedical research. Electrophoresis 25, 1949–1963. https://doi.org/10.1002/elps.200405825

Zang, H., Li, L., Wang, F., Yi, Q., Dong, Q., Sun, C., Wang, J., 2012. A method for identifying the origin of chondroitin sulfate with near infrared spectroscopy. J. Pharm. Biomed. Anal. 61, 224–229. https://doi.org/10.1016/j.jpba.2011.12.011

Zappe, A., Miller, R.L., Struwe, W.B., Pagel, K., 2021. State-of-the-art glycosaminoglycan characterization. Mass Spectrom. Rev. https://doi.org/10.1002/mas.21737

Zhang, F., Yang, B., Ly, M., Solakyildirim, K., Xiao, Z., Wang, Z., Beaudet, J.M., Torelli, A.Y., Dordick, J.S., Linhardt, R.J., 2011. Structural characterization of heparins from different commercial sources. Anal Bioanal Chem 401, 2793–2803. https://doi.org/10.1038/mp.2011.182.doi

Zhang, H., Uchimura, K., Kadomatsu, K., 2006. Brain Keratan Sulfate and Glial Scar Formation. Ann. N. Y. Acad. Sci. 1086, 81–90. https://doi.org/10.1196/annals.1377.014

Zhang, Z., Li, B., Suwan, J., Zhang, F., Wang, Z., Liu, H., Mulloy, B., Linhardt, R.J., 2009. Analysis of Pharmaceutical Heparins and Potential Contaminants Using 1H-NMR and PAGE. J. Pharm. Sci. 98, 4017–4026. https://doi.org/10.1002/jps

Ziegler, A., Zaia, J., 2006. Size-exclusion chromatography of heparin oligosaccharides at high and low pressure. J. Chromatogr. B Analyt. Technol. Biomed. Life. Sci. 837, 76–86. https://doi.org/10.1016/j.jchromb.2006.04.013

Zumbühl, S., Brändle, A., Hochuli, A., Scherrer, N.C., Caseri, W., 2017. Derivatization Technique To Identify Specifically Carbonyl Groups by Infrared Spectroscopy: Characterization of Photooxidative Aging Products in Terpenes and Terpeneous Resins. Anal. Chem. 89, 1742–1748. https://doi.org/10.1021/acs.analchem.6b04008

Chapter 8: Appendix

Table A.1: Table of heparins used and literature in which they were characterised. PMHs are also "pharmaceutical heparins"

Name	Serial Number	Refs	Name	Serial Number	Refs	Name	Serial Number	Refs
PMH-1	A01106023		BMH-1	G10803		OMH-1	G10786	
PMH-2	A3100A		BMH-2	G12014		OMH-2	G11447	
PMH-3	A3201QA		BMH-3	G12015		OMH-3	G11448	
PMH-4	A3288QA		BMH-4	G12017		OMH-4	G11449	
PMH-5	A353PSA		BMH-5	G12018		OMH-5	G11480	
PMH-6	A3787		BMH-6	G12019		OMH-6	G11482	
PMH-7	A3934		BMH-7	G12020		OMH-7	G11489	
PMH-8	A4077		BMH-8	G12021		OMH-8	G11936	
PMH-9	A4427		BMH-9	G12022		OMH-9	G11937	
PMH-10	A4464		BMH-10	G12028		OMH-10	G12024	
PMH-11	A5229		BMH-11	G12657		OMH-11	G12025	
PMH-12	A5302		BMH-12	G12658		OMH-12	G12026	
PMH-13	A5306		BMH-13	G12659		OMH-13	G12028	
PMH-14	A5319		BMH-14	G12660		OMH-14	G12029	
PMH-15	A5363		BMH-15	G12661		OMH-15	G12077	
PMH-16	A5444		BMH-16	G12662		OMH-16	G12205	
PMH-17	A5468		BMH-17	G12663		OMH-17	G12206	
PMH-18	A5569		BMH-18	G12664		OMH-18	G12207	(Colombo et al., 2021; Rudd et al., 2012; Mauri et al., 2019)
PMH-19	A5595		BMH-19	G12665		OMH-19	G12288	
PMH-20	A5605		BMH-20	G12666		OMH-20	G12301	
PMH-21	A5613		BMH-21	G12667		OMH-21	G12302	
PMH-22	A5615		BMH-22	G12668		OMH-22	G13096	
PMH-23	A5668		BMH-23	G12669	(Colombo et al., 2021; Rudd et al., 2012; Mauri et al., 2019)	OMH-23	G13097	
PMH-24	A5675		BMH-24	G12802		OMH-24	G13098	
PMH-25	G10921		BMH-25	G12803		OMH-25	G13689	
PMH-26	G10922		BMH-26	G12804		OMH-26	G13690	
PMH-27	G11429	(Colombo et al., 2021; Rudd et al., 2012; Mauri et al., 2019)	BMH-27	G13420		OMH-27	G13691	
PMH-28	G1143		BMH-28	G13420		OMH-28	G1371	
PMH-29	G11444		BMH-29	G13421		OMH-29	G1372	
PMH-30	G11627		BMH-30	G13421		OMH-30	G1373	
PMH-31	G11629		BMH-31	G13421		OMH-31	G8117	
PMH-32	G11653		BMH-32	G13422		OMH-32	GUK	
PMH-33	G11657		BMH-33	G13422		OMH-33	NA	
PMH-34	G11661		BMH-34	G13422				

Name	Serial Number	Refs	Name	Serial Number	Refs	Name	Serial Number	Refs
PMH-35	G12006		BMH-35	G13423		BLH-1	G10787	
PMH-36	G12007		BMH-36	G13423		BLH-2	G11099	
PMH-37	G12009		BMH-37	G13504		BLH-3	G11100	
PMH-38	G12010		BMH-38	G13505		BLH-4	G11101	
PMH-39	G12011		BMH-39	G13506		BLH-5	G11388	
PMH-40	G12012		BMH-40	G1467		BLH-6	G11389	
PMH-41	G12013		BMH-41	G1468		BLH-7	G11463	
PMH-42	G13100		BMH-42	G1469		BLH-8	G11757	
PMH-43	G13425		BMH-43	G1470		BLH-9	G11758	
PMH-44	G13426		BMH-44	G1471		BLH-10	G11759	(Colombo et al., 2021; Rudd et al., 2012; Mauri et al., 2019; Bertini et al., 2017)
PMH-45	G13427		BMH-45	G1472	(Colombo et al., 2021; Rudd et al., 2012; Mauri et al., 2019)	BLH-11	G121	
PMH-46	G13456		BMH-46	G1473		BLH-12	G12271	
PMH-47	G6202		BMH-47	G1474		BLH-13	G12272	
PMH-48	G6203		BMH-48	G1475		BLH-14	G12275	
PMH-49	G7765		BMH-49	G1476		BLH-15	G17273	
PMH-50	G7768		BMH-50	G7773		BLH-16	G17274	
PMH-51	GUK1		BMH-51	G7841		BLH-17	G4141	
			BMH-52	G7844		BLH-18	G46	
			BMH-53	G7845		BLH-19	G65	
			BMH-54	G7846				
			BMH-55	G7847				
			BMH-56	G7848				
			BMH-57	G7850				

Table A.2: Table of GAGs used.

Sample	Serial Number	Sample	Serial Number	Sample	Serial Number	Sample	Serial Number
HA-1	G1039	HS-1	G101	DS-1	10596	CS-1	DFSTDO
HA-2	G1093	HS-2	G104	DS-2	A1908	CS-2	G11251I
HA-3	G13998	HS-3	G107	DS-3	G1021	CS-3	G1381
HA-4	G1824	HS-4	G108	DS-4	G1084	CS-4	G1673A
HA-5	G28	HS-5	G116	DS-5	G33	CS-5	G267
HA-6	G45	HS-6	G120	DS-6	G5026	CS-6	G6866
HA-7	G5673	HS-7	G13014	DS-7	G593	CS-7	G7222
HA-8	G5674	HS-8	G13015	DS-8	G6322	CS-8	G8501
HA-9	G5733	HS-9	G1567	DS-9	G6987	CS-9	G8585
HA-10	G9980	HS-10	G1568	DS-10	G904	CS-10	G8783
		HS-11	G1807	DS-11	G904/1	CS-11	G91039
Sample	**Serial Number**	HS-12	G34	DS-12	G904D44	CS-12	G9131
CSA-1	G1125B	HS-13	G41	DS-13	G904D47	CS-13	G9132
CSA-2	G167	HS-14	G42	DS-14	G904D51	CS-14	G9133
CSA-3	G2426	HS-15	G5021	DS-15	G904D54	CS-15	G9175
CSA-4	G3432	HS-16	G5547	DS-16	GC180	CS-16	G9176
CSA-5	G4090	HS-17	G67	DS-17	GC181	CS-17	G9178
CSA-6	G920	HS-18	G7861	DS-18	GC182	CS-18	MPSTDN
CSA-7	YC152881701	HS-19	HHS-S	DS-19	GC185	CS-19	P6098
		HS-20	HS137				
Sample	**Serial Number**	HS-21	HS138	**Sample**	**Serial Number**		
CSC-1	G113	HS-22	HS139-140	OSCS-1	FDA		
CSC-2	G1223	HS-23	HS141	OSCS-2	G62439a		
CSC-3	G56	HS-24	HS149	OSCS-3	G62439b		
CSC-4	MPS2	HS-25	HS20594	OSCS-4	G62439c		
CSC-5	MPS3	HS-26	HS271192	OSCS-5	G6243b1		
		HS-27	HS9335	OSCS-6	G6296		
		HS-28	HS934	OSCS-7	G6469		
		HS-29	LOTUKa	OSCS-8	H0M191		
		HS-30	O258				
		HS-31	OBGR				

Crude marine GAG samples were a generous gift from Courtney Mycroft-West and are characterised in (Mycroft-West, 2021).

CS utilised for the generation of new OSCS in house was acquired from Carbosynth (catalogue No. YC15288; Batch No. YC15288.1701). and found to have comparable electrophoretic migration and disaccharide quantities to standard CS (Mycroft-West, 2021).

:
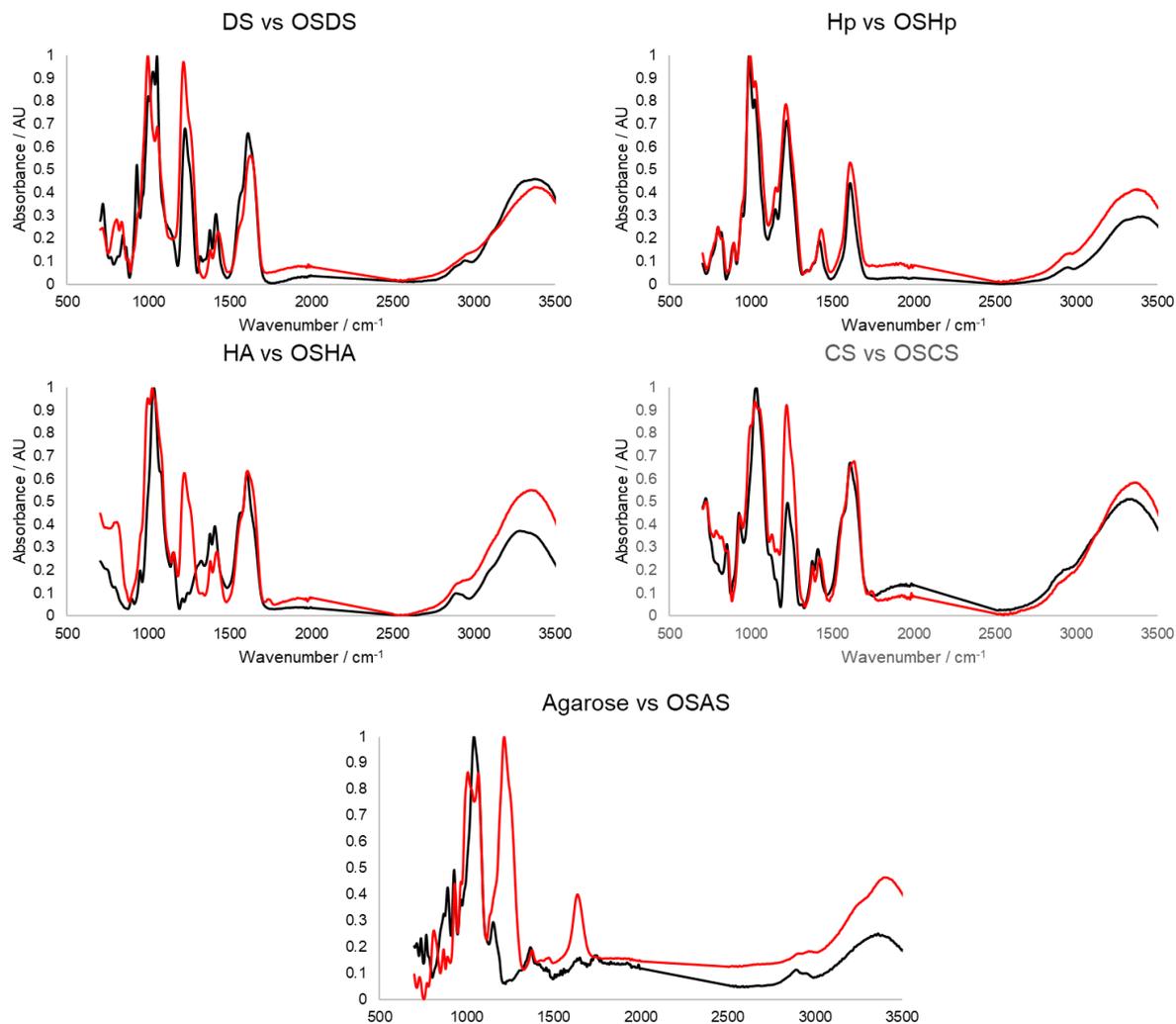
Figure A.1: IR spectra of chemically sulphated polysaccharides that were used to form contamination series. Precursor: black. OS- red. The presence of sulphate or increased sulphated is indicated by the presence or an increase in the intensity at the sulphate band (~1250cm-1). Modifications to the main carbohydrate band at ~1000 cm-1 may also be observed (Amarasekara et al., 2007). All spectra are the average of 5 repeats of 3 scans and in the region of 700 to 3700 cm-1.. Spectra were smoothed, baseline corrected with a 7th order polynomial and normalised (0-1).
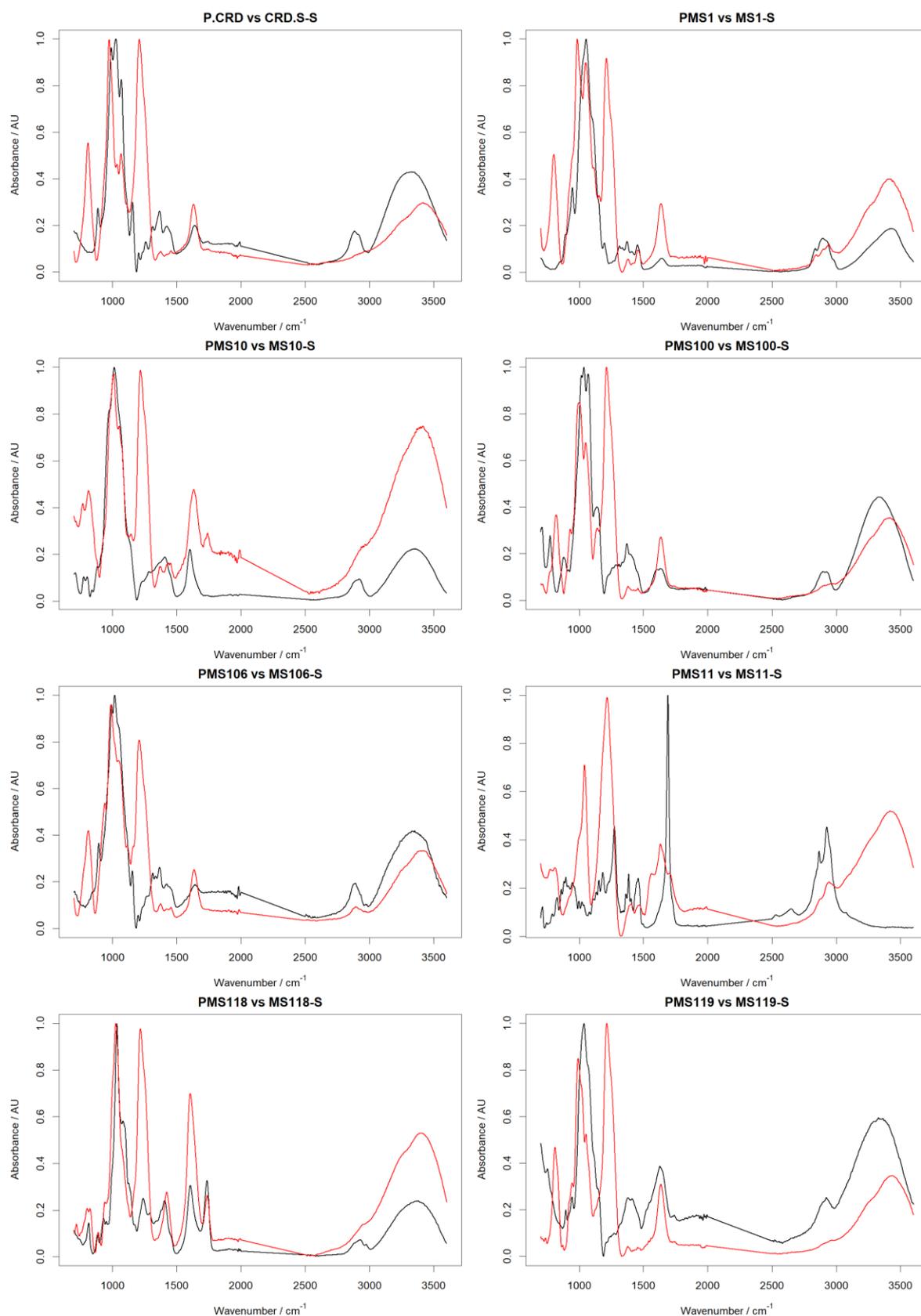
Figure A.2.1: IR spectra of chemically sulphated polysaccharides that were used in Table 4.1. Precursor: black. OS- red. The presence of sulphate or increased sulphated is indicated by the presence or an increase in the intensity at the sulphate band (~1250 cm⁻¹). Modifications to the main carbohydrate band at ~1000 cm⁻¹ may also be observed (Amarasekara et al., 2007). All spectra are the average of 5 repeats of 3 scans and in the region of 700 to 3700 cm⁻¹. Spectra were smoothed, baseline corrected with a 7th order polynomial and normalised (0-1).
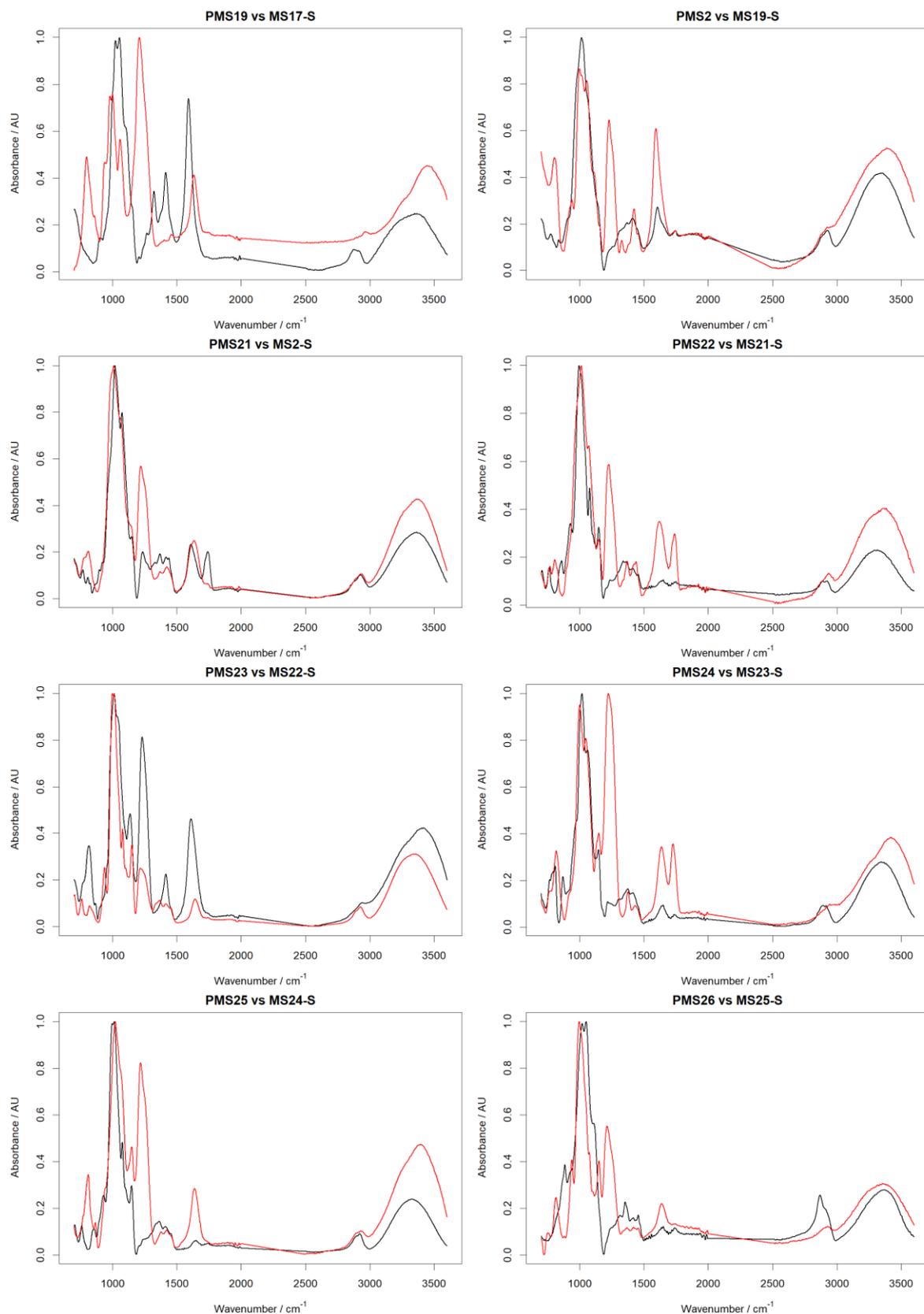
Figure A.2.2 IR spectra of chemically sulphated polysaccharides that were used in Table 4.1. Precursor: black. OS- red. The presence of sulphate or increased sulphated is indicated by the presence or an increase in the intensity at the sulphate band (~1250cm^{-1}). Modifications to the main carbohydrate band at ~1000 cm^{-1} may also be observed (Amarasekara et al., 2007). All spectra are the average of 5 repeats of 3 scans and in the region of 700 to 3700 cm^{-1}. Spectra were smoothed, baseline corrected with a 7th order polynomial and normalised (0-1).
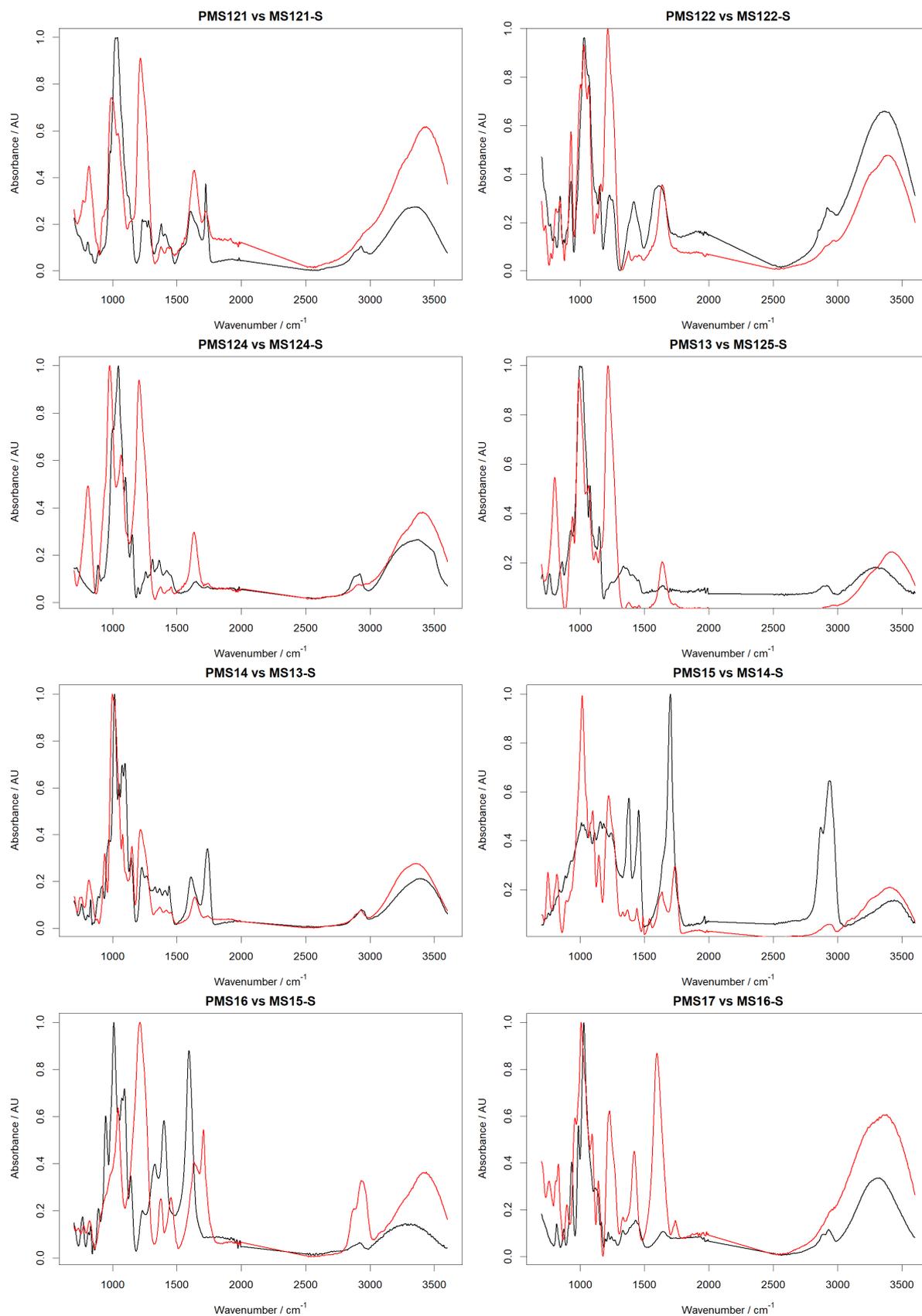
Figure A.2.3: IR spectra of chemically sulphated polysaccharides that were used in Table 4.1. Precursor: black. OS- red. The presence of sulphate or increased sulphated is indicated by the presence or an increase in the intensity at the sulphate band (~1250cm^{-1}). Modifications to the main carbohydrate band at ~1000 cm^{-1} may also be observed (Amarasekara et al., 2007). All spectra are the average of 5 repeats of 3 scans and in the region of 700 to 3700 cm^{-1}. Spectra were smoothed, baseline corrected with a 7th order polynomial and normalised (0-1).

[VI]
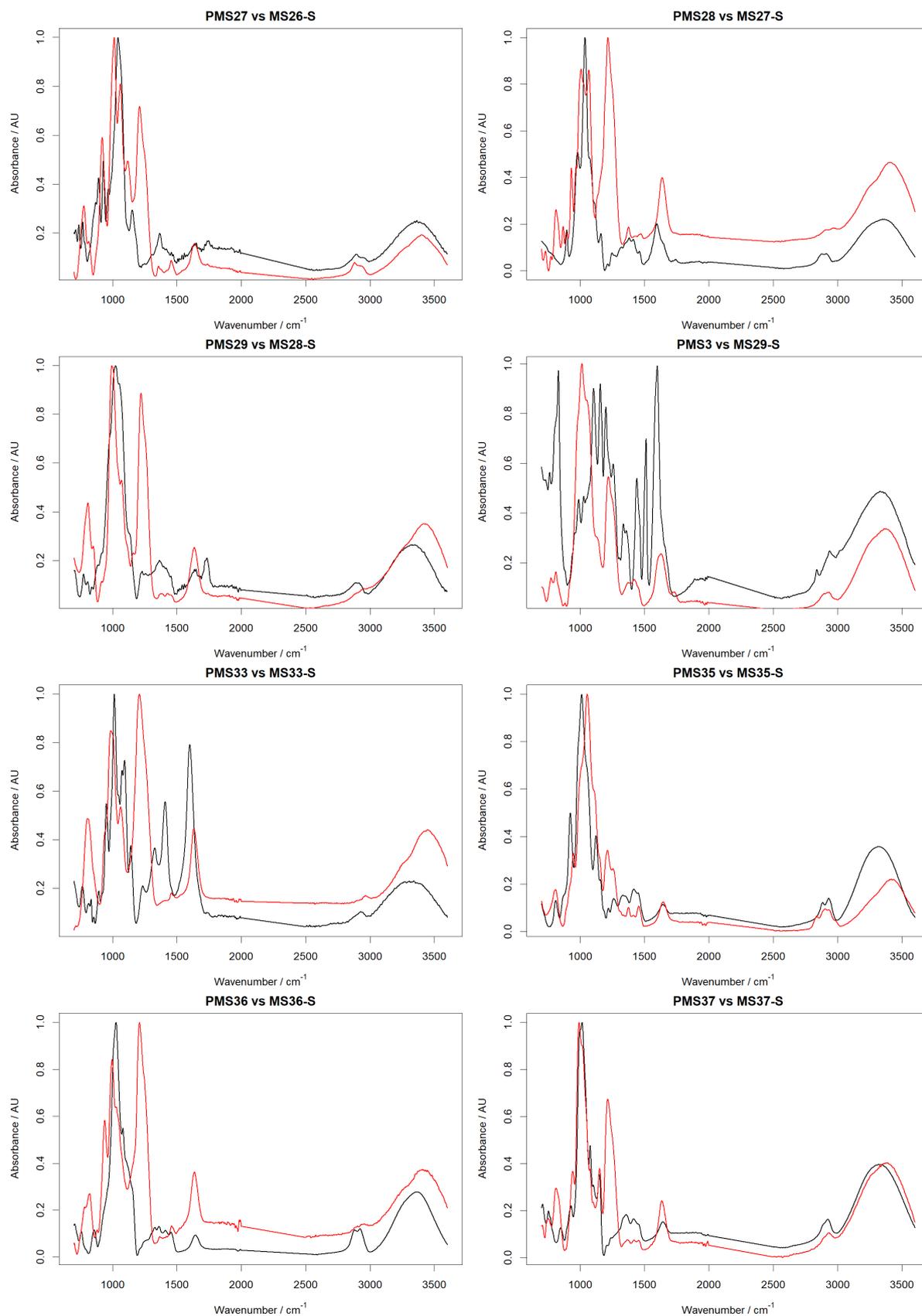
Figure A.2.4: IR spectra of chemically sulphated polysaccharides that were used in Table 4.1. Precursor: black. OS- red. The presence of sulphate or increased sulphated is indicated by the presence or an increase in the intensity at the sulphate band (~1250cm[-1]). Modifications to the main carbohydrate band at ~1000 cm[-1] may also be observed (Amarasekara et al., 2007). All spectra are the average of 5 repeats of 3 scans and in the region of 700 to 3700 cm[-1]. Spectra were smoothed, baseline corrected with a 7th order polynomial and normalised (0-1).
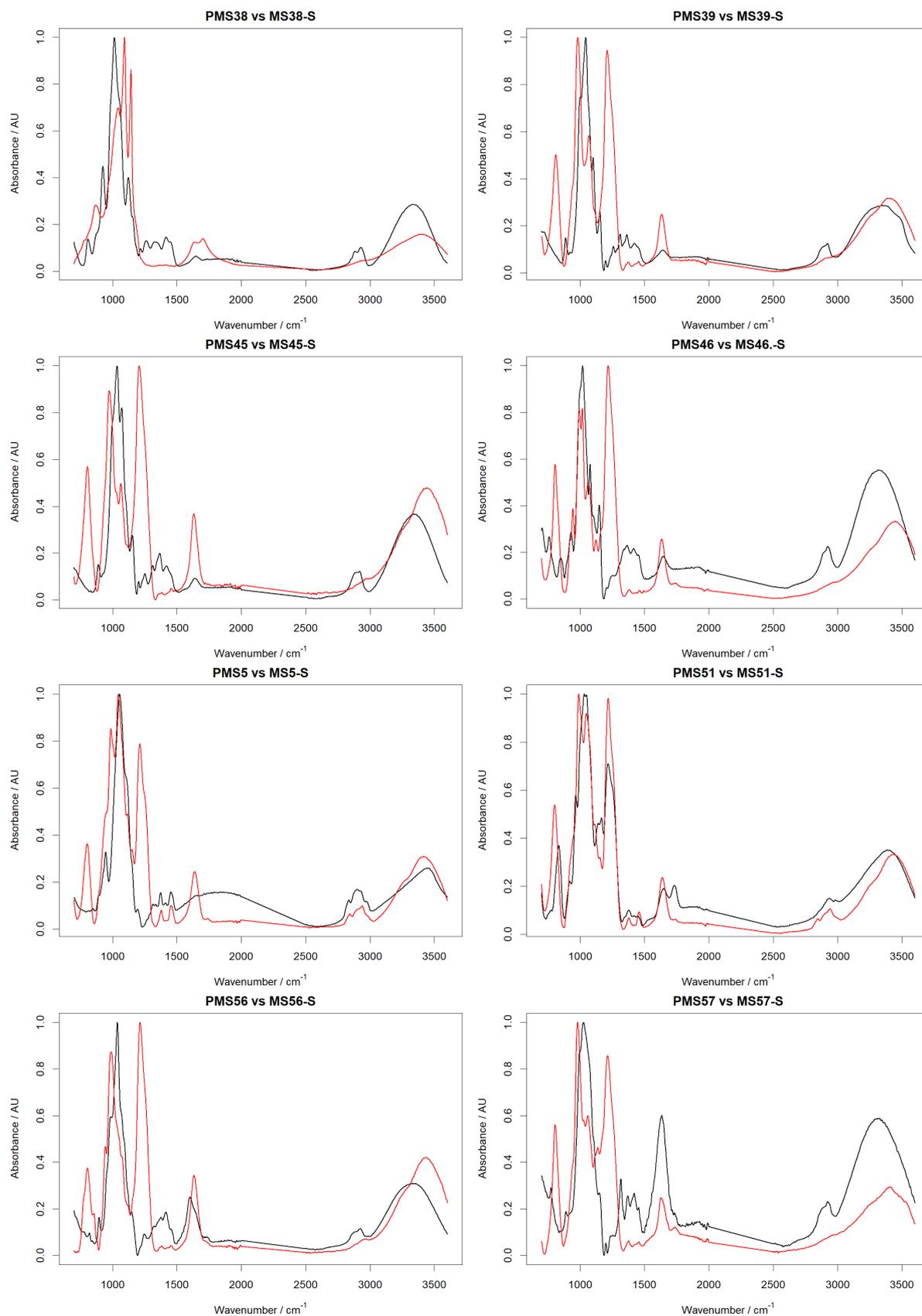
Figure A.2.5: IR spectra of chemically sulphated polysaccharides that were used in Table 4.1. Precursor: black. OS- red. The presence of sulphate or increased sulphated is indicated by the presence or an increase in the intensity at the sulphate band (~1250cm^{-1}). Modifications to the main carbohydrate band at ~1000 cm^{-1} may also be observed (Amarasekara et al., 2007). All spectra are the average of 5 repeats of 3 scans and in the region of 700 to 3700 cm^{-1}. Spectra were smoothed, baseline corrected with a 7th order polynomial and normalised (0-1).
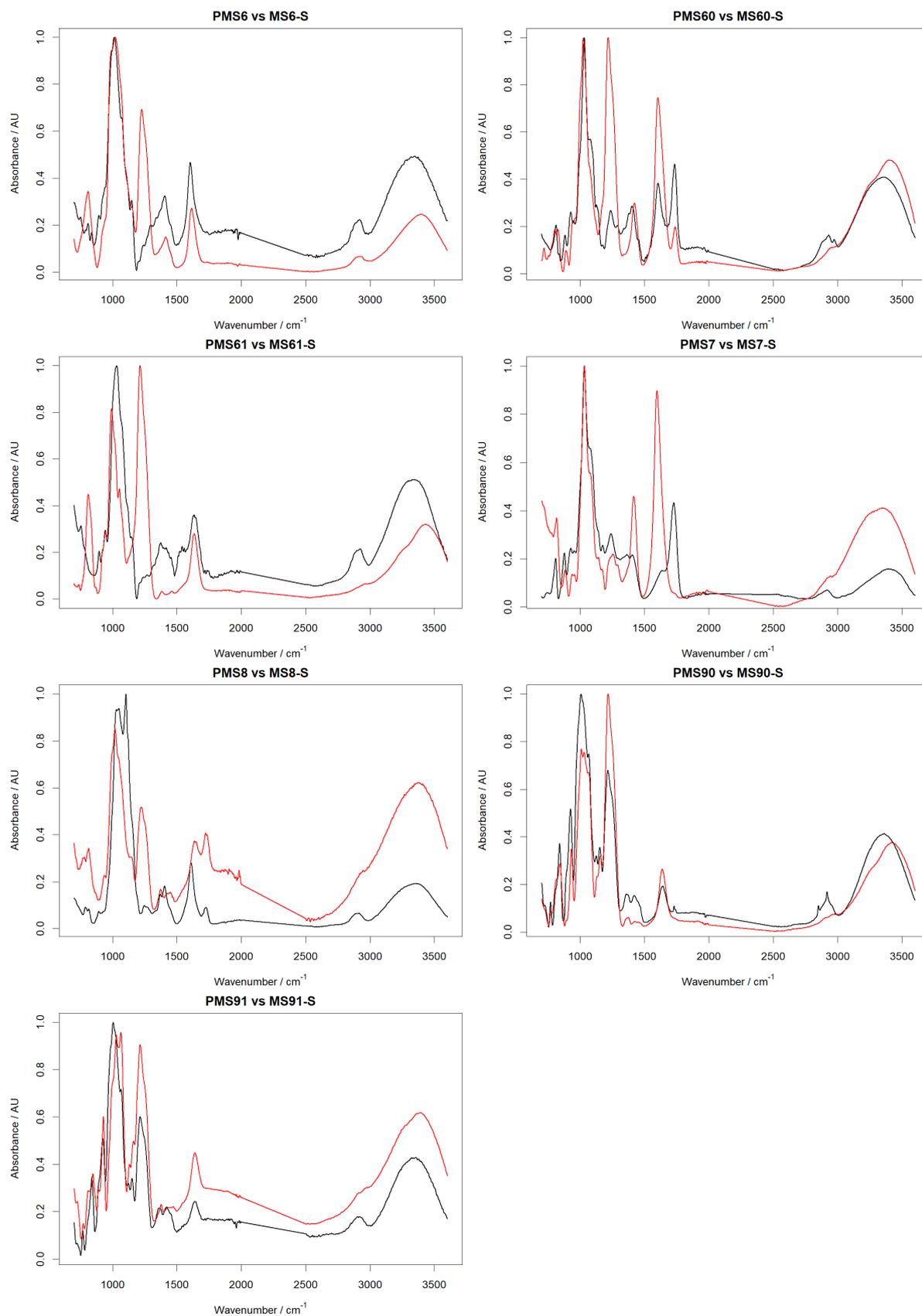
Figure A.2.6: IR spectra of chemically sulphated polysaccharides that were used in Table 4.1. Precursor: black. OS- red. The presence of sulphate or increased sulphated is indicated by the presence or an increase in the intensity at the sulphate band (~1250 cm^{-1}). Modifications to the main carbohydrate band at ~1000 cm^{-1} may also be observed (Amarasekara et al., 2007). All spectra are the average of 5 repeats of 3 scans and in the region of 700 to 3700 cm^{-1}. Spectra were smoothed, baseline corrected with a 7th order polynomial and normalised (0-1).

Table A.3: Table of polysaccharides used. Codes can be used to identify polysaccharides in **Figures A.2.**

Code	Polysaccharide	Code	Polysaccharide
MS1	Tylose (MH300)	MS26	Hydoxyethyl cellulose
MS10	Gum Arabic	MS27	Agarose
MS100	Arabinogalactan	MS28	Xylan
MS106	Hydroxypropyl cellulose	MS29	Arabic acid
MS11	Gum Rosin	MS33	Polygalacturonic acid
MS118	Propylene glycol Alginate (DEXTRA)	MS35	Methyl cellulose
MS119	Tamarind gum	MS36	b-cyclodextrin polymer
MS121	Welan gum	MS37	Pullulan
MS122	K-carrageenan	MS38	Levan
MS124	paramylon	MS39	b-1-3-glucan
MS125	α-Cellulose	MS45	Laminarin
MS13	Starch	MS46	Maltoheptaose
MS14	Pectin	MS5	(Hydroxypropyl)methyl cellulose
MS15	Gum Mastic	MS51	Fucogalactan (DEXTRA)
MS16	Pottasium Pectate	MS56	Psyllium Seed Gum (DEXTRA)
MS17	Inulin	MS57	Scleroglucan (DEXTRA)
MS19	Carboxymethyl cellulose	MS6	Gellan Gum
MS2	Gum Ghatti	MS60	Propylene glycol Alginate (DEXTRA)
MS21	Tragacanth	MS61	Tamarind Gum
MS22	Amylopectin	MS7	Alginic acid
MS23	Karaya Gum	MS8	Xanthan Gum
MS24	Guar	MS90	Chitin
MS25	Dextrin Type1	MS91	Chitosan
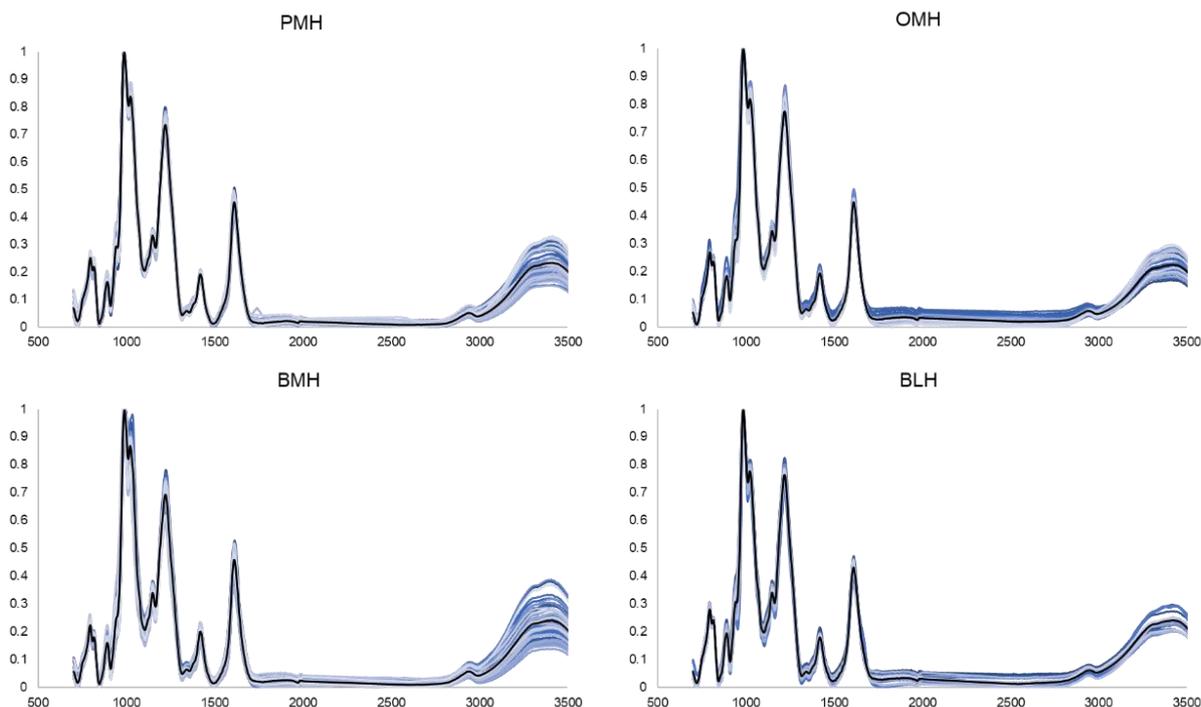
Figure A.3: IR spectra of heparin libraries. Samples: blue, average: black. PMHs are also "pharmaceutical heparins". All spectra are the average of 5 repeats of 3 scans and in the region of 700 to 3700 cm^{-1}. Spectra were smoothed, baseline corrected with a 7th order polynomial and normalised (0-1).

Figure A.4: IR spectra of GAG libraries. Samples: blue, average: black. All spectra are the average of 5 repeats of 3 scans and in the region of 700 to 3700 cm^{-1}. Spectra were smoothed, baseline corrected with a 7th order polynomial and normalised (0-1).

Table A.4: Table of crude heparins used and characterisation data. Samples were characterised in (Mauri et al., 2017c).

Serial No.	Heparin Level	DS Level	CS Level	IdoA2S-GalNac (DS-2S)	GalNac-6S (CS-6S)	GlcN,3S,6X	GlcA-GlcNS,35,6X	IdoA2S-GlcN	DoS	GlcNAc,6X-G	GlcNS,6X-GlcA	GlcNS,6X-IdoA	GlcNS,6X-IdoA2S	IdoA2S-GlcNH2,6X	GlcNAc,6X	GlcNS,6X	GlcNx,6S	GlcA-GlcNAc,6X	GlcA-GlcNS,6X	IdoA-GlcNx	IdoA-GlcNx,6S	GlcNH2,6X	GlcA,2OH	GlcA,2S	IdoA,2OH	IdoA,2S	LR	GNR
G12131	84.6	13.5	1.8	8.8	18.2	4.9	4.1	72.3	2.43	11.9	9.7	9	61.6	3.2	12.6	80.2	77.7	6.6	5.5	2.4	5.8	2.3	16.2	0	8.2	75.5	6.1	2.1
G12132	89.5	8.5	2	13.1	18.7	5.3	4	69.7	2.34	15	10.1	8.5	56.5	2.3	16.4	76.4	74.6	6.8	7.1	3.1	7	1.9	17.9	0	10.2	72	5.7	2.1
G12133	89.5	8.5	2	13.1	18.7	5.3	4	69.7	2.34	15	10.1	8.5	56.5	2.3	16.4	76.4	74.6	6.8	7.1	3.1	7	1.9	17.9	0	10.2	72	5.7	2.1
G12134	89.8	8.7	1.5	6.6	18.7	4.8	4	71.5	2.37	14.9	10	9.7	57	2.6	16.4	77	76.3	6.1	7.4	2.5	5.8	1.8	17.6	0	8.3	74.1	5.8	2.1
G12135	88	11.4	0	12.9	0	5.3	3.1	72.7	2.44	12.3	9.7	7.8	61.4	3	13.2	79.1	79	7.1	4.6	2.6	6.8	2.4	14.9	0	9.4	75.7	6.2	1.8
G12136	88.2	10.2	1.6	5.6	0	3.7	3.4	70.7	2.32	14	8.4	8.5	60.5	2.8	16.1	77.8	79	7.6	5.9	3.1	6.5	2.3	16.9	0	9.6	73.5	6.2	1.9
G12137	91.3	6.1	2.6	0	19.8	3.7	4.1	67.8	2.32	14.2	8.4	9.2	55.9	2.4	15.8	77	74.1	6.9	7.4	3.6	7.9	3	18.4	0	11.5	70.1	6.5	2.3
G12138	91.7	7.9	0	11.6	30.9	5.9	4.1	69.8	2.39	12	8.7	9.1	60	3	13.9	77.2	72.8	7.3	9	3	7.1	3	16.9	0	10.3	72.8	6	2.6
G12139	94.4	4.7	0.9	0	43.6	4.4	3.1	69.1	2.27	16.1	9.1	8.8	56.6	1.8	17.7	76.4	70.9	6.8	9	2.4	7.2	1.5	18.9	0	10.2	70.9	5.9	2.6
G12141	83.5	14.9	1.7	11.3	0	4.1	2.6	73.7	2.45	13.2	8.7	8.1	62.8	3.1	13.8	80	79.8	8	4.2	2.4	6	2.1	14.8	0	8.4	76.8	5.6	2.2
G12142	81.8	16.6	1.6	12.5	0	5.1	3.9	73.6	2.43	13.5	10.2	8.3	59.5	2.9	14.5	77.8	78	6.4	4.5	2.1	6.6	2.6	14.8	0	8.7	76.5	5.7	2.1
G12147	85.2	8.7	6	6.1	41.7	4.1	3.5	64.2	2.14	20	10.7	9.1	53.7	1.9	21.1	72.9	66.5	7.9	10.5	4	8	1.8	21.9	0	12	66	6.2	2
G12149	81.8	7.2	2	8.2	26.3	4.5	2.6	71.3	2.35	14.8	10.7	8.6	57.2	2.3	16.5	77	75	6.6	6.4	3.5	7.3	2	15.8	0	10.6	73.6	5.7	2.3
G12150	90.5	8.1	1.4	10.9	0	4.5	3.7	69.1	2.31	13.9	9.7	9.7	57.6	2.3	15.3	77.9	72.9	6.6	7.8	3.1	7.3	2.2	18.1	0	10.4	71.4	6.4	2.3
G12151	91.9	5	3	0	33.1	4.6	3	66.4	2.19	18.3	9.2	10.6	53.1	1.8	20.1	74.1	67.2	6.6	11.1	3	8.1	1.3	20.7	0	11.1	68.2	5.2	3
G12152	79.9	13.5	6.6	7.1	46.4	3.7	4.5	66.2	2.2	17.5	10.3	8.8	57.6	1.6	18.4	76.8	68.4	6.6	10	3.4	7	0	21.8	0	10.4	67.8	6.1	2.2
G12153	83.7	7.8	6	0	32.1	4.1	2.9	68.7	2.2	18.5	8.9	9.6	56.1	0	19.4	74.7	67.3	8.2	7.8	3.2	8.3	1.8	18.9	0	11.5	69.6	5.6	0
G12154	86.1	7.8	6.2	0	24.6	3.9	5.8	65.6	2.21	16.8	9.6	7.6	57	2.2	19.3	76.8	69.4	6.5	9.8	3.8	6.9	0	22	0	10.7	67.3	6.2	0
G12155	89.4	8.3	2.3	0	0	4.6	4.5	66.5	2.22	16.8	9.2	9.6	55.8	1.9	18.7	74.5	69.9	7.4	8.6	3.5	7.6	2.2	20.4	0	11.1	68.5	6	2.4
G12158	87.2	9.6	3.2	9.7	37.5	5.5	4	69	2.32	14.8	9.8	9.5	56.8	2	16.5	76.5	73.2	7.5	7.6	2.2	7.8	1.5	19.1	0	10	70.9	6.4	1.9
G12161	89.1	12	1.3	12.3	0	5.9	4	72.9	2.46	11.8	10	8.5	60.7	2.9	12.2	79.3	79.2	7	5.1	2.6	6.3	2.6	15.6	0	8.5	75.8	5.4	2
G12163	89.1	9.3	1.6	12.8	42.7	5.5	4	68.2	2.32	15	9.8	8.7	56.6	2.7	16.9	75.5	74.7	6.7	8.1	3	7.6	2	18.5	0	10.6	70.9	6.4	2.5
G12164	89.8	9.4	0	0	0	4.5	3.6	71.6	2.34	14.1	9.1	8.8	60.7	1.6	16.9	75.5	73.9	6.7	6.2	3	7.1	1.9	16.5	0	10.3	73.2	6.4	2.9
G12166	98.44	1.56	0	0	0	5.4	3.02	72.35	2.39	14.75	9.07	8.47	60.12	1.74	14.75	77.81	75.97	7.41	6.28	2.86	6.34	2.04	16.71	0	9.2	74.1	4.69	2.17
G12170	87.7	9.9	2.4	12.1	0	5.2	4.8	67.8	2.32	13.4	9.7	8.8	57.9	2.3	15.4	77.4	73.7	6.9	6.9	3.8	7.4	0	18.6	0	11.3	70.1	6.4	2.6
G12174	99.17	0	0	0	32.8	5.13	2.61	75.77	2.43	14.47	10.56	7.97	59.82	ND	14.47	80.4	75.92	5.76	6.61	2.76	5.73	0	14.99	0.75	8.49	71.5	5.63	2.76
G12175	87.9	10	2.2	7.5	49.7	5.9	3.4	69.6	2.33	13.7	8.8	9.9	58.3	1.9	14.9	76.8	73.2	7.5	7.3	3.1	7.2	2.4	18.2	0	10.3	71.5	6.3	1.9
G12177	88.4	9.3	2.3	14.6	33.7	5.2	4	68.9	2.3	14.8	9.5	8.7	57.9	2.5	17.8	76	73.1	7.2	8.2	3	7.5	2.2	19.2	0	10.5	70.3	6.2	2.1
G12179	87.2	9.7	3.2	11.5	23.6	4.5	3.4	67.4	2.25	16.1	9.1	8.1	58.5	2.2	16.1	75.5	71.1	8.8	8.8	3.3	7.6	2.2	19.5	0	10.9	69.6	5.4	2.4
G12180	100	0	0	0	0	5.2	2.9	72.9	2.42	12.6	10.5	9.6	58.8	0	13.7	81.1	77	6.6	6.6	2.2	7.1	2	16.1	0	9.4	73.5	4.4	1.7
G12181	98.67	0	0	0	0	5.03	2.53	75.31	2.46	12.72	10.25	8.64	59.59	1.46	13.62	80.37	78.9	6.04	5.46	2.64	6.55	0.99	14.04	0	9.19	76.8	4.36	2.02
G12181	82.43	16.52	1.05	10.15	0	5.4	3.72	71.39	2.39	15.42	9.59	9.22	58.41	1.76	15.42	77.73	77.74	6.41	6.36	3.01	7.35	1.45	16.49	0	10.4	73.2	6.39	2.02
G9907	94.3	4.6	0	0	34.7	4.1	2.8	67.7	2.29	15.7	9.5	10.1	56.5	2.8	17.1	77.4	73.7	6.8	7.5	3.5	8.1	1.6	17.9	0	11.7	70.4	5	2.4
G9908	89.8	6.5	6.6	0	0	3.1	2.8	68	2.24	17.6	10.1	9.8	56.4	1.4	18.4	76.3	70.5	6.7	10.5	3.5	6.8	1.1	20.2	0	10.9	69.4	5.1	2.4
G9909	85.8	7.1	7.1	0	40.6	4.1	2.8	67.9	2.26	15.5	10.3	9.5	56.9	1.8	17.6	73.7	70.4	6.7	9.7	3.2	7.8	1.1	19.2	0	11	69.8	5.4	3.1
G9911	86.6	5.9	7.5	0	38.2	4.2	2.8	67.5	2.24	16.4	9.4	9.7	56.8	1.5	17.9	76.4	69.9	6.7	11	3.2	6.9	1.5	20.9	0	10.1	69.1	4.7	2.4
G9912	90.2	9	2.5	0	28.8	4.3	3.5	68.9	2.3	15.1	9.6	8.7	58	2.2	15.8	78.1	73.6	6	6.8	3.1	7.9	1.8	17.6	0	11.3	71.1	6.6	2.3
G9913	90.2	7.8	2.7	0	41.8	4.3	2.9	66.7	2.32	16.9	10	9.5	59.5	1.8	16.2	77	73.6	7.2	6.5	3.5	7.1	1.8	17.6	0	11.2	71.1	6.5	2.6
G9914	91.3	9.2	2.7	0	33.8	4.9	3.7	65.9	2.31	15.3	10.1	9.3	57.3	2.7	17.7	76.8	74.8	8.4	7.8	3.3	7.8	2	17	0	11.2	71.2	6.5	2.7
G9915	83.2	6.7	1.9	0	27.1	4	3.7	70.1	2.27	16.1	10.1	9.5	55.2	2.3	13.7	75.2	73.6	7.7	6.9	3.4	7.4	2	17.4	0	10.2	72.4	5.9	2.1
G9916	89.3	8.3	2.4	0	25.8	5.5	4.2	69	2.36	14.2	9.6	10	56.5	2.4	17	76.3	76.9	7.4	6.3	2.8	7.9	1.9	17.9	0	10.7	71.4	5.2	2.6
G9917	87.3	10.5	2.1	9.2	17	4.5	3.9	69.2	2.36	14.7	9.8	10.2	58.5	2.2	16.1	78.3	73.2	7	5.4	3.4	7.8	1.8	18.2	0	11.2	71.9	5.6	2.7
G9918	92.7	6	0	0	0	3.4	4.1	70.8	2.31	13.9	10.7	10	58.7	2.1	18.6	78.2	76.2	7.8	5.4	3.2	6.5	1.4	20.1	0	9.6	70.3	6	2.9
G9919	95	4.3	0	0	39.1	3.6	4.2	70	2.32	13.9	10	9.3	59.2	1.6	15.3	78.7	71.3	6.3	6.1	3.2	7.9	1.9	16.6	0	11.2	72.2	4.3	2.3
G9920	90.2	9	0	34.4	0	3.8	2.3	68.6	2.26	15.5	10.1	9.5	60.6	2.4	15.3	78.8	72.9	7.9	8.2	3.4	6.6	2.6	18.2	0	10.5	71.6	5.1	2.6
G9921	93.1	6.5	0	0	38.2	4.1	2.8	68.9	2.38	13.8	9.4	9.1	59.8	2.1	15.1	79.2	73.1	6	6.9	3.2	7.9	1.8	17.6	0	11.4	72.4	5.5	2.6
G9923	83.6	13.5	0	33.8	23.3	4.3	3.5	69.4	2.3	16.9	9.6	9.6	57	2.4	18.6	76.3	74.8	7.8	9.6	3.3	7.6	1.9	18.2	0	10.8	68.8	6.3	2.6
G9924	88.6	6.2	5.2	27.1	0	4.2	3.7	68.1	2.27	16	10.8	9.5	57.2	1.9	17.7	76.8	71	7.8	9.2	3.8	7.4	1.2	20.5	0	11.4	69.7	5.4	2.7
G9925	87.7	11	1.2	0	34.4	5.1	4.9	66.9	2.29	15.4	10.8	9	57.2	2.7	17	78.8	73.1	7	6.2	3.5	6.6	1.9	17.6	0	12.4	69.6	5.4	2.1
G9926	90.2	9	0	34.4	28.4	4.7	3	69.2	2.33	14.7	9.8	10.2	56.5	2.4	16.1	78.8	71.9	7.1	6.5	3.1	7.9	1.6	17.1	0	11.3	71.6	5.6	2.6
G9927	90.4	7	2.5	41.8	0	4.1	3.3	66.7	2.32	17.6	9.9	10.7	57	2.1	19.2	75.6	77.9	6.8	7.4	3.9	7.6	1.7	20.7	0	11.4	68.8	6.5	2.6
G9928	91.3	6	2.7	33.6	41.8	3.8	2.9	65.9	2.34	16.9	9.4	10.2	58	1.9	18.6	75.8	74.8	9.2	6.4	3.8	7.3	2.1	20.6	0	10.8	68.1	6.3	2.7
G9929	90.4	7	2.7	27.1	23.7	3.7	3.2	67.8	2.25	16	10.1	10.1	56.6	1.9	17.7	76.8	69.7	7.7	6.4	3.6	7.5	1.9	19.4	0	11	69.7	5.4	2.2
G9930	89.2	8.2	1.6	0	0	5.3	3.9	66.9	2.29	15.4	9.8	9	56.1	2.7	17	75.8	73.1	7.7	6.5	3.5	8.9	2	18	0	12.4	69.6	5.4	2.1
G9932	92.7	7.3	0	0	35.9	4.5	3.2	69.2	2.33	14.7	9.8	10.2	56.5	2.4	14.7	78.8	71.6	7.9	6.7	3.9	7.9	1.8	17.1	0	11.3	71.6	5.6	2.6
G9933	89.2	7.6	3.2	33.3	0	3.2	3.3	68.1	2.23	16.6	10.7	10	55.9	2.1	18.6	75.7	68.5	6.5	6.3	3.4	6.8	1.4	20.1	0	9.6	70.3	5.5	2.3
G9934	98.1	1.4	0	0	0	4.2	2.3	70.8	2.36	13.9	10	9.3	59.2	1.6	15.3	79.8	75.3	7.9	6.1	3.2	7.9	2	16.6	0	11.2	72.2	4.3	2.3
G9935	98.3	1.7	0	0	34	4.8	3.4	70.9	2.32	14.8	9.7	9.6	58.1	2.7	16	77.3	74.3	7.1	5.4	3.2	7.3	2	15.9	0	10.5	73.6	6.5	2.8
G9936	87.8	8.4	3.7	7.5	16	4.3	3.1	68.6	2.3	15.5	9.3	9.2	56.3	2.6	16.9	78.8	72.9	6.6	8.2	3.4	6.6	1.3	18.2	0	10.9	71	5.8	2.6
G9937	86.5	11.2	2.6	8.8	20.1	3.1	2.9	65.5	2.16	18.8	10.7	9.8	51.6	1.7	20.4	75.2	67.6	7.4	10.9	3.9	7.7	1.3	21.1	0	11.7	67.2	4.1	2.8
G9938	89.6	7.8	2.6	13.1	30.3	4	3.9	68.4	2.27	15.1	10.3	9.8	57.1	2.1	17	77.2	71.3	7.2	8	4.1	7.2	1.9	18.8	0	10.8	70.4	6.8	3.7
G9939	91	6.9	2.1	12.7	31.3	3.6	3.4	69.8	2.28	15.1	10.1	10.1	58.4	2.5	16.9	78.8	71.8	7.9	7.9	3.9	6.6	2	17.6	0	10.5	71.9	6.8	2.5
G9940	87.9	8.9	0	11.7	22	4.2	3.8	67.6	2.33	13.8	9.3	8.6	60.6	2.6	15.3	78.8	72.9	7.4	6.6	3.2	7.7	1.6	17.8	0	10.9	71.3	6	2.4
G9941	90	8.9	0	0	13.4	4.5	3.4	68.9	2.3	14.6	9.7	9.8	60.6	2.2	16.3	78	72.9	7	7.1	3.9	8.2	1.6	17.5	0	11.4	71.1	6.4	2.5
G9942	91	7.3	1.6	7.6	20.1	3.1	3.8	67.4	2.29	15.1	10.2	10.2	58.3	2.2	16.7	78.5	71.7	7.9	6.8	3.9	8.2	1.6	18.5	0	10.7	68.9	6.6	2.6
G9943	83.6	13	3.5	10.8	26.3	3.9	3.5	66.1	2.23	14.1	9.8	10.4	57.6	2.3	16.2	80.6	69.3	8.6	4.4	4.1	7.6	2.1	19.9	0	11.7	69.4	6.2	3.7
G9944	88.9	10.5	0	9.1	0	4	3.2	73.2	2.42	11.7	9	9.8	62	2.7	13.1	79.9	77.6	7.3	4.4	2.3	6.8	2.2	15	0	9.1	75.9	5.5	2.3
G9945	90.5	8.9	0	8.3	0	3.9	2.7	70.2	2.34	12.5	9.4	9	59.8	2	14.1	79.9	73.7	7.2	6.4	3.5	7.3	2.1	16.3	0	10.8	72.9	5.7	2.6
G9946	88.8	8.7	0	11.9	25.6	4.4	3.2	69.8	2.32	14.7	10	9.3	57.2	2.7	16.8	76.9	74.8	7.4	6.4	3.6	7.5	1.9	17.2	0	11	71.8	5.8	2.2
G9947	93.3	5.6	1.1	0	0	3.8	2.8	67.6	2.28	14.1	10.2	9.6	58.3	2	15.6	78	72.4	8.6	7.5	3	8.2	1.6	19	0	11.2	69.8	5.5	2.4
G9948	82.6	9.8	7.7	7.6	35.9	3.2	3.3	67.4	2.2	19	10.2	10.2	58.3	2.2	20	78.5	72.4	8.3	8.3	3.4	7.2	1.8	19.6	0	10.6	69.8	7.1	2.7
G9949	91.1	7.4	1.5	10.8	0	4	2.4	71.8	2.37	12.9	9.4	8.4	56.8	2.4	13.4	80	74.7	6.3	7.3	3.4	6.8	2.6	16	0	10.6	73.8	5.2	2.6
G9950	87.8	10.4	1.7	12.8	34	4.8	3.4	70.9	2.35	14.8	9.7	9.6	58.1	2.7	16	77.3	74.3	7.1	5.4	3.2	7.3	2	15.9	0	10.5	73.6	6.5	2.8
Average	89.41	8.23	2.25	6.21	18.16	4.49	3.53	69.40	2.31	14.98	9.79	9.17	57.92	2.17	16.31	77.39	73.37	7.24	7.11	3.20	7.24	1.78	17.96	0.01	10.45	71.56	5.85	2.33
SD	4.08	3.36	1.92	5.35	16.31	0.66	0.64	2.30	0.07	1.81	0.59	0.68	2.07	0.57	1.89	1.61	3.15	0.64	0.63	0.47	0.64	0.62	1.84	0.09	0.90	2.36	0.65	0.52
Min	79.90	0.00	0.00	0.00	0.00	3.10	2.30	64.20	2.14	11.70	8.40	7.60	51.60	0.00	12.20	72.90	66.50	5.73	5.76	2.00	5.73	0.00	14.04	0.00	8.20	66.00	4.10	0.00
Max	100.00	16.60	7.70	14.60	49.70	5.90	5.80	75.77	2.46	20.00	11.10	10.90	62.80	3.20	21.10	81.10	79.80	8.90	8.80	4.10	8.90	3.00	22.00	0.75	12.40	76.80	7.10	3.70

Table A.5: Raw predicted values for crude heparins using PCR. Raw predictions compared with expected, descriptive statistics for which are found in **Table 5.3**

Sample	DS		CS		DS2S		CS6S		GlcN,3-S		GlcUa-GlcN,3-S		IdoA2S-GlcN		DOS		GlcNAc-GlcA		GlcNS-GlcA		GlcNS-IdoA		GlcNS-IdoA2S		I2S-GlcNH2	
	Orig	Pred	Orig	Pred	Orig	Pred	Orig	Pred	Orig	Pred	Orig	Pred	Orig	Pred	Orig	Pred	Orig	Pred	Orig	Pred	Orig	Pred	Orig	Pred	Orig	Pred
Crude T-1	7.8	9.4	2.6	2.9	13.1	8.0	30.3	27.7	4.0	4.5	3.9	3.8	68.4	68.7	2.3	2.3	15.1	15.3	10.3	10.3	9.8	9.1	57.1	57.5	2.1	13.0
Crude T-2	6.9	8.3	2.1	1.7	12.7	4.8	31.3	15.5	3.6	4.5	3.4	3.6	69.4	69.8	2.3	2.3	15.1	15.2	10.1	10.1	8.9	9.1	58.4	58.1	2.5	15.3
Crude T-3	9.7	9.2	2.4	2.0	11.7	7.1	22.0	18.8	4.2	4.6	3.8	3.7	68.7	69.7	2.3	2.3	13.8	14.8	9.3	9.3	8.6	9.1	60.6	58.2	2.6	14.8
Crude T-4	8.9	5.5	1.2	0.7	0.0	6.7	0.0	15.8	4.0	5.0	3.4	3.4	68.9	71.2	2.3	2.4	14.6	13.7	9.7	9.7	8.8	8.9	59.1	58.7	2.2	12.3
Crude T-5	7.3	8.4	1.6	2.0	7.6	5.8	13.4	17.4	4.5	4.6	3.8	3.6	68.8	69.6	2.3	2.3	15.1	14.9	10.2	10.2	8.5	9.2	58.3	58.0	2.1	13.9
Crude T-6	13.0	12.8	3.5	3.1	10.8	8.0	26.3	20.7	3.9	4.3	3.5	4.0	66.1	68.5	2.2	2.3	14.1	15.2	9.8	9.8	10.4	9.3	57.6	58.3	2.3	18.2
Crude T-7	10.5	10.8	0.0	1.2	9.1	9.8	0.0	11.6	4.2	4.8	3.2	3.7	73.2	71.2	2.4	2.4	11.7	13.8	9.0	9.0	9.8	8.8	62.0	59.5	2.7	17.8
Crude T-8	8.9	11.0	0.0	1.9	8.3	7.0	0.0	10.6	3.9	4.6	2.7	3.8	70.2	69.9	2.3	2.3	12.5	14.0	9.4	9.4	9.0	9.3	59.8	59.2	2.7	17.9
Crude T-9	8.7	9.0	2.5	2.0	11.9	7.1	25.6	17.7	4.4	4.6	3.2	3.6	69.8	69.8	2.3	2.3	14.7	14.7	10.0	10.0	9.3	9.1	57.2	58.2	2.0	14.1
Crude T-10	5.6	8.1	1.1	2.3	0.0	1.9	0.0	16.5	3.8	4.1	2.8	3.6	67.6	68.7	2.3	2.3	14.1	16.7	10.2	10.2	9.6	9.3	58.3	56.9	2.2	14.4
Crude T-11	9.8	11.1	7.7	6.9	10.8	3.8	35.9	45.9	3.2	3.6	3.3	4.0	67.4	63.8	2.2	2.1	19.0	19.7	10.2	10.2	8.1	9.8	56.8	53.3	2.4	6.2
Crude T-12	7.4	7.4	1.5	2.4	0.0	1.4	0.0	9.2	4.0	4.5	2.4	3.6	71.8	68.6	2.4	2.3	12.9	14.3	9.4	9.4	8.4	9.8	62.6	58.4	2.0	13.8
Crude T-13	10.4	7.2	1.7	1.6	12.8	5.7	34.0	16.5	4.8	4.7	3.4	3.5	70.9	70.1	2.4	2.3	14.8	14.6	9.7	9.7	9.6	9.1	58.1	58.2	2.7	13.4

Sample	GlcNS		GlcA6S		GlcA-GlcNAc		GlcA-GlcNS		IdoA-GlcN		idoA-GlcN6S		GlcNH2		GlcA2OH		GlcA2S		IdoA2OH		IdoA2S		LR		GNR	
	Orig	Pred	Orig	Pred	Orig	Pred	Orig	Pred	Orig	Pred	Orig	Pred	Orig	Pred	Orig	Pred	Orig	Pred	Orig	Pred	Orig	Pred	Orig	Pred	Orig	Pred
Crude T-1	77.2	76.8	71.3	72.2	8.0	8.0	6.9	7.0	3.6	3.3	7.2	7.2	1.9	1.7	18.8	18.6	0.0	0.0	10.8	10.5	70.4	70.9	6.8	6.4	2.6	2.4
Crude T-2	77.5	77.5	71.8	73.7	7.9	7.9	6.3	7.0	3.9	3.3	6.6	7.2	2.0	1.8	17.6	17.6	0.0	0.1	10.5	10.5	71.9	71.8	6.6	6.1	2.5	2.6
Crude T-3	78.8	77.4	74.8	73.8	6.6	6.6	7.4	7.0	3.2	3.2	7.7	7.2	1.6	1.9	17.8	17.7	0.0	0.0	10.9	10.4	71.3	71.9	6.0	6.2	2.0	2.5
Crude T-4	78.0	78.3	72.9	75.3	7.1	7.1	7.0	6.7	3.2	3.0	8.2	7.0	1.6	1.6	17.5	16.8	0.0	0.1	11.4	10.0	71.1	73.1	6.4	6.0	2.5	2.5
Crude T-5	77.0	77.4	71.7	73.7	6.8	6.8	7.9	7.0	3.9	3.2	6.8	7.2	1.8	1.9	18.5	17.8	0.0	0.0	10.7	10.4	70.9	71.8	6.6	6.0	2.6	2.4
Crude T-6	77.9	76.7	69.3	72.9	8.6	8.6	7.8	7.3	4.1	3.4	7.6	7.3	2.1	2.4	19.9	18.3	0.0	0.0	11.7	10.7	68.4	71.2	6.2	6.6	3.7	2.5
Crude T-7	80.6	78.3	77.6	76.4	4.4	4.4	7.5	6.9	2.3	3.0	6.8	6.9	2.2	2.1	15.0	16.4	0.0	0.0	9.1	9.9	75.9	73.7	5.5	6.2	2.3	2.4
Crude T-8	79.9	77.5	73.7	75.2	6.4	6.4	7.2	7.2	3.5	3.1	7.3	7.2	2.1	2.6	16.3	17.1	0.0	0.0	10.8	10.3	72.9	72.7	5.7	6.1	2.6	2.3
Crude T-9	76.9	77.5	74.8	74.1	6.4	6.4	7.6	7.0	3.6	3.1	7.5	7.1	1.9	1.9	17.2	17.6	0.0	0.0	11.0	10.3	71.8	72.0	5.8	6.0	2.2	2.4
Crude T-10	78.5	76.9	72.4	72.0	7.5	7.5	8.6	7.1	3.0	3.5	8.2	7.4	2.2	1.6	19.0	18.5	0.0	0.1	11.2	10.9	69.8	70.5	5.5	5.9	2.4	2.6
Crude T-11	75.1	74.1	68.5	65.3	8.3	8.3	8.0	7.5	3.4	3.7	7.2	7.8	1.8	0.9	19.6	22.8	0.0	0.0	10.6	11.6	69.8	65.6	7.1	5.7	2.7	1.9
Crude T-12	80.0	76.9	74.7	73.4	7.3	7.3	6.3	7.6	3.4	3.2	6.8	7.6	2.6	2.7	16.0	18.1	0.0	-0.1	10.2	10.8	73.8	71.3	5.2	5.6	2.6	2.2
Crude T-13	77.3	77.7	74.3	74.2	5.4	5.4	7.1	6.9	3.2	3.2	7.3	7.1	2.0	1.8	15.9	17.5	0.0	0.0	10.5	10.3	73.6	72.2	6.5	6.0	2.8	2.5

Table A.6: Raw predicted values for crude heparins using PLS. Raw predictions compared with expected, descriptive statistics for which are found in **Table 5.5**.

Sample	DS		CS		DS2S		CS6S		GlcN,3-S		GlcUa-GlcN,3-S		IdoA2S-GlcN		DOS		GlcNAc-GlcA		GlcNS-GlcA		GlcNS-IdoA		GlcNS-IdoA2S		I2S-GlcNH2		GlcNAc	
	Orig	Pred	Orig	Pred	Orig	Pred	Orig	Pred	Orig	Pred	Orig	Pred	Orig	Pred	Orig	Pred	Orig	Pred	Orig	Pred	Orig	Pred	Orig	Pred	Orig	Pred	Orig	Pred
Crude T-1	7.8	10.3	2.6	3.5	13.1	9.5	30.3	30.9	4.0	4.4	3.9	3.8	68.4	68.3	2.3	2.3	15.1	16.0	10.3	10.3	9.8	9.1	57.1	56.9	2.1	13.0	17.0	17.5
Crude T-2	6.9	7.7	2.1	1.7	12.7	5.6	31.3	17.1	3.6	4.3	3.4	3.5	69.4	69.3	2.3	2.3	15.1	14.9	10.1	10.1	8.9	9.2	58.4	58.1	2.5	14.9	16.9	16.3
Crude T-3	9.7	9.5	2.4	2.1	11.7	8.1	22.0	18.9	4.2	4.5	3.8	3.6	68.7	69.5	2.3	2.3	13.8	14.7	9.3	9.3	8.6	9.1	60.6	58.3	2.6	15.0	15.3	16.1
Crude T-4	8.9	6.6	1.2	0.9	0.0	7.2	0.0	15.2	4.0	4.9	3.4	3.5	68.9	71.0	2.3	2.4	14.6	14.1	9.7	9.7	8.8	9.0	59.1	58.4	2.2	12.7	15.3	15.4
Crude T-5	7.3	8.4	1.6	2.2	7.6	6.5	13.4	19.2	4.5	4.4	3.8	3.6	68.8	69.3	2.3	2.3	15.1	15.0	10.2	10.2	8.5	9.2	58.3	57.9	2.1	13.5	16.7	16.4
Crude T-6	13.0	12.8	3.5	2.8	10.8	9.2	26.3	17.7	3.9	4.1	3.5	3.7	66.1	68.7	2.2	2.3	14.1	14.8	9.8	9.8	10.4	9.2	57.6	58.8	2.3	18.0	16.2	16.1
Crude T-7	10.5	10.7	0.0	0.7	9.1	10.1	0.0	9.4	4.2	4.8	3.2	3.6	73.2	71.2	2.4	2.4	11.7	13.2	9.0	9.0	9.8	9.0	62.0	59.9	2.7	18.1	13.1	14.4
Crude T-8	8.9	11.0	0.0	1.3	8.3	7.3	0.0	4.1	3.9	4.5	2.7	3.5	70.2	70.5	2.3	2.4	12.5	13.3	9.4	9.4	9.0	9.2	59.8	60.1	2.7	17.6	14.4	14.4
Crude T-9	8.7	9.2	2.5	2.0	11.9	7.8	25.6	17.9	4.4	4.5	3.2	3.6	69.8	69.7	2.3	2.3	14.7	14.7	10.0	10.0	9.3	9.1	57.2	58.3	2.0	14.3	16.8	16.0
Crude T-10	5.6	8.3	1.1	1.9	0.0	2.4	0.0	17.4	3.8	4.1	2.8	3.3	67.6	68.3	2.3	2.3	14.1	15.6	10.2	10.2	9.6	9.4	58.3	57.6	2.2	14.1	15.6	17.0
Crude T-11	9.8	9.9	7.7	7.2	10.8	5.0	35.9	49.3	3.2	3.6	3.3	3.8	67.4	63.9	2.2	2.1	19.0	19.4	10.2	10.2	8.1	9.6	56.8	53.7	2.4	7.2	20.0	21.2
Crude T-12	7.4	8.0	1.5	1.1	0.0	0.8	0.0	-7.9	4.0	4.4	2.4	3.1	71.8	70.7	2.4	2.4	12.9	13.0	9.4	9.4	8.4	9.4	62.6	60.4	2.0	14.1	13.4	13.9
Crude T-13	10.4	7.7	1.7	1.6	12.8	6.5	34.0	16.9	4.8	4.6	3.4	3.5	70.9	69.9	2.4	2.3	14.8	14.6	9.7	9.7	9.6	9.1	58.1	58.2	2.7	13.5	16.0	16.0

Sample	GlcNS		GlcA6S		GlcA-GlcNAc		GlcA-GlcNS		IdoA-GlcN		idoA-GlcN6S		GlcNH2		GlcA2OH		GlcA2S		IdoA2OH		IdoA2S		LR		GNR	
	Orig	Pred	Orig	Pred	Orig	Pred	Orig	Pred	Orig	Pred	Orig	Pred	Orig	Pred	Orig	Pred	Orig	Pred	Orig	Pred	Orig	Pred	Orig	Pred	Orig	Pred
Crude T-1	77.2	76.5	71.3	71.8	8.0	8.0	6.9	7.2	3.6	3.3	7.2	7.3	1.9	1.5	18.8	18.9	0.0	0.0	10.8	10.6	70.4	70.5	6.8	6.3	2.6	2.3
Crude T-2	77.5	77.6	71.8	73.1	7.9	7.9	6.3	7.2	3.9	3.4	6.6	7.3	2.0	1.8	17.6	17.8	0.0	0.1	10.5	10.7	71.9	71.4	6.6	6.1	2.5	2.7
Crude T-3	78.8	77.7	74.8	73.7	6.6	6.6	7.4	7.1	3.2	3.2	7.7	7.2	1.6	1.9	17.8	17.7	0.0	0.0	10.9	10.4	71.3	71.8	6.0	6.1	2.0	2.4
Crude T-4	78.0	78.1	72.9	75.5	7.1	7.1	7.0	6.8	3.2	3.0	8.2	7.1	1.6	1.6	17.5	16.8	0.0	0.0	11.4	10.0	71.1	73.0	6.4	5.8	2.5	2.3
Crude T-5	77.0	77.4	71.7	73.2	6.8	6.8	7.9	7.1	3.9	3.3	6.8	7.3	1.8	1.8	18.5	18.0	0.0	0.0	10.7	10.5	70.9	71.4	6.6	6.0	2.6	2.4
Crude T-6	77.9	77.4	69.3	73.0	8.6	8.6	7.8	7.4	4.1	3.4	7.6	7.3	2.1	2.3	19.9	18.0	0.0	0.0	11.7	10.7	68.4	71.4	6.2	6.3	3.7	2.5
Crude T-7	80.6	78.6	77.6	76.6	4.4	4.4	7.5	7.0	2.3	3.0	6.8	7.0	2.2	2.2	15.0	16.2	0.0	0.0	9.1	10.0	75.9	73.8	5.5	6.2	2.3	2.5
Crude T-8	79.9	78.6	73.7	75.7	6.4	6.4	7.2	7.2	3.5	3.1	7.3	7.1	2.1	2.6	16.3	16.6	0.0	0.0	10.8	10.2	72.9	73.3	5.7	5.8	2.6	2.4
Crude T-9	76.9	77.6	74.8	74.0	6.4	6.4	7.6	7.1	3.6	3.2	7.5	7.2	1.9	1.9	17.2	17.6	0.0	0.0	11.0	10.3	71.8	72.0	5.8	6.0	2.2	2.4
Crude T-10	78.5	77.3	72.4	71.3	7.5	7.5	8.6	7.4	3.0	3.7	8.2	7.5	2.2	1.8	19.0	18.6	0.0	0.1	11.2	11.2	69.8	70.3	5.5	6.1	2.4	2.8
Crude T-11	75.1	73.9	68.5	65.1	8.3	8.3	8.0	7.7	3.4	3.8	7.2	7.8	1.8	1.1	19.6	22.6	0.0	0.0	10.6	11.6	69.8	65.8	7.1	6.2	2.7	2.2
Crude T-12	80.0	79.0	74.7	75.7	7.3	7.3	6.3	7.2	3.4	3.0	6.8	7.2	2.6	2.8	16.0	16.5	0.0	-0.1	10.2	10.3	73.8	73.3	5.2	4.9	2.6	2.2
Crude T-13	77.3	77.7	74.3	74.1	5.4	5.4	7.1	7.0	3.2	3.2	7.3	7.2	2.0	1.7	15.9	17.5	0.0	0.0	10.5	10.4	73.6	72.0	6.5	5.9	2.8	2.4

Figure A.5: COS-Firs difference spectra of sub 1% (w/w) OSCS in heparin. Heparin: black, OSCS: red, variable region: blue
Spectra were calculated as explained in **Section 2.2.6**. Signals attributed to OSCS can be observed down to 0.06% (w/w) OSCS.