Development of Techniques for the Analysis of Protein:Glycosaminoglycan Interactions in Solution.

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

Written

By

Ashley Hughes

September 2014

Abstract:

Glycosaminoglycans (GAGs), such as Heparan Sulfate (HS), are vital components, of all multicellular organisms, with confirmed roles in development, cell-cell communication and diseases such as dementia and cancer. As such, the interaction of this important group of molecules with proteins is an important process to comprehend and improvements to current techniques are sought. To achieve this HS and its structural analogue heparin, were investigated in various biological systems to simultaneously improve our collective knowledge of the biological systems employed and to improve techniques available to study them. Hen Egg White Lysozyme (HEWL), a model amyloid forming protein, is found to be destabilised with a complex of HS in the zinc (HS(Zn)) from 67.7 °C to 57.6 °C compared to 65.3 °C in the presence of HS in the Na form (as measured by differential scanning fluorimetry). These complexes were studied structurally using synchrotron radiation circular dichroism (SRCD) and found to contain extremely similar secondary structure. To differentiate these similar structures principal component analysis (PCA) of thermal and UV degradations of the complexes using CD were analysed. The HS and HS(Zn) ligands were hypothesised to interact with the two tryptophan residues (Trp) in the active site of HEWL. To confirm this, magnetic circular dichroism (MCD) was also developed as a technique to observe direct Trp:ligand interactions. MCD demonstrated an interaction between HS/HS(Zn) and Trp in HEWL using a signal at the novel position of 286 nm. Antithrombin (AT) was studied in relation to the active pentasaccharide drug and inactive shrimp heparinoids (SH) that contain the important 3-O-sulfate, thought to be key for AT activity against factor Xa. These active and inactive complexes were found to be structurally similar

again when studied with SRCD so were subjected to UV degradation. The active and non-active compounds were finally differentiated via their induced stability in AT to thermal degradation. This important correlation challenges the current dogma, which relates activity with a particular pentasaccharide sequence. 2D generalised correlation techniques were applied to the data obtained in this thesis increasing the resolution of the data significantly, allowing for <1nm resolution in some examples. A separate method utilizing the same analysis techniques reveals subtle features within a spectrum, thereby increasing the information available and will be useful for future developments of algorithms assuring improved reliability. A new method of data handling is also proposed for Raman optical activity data, which allows for the correction of baseline drifts, enabling a series of data to be directly compared.

Acknowledgments:

As the last 8 years of education and progress towards the beginning of an academic career draws to an end, I would firstly like to give my thanks to my parents who have supported me at every step throughout the process, and despite my refusal to ‘grow up’ or stop ‘dressing like a slob’, I hope they will be proud of these achievements.

I also have to extend my deepest gratitude to all of those who have guided me through this PhD from the hopelessly amateur beginnings, fresh from undergraduate, to the semi-competent scientist I currently am. Without the patience and advice of Ed Yates and Tim Rudd especially, this thesis would not have been possible.

Thanks also to collaborators and friends for their scientific discussions and advice. This is aimed particularly at M. Lima of São Paulo and G. Siligardi of Diamond Light Source, but also the many people I worked with over these last four years.

A big thank you goes out to all those who have kept me relatively sane in the last few years by being by my side, pint in hand, talking rubbish, yelling at Leeds United/Leeds Rhinos and playing ridiculous air guitar. Particular culprits are Simon Hare and Andy ‘Bob’ Roberts in Leeds while in Liverpool all of Labs B and D deserve a mention for their company in the boozer, as well as scientific discussion. Thanks to Liane for being by my side and supporting me whilst I stressed over the writing of this thesis. I’ll look forward to supporting you in the same way!

Finally, thank you to the MRC and Diamond Light source for funding.

For Daz

Contents:

Figure list:

Figures 1-1 to 1-7 describe various aspects of amino acid, protein and glycosaminoglycan structure:

Figure 1-8 describes the CD effect on samples:

Figures 3-1 to 3-7 investigate the SRCD spectra of HEWL complexes:

Figures 4-1 to 4-11 relate to the development of MCD as a tool for probing direct Trp:ligand interactions:

Figures 5-1 to 5-7 investigate the interactions between AT and various active and non-active GAGs:

Figures 6-1 to 6-20 show the various covariance matricies produced via the four proposed methods:

List of tables:

List of abbreviations:

A: Adenine

AD: Alzheimer's Disease

AT: Antithrombin

C: Cytosine

CD: Circular Dichroism

CS: Condroitin Sulfate

DS: Dematan Sulfate

DSF: Differential Scanning Fluorimetry

ECM: Extra Cellular Matrix

FGF: Fibroblast Growth Factor

G: Guanine

GAG: Glycosaminoglycan

Gal: Galactose

HBP: Heparin Binding Protein

HEWL: Hen Egg White Lysozyme

HIT: Heparin-Induced Thrombocytopenia

HS: Heparan Sulfate

HY: Hyaluronic acid

IR: Infrared

KS: Keratan Sulfate

LMWH: Low Molecular Weight Heparin

MCD: Magnetic Circular Dichromism

NA: N-Acetylation

NMR: Nuclear Magnetic Resonance

NS: N-Sulfation

PAPs: 3'phosphoudenosine 5'phosphosulfate

PCA: Principal Component Analysis

PD: Parkinson's Disease

PG: Proteoglycan

ROA: Raman Optical Activity

SAX: Small Angle X-Ray Scattering

SD: Standard Deviation

SH: Shrimp Heparinoids

SOD1: Super Oxide Dismutase 1

SRCD: Synchrotron Radiation Circular Dichroism

T: Thymine

UFH: Unfractionated Heparin

UV: Ultraviolet

VCD: Vibrational Circular Dichroism

Chapter 1: General Introduction.

1.1. Protein structure.

Proteins are versatile macromolecules that serve crucial functions in all biological systems. They are made up of amino acids and are encoded for by genes contained within the DNA sequence of an organism.

DNA is the basic code for proteins consisting of the bases adenine (A), thymine (T), guanine (G), cytosine (C) where A hydrogen bonds to T and G to C. The bases themselves are sequentially arranged along a sugar phosphate backbone, and sequences of 3 bases are used to encode for individual amino acids. These genetic codes, the order in which bases appear along the DNA strand, are translated via the evolutionary older mRNA, and tRNAs, into a linear sequence of amino acids. These then fold into unique conformations in order to perform their biological functions within the organism.

There are 20 amino acids that make up the proteins for all organisms, and each amino acid has a unique side chain with its own properties and functions. The majority of amino acids are found in the L isomer, and at physiological pHs exist as zwitterions of free amino acids. The basic structure of an amino acid is shown in Figure 1-1.

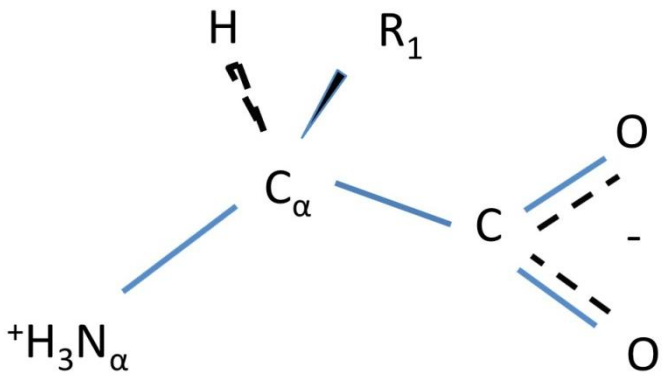

Figure 1- 1: The basic structure of an L-amino acid, the R_1 group is the location of the side chain that makes each of the 20 amino acids unique. In this representation, the C-H bond is facing away from the observer into the page.

The pK_a values of the amino group and carboxylic acid result in a dipolar molecule with a positive charge on the nitrogen and a negative charge on the carboxylic groups [1]. These monomers are bonded together via the loss of water to form polypeptides (below ~50 amino acids, and referred to as proteins above that). The bond is created through linking the carboxyl group to the amino group to form a peptide bond (see Figure 1-2).

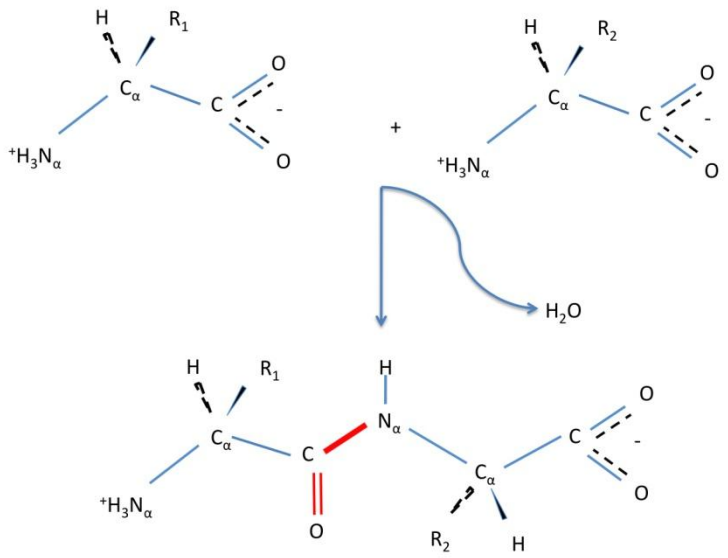

Figure 1- 2: Two amino acids are covalently bonded together through the loss of water and the creation of the peptide bond (Red bond).

There are several structurally important aspects of the peptide bond that ultimately influence protein structure. It is an essentially planar bond, resulting in a flat uniform backbone from which the amino acid side chains protrude. Another key feature is the partial double bond nature of the C-N within the peptide bond. This bond resides somewhere between a single bond and a double bond (Figure 1-3A) as demonstrated by the bond lengths. A C-N single bond would have a length of 1.49 Å while a double bond would be shorter at 1.27 Å, the peptide bond meanwhile resonates at 1.32 Å, demonstrating double bond features and therefore restricts the rotation around this bond.

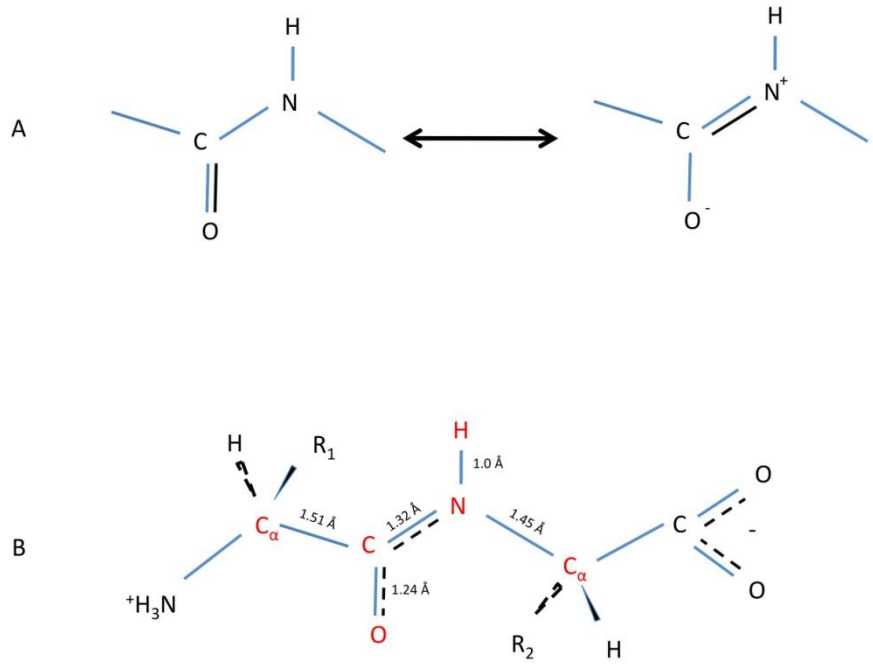

Figure 1- 3: A: The resonant nature of the double bond in equilibrium leading to the planar peptide bond. B: Shows the six atoms that are in the same plane with one another over a peptide bond between two amino acids. The peptide bond has a partial double bond demonstrated by its short C-N bond length. Rotation is allowed around the N-C_α and the C – C bonds.

The peptide bonds do have some rotational flexibility around the N → Cα (Phi) and Cα→ C^1 (Psi) bonds. This rotational flexibility is what gives proteins the ability to fold to form the common secondary structures. The folding is dictated by the properties of the amino acid side chains and, the order in which they are added to the

polypeptide chain. These are then stabilised by many interchain and chain-solvent interactions.

The side chains of amino acids are summarised in Table 1-1. They fall into several categories that are broadly, hydrophobic, charged and uncharged, some however, have specialised functions within proteins.

Table 1- 1: A list of the 20 amino acid properties.

Group	Name	Functional Group	Notes
Special case	Glycine	-	Unique as Achiral
Hydrophobic	Alanine	-	-
Hydrophobic	Valine	-	Non-polar, Hydrophobic
Hydrophobic	Leucine	-	Non-polar, Hydrophobic
Hydrophobic	Isoleucine	-	Non-polar, Hydrophobic, Additional chiral center
Hydrophobic	Methionine	-	Non-polar, Hydrophobic
Special Case	Proline	-	Side chain linked to C_α and N_α
Hydrophobic	Phenylalanine	-	Hydrophobic
Hydrophobic	Tyrosine	O-H	Hydrophobic
Hydrophobic	Tryptophan	N-H	Hydrophobic
Uncharged	Serine	O-H	Hydrophilic, Site of glycosylation
Uncharged	Threonine	O-H	Hydrophilic, Additional chiral center
Uncharged	Glutamine	NH_2	Hydrophilic
Uncharged	Asparagine	NH_2	Hydrophilic
Special case	Cysteine	S-H	Two can Oxidise to form disulfide bond
Charged	Lysine	NH_3^+	Full positive charge
Charged	Arginine	NH_2/NH_3^+	Full positive charge
Charged	Histidine	-	Full positive charge/No charge at pH 7, pK_a 6.0
Charged	Glutamic Acid	COO^-	Negative charge
Charged	Aspartic Acid	COO^-	Negative charge

The amino acids encoded for by the genetic code are joined together with peptide bonds to form the primary sequence of proteins. These amino acids must then fold together into local secondary structures to ultimately form the overall tertiary structure of the folded protein. This is achieved through various interchain and solvent-based interactions, predominantly through non-covalent bonding. Many of these interactions are heavily dependant on charges carried by the amino acid groups, and these are summerised in Table 1-2.

Table 1- 2: A list of (protein side chain) pK_a values, and their charge transitions at this value.

Group	pKa	Charge
Aspartic Acid	4.1	0 → -
Glutamic Acid	4.1	0 → -
Histidine	6.0	+ → 0
Amino Group	8.0	+ → 0
Cysteine	8.3	0 → -
Lysine	10.8	+ → 0
Tyrosine	10.9	0 → -
Arginine	12.5	+ → 0

The charges on the chemical groups listed in Table 1-2 contribute to the folding of a protein through non-covalent forces. The non-covalent forces are relatively small, but numerous so collectively are essential for correct folding of proteins. Forces affecting the correct folding are Van der Waals forces, hydrogen bonding and dispersion forces. Dispersion forces are the electrostatic attraction of molecules that are given small dipole moments by the oscillation of the electrons, these are

countered by the repulsion of electron shells from one another, although the overall effect is a net attraction. One of the most important forces in protein secondary structures are hydrogen bonds, and a vital acceptor/donator pair for this is the protein backbone which has a small negative charge on the C=O and a small positive charge on the amino group allowing a hydrogen bond to form within the backbone and stabilizing the secondary structures formed. The major hydrogen bond forming groups are described in Table 1-3, along with the bond distances (between donor and acceptor atoms) as a direct measure of bond strength.

Table 1- 3 A list of important groups involved in inter-chain hydrogen bonding in protein structure and their relative bond lengths (Å), representing the bond strengths. The amide carbonyl hydrogen bond is essential in protein secondary structure.

Name of hydrogen bond	Groups	Distance between atoms (Å)
Hydroxyl-hydroxyl	O-H…….O-H	2.8 ±0.1
Hydroxyl – carbonyl	O-H……C=O	2.8 ±0.1
Amide – carbonyl	**N-H……C=O**	**2.9 ±0.1**
Amide – hydroxyl	N-H…….O-H	2.9 ± 0.1
Amide – imidazole nitrogen	N-H…….N=	3.1 ± 0.2
Amide - sulfur	N-H…….S-	3.7

These hydrogen bonds serve to stabilise interactions within the protein chain to form the essential secondary structures in proteins. These can broadly be summarised as α-helix, β-sheet, β-turns and random coils. Alpha helical structures are tight coils of the protein backbone with the amino acid side chains sticking outwards at regular

intervals. Although left-handed helicies are possible, they exist predominantly in a right-handed form, as this is more energetically favourable. In this arrangement the amino acids are tightly packed, with one amino acid being 1.5 Å higher up the helix than the previous at a 100° angle. This twist in the amino acid at each residue results in one full turn of the helix taking on average 3.6 amino acids and enables the CO and NH groups to line up, allowing extensive hydrogen bonding. The CO group in the backbone hydrogen bonds to the NH group of the amino acid 4 places further down the sequence and the alignment of all the amino acids enables every amino acid to be hydrogen bonded to another, excepting the extremities of a helix. This results in a stable structure and can be seen in Figure 1-4.

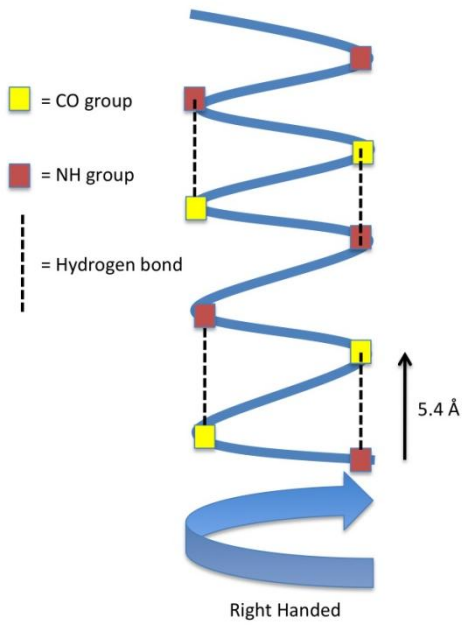

Figure 1- 4: A representation of a right-handed α-helix and the hydrogen bonding that stabilises the structure.

β-sheets are strands of a protein that are in a flat extended form and have hydrogen bonding stabilizing them. These structures have the amino acids in a much more elongated form compared to α-helix structures with an average of ~3.5 Å between each residue. The amino acids side chains in β-sheets alternate and protrude from opposing sides of the plane of the backbone of the strand. There are two forms of β-sheet; parallel where the two strands both run in the same direction, and antiparallel where the direction of the protein backbone is in opposite directions. β-sheets in proteins can exist in one, the other, or both forms within the same larger structure within a protein and are commonly found with 4 or more strands stacked upon one another. The hydrogen bonding is different for each example with the antiparallel sheet being a simple example of each CO group lining up and bonding with its opposite NH group. However, the parallel β-sheet each amino acid bonded to two other amino acids, the NH group bonds with the CO of an amino acid opposite and the CO group of the same residue is then bonded to the NH group of an amino acid two residues further down the chain. This can be seen in Figure 1-5.

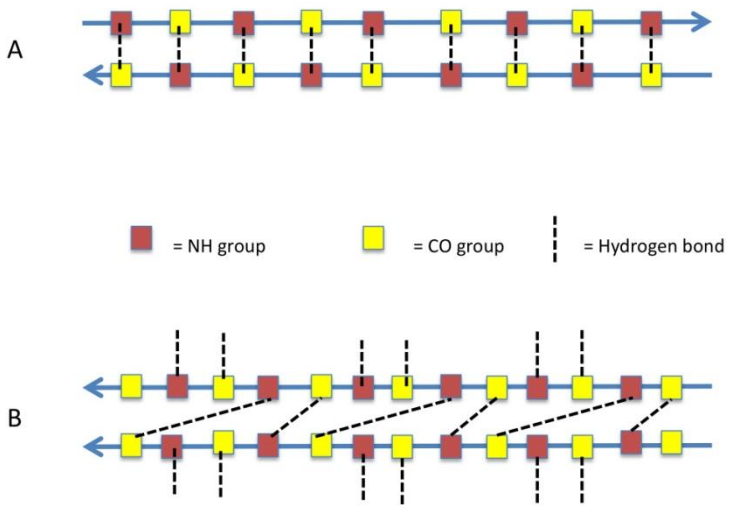

Figure 1- 5: A representation of the hydrogen bonding present in anti-parallel (A) and parallel (B) β-sheet secondary structures.

Another common structure is the β-turn in which the direction of a protein backbone can be reversed, essential in forming globular proteins. These are stabilised by the use of a hydrogen bond between the CO group of amino acid i and amino acid i_{+3} shown in Figure 1-6.

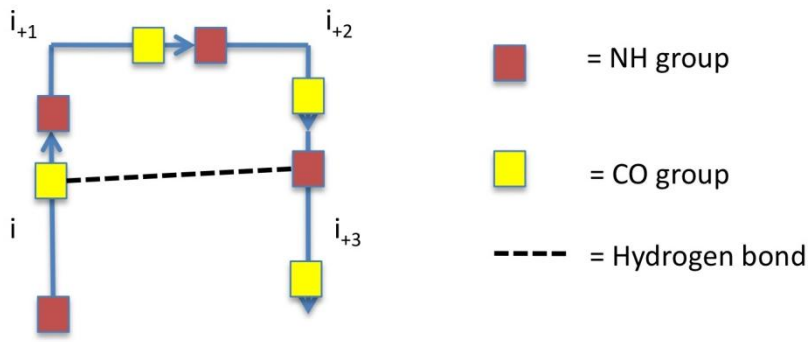

Figure 1- 6: A representation of a β-turn in which the carboxyl group of the amino acid i is hydrogen bonded to the amino group of i_{+3}.

There are also more loosely defined structures referred to collectively as random coil, or loops. There are no regular or periodic structures within these loops, however, they are often rigid and defined. They are found at the surfaces of proteins and can be important in protein:protein and protein:ligand interactions.

Tertiary structure allows pairs of cysteine amino acids to come into contact with one another and form disulfide bonds through an oxidation reaction. These are used to stabilise structures within the protein, and are especially prevalent in extra-cellular proteins. Once the production of a protein is completed, and as it folds into its native state, the proteins can then be modified in various ways, for example glycosylation and phosphorylation. These modifications can further modify the structure, function and stability of the protein.

Hydrophobic forces also play a significant role in correct protein folding in addition to the binding energies discussed. Three forces also have a large impact on the overall stability of the protein as the hydrophobic amino acids are preferentially buried in the center of proteins to avoid contact with polar solvents, such as in aqueous environments. An overview of the non-covalent forces acting upon a protein are presented in Table 1-4.

Table 1- 4: An overview of the energies of non-covalent interactions that act upon a protein to create the native conformation in an aqueous solution. The hydrophobic effect value given is for a single CH_2 bond [2].

Type	Example	Binding Energy (Kcal/mol)
Dispersion forces	Aliphatic hydrogen	-0.03
Electrostatic interaction	Salt bridge	-5
Hydrogen bond	Protein backbone	-3
Hydrophobic forces	Phenylalanine side chain	1

This thesis will describe studies of the average secondary structures of proteins, and how they change upon an interaction with a GAGs in solution for which there is a dearth of available techniques. The main technique used for this purpose is circular dichroism (CD) and is described in section 1.4 of this general introduction.

1.2. Glycosaminoglycans.

Glycosaminoglycans (GAGs) are long linear negatively charged polysaccharides. GAGs are thought to have evolved by the evolutionary clade, Eumetazoa [3]. They have become an essential component of the extra cellular matrix (ECM), present on practically all mammalian cells [4]. There are several polysaccharides within the GAG family classification; hyaluronic acid (HY), the two main chondroitin sulfates A and C (CS-A and CS – C), dermatan sulfate (DS), keratan sulfate (KS), heparan sulfate (HS) and heparin. With the exception of HY, the only unsulfated GAG in the family all other members of the GAG family contain some level of O-sulphation. The development of HS regulatory proteins has been associated with the onset of multi-cellular life. Some simple multi-cellular organisms have been shown to possess components of the HS biosynthetic equipment, and they have even been found in some unicellular organisms [5]. They are naturally found attached to proteins forming proteoglycans (PG), and each PG can have up to as many as 100 GAG chains attached [4]. HS and heparin are unique in that they also poses varying degrees of N-sulfation through their chains [6].

This work focuses on the GAG HS, and the closely related heparin. They share the same backbone disaccharide repeating units of β-D-glucuronic acid-(1→4)-α-D-glucosamine or α-L-iduronic acid-(1→4)-α-D-glucosamine (Fig 1-7). These disaccharides can be modified through the additions of sulfate groups at specific positions. For HS and heparin these are on the oxygen on the C-2 carbon of glucuronic acid and iduronic acid and the glucosamine can commonly be O-sulfated

at the C-2 carbon position on the amine group, as well as at the C-6 carbon position. In rarer circumstances the glucosamine can also gain an extra sulfation modification at the C-3 carbon oxygen position [6]. Since it shares a common underlying structure, shares similar activates and is relatively abundant, heparin, the pharmaceutical, is used as a proxy for HS in many biochemical investigations [7,8]. HS is found attached to a protein component, an example of this is the protein family of syndicans. These are a family of PGs that have been linked with cell proliferation, cell-cell adhesion and ligand binding through their HS components [9].

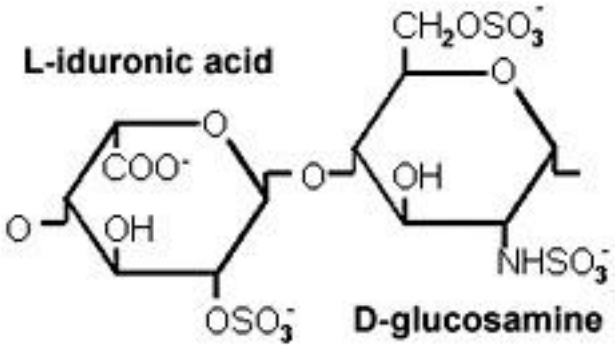

Figure 1- 7: A representation of the major repeating unit within heparin and Heparan sulfate, in its tri-sulfated form. Image is reproduced from reference [10]

Heparin and HS biosynthesis takes place within the lumen of the endoplasmic reticulum-trans Gogi network inside cells. Heparin is found in mast cells of connective tissue, and HS and heparin are differentiated in various ways, one of which is the N-sulfation (NS) – N-acetylation (NA) ratios, with heparin having a <2:1 ratio of NS to NA [11]. Heparin is also cleaved from its PG by a specific endo-β-D-glucuronidase [12] to produce heparin fragments of 3 – 30 kDa. HS

polysacchairde chains are longer than heparin with an average of ~30 kDa, as opposed to the ~15 kDa of heparin chains [13] and, unlike their heparin counterparts, HS chains have some organisation present, with NA and NS domains being apparent [14]. Heparin and HS are made up of combinations of 4 hexuronic acids and 6 different glucosamines in disaccharide units and share the same biosynthetic pathway [15].

The initiation of the heparin and HS biosynthetic pathway is the xylosylation of a serine residue, followed by addition of a β(1→4) Gal, a β(1→3) Gal, β(1→3) GlcA and finally an α(1→4) linked GlcNAc residue [16]. This final GlcNAc monosaccharide is what determines that this will ultimately become a HS/heparin chain, while the addition of a GalNAc instead would result in a CS or DS chain being polymerised. Following the GlcNAc addition, there is a stepwise addition of the unmodified backbone units of GlcA and GlcNAc to the non-reducing end of the polysaccharide chain. The alteration of monosaccharides is achieved through apparent substrate specificity of the enzymes [17,18,19]. This polysaccharide chain is then modified sequentially by a number of enzymes, the whole process taking <1 minute from start to finish [20]. The first modification step, and a key regulatory step in the overall modification of the chain, is the deacetylation of the N-actylyglucosamine residues. This step is a key regulatory step as it is a prerequisite of N-sulfation, which is the next step performed on the now free amino groups. These two reactions are thought to be coupled *in vivo* [21] and are catalysed by the same protein [22]. Once the N-sulfation of the glucosamine is completed, another key regulatory step that, if not performed, prevents the subsequent modification of

the D-glucuronic acid [23] is epimerisation at the C-5 position to form L-iduronic acid (IdoA). The IdoA can then be modified at the C-2 position with O-sulfation, and also the GlcNS can be O-sulfated at the C-6 position [15,20,23,24,25]. The transfer of sulfates onto their respective groups requires a 3′-phosphoadenosine 5′-phosphosulfate (PAPs) as a sulfate donor. This is known to influence the properties of the degree of N-sulphation within a HS chain [26]. In addition to these modifications, there are two rarer modifications that can occur, the glucuronic acid can be O-sulfated at the C-2 position and the C-3 of GlcN. The biosynthesis of HS and heparin are likely to entail some distinct features, not accounted for here, because heparin lacks the domain structure of HS. The domains of HS allow for active NS regions of highly modified sugars, hypothesized to be the regions of most biological activity [27] to be separated by the more flexible, but much less modified, NA domains. The ratio in HS for NS and NA units is reported to be around 50:50, but this can change depending on the tissue type, and this ratio is observed in tissues. However, when the required components are added together *in vitro,* a ratio approaching 75:25 (NS:NA) appears implying some external regulation mechanism [28].

The above mechanism is the conventional view of HS and heparin biosynthesis, it is however not the only proposed route for creating the required products. A new mechanism has been proposed following an overview of the available data in which the large redundancy in the structural geometries was noted for different sulfation patterns [29,30]. The analysis conducted was based upon the known substrate specificity for the enzymes and employing an approach based on graph theory to

reveal an efficient tree structure for the biosynthetic pathway. This tree structure has two main branches based around the initial N-sulfation event. The division of the biosynthetic pathway in this manner, coupled with the restrictions on substrate specificities leads to the abundant structures all being on a single branch, whilst rarer structures lie on the minor branch [5]. This branching of the biosynthetic pathways requires that the C-5 epimerase have dual substrate specificity and is able to convert both the GlcA-GlcNA and GlcA-GlcNS into their respective IdoA forms, or the existence of a separate epimerase for this particular substrate. The product of this hypothesized epimerase has been observed however [31]. There is evidence to support the hypothesis that the same epimerase acts on both substrates as a knockout of this enzyme results in no IdoA being detected at all, and an increase in GlcA-2S products [32]. The synthetic route is also supported by analysis of the HS present in the lethal knockout of the C-5 epimerase [33]. This proposed route of biosynthesis has two major benefits for the biological system, firstly it is very efficient in that all 6 common structures are attainable through just five enzymes and 4 steps. This system also provides a mechanism which creates a large variety of possible structures through alterations in glycosidic linkage positions or ring conformations at each step, so again very efficient if the overall aim of the process is to create a particular conformation and charge density combination [5]. The idea is then carried on to a hypothetical situation where single position frame shifting is able to take place, providing a hypothetical mechanism whereby the blocking or modifications on chains can be attained and therefore a possible mechanism by which the domain structure of HS could be attained.

The synthesis of heparin and HS is by no means a simple process, and there are many things yet to be understood about the process, which has no apparent template or known means of regulation. It has been suggested, however, that there exists some form of feedback mechanism whereby changes to the sulf enzymes (that can remove sulfates from HS chains after they have been released from the Golgi) can influence the biosynthesis of subsequent HS chains [34].

The biosynthetic pathway and its regulation lead ultimately to the sulfation sequence of heparin and HS molecules. There has been much work in studying this sulfation sequence in an attempt to link this aspect to biological activity in a sequence specificity manner [35,36,37]. This cumulated in the now widely used pharmaceutical based on an antithrombin (AT) binding pentasaccharide sequence, based around the rare 3-O-sulfate upon the GlcNS saccharide, that is known to increase the activity of this important blood anti-coagulant enzyme [38]. The details of this discovery are discussed within the introduction to Chapter 5. There has, in subsequent studies, been evidence of a correlation not only of the sulfation pattern, but also charge density and binding [39], although this correlation could be attributed to binding experiments that contain an inherent bias towards more charged molecules [36]. The most likely requirement for biological activity appears to be a combination of several factors; appropriate conformation, charge density and molecular flexibility to allow binding to the complementary protein surface via interactions, predominantly, with lysine and arginine side chains [6]. This idea is supported by the observation that non-GAG plant polysaccharides that have similar charge density and conformation to active GAGs are capable of inducing activity in Fibroblast

Growth Factor 2 (FGF2) cell signalling and also in vascular endothelial growth factor [40,41,42]. The pentasaccharide sequence, the sugar that has a reported specific sequence – activity relationship for to AT changes its binding properties when it is extended, and is able to have some leeway regarding sulfation positions, again lending support to the importance of the overall conformation, and charge distribution [43,44]. The issue is further confounded by the additional knowledge that not only will the overall charge density and sulfate sequence of a HS molecule alter its biological activity, but so will a change in its cation state [45,46].

A difficulty faced when attempting to address the question of sequence-activity relationship is the use of heparin libraries of limited structural diversity [47]. Another is the difficulty of obtaining pure oligosaccharides from an initially mixed pool. There are currently few examples beyond the AT example that demonstrate a sequence specificity for protein binding and activity [48] and this is further addressed in Chapter 4. HS and heparin protein binding is also complicated by overlapping protein binding sites [49], as well as a degree of redundancy [50], for example in the case with the hepatocyte growth factor requiring any two sulfates in the disaccharide to bind, regardless of position [43].

1.3. The heparan sulfate interactome.

HS interactions with their various protein ligands are enabled through location of their core proteins, of which there are two main families; the syndican family of which there are four, and the glipican family there are six. The syndicans are a

family of proteins that contain a transmembrane core protein to which either HS or CS chins are attached. These polysaccharide chains then extend out into the pericellular matrix [51]. The other major family of proteins that HS is associated with is the glypicans. These are a group of proteins that are anchored to the plasma membrane via a hydrophobic region at the C-terminal. HS attachment is restricted to only the last 50 amino acids, keeping the chains relatively close to the membrane [52]. HS chains are also known to be present on a range of proteins present in the ECM such as perlecan, agrin, and collagen XVIII [53].

Heparin and HS bind to, and affect the biological activities of, a wide range of proteins primarily in the ECM. In an extensive systems biology investigation into the proteins that interact with heparin and HS, 435 proteins were found [54], covering a wide range of biological activities. This expands on the 216 proteins found to bind HS just 3 years previously [55]. The group of proteins that interact with HS and heparin, termed the 'HS interactome', includes many important classes of biological activity. HS has been shown to have a role in the infection of mammals [56], and plays a key role in embryonic development [57]. In addition, HS has been shown to have biological activity in the inflammation and immune response [58,59], angiogenesis [60], morphogenesis [61], ECM assembly [62] and the coagulation cascade [63].

The bioinformatic approach used in the study of heparin/HS interaction networks has revealed several key pieces of information; First it is noted that the interactions formed by HS and heparin with ECM proteins are very highly clustered in

comparison to the interaction networks of non-heparin binding proteins. This places HS at the centre of many interaction nodes between signaling cascades, potentially playing an important role in the mediation of information between the extra and intracellular signaling pathways [54].

Further analysis of the heparin binding proteins (HBP) shows a structurally very diverse group of proteins. This could be a result of the many modes of binding thought to be available through the structural diversity available in GAGs. Analysis of the evolution of these HBPs shows a strong Pearson's correlation coefficient between their evolution and the complexity of the organism, suggesting a need for HS interactions in higher animals. This suggestion is reinforced when studying the presence of HS biosynthetic enzymes, again their presence correlating with higher organisms [54].

1.4. Methods overview.

There are many methods for studying protein structure and these can come in various degrees of resolution, precision, and accuracy. Often seen as a gold standard in protein structural analysis is X-ray crystallography. This technique relies on the creation of crystals which are then subjected to high intensity X-rays which interact with the atoms of the protein and diffract to give a specific pattern that is directly linked to the organisation of the atoms within the crystal, and therefore the structure of the protein. While this technique provides extremely good resolution, in the best case resolving every atom individually, it suffers from several issues. Firstly, in

relation to the work to be undertaken herein, cystallisation cannot be achieved when working with highly flexible and heterogeneous GAG polysaccharides. It also suffers in that proteins are studied in non-native conditions that lack buffer, salt, and physiological pH conditions.

Nuclear Magnetic Resonance (NMR) is another high-resolution technique that is suited to studying protein structure. This technique, unlike x-ray crystallography, is a solution-based method, enabling physiological conditions to be mimicked during experiments. This method, in simple terms, probes the local environment of NMR active nuclei (eg ^{1}H, ^{13}C, ^{15}N), which is altered depending on how they are interacting with the rest of the protein or solvent and thereby provides information on protein conformation. This technique can be used to observe interactions in solution, however, a major drawback is the broad signals for large and immobile proteins and polysaccharides. It is difficult to study protein:GAG interactions using this technique for several reasons such as the size limit (30 kDa) and fast relaxation times of proteins, the multiple binding sites on GAG polymers combined with sample precipitation and the need for isotropically labeled protein make it a challenging technique. The signals from both components overlap, the size and heterogeneity of HS samples with multiple binding modes and line broadening. It has been possible to study some protein:GAG oligosaccharide interactions via NMR for heparin/HS interactions with FGF 1 and 2 [64,65] and has also been achieved with a tetrasaccharide – protein interaction following cross-linked basic side chains [66]. NMR is also one of the key techniques for studying the conformation of GAGs in solution. It is able to provide information on ring puckering, glycosidic linkage

conformation and average sulfation positions with a sample [29,30,67,68,69,70]. NMR is also capable of investigating the overall flexibility of GAGs, as shown in a study on the flexability of the NA domain in HS with the use of ^{15}N labeled HS [43]. There are efforts to advance the structural information available through NMR via chemical and enzymatic methods to synthesise heparin oligosaccharides that are ^{13}C and ^{15}N labeled [71,72]. NMR requires relatively high concentrations of sample (0.2 – 2 mM [73]), which can lead to aggregation in some proteins. For structural work to take place, NMR stable isotopes (^{13}C and ^{15}N) are required, combined with multidimensional NMR experiments [73].

Small angle x-ray scattering (SAX) is another option to look at protein conformation in solution through its overall shape in solution and is very low resolution. It is however very useful for studying aggregation. This technique can provide information regarding the overall shape of a molecule in solution. SAX is currently limited to this type of analysis as the molecule is free to rotate in solution resulting in the gathering of the average scattering across all possible orientations and is thus a low-resolution technique.

X-ray scattering is a technique that has been used to study the properties of some heparin oligosaccharides, finding the heparin to be semi-rigid, extended molecules in solution [74]. Another little used technique being developed to study GAGs in solution is Terahertz (10-12 Hz) spectroscopy which can observe the fundamental vibrational modes of heparin present at physiological temperatures [75]. The basis of this technique is that electromagnetic radiation in the THz frequency range

corresponds to energies of large scale molecular movements of macromolecules (or other materials) such as flexing of protein domains, or bending of linear polysaccharides, which are sometimes termed phonon modes. A frequency of about 6 THz corresponds to energies (related by kT) that can be provided by ambient (i.e. biologically relevant) temperatures hence, absorbance in this range relates to the large scale flexing motions that occur in these molecules under natural circumstances. These molecular motions, of relevance to biological activities through such properties as enzyme activities, protein and polysaccharide dynamics and binding properties, cannot be observed by other spectroscopic methods, which all operate at higher energy levels.

Another technique which has been developed in the structural biology field is cryo-electron microscopy (cryo-EM), and recently it has been used to identify and characterise protein:GAG interactions [76,77]. This technique enables atomic or near-atomic resolution of GAG interactions at liquid nitrogen temperatures. It has also been used to study disordered protein and the variability found in amyloid fibrils [78]. Whilst cryo-EM can offer increased resolution of protein-GAG interactions, it lacks the ability of solution state techniques to dynamically investigate the stability of complexes or the binding kinetics of protein-ligand interactions.

Infrared spectroscopy (IR) has previously been used in the past to study protein conformation via the stretching of the C=O bonding in the protein backbone within the amide I region (~1650 cm^{-1}). It gives distinct spectral changes for the different protein structures. However, these regions also contain contributions from the amide

stretches arising from the N-acetyl groups of the GAGs. This technique can be done in an optically active manner as vibrational circular dichroism (VCD). VCD can be used to study protein conformation and has a major benefit of having no significant contribution from the GAG components [79]. However, VCD requires very large amounts of material and high intensity light to overcome its inherently weak signals.

The GAG component of these studies can also be observed using several techniques. The most prominent and structurally useful of which is NMR. The signals originating from heparin have been fully assigned in NMR [30]. NMR has long been used to provide conformational information on polysaccharides and heparin, giving information on glycosidic linkages, sugar ring puckering and flexibility of the molecule [67,70,80]. NMR has been used within publications associated with this work only to characterise compounds and will not be discussed extensively. All NMR work was conducted in conjunction with Dr T. Rudd (Diamond Light Source), as part of ongoing collaborations.

Another widely used technique for the study of GAGs such as heparin and HS is mass spectrometry. This technique is based around the calculation of mass to charge ratios from a sample often using the length of time it takes to move from the sample stage to a detector (time of flight), and from this the deduction of possible structures that could give rise to that size fragment. This technique has been used extensively to try and deduce sequence information from the sulfation pattern within heparin and HS samples and trying to link these to biological activities. It has so far yielded little insight into any sequence:activity relationships. Ion-mobility mass spectrometry has

been used previously to gain structural information on GAGs when binding cations [81].

A technique that is currently being subjected to scrutiny for its potential to provide structural information on GAGs is Raman Optical Activity (ROA). This technique will be discussed in more detail in chapter 7 but, is of interest because it is thought to be directly sensitive to the overall conformation of complex GAGs such as heparin and HS in solution, under physiological conditions. This method is of particular interest because it is the only technique that directly reports the helicity of GAG molecules in solution, thereby giving some information on long range conformations [82].

In order to identify interactions of interest, the high throughput assay, differential scanning fluorimetry (DSF), has been used in this thesis extensively. This is a technique developed to study protein stability changes upon ligand binding [83], and has since been developed further to follow GAG binding to proteins [84,85]. DSF works through the binding of a hydrophobic dye to exposed aromatic residues within the protein as the protein is subject to a thermal degradation. The dye, SyproOrange, is self-quenching in an aqueous environment and, when stabilised by an interaction with the hydrophobic regions of the protein, usually buried in the core of a protein, and thus exposed when that protein unfolds, emits fluoresence at 610 nm, after being excited at 492 nm [83]. This technique can be carried out in a RT-PCR machine, and the fluorescence data are then analysed to discover the temperature at which the protein denatures.

The technique most extensively investigated within this thesis CD spectroscopy. The origins of the technique are based around an observation in 1895 concerning the optical rotation dispersion of an absorption band in samples, and were termed the "Cotton effect". Eventually this led to the technique today known as circular dichroism, which exploits Cotton effects in samples to gain molecular information. The technique usually covers two main wavelength regions for structural protein work, far-UVCD (180 – 260 nm) and near-UVCD (250 -350 nm) for electronic transitions but, also strays into the IR for VCD, and also available is fluorescence – detection CD. Far UVCD is a spectroscopic technique that probes chiral chromophores in the backbone of proteins, providing information on the structure of the protein being observed through variations in their absorptive properties. The CD arises from two circularly polarised components initially of equal magnitude, one rotating counter-clockwise (left handed, L) and one clockwise (right handed, R) which can be differentially absorbed by a chiral molecule, or a molecule that is placed in an asymmetric environment through ligand binding or other structural arrangements. A CD spectrum is reported as the difference in absorption observed between these two components, equation 1-1.

$$1.1 \quad \Delta A = A_L - A_R$$

These data are reported within this text in terms of millidegrees, and a representation is shown in Figure 1-8 on the next page.

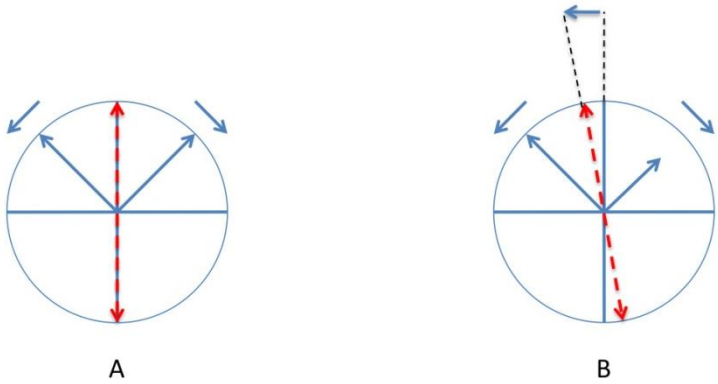

Figure 1- 8: A representation of the circular dichroism effect at the point of interaction with a sample. In A, both left and right-handed circular polarised light (blue arrows) are absorbed equally resulting in zero bias towards either vector. In B the absorption of right circularly polarised light is higher than left circular polarised light and therefore the sum of these absorptions results in the vector being rotated to the left by X millidegrees, giving rise to the millidegrees units commonly used for CD spectra.

The method of collection for the equipment at the purpose built CD beamline (B23, Diamond Light Source, Harwell, UK, synchrotron radiation circular dichroism, SRCD) is one of modulation between the L and R polarised light, rather than the alternatives of subtraction. CD as a technique is complementary to high-resolution NMR. CD can be carried out rapidly, a good quality spectra can be obtained <30 minutes, and only requires small amounts of material for far-UV work (~0.3 µM concentrations, for a 15 kDa protein in a 0.02 cm path length cell).

The far UV region (180 – 260 nm) is sensitive to average secondary structures within proteins. It is able to distinguish between α-helix, β-sheet (parallel and anti-parallel), β - turns and unordered structures. There are two principal contributions to the spectrum originating from the peptide bond. There is a weak, broad n → π* transition centered around 220 nm and a more intense π → π* transition at 190 nm [73]. The transitions for the α-helical spectrum were assigned in the 1960's. They were taken to be an n_1 → π* transition, between 210 – 230 nm, and observed experimentally at 222 nm (negative sign). There are transitions predicted at 198 nm but observed at 206 nm (negative sign) for a π^o → π* transition of parallel polarisation and also an associated perpendicular polarised transition at 191 nm, observed at 190 nm (positive sign). There was also a predicted transition, not observed at that time at 185 nm again from a π^o → π* of a negative sign [86,87]. The transition at 222 nm originates from the promotion of an electron from the oxygen non-bonding atomic orbital to an antibonding molecular orbital involving oxygen, carbon and nitrogen [87]. The transitions at 206 and 191 nm result from the splitting of a single absorption band [87].

There are several definable far UV contributions from the different secondary structures summarised in Table 1-5. The contributions from α-helical proteins consist of negative contributions at 222 and 208 nm, and a sharp positive band at 193 nm. β-sheet has a broad negative contribution at 218 nm and a positive feature at 195 nm, while disordered proteins assignments are given as 217 (positive) and 200 (negative) nm. The assignments of feature positions for proteins structures are taken from Greenfield [88]. The assignment for unordered structure however can be

difficult currently. As previously shown several proteins that are classified as unordered or have been denatured through different mechanisms actually have different CD spectra [89,90].

Table 1- 5: A table summarising some feature positions for some basic protein structures in the far UV (180 - 260 nm) CD as stated by Greenfield [88].

Structure	Feature Position	Sign	Notes
α-helix	193 nm 208 nm 222 nm	Positive Negative Negative	Strong, Sharp.
β-sheet	195 nm 218 nm	Positive Negative	 Broad
Unordered	200 nm 217 nm	Negative Positive	Strong Weak, Broad

There can also be a contribution from amino acid side chains in certain circumstances [91,92]. There can be a positive band at ~224 nm for aromatic side chains, while sulfur containing side chains have a weak negative band at ~228 nm [93]. A study using small helical peptides found that tryptophan contributed a negative induced CD band at 222 nm (the point of a significant negative feature in α-helical proteins) and a positive one for tyrosine [92]. These contributions can affect secondary structural analysis, of which there are several widely used algorithms.

The sensitivity of CD data to protein structure has been known since the technique was first developed in the 1960s [87,94] and the theory of the origins of signals has led ultimately to algorithms that can predict protein secondary structures [95,96]. These include SELCON, CDSSTR, and CONTIN amongst others [97]. In this work the dedicated software CDNN (Version 2.0) has been used due to the ease of use, availability, suitable reference set of proteins, which include some proteins that have been used throughout this PhD project, and through manufacturers information [98].

While these algorithms are sensitive to small changes within a protein structure, especially regarding the α-helical content to which CD is particularly sensitive: these algorithms are mostly based upon reference libraries that are taken from crystal structures where NMR structures are unavailable. This means that any absolute values for structures will inherit any mistakes in the crystal structure through crystal packing or inherent from the conditions to which the proteins are subjected to during crystallisation. This is especially apparent for disordered structures, which, through their inherent flexibility, are generally difficult to crystallise. These algorithms tend to require detailed information on protein conditions such as path length and the exact concentration for the correct prediction of structure. Recently, however, the need for precise knowledge of concentration has been reduced, following development of an alternative scaling method for the prediction of α-helical content [99]. The far UVCD spectra of protein-heparin complexes do contain contributions from heparin. However, these are negligible in comparison to the signal from the peptide bond so can be discounted from the analysis. The first observations of this kind were in the 1960s when oligosaccharides were investigated, giving a CD band

at 210-212 nm, assigned to the $\pi \rightarrow \pi^*$ transition of the 2-acetamido group [95,100]. The technique has been utilized to investigate the structure of heparin in recent times along with complementary techniques [68]. While the contributions from heparin overlap with peptide bond signals and the technique, as a whole, is a low-resolution structural technique, nevertheless, it does allow the structural study of proteins in solution in the presence of heparin with relative ease. The technique is also carried out under physiological mimicking conditions and allows for dynamic processes to be observed in solution and so potentially contains considerable structural information, especially if used in conjunction with other complementary techniques.

The near UV region (250 – 350 nm) is sensitive to the aromatic amino acids. These amino acids have characteristic regions where they exhibit a maximal signal. Phenylalanine exhibits weak bands between 255 and 275 nm, tyrosine between 275 and 282 nm with a shoulder appearing at longer wavelengths. The strongest contributor to this region is tryptophan, which has a maximal signal at ~290 nm. As these signals derive from individual amino acids, the signal intensity for each band is entirely reliant on the amino acid sequence for each protein. The shape of these features and their amplitude also depends on their local conditions such as H-bonding, position in the protein (buried or exposed to solvent) and any molecules that may affect their electronic properties (generally within 1 nm).

Synchrotron radiation CD (SRCD) was developed in the 1980s [101,102] and has many advantages over benchtop CD instruments. The basis of these advantages lies in the increased photon flux that is available in SRCD [90], over 1000 fold increase

in intensity below 200 nm when compared to xenon arc light sources [73,103]. This increased flux provides several benefits. The SRCD light sources provide brighter light, especially at lower wavelengths (< 200 nm) where bench top CD instruments with Xeon lights begin to lose their intensity. This brighter light results in higher signal to noise ratios, providing several advantages; less material may be used to gain the same quality spectrum as normal CD, the same quality spectrum may be achieved in less time (approximately 10% of the time [90]) and also using the same quantities of protein as a standard bench top CD machine, higher quality spectra may be obtained, allowing for the observation of subtle differences to be measured accurately. A SRCD source can reach wavelengths as low as 140 nm [90] for dehydrated samples and 168 nm for aqueous solutions, which enables a better differentiation between types of β-sheet and other structures, especially unordered. Module B at B-23 (Diamond Light Source Ltd.), the equipment used for the CD work within this thesis, allows measurements between 165 and 650 nm. It also has several other advantages, both over other CD beamlines and benchtop CD instruments. It has a temperature controlled sample holder (5 – 95 °C), automated scripts and a variable beam size as small as 0.25 mm^2 [104]. The small beam size has the added benefits of a high photo flux density (75 $X10^{11}$ photons s^{-1} at 200 nm) and also allows the measurement of long path length cells (up to 10 cm) while maintaining a low volume [104].

These increased wavelength ranges (down to 168 nm) have great advantages when obtaining structural information on proteins. It has been shown to improve structural analysis of proteins in comparison to their crystal structures [105] and is possible in

aqueous solutions [106]. It has been shown that below 190 nm, SRCD has a large advantage in terms of signal to noise in comparison with benchtop CD instruments [107]. In this body of work the wavelength range generally employed will be 185 nm – 260 nm.

1.5. Aims and scope of work covered by this thesis.

As stated in this introduction the study of proteins:GAG interactions is a challenge that has not yet been met, and the aim of this body of work will be twofold, firstly CD, a technique capable of studying protein:ligand interactions in solution through the detection of protein average secondary structural changes, will be improved through either data analysis methods or the improvement of methodology in the collection of data. Secondly, new methods in addition to standard near and far UVCD shall be developed for use in the study of protein:GAG interactions. Whilst the aim is to achieve these two goals primarily, any additional information that can be gained investigating relevant biological systems of interest will also be of interest to the field and as such systems used have been selected based on their cost effectiveness and their interest to the field.

The work in following chapters will comprise of these themes:

Chapter 3: This chapter, using the amyloid forming protein Hen egg white lysozyme (HEWL) looks into the effect of Zinc Chloride ions on the interaction HS with the

protein, utilizing little used denaturation techniques to differentiate apparently similar complexes.

Chapter 4: Employs the interactions observed in chapter 3 and also an example of super oxide dismutase 1 (SOD1) with a non-GAG ligand to demonstrate the potential of using magnetic circular dichroism (MCD) in the study of protein:GAG interactions.

Chapter 5: AT, the important protein within the blood clotting cascade, will be the focus of this chapter. A range of heparin like, and heparin based, polysaccharides will be tested for activity against anti factor Xa with the classic pentasaccharide as a positive control. Again underused denaturation techniques are employed to differentiate complexes.

Chapter 6: This takes SRCD data from chapters 3 and 5 and applies generalised 2D correlation techniques to add resolution and information to the data already attained.

Chapter 7: This chapter presents two improvements to data handling, specifically for the handling of ROA data regarding baseline correction and the normalisation of CD data before further processing.

Chapter 2: Materials and methods.

2.1. Materials.

Hen egg white lysozyme was obtained from Fluka, (Sigma-Aldrich, Gillingham, Dorset, UK). Heparin used is from porcine intestinal mucosa, stored as a powder (Celsus Laboritories, Inc. Lot- PH-42800 activity (anti-Xa) 201 IU/mg). Heparan sulfate, sodium salt from bovine kidney, was obtained from Sigma-Aldrich (Gillingham, Dorset, UK). Super-oxide dismutase 1 was expressed in *E.Coli* BL21 and kindly donated by Dr G Wright (University of Liverpool, UK). 5-flourouridine was also kindly donated by Dr G Wright (University of Liverpool, UK). Fibroblast Growth Factor proteins were expressed in *E.Coli* BL21 cells, donated kindly by Mr Y Lee (University of Liverpool, UK). Resin used for antithrombin purification is the matrix from a HiTrap heparin affinity column.

2.2. Methods.

2.2.1. Synchrotron Radiation Circular Dichroism (SRCD): Basic set up and far UV SRCD.

What follows is an overview of standard conditions used during a SRCD experiment on B-23, the details of which will be presented in each individual experiment since the conditions vary depending on the sample used and experiment performed.

A standard SRCD experiment to probe protein secondary structure would be conducted between 185 and 260 nm when in the presence of NaCl.

When undertaking all spectroscopic experiments, thought must be given to the Beer-Lambert rule in equation 2-1:

$$2.1: \quad A = L\,\varepsilon\,c$$

Where A= absorbance, L=path length (cm^{-1}), ε=molar absorption of molecule at wavelength ($L \cdot mol^{-1} \cdot cm^{-1}$) and c = concentration ($mol \cdot L^{-1}$).

To attain the ideal recording conditions, the detector voltage must be kept below 600 mA whilst maintaining a high signal to noise ratio. To achieve these conditions in an experiment, three properties can be altered (within the setup available at B23), path length, concentration, and integration time.

Integration time refers to the amount of time (in seconds) spent recording one data point at a single wavelength. The path length is distance in the sample through which light passes before the detector, usually measured in cm, and the concentration refers to the Molarity of the sample in the light path. The CD signal amplitude can be calculated via the number of amino acids present in peptide, since the main chromophores being probed at the wavelengths (180-260 nm, far UVCD) is the peptide bond (see Chapter 1.4).

Standard cells available for experiments range from 0.01 mm to 0.5 mm CaF_2 and 0.2 mm to 1 cm quartz cells. The path length of cells are nominally marked but accuracy decreases at the shorter length. The actual path length can be measured by utililising the interference waves in a Fourier transform IR instrument. A standard concentration in a 0.02 mm CaF_2 cell with a 15 kDa protein would be around 0.5 mg/ml. Two common buffer systems used for CD work are 12.5 mM (NaCl) PBS and 20 mM phosphate buffer. In order to obtain structural data below 190 nm a buffer low in absorbing chromophores must be used. This especially applies to concentrations of Tris and chloride which in particular are not tolerated, as they absorb UV light in the region below 200 nm, leading to a poor signal to noise ratio. A standard secondary structural experiment consists of an average of 4 scans, in order to reduce noise in the spectra. The benefit of N scans is the reduction in noise, which is related to the square root the number of scans (See Equation 2.2). There are limitations to the benefit of additional scans, however, as UV light is known to degrade proteins, and takes longer to acquire, so a compromise must be found.

2.2: $n \propto \sqrt{s}$

n = signal to noise ratio is proportional to the number of scans, s = scan number

The standard equipment setup on module B at B-23 would be 0.5 mm slits and 1 second integration time, at 20°C.

2.2.2. Near UV SRCD.

Conducting SRCD experiments in the near UV region (250 – 310 nm), that probes aromatic compounds within a sample requires different experimental conditions to optimise data acquisition. There are several problems and benefits to this region of the spectrum in comparison to a standard far UV experiment (section 2.2.1), while the main drawback of the signal from aromatic side chains is their low abundance in comparison to the chromophore for far UVCD. To counter this, a longer integration time, increased path length and higher concentrations can be used. This region of the spectrum does not contain information on secondary structures of the protein. However, one advantage of this type of spectroscopy is its tolerance for salt, which does not absorb in this region, and the energy levels of the light involved at these wavelengths are lower (Equation 2.3) - meaning that UV damage to protein samples is not usually a consideration. Another benefit, relevant to the majority of work undertaken here, is that GAGs do not have a contributing signal at these wavelengths. This being noted, it is usual to use the same buffers, so that complementary far UV experiments maintain consistent experimental conditions.

$$2.3: E = h\nu$$

E =energy (J), h =Planck's constant ($6.62606957 \times 10^{-34}$ J), ν = frequency (s^{-1})

A standard near UV experiment would have a 0.5 mm slit, 3 second integration time, 1 cm quartz cell and 0.5 mg/ml concentration of protein.

2.2.3. Degradation of proteins via far UV SRCD.

UV degradation is conducted in the far UV region, in the same conditions outlined within section 2.2.1, however, many more scans are required in order to degrade the protein, and each scan is taken individually through the series. A typical experiment would require between 20 and 30 scans to achieve the desired degradation, which typically equates to around 2 hours exposure to the high intensity high photon flux radiation available at B23 (2×10^{13} photons s^{-1} mm^{-2} [108]). The number of scans required can depend on how sensitive each individual protein is to UV degradation and exposure can be adjusted by reducing slit width and integration time.

2.2.4. Degradation of proteins through a thermal gradient within SRCD.

A temperature series is stipulated between the regions of interest, and can be controlled in 0.1°C steps. When conducting these experiments, controls or consideration must be given to the effect that continuous exposure to UV light can have on the sample, as described in section 2.2.3. Typically a temperature series involving 5°C steps are employed. Each temperature is held for two minutes before data collection commences and the next step is begun. The temperature is controlled via a Peltier temperature controller (Quantum Northwest, Liberty Lake, USA). This can be controlled either manually or via the use of a script that can be incorporated into the OLIS™ data collection software. This system pumps water into the cell holder and the temperature is changed to the appropriate value. This then heats the

cell and sample within. Allowing adequate time for this process is essential for accurate results.

2.2.5. Magnetic circular dichroism.

The experimental setup at B23 can be modified in order to employ MCD. There is one major change to the equipment, the addition of a 1.4 Tesla magnet (permanent magnet for MCD, OLIS, Bogart, USA) sample holder in place of the Peltier temperature controller. The dimensions of this sample holder restricts the user to using only 1 beam during the experiments and cells with 1 cm^2 dimensions. Experimental design in the near UV (270-320 nm) must only account for the concentration of tryptophans within the samples as this is the major contributing factor to spectra in this region, for full details see Chapter 4.

An experiment is conducted first using a non-magnetic control sample holder to record the full aromatic spectra, and then again in the presence of the magnet. The magnet is first orientated either N → S (in relation to the light from source to detector) or S → N and then reversed for the second recording. As this is done, it is important that all other possible variables are maintained, including the orientation of the cell in relation to the propagation of light.

The MCD spectrum is then obtained by the removal of standard CD contributions via the deduction of the N→S spectrum from the S→N. This results in a 2 x MCD spectrum, which is then halved to give the MCD spectrum for the sample. A

background of buffer alone must be recorded as the buffer is not deducted during the above process, rather contains a induced drift, see Chapter 5, Figures 4-8 and 4-9 for an example of how this effects data.

2.2.6. Differential Scanning Fluorimetry.

DSF is a technique that can be used to survey large numbers of compounds for interactions with a target protein. It works on the basis of the dye SyproOrange binding to hydrophobic regions of a protein that become exposed as the protein unfolds from thermal stress. Upon binding to hydrophobic regions of the protein the fluorescence 620 nm emission of SyproOrange, which is excited at 492 nm, increases [83]. This dye is only capable of detecting proteins with a hydrophobic regions, that is, globular proteins. This normally restricts use to those proteins with a molecular weight of 11 kDa and above.

For each protein investigated it is standard practice to test a range of concentrations in order to attain a high quality, repeatable, melting curve for analysis with compounds present. A typical value for this type of experiment would be 10 µM for a 15 kDa protein. Preparation of the dye itself involves diluting the neat dye by adding 0.02 µl per 1 µl of fresh H_2O (50 x working concentration). It can then be mixed with the buffer of choice for the experiment; this solution is at 10 x working concentration. The experiment is conducted in a 7500 fast PCR instrument ramping the temperature from 30 to 90°C in 0.5°C steps. At each point the temperature is held for 5 seconds before flouresence is recorded for 30 seconds. The fluorescence data

are then analysed by obtaining the first derivative of the curve for each sample well, giving the point at which 50% of the protein is denatured. These experiments are typically conducted as triplicate repeats of a single sample, which are then repeated again in triplicate, giving 9 readings from 3 independent samples. This ensures that standard deviations can be found and then used to find and significant differences using Student's t-test (Equation 2-4).

Equation 2-4: *Independent two sample students t test, equal sample size equal variance:*

$$t = \frac{\bar{x}_1 - \bar{x}_2}{{}^{s}x_1x_2 \cdot \sqrt{\frac{2}{n}}}$$

where ${}^{s}x_1x_2$ *is:*

$${}^{s}x_1x_2 = \sqrt{\frac{1}{2}(s^2x_1 + s^2x_2)}$$

and the degrees of freedom for significance testing is defined as:

$$d.f = 2n - 2$$

Where x_1 and x_2 are the sample means of groups 1 and 2, s is the standard deviation of that mean and n the number of measurements in that group.

2.2.7. Changing the cation state of heparan sulfate and heparin molecules.

In order to change the cation state of heparin and HS molecules, ion exchange resin (Alfa Aesar, Dowex 50W x8 200-400 (H), Lancs, UK) must be used. To do this the resin must be first put into either their H^+ or cation form.

Cation forms of HS were prepared from HS (10 mg/ml, in deionised water), shaken with cation-cation exchange resin, spun and the supernatant removed, quantified (Absorbance at 232 nm) and employed for experiments.

2.2.7.1. Creation of H^+ resin beads.

1. Wash resin with H_2O for 30 minutes
2. Wash through a filter 3 times with a 1 M HCl solution
3. Wash 3 times with H_2O
4. Check H_2O elute below the filter is neutral with pH paper to ensure excess acid has been removed.

2.2.7.2. Creation of resin in cation forms.

1. Wash H^+ form resin with H_2O
2. Incubate with 2 M chloride salt (eg zinc chloride ions), whilst stirring with magnetic stirrer.
3. Wash through a filter

4. Wash with H_2O to remove excess salt
5. Repeat steps 2-4 to ensure complete coverage.

2.2.7.3. Modification of heparin/HS cation state.

1. Make a heparin solution of 10 mg/ml in H_2O
2. Add a spatula of the pre-made ion form resin beads (section 2.2.7.2) per each 500 µl of heparin solution.
3. Spin solution and remove the supernatant
4. Recover the beads for re-use in section 2.2.7.1.

2.2.7.4. Isolation of shrimp heparin compounds.

Shrimp heparin (SH) was isolated via a homogenised shrimp head:acetone solution (10:4 v/v) for lipid removal, ground and dissolved in 1 M NaCl adjusted to pH 8.0 in preparation for a 24 h proteolytic digestion of the solution at 60°C, the result of which was filtered and the negatively charged components of the solution removed via ion-exchange resin. The resulting elute was precipitated with cold methanol at 4 °C over 48 hours, filtered, and the precipitate collected. This was then subsequently fractionated using DEAE-sephacel. Full details of SH purification are provided in the Brito *et al* 2008 publication [109].

2.2.8. Purification of AT from human plasma.

All following steps are completed at 4°C. All purification buffers contain 0.1 M Tris-HCl buffer at pH 7

1. Run 2 x column volumes of 0.02 M NaCl buffer on heparin Sepharose column.
2. Load human plasma onto heparin Sepharose column (2 ml plasma for 5 ml column volume)
3. Wash with 0.02 M NaCl buffer 3X column volume
4. Wash with 0.5 M NaCl buffer 3X column volume
5. Repeat steps 2-4
6. Elute AT with 2.0 M NaCl solution.
7. Concentrated with SpinX UF spin concentrator columns 30 kDa cut off at 4°C
8. Take 280 nm absorbance to check protein concentration.
9. Confirm protein in SDS page gel.

This method was developed by and completed at the University of São Paulo by Dr M. Lima as part of a collaborative project.

2.3. Data handling and analysis methods.

2.3.1. Circular dichroism data handling.

The raw data, were recovered from the Olis software as several spreadsheets. The data were loaded into the R statistical software package as a matrix. The scans were then averaged together to form the mean spectra across the data. The same was then done to the baseline spectra consisting of all components minus sample, i.e. cell, air, and buffer which is then deducted from the sample spectra. The resulting spectrum was then normalized to a value at the beginning of the spectrum where there are no chromophores and the data were set to zero. The normalisation of data has been to the average value between 250-255 nm as this was found to be more accurate than normalisation to 260 nm alone, this is described in Chapter 7. For near UV work spectra were normalized to 330 nm.

All data handling was performed within R statistical programming software [110].

2.3.2. Estimation of protein secondary structure.

Pre-processed far UVCD data were entered into the neural networks based software CDNN (V 2.1) [98]. Fits were considered reliable when the result was between < 90% and > 110%. These structural estimations were checked against PDB structures where possible.

Chapter 3: A zinc complex of heparan sulfate destabilises lysozyme and alters its conformation.

3.1. Introduction.

3.1.1. Hen egg white lysozyme.

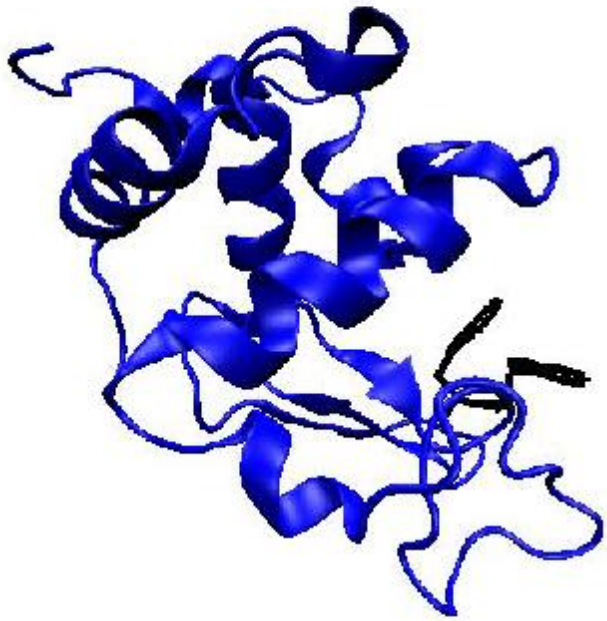

Figure 3- 1: The crystal structure of HEWL (pdb number 4AXT) showing two exposed Trp residues [black] in the carbohydrate binding cleft.

HEWL is a small, well studied globular enzyme and has been the subject of over 6000 articles since its discovery in 1922 [111]. It is highly conserved between many species and is expressed in the secreted fluids as an antibacterial innate defense in, for example, tears, milk and saliva. It is a muramidase that cleaves the glycosidic

linkage of carbohydrates, specifically the 1,4 beta linkage between N-acetylmuramic acid and N-acetyl-D-glucosamine in bacterial cell surface polysaccharides. HEWL is a 14.2 kDa protein consisting of 129 amino acids, with 4 disulfide bonds. The protein has two main domains, divided roughly along the axis of the active site cleft. The alpha domain (amino acids: 1-36 and 85-129) is helical, containing 4 helices, while the beta domain, which includes the central region of the primary amino acid sequence (amino acids: 37-84) consists of a three-stranded β-sheet, a loop and a single helix [111]. The overall structure of HEWL is shown in Figure 3-1. HEWL is a protein that is susceptible to aggregation, an intrinsic property of many proteins [112]. HEWL became a major target for the study of protein aggregation since point mutations in it were discovered to be involved in hereditary systemic amyloidosis in 1993 [111,113]. In general, the aggregation of proteins can lead to physiological problems, which include Parkinson's disease (PD), Alzheimer's disease (AD) and congestive heart failure. Aggregation has been proposed in both a beneficial role [114] and as a harmful process resulting in cytotoxicity arising from soluble oligomers, as opposed to fibrils [115,116]. It has been suggested that the folding of proteins into their active forms can take several routes to their resulting active conformations [117]. The formation of amyloid plaques has also been identified as a multiple phase process [118], while HEWL has been proposed as a paradigm for protein folding and misfolding [119], having its core structures of amino acid regions 26-123 and 32 to 108 heavily implicated in amyloid formation [120]. The creation of amyloid fibrils involves the interplay of several factors. A key factor in amyloid formation is the native state stability of the protein [112], promoted by the presence

of "seeds" [121] and this seeding is not confined to proteins of the same sequence [121,122].

3.1.2. GAGs are associated with amyloid fibrils and are influenced by cation state.

GAGs, in particular HS, are associated with the aggregation of proteins and are a consistent factor in amyloid plaque formation, regardless of disease and protein component [123,124,125]. It has been reported that HS is involved in the early formation of amyloid plaques, including those of AD and PD [126,127], and has also been proposed as an initiator of these plaques [125]. The scrutiny of GAGs due to their association with plaque formation [127,128] has led to the idea that the interactions between these sugars, widely expressed on mammalian cells and a ligand for many proteins [129,130], could be important for the formation, development and stability of the amyloid fibrils formed during these disease processes. GAGs themselves have been shown to alter conformation and activity when substitution pattern and cation state alters [45]. The binding of these cations can also be selective for particular HS structures [68]. In addition it is known that zinc ions selectively bind to heparin in a manner that is above that of just simple electrostatic interactions [131]. The conformation of heparin is known to change with the addition of metal ions, as shown by ion mobility mass spectrometry [45,81] and NMR [67,70,132]. The conformational change induced in heparin has led to work that showed examples of signaling and non-signaling determined solely on the cation state of the heparin in FGF systems [46]. One of the most in-depth studies of cation-heparin binding

involves copper ions, which are shown to specifically coordinate with heparin in a high affinity manner initially before a less specific binding takes place [68].

Zinc (Zn) has been shown to have a variety of effects on amyloid diseases [133,134], and is the most abundant trace metal in the brain, varying between 0.15 and 0.20 mM in brain tissue. It has been reported that Zn ions bind to Aβ peptides [135,136] with three histidine residues being heavily implicated in coordinating the binding of zinc to Aβ [136]. Zinc is also reported to inhibit Aβ toxicity at μM concentrations [137]. A complex of modified heparin:Zn ligands with HEWL showed some initial promise in the DSF [83,84] assay (data not shown), which measures the thermal stability of a protein indirectly by reporting the fluorescence emitted by the dye SyproOrange. SyproOrange is self-quenching in an aqueous environment, only emitting a signal when stabilised by aromatic amino acids, which are found commonly in the hydrophobic core of proteins. This metal ion was therefore studied using the natural ligand HS, as an investigation into the effect cation binding has upon the important HEWL:HS interactions, an important model amyloid forming protein. The following data will establish initially that there is an interaction between GAG compounds and HEWL in Table 3-1 through the use of DSF. This is then confirmed by far UV SRCD in Figure 3-2A and near UV SRCD in Figure 3-2B. The observed changes are confirmed to be stemming from HEWL-HS(Zn) and not Zn cations in Figure 3-3. The stability of the protein and complexes are investigation through temperature and UV degradation in Figures 3-4 and 3-5 before being further analysed in Figures 3-6 and 3-7. The spectra in Figure 3-2A are analysed for secondary structure content in table 3-2.

3.2. Results and discussion.

3.2.1. Discussion of thermal stability investigations.

Table 3- 1: The thermal stability of HEWL alone and in the presence of a ligand. HEWL at a concentration of 10 µM (0.14 mg/ml) in the presence of an excess of GAG, based where necessary on average molecular weight (Mw). The protein was subjected to a step-wise temperature gradient, from 32 °C to 85 °C in 0.5 °C steps. There was an initial 2 min incubation period at 31 °C and 5 s between each temperature increase to allow equilibration. Data were collected for 30 s at each temperature. First derivatives of the melting curves were employed with the Solvitzky-Golay smoothing method, using a second order polynomial.

Sample	Melting Temperature
HEWL	67.7 (±0.7)°C
HEWL:$ZnCl^{2+}$	72.1 (±0.3)°C
HEWL:HS	65.3(±3.6)°C
HEWL:HS(Zn)	57.6(±1.6)°C

The initial observations using the DSF thermal stability assay (Table 3-1), demonstrated dramatic changes in the stability of HEWL. The presence of HS(Zn) with HEWL reduced the stability of the protein by 10.1°C (67.7 to 57.6 °C). This observation is not repeated by either HS alone (stability reduced by 2.4 °C) or by $ZnCl_2$ ions that stabalise the protein by 4.4 °C. Altered stability induced by GAGs such as heparin or HS is not new, since stabilisation has been demonstrated

extensively in some FGFs [139,140], as well as in AT [141] (See Chapter 5). It is, however, unusual to see a destabilisation of a protein by HS, but here it is demonstrated that significant denaturation of HEWL by HS saturated with zinc ions (HS(Zn)) occurs, and this is unique to the zinc complex since HS alone has minimal impact on the overall stability of the complex. The T_m for the HEWL:HS complex has a large standard deviation (SD) from the mean (2.25 times higher than HS(Zn), the next highest SD), and this can be attributed to comprising a very heterogeneous mixture of potential ligands and, while it has not been subjected to ion exchange, it is predominantly in the Na form (according to the manufacturer). The HS(Zn) conversely might be strongly inclined to fewer possible conformations as the polysaccharide chain approaches saturation with Zn ions, and therefore would have a smaller deviation from the mean as the compound would become more homogeneous relative to that of unmodified HS. This hypothesis is further evinced by the NMR studies conducted on these samples (supplementary data, [142]) and it will not be discussed in detail here. However, in summary, it was shown that Zn ions bind preferentially to IdoA, possibly interacting with the carboxylic acid. There is no sign of conformational change along the chain; NOEs stemming from the glycosidic linkage region which does not change. These remained consistent in the presence and absence of Zn (NMR experiment design and analysis conducted at the Ronzoni Institute, Milano, by Dr T Rudd).

3.2.2. Discussion of SRCD investigation of complexes.

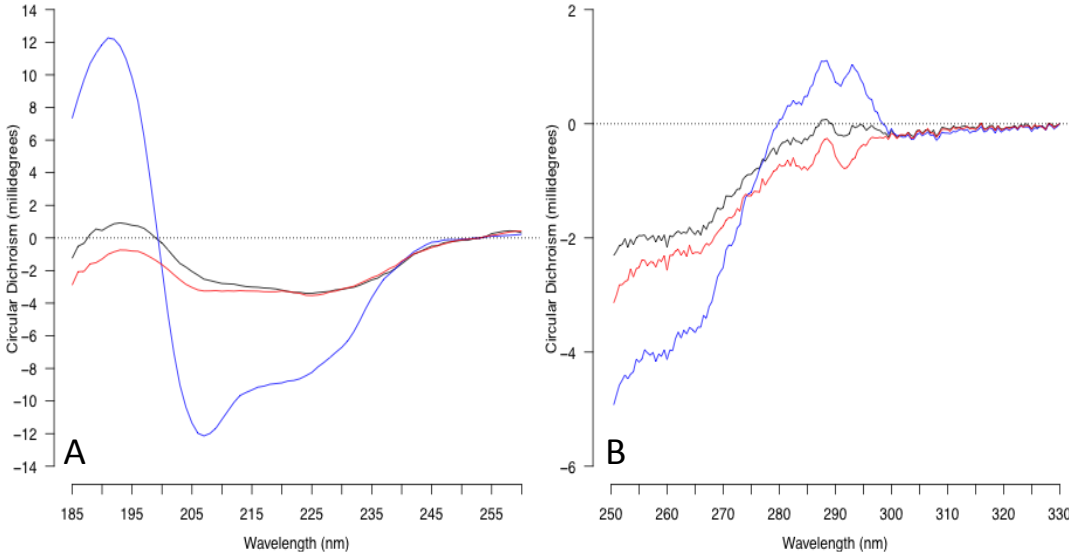

Figure 3- 2: Panel **A**: Far UV SRCD spectra (185–260 nm) of HEWL (0.50 mg/ml) [blue], HEWL with HS (0.08 mg/ml) [red] and HS(Zn) (0.08 mg/ml) [black]. Panel **B**: Near UVCD spectra (250–330 nm) of HEWL (0.25 mg/ml) [blue], HEWL with HS (0.04 mg/ml) [red] and HEWL with HS(Zn) (0.04 mg/ml) [black]. The Trp residues of the active site reportedly give rise to a peak at 291 nm and a trough at 293 nm [138].

These samples were then subjected to far UVCD (185 -260 nm) at Diamond Light Source, with the hypothesis that these complexes would yield different secondary structures. The results are shown in Figure 3-2A and the secondary structure evident from these spectra are shown in Table 3-2 as derived from the CDNN software [98]. As can be seen from these results, the structures induced by these ligands, with significantly different thermal stability profiles, are remarkably similar. There is a general loss of α-helical content with an increase in antipararellel β-sheet in both experiments involving HS ligands. HS and HS(Zn), interacting with HEWL results

in a loss of 18% of α-helix, and are replaced by increases of 22 and 27% of antiparallel ß-sheet respectively.

Table 3- 2: Secondary structure of HEWL, HEWL:HS and HEWL:HS(Zn) as determined using the CD spectra in Figure 3-2 using CDNN version 2.1 [98]. Values given by software are presented in black while the difference compared to HEWL alone are presented in red text.

	HEWL	HEWL:HS	HEWL:HS(Zn)
Helix	27%	9% (-18%)	9% (-18%)
Antiparallel	12%	34% (+22%)	39% (+27%)
Parallel	7%	4% (-3%)	4% (-3%)
β-Turn	23%	20% (-3%)	18% (-5%)
Random Coil	33%	34% (+1%)	31% (-2%)
Total Sum	100%	100%	102%

As these complexes were similar in secondary structure, the surface interactions were probed in Figure 3-2B by near-UVCD. This area of the CD spectrum relates to transitions observed in the aromatic residues of the protein, and in this case has a particular wavelength (peak at 291 nm and a trough at 293 nm [138], the spectra here slightly disagree due to different buffer conditions which can change the local environment of the chromophore and therefore the peak position) which previously has been associated with the two (of six) Trp residues within the active site (Figure 3-1) [143]. The data presented in Figure 3-2B indicate that, even though again similar, there are differences in the mode of binding of the HS(Zn) ligand in and around the binding site when compared with standard HS, which is in combination with other possible binding sites elsewhere on the surface of the protein. The interactions with Trp are investigated further in Chapter 4 by MCD.

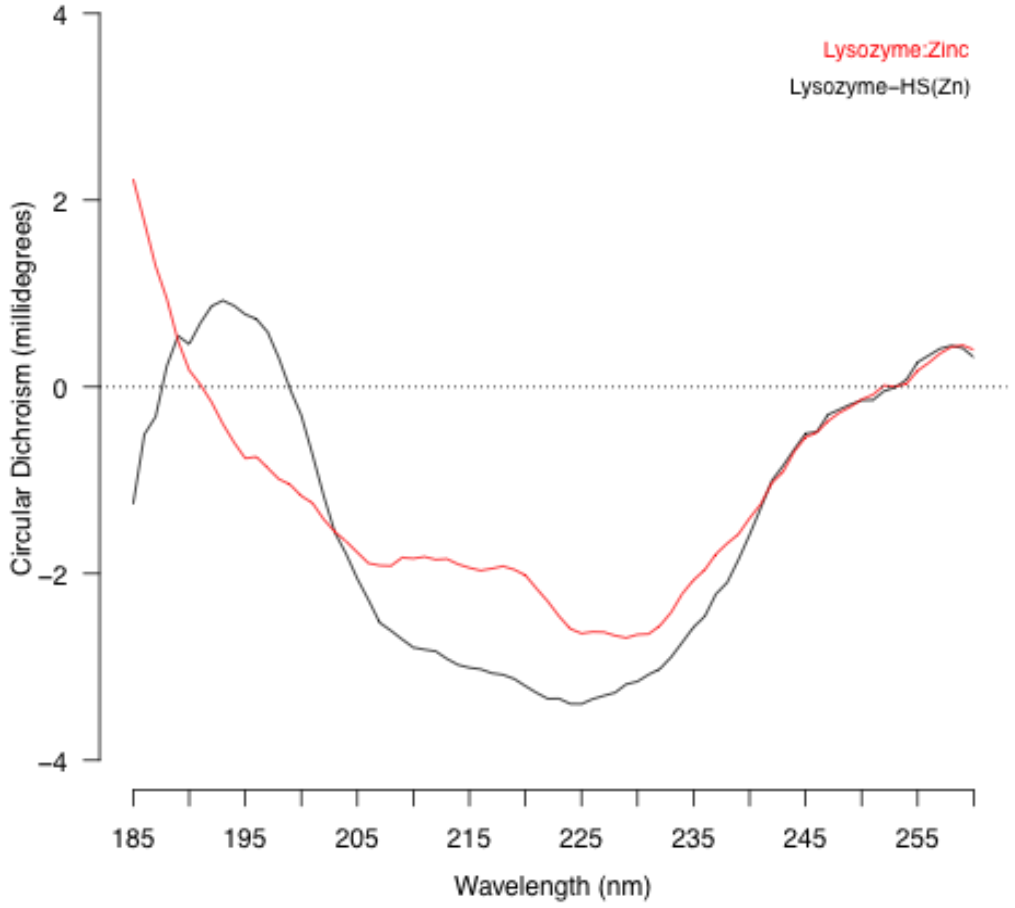

Figure 3- 3: Far UV SRCD spectra (185–260 nm) of HEWL (0.50 mg/ml) with either zinc cations ($ZnCl_2$) (at equimolar ratio of typical HS cation content, red) or with HS(Zn) (0.08 mg/ml, black).

Zinc itself is not the cause of the structural changes in HEWL (Figure 3-3), nor the stability difference as shown in Table 3-1. Figure 3-3 shows the effect of Zn ions on the conformation of HEWL and, as can be observed, is distinct from that of HEWL:HS(Zn), demonstrating that it is the combination of HS with Zn that is the cause of the observed effects.

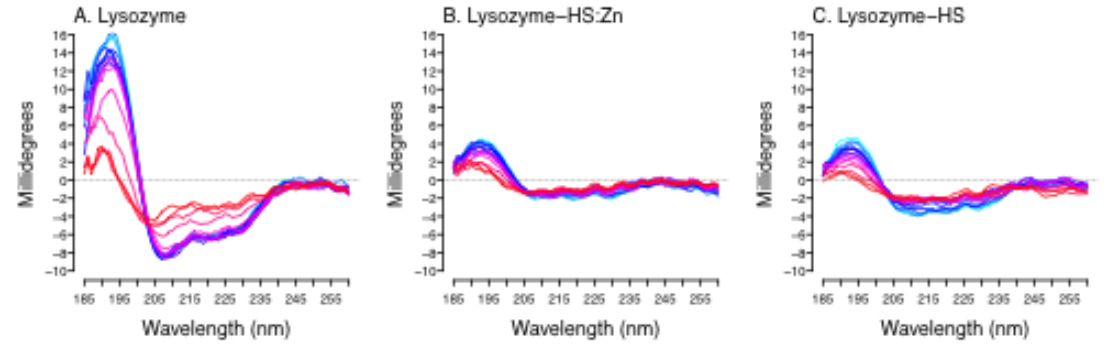

Figure 3- 4: Exposure of HEWL and HEWL:HS (HEWL 0.5 mg/ml, HS/HS(Zn) 0.08 mg/ml) complexes to a temperature gradient (A–C). The temperature ranged from 10 to 85°C in 5°C increments (progressively blue to red) resulting in distinct thermal denaturation profiles.

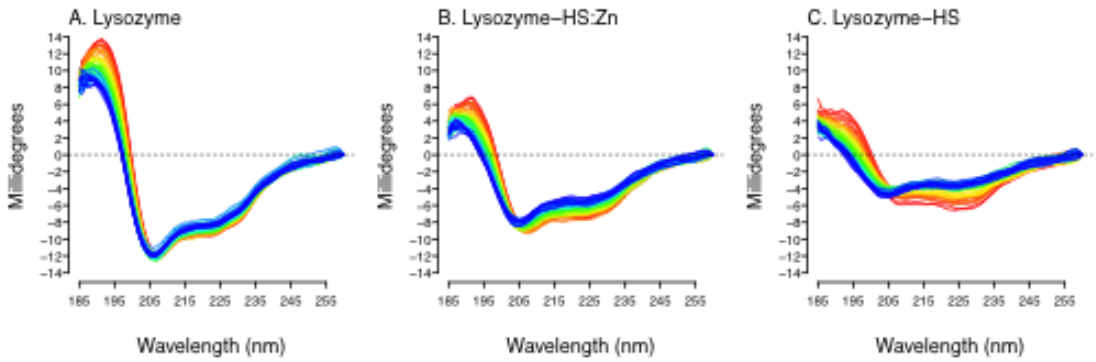

Figure 3- 5: Exposure of the HEWL and HEWL complexes to prolonged UV irradiation resulted in denaturation. The stability of complexes of HEWL alone (A) with HS(Zn) (B) and native HS (C) to degradation by UV radiation was followed over 60 scans between 180 and 260 nm (experimentally, only 185 – 260 nm are plotted), which can detect the changes in secondary structure produced by UV-induced denaturation. The red line indicates the starting spectrum, proceeding progressively to blue.

Further ways to distinguish similar complexes involve the degradation of the protein in the presence of ligands through either UV irradiation or thermal degradation (Figures 3-4 and 3-5). The raw data of these types of experiments can be difficult to interpret therefore principal component analysis (PCA) has been applied to them. PCA, (used in chapters 3 and 5 on lysozyme and antithrombin proteins complexed with GAGs) is a mathematical technique designed to simplify complex data sets. The analysis technique allows features which differ between spectra to be revealed, which in complex data set may be impossible by eye. This allows for patterns, groupings and kinetic vectors to be revealed, thus adding information to the basic spectra recorded. The result is a list of weighted principal components (PC) that explain a percentage of variance in the data. The weighting of the PC is termed the loading, and these can be plotted in scree plot and the most important PCs chosen for plotting. The original dataset, V, can be described by the summation of loaded contributions ($\underline{w}_n$) from the various PC's (P_n) and the remaining variance stemming from noise, e, thus:

$$V = w_1.P_1 + w_2.P_2 + w_3.P_3 + \ldots + w_n.P_n + e$$

By analysing data in this way, the resulting data stem from two key points, firstly that no assumption is made and secondly that each PC is independent of all other PCs and complex data sets can be reduced to just a couple of variables which explain a majority of the variance in the data observed.

By applying this technique to the data in Figure 3-5, differences in the pathways taken by each sample can be observed, especially in this instance between the thermal degradation of HEWL in Figure 3-4A, compared to the two examples in a complex with the HS ligands (Figure 3-4B and 3-4C). This demonstrates that for HEWL there is a two step process centered at 65°C, while the samples containing HS and HS(Zn) show continuous degradations. The data presented are further investigated by the variable reduction technique, PCA in Figure 3-6.

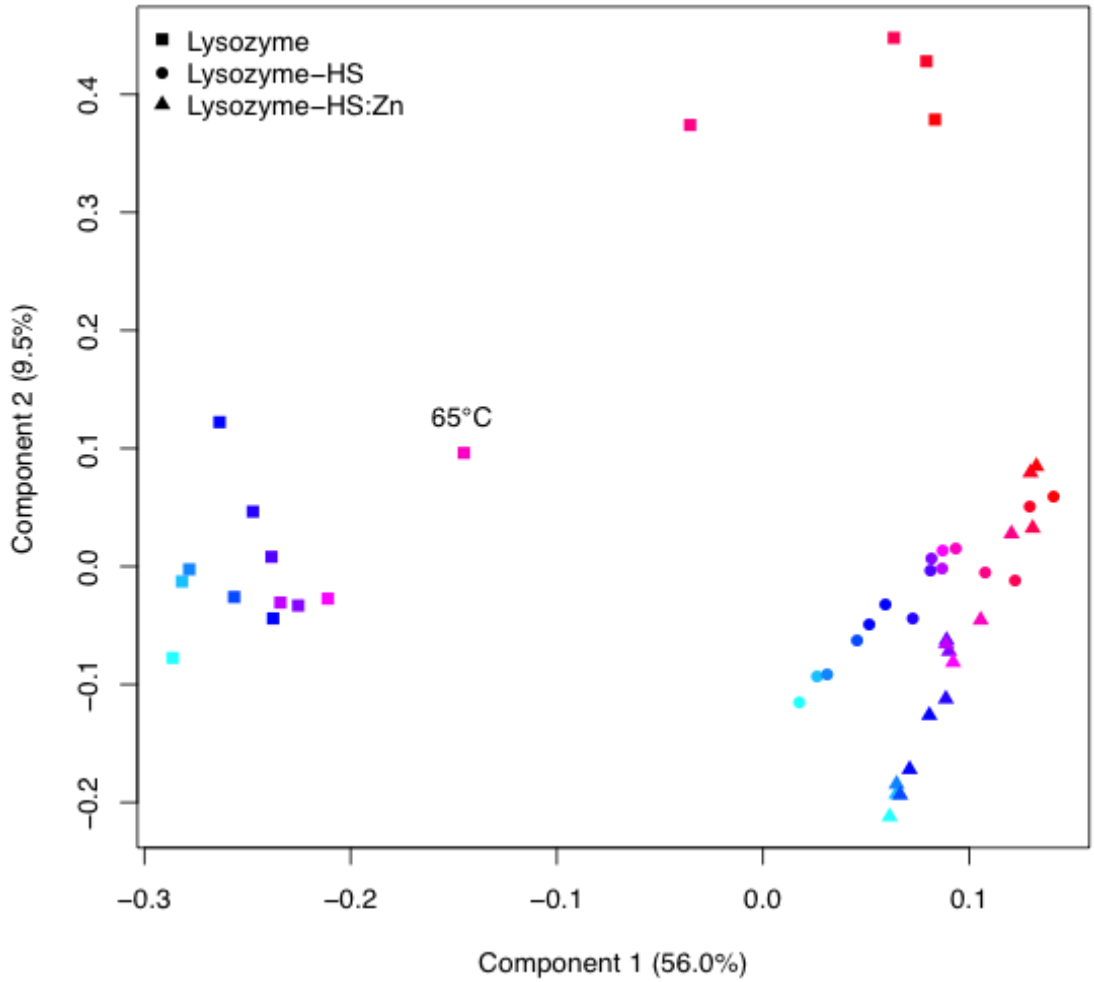

Figure 3- 6: Principal component analysis (PCA) of the SRCD spectra of HEWL and HEWL:HS complexes undergoing thermal denaturation (Figure 3-4). The colours progress from blue (10°C) to red (85°C)

The PCA analysis of the temperature degradation, shown in Figure 3-6 further shows the difference in stability and through the two stage unfolding of HEWL alone (squares), confirming the DSF investigations, with a major transition of conformational states occurring at 65 °C. The two examples with ligands are again similar when utilizing these two principal components but, interestingly, they both

lose the two-step degradation, rather, exhibit a more directed kinetic pathway to similar endpoints. This pathway is subtly different between the two ligands indicating a different interaction in each case.

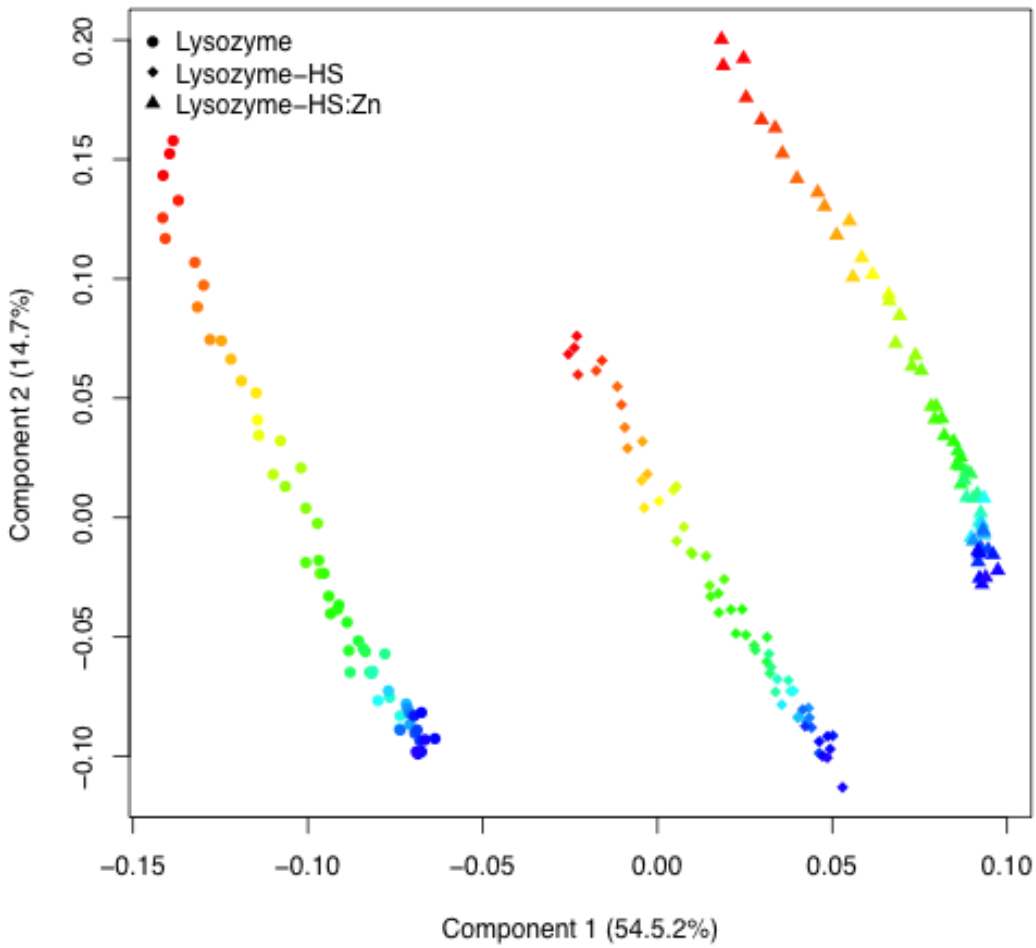

Figure 3- 7: PCA of the far UV SRCD spectra of HEWL and HEWL:HS complexes undergoing UV denaturation shown in Figure 3-5. HEWL, free and bound, undergoes a gradual denaturation when exposed to UV radiation. The process is linear with two components describing the event. The plot progresses from scan 1 (red) to 60 (blue).

All three systems appear to be equally susceptible to UV degradation (PCA, shown in Figure 3-7) and each have the same route towards their endpoints with equal additions of component 1, and equal losses of component 2 in each case. This PCA plot does, however, highlight the structural differences between the three samples even though quantitative structures cannot be deduced from these data. The alteration of temperature degradation (and DSF results) but not the UV irradiation result for HS(Zn) suggests that the binding of the HS(Zn) ligand increases the accessibility of the hydrophobic regions of the protein to the surrounding buffer through a conformational change. It can be suggested that, owing to the consistency in the stability of the proteins to UV irradiation, the core structure remains intact in each case, suggesting that the zinc form of HS is affecting the surface only. This could lead to a more flexible surface structure capable of adopting several conformations that may be more susceptible to binding other proteins – especially if hydrophobic regions of the protein become exposed to the buffer. The UV degradation data presented here are further investigated in the chapter covering data analysis (Chapter 6).

3.2.3. General discussion of results.

Here, it has been shown that a complex of HS and HS(Zn) bind to a model amyloid forming protein, HEWL, altering its secondary structure to that resembling a non-cytotoxic amyloid complex of heparin and apomyoglobin at pH 5.5 [114]. This structural change consists of a large loss of α-helical content, replacing it with β-sheet. It can be hypothesized that these structural changes are taking place within the

α domain, as this is the domain that contains the majority of α-helices in HEWL [111]. This α domain contains 2 of 3 predicted amyloid forming hotspots (by hydrophobic cluster regions) [111,144]. The greater flexibility of the surface could open up the core elements of HEWL for further interactions, these core elements (amino acids 26–123 and 32-108) have been identified previously as being core structures in lysozyme amyloid formation [120]. The far UVCD spectra induced by HS and HS(Zn) also closely resemble amyloid HEWL created using high concentrations of ethanol [145]. The folding/unfolding of HEWL is known to be pH dependent [111] and, in acidic conditions the β domain unfolds first, followed by the α-domain [111,146]. The stability of HEWL is also known to change dramatically with pH, especially at lower pH's (several studies summarised in Figure 3-7 [111]). This property was known early on; in 1976 it was said that 'charge balance in a protein plays an important role in maintaining its conformation' [147], and this could be the situation being observed here. HS and heparin are two of the most negatively charged natural molecules, de-N-sulfated heparin having a pI between 1 and 3 [148].When these molecules bind to proteins the properties of the protein within the complex could be changed via the alteration of the pI, thereby something that was initially stable at higher pH's would now have altered surface charges due to the combined effect of its positive charges becoming involved in electrostatic binding with the negatively charged HS or HS(Zn) and the presence of a negatively charged polysaccharide. This altered pI would then be more susceptible to destabilisation and conformational change at pH values at which it was initially stable. Therefore HEWL, which has a natural pI predicted to be at 11.35 [149], could conceivably

become unstable at physiological pHs when bound to the negatively charged HS, shifting its pI towards the acidic.

There are several hypotheses that could be tested in future work following from this study. Firstly, the mode of binding in the active site could be investigated to confirm whether the HS and HS(Zn) ligands are indeed binding in subtly different ways. This has been attempted via MCD in Chapter 4. MCD is a method capable of selectively observing Trp residues. Another potential method is the observation of the protein in near UVCD (250-310 nm), NMR or by tryptophan fluorescence after the blocking of Trp with an organic reagent as described by Ladner *et. al.* [150]. Secondly, the pI of the complex could be tested with simple absorption at 600 nm to obtain the optical density of a solution containing lysozyme across a pH range, and compared to the results in the presence of a heparin/HS ligand. Another way to test the pI of complexes would be with isoelectric focusing. In order to do this, the protein would need to be crosslinked to the heparin/HS ligand via EDC or another crosslinking method to avoid the high current involved causing disassociation of the protein from the ligand.

Chapter 4: Magnetic Circular Dichroism as a technique to study tryptophan:ligand interactions.

4.1.1. Introduction to MCD.

MCD is a technique arising from Faraday's observation in 1846 that the plane of polarised light rotates when subjected to the poles of an electromagnet [151,152]. These differences in optical rotation (OR) only occur when subjected to a longitudinal magnetic field, one that is in the same direction as the propagation of light along the Z-direction. When measured at wavelengths in which a chromophore from the sample absorbs this then gives rise to MCD. The Magnetic Optical Rotation (MOR), which is the rotation of polarised light through a sample in the presence of an applied magnetic field, and MCD signals are additive to the natural OR and standard CD within a sample, and the strengths of both are in direct proportion to the magnetic field applied to the sample. This means that when H_z (H – the magnetic field, along the Z axis) equals zero, then the MOR and the MCD also equal zero. Natural OR results from very low symmetry within a sample leading to a difference in the refractive indices of the sample to left and right polarized light. The angle of rotation of polarised light propagated though a transparent material (a protein in this work) and magnetic field is described by equation 4.1:

$$4.1: \beta = VBd$$

This is where β is the angle of rotation in radians, *V* is the Verdet constant for the material (rad T^{-1} M^{-1}), B is the strength of the magnetic field in the direction of the light propagation (in Tesla) and *d* is the path length where the light, sample and magnetic field interact (in meters). This effect is shown in Figure 4-1.

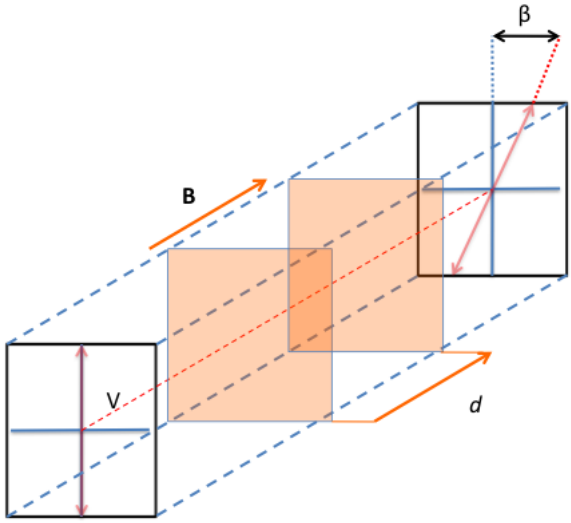

Figure 4- 1: A schematic illustration of the origins of MOR, the principle on which MCD is based. The letters refer to parameters in Equation 4-1. The red arrows refer to the direction of the polarised light as it is altered by passing through the magnetic field.

In MOR a substance with a positive Verdet constant would undergo a L-rotation (anticlockwise, relative to the observer facing the light) when the direction of propagation is parallel with the magnetic field. In the presence of a magnetic field anti-parallel to the direction of propagation a R-rotation occurs (clockwise). Therefore if light is passed through a sample in these conditions and then reflected back through the sample the rotation is doubled. In our experimental set up this is

achieved through the reversal of the static magnet employed to reverse the orientation of the poles relative to the propagation of the light.

MCD differs from MOR in that it occurs, like conventional CD, only where an appropriate chromophore absorbs, while MOR is observable at all wavelengths. MCD differs from conventional CD in that the signals rely on the splitting of ground and excited states of chromophores by an applied magnetic field. MCD has three potential contributions (components) that arise from three separate possible electron states, referred to as components A, B and C. The MCD signals observed in the near UV arise form components A and B, they originate through $\pi \rightarrow \pi^*$ transitions, with fine detail added to the features by the ^1L_b vibrational transitions at 292 and 286 nm, and the negative feature at 275 nm is associated with the ^1L_a band [153,154]. The transition exists in a triply degenerate excited state, within which the angular momentum of the electrons change depending on the local environment [155,156]. With no magnetic field present, the net moment is zero. However, when a magnetic field is applied, these are then split into each of the 3 states. This means that 3 different transitions *x, y, z* are now possible and no longer cancel each other out as the refractive indices to circularly polarised light ($\eta+$ and $\eta-$) are unequal. These 3 transitions can be observed as the 3 components, A, B and C. (Components A and B shown in Figure 4-2). Component C is temperature (1/T) dependent [155] and can be isolated through a temperature gradient, it is also the weakest contributor. It will not be considered within this work, which involves organic biological molecules under or close to physiological conditions, because it originates in a paramagnetic ground state in free radical containing species or metal ions.

The magnetic transitions observed in MCD arise from the different energy levels set up by the Zeeman effect on the electron orbitals and result in components A and C. In component A the orbitals of the excited states become split into a higher and lower energy state that interact differently with left and right polarised light. In component C the ground states of these electrons are split, in a manner that is directly proportional to the inverse of the temperature, and thus can be differentiated from other components via a temperature gradient. The B component is one in which two separate excited state orbitals have energy levels very close to one another which allows mixing of the states. Components A and B are shown in Figure 4-2.

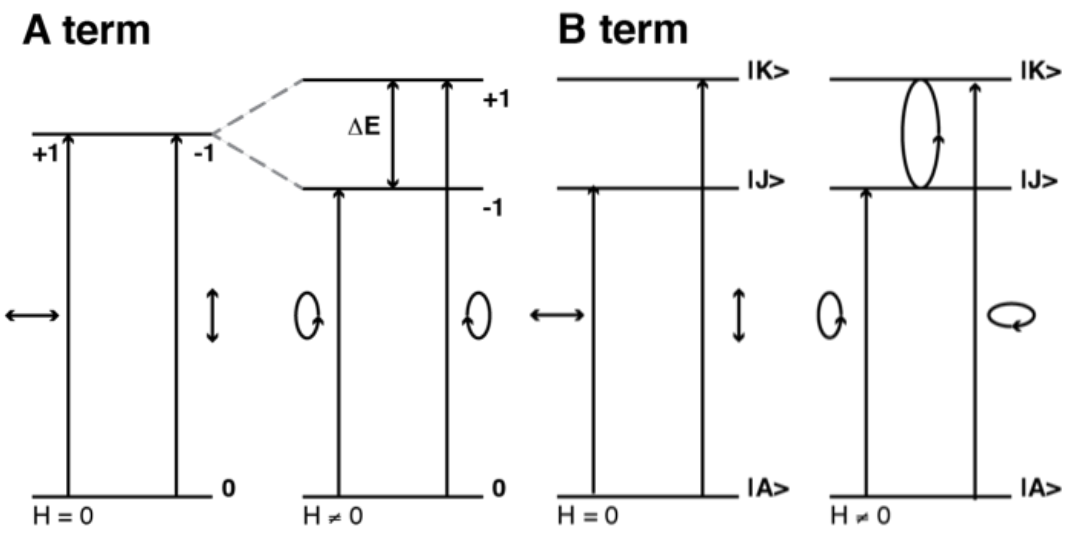

Figure 4- 2: Schematic of the split electron energies as a result of the Zeeman effect, in the presence of a magnetic field, giving rise to a A and B components in MCD spectra.

4.1.2. MCD in the near UV (270 – 350 nm).

A standard MCD spectrum in the near UV (270-350 nm) of a protein will typically contain 5 contributions. The signal arising from tryptophan (Trp) is by far the most significant. It gives rise to the only positive features by any amino acid in the near UV, making it distinctive [156]. The Trp features comprise a sharp positive feature between 292 and 295 nm ($\Delta EM = 2.35\ M^{-1}\ cm^{-1}\ T^{-1}$), a positive shoulder at 284 nm and a broad negative feature with a maximum at 267 nm. Other contributors in this region are from phenylalanine (Phe) and tyrosine (Tyr). The contribution from Phe is via a weak, negative feature at 267 nm ($\Delta EM = -0.055\ M^{-1}\ cm^{-1}\ T^{-1}$) and Tyr, which exhibits a stronger band, comprising a single minimum negative feature at 275 nm ($\Delta EM = -0.38\ M^{-1}\ cm^{-1}\ T^{-1}$). It has been shown in previous studies that TRp quantification can be achieved through measurement of the maximal peak value at the feature between 292 and 295 nm [154,157,158]. It has also been observed that this feature can move between protein samples, so locating the maxima is essential [154]. The shift in the feature observed at 292-295 nm demonstrates, to some extent, that Trp residues are sensitive to their local environment [154,159]. Other evidence for this sensitivity comes from a small shift in peak position which can be observed by changing the solvent, and a consistent blue shift of the positive spectral feature to 292.1 nm when the protein becomes unfolded [154]. The Tyr feature at 275 nm gradually red shifts as the pH increases as the chromophore becomes ionised (pK tyr ~10.1), up to a maximum of 292 nm [154,157]. This can be exploited for the determination of Trp:Tyr ratios within a protein [160]. The local conditions of tryptophan have been utilized in the observation of quaternary structure in

heamoglobin using bands at 292 nm and 286 nm [161], and also on the same protein for time resolved MCD study of a hydrogen bond formation [162]. The technique has also more generally been used to study a wide range of proteins containing metal ions, especially haem groups [163].

Artifacts can appears in MCD spectra with distortions to the direction, strength and focal point of the applied magnetic field between samples. Alternatively artifacts can arise from changes in the beam position relative to the magnetic field giving rise to much the same effects. These issues cannot occur with spectra obtained in this body of work owing to the use of a static, permanent magnet (i.e. not an electromagnet) and the precise optical benches employed to stabilise the beam on B23 at Diamond light source. Any signals arise directly from the interaction of light between the sample and the magnetic field.

Since it has been shown previously that Trp residues can be sensitive to their local environment [161,162], and the main MCD feature has been used to identify ligand binding [162,164], it was proposed to investigate whether this technique was suitable for the observation of GAG:protein interactions with specific Trp residues.

4.2. Results.

Initially, it was confirmed that the MCD signals observed between 250 and 350 nm are features that arise from observing the Trp amino acid. This is demonstrated in Figure 4-3.

4.2.1. Free Trp amino acid signal in MCD and its sensitivity to environmental changes.

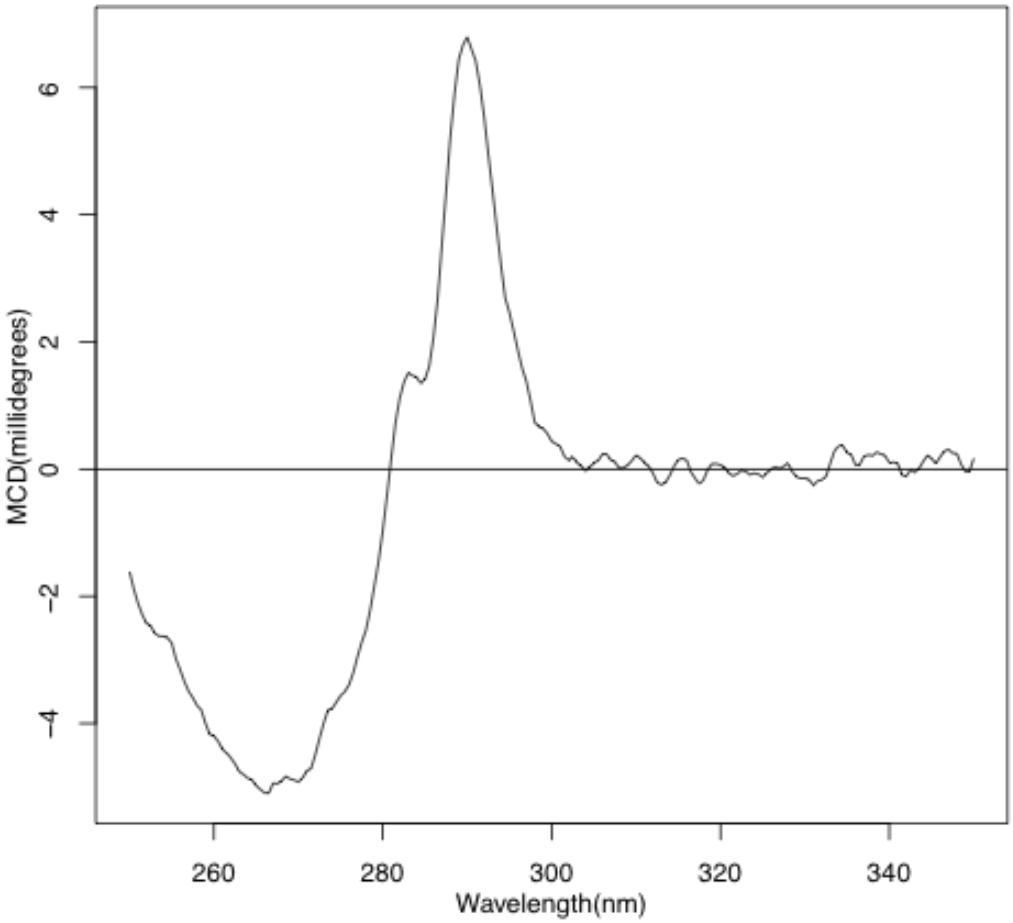

Figure 4- 3: MCD spectrum (250-350 nm) of tryptophan amino acid (14.8 µg/ml / 71 µM, PBS). Averaged over 4 scans, 1 cm path length, 1 second integration time, 1 mm slit, 1.5 T magnetic field strength.

This spectrum (Figure 4-3) is similar to a spectrum of tryptophan reported in the literature. Trp residues indeed exhibit a strong MCD spectrum in the near UV [158]. There are 3 features present, as reported in the literature, and these are thought to arise from two transitional modes. The large negative feature at 270 nm and the positive 292 nm feature form a B component contribution while, the harder to define, positive shoulder at 286 nm is thought to originate from an A component. The main positive feature at 292 nm has previously been used when quantifying protein concentration in solution [154,155,157,158,159] and has also been utilised in the determination of tryptophan-tyrosine ratios in proteins [160]. The positive feature present at 286 nm is followed by a broad negative feature at 270 nm, both of which are present in the previously reported spectra of Trp [154,155,156,165]. The reported sensitivity of Trp to its local environment was then confirmed by changing the solvent used to dissolve the amino acid, as shown in Figure 4-4.

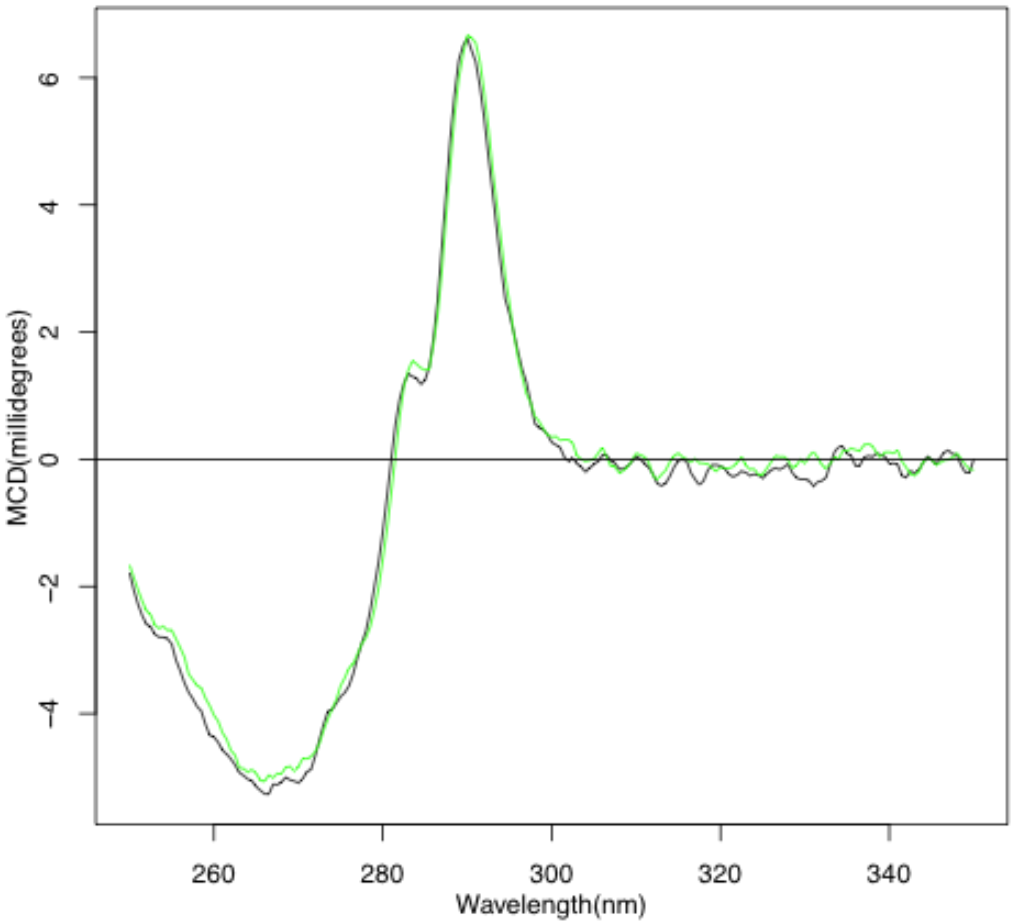

Figure 4- 4: MCD spectrum (250 – 350 nm) of Trp, PBS (black line) and Trp in 25% ethanol / 75% PBS (green line). 14.8 µg/ml (71 µM), 1 cm path length, 1 second integration, 1 mm slit. Note the difference at 286 nm, possible due to affecting the vibrational fine detail of the chromophore.

Altering the solvent environment for the amino acid alters the MCD signal of Trp (Figure 4-4). This has manifested itself in a slightly broader peak at 292 nm, and a more intense positive 286 nm feature. The broad negative feature at 267 nm remains unchanged, suggesting that this may not be where observable changes will occur in

future experiments. Interestingly, the intensity of the 292 nm peak remains constant in this instance, suggesting that changes which may be observed during ligand binding will be to the position of the 292 nm peak, rather than to its intensity, as was observed during unfolding experiments reported in the literature [158]. The positive feature at 286 nm also appears to be sensitive to changes in the amino acid environment, suggesting that this little studied feature could be of interest in later experiments.

After confirming that Trp is indeed sensitive to its local environment, several proteins were proposed for assessing this approach. One of these involves the use of FGF2. This protein was used for its defined interaction with heparin. FGF-2 has a single Trp four amino acids apart from the heparin-binding site at the start of a β-strand [129]. It was reported that heparin binding to FGF-2 induces structural changes within the protein, however, no direct interaction between this Trp and heparin is thought to take place [139,166]. This system should allow for the observation of changes to the environment of tryptophan occurring from the restructuring of the protein.

The second study utilises SOD1. This protein also has a single Trp residue, and has previously been crystallised with several ligands interacting with the tryptophan. One of these ligands, 5-fluorouridine, was selected and investigated further here as an example, in which a confirmed, direct interaction takes place.

Lastly, HEWL was used. Although this protein contains six Trp residues, previous work has provided evidence for a direct HS:Trp interaction within the active site by its substrate (Chapter 3, [142]), where two Trp residues (W62 and W63) are present [143]. This work suggested that, HS and HS in the zinc form (HS(Zn)) have distinct interactions with the surface Trp residues within the binding cleft.

4.2.2. MCD spectra of FGF2 interactions with heparin.

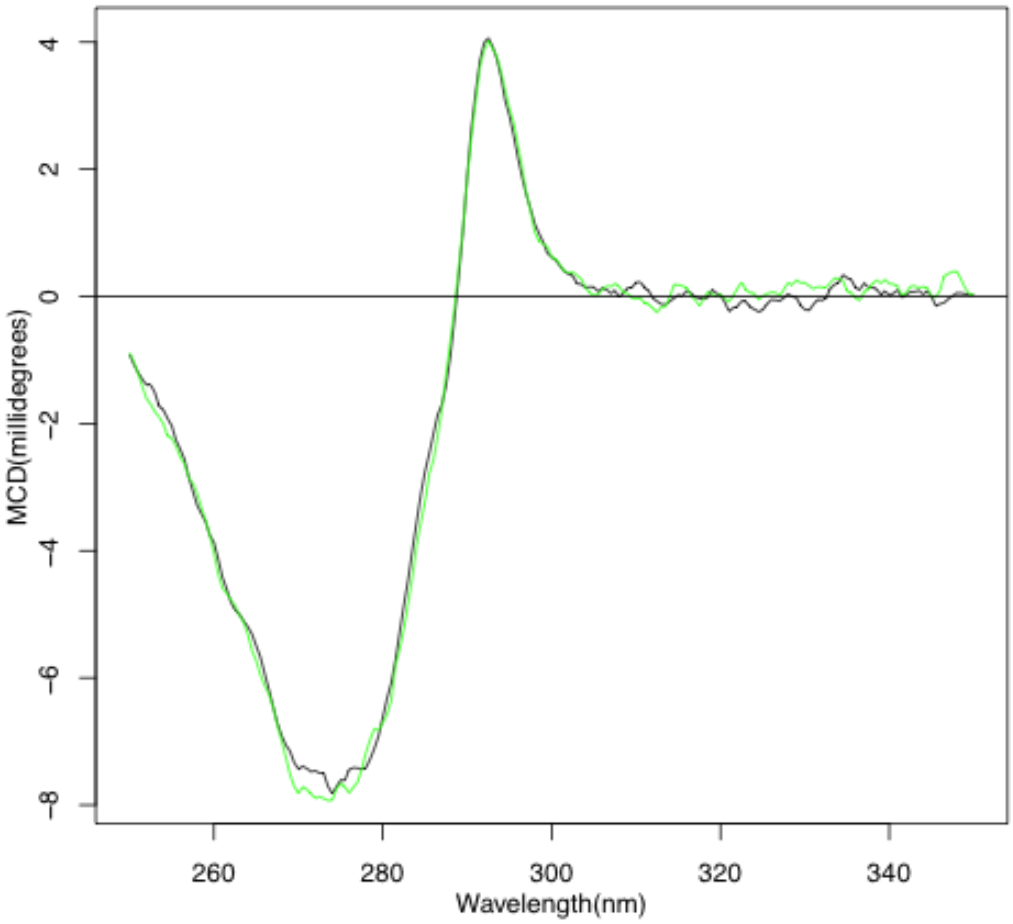

Figure 4- 5:MCD spectrum (250 – 350 nm) of FGF-2 alone (black line) and in the presence of heparin (green line) in PBS, with FGF-2 at 0.7 mg/ml and heparin at 3.5 mg/ml. 1 cm path length, 1 mm slit, and 1 second integration time were employed.

Initially, the MCD spectrum of FGF-2 was recorded in the presence and absence of heparin (Figure 4-5) in conditions that replicate previous work [140,166]. As shown in Figure 4-5, despite the known interaction between the ligand and the protein, there

were no changes to the spectra. This indicates that there is no direct interaction between the Trp in this protein and the ligand as expected, and that changes to either the main 292 nm MCD feature, or the 286 nm shoulder will be due to direct changes in the local environment of Trp and its interactions and not the result of global secondary structural changes as a consequence of binding. It is interesting to note that this is the only case in which the shoulder at 286 nm is absent, indicating that the local environment of the Trp may have a large influence on the observed A component, whether it be from local interactions with the conformation of the protein around the amino acid, or through interactions with external factors, buffer or ligands. The proposed binding site is 4 amino acids removed from the Trp position in the sequence at the start of a segment of β-strand. This result suggests that, if there are to be any interactions that are observable by MCD, they either need to be direct, or closer than 4 amino acids, thereby altering the electronic configuration of the delocalised π-π electron cloud present above and below the indole ring of Trp. This result also suggests that MCD is not sensitive to the secondary structure changes known to occur during this interaction.

The hypothesis that a direct interaction was required for an altered MCD signal was then investigated with SOD1, a protein which is known to have a direct Trp:ligand interaction. The interaction of the single Trp in SOD-1 with various ligands was confirmed in crystallographic studies [167,168] in which it was demonstrated that four ligands interacted with the sole Trp residue within the “drug pocket”. This interaction was through π-π stacking of the ligands with the amino acid side chain. In

this study one of these ligands has been selected and its interactions with SOD-1 investigated (Figures 4-6 - 4-10).

4.2.3. Confirmation of SOD1 interactions with ligand in solution.

Initially, investigations sought to confirm that the selected ligand, 5-fluorouridine interacted with SOD-1 in solution. This was achieved using three methods; far UVCD, near UVCD spectroscopy and a thermal degradation of the protein in the absence and presence of the ligand, tracked using far UVCD.

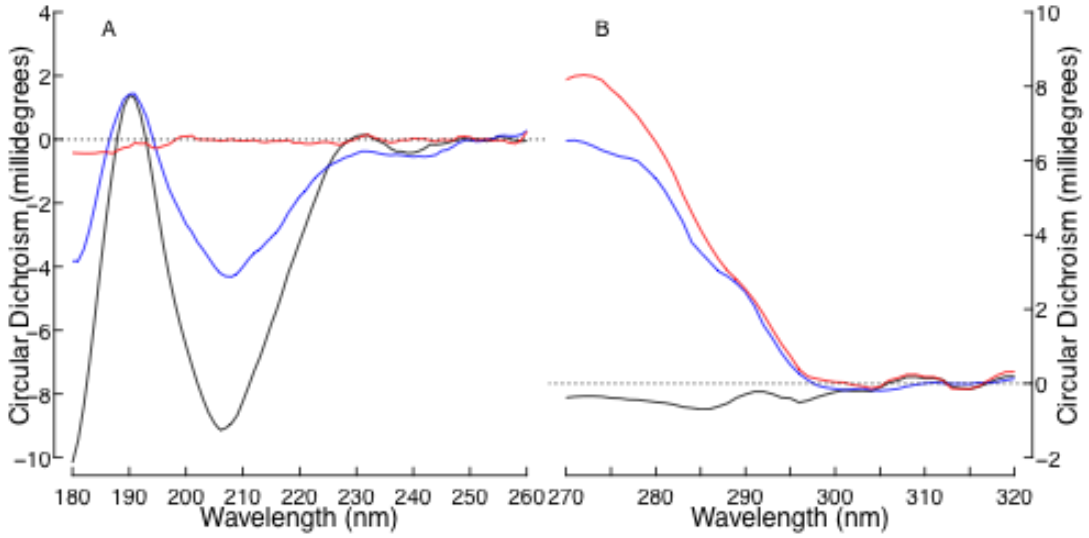

Figure 4- 6: Figure 4-6A shows the far UVCD (180-260 nm) spectrum of SOD-1 alone (black line 0.03 µM) and in the presence of the ligand 5-flurouridine (blue line, 1:1 Molar ratio). The red line is the ligand alone, recorded with 1 second integration time, 0.1 mm path length. B shows the near UVCD (270 – 320 nm) of the same samples (SOD-1: 0.16 µM, ligand: 1:1 molar ratio) in panel A. These spectra were recorded using 1 mm path length cell and 3 second integration time. All samples were recorded in 20 mM phosphate buffer.

Figure 4-6, panel A, shows the far UVCD of SOD-1 in the presence (blue line) and absence of the ligand (black line), and also the ligand alone (red line). The near UVCD spectrum is also shown in Figure 4-6B, with matching colour coding. Figure 4-6A shows first that, when SOD-1 is in the presence of the ligand, there is a substantial spectral change around the large negative feature at 205 nm. This is manifest in a positive contribution, but one that does not, or very minimally, affect the positive feature present at 190 nm. This suggests a possible increase in β-turns,

known to exhibit a positive feature at 205 nm. The observed spectral change can only be attributed to alterations in the protein conformation because there are no chiral chromophores in the ligand that are active at these wavelengths as demonstrated by the featureless spectrum of the ligand in Figure 4-6A (red line). The near UVCD spectrum is dominated under these specific conditions by the signal contribution from the ligand, peaking at ~275 nm. There is a change within the protein:ligand spectrum in comparison to either the protein or ligand spectra alone, providing evidence of an interaction, in agreement with the far UVCD experiment.

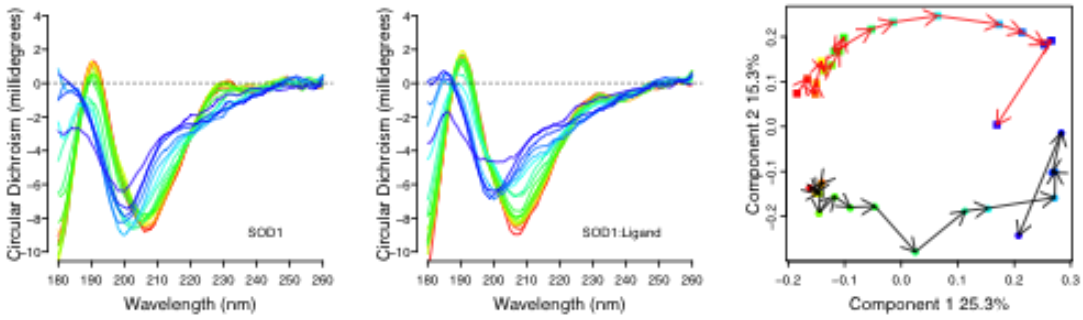

Figure 4- 7: A SOD-1 unfolding pathway as tracked by far UVCD spectroscopy (180-260 nm) in a temperature series from 10 °C (red) to (blue) 90 °C in 5°C steps, followed by a final 20 °C step. SOD-1 at 0.03 µM in a 0.1 mm path length cell and a 1 second integration time. B: A repeat of sample A conditions in the presence of the ligand, 5-flourouridine, 1:1 molar ratio. C: PCA analysis of data presented in A and B showing the structural changes between each temperature step in the experiment. This experiment shows that the unfolding pathways are subtly different in the presence of the ligand but, generally, they follow a similar pathway, the differences occurring mostly in component 2.

Figure 4-7 shows the denaturation profile of SOD-1 as tracked by far UVCD in the absence or presence of the ligand. The temperature series employed during this experiment was 10°C (red) to (blue) 90 °C (in 5°C steps), and a final step where the temperature was returned to 20 °C. The sample was held at each temperature to equilibrate before a single scan was recorded (180-260 nm) to track the structure within the protein at that temperature. Although a single scan is susceptible to an increased noise (twice as much noise compared to 4 scans), this method minimises sample damage via the UV degradation method at each point, and can be used to reveal the unfolding pathway of the protein when subjected to a temperature gradient. This pathway was then analysed using principal component analysis (Figure 4-7C) where a direct comparison of the unfolding pathways can be observed. Both samples have broadly similar unfolding pathways, however, these do not share the same secondary structure – in agreement with the previous experiment (Figure 4-6A). There is a slight difference between SOD-1 alone (red line) which gains more of component 2 as it unfolds and SOD-1 with ligand which does not. However, both samples do start and finish at similar points around the dominant contributor, component 1. Upon returning to 20°C, both samples exhibit the same refolding mechanism. The slightly altered thermal denaturation pathways, coupled with the distinct spectral features in a region that reports protein secondary structure, alongside that of an alteration in the spectrum from the ligand in the near UV, provide evidence of an interaction occurring under these conditions. This provides added confidence in the proposed experiment to use this ligand protein combination for further MCD studies as an exemplar of a ligand that is known to interact directly with a Trp residue.

4.2.4. MCD spectra of SOD1 in the presence and absence of ligand.

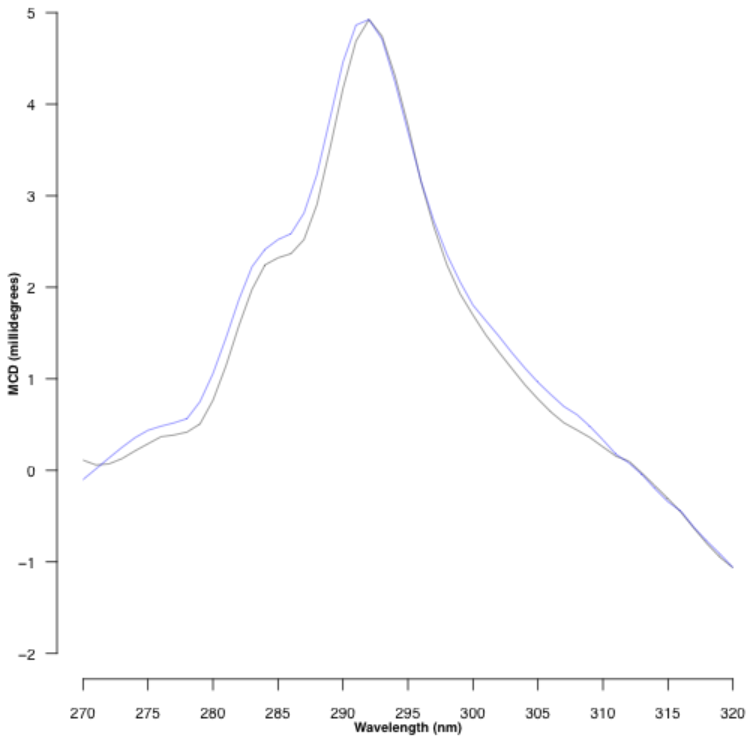

Figure 4- 8: MCD (270-320 nm) spectra of SOD-1 alone (0.15 mM, black line) and SOD-1 in the presence of the ligand, 5-flourouridine (1:1 molar ratio, blue line, normalised to SOD-1 maximal signal). Samples recorded with 1 mm path length cell, 3 second integration time. Spectra normalized to maximal value of the positive feature at 292 nm. This shows a significant ($P<0.001$, 286 nm) change in the presence of the ligand demonstrating an interaction between ligand and protein. Spectra are an average of 4 complete cycles of north to south and south to north orientations of magnet.

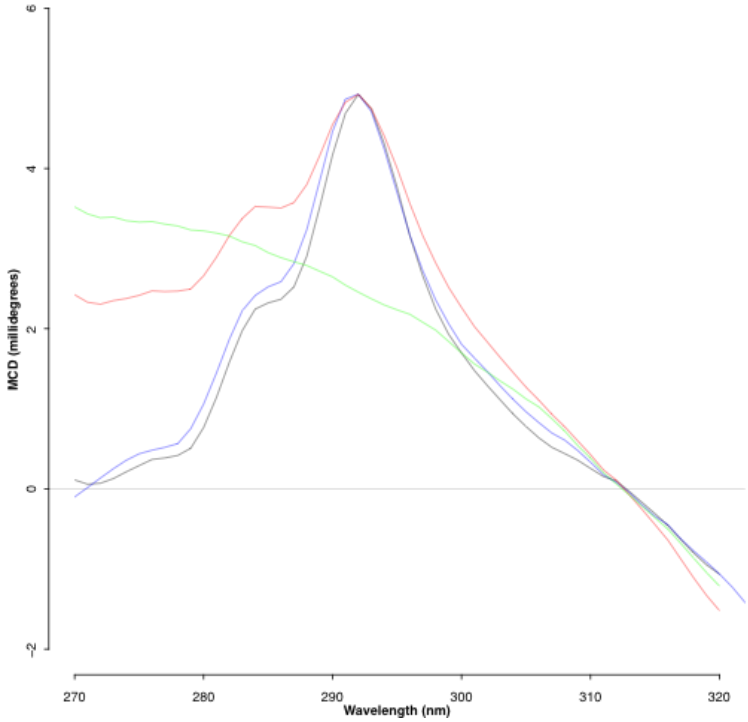

Figure 4- 9: MCD (270-320 nm) spectra of SOD-1 alone (0.15 mM, black line) and SOD-1 in the presence of the ligand, 5-flourouridine (1:1 molar ratio, blue line, normalised to SOD-1 maximal signal). Ligand alone (green line) and ligand + protein, (red line) to demonstrate the additive theoretical spectra normalised to maximal SOD-1 signal. Samples recorded with 1 mm path length cell, 3 second integration time.

The investigation of SOD-1 with its ligand, 5-flourouridine, demonstrates that an MCD spectral change (Figure 4-8) occurs upon introduction of a ligand that is known to interact directly with the Trp of SOD-1 (SOD-1: black line, SOD-1:Ligand: blue line). The ligand alone has a broad positive feature (Figure 4-9, green line). The two main features of these changes are the peak broadening of the positive 292 nm feature, and the shoulder centred around 284 nm, which is significantly

different from that of the SOD-1 alone (P<0.0001, 284 nm). Shown in Figure 4-9 is the theoretical spectrum of the ligand in the presence of protein in which no interactions are taking place, a purely additive case that has been treated in the same way as the observed data. This theoretical spectrum is different to the observed Trp:ligand spectrum indicating an interaction between the ligand and the Trp. While it can be seen in Figure 4-9 (green line) that the ligand itself has a MCD signal, it is featureless, so changes observed at 286 nm are thought to come directly from the Trp amino acid residue, consistent with other changes observed within this chapter. Studying the data with the benefit of having other data sets and published data it becomes apparent that there should not be a negative feature above 310 nm, and the features at 275 for Tyr and Trp have negative features, as shown for the dominant Trp in Figures 4-2 and 4-3. In addition to this, the height of the features at 292 nm is greater than expected from a Trp at 0.15 mM in a 1.4 T magnet at 1 mm path length.

The possibility of light scattering as a source of this result was investigated, however, there is no indication within this data set that this phenomenon has occurred. The voltages were comfortably below the upper accepted levels (max voltages, <400 mV) and there is no sign of scattering within the data shown in Figure 4-6B which is a recording of the same sample in the same cell subsequent to the MCD measurements in the presence of a dummy magnet. It was further investigated whether this could be an induced effect within the buffer system in the presence of a magnetic field, and would therefore need to be corrected for.

It was decided to treat the ligand spectrum, which appears to be totally featureless, as a baseline and subtract it (see Figure 4-10).

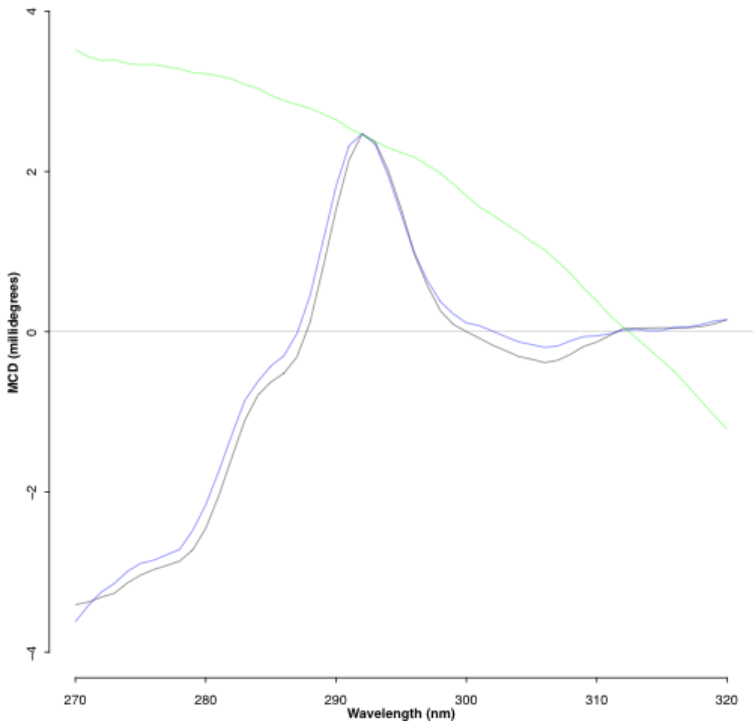

Figure 4- 10: MCD (270-320 nm) spectra of SOD-1 alone with the green line subtracted (0.15 mM, black line) and SOD-1 in the presence of the ligand, 5-flourouridine with the green line subtracted (1:1 molar ratio, blue line, normalised to SOD-1 maximal signal). Ligand alone, (green line) herein used as a baseline for the other two spectra. Samples recorded with 1 mm path length cell, 3 second integration time.

As can be seen in Figure 4-10, the ligand spectrum has been subtracted the experimental spectra and this has given a new set of MCD spectra for the SOD-1 and SOD-1:ligand. There are now no features early in the spectra and the broad feature between 270 and 280 nm now has a negative sign. After having the ligand subtracted

as a background the signal intensity at 292 nm is now at the expected amplitude for a Trp at 0.15 mM in a 1 mm path length cell with a 1.4 T magnetic field. The signal size of ~2 units in Figure 4-10 equates to ~ 0.1 unit for 1 µM of Trp present (for a 1cm path length cell), and if compared with Figure 4-3, which has a signal size of ~6.5 from a 1 cm path length at 71 µM, this also equates to ~0.1 unit per 1 µM of Trp in a 1 cm path length cell. These observations combined indicated that Figure 10 is the true MCD spectrum of SOD-1 and SOD-1:ligand. The fact that the ratios between peak heights and features are not altered by subtracting the same spectra from both the previous analyses confirms that these spectra demonstrate an interaction between the protein and ligand, that is statistically significant at 286 nm.

4.2.5 MCD spectra of HEWL in the presence and absence of GAG ligands

An additional protein, HEWL, has been studied as an example of protein:heparan sulfate interactions with Trp. The protein has six Trp, two of which are situated in the active site [143], the proposed binding site for HS. Earlier work discussed in chapter 3 and in Hughes *et al* (2012, [142]) suggests that, although their interactions with HEWL result in similar secondary structure changes, (Chapter 3), there are differences in how these ligands (HS and HS(Zn)) interact with the protein, see Figure 3-2 of Chapter 3. Here the MCD spectra for this protein in the absence and presence of the two ligands (HS and HS(zn)) is presented (Figure 4-11), which again show changes in the spectrum comparable with the contribution of Trp, and consistent with a direct interaction of the HS ligands with the Trp residues at positions 62 and 63.

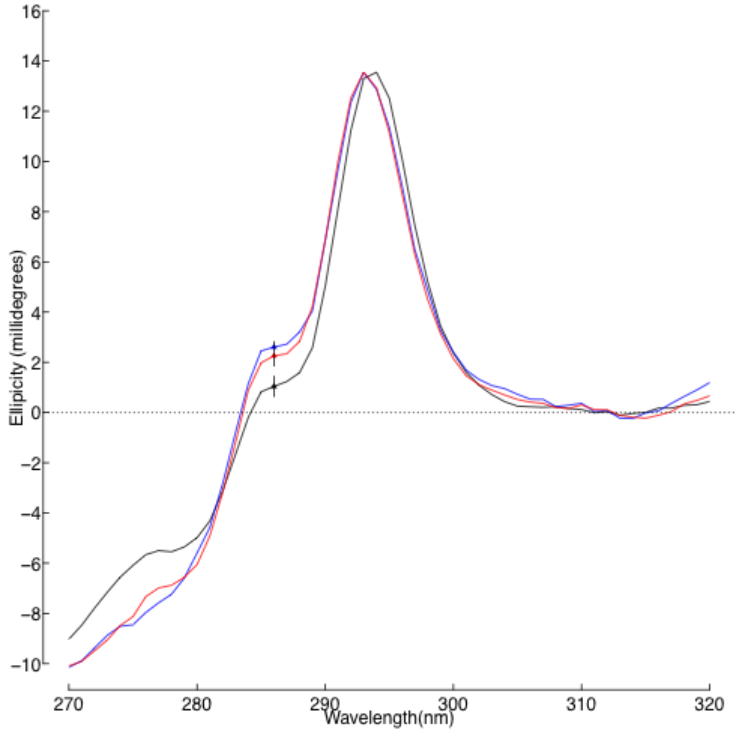

Figure 4- 11: MCD spectra of HEWL 0.4 mg/ml, 12.5 mM NaCl PBS, 1 cm path length, 3 second integration time, HS and HS(Zn) at 0.03 mg/ml. Spectra of HEWL:HS (red line) and HEWL:HS(Zn) (blue line) have been adjusted to equal the maxima value of HEWL alone (black line). Recorded using a 1.4 T magnet. Standard deviations are shown at 286 nm, where signals involving both ligands are statistically different from the signal of HEWL alone ($p \leq 0.0002$).

Figure 4-11 shows the spectrum of HEWL, which is consistent with the literature spectra [158,160]. The samples used to collect the data in the presence of ligand, in both cases had a tendency to aggregate within the cell, causing some scattering noise. To deal with this, samples were centrifuged prior to measurement and the

supernatant removed for subsequent measurement. This resulted in some loss of protein in the pellet. The data have therefore been normalized to the maximum amplitude of the HEWL alone sample (Figure 4-11, black line). The result is that only the shape and position of features is relevant here, and not the values of the feature amplitudes – although relative amplitudes within a spectrum should be considered. Figure 4-11 shows that with the addition of HS(zn) (blue line) several features change in the spectrum. These are manifested in the red shift of the broad negative feature ~267 nm. This feature is a result of both Trp and Tyr amino acids (with negative contributions at 267 and 275 nm respectively). There is also an increase in the intensity of the shoulder, apparent at 286 nm, and a slight blue shift of the main positive peak present. These changes are consistent with the hypothesis that a ligand-Trp interaction is causing an alteration in the Trp contribution. It is known that the Tyr negative band typically present at 275 nm is able to red shift progressively to 292 nm (pH 12.1) above pH 8 due to it becoming ionized. If this were the case then that would be expected to have a negative influence on the shoulder at 286 nm, leading to a negative contribution, not the positive one that is observed here. If it were the case that the Tyr had blue shifted, then we would not expect the broad negative feature as a whole to be red shifted. It is also to be considered that the relative contributions of these signals, since tyrosine is a relatively weak contributor to the near UV MCD signals, and in this protein being at a 2:1 ratio lower abundance than Trp, it can be asserted that changes in this spectrum are to be assigned to contributions from Trp. The small blue shift observed in the positive feature at 292 nm could be attributed to the resulting conformational changes, as previously observed. It is of note that previous experiments have shown

that at high Tyr to Trp ratios (>35:1) causes the 292 nm feature to red shift by 1 nm, indicating again that this shift is not due to the possible alteration of the local environment of the environment of Tyr residues. Figure 4-11 also shows that HS (red line), induces very similar changes to HS(Zn). These changes include the common feature at 286 nm where, again, the change relative to HEWL alone is statistically significant ($p \leq 0.0002$).

4.3. Discussion of results; MCD is suitable for showing direct Trp:GAG interactions

Here it is has been shown that we can observe ligand binding at ~284-286 nm in MCD spectra and that this area is susceptible to local environmental changes. Figure 4-4 demonstrates this susceptibility, the changes observed are probably due to hydrogen bonds that are being formed or broken via a direct interaction between the Trp and ligand. MCD has been shown previously to demonstrate protein:ligand interactions [164]. The authors observed a decrease in signal amplitude at 292 nm and claimed that the simplest explanation would be that this is the result of a ligand-binding event, and the 292 nm feature has subsequently been used again to monitor hydrogen bonding [162]. However, this has not been observed to be the case in the work conducted herein, one clear example would be the SOD experiment, where there is a proven interaction with the Trp through crystallographic studies and yet no positional change in the feature at 292 nm. This lead us to propose a new area of interest within a spectrum at 286 nm, so far not explored using this technique, as a sensitive region of the spectrum for detection of Trp:ligand interactions.

Any further work on this topic would focus on establishing additional proof of what was observed here, and the mechanism by which the signal change at 286 nm is manifested. Definitive proof of direct Trp:ligand interactions could be provided by the heavy labelling of the nitrogen atom (^{15}N) in a specific tryptophan residue within a protein, such as HEWL, and performing a parallel experiment using NMR to observe the direct interaction in solution at the atomic level. The heavy labelling of a Trp could prove to be one method of isolating a single Trp in a protein such as HEWL which has six Trps, by the ^{15}N effecting the electronic configuration of the amino acid and therefore altering the signal to provide a means of selecting solely that contribution. An alternative method of obtaining further evidence of what has been observed here would be to modify exposed Trp residues through indole chemistry to add a blocking methyl group to the amino acid, thereby preventing binding because of steric hinderence [150].

Chapter 5: Investigating the interactions between non-active heparin like compounds derived from the shrimp *L. Vannamei* and Antithrombin.

5.1. Introduction.

This chapter will investigate the structural effects of the binding of antithrombin (AT) to heparin and the synthetically derived pentassacharide, which is described below. The aim of this investigation is to determine the structural differences in the interactions of AT with several heparin derivatives of different structures. The Pentasaccharide refered to herein is a pharmaceutical drug sometimes known as Arixtra or Fondaparinux, Figure 5-1.

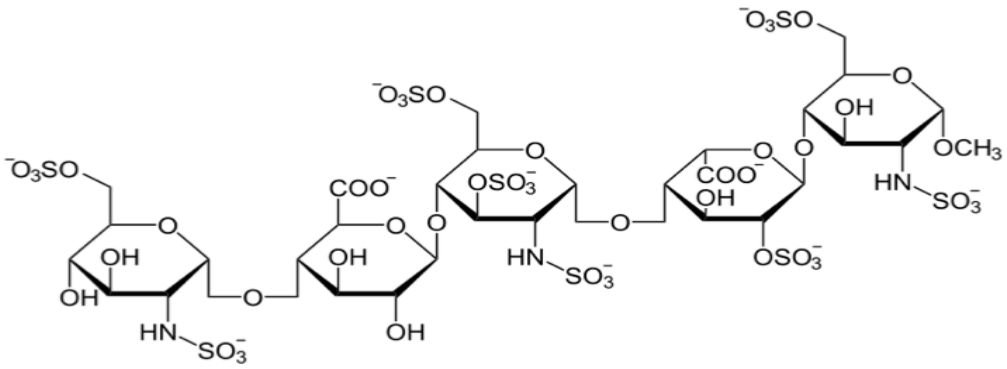

Figure 5- 1: The structure of the chemically synthesised pentasaccharide sequence showing the important rare 3-O-sulfate group.

Investigating the interactions between non-active heparin like compounds derived from the shrimp *L. Vannamei* and Antithrombin.

5.1.1. Antithrombin.

AT is a serine protease inhibitor synthesised in the liver and discovered in the early 20th century. It is a 58 kDa protein that contains four glycosylation sites in its major form, 432 amino acids and three disulphide bridges. The carbohydrate component accounts for 15% of the total mass, and human AT [169] is present at between 14 and 0.12 mg/ml in serum samples [170,171]. The main targets for the activity of AT are thrombin, factor IIa and factor Xa, however, to a lesser extent, it is also known to inhibit urokinase, trypsin, plasmin, kallikrein, factors IXa, XIa, XIIa, and tPA [172,173,174,175]. AT circulates in the blood with minimal apparent activity until it binds with its presumed ligand, HS, on the endothelial blood vessel cell walls. Upon activation, AT binds thrombin to form a tightly bound complex that is then removed from circulation [176,177]. The addition of heparin to AT can increase its activity against thrombin up to 4,000 fold. The activation of AT by full length heparin against factor Xa is around 500 fold higher [178]. Factor Xa is a key factor in the intrinsic coagulation pathway processing prothrombin into thrombin, a key step in blood coagulation. The activity of AT against factor Xa decreases the rate of coagulation.

5.1.2. Heparin, and the discovery of the pentasaccharide sequence.

Heparin, first discovered in 1916 [179] for its anticoagulant properties, was associated with a heparin co-factor in 1939 [180], which was confirmed to be AT in

the 1950's [181,182]. In the following years attempts were made to relate the activity of heparin to its size, degree of sulphation and N-sulfate content [183,184] before the discovery of a high affinity fraction in 1976, by separation using analytical centrifugation [185]. This result appeared to suggest some specificity within the heparin samples for AT, implying a preferential binding to a specific sulphation pattern [185]. The fractionation of heparin led to an increase in activity of the active fraction from 155 U/mg to 248-388 U/mg (high affinity fraction, HA). The unbound heparin (low affinity fraction) on the other hand lost virtually all activity, reducing it to 19-52 U/mg. A high affinity fraction was also collected in a separate study using affinity chromatography [186]. The resulting work on this fraction to isolate an active component took the form of enzymatic degradation prior to affinity chromatography, which demonstrated that small fragments of heparin as low as hexasaccharides could bind to AT [187]. In the same year an investigation into the effect of size on AT anti Xa activity demonstrated that, while thrombin inactivation by AT required a large heparin fragment, the inhibition of factor Xa was size independent [188]. In 1978, it was observed that the Xa activity of heparin could not be attributed solely to the availability of binding sites related to the length alone, leading to the suggestion that a specific dodecasaccharide sequence may be responsible for activity [189]. Although a specific sequence for Xa activity was proposed, the authors warned against the interpretation of a 'specific dodecasaccharide sequence' as meaning only one possible binding sequence [189]. It was also suggested that the low affinity fraction, which maintains some activity, sometimes attributed to contamination, may actually be 'ascribed to sequences that

bind less strongly… …and result in partial activation of AT' [189]. Progress towards the discovery of an active sequence in heparin was made by Rosenberg, who observed an enrichment of the saccharides D-glucuronic acid and N-acetyl-D-glucosamine [190,191], proposing that a tetrasaccharide was required for activity. This tetrasaccharide was also found in the low affinity fraction however, against the supposition that a more defined structure was required. This theory was further developed by the proposal of a complete sequence that included the latter tetrasaccharide in 1979 [192], later refined to an octasaccharide [193].

Parallel to these investigations was a drive by the Choay laboratory to find the smallest heparin fragment capable of binding to AT and inhibiting the activity of factor Xa. They approached the task using both partial nitrous acid fragmentation [194] and enzymatic digestion of heparin [195]. These investigations lead to the observation of a new signal at 57.7 ppm in ^{13}C NMR experiments, exclusively present in the active octossacharides [194,196]. Subsequent experiments, to identify the missing structural element absent from previous analysis of the active fragment, led to the discovery of the important 3-O-sulfate on the 3,6-di-O-sulfated D-glucosamine residue [38]. Within the octasaccharide unit a small amount of hexasaccharide was observed. Investigating further, it was observed that this contained the NMR peak previously observed at 58 ppm [197]. This then led, via two independent routes, to the proposal of an active pentasacchaide with the same affinity as the previously proposed octasaccharide [198,199]. This hypothesis was tested via a synthesised pure compound in 1991 [200], and was found to maintain the

same high affinity for AT [201]. Two tetrassacharides were also tested, which showed significantly lower affinity (2 mM *v* 50 µM disassociation constant), showing that a pentasacchaide was indeed the minimum sequence required to bind to AT [202].

Research to this point arrived at the conclusion that a particular heparin sulfation sequence was responsible for AT activation and binding. This sequence was marked by the rare 3-O-sulfate within the synthesised pentasaccharide and once synthesised on a commercial scale, became a successful anticoagulant drug targeting the AT anti-Xa activity. While this treatment in most cases is highly effective, there are occasions in which adverse reactions can occur. It is known, for example, that patients who have undergone vasectomies can develop an allergic reaction to the antidote of heparin based anticoagulants, protamine sulfate [203]. Importantly, it is also difficult to neutralise pentasaccharide with protamine due to its small size, which is currently used for this purpose after surgery and has led to bleeding complications [204,205,206,207].

5.1.3. Potential therapeutic targets of heparin based drugs.

Alternative heparin based drugs have been developed in the form of low molecular weight heparins (LMWH). The main driver towards LMWH, each with a unique activity profile, was to avoid heparin-induced thrombocytopenia (HIT) caused by antibodies to the aggregates of heparin and platelet factor IV, which in turn can lead

to HIT thrombosis, and simultaneously be capable of selectively targeting AT and being too small to interact with thrombin or platelet factor IV. Heparin also exhibits many alternative activities that are thus far under-exploited [208]. Heparin can have activity as an anti-inflammatory agent, cause angiogenesis and also interact with a wide range of proteins involved in many disease states [54] including cancer, and is actually becoming a target for some cancer therapies [209]. There is hence a need for new compounds that both exhibit antithrombotic activity, but without the problems of the existing array of heparin based compounds such as the pharmaceutical agent pentasaccharide. There is also a demand for drugs that lack this activity, but maintain useful activities against other therapeutic targets.

5.1.4. Shrimp heparinoids: An apparent enigma.

Shrimp heparinoids (SH) purified from the viscera of the Atlantic shrimp *Litopenaeus vannamei* contained the distinctive 3-O-sulfate thought to be the marker for AT activity [210], yet these lacked the increased AT activity expected [109]. However, they do exhibit interesting properties that suggest they could be used as a basis for novel drug candidates targeting a range of desirable activities, including anti-angiogenic effects, and anti-inflammatory effects whilst maintaining the desirable negligible *haemorrhagic* activities [109,210]. Despite its lack of AT activity the SH compounds were eluted in the same fraction as the classic high affinity heparin fraction at 0.75-1.5 M NaCl. This result led to the current investigation to determine the characteristics of active and inactive AT interactions

with the polysacchairdes SH1 and SH2, unfractionated heparin (UFH), and the classic pentasaccharide sequence. It was hypothesised that the resulting secondary structures would be distinct between active and non-active heparin structures.

In line with our initial hypothesis that the interaction of active and inactive sugars with AT would result in AT adopting distinct secondary structures, experiments were designed to investigate this. The results of this investigation are presented below. Initially, SRCD of AT complexed with the various sugars that induce a range of anti Xa activities in AT were investigated (Figure 5-1). The sugars used in this study are the synthetic pharmaceutical Pentasaccharide, UFH and SH1 and 2. SH1 and 2 are sugars derived from shrimps, and are composed of the heparin disaccharide repeat with the only difference being the size of the molecule in comparison to UFH. They also contain the 3O sulfate described as been responsible for AT anti Xa activity.

5.2. Results.

5.2.1. SRCD spectra and analysis.

5.2.1.1. The effects of inactive heparin containing 3-O-sulfation on the structure of AT. Comparison to UFH and highly active synthetic pentasaccharide.

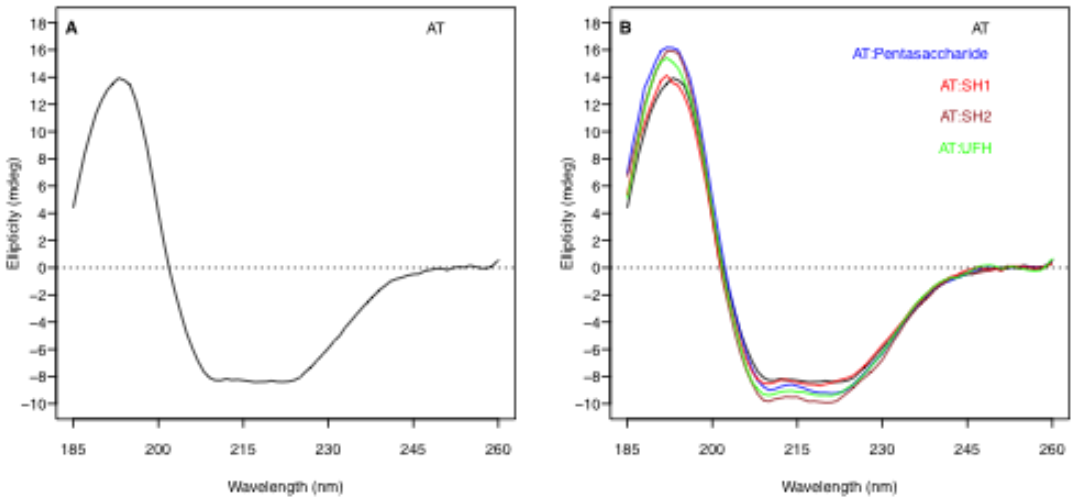

Figure 5- 2: A: SRCD spectra (185 – 260 nm) of native AT (0.4 mg/ml 12.5 mM NaCl PBS buffer, 0.1 mm path length, 0.5 mm slit and average of 4 scans). B: SRCD spectra of AT with various GAG derivatives (0.8 mg/ml), under the same conditions as A.

Table 5- 1: Secondary structural analysis of the spectra provided in Figure 5-2. Analysis conducted using the CDNN software, with complex base set reference proteins (version 2.1 [98]).

Sample	AT	AT:pentasaccharide	AT:UFH	AT:SH1	AT:SH2
Helix	80.70%	85.90%	85.80%	80.90%	87.30%
Antiparallel	0.10%	0.10%	0.10%	0.10%	0.10%
Parallel	2.20%	1.80%	1.70%	2.10%	1.50%
Beta-Turn	9.80%	8.70%	9.10%	9.90%	8.80%
Random. Coil	8.00%	6.30%	5.30%	7.30%	4.80%
Total Sum	100.80%	102.80%	102.00%	100.30%	102.5%

The complexes were further investigated via UV degradation over the course of 35 scans, and their secondary structures were tracked to determine how the degradation pathways differed when complexed with the various GAGs, as a means of differentiating otherwise quite similar CD spectra.

5.2.1.2. Comparing and differentiating AT:GAG complexes.

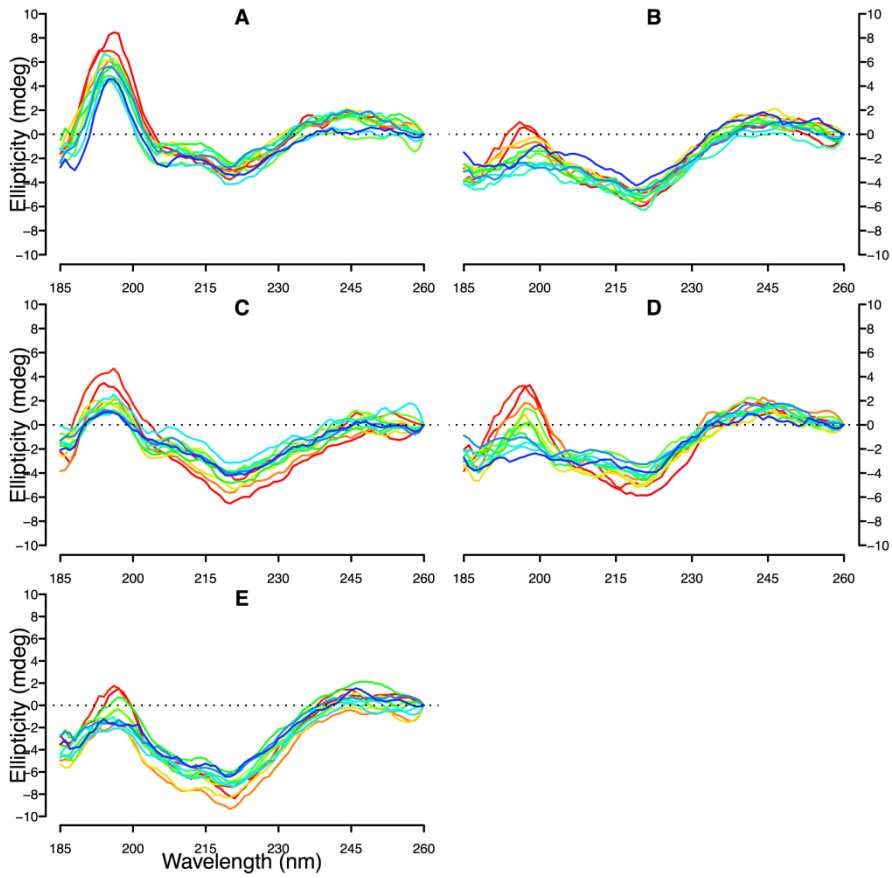

Figure 5- 3: SRCD spectra (185 – 260 nm) of native AT over a 35 scan degradation (0.4 mg/ml 12.5 mM NaCl PBS buffer, 0.1 mm path length, 0.5 mm slit. Every 3rd scan is shown, red to blue) alone and with various GAG oligosaccharides (0.8 mg/ml). The panels represent the following samples **A**, AT; **B**, AT:pentasaccharide; **C**, AT:SH1; **D**, AT:SH2 and **E**, AT:UFH.

To allow for a direct comparison between the unfolding pathways the data from Figure 5-3 above were analysed using PCA. The initial data analysis showed that the first three components accounted for the major contributions to the observed variation in the samples (22.5%, 12.3% and 9.4% respectively, 44.2% cumulatively)

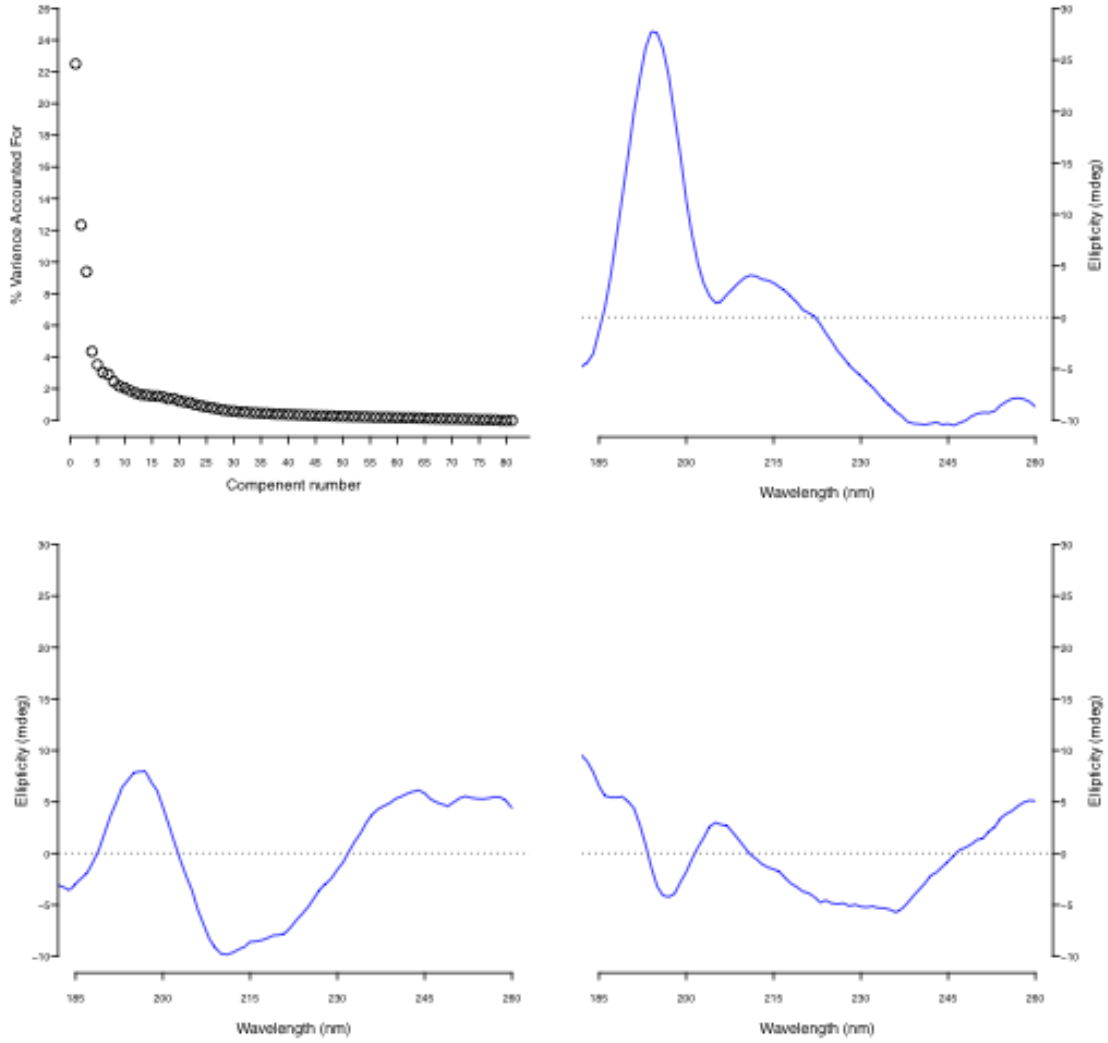

Figure 5- 4: A Top left is the scree plot for the amount of variance accounted for by each component to explaining the variance from all of the spectra shown in Figure 5-2. Top Right, component 1, bottom left, component 2, bottom right, component 3. The components 1, 2 and 3 explain 22.5%, 12.3% and 9.4% of the overall variance observed in the raw data in Figure 5-3 respectively.

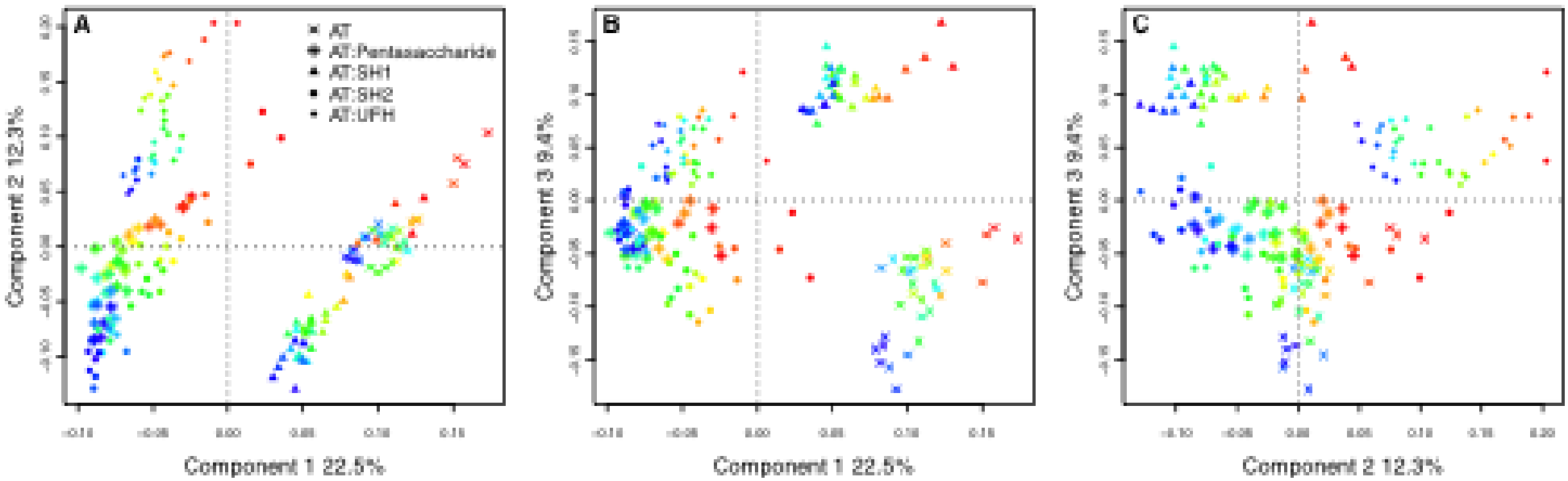

Figure 5- 5: PCA of AT with various GAG oligosaccharides showing the degradation process over the course of 35 scans (blue to red) in the far UV (180-260 nm). A: shows component 1 *v* component 2, B: is component 1 *v* 3 and C: is component 2 *v* 3. The components 1, 2 and 3 explain 22.5%, 12.3% and 9.4% of the overall variance observed in the raw data in Figure 5-3.

5.2.2. DSF and activity (against factor Xa) results.

5.2.2.1. Stabilisation, not conformation, determines activity.

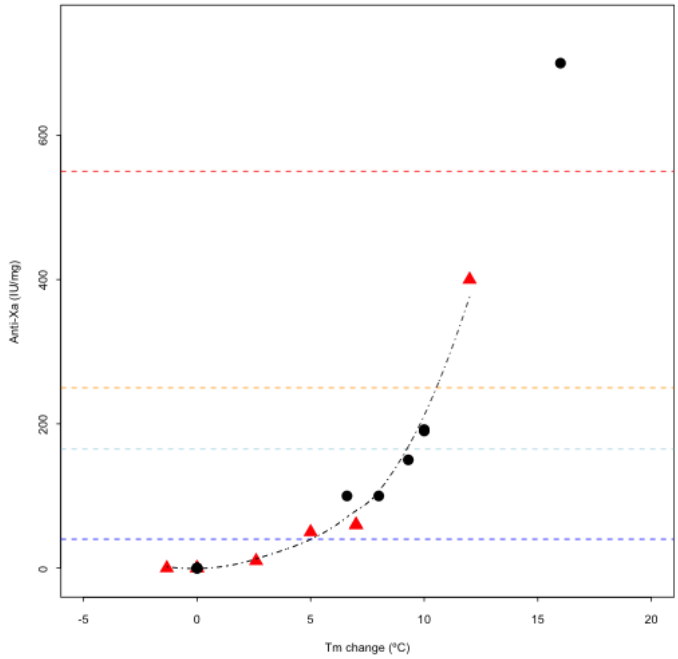

Figure 5- 6: Relationship between anticoagulant activity (anti-factor Xa) and the melting temperature (Tm, °C) of native AT in the presence of 17 polysaccharide derivatives (5 plant and 12 GAG saccharides). Activity values are either measured in the laboratory (by collaborators, triangles) or by using manufacturers values (circle). The dashed line is a line of best fit, with six values coinciding at (0,0) excluding the pentasaccharide in the upper right corner due to it being the only pure, non-heterogeneous, compound. Reproduced from [141].

Table 5- 2: The stability (°C) of AT alone and in the presence of various polysaccharide ligands as measured by DSF. Also listed is the anti Xa activity of these AT:ligand complexes measured experimentally.

Sample	Melting Temperature (°C)	Anti-Xa Activity (IU/mg)
AT	49.5 (±2.0)	-
AT:Pentasaccharide	61.9 (±0.8)	700
AT:UFH	57.1 (±0.5)	190
AT:SH1	54.5 (±1.1)	≤10
AT:SH2	53.8 (±1.0)	≤10

5.3. Discussion.

5.3.1. Justification for using freshly purified human AT.

AT was purified fresh from plasma using a salt gradient eluting from heparin affinity beads. This technique was used because native AT from plasma was required, allowing the protein to be kept in solution and at constant pH throughout all experiments. The treatment of AT purchased from commercial sources involves pasteurisation and lyophilisation. This has been shown to result in a protein that, while maintaining activity, possesses different secondary structure and also altered thermal stability measured using DSF [83,84] (data not shown and conducted by Ms N. Veraldi, Ronzoni Institute, Milan). The native AT purified from human plasma also maintained all complex glycosylation, including the rarer glycosylated form where the glycosylation site, Asn 135, which arises in 5-10% of cases [211,212]. The SRCD experiments were conducted in 12.5 mM PBS, so allowing the experimental conditions to approach the biological state without compromising the high quality structural data by excessive UV absorption caused by chloride ions (salt concentrations about 10 x less than physiological [213], and AT at 0.4 mg/ml; 3 x lower than the physiological levels [214]).

5.3.2. Discussion of SRCD experiments: Structural similarities induced in AT between active and non-active compounds.

The initial experiment, recording the SRCD spectrum of AT alone in solution agrees with previous work that states AT consists mainly of α-helical protein with some β-

sheet (Figure 5-2A, Table 5-1). When the spectrum was recorded in the presence of pentasaccharide (Figure 5-2B, blue line) there was an expected change in the increase of alpha helix as the reported reactive centre loop is expelled and converted from β-sheet to α-helix [215] (Table 5-1). It has been observed that this change resulted in a 5.2% increase in α-helical content, at the expense of random coil, parallel β-sheet and β-turns (Table 5-1). The average structural changes induced in AT by UFH were similar which, being an active compound, was expected. The inactive compound shrimp heparin 1 (SH1, Figure 5-2B, brown line) showed some evidence of interaction with AT, resulting in a small conformational change, slightly increasing in α-helical content (Table 5-1). The SRCD evidence confirms the interaction as SRCD is known to be extremely sensitive to detecting small changes in protein structure [90]. It must be considered that the purification process of the SH compounds included an AT affinity step, indicating that, despite the small conformational changes observed in SRCD these compounds have been purified on the basis of interactions with AT. Further evidence of the interaction was provided in the form of DSF measurements showing a subtle stabilisation of AT (from 49.5°C ±2.0, to 53.8 °C ±1.0). The observation presented below, however, led to the previous hypothesis, that these inactive shrimp heparins would induce distinct secondary structures, being rejected and a new explanation of the significance of structural changes induced in AT by heparin derivatives being sought.

The compound labelled SH2 induced remarkably similar secondary structural changes on average to that of pentassacharide and UFH. This compound resulted in a 6.6% rise in α-helical content, compared to 5.2% with pentasaccharide. However,

since the measured anticoagulant activity of SH2 is below 10 IU/mg of activity, this result was unexpected.

5.3.3. Differentiation of similar complexes by UV degradation.

It was then decided that an alternative method was required to differentiate these complexes so a UV degradation experiment was conducted in order to track their unfolding kinetics, the results of which are shown in Figure 5-3. These results do show some differences between the complexes, especially around the positive feature at 190-200 nm and the 220 nm negative feature. These data were collated into a single matrix and PCA was performed on the spectra to allow a direct comparison of the changes that occur during the perturbation. These changes are explained in terms of independent components that together account for 100% of the variance observed from the mean "spectrum" of all the spectra input into the matrix (created via the mean centring of the data). These components are weighted in terms of how much variance each explains and can be used to follow directly the process occurring during a stepwise structural change such as this. This example of PCA first indicates that there are many complex processes occurring simultaneously, evinced by the low weighting apportioned to even the first component, 22.5% *v* 56 and 54.5 % observed for component 1 in Chapter 3 [142]. There are 3 major components within this data set, shown in the scree plot (Figure 5-4A). These are shown in panels Figure 5-4B, C and D for components 1, 2 and 3 respectively. They are weighted as explaining 22.5, 12.3, and 9.4% of the total variance, collectively covering 44.2% of the total variance observed in the raw data in Figure 5-3. A structural assessment of what each component represents cannot be attempted. This is largely caused by the fact that

during this analysis, if a certain structure were lost from the protein in solution then the resulting component would be the inverse of the typical spectral shape of that structure. These details would be missed by conventional structural analysis. This assumption would also render conventional structural analysis of the raw data from Figure 5-3 at different time points inaccurate, because the state of the protein is not known at each point, regarding protein length especially since high energy UV irradiation could break the protein backbone, so the input parameters regarding protein size and therefore molar concentration would be inaccurate.

Therefore, a more qualitative, rather than quantitative approach was taken to understanding the changes occurring within the spectra. Using this approach it can be stated that within component one (Figure 5-4), the major positive feature present at ~195 nm, which is in line with the addition of β-sheet as stated by Table 1-3, Chapter 1 (which states β-sheet has a positive sign focused at 195 nm) and the positive feature around 215 nm could be representative of an increase in β-turns. The second component is dominated by a classical α-helix structure, with a positive feature at ~195 nm being lower than that expected of an entirely helical spectrum suggesting a negative influence around this point. This again could be from increased levels of β-turn, or possible unordered structure. The third component is rather more complex, possibly comprising either a mixture of different defined structures, or some new undefined structure induced the high intensity UV. There are many different structures, defined broadly as "unordered", that have distinct CD profiles (see Figure 5-4 of the review by Wallace and Janes 2010 [90]). These qualitative assessments must be taken as best estimate interpretations given the available data thus far, and it

must be remembered (evinced by the low weighting of components) that this process is a complex one.

The PCA analysis shows that, broadly, all five complexes undergo similar unfolding processes demonstrated in Figure 5-5A with comparable increases of both component 1 and 2 during the denaturation process. This indicated the addition of alpha helix and a reduction of unordered structure. The major difference between the samples when taking these two components into consideration is the extents to which they denature during the 35 scans. It can be said that AT unfolds in an independent manner to the AT:polysacchaide complexes, shown by its initial period of resistance to UV stress. The complexes, in contrast, undergo a continuous step-wise unfolding with no initial resistance. This panel (Figure 5-5A), representing 34.8% of the total variance, again suggests that the interactions between these complexes are in fact very similar. When observing the more subtle changes that are accounted for by the (difficult to define in secondary structural terms) component 3 (Figures 5-5B and 5-5C), differences in the unfolding begin to become apparent. It is observable again, that AT alone undergoes a different unfolding process to that of AT complexed with GAGs. This is demonstrated by the lack of any change in either of the major components 1 and 2 during the early stages of the process, but does reveal that some more subtle changes are occurring early in the process accounted for by component 3. This then stops being a contributing factor roughly half way through the process employed here, suggesting a two stage unfolding event for AT alone when subjected to UV irradiation. Component 3 generally has only a minor influence on the unfolding process of AT when complexed, but, by the selective removal of component 1 and 2, it is possible to observe each complex individually showing that,

despite being similar, they are distinct. This is except for AT:pentasaccharide and AT:SH2, which have some major overlaps in their common structures and their unfolding pathways for all combinations of components 1, 2 and 3. Their only difference is the extent to which each component is changed in each case, not the direction or positions within each combination of components. The result is that these complexes share similar structural interactions, especially AT:Pentasacchaide and AT:SH2. This is contrary to the original hypothesis considering the large differences of activity induced in AT against factor Xa. The data, suggesting differential stabilities as measured by the intense denaturation process by far UV irradiation, were investigated further. To do this, initially, these complexes were investigated using DSF and a thermal denaturing technique. It was observed that the activity of these compounds were closely linked to the thermal stability induced in AT when complexed together (Table 5-2). This trend was then further investigated with a collection of GAGs and plant polysaccharides, both sulfated and non-sulfated. This collection of polysaccharides were selected to investigate the independence of the stability from a particular sequence, or structure in the determination of activity. They were used to demonstrate that AT activity does not require a particular GAG sequence, or even a GAG. The data were taken from reference [141], but these compounds will not be discussed within this thesis. The thermal stability data were then plotted against either their reported (by manufacturer), or measured, anti factor Xa activity (Figure 5-6, Table 5-2). The data presented provide strong support of the revised hypothesis; namely, that the ability of a ligand to stabilise AT when complexed determines the level of activity and not the induced secondary structure changes.

The analysis also raises some important questions regarding the mechanism by which pentasaccharide interacts with AT. This product has until now been described as a synthetic mimic of a natural fraction of heparin that binds selectively to AT and causes this conformational change in the same manner as the natural ligand. This will form part of the investigation for Chapter 6, looking at the degradation data in more detail through generalised 2D spectroscopy.

5.3.4. AT activity is directly related to complex stability and not solely to a sulfation sequence.

These results suggest many questions regarding the core assumptions that have been made around the basic mechanism behind AT activation. The results presented here do not disagree with the structural work done previously regarding the expulsion of the reactive loop when bound to active heparin fragments, however, they do call into question the assertion that the 3-O-sulfate is the marker for activity and indeed that the induced secondary structural changes are the underlying factor behind AT activation. The sarccharides derived from shrimp used in this work contain the 3-O-sulfate group and, and with the exception of the size of the molecule, are very similar to heparin in composition and basic structure sharing an identical dicaccharide repeat unit and linkages. Beyond a minimal level size, as noted in the literature [188], the chain length is not a factor in anti Xa activity with AT, unlike with thrombin activity, so these shrimp heparinoids can be thought of as ‘standard’ heparin molecules, such as the UFH used here, which is usually derived from mammalian sauces [210,216]. These heparinoids have shown during purification to have high affinity for AT, eluting in the same fraction as HA heparin. These collective results open up new

avenues for anticoagulant drug design, including routes for small molecule activation of AT, should a molecule be found that can stabilise AT in a similar manner to the results reported here. The results also lend themselves to DSF being used as a medium to high throughput assay with which to screen potential ligands for AT activity before the more time consuming and costly anti Xa activity assays take place. It also raises the question of the mechanism by which the presumed natural ligand HS activates this enzyme in the blood and a perhaps more important question regarding our understanding of the mechanism of action of the current drugs. It is possible that there are many more factors that influence the activity and the resulting clinical response than those considered so far, especially as our understanding of its natural source material is still being expanded. This is not to say that the current drugs available in the form of the synthesised pentasaccharide or range of LMWHs are bad or ineffective *per se*, rather that it demonstrates a lack of real understanding regarding their mechanism when admistered to patients, which is an important point to realise and if possible address. Doing so would be of vital importance in improving these drugs and in the development of novel compounds in which it might be desirable to altogether avoid these activities. The idea of heparin being a redundant system, and therefore lacking obvious structure (sulfation sequence) to activity relationship, is not a novel one. It has been observed for instance in FGFs that the stabilisation of the protein is required for its activity, with no clear correlation with sulfation patterns within the molecules [46,50,139].

5.3.5. Future work, the creation of active compounds from non-active heparin fractions.

This new AT hypothesis, that the stabilisation directly correlates with its activity, can be tested via covalent cross linking of inactive, or partially active, heparins with the protein. This should induce the required stability and a preliminary experiment was carried out using EDC (1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide) a zero-size crosslinking agent, and crosslinking protocols previously optimised. The crosslinking of UFH and non-affinity (NA) to AT, and the gel filtration purified crosslinked product were measured by DSF. The result indicated a change of the AT:UFH complex from its original stability to a stability close to that of pentasacchaide (See Figure 5-7), its activity has yet to be assessed.

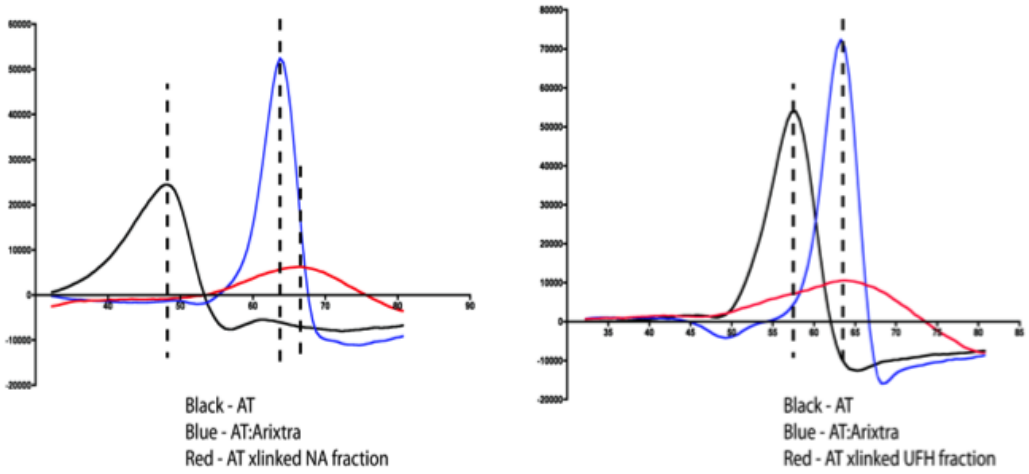

Figure 5-7: Preliminary 1st derivative curves from DSF results from the covalent crosslinking of inactive (NA, red line left panel) and partially active (UFH, red line, right panel) heparin compounds to AT showing a comparable stabilisation to that of the active, non-crosslinked, pentasaccharide (Arixtra, blue lines). Y-axis represents arbitrary values from 1st derivative measurement and X-axis represents temperature in °C.

Alongside these EDC crosslinking experiments several others could be proposed to test these hypotheses, including selectively labelling AT in the heparin binding site and probing with NMR to discover whether these compounds bind in the same site or have various binding sites and modes across the protein. The crosslinking data provide a novel route for drug design should these compounds prove to be effective in activating AT against factor Xa for instance, heparin is open to chemical modification, for example through the addition of thiol groups, which are very reactive, opening up many potential additional groups that may be added. This could be used to localise the heparin molecule and provide to localised anticoagulant activity.

Considering the direction the field has taken since the original discovery of the pentasacchaide that led to the hypothesised structure (sulfation pattern) function relationship, then these results add to a growing body of evidence [36,45,46,139] that suggests this is incorrect. In the proceeding ~30 years much effort has been expended trying to find more examples of these relationships to no avail. It is therefore becoming apparent that an alteration of mind set may be required to advance the field, and a shift away from the template driven genetics style to one of a more redundant and flexible system that is capable of affecting any responses to a change in physiological needs via subtle alterations to the GAGs.

5.4. Author contributions.

This chapter has been a large collaborative project between A Hughes at Liverpool University/Diamond Light Source and Dr M Lima of UNIFESP, São Paulo, Brazil. As such contributions are listed below;

AT collection from Human plasma: Dr M Lima.

SRCD experimental design: A Hughes

SRCD experimental data collection: A Hughes / Dr M Lima

SRCD analysis: A Hughes

DSF experimental planning, data analysis and collection: A Hughes (Excepting Figure 5-6)

EDC crosslinking optimisation: A Hughes / Dr M Lima

AT:GAG cross linking for Figure 5-6: Dr M Lima.

Chapter concept and direction: A Hughes / Dr M Lima.

Chapter 6: An investigation into the application of covariance matrix data analysis during a perturbation.

6.1. Introduction.

6.1.1. Covariance matrices and the basic method.

This chapter will aim to investigate the suitability of applying covariance matrices as a data analysis tool for increasing the resolution and information available in a technique that is low resoution in nature. It will focus initially on the 35 scan AT UV degradation shown in Chapter 5, and then will be selectively applied to the 60 scan degradation of HEWL investigated in Chapter 3.

Tracking a perturbation, such as a UV degradation, via 2D correlation of the data matrix can be a useful method for distinguishing complexes with similar CD spectra, in that they may degrade by distinct routes, giving rise to different CD spectra. Two examples of this type of perturbation will now be analysed using a number of variations of the methods first suggested by Noda [217]. This process of matrix manipulation can be employed to probe both synchronous and asynchronous correlations within a data set, and was first developed by Noda *et al*. These two types of covariance matrix in their simplest form can give the user different information, and these will be described below.

The synchronous covariance matrix reveals those wavelengths that change in phase with the perturbation, while the asynchronous covariance matrix selects those that

change out of phase during the perturbation. This means that the synchronous covariance matrix highlights correlations that are happening simultaneously (in phase) with the progression of the perturbation being studied such as temperature, concentration or time series for example. In this case the perturbation is the scan number – the amount of UV exposure. The Asynchronous covariance matrix highlights those changes that occur sequentially to one another, but not coincidentally.

The covariance matrix, for both synchronous and asynchronous correlations, is derived from the multiplication of a mean centered matrix [2] by the transpose of that matrix (Equation 6.1 and 6.2 below, 6.1 – for synchronous covariance and 6.2 for asynchronous, HN – Hilbert-Noda matrix), thus multiplying every cell within the matrix by every other one and thereby creating a covariance matrix.

$$6.1: M*M^T$$

$$6.2: M*HN*M^T$$

The Hilbert-Noda transformation matrix is designed to move changes out of phase with one another, and does so by multiplying M by a number (derived from the term 6.3 below) before its multiplication with M^T. This also assigns all diagonal peaks as '0' explaining the reason why these matrices feature no diagonal peaks. The details of these are explain in detail in Noda and Ozaki's book [217]

$$6.3: (1/\pi(k - j))$$

In term 6.3 the values on the Hilbert-Noda transformation matrix are calculated. In this equation k refers to the column number and j the row number.

The resulting covariance matrices (from either Equation 1 or 2) are then plotted by X Y and Z (as colour to represent the degree of correlation). Initially, this approach will be used and discussed. Alternative methods will then be explored as possible routes to improving resolution, and allowing a more direct comparison between two perturbations. The more standard method introduced above covers Figures 6.1-6.4 and is described as method 1 in Table 6-1, while later figures refer to alternative methods, of which there are four, described below.

6.1.2. Alternative covariance matrices.

The alternative methods developed for this thesis consist of four approaches. First an attempt was made to observe how each covariance matrix differs from other covariance matrix. This was achieved through the deduction of one covariance matrix from the covariance matrix to be compared to. Thereby revealing the differences between the two perturbations. This is method 2 in table 6-1.

The third approach, attempts to correlate two perturbations together using an alternative mean centering (MC). This method has an initial MC step across the two perturbations (A + B) before perturbations A and B are then correlated independently of each other. This method of handling the data removes the features in common between A and B, and should highlight features that differ from each other. The usual MC method for correlations is for the MC to be through the single perturbation

being studied. For an example of the alternative method proposed, if comparing AT to AT:pentasaccharide, the usual process would be to mean centre the AT spectrum and separately mean centre the AT:pentasaccharide spectrum, but in this situation a single mean centre is created across both perturbations, which is then deducted from each and then the two correlation matrices are created, as summarised below;

- Matrix A
- Matrix B
- A is combined with B to give the matrix AB
- This is then mean centered AB(MC:AB)
- This is then separated into each component again as the A and B dimensions are unchanged: A(MC:AB), B(MC:AB)
- Then each matrix (A / B) is correlated in the usual fashion presented in Equations 6.1 and 6.2

The fourth method is similar to the first, however, in this case, one of matrices is prenormalised to the maximum amplitude of the other matrix, thus eliminating amplitude as a contributing factor to the resulting comparison. This method of comparison will lose information regarding the extent to which changes happen at a particular wavelength, but will provide a better representation of the wavelengths at which the correlations differ in each given situation.

A final (fifth) analysis is conducted across all AT degradation spectra in the same manner as method one. In this instance only the wavelengths 185-250 nm are considered. This is to investigate whether having equal variance at each position

across the spectrum affects the final result. The increased variance at 260 nm is discussed in Chapter 7.

As this is a demonstration of an analysis technique and not a full analysis of the data at hand, examples from each method of analysis will be analysed for AT perturbations and from the methods giving the best results in the AT example will be applied to the HEWL perturbations. Methods are summarised in Table 6-1.

Table 6- 1: Summary of the methods utilised with results section.

Method number	Description
1	Basic covariance matrix of perturbation MC across single perturbation
2	Deduction of two covariance matricies created using method 1 from one another
3	Alternative MC technique. Individual covariance matrices are created from data sets mean centered across two perturbations to be compared.
4	As method 2, except the covariance matrices are normalised prior to deduction from one another.
5	A repeat of method 1, but excluding contributions above 245 nm from the creation of the covariance matrix.

6.2. Results.

6.2.1. AT perturbation analysed by method 1.

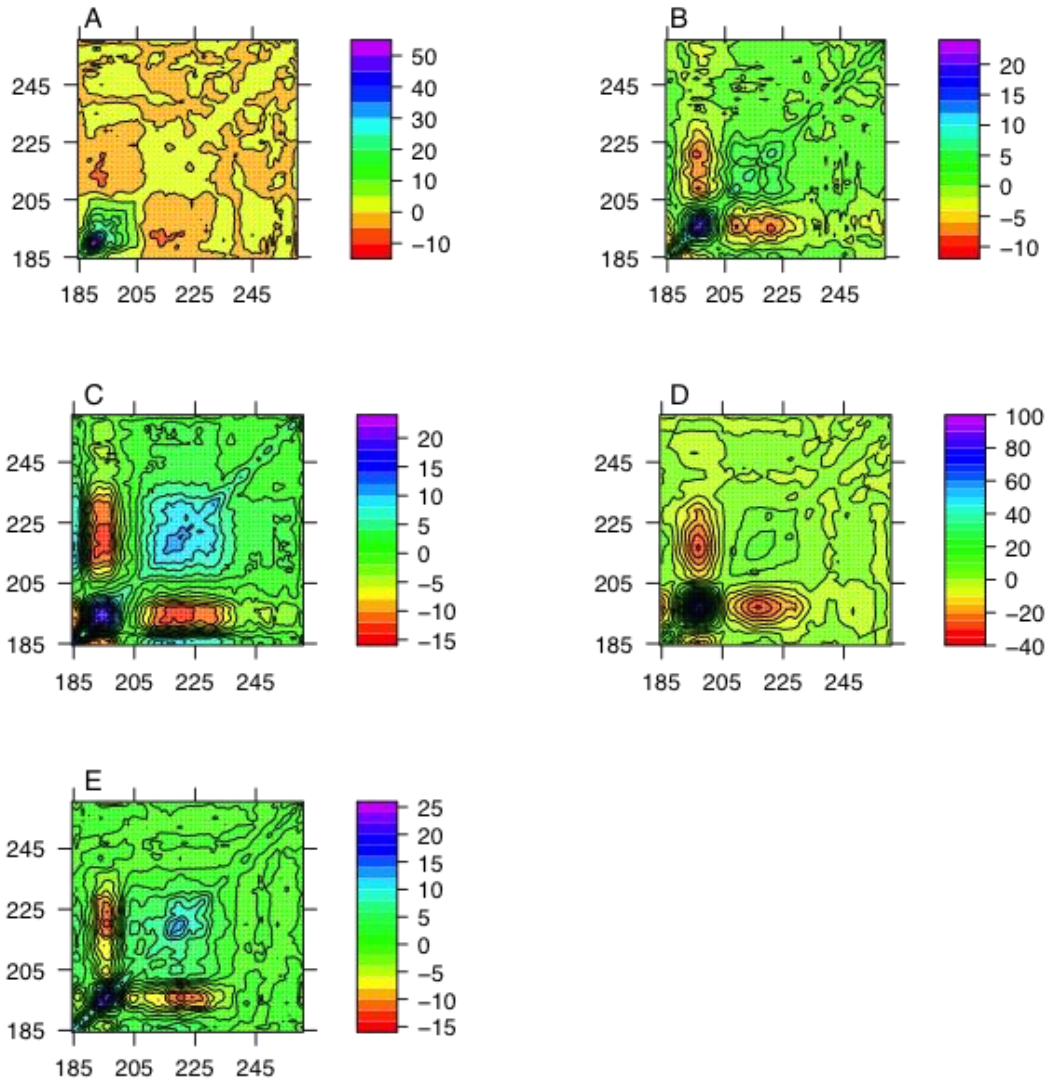

Figure 6- 1: A summary of the synchronous covariance matrices of the 35 scan perturbation of AT alone (A) and in the presence of the ligands pentasaccharide (B), SH1 (C), SH2(D), and UFH (E). As shown in Figure 5-2.

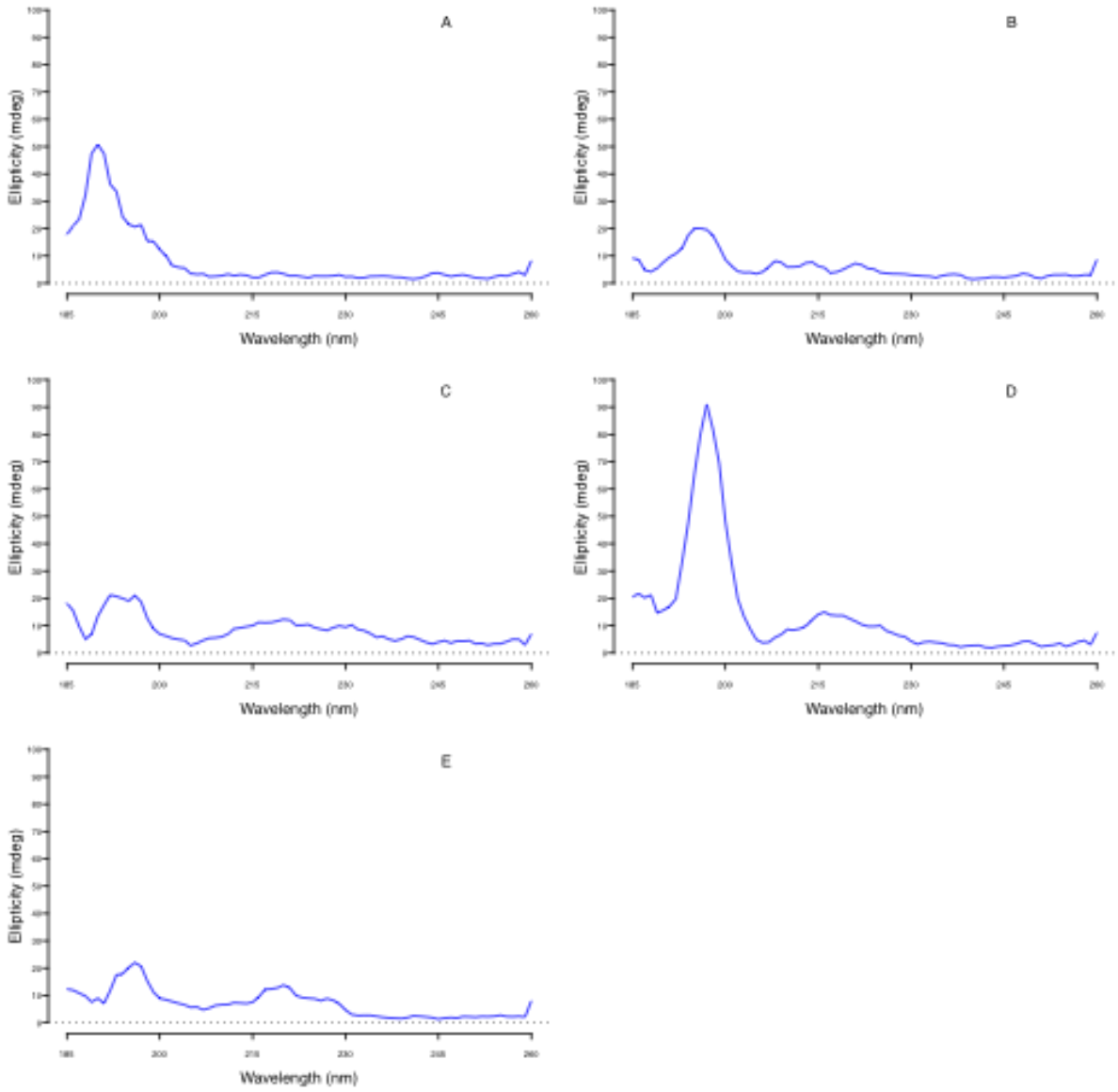

Figure 6- 2: The diagonals of the synchronous covariance plots shown in Figure 6-1. This demonstrates variance within the dataset for each complex. AT alone (A) and in the presence of the ligands pentasaccharide (B), SH1 (C), SH2(D), and UFH (E).

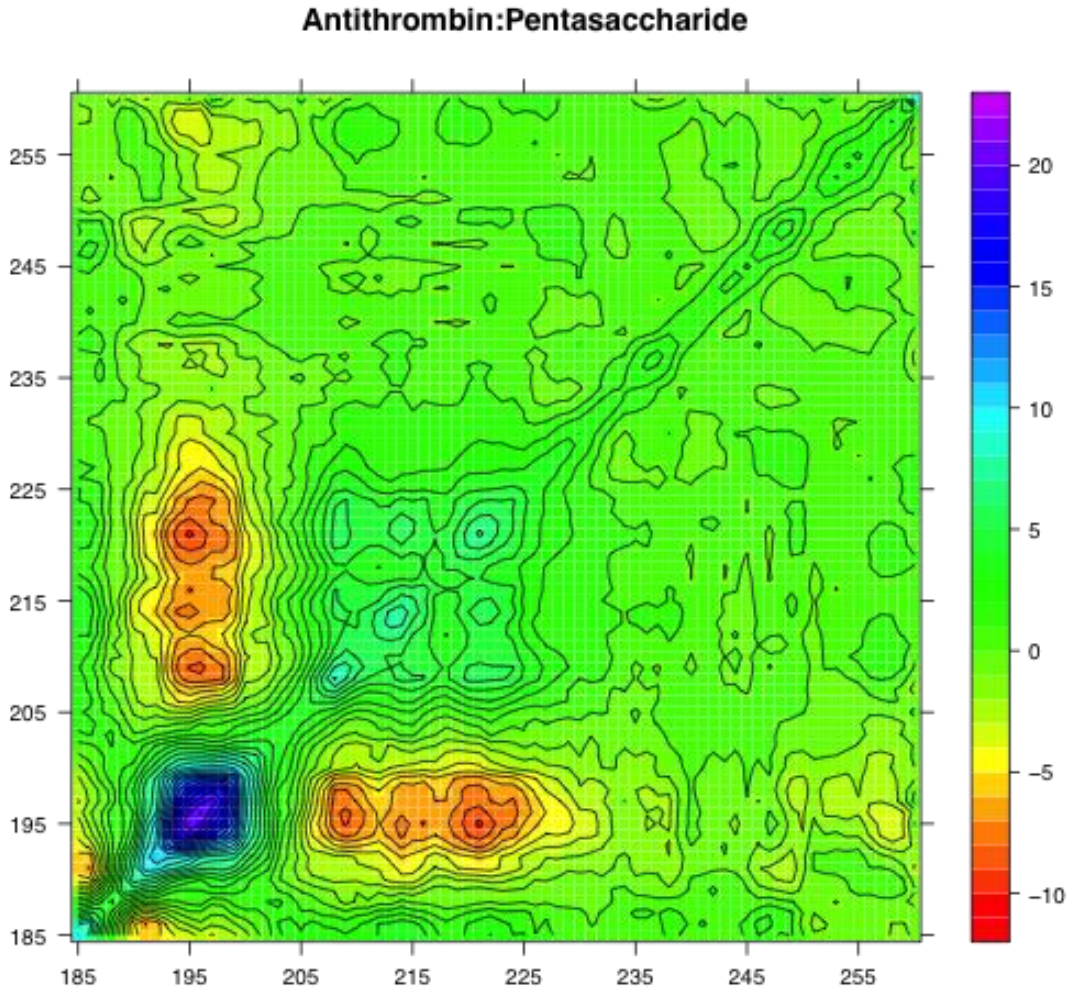

Figure 6- 3: Synchronous covariance matrix plot of AT in the presence of pentasaccharide undergoing a 35 scan UV degradation perturbation represented in Chapter 5 Figure 5-2B. This is a more detailed expansion of Figure 6-1B.

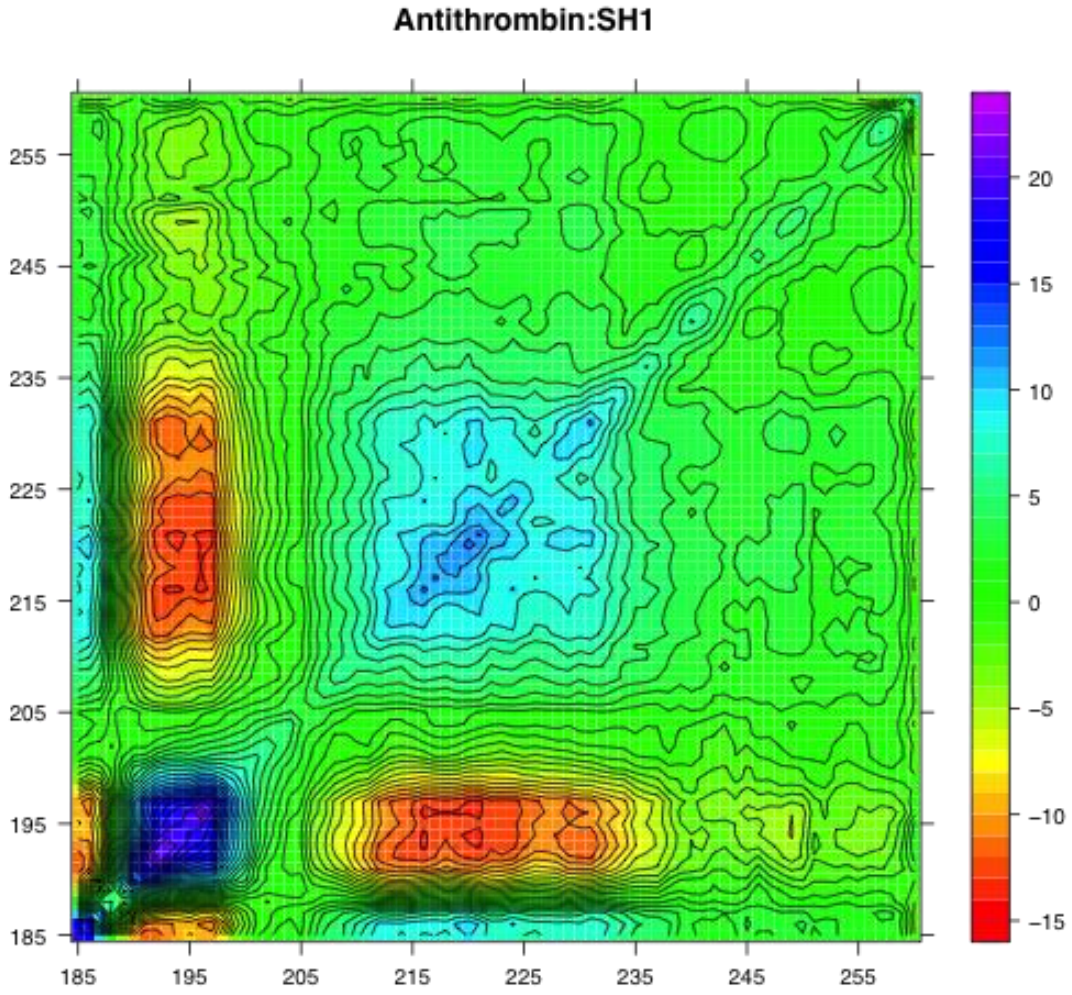

Figure 6- 4: Synchronous covariance matrix plot of AT in the presence of SH1 undergoing the 35 scan UV degradation perturbation represented in Chapter 5 Figure 5-2C. This is a more detailed expansion of Figure 6-1C.

6.2.2. Method 2. Deduction of one AT covariance matrix from another.

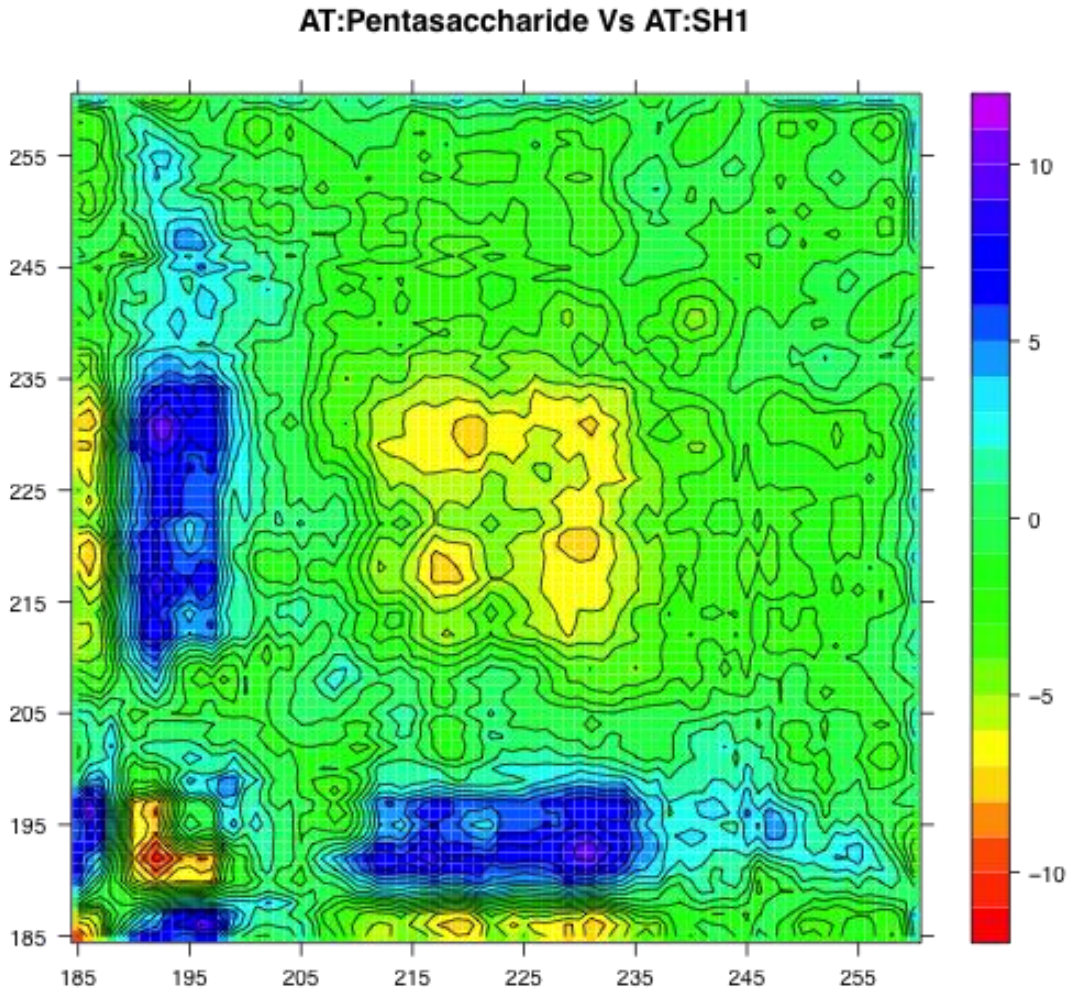

Figure 6- 5 AT:pentasaccharide covariance matrix (Figure 6-3) minus AT:SH1 (Figure 6-4) covariance matrix, revealing differences in correlated areas during 35 scan UV degradation experiment.

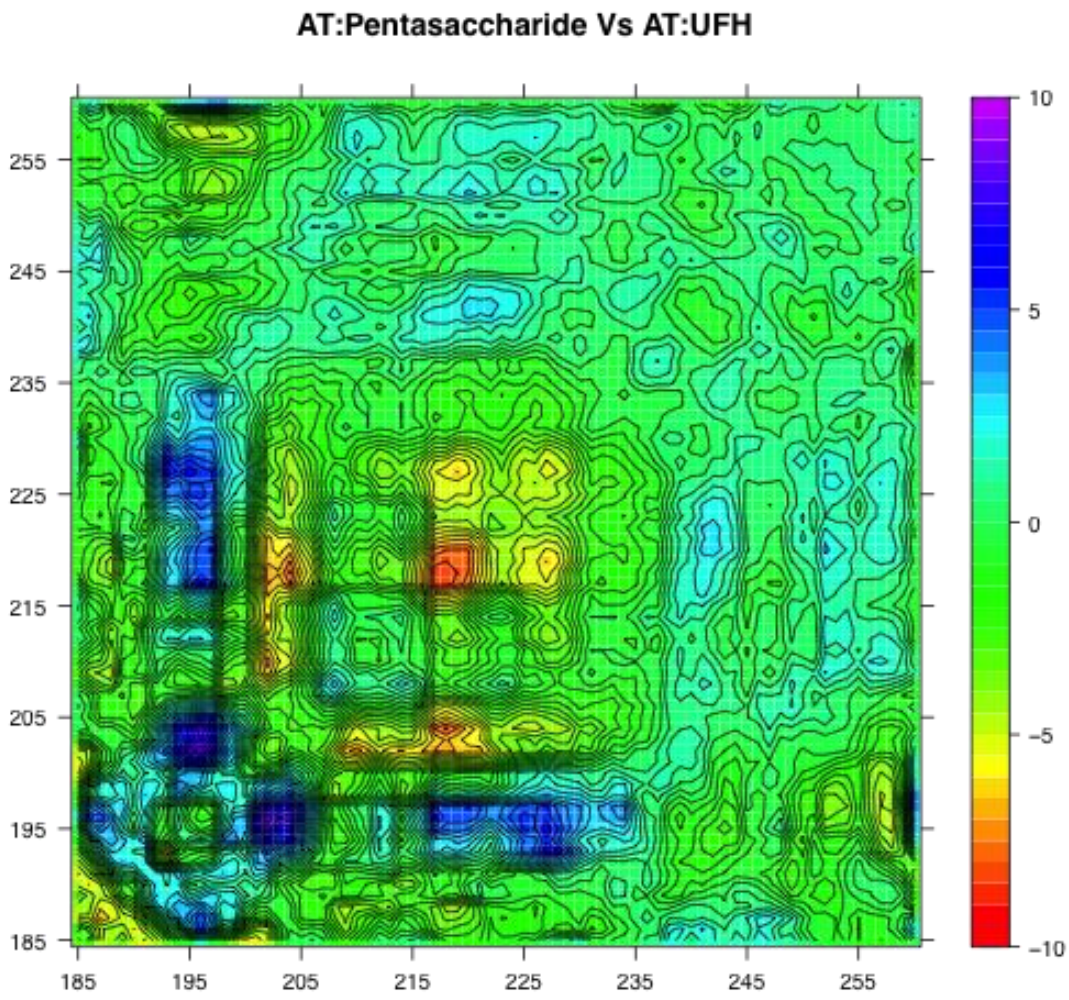

Figure 6-6: AT:pentasaccharide correlation matrix (Figure 6-3) minus AT:UFH covariance matrix (not presented), revealing differences in correlated areas during the 35 scan UV degradation.

6.2.3. Method 3. Mean centring across two AT perturbations to be compared.

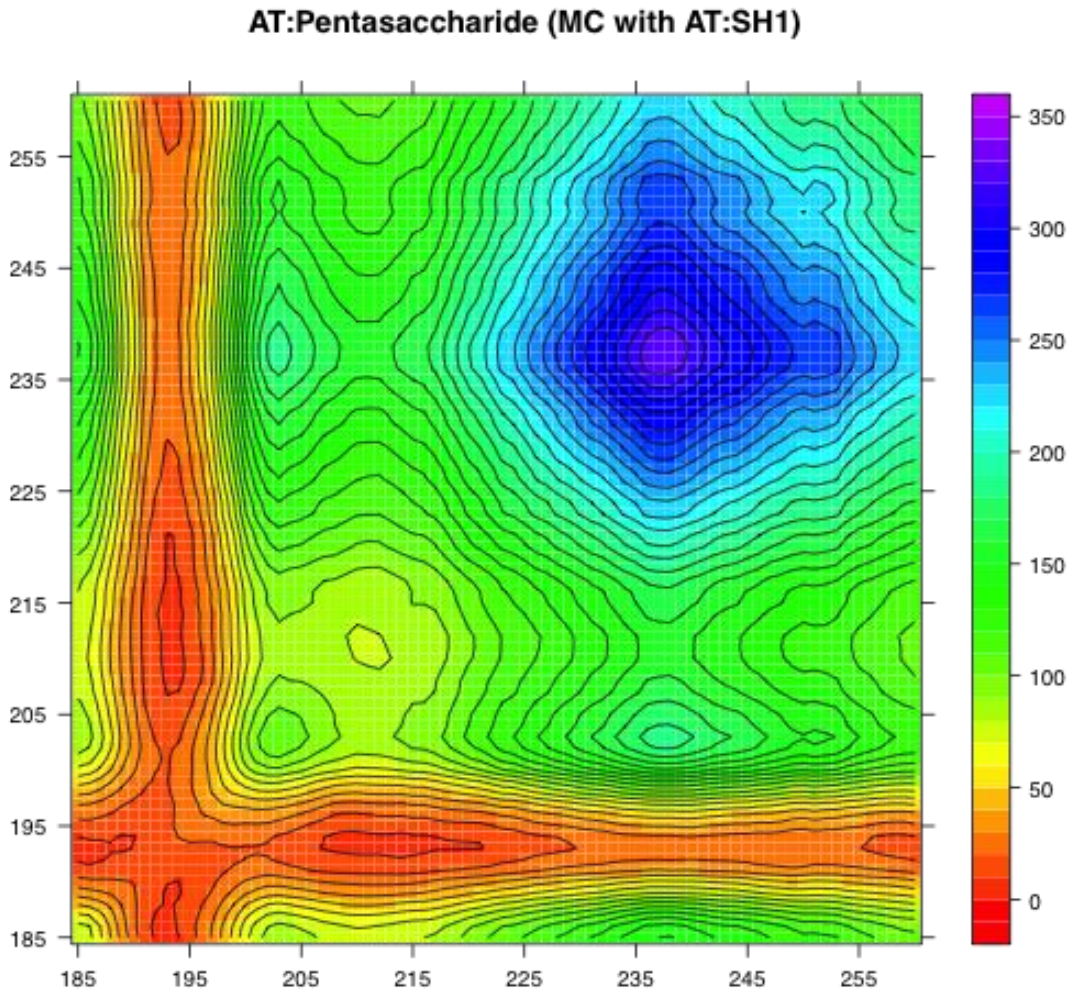

Figure 6- 7 AT:pentasaccharide synchronous covariance matrix when mean centered in combination with the AT:SH1 perturbation. Perturbations are over 35 scans of UV degradation.

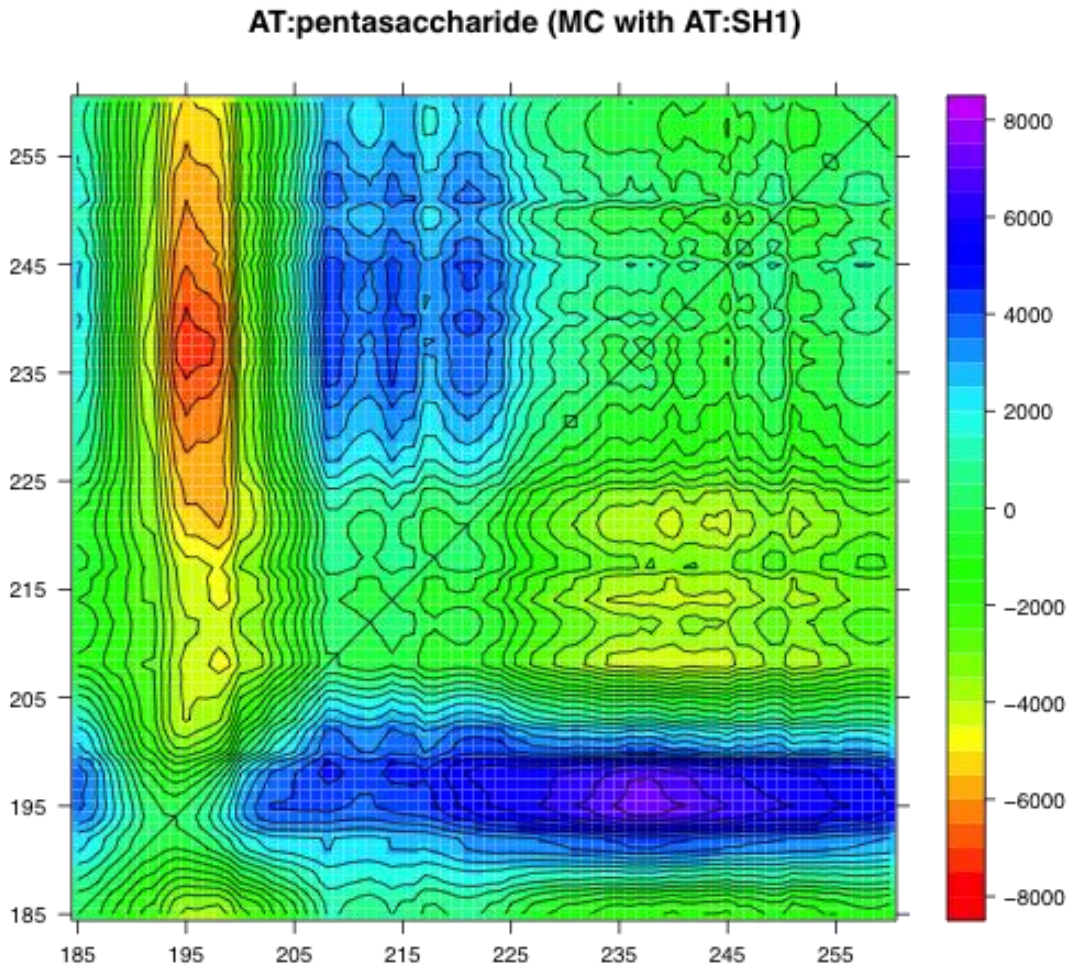

Figure 6- 8: AT:pentasaccharide asynchronous covariance matrix when mean centered in combination with the AT:SH1 perturbation. Perturbations over a 35 UV degradation experiment

6.2.4. Method 4. Comparison of the differences in two AT covariance matricies deducted from one another after prenomalisation.

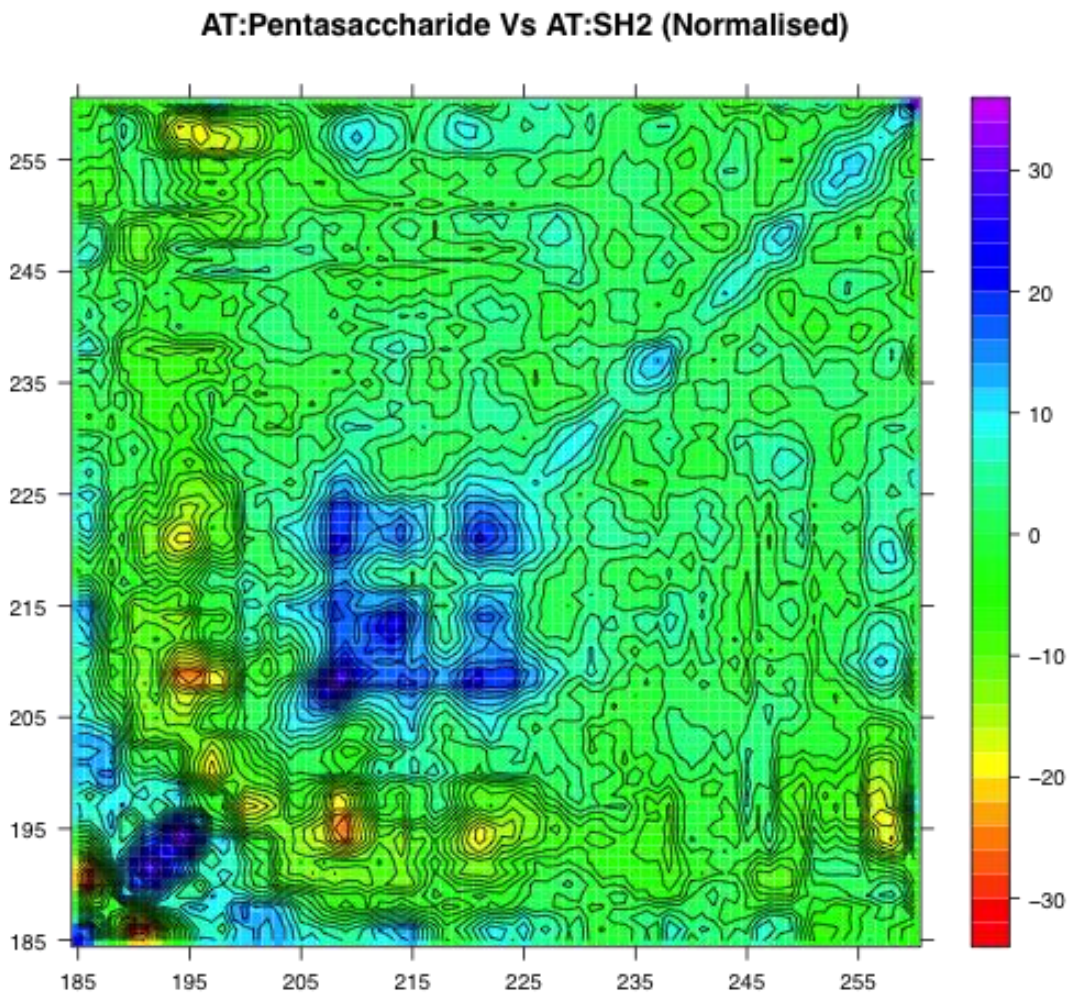

Figure 6- 9: Synchronous covariance matrix of AT:pentasacharide (prenormalised to the maximal signal from AH:SH2) with AT:SH2 deducted.

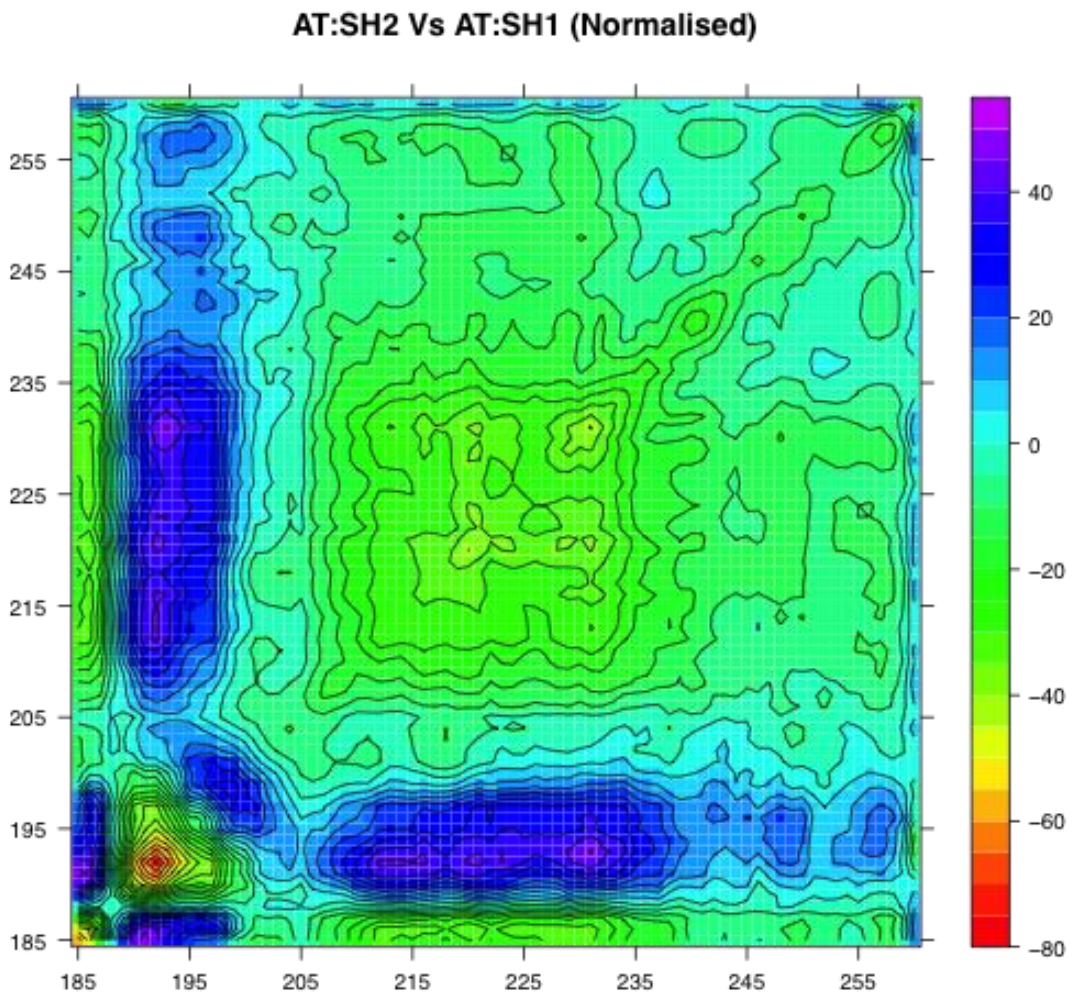

Figure 6- 10: Synchronous covariance matrix of AT:SH2 minus AT:SH1 covariance matrix (prenormalised to the maximal signal from AH:SH2).

6.2.5. Method 5. Checking for any artefacts stemming from added noise at 260 nm.

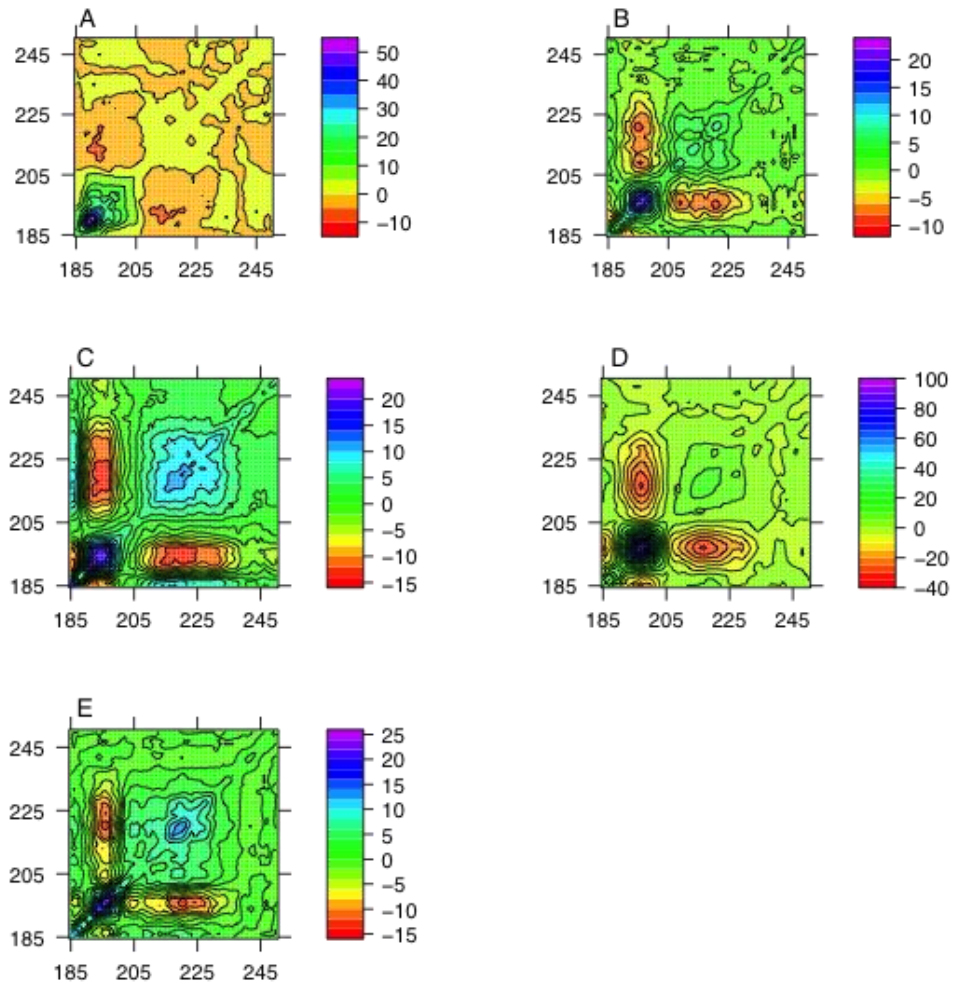

Figure 6- 11: A summary of the synchronous correlations of the 35 scan perturbation of AT alone (A) and in the presence of the ligands pentasaccharide (B), SH1 (C), SH2 (D), and UFH (E). These correlations only account for the spectrum up to 250 nm.

6.2.6. HEWL subjected to method 4, the direct comparison of two covariance matricies after prenormalisation.

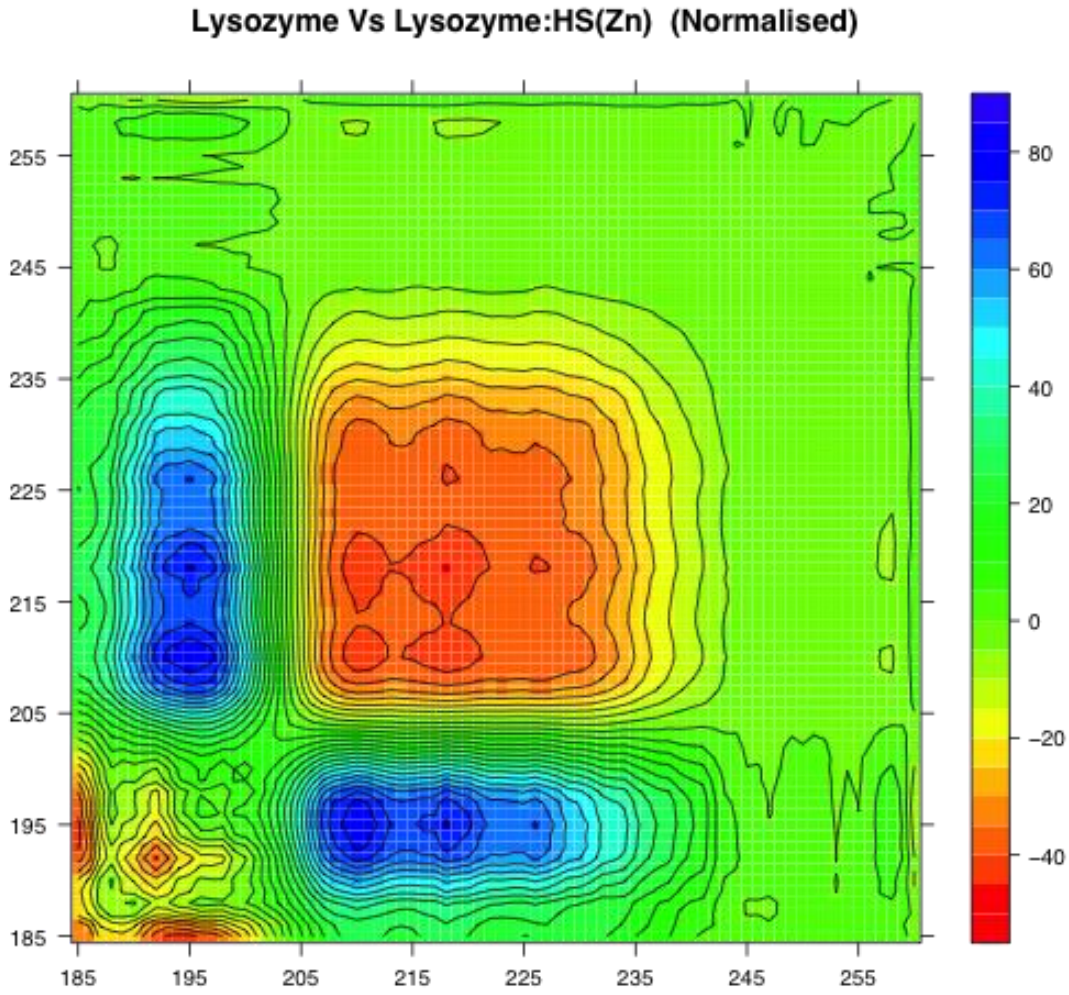

Figure 6-12: Covariance matrix of HEWL minus covariance matrix of HEWL:HS(Zn) following pre-normalisation. The perturbation being followed is a 60 scan UV degradation shown in Chapter 3, Figure 3-5.

6.2.7. HEWL subjected to method 3, MC across two perturbations to be compared.

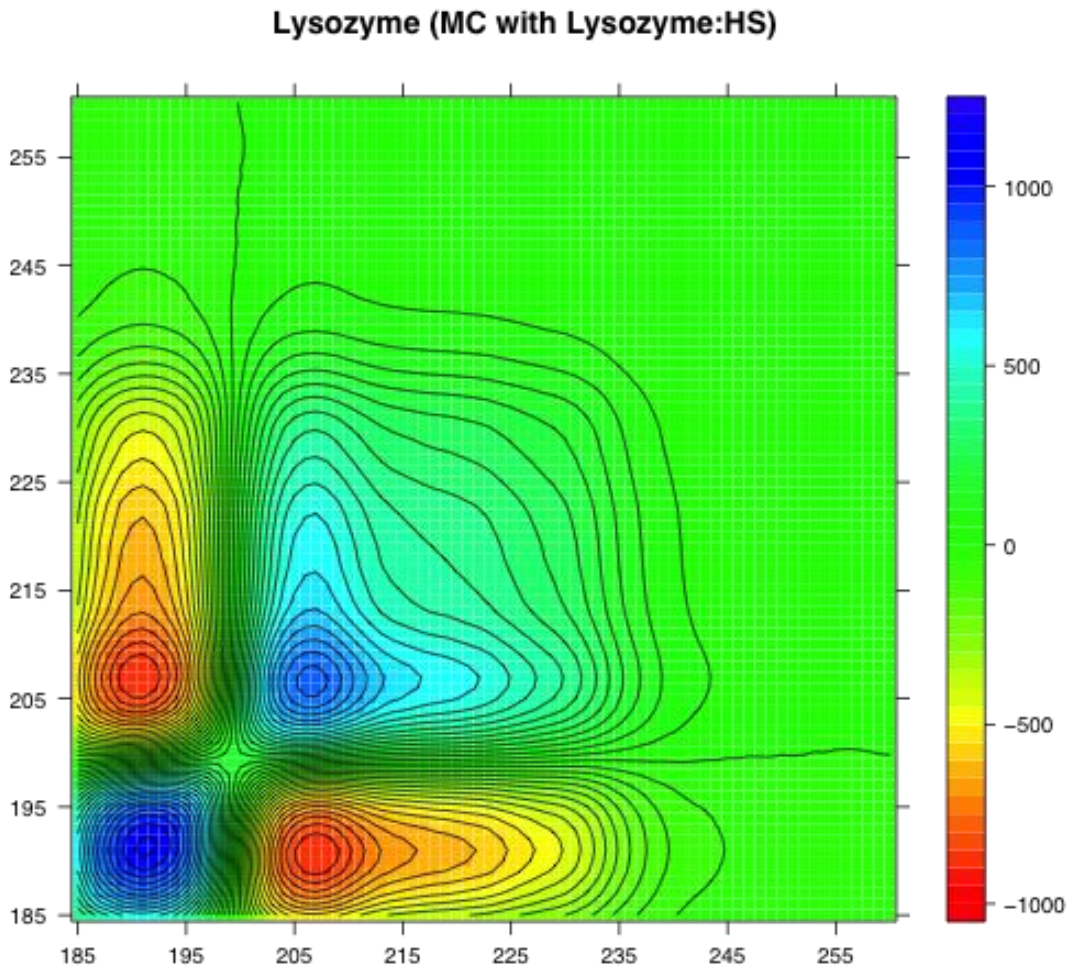

Figure 6- 13: Synchronous covariance matrix of HEWL undergoing a 60 scan UV degradation, mean centered with a 60 scan UV degradation of HEWL:HS, raw data from Chapter 3 Figure 3-5.

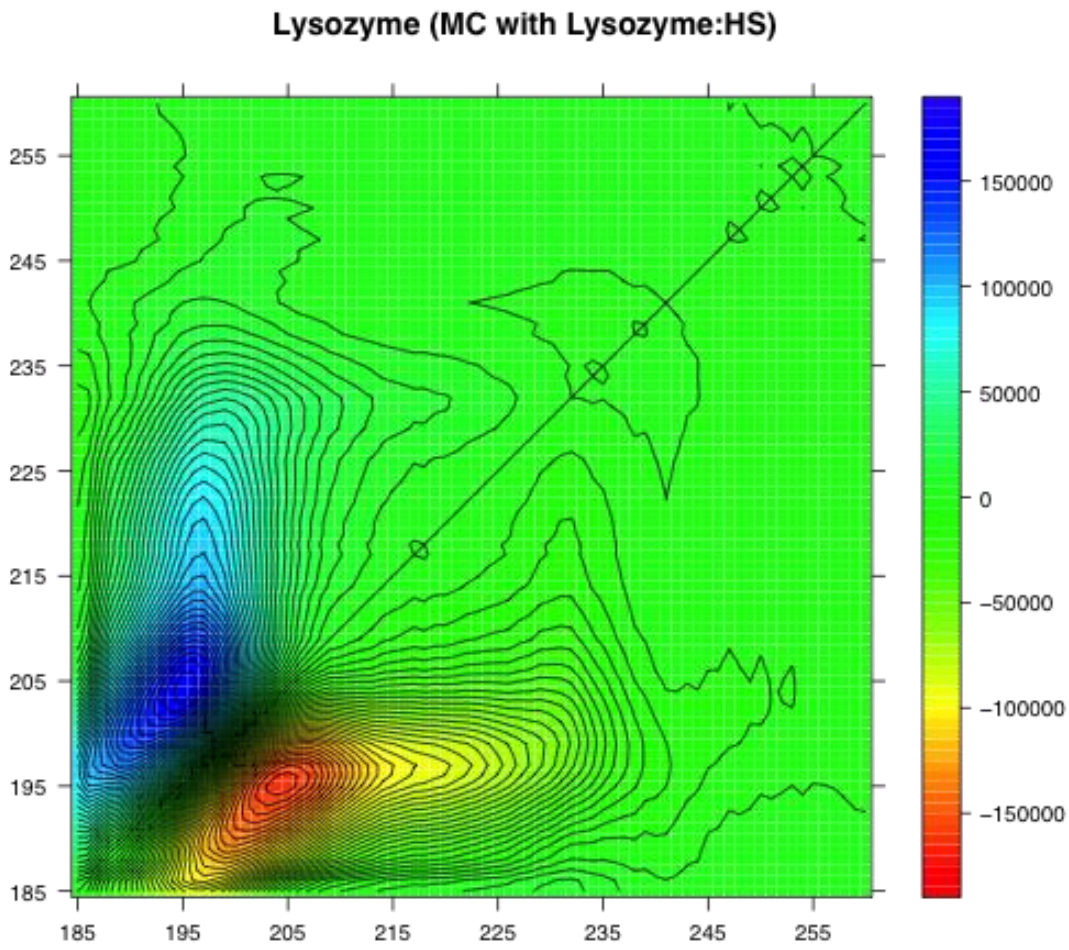

Figure 6- 14: Asynchronous covariance matrix of HEWL undergoing a 60 scan UV degradation mean centered with a 60 scan UV degradation of HEWL:HS, raw data from Chapter 3 Figure 3-5.

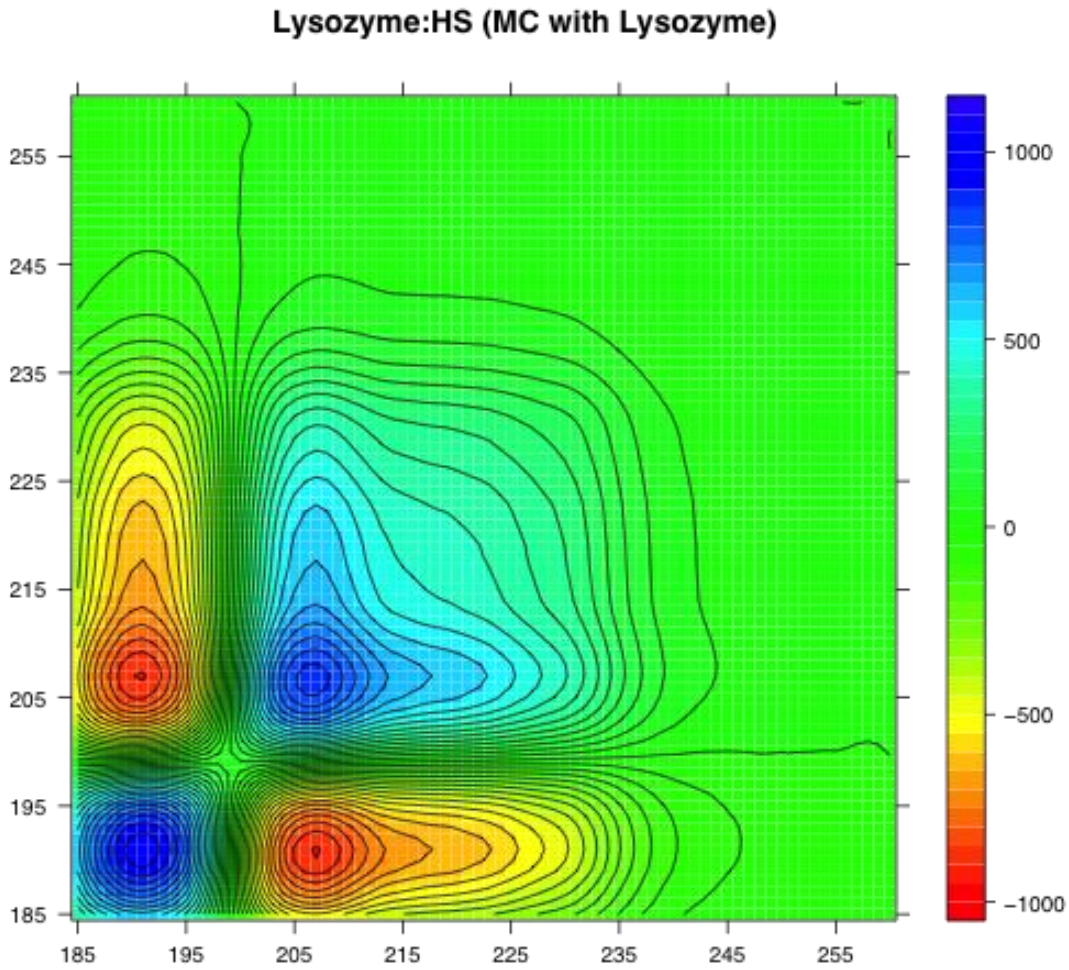

Figure 6- 15: Synchronous covariance of HEWL:HS undergoing a 60 scan UV degradation mean centered with a 60 scan UV degradation of HEWL, raw data from Chapter 3 Figure 3-5.

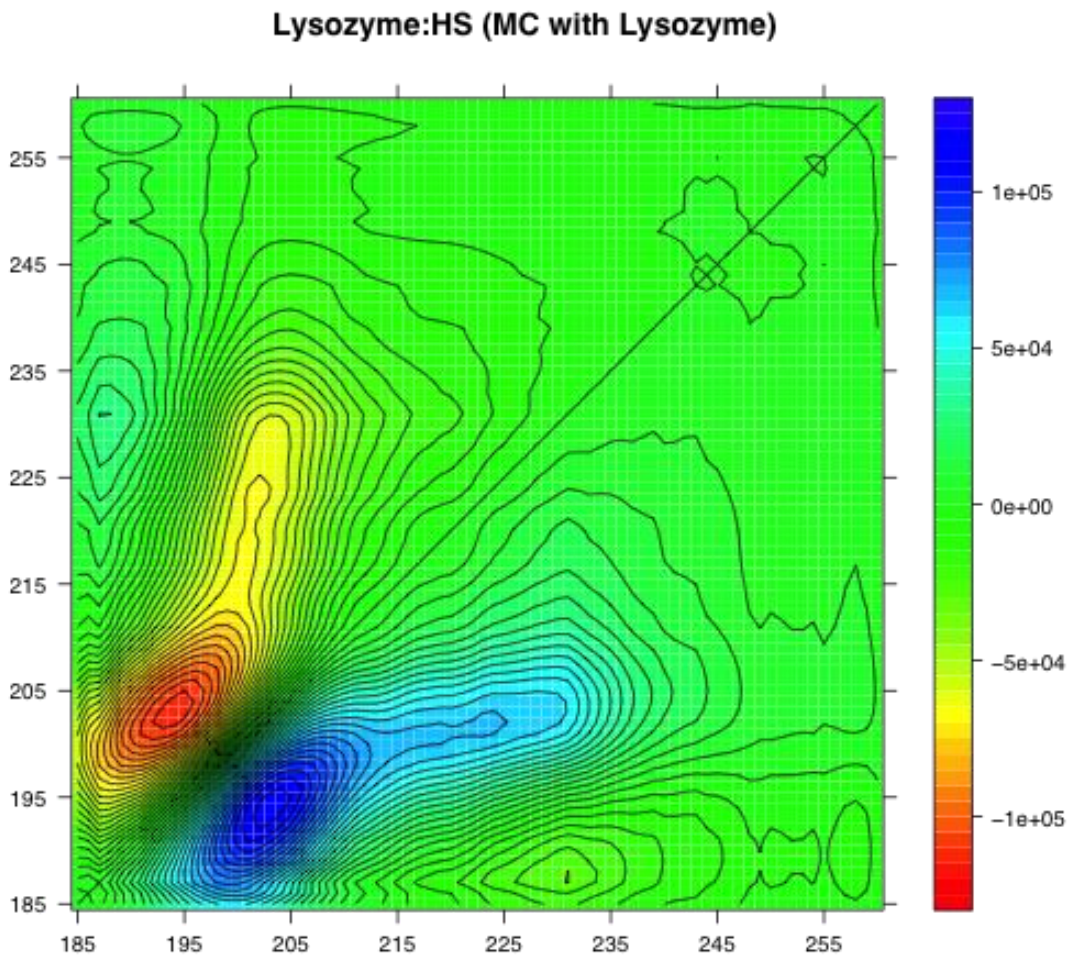

Figure 6- 16: Asynchronous covariance matrix of HEWL:HS undergoing a 60 scan UV degradation mean centered with a 60 scan UV degradation of HEWL, raw data from Chapter 3 Figure 3-5.

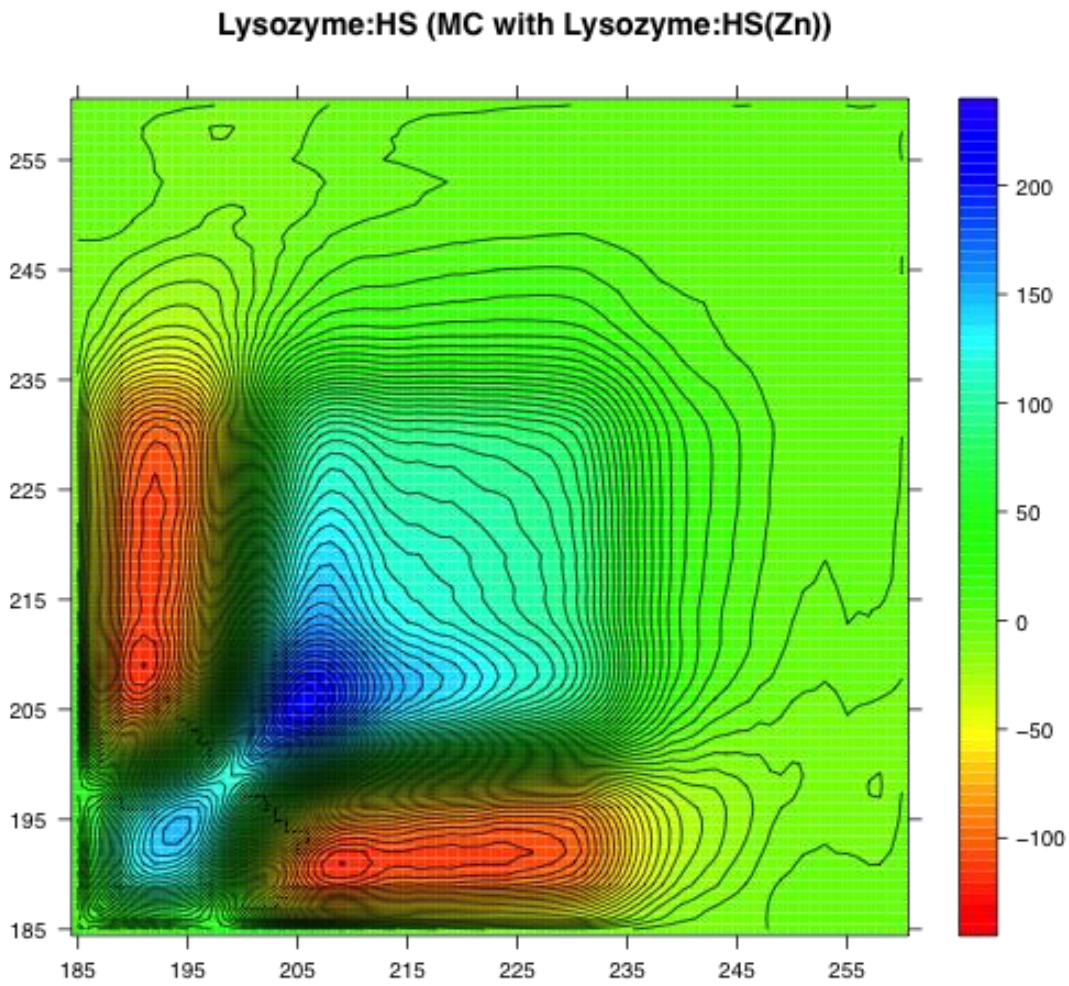

Figure 6- 17: Synchronous covariance matrix of HEWL:HS undergoing a 60 scan UV degradation mean centered with a 60 scan UV degradation of HEWL:HS(zn), raw data from Chapter 3 Figure 3-5.

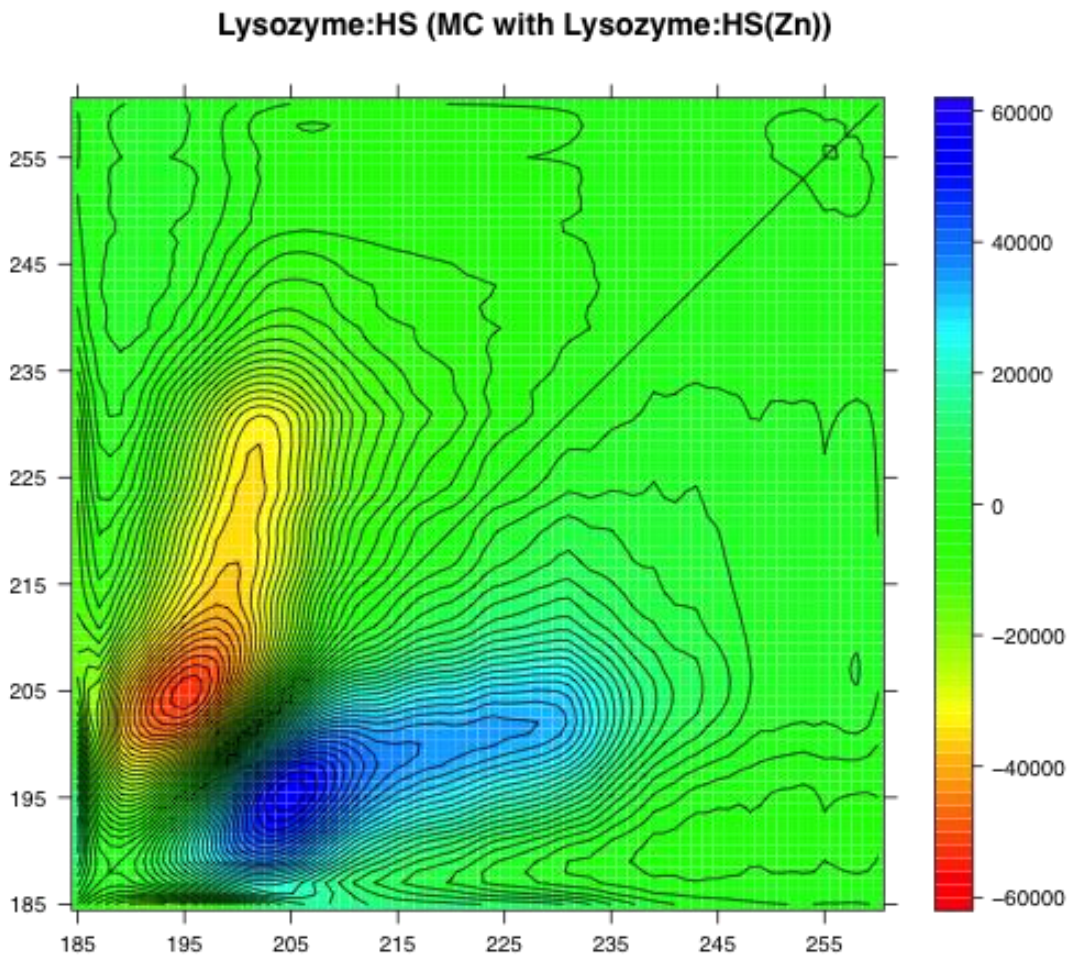

Figure 6- 18: Asynchronous covariance matrix of HEWL:HS undergoing a 60 scan UV degradation mean centered with a 60 scan UV degradation of HEWL:HS(zn), raw data from Chapter 3 Figure 3-5.

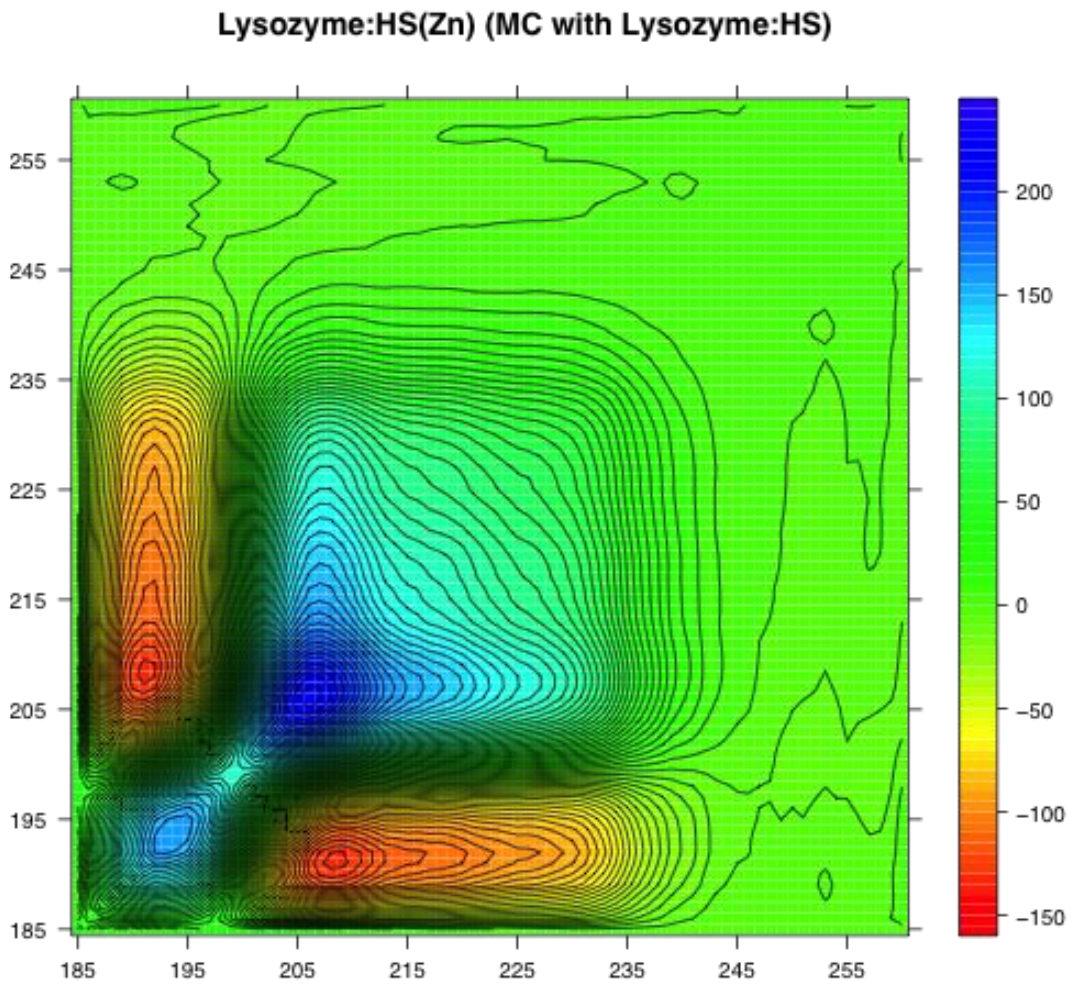

Figure 6- 19: Synchronous covariance matrix of HEWL:HS(Zn) undergoing a 60 scan UV degradation mean centered with a 60 scan degradation of HEWL:HS, raw data from Chapter 3 Figure 3-5.

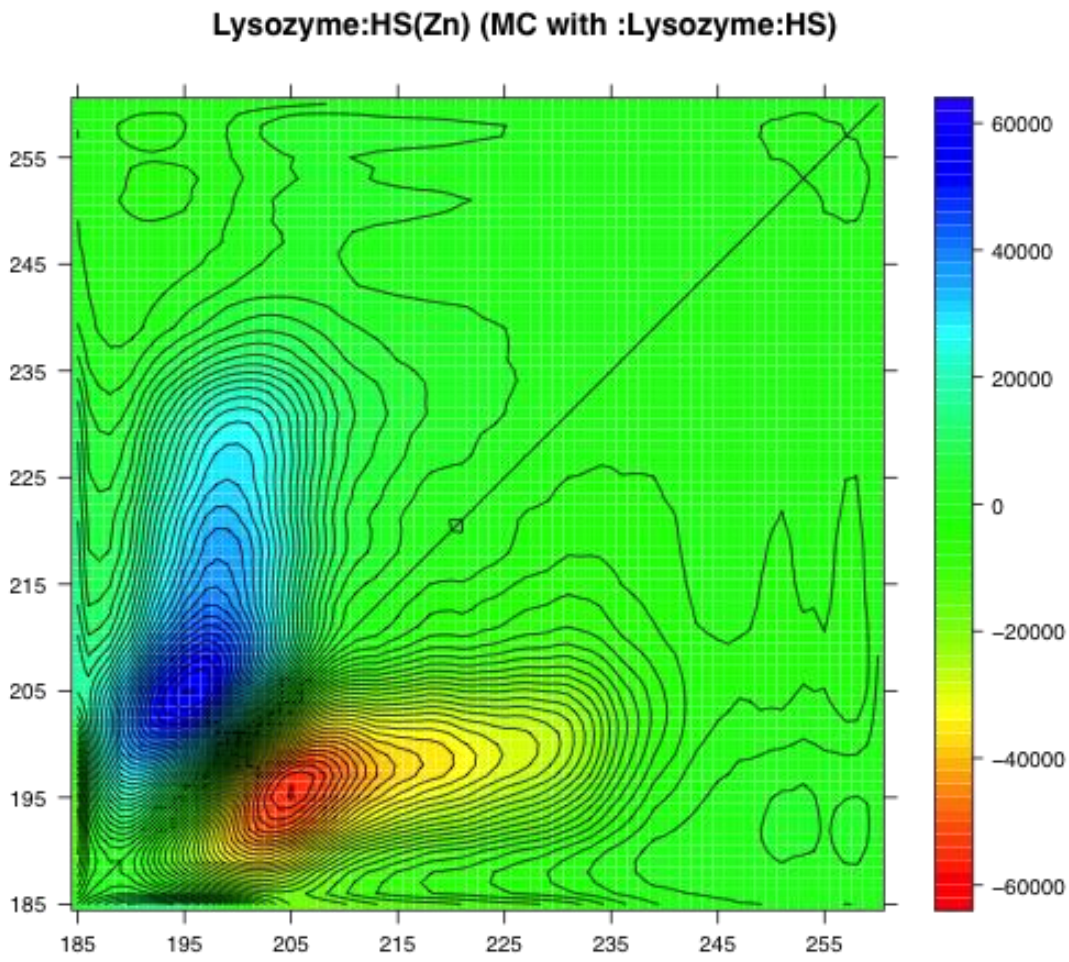

Figure 6- 20: Asynchronous covariance matrix of HEWL:HS(Zn) undergoing a 60 scan UV degradation mean centered with a 60 scan UV degradation of HEWL:HS, raw data from Chapter 3 Figure 3-5.

6.3. Discussion of results.

6.3.1. General covariance matrix interpretation rules and method 1, basic covariance matricies applied to AT perturbations.

The synchronous matrix summarized in Figure 6-1, from the first (standard) analysis method, highlights wavelengths that change in phase with one another. The peaks on the diagonal of these plots (Figure 6-2) reveal their auto-correlations, showing the spectral positions at which most change occurs within a spectrum and are always positive. The cross-peaks (off the diagonal, Figure 6-1) show which wavelengths are correlated together in the covariance matrix. These can be either positive or negative providing information on the direction of the change. For example, positive cross-peaks will indicate that both wavelengths change in the same direction; either positively or negatively, while a negative cross peak indicates that the two correlated wavelengths change in opposite directions; one in a positive fashion and the other in a negative direction.

The asynchronous covariance matrices are more relevant to the third analysis technique (alternative MC method). It reveals those wavelengths that change out of phase with the perturbation. More interestingly, this correlation can reveal the sequential order of events happening via their negative or positive sign (see summary in Figure 6-21). The rule followed, as shown by Noda ([217] Chapter 2.3.2), is that if the sign of the cross-peak is positive, then the first variable happens before that of the second, and the reverse is true if the sign is negative. This rule is reversed, however, if there is a corresponding peak in the synchronous matrix that has a

negative sign (See Figure 6-21). The convention is to observe the upper left half of a covariance matrix, above the diagonal. In the example below the asynchronous correlation has two positive features in the upper left while the lower right contains the negative features. The rule set out above are for spectra in their simplest cases, where the amplitude of individual peaks change without overlap or position shifting, to be discussed later.

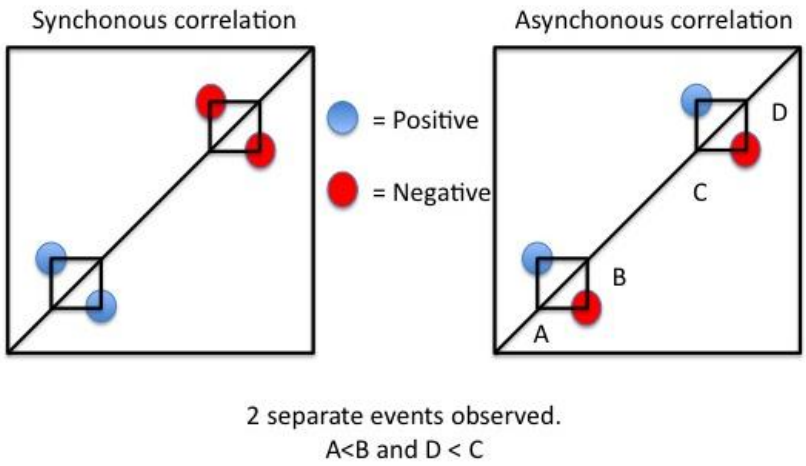

Figure 6- 21: A schematic diagram of the rule used to determine the order of events when analysing covariance matrices. The order of events is determined by the sign of the cross peaks in the asynchronous correlation relative to the sign in the upper left triangle synchronous correlation. If this is positive as per A and B, then A happens before B if there is a positive feature (or no feature) in the synchronous spectrum. This rule, however, is reversed if there is a negative feature present in the synchronous covariance matrix, as in the case of C and D.

The auto peaks on the diagonals shown in Figure 6-2 reveal that AT is different to AT in all of the complexes, and comparing the synchronous data shown in Figure 6-1

it is immediately apparent that there are differences between the complexes during the unfolding pathway caused by UV degradation. It is also noteworthy that, contrary to how the spectra appear in the raw data (Chapter 5, Figure 5-2), there is no evidence in these correlations (Figure 6-1) of any more complex features, such as line broadening (two overlapping peaks that change in opposing directions), band shifting, or band shifting coupled with intensity changes (for reference [217] Chapter 4.3). This leads to the conclusion that all of the features observed with the AT dataset are arising from independent events on static features. This observation could be significant in respect to algorithms for secondary structure, suggesting that they can be used for analysis of these types of data. There appears to be an artifact present in all correlations at 260 nm, probably due to the initial noise (discussed in Chapter 7) but does not appear to contribute to any cross peaks so has been ignored (for conformation see section 6.3.5.), rather, the analysis has picked up on this wavelength as an area of variation.

The diagonal of the AT correlation (Figure 6-2A) separates it from the other correlations by being devoid of correlations, beyond a feature at 190 nm. This single feature is relatively intense in comparison to all other correlations, with the exception of SH2 (Figure 6-2D), and it is also broad so could contain several contributions around that spectral area. The AT:pentasaccharide (Figure 6-2B) complex shows that there are several more points of change with a major correlated change at 196 nm and then three minor features at 208, 214 and 221 nm that are unique to this complex. AT:SH1 (Figure 6-2C) has changes at 192, 196, 220 and 231 nm. The double feature around the 195 nm region is unique to this complex and contributes to several unique cross peaks discussed later. AT:SH2 (Figure 6-2D) has only two

changes at 197 nm and 216 nm. The 197 nm correlation is the largest and best resolved correlation out of all of those analysed using this method. The feature at 216 nm is again broad and unresolved, possibly owing to several binding modes. AT:UFH (Figure 6-2E) comprises a feature at 196 nm and at 220 nm with a possible feature comprising a shoulder present at 228 nm.

The initial observations mark the AT:pentasaccharide as the most defined complex. There also seems to be a common feature between most of the complexes with the largest portion of variance being centered at 196 nm. This observation is to be expected as many structures have features in this region of the spectrum. This observation excludes the large and defined feature of AT:SH2, which is slightly shifted to 197 nm and AT:SH1 which, while containing the 196 nm also has a separate 192 nm feature. AT alone is the sole perturbation to exhibit its major change at 190 nm, immediately suggesting a different route during the perturbation. There is also an apparant link between AT:SH1 and AT:UFH in that they share the 220 nm feature in addition to the 196 nm feature.

AT:Pentasaccharide (Figure 6-3), is one of the best resolved correlations in this series. There are observable cross peaks at 186, 195 (x 3) and 196 nm. The values are obtained by referring to peak tops at the highest resolution, although these features obviously overlap significantly with one another. These features are respectively correlated with 191, 209, 213, 216 and 221nm as demonstrated through their cross-peaks.

The “double” set of cross peaks that run parallel to each other (Figure 6-4, ~195 nm) in AT:SH1 might also suggest that there are two distinct populations of interactions within this solution, one independent of all others observed here with its major changes at 192 nm and 231 nm, and the parallel one which has similarities with AT:UFH with its major changes at 196 nm and 220 nm.

6.3.2. Discussion of method 2. The deduction of one AT covariance matrix from another.

These results can be analysed further using what was referred to as the second method in Table 6-1. This method entails deducting the covariance matrix of one AT:GAG complex perturbation from that of AT:GAG complex perturbation. This analysis method has some limitations in that some artifacts are apparent – this especially appears to be the case at the higher wavelengths where no signals are expected – so, for this analysis all peaks above 235 nm will be ignored, and some attempt to remove signals of similar amplitude when they arise in illogical places will be made. There are many more signals originating at 185/186 nm in the following data, revealing potentially important differences between some spectra. These features are reliable owing to the data being collected reliably to 183 nm below the 700 mV limit for the detector (data not shown). These data were then analysed from 185 nm, thus avoiding any artifacts from a saturated detector.

Figures 6-5 and 6-6 correspond to this data analysis method with the first complex quoted in the main title being the matrix from which the covariance matrix of the second complex is deducted.

Figure 6-5 compares the relatively highly resolved AT:pentasaccharide with that of AT:SH1. The differences between these correlations are listed in Table 6-2.

Table 6- 2: The resolvable features present in Figure 6-5

Wavelength(s)	Auto/Cross peak	Sign	Notes
185 nm	Auto peak	Negative	-
192 nm	Auto peak	Negative	-
198 nm	Auto peak	Positive	-
218 nm	Auto peak	Negative	-
231 nm	Auto peak	Negative	-
186 / 196 nm	Cross Peak	Positive	-
186 / 219 nm	Cross Peak	Negative	-
186 / 229 nm	Cross Peak	Negative	-
186 / 231 nm	Cross Peak	Negative	-
192 / 196 nm	Cross Peak	Negative	-
193 / 217 nm	Cross Peak	Positive	-
193 / 231 nm	Cross Peak	Positive	-
197 / 217 nm	Cross Peak	Positive	-
197 / 231 nm	Cross Peak	Positive	-
220 / 230 nm	Cross Peak	Negative	-

This set of data reveals many differences between the two correlations, and adds resolution when studying how these two complexes respond to UV irradiation. The auto correlation peaks that are negative all refer to features that are related to

AT:SH1 while, conversely, the positive auto correlations refer to those stemming originally from AT:pentasaccharide. This is due to the amplitudes in the original (method 1) covariance matrices. The resolved and broad unresolved feature in the AT:SH1 correlation at 186 nm are here resolved into 4 separate features that mark part of this complex as being different from that of AT:pentasaccharide. In the initial analysis of Figure 6-4 (AT:SH1) it was suggested that there may be two sets of interactions occurring.

Figure 6-6 investigates the differences between AT:Pentasaccharide and AT:UFH. The differences are listed in Table 6-3.

Table 6- 3: The resolvable features present in Figure 6-6

Wavelength(s)	Auto/Cross peak	Sign	Notes
185 nm	Auto peak	Negative	-
192 nm	Auto peak	Positive	-
198 nm	Auto peak	Negative	-
218 nm	Auto peak	Negative	-
187 / 196 nm	Cross peak	Positive	-
193 / 227 nm	Cross Peak	Positive	-
195 / 227 nm	Cross Peak	Positive	-
196 / 202 nm	Cross Peak	Positive	-
196 / 204 nm	Cross Peak	Positive	Shoulder
196 / 212 nm	Cross Peak	Positive	-
196 / 217 nm	Cross Peak	Positive	-
196 / 234 nm	Cross Peak	Positive	-
202 / 210 nm	Cross Peak	Negative	-
202 / 214 nm	Cross Peak	Negative	-
202 / 218 nm	Cross Peak	Negative	-
204 / 218 nm	Cross Peak	Negative	-
208 / 214 nm	Cross Peak	Positive	-
208 / 223 nm	Cross Peak	Positive	-
219 / 226 nm	Cross Peak	Positive	-

Considering the changes listed in Table 6-3 the first noticeable thing is that there are many subtle differences between the spectra. Some of the apparent changes, however, are explained by differences in scale; the difference at 195 and 196 nm

occurs in both spectra, with ~5 units difference between them. This indicates that similar changes in the raw data occur around this point between the spectra. However, it does show that those in AT:UFH are more heavily correlated than those in AT:pentasaccharude, as expected from the more extensive degradation apparent in the PCA analysis shown in Figure 5-4 of Chapter 5. However, there are many small differences observable above 200 nm, showing this region to be of particular interest in separating out similar interactions during this degradation.

In summary, this method of investigating correlations along a perturbation appears to be capable of revealing more highly resolved features within comparisons of a similar scale. This is exemplified by the added detail contained in Figure 6-5 AT:pentasaccharide v AT:SH1. It suffers, however, with a scaling issue in its current form. Normalisation (Method 4, Table 6-1) may add better feature resolution, but lose information regarding the extent to which changes have occurred. This is arguably less important when there are other techniques that are better suited to interpreting the extent of changes occurring within the original raw data.

6.3.3. Discussion of AT perturbations subjected to analysis method 3, MC of data across two perturbations to be compared.

As can be seen in Figures 6-7 and 6-8, there appears to be an overall loss in resolution when looking at the synchronous covariance matrix. However, this type of analysis appears to reveal some asynchronous correlations not present in other forms of covariant analysis (Figure 6-8). These types of correlations can reveal information regarding the sequential order of events taking place within a perturbation, and report

those events that happen out of phase with the progression through the perturbation (explained in section 6.3.1. Figure 6-21).

Figures 6-7 and 6-8 show the synchronous and asynchronous correlations of the perturbation of AT:pentasaccharide when mean centered together with AT:SH1. This case was chosen as an example as it reveals differences between the active complex of AT:pentasaccharide and the structurally similar but inactive complex of AT:SH1 (See Chapter 5). This method of MC across two perturbations before the independent correlations are made removes the common features between them. These correlations show no signs of the complicated features that are present in the HEWL examples of this analysis (Figures 6-13 - 6-20) and are explained later in this chapter (Section 6.3.8), meaning that the sequential information can be gathered from the asynchronous data.

Figure 6-7, showing the synchronous covariance matrix of AT:pentasaccharide when mean centered with AT:SH1 has two auto correlations at 203/235 nm. There are also cross peaks between 187/194 nm (negative sign), 193/212 nm (broad along the 193 axis, negative sign) and finally between 203/237 nm (positive sign).

Shown in Figure 6-8 is the asynchronous correlation for the perturbation presented in Figure 6-7. It is immediately apparent that this correlation is relatively information rich, especially when it is considered that other analysis techniques do not present the user in these examples with sequential information regarding the order of events. There are 7 cross peaks when all features that appear below 245 nm are considered, these arise from ten positions on the diagonal, referred to here as A-J. There are

features present at 185 (A), 195 (B), 198 (C), 208 (D), 209 (E), 214 (F), 221 (G), 237 (H), 240 (I), and 244 nm (J). when the center of peak tops are considered it can be deduced that there are 4 independent events taking place, and the order given to them as described in section 6.3.1 by taking into account the signs of the features. From this we conclude that A/H precedes a change in B, that a change in D happens before C and that E/F/G happen before I, while finally G also changes before J.

6.3.4. Discussion of AT perturbations subjected to analysis Method 4; The prenormalisation of covariance matrices before deduction from one another.

An alternative method of analysis is shown in Figures 6-9 – 6-10. These figures feature comparisons between complexes that initially had problems with scaling due to the larger contributions of AT:SH2 to the subsequent covariance matrix. Here all feature heights have been prenormalised to the same scale before deductions were made. This removes any information regarding the extent of each change but, as can be observed in these example figures, a much higher resolution is achieved in comparison to those achieved in method 2. This is especially evident through the large broad feature along the axis of 199 nm from AT:SH2. These features are now resolved into several individually identifiable peaks.

6.3.5. Discussion of AT perturbations subjected to method 5, accounting for the extra noise inherent in the 260 nm data point.

Finally, as there appear to be artifacts observed around ~260 nm, easily seen in Figure 6-2, a variant of the original (method 1) method was tested. This method (method 5, Table 6-1.) takes into account the need for equal variance along the data vector when mean centering is used, therefore, as noted in Chapter 7, there is increased noise initially during the recording of a spectrum. To eradicate the effect of this noise, only the wavelengths 185 to 250 nm have been considered (Method 5), thereby ensuring the equal variance consideration is fulfilled. As can be observed by the summary in Figure 6-11, there is no difference between this and the summary shown in Figure 6-1. This shows that, despite the observed artifact in Figure 6-2 and in some of the subsequent correlations, this has no material effect on the interpretation, as long as the user is aware of the possibility of these features arising.

6.3.6. Discussion of method 1, basic covariance matrices applied to HEWL perturbations.

The analysis of data from the HEWL 60 scan UV degradation presented in Figure 3-5 of Chapter 3 is now analysed in the same way. This enables the comparison of two different systems to further assess the suitability of this method as a technique capable of adding information to the raw data presented, and also to the PCA that was carried out on this type of data and this is shown in Figure 3-7. Data is not presented here for Method 1 for the HEWL perturbations as the correlations given are of broad, uninformative features in the synchronous covariance matricies and no

features present in the asynchronous. Although not presented and of low resolution, this did have some differences in their absolute feature positions, thereby providing additional information to supplement the PCA analysis previously published on this data set [142].

6.3.7. Method 3 applied to HEWL degradation data. A method to show subtle events arising from factors other than band intensity changes within a data set.

The following method of analysis, shown in Figures 6-13 to 6-20, reveals that these UV degradations are much more complex than previously shown (PCA analysis, Chapter 3), offering valuable information in the patterns created. This method MC across the two perturbations being investigated before separate covariance matricies are created in order to remove any common features in the two spectra. The patterns, not observed in the AT data, will now be described briefly.

The patterns in the synchronous and asynchronous covariance matrices demonstrate that the rules previously discussed in Figure 6-21 cannot be used in the analysis of Figures 6-13 to 6-20. The synchronous four leaf clover patterns are indicative of two complex spectral features, either two separate bands with intensity changes, or a single band shifting position. These two possibilities are then differentiated via the asynchronous matrix, with the band shifting position giving rise to the elongated features parallel to the diagonal, whereas the simpler two overlapping bands changing in intensity would give rise to two “spots” on the asynchronous matrix [217]. The four-leaf clover might be observed in Figure 6-15.

Another process can be identified where there is a band shifting position, coupled with an intensity change. Again identified via a four-leaf clover pattern on the synchronous matrix, this time the pattern may be asymmetrical giving rise to an "angel" pattern (Figure 6-13, 185 – 220 nm). This can be confused with other features, so again the asynchronous covariance matrix assists the identification where there is an elongated feature parallel to the diagonal. This time however, one "leg" will be longer than the other, biased towards the dominant auto peak (angel's legs) on the synchronous matrix (Figure 6-14, 185 – 220 nm). There is one more process that is identifiable – line broadening, which is described in the literature [217] and does not feature in this set of data.

These patterns are very sensitive, with 0.5% shifts in intensity and position being reported to give rise to these distinctive patterns [217]. There is information contained within these patterns, the centre of the four-leaf clover is the average feature position during the perturbation, and the signs of the elongated features in the asynchronous give information on the direction of the band shifting. The features themselves are insensitive generally to the extent of shifting, with the only direct correlation coming from the amplitude of the features.

The synchronous covariance shown in Figure 6-13 is of HEWL mean centered with HEWL:HS. There is a four-leaf clover pattern, centered at 199 nm. The elongated features in Figure 6-14 show this to be a single band shift coupled with an intensity change. The feature is shifting from the left to the right – deduced via the positive feature being above the diagonal.

There are again the same features present in Figures 6-15 and 6-16, with the clover being centered at 199 nm. The asynchronous this time suggests that the movement of this feature is from right to left in the spectra. The scale of this matrix indicates that, in this instance, the shift of the position is larger than for HEWL.

The following four figures (Figures 6-17 – 6-20) come from HEWL in both ligand states mean centered against one another. This case appears to be slightly more complex though, it is apparent that there is some band shifting taking place, and initially appears as if it is again coupled with band intensity changes. However, when looking at the asynchronous correlations, the elongated features do not appear to be biased towards one side of the four-leaf clover, indicating instead that it is a simple position shift in a band, with no intensity changes. If this is the case, then there is extra information contained within this analysis that cannot be elucidated currently. The HS form has a band shift from the right to the left of the spectra, while the HS(Zn) form from left to right, and in this instance there is about equal distance in the shift. This analysis is hindered by the position of the features been cut off by the edges of the plot, leading to some distinctive components not being present as described by Noda [217].

The analysis undertaken here has shown itself capable of adding greater resolution to spectra following a perturbation. This improved resolution has enabled the distinction between complexes that hitherto were indistinguishable structurally via the standard data analysis techniques of protein structure deconvolution, degradation via UV irradiation and the analysis of the raw data by PCA. It is possible with this technique to distinguish several independent contributing features in and around the

usually complex features apparent at 190-195 nm. The improved resolution however opens up several more challenges that will need to be overcome in order to further the assignment of structural features that are changing during a perturbation. There could be an issue with certain structures, such as an α-helix having multiple and differing local environments for each chain (in this example there are several separate helices with one becoming extended upon ligand binding), which may shift the peak positions of its transitions. As mentioned previously, these overlap with major features for other known structural elements making the assignment at this position difficult.

6.3.8. Discussion of HEWL perturbations analysed by Method 4; the prenormalisation of covariance matricies before deduction from one another.

If the covariance matrices created using Method 1 were probed for differences through normalization of the covariance matrices and then deducted from one another to allow for a direct comparison, subtle differences reveal themselves. Figure 6-12 is highly resolved, and is comparing HEWL and HEWL:HS(Zn). There are three auto correlations at 185, 192 and 218 nm. Then, in addition, a cross peak between 185 / 194 nm, and three from 195 / to 210, 218 / 226 nm, with a final cross peak between 218/226 nm.

6.3.9. General discussion of results.

It would appear that a combination of the methods presented in this chapter would be the best approach in any future work. The standard way of analyzing the

perturbations by 2D correlations, while adding information to supplement other analysis techniques, still gives the lowest resolution of all the methods investigated in this chapter. Taking these correlations and deducting the perturbations to be compared does increase the resolution, however, these tend to be biased towards the correlation with the largest scale. This is helpful to compare the extent of change at given correlations, but other analyses such as PCA could do this job. The method whereby two covariance matrices, that have been normalised to the highest peak from both spectra offers the best resolution of the differences and should be used for that purpose (Method 4). This method is well understood as each difference covariance matrix created with Method 4 stems from those created in the well-studied first method.

Mean centering across two perturbations to be compared before analyzing the perturbations independently by 2D correlations, offers an alternative route, and within these two experiments appears to supplement the other data significantly. This method works on the basis of eliminating, or reducing, features that are in common between the two perturbations before correlations are applied and thereby reveals more subtle differences that are hidden underneath the dominating major features. This serves to reveal previously unobserved features, especially apparent in these two examples. In the AT data series, new features are revealed, most noticeably in the asynchronous matrix, whereby some sequential information along the perturbation can be elucidated. This could provide in-depth structural information on the processes being observed when compared to previous analysis techniques. In the lysozyme experimental set, the analysis also reveals some subtle band shifting and overlapping signals with high sensitivity. This information will be vital if

quantitative secondary structures are to be elucidated at each step of a degradation, because algorithms assume that features have independent peaks, whose intensity changes depend on the contribution of that feature. In this study, it would be inappropriate to apply these structural analyses to the HEWL perturbations. Conversely, the lack of the distinctive features in the AT perturbations means that structural information can be gained at any stage during the perturbation, and any analysis that is applied is limited by signal to noise ratios and the quality of the algorithm without being complicated by any line broadening, shifting or overlapping features that change intensity. This also means that the peak tops in the AT case are reliable indicators of separate features, so much higher resolution can be obtained. For example, features differing by 1 or 2 nm can reliably be taken as two separate features rather than just one broad feature with low resolution.

In future analysis and data mineing of a peturbation should include a combination of Methods 1, 3 and 4. This would give the best range of results and information to added to more standard analysis techniques, although care should be taken in the appropriate pretreatment of data [2].

Chapter 7. Improvements to data handling for Raman optical activity and SRCD spectroscopy.

This chapter will describe two developments in the handling of general spectroscopic data, including those that involve GAGs. The first development will look at Raman and ROA spectra, where a method has been developed for the correction of baseline drifts and the normalisation of data. The method is intended for Raman and ROA data but, in principle, could be applied to any spectroscopic technique where the need arises. Secondly, the new method for normalisation of SRCD will be demonstrated. The need arose after the observation of an increase in noise at the initial data point (260 nm) within a SRCD spectrum. This will be demonstrated here, with the new method as a comparison.

7.1. Development of a baseline subtraction method for Raman and Raman Optical Activity spectroscopy.

Raman spectroscopy is the basis for ROA and stems from inelastic scattering from a molecule whose polarisability changes through interaction with the electric field of the incident radiation. The inelastic scattering originates through Stokes and anti-Stokes contributions, see Figure 7-1. There have been Raman spectra reported for monosaccharides [218] and polysaccharides in the Raman with assignments [219]. This technique is highly complementary to IR spectroscopy and can give information on conformational states of molecules, with signals arising over very short time scales, for example at 1000 cm^{-1} a transition will take ~10^{-14-} s [220].

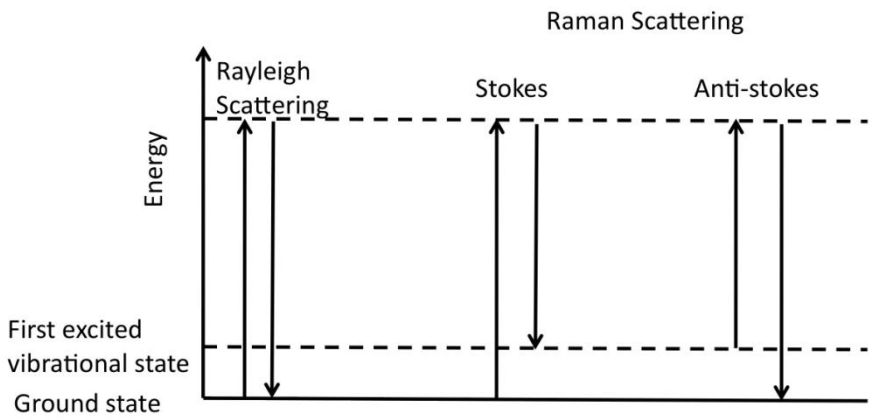

Figure 7- 1: The energy transitions involved in Raman scattering compared to the more common phenomenon Rayleigh scattering. Raman scattering is split into two categories, Stokes, and anti-Stokes. ROA is the result of the differences between optically active Raman scattering molecules in the presence of L and R polarised light.

ROA is an extension of the Raman and stems from the difference in intensity between Raman scattering of right and left handed circulary polarised radiation from a chiral molecule, much like that reported for SRCD. The signal arising from this technique is very weak, however, the potential information is potentially of great value in attempts to elucidate some form of structure-function relationship for the GAG component of protein-GAG interactions. ROA is highly sensitive to conformational features [221], for example the geometry of the glycosidic linkage and sugar puckering [222,223,224,225,226]. Of particular interest is the possibility that this technique could report helicity in solution for polysaccharides [82]. With this in mind, and following on from recent developments where the first full length GAGs were investigated in solution by ROA [227,228], a set of experiments were devised to attempt to better assign features observed in carbohydrate (heparin) ROA

spectra via the 2D heterocorrelation of a perturbation recorded in the ROA with reference to NMR where signal positions for each atom are fully mapped in heparin [30]. The perturbation employed was the addition of a cation manganese (Mn^{2+} in the form of $MnCl_2$) that was shown to interact with, and alter the conformation of, heparin (Figure 7-2, and NMR (data not shown)). This perturbation was performed and data recorded. However, owing to a malfunction in one of the components of the ROA equipment the data were not of sufficient quality to use as intended. The analysis of the ROA data has led to a method for baseline correction in the technique and this will be described below.

Figure 7-2 shows the raw data for the Raman (upper panel) and ROA (lower panel) spectra for the titration of $MnCl_2$ into heparin in the original form of handling the data, normalised to a featureless point of the ROA spectrum. The Raman spectra should contain only positive points, while ROA should consist of negative and positive features based around zero units.

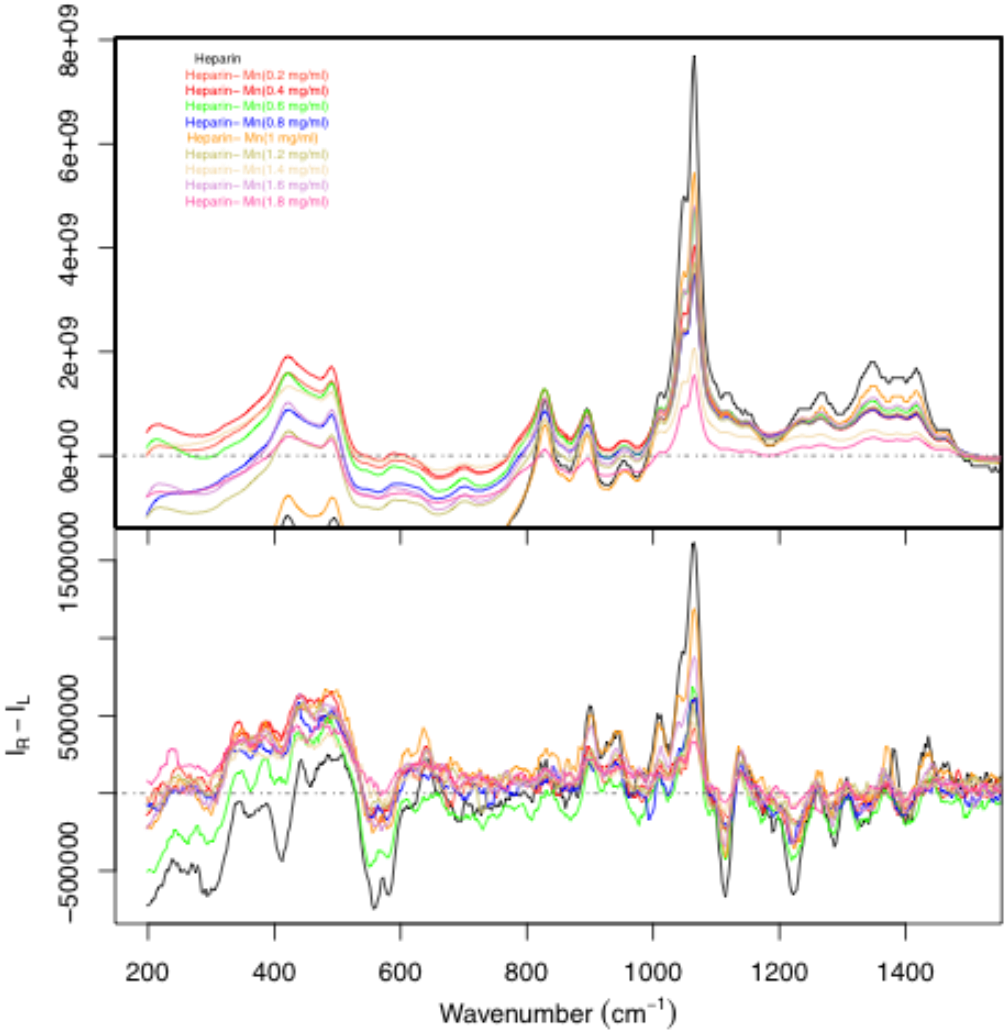

Figure 7- 2: Titration of $MnCl_2$ (0 – 1.8 mg /ml) in 0.2 mg/ml steps into a heparin solution (300 mg/ml) in 90% H_2O and 10% D_2O, recorded for 18300 scans, 1.029 s exposure per scan and 0.30 W laser strength (532 nm). The data have been normalised at 1592 cm^{-1}. All samples were centrifuged for 10 minutes at 30k rcf and pre-illuminated using a full strength laser for 30 minutes prior to recording.

The spectra presented in Figure 7-2 have a clear baseline drift despite being normalised to a featureless region of the spectrum (1592 cm^{-1}). In order for comparisons between spectra to be made, this drift needs to be corrected for. The first step in this for the Raman and ROA data collected here is to perform a spline fit (using the Forsythe, Malcolm and Moler method) for the x-axis data collected in order to assign the data points to the same position across all spectra. This is necessary because each spectrum contains a recorded wavelength to 3 decimal places for each data point, and is different in each spectrum. After completion of this step a background spectrum is created for each individual spectrum, here presented for the spectrum of heparin (Figure 7-3) and deducted from the raw data.

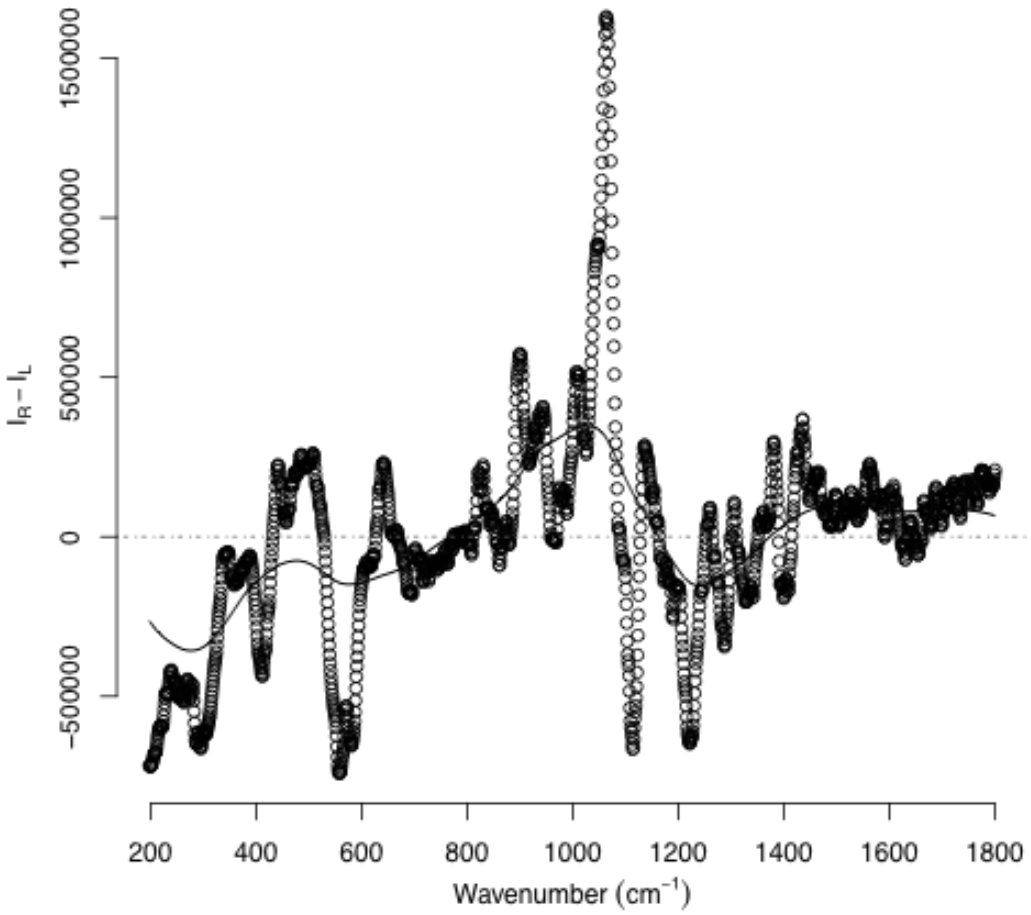

Figure 7- 3: Circles: Raw data of ROA spectrum of heparin (300 mg/ml) in 90% H_2O and 10% D_2O. ROA recorded for 18300 scans, 1.029 s exposure per scan and 0.30 W laser strength (532 nm). Solid black line: Raw data smoothed using a forwards and reverse pass filter of 200 data points using the R function “filtfilt” [229].

The raw ROA data were taken and smoothed by applying a forwards and reverse pass filter over 200 data points to create the solid black line shown in Figure 7-3. This solid line for each individual spectrum was then deducted from the corresponding raw data spectrum to apply a baseline correction; this serves to flatten the spectra, correcting for drifts introduced during the recording of the data. These data were then smoothed again over 2 data points in the same manner, to reduce noise in the spectrum, at the expense of resolution. This smoothed data is then deducted from the baseline corrected spectrum and is shown in Figure 7-4.

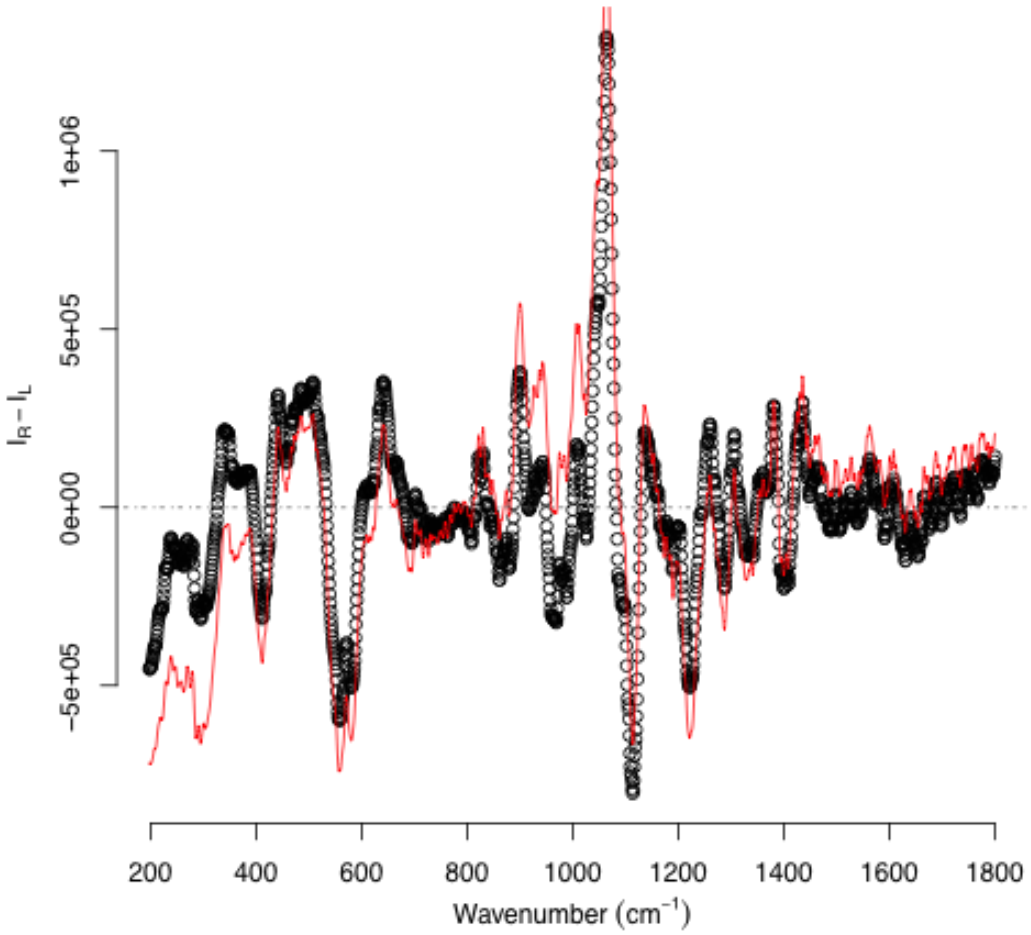

Figure 7- 4: Circles: Baseline corrected data of the ROA spectrum of heparin solution (300 mg/ml) in 90% H_2O and 10% D_2O, recorded for 18300 scans, 1.029 s exposure per scan and 0.30 W laser strength (532 nm). These data have been smoothed again over 2 data points to reduce noise. Solid red line: Original raw data presented in Figure 7-2 for comparison.

The data presented in Figure 7-4 were then taken and the average value between 1700 and 1800 cm^{-1} found, and deducted from the spectrum to give the data presented in Figure 7-5 (black circles). This sets the spectra to zero at a featureless region of the spectrum.

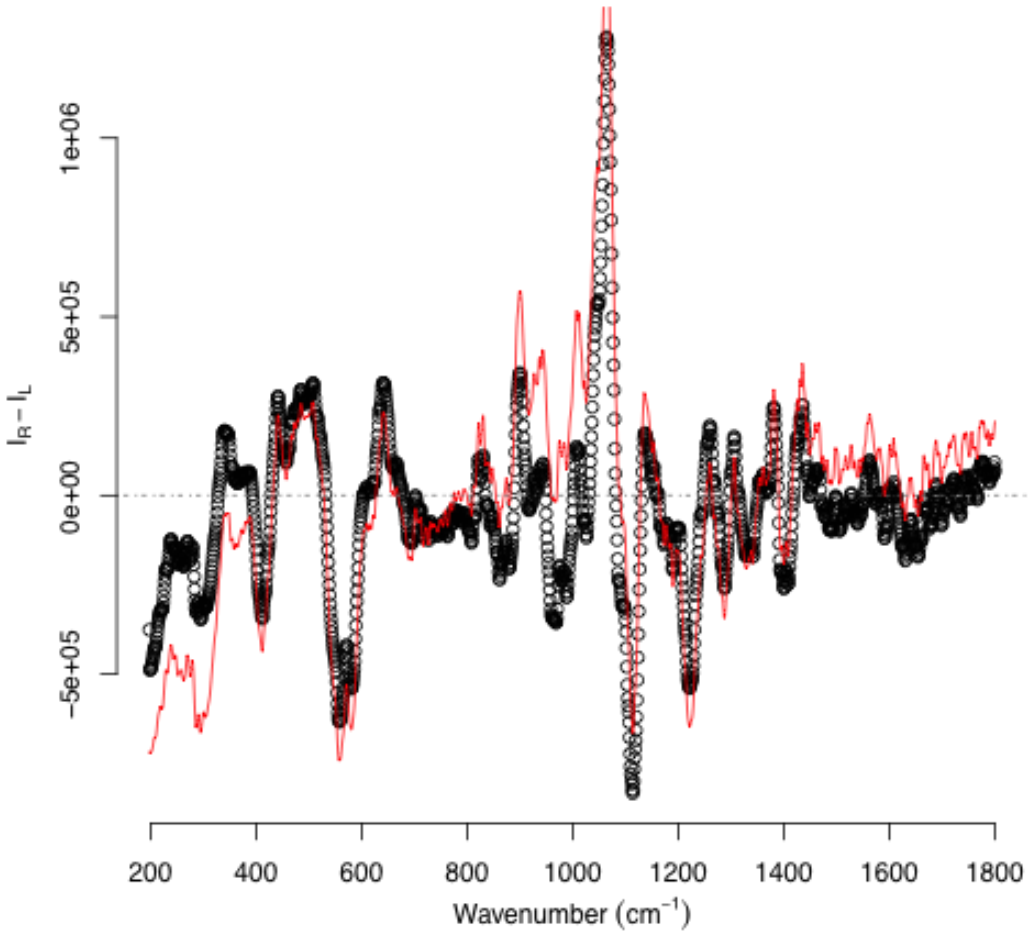

Figure 7- 5: Circles: Baseline corrected and normalised data (1700 – 1800 cm^{-1}) of a heparin solution (300 mg/ml) in 90% H_2O and 10% D_2O, recorded for 18300 scans, 1.029 s exposure per scan and 0.30 W laser strength (532 nm). These data have been smoothed again over 2 data points to reduce noise. Solid red line: Original raw data presented in Figure 7-2 for comparison.

This process was then completed across all the ROA spectra shown in Figure 7-2, and the results are shown below in Figure 7-6. The top panel shows the unprocessed data zeroed at 1592 cm^{-1} and lower panel, the baseline corrected spectra.

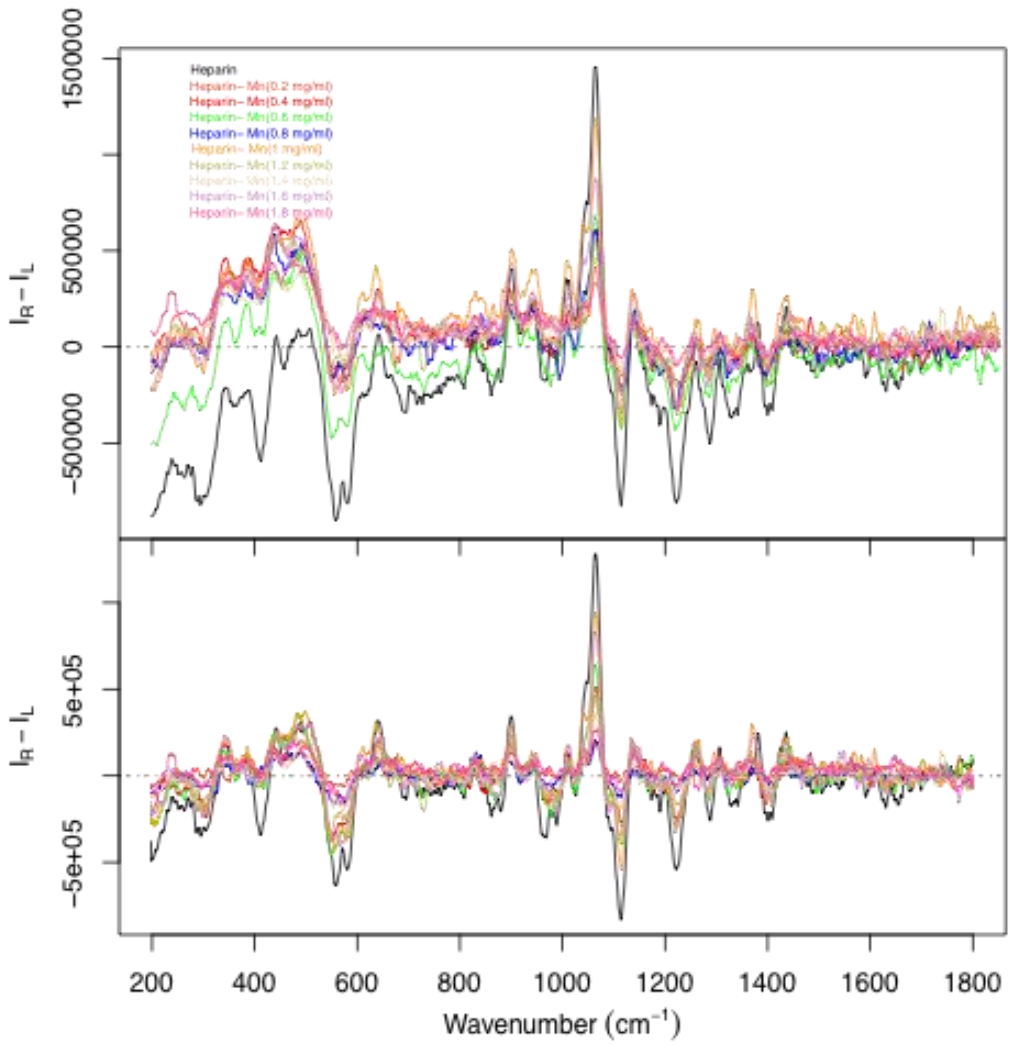

Figure 7- 6: ROA spectra of a titration of $MnCl_2$ (0 – 1.8 mg /ml) in 0.2 mg/ml steps into a heparin solution (300 mg/ml) in 90% H_2O and 10% D_2O, recorded for 18300 scans, 1.029 s exposure per scan and 0.30 W laser strength (532 nm). All samples were spun for 10 minutes at 30k rcf and pre-illuminated using a full strength laser for 30 minutes prior to recording. Top panel: The data were normalised to 1592 cm^{-1}. Lower panel: The data analysis method outlined in Figures 7-3 – 7-5 was applied then normalised to the average value between 1700 and 1800 cm^{-1}.

Shown in Figure 7-6 is a comparison between the original and the new methodologies for dealing with ROA spectra. This new methodology gives a much flatter baseline – especially apparent towards the lower wavenumbers (200 – 800 cm^{-1}) further away from the normalisation performed in the original data analysis (Figure 6, upper panel). The new baseline correction method was then applied to the Raman data with only minor changes to procedure and shown in Figure 7-7.

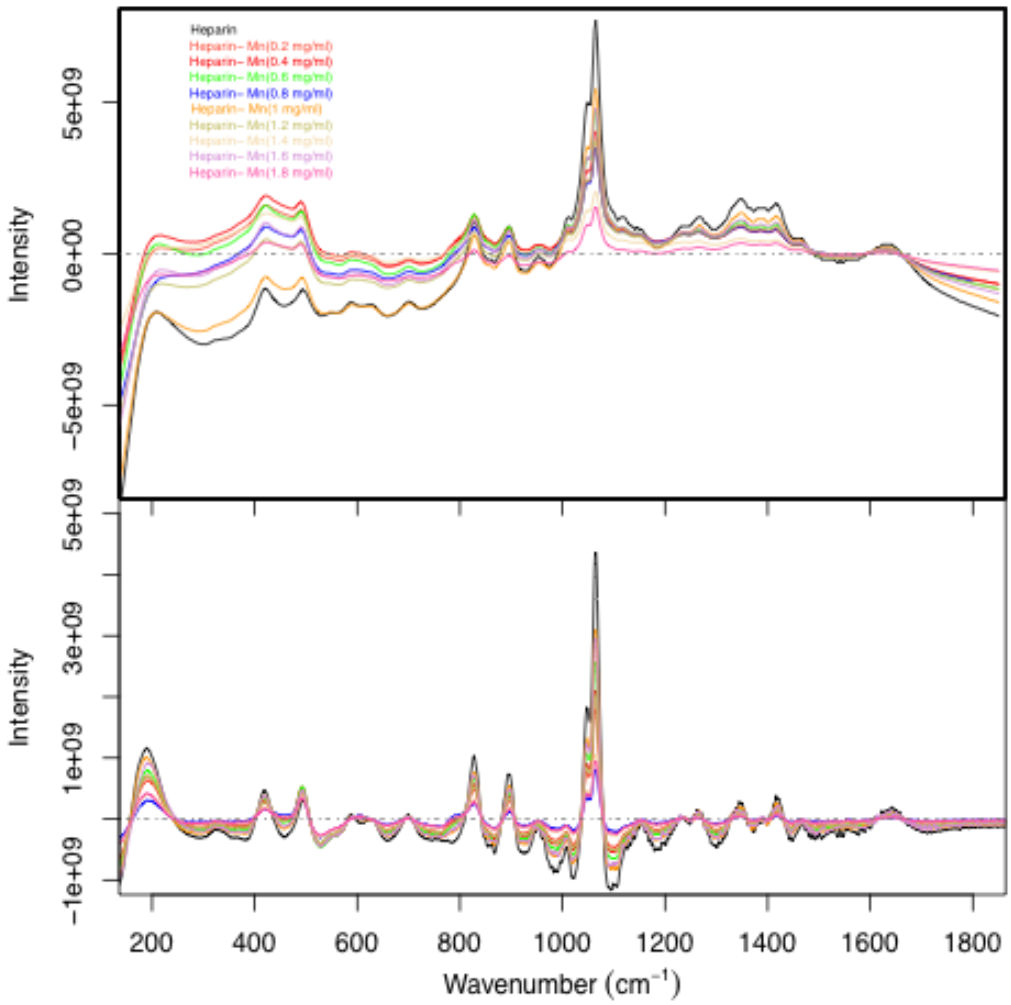

Figure 7- 7: ROA spectra of a titration of Mn^{2+} (0 – 1.8 mg /ml) in 0.2 mg/ml steps into a heparin solution (300 mg/ml) in 90% H_2O and 10% D_2O, recorded for 18300 scans, 1.029 s exposure per scan and 0.30 W laser strength (532 nm). All samples have been spun for 10 minutes at 30k rcf and pre-illuminated using a full strength laser for 30 minutes prior to recording. Top panel: Data have been normalised to 1592 cm^{-1}. Lower panel: data analysis method outlined in Figures 7-3 – 7-5 were applied to correct for baseline via the use of a two pass filter over 100 data points, and normalised to the average value between 1700 and 1800 cm^{-1}.

Figure 7-7 shows the result of applying the same baseline correction method to the Raman spectra, in this instance spectra were smoothed over 100 data points and not 200. There is less need to reduce noise owing to the extra intensity of the Raman signal compared to its optically active counterpart. As can be seen the correction method appears to improve the quality of the baseline across all spectra presented. However, there are still some errors, as is apparent from the negative signals shown around the 800 – 1200 cm^{-1} region. This has been accounted for in Figure 7-8 by simply taking the square root of the square of the data shown in Figure 7-7 (lower panel) to make all data zero or positive.

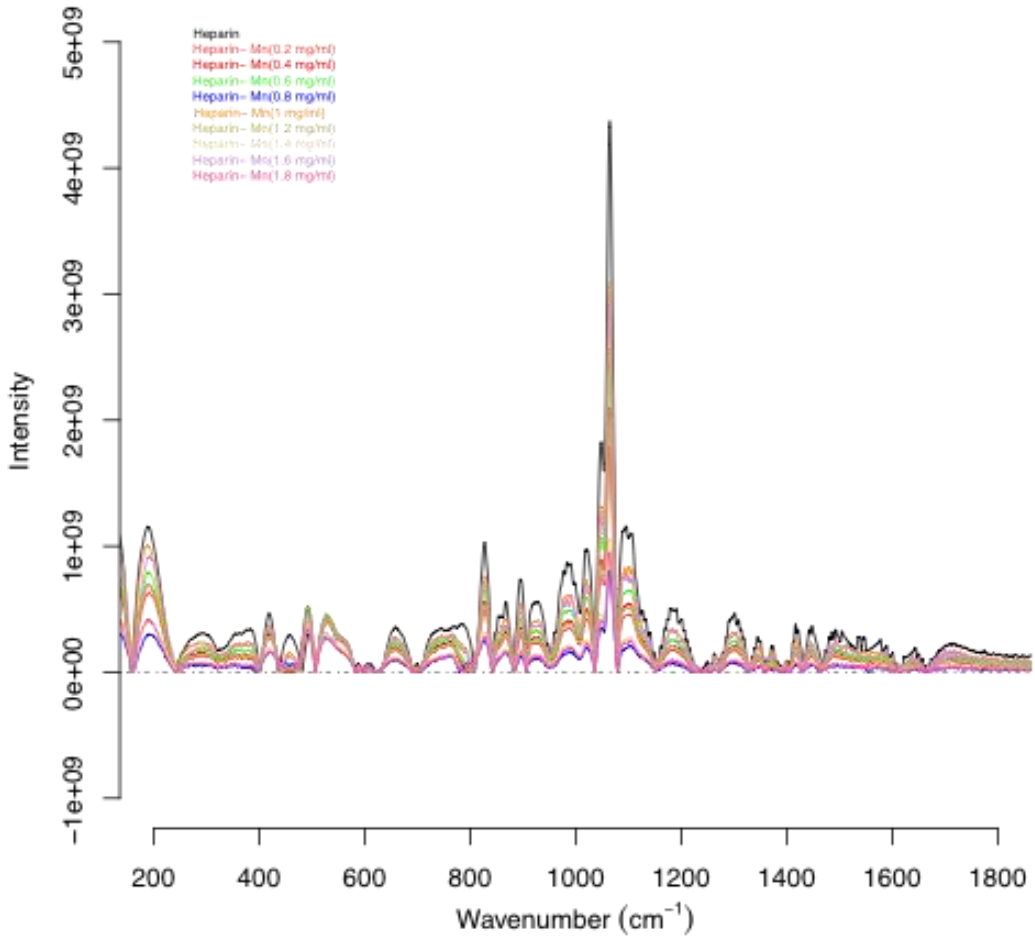

Figure 7- 8: ROA spectra of a titration of Mn^{2+} (0 – 1.8 mg /ml) in 0.2 mg/ml steps into a heparin solution (300 mg/ml) in 90% H_2O and 10% D_2O, recorded for 18300 scans, 1.029 s exposure per scan and 0.30 W laser strength (532 nm). All samples have been spun for 10 minutes at 30k rcf and pre-illuminated by a full strength laser for 30 minutes prior to recording. Data analysis method outlined in Figures 7-3 – 7-5 were applied to correct for baseline via the use of a two pass filter over 100 data points, and normalised to the average value between 1700 - 1800 cm^{-1}. Then the data were squared and the square root is presented here.

Figure 7-8 shows the Raman spectra after baseline correction and the additional data handling to remove negative signals from the spectra. These data are much improved compared to the original method of handling data (Figure 7-7, upper panel), however, there is still obviously some room for improvement. Namely, the normalisation of the data appears to be incorrect, as the edge of a feature apparent at 1700 cm^{-1} is being used as part of the normalisation. To correct for this it might be necessary to include more wavenumbers to find a truly featureless region at these wavelengths in the Raman.

In this section a technique for handling data from Raman and ROA experiments has been proposed with the purpose of correcting erroneous baselines. While the experimental series itself cannot be used due to a malfunctioning component in the ROA equipment at the time of recording, it can be observed that this method is an improvement upon simply applying a normalisation to the spectra. This data handling technique is not limited to ROA spectra, and can be applied to any similar data sets such as IR spectroscopy. As this methodology smoothes over a large number of data points to create a background file, resolution should not be lost in the deduction of a baseline and therefore maximum information is still maintained for interpretation.

7.2. Development of an improved normalisation method for far-UV SRCD.

Far UVCD requires the normalisation of data to a point at which there is no signal to allow for direct comparisons to be made. The data point used has generally been 260 nm, however, during the period of this body of work it was noted that this initial point has a large error (Table 7-1), which can introduce artifacts into subsequent analysis such as elucidating secondary structure information from spectra. Therefore, alternative to normalisation at 260 nm, the averages between 250 and 255 nm, an area of the spectrum that is still devoid of protein signals in far UVCD. The standard deviations of ten randomly selected spectra from the SRCD work during the PhD project are presented in Table 7-1.

Table 7- 1: A comparison between a selection of 10 spectra taken over the course of this body of work to compare between the error of the original method of normalising data at 260 nm, and the improved method of normalising data to the average of 250 -255 nm. Values given are the SD of the mean CD signal in that sample at the wavelengths stated.

File	Sample	Buffer	Path length	260 nm (SD)	250 – 255 nm (SD)
151212b05	IL-8 1 mg/ml	137 mM NaCl PBS	0.2 mm	1.026	0.424
110513b22	Galectin 0.5 mg/ml	137 mM NaCl PBS	0.1 mm	0.491	0.083
110513b01	Galectin (0.5 mg/ml), “E” (0.5 mg/ml)	137 mM NaCl PBS	0.1 mm	0.360	0.111
110513b26	FGF2 (26 µM)	137 mM NaCl PBS	0.1 mm	0.459	0.101
260514b05	Antithrombin 0.4 mg/ml	12.5 mM NaCl PBS	0.2 mm	0.840	0.046
260514b12	Antithrombin (1.7 mg/ml), NA heparin (1:1)	12.5 mM NaCl PBS	0.2 mm	0.472	0.085
071211b04	HEWL 0.5 mg/ml	12.5 mM NaCl PBS	0.2 mm	0.201	0.096
071211b10	HEWL (0.5 mg/ml), HS (0.08 mg/ml)	12.5 mM NaCl PBS	0.2 mm	0.377	0.064
040512b33	Antithrombin 0.4 mg/ml	12.5 mM NaCl PBS	0.1 mm	0.168	0.044
040512b21	Antithrombin (0.4 mg/ml), UFH (0.8 mg/ml)	12.5 mM NaCl PBS	0.1 mm	0.667	0.063
			Average	0.506	0.112
			SD	0.270	0.112

As shown by Table 7-1, the method of normalising data as used at the beginning of this body of work [142] is less reliable at creating a zero point from which spectra can develop their features. The alteration of this method to one where the average of 250 – 255 nm is used reduced the error incurred by over 75% and has since been used for data handling [141]. This data handling point is a crucial step when processing circular dichroism data as algorithms, such as the one used predominantly for the work within this project [98]. It requires that all spectra are normalised to zero at 260 nm, and any added error when applying this requirement could lead to an amplification of errors when deducing the average secondary structures from spectra.

Chapter 8: Conclusions.

This thesis covers several biological systems relevant to the study of protein:GAG interactions and methods by which their study can be improved. Initially a model amyloid forming protein, HEWL, was studied. It is known that HS is associated with amyloid plaques and promotes their growth [126,127,128]. The result of this initial investigation was then carried forward into the development of a new technique to study Trp:GAG interactions, MCD. This technique has shown itself to be capable of demonstrating direct interactions between Trp residues and the HS ligands with our system, thereby demonstrating strong support for the hypothesized binding site within the substrate binding cleft of HEWL. Another system utilised during this body of work is the much studied AT interaction with the pentasaccharide in combination with various other heparin derivatives and also non-active (against anti factor Xa) SHs that contain the rare 3-O-sulfate, which is often portrayed as the marker for active heparin fragments. Generalized 2D correlation techniques [217] were employed to attempt to increase the amount and quality of information gathered from CD experiments. Separately a simple data handling method for ROA and Raman data has been proposed in order to correct for baseline drifts within a data set to allow for direct comparisons between a series of measurements.

The HEWL complexes, investigated in Chapter 3, were all created at pH 7, in a 12.5 mM NaCl PBS buffer. The aggregation, and destabilisation, of HEWL would be interesting to investigate further and may have implications for potential routes to amyloid formation within physiological conditions. It has been noted that proteins at their pI form amyloid-like aggregates when studied using various techniques

[230,231]. This could be investigated as a possibility within the system used here. We observe the creation of a large quantity of unordered and β-sheet structure (+22 and 27% for HS and HS(Zn) ligands) which, would be consistent with the particles observed previously [230,231]. The pI of HEWL is close to 10.5, far removed from the natural pI. However, the pI of heparin molecules is extremely acidic [148], so combined could conceivably shift the pI of the complex into the physiological range through the covering of charges upon binding and introduction of acidic groups on the GAG. This hypothesis is supported by preliminary data, not shown here, by optical density (600 nm), polarised microscopy, and dynamic light scattering (personal communication, Joseph Holman) of HEWL:heparin complexes over a range of pHs in a universal buffering system.

The proposed binding site for HS on HEWL, from the near-UV data (Chapter 3, Figure 3-2) suggesting an interaction with Trp residues, contains two exposed Trp residues, out of six in total. To investigate this further MCD was utilised, in Chapter 4, as a possible tool for investigating the tryptophan amino acid selectively. This provided evidence of a direct Trp:HS interaction in both cases. The MCD spectra of the two ligands (HS and HS(Zn)) were not statistically different from one another ($p>0.05$). The possibility remains that this technique may be capable not only of confirming an interaction with those amino acids, but also differentiating between two extremely similar interactions, through a more through data collection process, such as more scans (in this work n=4) to increase the confidence of the measurement. Further investigation of the change in spectral shape ~286 nm could be conducted through NMR, and also through steric hindrance of the binding site via indole chemistry adding chemical groups to the indole ring of Trp [150].

Chapter 5 focuses on AT and a range of polysaccharides with varying degrees of anti factor Xa activity, an important function for AT in the blood clotting cascade and the target for several anticoagulant drug therapies. Here it is shown that the classic pentasaccharide sequence, long been held as the exemplar of what is possible with heparin based interactions if the technology and methodology can be developed to sequence heparin/HS sequences to an adequate extent, gives the same secondary structural changes in AT as the non-active heparinoid compounds from the shrimp *L. Vannamei*. These non-active shrimp compounds contain the same 3-O-sulfate in glucosamine thought to be central to heparin activity in AT and also elute from an AT affinity column at similar binding affinities. However, these compounds, which are not active, induce the same secondary structures in the protein implying that, while a valid drug, the pentasaccharide is not the sole sulfation sequence capable of inducing anti Xa activity in AT. Rather, it appears there is much degeneracy between compounds giving a range of activities as required, allowing for a flexible system. This assertation is further supported by the observation of the properties of complexes formed from a set of generic HS chains with FGF that differ only in their cation state, with their overall sulfation remaining the same the change of activity on cation state alone [45,46] as well as observation that even non-native plant polysaccharides can induce activities on some targets [139] is further evidence that HS activity is not solely sulfate sequence based, but a combination of factors.

The stability assay DSF has shown itself to be useful as a quick screen for compounds that may have anticoagulant activity. This is important in the search for new anticoagulant drugs to improve upon Arixtra as this has suffered from some problems such as difficulty in neutralization during surgery. In order to progress the

field a mind-set change is required away from the genetics doctrine of a sequence of repeated chemical units leading to a specific outcome to embrace one of redundancy and subtle changes modifying delicate balances within complex biological systems as the organism requires. This could be achieved through a change in cation state, or the removal of sulfates upon a GAG chain by the sulf enzymes as required.

Chapter 6 takes the degradation data from Chapters 3 and 5 and applies generalised 2D correlation techniques to them. The result of deducting two normalised covariance matrices from one another is that the resolution in each correlation can be increased to 1 nm in the 185-195 nm areas of the spectra. This region is where multiple signals overlap with one another enabling on a simplistic level, the distinction between extremely similar complexes. When these same degradations are mean centered across the two spectra to be compared, all common features are removed revealing the differences between the two. This has generated the useful feature of revealing very subtle band shifting events within the spectra. This could be extremely useful for the observer as it reports whether the signals observed in the initial 2D analysis are reliable independent features that can be deconvoluted and interpreted for secondary structural analysis. With the increased resolution, one idea for future work is to selectively label a protein with a heavy isotope eg ^{13}C, and thereby alter subtly the electronic properties of the atoms, and selectively observe a particular region of a protein. This could be extremely useful, for example in the AT case, as it would be useful to know if the secondary structural changes observed in Chapter 5 occur in the same region of the protein.

Chapter 7 provides an additional data handling technique suitable for any spectral data set, but designed for ROA data. A simple method has been developed to correct for baseline drift through a spectrum to allow for comparisons to be made throughout an experimental series recorded in ROA and Raman. The methods need to be improved, especially for the Raman correction as the final square and square root step (See Figure 7-8 of Chapter 7) introduces some artifacts in the removal of negative features, a Pareto scaling method, or some other form of weighted correction would greatly improve this correction method.

Overall, the future of CD should be considered fairly bright if used, as it should be, in conjunction with complementary techniques. With the new data analysis methods it is now possible to gain high-resolution information on data series. Here, degradation data has been used but any series of data such as a ligand titration would be perfectly acceptable. This enables the deconvolution of the information rich 185-195 nm region of the spectrum where many contributions overlap and could improve greatly the reliability of the structural information gained if used together with high quality algorithms. There is also the possibility with this higher resolution to selectively look at a particular region of the protein through labels that effect the electronic configuration of the protein and therefore the absorption properties. Away from CD, ROA is another promising technique that unfortunately was unable to be investigated further during the course of this PhD, but with its capacity to prove GAG conformation and helicity in solution could go some way to building an appreciation for the requirements between active and non-active compounds. This would be an ideal opportunity to utilise generalised 2D correlation techniques to

heterocorrelate between the unassigned ROA spectra and the previously assigned NMR spectra, thus assigning changes observed in one to known chemical groups.

In conclusion, CD and other techniques reported in this thesis offer additional capabilities for the study of protein:GAG complexes in solution.

References:

[1] J.M. Berg, J.L. Tymoczko, L. Stryer, L. Stryer, Biochemistry, 5th ed., W.H. Freeman, New York, 2002.
[2] R.A. van den Berg, H.C. Hoefsloot, J.A. Westerhuis, A.K. Smilde, M.J. van der Werf, Centering, scaling, and transformations: improving the biological information content of metabolomics data, BMC Genomics 7 (2006) 142.
[3] J. Filmus, M. Capurro, J. Rast, Glypicans, Genome Biol 9 (2008) 224.
[4] J.E. Scott, Extracellular matrix, supramolecular organisation and shape, J Anat 187 (Pt 2) (1995) 259-269.
[5] T.R. Rudd, E.A. Yates, A highly efficient tree structure for the biosynthesis of heparan sulfate accounts for the commonly observed disaccharides and suggests a mechanism for domain synthesis, Mol Biosyst 8 (2012) 1499-1506.
[6] T.R. Rudd, M.A. Skidmore, M. Guerrini, M. Hricovini, A.K. Powell, G. Siligardi, E.A. Yates, The conformation and structure of GAGs: recent progress and perspectives, Curr Opin Struct Biol 20 (2010) 567-574.
[7] S. Guimond, M. Maccarana, B.B. Olwin, U. Lindahl, A.C. Rapraeger, Activating and inhibitory heparin sequences for FGF-2 (basic FGF). Distinct requirements for FGF-1, FGF-2, and FGF-4, J Biol Chem 268 (1993) 23906-23914.
[8] S.J. Patey, E.A. Edwards, E.A. Yates, J.E. Turnbull, Heparin Derivatives as Inhibitors of BACE-1, the Alzheimer's β-Secretase, with Reduced Activity against Factor Xa and Other Proteases, Journal of Medicinal Chemistry 49 (2006) 6129-6132.
[9] D.J. Carey, Syndecans: multifunctional cell-surface co-receptors, Biochem J 327 (Pt 1) (1997) 1-16.
[10] A. Pestronk, http://neuromuscular.wustl.edu/pathol/diagrams/gangliosides.htm, 2015.
[11] J.T. Gallagher, A. Walker, Molecular distinctions between heparan sulphate and heparin. Analysis of sulphation patterns indicates that heparan sulphate and heparin are separate families of N-sulphated polysaccharides, Biochem. J. 230 (1985) 665-674.
[12] K.G. Jacobsson, U. Lindahl, Degradation of heparin proteoglycan in cultured mouse mastocytoma cells, Biochem. J. 246 (1987) 409-415.
[13] I. Capila, N.S. Gunay, Z. Shriver, G. Venkataraman, Chapter 3 - Methods for Structural Analysis of Heparin and Heparan Sulfate, in: H.G. Garg, R.J. Linhardt, C.A. Hales (Eds.), Chemistry and Biology of Heparin and Heparan Sulfate, Elsevier Science, Amsterdam, 2005, pp. 55-77.
[14] J.T. Gallagher, M. Lyon, W.P. Steward, Structure and function of heparan sulphate proteoglycans, Biochem. J. 236 (1986) 313-325.
[15] U. Lindahl, M. Kusche, K. Lidholt, L.G. Oscarsson, Biosynthesis of heparin and heparan sulfate, Ann N Y Acad Sci 556 (1989) 36-50.
[16] B. Bray, D.A. Lane, J.M. Freyssinet, G. Pejler, U. Lindahl, Anti-thrombin activities of heparin. Effect of saccharide chain length on thrombin inhibition by heparin cofactor II and by antithrombin, Biochem J 262 (1989) 225-232.

[17] J.E. Silbert, Incorporation of 14c and 3h from Nucleotide Sugars into a Polysaccharide in the Presence of a Cell-Free Preparation from Mouse Mast Cell Tumors, J Biol Chem 238 (1963) 3542-3546.
[18] T. Helting, U. Lindahl, Occurrence and biosynthesis of beta-glucuronidic linkages in heparin, J Biol Chem 246 (1971) 5442-5447.
[19] T. Helting, U. Lindahl, Biosynthesis of heparin. I. Transfer of N-acetylglucosamine and glucuronic acid to low-molecular weight heparin fragments, Acta Chem Scand 26 (1972) 3515-3523.
[20] M. Hook, U. Lindahl, A. Hallen, G. Backstrom, Biosynthesis of heparin. Studies on the microsomal sulfation process, J Biol Chem 250 (1975) 6065-6071.
[21] K.J. Bame, R.V. Reddy, J.D. Esko, Coupling of N-deacetylation and N-sulfation in a Chinese hamster ovary cell mutant defective in heparan sulfate N-sulfotransferase, J Biol Chem 266 (1991) 12461-12468.
[22] Z. Wei, S.J. Swiedler, M. Ishihara, A. Orellana, C.B. Hirschberg, A single protein catalyzes both N-deacetylation and N-sulfation during the biosynthesis of heparan sulfate, Proc Natl Acad Sci U S A 90 (1993) 3885-3888.
[23] I. Jacobsson, U. Lindahl, Biosynthesis of heparin. Concerted action of late polymer-modification reactions, J Biol Chem 255 (1980) 5094-5100.
[24] A. Hagner-Mcwhirter, U. Lindahl, J. Li, Biosynthesis of heparin/heparan sulphate: mechanism of epimerization of glucuronyl C-5, Biochem J 347 Pt 1 (2000) 69-75.
[25] J. Riesenfeld, M. Hook, U. Lindahl, Biosynthesis of heparin. Assay and properties of the microsomal N-acetyl-D-glucosaminyl N-deacetylase, J Biol Chem 255 (1980) 922-928.
[26] P. Carlsson, J. Presto, D. Spillmann, U. Lindahl, L. Kjellen, Heparin/heparan sulfate biosynthesis: processive formation of N-sulfated domains, J Biol Chem 283 (2008) 20008-20014.
[27] M. Lyon, J.T. Gallagher, Bio-specific sequences and domains in heparan sulphate and the regulation of cell growth and adhesion, Matrix Biol 17 (1998) 485-493.
[28] J. Riesenfeld, M. Hozok, U. Lindahl, Biosynthesis of heparan sulfate in rat liver. Characterization of polysaccharides obtained with intact cells and with a cell-free system, J Biol Chem 257 (1982) 7050-7055.
[29] T.R. Rudd, M.A. Skidmore, S.E. Guimond, C. Cosentino, G. Torri, D.G. Fernig, R.M. Lauder, M. Guerrini, E.A. Yates, Glycosaminoglycan origin and structure revealed by multivariate analysis of NMR and CD spectra, Glycobiology 19 (2009) 52-67.
[30] E.A. Yates, F. Santini, M. Guerrini, A. Naggi, G. Torri, B. Casu, 1H and 13C NMR spectral assignments of the major sequences of twelve systematically modified heparin derivatives, Carbohydr Res 294 (1996) 15-27.
[31] S. Yamada, Y. Yamane, H. Tsuda, K. Yoshida, K. Sugahara, A major common trisulfated hexasaccharide core sequence, hexuronic acid(2-sulfate)-glucosamine(N-sulfate)-iduronic acid-N-acetylglucosamine-glucuronic acid-glucosamine(N-sulfate), isolated from the low sulfated irregular region of porcine intestinal heparin, J Biol Chem 273 (1998) 1863-1871.

[32] L. Jin-Ping, G. Feng, D. Kamel El, J. Markku, L. Ulf, Characterization of the d-Glucuronyl C5-epimerase Involved in the Biosynthesis of Heparin and Heparan Sulfate, Journal of Biological Chemistry 276 (2001) 20069-20077.

[33] J. Jia, M. Maccarana, X. Zhang, M. Bespalov, U. Lindahl, J.P. Li, Lack of L-iduronic acid in heparan sulfate affects interaction with growth factors and cell signaling, J Biol Chem 284 (2009) 15942-15950.

[34] W.C. Lamanna, M.A. Frese, M. Balleininger, T. Dierks, Sulf loss influences N-, 2-O-, and 6-O-sulfation of multiple heparan sulfate proteoglycans and modulates fibroblast growth factor signaling, J Biol Chem 283 (2008) 27724-27735.

[35] J. Turnbull, A. Powell, S. Guimond, Heparan sulfate: decoding a dynamic multifunctional cell regulator, Trends Cell Biol 11 (2001) 75-82.

[36] A.K. Powell, E.A. Yates, D.G. Fernig, J.E. Turnbull, Interactions of heparin/heparan sulfate with proteins: appraisal of structural factors and experimental approaches, Glycobiology 14 (2004) 17R-30R.

[37] I. Capila, R.J. Linhardt, Heparin-protein interactions, Angew Chem Int Ed Engl 41 (2002) 391-412.

[38] U. Lindahl, G. Backstrom, L. Thunberg, I.G. Leder, Evidence for a 3-O-sulfated D-glucosamine residue in the antithrombin-binding sequence of heparin, Proc Natl Acad Sci U S A 77 (1980) 6551-6555.

[39] U. Lindahl, Heparan sulfate-protein interactions - A concept for drug design?, Thrombosis and Haemostasis 98 (2007) 109-115.

[40] S. Matou, S. Colliec-Jouault, I. Galy-Fauroux, J. Ratiskol, C. Sinquin, J. Guezennec, A.-M. Fischer, D. Helley, Effect of an oversulfated exopolysaccharide on angiogenesis induced by fibroblast growth factor-2 or vascular endothelial growth factor in vitro, Biochemical Pharmacology 69 (2005) 751-759.

[41] L. Borsig, L. Wang, M.C.M. Cavalcante, L. Cardilo-Reis, P.L. Ferreira, P.A.S. Mourão, J.D. Esko, M.S.G. Pavão, Selectin blocking activity of a fucosylated chondroitin sulfate glycosaminoglycan from sea cucumber: Effect on tumor metastasis and neutrophil recruitment, Journal of Biological Chemistry 282 (2007) 14984-14991.

[42] C.R. Parish, C. Freeman, K.J. Brown, D.J. Francis, W.B. Cowden, Identification of sulfated oligosaccharide-based inhibitors of tumor growth and metastasis using novel in vitro assays for angiogenesis and heparanase activity, Cancer Research 59 (1999) 3433-3441.

[43] J.A. Deakin, B.S. Blaum, J.T. Gallagher, D. Uhrín, M. Lyo, The binding properties of minimal oligosaccharides reveal a common heparan sulfate/dermatan sulfate-binding site in hepatocyte growth factor/scatter factor that can accommodate a wide variety of sulfation patterns, Journal of Biological Chemistry 284 (2009) 6311-6321.

[44] M. Guerrini, S. Elli, P. Mourier, T.R. Rudd, D. Gaudesi, B. Casu, C. Boudier, G. Torri, C. Viskov, An unusual antithrombin-binding heparin octasaccharide with an additional 3-O-sulfated glucosamine in the active pentasaccharide sequence, Biochem J 449 (2013) 343-351.

[45] T.R. Rudd, S.E. Guimond, M.A. Skidmore, L. Duchesne, M. Guerrini, G. Torri, C. Cosentino, A. Brown, D.T. Clarke, J.E. Turnbull, D.G. Fernig, E.A. Yates,

Influence of substitution pattern and cation binding on conformation and activity in heparin derivatives, Glycobiology 17 (2007) 983-993.
[46] S.E. Guimond, T.R. Rudd, M.A. Skidmore, A. Ori, D. Gaudesi, C. Cosentino, M. Guerrini, R. Edge, D. Collison, E. McInnes, G. Torri, J.E. Turnbull, D.G. Fernig, E.A. Yates, Cations modulate polysaccharide structure to determine FGF-FGFR signaling: a comparison of signaling and inhibitory polysaccharide interactions with FGF-1 in solution, Biochemistry 48 (2009) 4772-4779.
[47] T.R. Rudd, E.A. Yates, Conformational degeneracy restricts the effective information content of heparan sulfate, Mol Biosyst 6 (2010) 902-908.
[48] V. Tiwari, C. O'Donnell, R.J. Copeland, T. Scarlett, J. Liu, D. Shukla, Soluble 3-O-sulfated heparan sulfate can trigger herpes simplex virus type 1 entry into resistant Chinese hamster ovary (CHO-K1) cells, J Gen Virol 88 (2007) 1075-1079.
[49] B. Lindahl, C. Westling, G. Giménez-Gallego, U. Lindahl, M. Salmivirta, Common binding sites for β-amyloid fibrils and fibroblast growth factor-2 in heparan sulfate from human cerebral cortex, Journal of Biological Chemistry 274 (1999) 30631-30635.
[50] E.A. Yates, S.E. Guimond, J.E. Turnbull, Highly diverse heparan sulfate analogue libraries: providing access to expanded areas of sequence space for bioactivity screening, J Med Chem 47 (2004) 277-280.
[51] A.N. Alexopoulou, H.A. Multhaupt, J.R. Couchman, Syndecans in wound healing, inflammation and vascular biology, Int J Biochem Cell Biol 39 (2007) 505-528.
[52] S. Saunders, S. Paine-Saunders, A.D. Lander, Expression of the cell surface proteoglycan glypican-5 is developmentally regulated in kidney, limb, and brain, Dev Biol 190 (1997) 78-93.
[53] M.A. Nugent, J. Zaia, J.L. Spencer, Heparan sulfate-protein binding specificity, Biochemistry (Mosc) 78 (2013) 726-735.
[54] A. Ori, M.C. Wilkinson, D.G. Fernig, A systems biology approach for the investigation of the heparin/heparan sulfate interactome, J Biol Chem 286 (2011) 19892-19904.
[55] A. Ori, M.C. Wilkinson, D.G. Fernig, The heparanome and regulation of cell function: structures, functions and challenges, Front Biosci 13 (2008) 4309-4338.
[56] Y. Chen, M. Gotte, J. Liu, P.W. Park, Microbial subversion of heparan sulfate proteoglycans, Mol Cells 26 (2008) 415-426.
[57] N. Perrimon, M. Bernfield, Specificities of heparan sulphate proteoglycans in developmental processes, Nature 404 (2000) 725-728.
[58] C.R. Parish, The role of heparan sulphate in inflammation, Nat Rev Immunol 6 (2006) 633-643.
[59] T.M. Handel, Z. Johnson, S.E. Crown, E.K. Lau, A.E. Proudfoot, Regulation of protein function by glycosaminoglycans--as exemplified by chemokines, Annu Rev Biochem 74 (2005) 385-410.
[60] K.A. Uniewicz, D.G. Fernig, Neuropilins: a versatile partner of extracellular molecules that regulate development and disease, Front Biosci 13 (2008) 4339-4360.
[61] T.Y. Belenkaya, C. Han, D. Yan, R.J. Opoka, M. Khodoun, H. Liu, X. Lin, Drosophila Dpp morphogen movement is independent of dynamin-

mediated endocytosis but regulated by the glypican members of heparan sulfate proteoglycans, Cell 119 (2004) 231-244.

[62] T.M. Ritty, T.J. Broekelmann, C.C. Werneck, R.P. Mecham, Fibrillin-1 and -2 contain heparin-binding sites important for matrix deposition and that support cell attachment, Biochem J 375 (2003) 425-432.

[63] H. Yu, E.M. Munoz, R.E. Edens, R.J. Linhardt, Kinetic studies on the interactions of heparin and complement proteins using surface plasmon resonance, Biochim Biophys Acta 1726 (2005) 168-176.

[64] S. Guglier, M. Hricovini, R. Raman, L. Polito, G. Torri, B. Casu, R. Sasisekharan, M. Guerrini, Minimum FGF2 binding structural requirements of heparin and heparan sulfate oligosaccharides as determined by NMR spectroscopy, Biochemistry 47 (2008) 13862-13869.

[65] M. Guerrini, T. Agulles, A. Bisio, M. Hricovini, L. Lay, A. Naggi, L. Poletti, L. Sturiale, G. Torri, B. Casu, Minimal heparin/heparan sulfate sequences for binding to fibroblast growth factor-1, Biochem Biophys Res Commun 292 (2002) 222-230.

[66] B.S. Blaum, J.A. Deakin, C.M. Johansson, A.P. Herbert, P.N. Barlow, M. Lyon, D. Uhrin, Lysine and arginine side chains in glycosaminoglycan-protein complexes investigated by NMR, cross-linking, and mass spectrometry: a case study of the factor H-heparin interaction, J Am Chem Soc 132 (2010) 6374-6381.

[67] D. Mikhailov, K.H. Mayo, I.R. Vlahov, T. Toida, A. Pervin, R.J. Linhardt, NMR solution conformation of heparin-derived tetrasaccharide, Biochem J 318 (Pt 1) (1996) 93-102.

[68] T.R. Rudd, M.A. Skidmore, S.E. Guimond, M. Guerrini, C. Cosentino, R. Edge, A. Brown, D.T. Clarke, G. Torri, J.E. Turnbull, R.J. Nichols, D.G. Fernig, E.A. Yates, Site-specific interactions of copper(II) ions with heparin revealed with complementary (SRCD, NMR, FTIR and EPR) spectroscopic techniques, Carbohydr Res 343 (2008) 2184-2193.

[69] T.R. Rudd, D. Gaudesi, M.A. Skidmore, M. Ferro, M. Guerrini, B. Mulloy, G. Torri, E.A. Yates, Construction and use of a library of bona fide heparins employing 1H NMR and multivariate analysis, Analyst 136 (2011) 1380-1389.

[70] G. Gatti, B. Casu, G.K. Hamer, A.S. Perlin, Studies on the Conformation of Heparin by 1H and 13C NMR Spectroscopy, Macromolecules 12 (1979) 1001-1007.

[71] K.E. Kövér, P. Groves, J. Jiménez-Barbero, G. Batta, Molecular recognition and screening using a 15N group selective STD NMR method, Journal of the American Chemical Society 129 (2007) 11579-11582.

[72] Z. Zhang, S.A. McCallum, J. Xie, L. Nieto, F. Corzana, J. Jiménez-Barbero, M. Chen, J. Liu, R.J. Linhardt, Solution structures of chemoenzymatically synthesized heparin and its precursors, Journal of the American Chemical Society 130 (2008) 12998-13007.

[73] S.M. Kelly, N.C. Price, The use of circular dichroism in the investigation of protein structure and function, Current Protein & Peptide Science 1 (2000) 349-384.

[74] S. Khan, J. Gor, B. Mulloy, S.J. Perkins, Semi-Rigid Solution Structures of Heparin by Constrained X-ray Scattering Modelling: New Insight into

Heparin-Protein Complexes, Journal of Molecular Biology 395 (2010) 504-521.

[75] G.M. Holder, A. Bowfield, M. Surman, M. Suepfle, D. Moss, C. Tucker, T.R. Rudd, D.G. Fernig, E.A. Yates, P. Weightman, Fundamental differences in model cell-surface polysaccharides revealed by complementary optical and spectroscopic techniques, Soft Matter 8 (2012) 6521-6527.

[76] W. Zhang, M. Heil, R.J. Kuhn, T.S. Baker, Heparin binding sites on Ross River virus revealed by electron cryo-microscopy, Virology 332 (2005) 511-518.

[77] F. Zhang, J. Aguilera, J.M. Beaudet, Q. Xie, T.F. Lerch, O. Davulcu, W. Colon, M.S. Chapman, R.J. Linhardt, Characterization of interactions between heparin/glycosaminoglycan and adeno-associated virus, Biochemistry 52 (2013) 6275-6285.

[78] J.L. Jimenez, G. Tennent, M. Pepys, H.R. Saibil, Structural diversity of ex vivo amyloid fibrils studied by cryo-electron microscopy, J Mol Biol 311 (2001) 241-247.

[79] T.R. Rudd, R.J. Nichols, E.A. Yates, Selective detection of protein secondary structural changes in solution protein-polysaccharide complexes using vibrational circular dichroism (VCD), J Am Chem Soc 130 (2008) 2138-2139.

[80] B. Casu, P. Oreste, G. Torri, G. Zoppetti, J. Choay, J.C. Lormeau, M. Petitou, P. Sinay, The Structure of Heparin Oligosaccharide Fragments with High Anti-(Factor-Xa) Activity Containing the Minimal Antithrombin-Iii-Binding Sequence - Chemical and C-13 Nmr-Studies, Biochemical Journal 197 (1981) 599-609.

[81] Y. Seo, M.R. Schenauer, J.A. Leary, Biologically Relevant Metal-Cation Binding Induces Conformational Changes in Heparin Oligosaccharides as Measured by Ion Mobility Mass Spectrometry, Int J Mass Spectrom 303 (2011) 191-198.

[82] A.F. Bell, L. Hecht, L.D. Barren, Polysaccharide vibrational Raman optical activity: Laminarin and pullulan, Journal of Raman Spectroscopy 26 (1995) 1071-1074.

[83] F.H. Niesen, H. Berglund, M. Vedadi, The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability, Nat Protoc 2 (2007) 2212-2221.

[84] K.A. Uniewicz, A. Ori, R. Xu, Y. Ahmed, M.C. Wilkinson, D.G. Fernig, E.A. Yates, Differential scanning fluorimetry measurement of protein stability changes upon binding to glycosaminoglycans: a screening test for binding specificity, Anal Chem 82 (2010) 3796-3802.

[85] K.A. Uniewicz, A. Ori, T.R. Rudd, M. Guerrini, M.C. Wilkinson, D.G. Fernig, E.A. Yates, Following protein-glycosaminoglycan polysaccharide interactions with differential scanning fluorimetry, Methods Mol Biol 836 (2012) 171-182.

[86] G. Holzwarth, P. Doty, The Ultraviolet Circular Dichroism of Polypeptides, J Am Chem Soc 87 (1965) 218-228.

[87] S. Beychok, Circular dichroism of biological macromolecules, Science 154 (1966) 1288-1299.

[88] N.J. Greenfield, Applications of circular dichroism in protein and peptide analysis, TrAC Trends in Analytical Chemistry 18 (1999) 236-244.

[89] K. Matsuo, Y. Sakurada, R. Yonehara, M. Kataoka, K. Gekko, Secondary-structure analysis of denatured proteins by vacuum-ultraviolet circular dichroism spectroscopy, Biophys J 92 (2007) 4088-4096.
[90] B.A. Wallace, R.W. Janes, Synchrotron radiation circular dichroism (SRCD) spectroscopy: an enhanced method for examining protein conformations and protein interactions, Biochem Soc Trans 38 (2010) 861-873.
[91] M.C. Manning, R.W. Woody, Theoretical study of the contribution of aromatic side chains to the circular dichroism of basic bovine pancreatic trypsin inhibitor, Biochemistry 28 (1989) 8609-8613.
[92] A. Chakrabartty, T. Kortemme, S. Padmanabhan, R.L. Baldwin, Aromatic side-chain contribution to far-ultraviolet circular dichroism of helical peptides and its effect on measurement of helix propensities, Biochemistry 32 (1993) 5560-5565.
[93] C. Krittanai, W.C. Johnson, Correcting the circular dichroism spectra of peptides for contributions of absorbing side chains, Anal Biochem 253 (1997) 57-64.
[94] N. Greenfield, G.D. Fasman, Computed circular dichroism spectra for the evaluation of protein conformation, Biochemistry 8 (1969) 4108-4116.
[95] M. Kajiyoshi, F.K. Anan, Conformation of biological macromolecules. Circular dichroism and magnetic circular dichroism studies of metmyoglobin and its derivatives, J Biochem 78 (1975) 1087-1095.
[96] S. Venyaminov, K.S. Vassilenko, Determination of protein tertiary structure class from circular dichroism spectra, Anal Biochem 222 (1994) 176-184.
[97] N. Sreerama, R.W. Woody, Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set, Anal Biochem 287 (2000) 252-260.
[98] G. Bohm, R. Muhr, R. Jaenicke, Quantitative analysis of protein far UV circular dichroism spectra by neural networks, Protein Eng 5 (1992) 191-195.
[99] V. Raussens, J.-M. Ruysschaert, E. Goormaghtigh, Protein concentration is not an absolute prerequisite for the determination of secondary structure from circular dichroism spectra: a new scaling method, Analytical Biochemistry 319 (2003) 114-121.
[100] K.O. Lloyd, S. Beychok, E.A. Kabat, Immunochemical studies on blood groups. XXXIX. Optical rotatory dispersion and circular dichroism spectra of oligosaccharides from blood-group Lewis substance, Biochemistry 7 (1968) 3762-3765.
[101] J. Clark Sutherland, E.J. Desmond, P.Z. Takacs, Versatile spectrometer for experiments using synchrotron radiation at wave-lengths greater than 100 nm, Nuclear Instruments and Methods 172 (1980) 195-199.
[102] P.A. Snyder, E.M. Rowe, The first use of synchrotron radiation for vacuum ultraviolet circular dichroism measurements, Nuclear Instruments and Methods 172 (1980) 345-349.
[103] J.C. Sutherland, Circular Dichroism and the Conformational Analysis of Biomolecules Edited by Gerald D. Fasman (Brandeis University). Plenum Press: New York. 1996. x + 738 pp. $125.00. ISBN 0-306-45142-5, Journal of the American Chemical Society 118 (1996) 599-633.

[104] R. Hussain, T. Javorfi, G. Siligardi, Circular dichroism beamline B23 at the Diamond Light Source, Journal of Synchrotron Radiation 19 (2012) 132-135.
[105] A. Toumadje, S.W. Alcorn, W.C. Johnson, Jr., Extending CD spectra of proteins to 168 nm improves the analysis for secondary structures, Anal Biochem 200 (1992) 321-331.
[106] K. Matsuo, R. Yonehara, K. Gekko, Secondary-structure analysis of proteins by vacuum-ultraviolet circular dichroism spectroscopy, J Biochem 135 (2004) 405-411.
[107] B.A. Wallace, R.W. Janes, Synchrotron radiation circular dichroism spectroscopy of proteins: secondary structure, fold recognition and structural genomics, Current Opinion in Chemical Biology 5 (2001) 567-571.
[108] G. Siligardi, R. Hussain, D. Myatt, T. Jávorfi, B23 circular dichroism beamline at the Diamond Light Source, Diamond Light Source Proceedings 1 (2011) null-null.
[109] A.S. Brito, D.S. Arimateia, L.R. Souza, M.A. Lima, V.O. Santos, V.P. Medeiros, P.A. Ferreira, R.A. Silva, C.V. Ferreira, G.Z. Justo, E.L. Leite, G.P. Andrade, F.W. Oliveira, H.B. Nader, S.F. Chavante, Anti-inflammatory properties of a heparin-like glycosaminoglycan with reduced anti-coagulant activity isolated from a marine shrimp, Bioorg Med Chem 16 (2008) 9588-9595.
[110] R.C. Team, R: A language and environment for statistical computing, (2014).
[111] A.J. Trexler, M.R. Nilsson, The formation of amyloid fibrils from proteins in the lysozyme family, Curr Protein Pept Sci 8 (2007) 537-557.
[112] A.K. Buell, A. Dhulesia, M.F. Mossuto, N. Cremades, J.R. Kumita, M. Dumoulin, M.E. Welland, T.P. Knowles, X. Salvatella, C.M. Dobson, Population of nonnative states of lysozyme variants drives amyloid fibril formation, J Am Chem Soc 133 (2011) 7737-7743.
[113] M.B. Pepys, P.N. Hawkins, D.R. Booth, D.M. Vigushin, G.A. Tennent, A.K. Soutar, N. Totty, O. Nguyen, C.C. Blake, C.J. Terry, et al., Human lysozyme gene mutations cause hereditary systemic amyloidosis, Nature 362 (1993) 553-557.
[114] S. Vilasi, R. Sarcina, R. Maritato, A. De Simone, G. Irace, I. Sirangelo, Heparin induces harmless fibril formation in amyloidogenic W7FW14F apomyoglobin and amyloid aggregation in wild-type protein in vitro, PLoS One 6 (2011) e22076.
[115] M. Bucciantini, E. Giannoni, F. Chiti, F. Baroni, L. Formigli, J. Zurdo, N. Taddei, G. Ramponi, C.M. Dobson, M. Stefani, Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases, Nature 416 (2002) 507-511.
[116] H.A. Lashuel, D. Hartley, B.M. Petre, T. Walz, P.T. Lansbury, Jr., Neurodegenerative disease: amyloid pores from pathogenic mutations, Nature 418 (2002) 291.
[117] C.M. Dobson, The structural basis of protein folding and its links with human disease, Philos Trans R Soc Lond B Biol Sci 356 (2001) 133-145.
[118] G. Plakoutsi, F. Bemporad, M. Calamai, N. Taddei, C.M. Dobson, F. Chiti, Evidence for a mechanism of amyloid formation involving molecular

reorganisation within native-like precursor aggregates, J Mol Biol 351 (2005) 910-922.
[119] G. Merlini, V. Bellotti, Lysozyme: a paradigmatic molecule for the investigation of protein structure, function and misfolding, Clin Chim Acta 357 (2005) 168-172.
[120] E. Frare, M.F. Mossuto, P. Polverino de Laureto, M. Dumoulin, C.M. Dobson, A. Fontana, Identification of the core structure of lysozyme amyloid fibrils by proteolysis, J Mol Biol 361 (2006) 551-561.
[121] L.A. Morozova-Roche, J. Zurdo, A. Spencer, W. Noppe, V. Receveur, D.B. Archer, M. Joniau, C.M. Dobson, Amyloid fibril formation and seeding by wild-type human lysozyme and its disease-related mutational variants, J Struct Biol 130 (2000) 339-351.
[122] H. Yagi, E. Kusaka, K. Hongo, T. Mizobata, Y. Kawata, Amyloid fibril formation of alpha-synuclein is accelerated by preformed amyloid seeds of other proteins: implications for the mechanism of transmissible conformational diseases, J Biol Chem 280 (2005) 38609-38616.
[123] N. Motamedi-Shad, E. Monsellier, S. Torrassa, A. Relini, F. Chiti, Kinetic analysis of amyloid formation in the presence of heparan sulfate: faster unfolding and change of pathway, J Biol Chem 284 (2009) 29921-29934.
[124] A.T. Alexandrescu, Amyloid accomplices and enforcers, Protein Sci 14 (2005) 1-12.
[125] J.B. Ancsin, Amyloidogenesis: historical and modern observations point to heparan sulfate proteoglycans as a major culprit, Amyloid 10 (2003) 67-79.
[126] A.D. Snow, R. Kisilevsky, J. Willmer, S.B. Prusiner, S.J. DeArmond, Sulfated glycosaminoglycans in amyloid plaques of prion diseases, Acta Neuropathol 77 (1989) 337-342.
[127] A.D. Snow, R.T. Sekiguchi, D. Nochlin, R.N. Kalaria, K. Kimata, Heparan sulfate proteoglycan in diffuse plaques of hippocampus but not of cerebellum in Alzheimer's disease brain, Am J Pathol 144 (1994) 337-347.
[128] A.D. Snow, H. Mar, D. Nochlin, R.T. Sekiguchi, K. Kimata, Y. Koike, T.N. Wight, Early accumulation of heparan sulfate in neurons and in the beta-amyloid protein-containing lesions of Alzheimer's disease and Down's syndrome, Am J Pathol 137 (1990) 1253-1270.
[129] A. Ori, P. Free, J. Courty, M.C. Wilkinson, D.G. Fernig, Identification of heparin-binding sites in proteins by selective labeling, Mol Cell Proteomics 8 (2009) 2256-2265.
[130] S. Sarrazin, W.C. Lamanna, J.D. Esko, Heparan sulfate proteoglycans, Cold Spring Harb Perspect Biol 3 (2011).
[131] R.F. Parrish, W.R. Fair, Selective binding of zinc ions to heparin rather than to other glycosaminoglycans, Biochem J 193 (1981) 407-410.
[132] B. Mulloy, M.J. Forster, Conformation and dynamics of heparin and heparan sulfate, Glycobiology 10 (2000) 1147-1156.
[133] A.I. Bush, W.H. Pettingell, G. Multhaup, M. d Paradis, J.P. Vonsattel, J.F. Gusella, K. Beyreuther, C.L. Masters, R.E. Tanzi, Rapid induction of Alzheimer A beta amyloid formation by zinc, Science 265 (1994) 1464-1467.

[134] P.W. Mantyh, J.R. Ghilardi, S. Rogers, E. DeMaster, C.J. Allen, E.R. Stimson, J.E. Maggio, Aluminum, iron, and zinc ions promote aggregation of physiological concentrations of beta-amyloid peptide, J Neurochem 61 (1993) 1171-1174.

[135] C.S. Atwood, R.C. Scarpa, X. Huang, R.D. Moir, W.D. Jones, D.P. Fairlie, R.E. Tanzi, A.I. Bush, Characterization of copper interactions with alzheimer amyloid beta peptides: identification of an attomolar-affinity copper binding site on amyloid beta1-42, J Neurochem 75 (2000) 1219-1233.

[136] P. Faller, C. Hureau, Bioinorganic chemistry of copper and zinc ions coordinated to amyloid-beta peptide, Dalton Trans (2009) 1080-1094.

[137] K. Jomova, D. Vondrakova, M. Lawson, M. Valko, Metals, oxidative stress and neurodegenerative disorders, Mol Cell Biochem 345 (2010) 91-104.

[138] W.J. Goux, T.M. Hooker, Jr., The chiroptical properties of proteins. II. Near-ultraviolet circular dichroism of lysozyme, Biopolymers 19 (1980) 2191-2208.

[139] T.R. Rudd, K.A. Uniewicz, A. Ori, S.E. Guimond, M.A. Skidmore, D. Gaudesi, R. Xu, J.E. Turnbull, M. Guerrini, G. Torri, G. Siligardi, M.C. Wilkinson, D.G. Fernig, E.A. Yates, 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 (2010) 5390-5397.

[140] R. Xu, A. Ori, T.R. Rudd, K.A. Uniewicz, Y.A. Ahmed, S.E. Guimond, M.A. Skidmore, G. Siligardi, E.A. Yates, D.G. Fernig, Diversification of the structural determinants of fibroblast growth factor-heparin interactions: implications for binding specificity, J Biol Chem 287 (2012) 40061-40073.

[141] M.A. Lima, A.J. Hughes, N. Veraldi, T.R. Rudd, R. Hussain, A.S. Brito, S.F. Chavante, I.I. Tersariol, G. Siligardi, H.B. Nader, E.A. Yates, Antithrombin stabilisation by sulfated carbohydrates correlates with anticoagulant activity, Medchemcomm 4 (2013) 870-873.

[142] A.J. Hughes, R. Hussain, C. Cosentino, M. Guerrini, G. Siligardi, E.A. Yates, T.R. Rudd, A zinc complex of heparan sulfate destabilises lysozyme and alters its conformation, Biochem Biophys Res Commun 425 (2012) 794-799.

[143] C.R. Beddell, C.C. Blake, S.J. Oatley, An x-ray study of the structure and binding properties of iodine-inactivated lysozyme, J Mol Biol 97 (1975) 643-654.

[144] N. Sanchez de Groot, I. Pallares, F.X. Aviles, J. Vendrell, S. Ventura, Prediction of "hot spots" of aggregation in disease-linked polypeptides, BMC Struct Biol 5 (2005) 18.

[145] S. Goda, K. Takano, Y. Yamagata, R. Nagata, H. Akutsu, S. Maki, K. Namba, K. Yutani, Amyloid protofilament formation of hen egg lysozyme in highly concentrated ethanol solution, Protein Sci 9 (2000) 369-375.

[146] M. Hirai, M. Koizumi, T. Hayakawa, H. Takahashi, S. Abe, H. Hirai, K. Miura, K. Inoue, Hierarchical map of protein unfolding and refolding at thermal equilibrium revealed by wide-angle X-ray scattering, Biochemistry 43 (2004) 9036-9049.

[147] T. Imoto, S. Moriyama, K. Yagishita, A study of the native-denatured (N in equilibrium with D) transition in lysozyme. III. Effect of alteration of net charge by acetylation, J Biochem 80 (1976) 1319-1325.
[148] J. Holman, M.A. Skidmore, T.R. Rudd, E.A. Yates, The latent ampholytic nature of glycosaminoglycan (GAG) oligosaccharides facilitates their separation by isoelectric focusing, Analytical Methods 2 (2010) 1550-1554.
[149] L.R. Wetter, H.F. Deutsch, Immunological studies on egg white proteins. IV. Immunochemical and physical studies of lysozyme, J Biol Chem 192 (1951) 237-242.
[150] C.L. Ladner, R.J. Turner, R.A. Edwards, Development of indole chemistry to label tryptophan residues in protein for determination of tryptophan surface accessibility, Protein Sci 16 (2007) 1204-1213.
[151] Buckingh.Ad, P.J. Stephens, Magnetic Optical Activity, Annual Review of Physical Chemistry 17 (1966) 399-&.
[152] M. Faraday, Faraday's diary of experimental investigation, in: M. Thomas (Ed.), Faraday's diary of experimental investigation, G. Bell and Sons, London, 1933.
[153] T.M. McFarland, J.E. Coleman, Magnetic circular dichroism of tryptophanyl residues in proteins. Azurin and carbonic anhydrases, Eur J Biochem 29 (1972) 521-527.
[154] B. Holmquist, B.L. Vallee, Tryptophan quantitation by magnetic circular dichroism in native and modified proteins, Biochemistry 12 (1973) 4409-4417.
[155] J.C. Sutherland, B. Holmquist, Magnetic circular dichroism of biological molecules, Annu Rev Biophys Bioeng 9 (1980) 293-326.
[156] B. Holmquist, Magnetic circular dichroism, Methods Enzymol 130 (1986) 270-290.
[157] G. Barth, W. Voelter, E. Bunnenberg, C. Djerassi, Magnetic circular dichroism studies. XVII. Magnetic circular dichroism spectra of proteins. A new method for the quantitative determination of tryptophan, J Am Chem Soc 94 (1972) 1293-1298.
[158] G. Barth, R. Records, E. Bunnenberg, C. Djerassi, W. Voelter, Magnetic circular dichroism studies. XII. The determination of tryptophan in proteins, J Am Chem Soc 93 (1971) 2545-2547.
[159] Holmquis.B, Magnetic Circular Dichroic Spectra of Tryptophan Containing Proteins, Federation Proceedings 30 (1971) 1178-&.
[160] G. Barth, E. Bunnenberg, C. Djerassi, Magnetic circular dichroism studies. XIX. Determination of the tyrosine: tryptophan ratio in proteins, Anal Biochem 48 (1972) 471-479.
[161] D.J. Vitale, R.A. Goldbeck, D.B. Kim-Shapiro, R.M. Esquerra, L.J. Parkhurst, D.S. Kliger, Near-ultraviolet magnetic circular dichroism spectroscopy of protein conformational states: correlation of tryptophan band position and intensity with hemoglobin allostery, Biochemistry 39 (2000) 7145-7152.
[162] R.A. Goldbeck, R.M. Esquerra, D.S. Kliger, Hydrogen Bonding to Trp β37 Is the First Step in a Compound Pathway for Hemoglobin Allostery, Journal of the American Chemical Society 124 (2002) 7646-7647.

[163] G. Zoppellaro, K.L. Bren, A.A. Ensign, E. Harbitz, R. Kaur, H.P. Hersleth, U. Ryde, L. Hederstedt, K.K. Andersson, Review: studies of ferric heme proteins with highly anisotropic/highly axial low spin (S = 1/2) electron paramagnetic resonance signals with bis-histidine and histidine-methionine axial iron coordination, Biopolymers 91 (2009) 1064-1082.
[164] M.S. Ali, B.C. Shenoy, D. Eswaran, L.A. Andersson, T.E. Roche, M.S. Patel, Identification of the tryptophan residue in the thiamin pyrophosphate binding site of mammalian pyruvate dehydrogenase, J Biol Chem 270 (1995) 4570-4574.
[165] D.D. Ulmer, B. Holmquist, B.L. Vallee, Magnetic circular dichroism of non-heme iron proteins, Biochem Biophys Res Commun 51 (1973) 1054-1061.
[166] R. Xu, T.R. Rudd, A.J. Hughes, G. Siligardi, D.G. Fernig, E.A. Yates, Analysis of the fibroblast growth factor receptor (FGFR) signalling network with heparin as coreceptor: evidence for the expansion of the core FGFR signalling network, FEBS J 280 (2013) 2260-2270.
[167] S. Antonyuk, R.W. Strange, S.S. Hasnain, Structural discovery of small molecule binding sites in Cu-Zn human superoxide dismutase familial amyotrophic lateral sclerosis mutants provides insights for lead optimization, J Med Chem 53 (2010) 1402-1406.
[168] G.S. Wright, S.V. Antonyuk, N.M. Kershaw, R.W. Strange, S. Samar Hasnain, Ligand binding and aggregation of pathogenic SOD1, Nat Commun 4 (2013) 1758.
[169] L.E. Franzen, S. Svensson, O. Larm, Structural studies on the carbohydrate portion of human antithrombin III, J Biol Chem 255 (1980) 5090-5093.
[170] G. Murano, L. Williams, M. Miller-Andersson, D.L. Aronson, C. King, Some properties of antithrombin-III and its concentration in human plasma, Thromb Res 18 (1980) 259-262.
[171] R.D. Rosenberg, K.A. Bauer, J.A. Marcum, Antithrombin - Iii - the Heparin - Antithrombin System, Journal of Medicine 16 (1985) 351-416.
[172] J. Travis, G.S. Salvesen, Human-Plasma Proteinase-Inhibitors, Annual Review of Biochemistry 52 (1983) 655-709.
[173] D. Menache, B.J. Grossman, C.M. Jackson, Antithrombin-Iii - Physiology, Deficiency, and Replacement Therapy, Transfusion 32 (1992) 580-588.
[174] D. Menache, Antithrombin-Iii - Introduction, Seminars in Hematology 28 (1991) 1-2.
[175] B. Lahiri, A. Bagdasarian, B. Mitchell, R.C. Talamo, R.W. Colman, R.D. Rosenberg, Antithrombin-Heparin Cofactor - Inhibitor of Plasma Kallikrein, Archives of Biochemistry and Biophysics 175 (1976) 737-747.
[176] K.A. Bauer, R.D. Rosenberg, Role of antithrombin III as a regulator of in vivo coagulation, Semin Hematol 28 (1991) 10-18.
[177] R. Carrell, R. Skinner, L. Jin, J.P. Abrahams, Structural mobility of antithrombin and its modulation by heparin, Thromb Haemost 78 (1997) 516-519.
[178] S.T. Olson, I. Bjork, R. Sheffer, P.A. Craig, J.D. Shore, J. Choay, Role of the antithrombin-binding pentasaccharide in heparin acceleration of antithrombin-proteinase reactions. Resolution of the antithrombin conformational change contribution to heparin rate enhancement, J Biol Chem 267 (1992) 12528-12538.

[179] J. McLean, The Thromboplastic Action of Cephalin, American Journal of Physiology 41 (1916) 250-257.
[180] K.M. Brinkhous, H.P. Smith, E.D. Warner, W.H. Seegers, The inhibition of blood clotting: An unidentified substance which acts in conjunction with heparin to prevent the conversion of prothrombin into thrombin, American Journal of Physiology 125 (1939) 683-687.
[181] F.C. Monkhouse, E.S. France, W.H. Seegers, Studies on the Antithrombin and Heparin Cofactor Activities of a Fraction Adsorbed from Plasma by Aluminum Hydroxide, Circulation Research 3 (1955) 397-402.
[182] D.F. Waugh, M.A. Fitzgerald, Quantitative Aspects of Antithrombin and Heparin in Plasma, American Journal of Physiology 184 (1956) 627-639.
[183] M. Petitou, B. Casu, U. Lindahl, 1976-1983, a critical period in the history of heparin: the discovery of the antithrombin binding site, Biochimie 85 (2003) 83-89.
[184] J.A. Cifonelli, The relationship of molecular weight, and sulfate content and distribution to anticoagulant activity of heparin preparations, Carbohydr Res 37 (1974) 145-154.
[185] L.H. Lam, J.E. Silbert, R.D. Rosenberg, The separation of active and inactive forms of heparin, Biochem Biophys Res Commun 69 (1976) 570-577.
[186] M. Hook, I. Bjork, J. Hopwood, U. Lindahl, Anticoagulant activity of heparin: separation of high-activity and low-activity heparin species by affinity chromatography on immobilized antithrombin, FEBS Lett 66 (1976) 90-93.
[187] J. Hopwood, M. Hook, A. Linker, U. Lindahl, Anticoagulant activity of heparin: isolation of antithrombin-binding sites, FEBS Lett 69 (1976) 51-54.
[188] L.O. Andersson, T.W. Barrowcliffe, E. Holmer, E.A. Johnson, G.E. Sims, Anticoagulant properties of heparin fractionated by affinity chromatography on matrix-bound antithrombin iii and by gel filtration, Thromb Res 9 (1976) 575-583.
[189] T.C. Laurent, A. Tengblad, L. Thunberg, M. Hook, U. Lindahl, The molecular-weight-dependence of the anti-coagulant activity of heparin, Biochem J 175 (1978) 691-701.
[190] R.D. Rosenberg, G. Armand, L. Lam, Structure-function relationships of heparin species, Proc Natl Acad Sci U S A 75 (1978) 3065-3069.
[191] R.D. Rosenberg, L. Lam, Correlation between structure and function of heparin, Proc Natl Acad Sci U S A 76 (1979) 1218-1222.
[192] U. Lindahl, G. Backstrom, M. Hook, L. Thunberg, L.A. Fransson, A. Linker, Structure of the antithrombin-binding site in heparin, Proc Natl Acad Sci U S A 76 (1979) 3198-3202.
[193] L. Thunberg, G. Backstrom, H. Grundberg, J. Riesenfeld, U. Lindahl, The molecular size of the antithrombin-binding sequence in heparin, FEBS Lett 117 (1980) 203-206.
[194] B. Casu, P. Oreste, G. Torri, G. Zoppetti, J. Choay, J.C. Lormeau, M. Petitou, P. Sinay, The structure of heparin oligosaccharide fragments with high anti-(factor Xa) activity containing the minimal antithrombin III-binding sequence. Chemical and 13C nuclear-magnetic-resonance studies, Biochem J 197 (1981) 599-609.

[195] J. Choay, J.C. Lormeau, M. Petitou, Low-Molecular Weight Oligosaccharides Active in Plasma against Factor Xa, Annales Pharmaceutiques Francaises 39 (1981) 37-44.

[196] J. Choay, J.C. Lormeau, M. Petitou, P. Sinay, B. Casu, P. Oreste, G. Torri, G. Gatti, Anti-Xa active heparin oligosaccharides, Thromb Res 18 (1980) 573-578.

[197] B. Meyer, L. Thunberg, U. Lindahl, O. Larm, I.G. Leder, The antithrombin-binding sequence of heparin studied by n.m.r. spectroscopy, Carbohydr Res 88 (1981) C1-4.

[198] J. Choay, J.C. Lormeau, M. Petitou, P. Sinay, J. Fareed, Structural studies on a biologically active hexasaccharide obtained from heparin, Ann N Y Acad Sci 370 (1981) 644-649.

[199] L. Thunberg, G. Backstrom, U. Lindahl, Further characterization of the antithrombin-binding sequence in heparin, Carbohydr Res 100 (1982) 393-410.

[200] P. Duchaussoy, P.S. Lei, M. Petitou, P. Sinay, J.C. Lormeau, J. Choay, The first total synthesis of the antithrombin III binding site of porcine mucosa heparin, Bioorganic & Medicinal Chemistry Letters 1 (1991) 99-102.

[201] J. Choay, M. Petitou, J.C. Lormeau, P. Sinay, B. Casu, G. Gatti, 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 (1983) 492-499.

[202] M. Petitou, P. Duchaussoy, I. Lederman, J. Choay, P. Sinay, Binding of heparin to antithrombin III: a chemical proof of the critical role played by a 3-sulfated 2-amino-2-deoxy-D-glucose residue, Carbohydr Res 179 (1988) 163-172.

[203] D. Vezina, P. Sheridan, R. Blain, K.D. Roberts, G. Bleau, Safety of protamine sulfate administration in vasectomized men, Contraception 41 (1990) 605-616.

[204] H.L. Pachter, T.S. Riles, Low dose heparin: bleeding and wound complications in the surgical patient. A prospective randomized study, Ann Surg 186 (1977) 669-674.

[205] G. Montalescot, U. Zeymer, J. Silvain, B. Boulanger, M. Cohen, P. Goldstein, P. Ecollan, X. Combes, K. Huber, C. Pollack, Jr., J.F. Benezet, O. Stibbe, E. Filippi, E. Teiger, G. Cayla, S. Elhadad, F. Adnet, T. Chouihed, S. Gallula, A. Greffet, M. Aout, J.P. Collet, E. Vicaut, A. Investigators, Intravenous enoxaparin or unfractionated heparin in primary percutaneous coronary intervention for ST-elevation myocardial infarction: the international randomised open-label ATOLL trial, Lancet 378 (2011) 693-703.

[206] E. Puymirat, N. Aissaoui, J.P. Collet, A. Chaib, J.L. Bonnet, V. Bataille, E. Drouet, G. Mulak, J. Ferrieres, D. Blanchard, T. Simon, N. Danchin, Comparison of bleeding complications and one-year survival of low molecular weight heparin versus unfractioned heparin for acute myocardial infarction in elderly patients. The FAST-MI registry, Int J Cardiol 166 (2013) 106-110.

[207] O. Jeong, S.Y. Ryu, Y.K. Park, Y.J. Kim, The effect of low molecular weight heparin thromboprophylaxis on bleeding complications after gastric cancer surgery, Ann Surg Oncol 17 (2010) 2363-2369.

[208] R. Lever, C.P. Page, Novel drug development opportunities for heparin, Nat Rev Drug Discov 1 (2002) 140-148.
[209] N. Afratis, C. Gialeli, D. Nikitovic, T. Tsegenidis, E. Karousou, A.D. Theocharis, M.S. Pavao, G.N. Tzanakakis, N.K. Karamanos, Glycosaminoglycans: key players in cancer cell biology and treatment, FEBS J 279 (2012) 1177-1197.
[210] J.L. Dreyfuss, C.V. Regatieri, M.A. Lima, E.J. Paredes-Gamero, A.S. Brito, S.F. Chavante, R. Belfort, Jr., M.E. Farah, H.B. Nader, A heparin mimetic isolated from a marine shrimp suppresses neovascularization, J Thromb Haemost 8 (2010) 1828-1837.
[211] A.J. McCoy, X.Y. Pei, R. Skinner, J.P. Abrahams, R.W. Carrell, Structure of beta-antithrombin and the effect of glycosylation on antithrombin's heparin affinity and activity, J Mol Biol 326 (2003) 823-833.
[212] S. Frebelius, S. Isaksson, J. Swedenborg, Thrombin inhibition by antithrombin III on the subendothelium is explained by the isoform AT beta, Arterioscler Thromb Vasc Biol 16 (1996) 1292-1297.
[213] R.M. Reynolds, P.L. Padfield, J.R. Seckl, Disorders of sodium balance, BMJ 332 (2006) 702-705.
[214] J. Conard, F. Brosstad, M. Lie Larsen, M. Samama, U. Abildgaard, Molar antithrombin concentration in normal human plasma, Haemostasis 13 (1983) 363-368.
[215] H.A. Schreuder, B. de Boer, R. Dijkema, J. Mulders, H.J. Theunissen, P.D. Grootenhuis, W.G. Hol, The intact and cleaved human antithrombin III complex as a model for serpin-proteinase interactions, Nat Struct Biol 1 (1994) 48-54.
[216] S.F. Chavante, A.S. Brito, M. Lima, E. Yates, H. Nader, M. Guerrini, G. Torri, A. Bisio, A heparin-like glycosaminoglycan from shrimp containing high levels of 3-O-sulfated d-glucosamine groups in an unusual trisaccharide sequence, Carbohydr Res 390C (2014) 59-66.
[217] I. Noda, Y. Ozaki, Two-dimensional correlation spectroscopy: applications in vibrational and optical spectroscopy, John Wiley & Sons, 2005.
[218] J.J. Cael, J.L. Koenig, J. Blackwell, Infrared and raman spectroscopy of carbohydrates : Part IV. Identification of configuration- and conformation-sensitive modes for D-glucose by normal coordinate analysis, Carbohydrate Research 32 (1974) 79-91.
[219] F. Cabassi, B. Casu, A.S. Perlin, Infrared absorption and raman scattering of sulfate groups of heparin and related glycosaminoglycans in aqueous solution, Carbohydrate Research 63 (1978) 1-11.
[220] L.D. Barron, L. Hecht, E.W. Blanch, A.F. Bell, Solution structure and dynamics of biomolecules from Raman optical activity, Prog Biophys Mol Biol 73 (2000) 1-49.
[221] J. Kaminsky, J. Kapitan, V. Baumruk, L. Bednarova, P. Bour, Interpretation of Raman and Raman optical activity spectra of a flexible sugar derivative, the gluconic acid anion, J Phys Chem A 113 (2009) 3594-3601.
[222] S.N. Rashkeev, G. Wendin, Electronic Raman continuum for YBa2Cu3O7-delta : Effects of inelastic scattering and interband transitions, Phys Rev B Condens Matter 47 (1993) 11603-11606.

[223] A.F. Bell, L.D. Barron, L. Hecht, Vibrational Raman optical activity study of d-glucose, Carbohydrate Research 257 (1994) 11-24.
[224] A.F. Bell, L. Hecht, L.D. Barron, Disaccharide Solution Stereochemistry from Vibrational Raman Optical Activity, Journal of the American Chemical Society 116 (1994) 5155-5161.
[225] F. Zhu, N.W. Isaacs, L. Hecht, G.E. Tranter, L.D. Barron, Raman optical activity of proteins, carbohydrates and glycoproteins, Chirality 18 (2006) 103-115.
[226] L.D. Barron, A.R. Gargaro, Z.Q. Wen, Vibrational Raman optical activity of carbohydrates, Carbohydr Res 210 (1991) 39-49.
[227] T.R. Rudd, R. Hussain, G. Siligardi, E.A. Yates, Raman and Raman optical activity of glycosaminoglycans, Chem Commun (Camb) 46 (2010) 4124-4126.
[228] N.R. Yaffe, A. Almond, E.W. Blanch, A new route to carbohydrate secondary and tertiary structure using Raman spectroscopy and Raman optical activity, J Am Chem Soc 132 (2010) 10654-10655.
[229] s. developers, Signal: Signal processing, (2013).
[230] M.R. Krebs, G.L. Devlin, A.M. Donald, Amyloid fibril-like structure underlies the aggregate structure across the pH range for beta-lactoglobulin, Biophys J 96 (2009) 5013-5019.
[231] M.R. Krebs, K.R. Domike, A.M. Donald, Protein aggregation: more than just fibrils, Biochem Soc Trans 37 (2009) 682-686.

Outputs:

First author peer reviewed publications:

- **A zinc complex of heparan sulfate destabilises lysozyme and alters its conformation.**

  Hughes AJ, Hussain R, Cosentino C, Guerrini M, Siligardi G, Yates EA, Rudd TR.

  Biochem Biophys Res Commun. 2012 Sep 7;425(4):794-9.

  DOI: 10.1016/j.bbrc.2012.07.154. .

- **Antithrombin Stabilisation by Sulfated Carbohydrates Correlates with Anticoagulant Activity**

  Marcelo A. Lima, **Ashley J. Hughes,** Noemi Veraldi, Timothy R. Rudd, Rohanah Hussain, Adriana S. Brito, Suely F. Chavante, Ivarne I. Tersariol,Giuliano Siligardi, Helena B. Nader, Edwin A. Yates

  Medicinal Chemistry Communications 04/2013; DOI:10.1039/c0xx00000x

  *Lima & Hughes Joint authors

- **Detection of interaction between protein trytophan residues and small or macromolecular ligands by synchrotron radiation magnetic circular dichroism**

  AJ Hughes, TR Rudd, GSA Wright, DG Fernig, G Siligardi, and E A. Yates

  Anal. Methods, 07/2015;

  DOI: 10.1039/C4AY02618G

Other peer reviewed publications

- **Heparan sulfate, its derivatives and analogues share structural characteristics that can be exploited, particularly in inhibiting microbial attachment.**

  Rudd TR, **Hughes A**, Holman J, Solari V, Ferreira Ede O, Domingues RM, Yates EA.

  Braz J Med Biol Res. 2012 May;45(5):386-91.

- **Analysis of the fibroblast growth factor receptor (FGFR) signalling network with heparin as coreceptor: evidence for the expansion of the core FGFR signalling network.**

  Xu R, Rudd TR, **Hughes AJ**, Siligardi G, Fernig DG, Yates EA.

  FEBS J. 2013 May;280(10):2260-70.

  DOI: 10.1111/febs.12201

- **A Non-Hemorrhagic Hybrid Heparin/Heparan Sulfate with Anticoagulant Potential**

  Adriana S. Brito, Rômulo S. Cavalcante1, Laís C.G.F. Palhares, **Ashley Hughes,** Giulianna P.V. Andrade1, Edwin A. Yates, Helena B. Nader, Marcelo A. Lima, Suely F. Chavante

  Carbohydrate Polymers 99 (2014) 372–378

- **Heparin Derivatives for the Targeting of Multiple Activities in the Inflammatory Response**

  N Veraldi, **AJ Hughes**, TR Rudd, HB Thomas, SW Edwards, L Hadfield, MA Skidmore, G Siligardi, JK Shute, A Naggi and EA Yates

  Accepted to Carbohydrate Polymers – 23rd September 2014