Structural characterisation of human immunoglobulin G Fc fragment and its interaction with heparin

Henry Frederick Earps

Thesis for submission towards the degree of

Master of Philosophy

July 2020

Keele University

Abstract

Human immunoglobulin (Ig) G is the most commonly found class of antibody in the human body and is involved in pathogen recognition as a part of the adaptive and innate immune response. Recognition occurs through the fragment antibody binding (FAB) region and interaction with the innate immune system occurs via interactions between the fragment crystallisable (Fc) region and cell surface Fc receptors as well as select proteins of the complement system. It is this interaction that allows IgG to activate the immune system. The aim of this work was to characterise, at the atomic level, the recognition of heparin analogues using IgG-Fc fragments.

One native crystal structure of purified IgG-Fc and two crystal structures of heparin soaked IgG were determined, one with purified IgG-Fc soaked in heparin of size degree of polymerisation (dp) 12 and a recombinant IgG-Fc soaked in a rough heparin fraction. These structures have been successfully solved in $P2_12_12_1$ using rigid body refinement. The native purified IgG-Fc was refined to 2.30 Å, with refinement statistics: Rwork 0.296 and Rfree 0.356. And the dp12-soaked purified IgG-Fc was refined to 2.15 Å with refinement statistics: Rwork 0.250 and Rfree 0.301. The rough heparin-soaked recombinant IgG-Fc was refined to 3.55 Å with refinement statistics: Rwork 0.325 and Rfree 0.297. Whilst no ligand binding was observed in the heparin soaked structures, the N-linked glycan of the Fc region is clearly defined.

The proportions of IgG subtypes in the human sera sample were also characterised through the use of matrix assisted laser desorption/ionisation (MALDI) and electrospray ionisation (ESI) mass spectrometry. Using these techniques, the sequence coverage of IgG-Fc subtypes in the sample of purified human sera used in this work were found to be: 84.5% IgG1-Fc, 28.2% IgG2-Fc, 12.1% IgG3-Fc and 11.3% IgG4-Fc respectively.

Table of Contents

List of Figures

Chapter 1

Chapter 2

Chapter 3

List of Tables

Chapter 2

Chapter 3

Acknowledgements

I would like to express my deepest gratitude to Dr. Annette Shrive and Prof. Trevor Greenhough for providing me with the excellent opportunity to undertake this research masters and to further develop a range of life skills, which has allowed me to grow personally and professionally.

I would like to thank Dr. Annette Shrive for her unwavering support, gentle guidance and for her patience throughout this process. I would also like to thank her for providing her wealth of knowledge on a wide variety of subjects and much needed advice. I would also like to thank Prof. Trevor Greenhough for his support and for sharing a fraction of his immense knowledge of protein crystallography. Thank you both for your patience and support throughout the years.

I would like to thank the Skidmore research group for providing the rough and dp12 heparin fractions that were the cornerstone of my research. In addition to this, they also provided the initial circular dichroism results that are the source of this undertaking.

I would like to acknowledge Ian Burns for providing his wealth of knowledge both in and out of the laboratory and for undertaking the initial crystal growth trials that provided a starting point for the work described in this thesis. In addition to this, I would like to thank Dr. Sarah Hart for introducing me to the world of mass spectrometry and for instructing me in the use of a range of analytical techniques. To my friends and colleagues in our research group, I would like to thank you for providing a pleasant and welcoming office environment and a helping hand whenever I needed it. Thank you to Jamie Littlejohn, Willian Neale, Harry Williams, Jenny Moran, Omar Alhamd and Sameer Mahmood. I would like to especially thank Jamie Littlejohn, William Neale and Harry Williams, without whom I would not have been able to carry out the work described below. Their guidance both in and out of the laboratory has been monumental in my development as both a researcher and an individual.

Finally, I would like to extend my most sincere thanks to my friends and family, without whose continued love and support I could not have done without.

Chapter 1 – Introduction and Background

1.1 The Immune System

The immune system is the most important factor in maintaining the integrity of host defence against the harmful pathogens to which the body is exposed almost constantly. To this end it has evolved a vast array protective mechanisms which enable it to control and remove these pathogens and the harmful toxins that they produce. The immune system is divided into two equally important halves called the adaptive and the innate immune systems, these names are derived from the specificity and mechanism of their actions. Whilst these two systems are very distinct there is actually frequent interaction between the two; the innate immune system encompasses the parts of the immune system such as neutrophils, monocytes, macrophages and the complement system and immediately respond to threats to the survival of an organism. Due to the fact that the immune system of invertebrates consist solely of innate immune system and because the components of innate immunity are highly conserved between both vertebrates and invertebrates; it can be deduced that it is extremely important for fighting off infections. The adaptive immune system is only found in higher animals such as humans and consists of specific antigen responses brought about by T- lymphocytes and B-lymphocytes. The T-lymphocytes or T-cells are characterised by their surface expression of the αβT-cell receptor (TCR) whose primary function is to recognise antigens in complex with class I and II MHC proteins. These mature T-cells can differentiate into several mature sub-types all of which have different functions and targets: CD8 T-cells target cells which are intracellularly infected, CD4 T-cells regulate the cellular and humoral immune responses such as activation of the B-lymphocytes. The smallest group are the natural killer cells which target cells that present glycolipid antigens which are presented by the CD1d molecules. This particular type of T-cell is characterised by its ability to rapidly produce large numbers of cytokines. The B-lymphocytes or B-cells make up about 15% of the blood leukocyte concentrations and are defined by their secretion of antibodies. B-cells mature from haematopoietic stem cells within the bone marrow and it is here that they produce their antibodies. B-cells can become activated by T-cells during an infection by a pathogen, this activation

causes them to mature which induces isotype switching in the antibodies produced by the B-cells by T-cell cytokines. This allows the T-cells to control this switching as well as the type of antibody produced. Isotype switching has been heavily associated with the B-cell memory response which allows for rapid production of antibodies in the event of secondary infection, it is the memory response which makes vaccination so important and effective. The innate immune response is characterised by a rapid but non-specific response which has the potential to damage healthy host tissues; the adaptive immune system is characterised by a slow but highly specific response. Whilst the initial response of the adaptive immune system is slow its memory allows its response to subsequent infections to be far more rapid however they are still not as immediate as the innate response. This lag period associated with the adaptive immune system highlights the importance of the interactions of the two systems (Delves 2017; Lawrence 2010).

1.1.1 An Introduction to Antibodies

Antibodies, also known as immunoglobulins (Ig), are the secreted form of B-cell receptors. They make up part of the immunoglobulin superfamily alongside other cell surface recognition proteins such as MHC Class I and II proteins. These proteins are soluble and make up 10%-20% of the total plasma content within human serum. These immune proteins are Y-shaped and made up of three recognisable components: two arm regions and a stem region (figure 1.1). The arms are referred to as the variable regions due to their large intermolecular variation, which is required in order for them to perform as antigen binders. The stem interacts with and binds to effector molecules of IgG; such as C1q receptors. Each Ig monomer is made up of these three domains; which are in turn made up of 4 protein chains: 2 heavy chains which have a Mr of 50 KDa each and 2 light chains which have and Mr of 25 KDa each. Both of the heavy chains are identical to each other and both of the light chains are identical to each other. Both of the light chains are associated with a different heavy chain; this produces two identical chains which consist of both a light chain and a heavy chain; it is this symmetry that gives Igs their two identical antigen binding sites. These multiple identical binding sites allow for greater binding affinity and strength of the antibody/antigen interaction due to the multiple contact points they provide (Vidarsson *et al*, 2014).

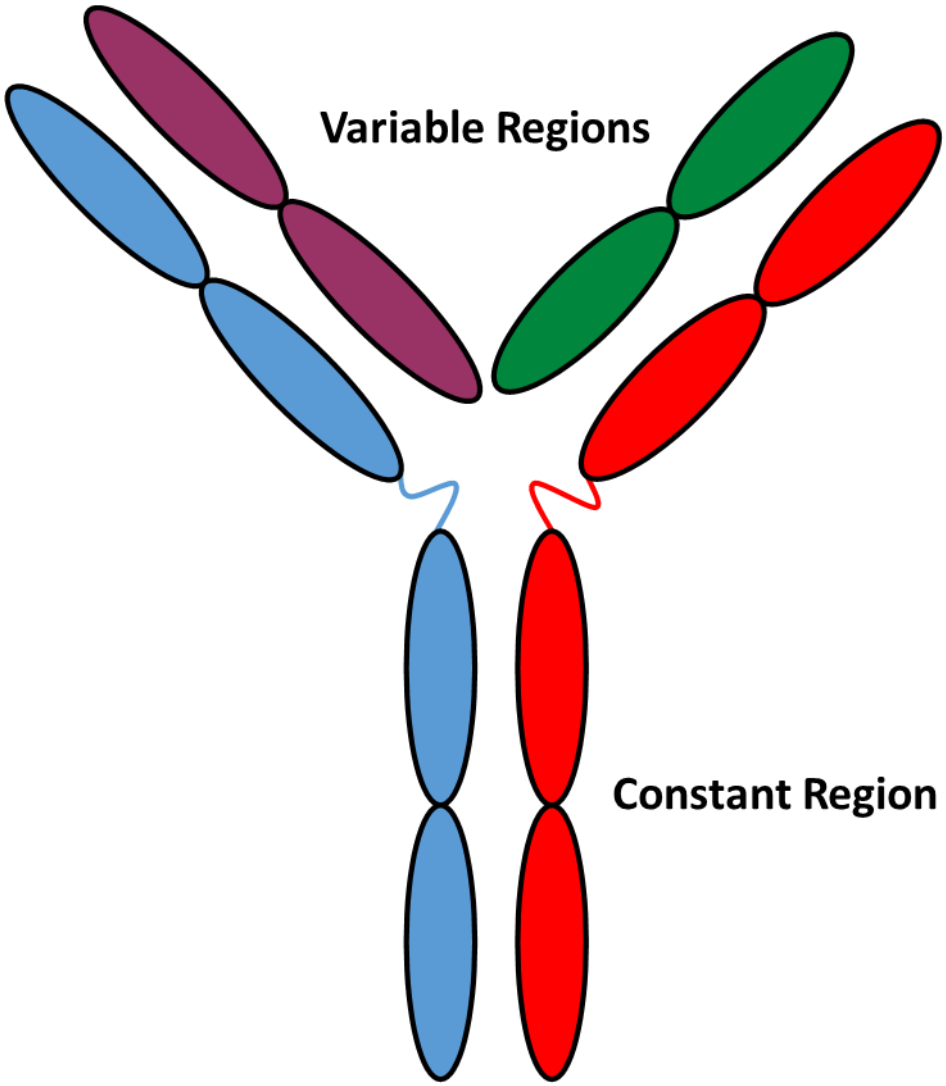


Figure 1.1 – The basic structure of an IgG antibody. The light chains of the variable domains are shown in purple and green and the heavy chains of both the variable and constant domains are shown in blue and red.

1.1.2 Overview of IgG

IgG is one of the most abundant proteins found in human serum making up about 10%-20% of the total plasma content. It is also the most abundant of the five human antibody classes: IgM, IgD, IgG, IgA and IgE. These glycoproteins are roughly composed of 82-96% protein and 4-18% carbohydrate in the form of an N-linked glycan at the Asn297 residue. There are a further four subclasses of IgG: IgG1, IgG2, IgG3 and IgG4; they are named in order of decreasing abundance and were first discovered in the 1960s through studies using specific rabbit antisera against human IgG proteins.

Whilst they are 90% identical on the amino acid level, they all have a unique profile for antigen binding, effector binding and half-lives. The heavy chain structures of the different subclasses of IgG are marginally different to each other and because of this their effector functions differ (Schur 1988). The different subclasses respond to different types of antigens as well as initiating different responses caused by different secondary signals which influence the differentiation of the B-cells, including recognition by pattern-recognition receptors. IgG1 is the most common type of antibody found within the body. As with all IgG1 subtypes they respond to soluble protein antigens/ membrane proteins however they are usually accompanied by lower levels of IgG3 and IgG4 (Ferrante *et al.* 1990). Due to the usually high abundance of IgG1, a lack of IgG1 present in primary and secondary infections can possibly result in a decrease in total IgG levels which can be associated with recurrent infections (Jefferis & Kumararatne 1990). The IgG2 subclass respond exclusively to the bacterial capsular polysaccharides (Ferrante *et al.* 1990; Barrett & Ayoub 1986; Schauer *et al.* 2003). A deficiency of this subclass can result in the absence of IgG anti-carbohydrate antibodies (Hammarstrom *et al.* 1983) however this response can be compensated for by a higher level of the IgG1 and IgG3 subclasses (Hammarstrom *et al.* 1987). This deficiency is also linked to an increased susceptibility to certain bacterial infections which implies that IgG2 has a role in defence against these bacteria (Kuijpers *et al.* 1992; Siber *et al.* 1980). Whilst the majority of glycan infections are reacted to by IgG2 there have been cases of IgG1 antibodies working against *Haemophilus influenzae* b polysaccharide (Ferrante *et al.* 1990). In normal immune responses IgG1 and IgG3 can also be observed especially against protein-conjugated glycans. IgG3 is particularly effective at the induction of effectors, they are pro-inflammatory antibody and as such its shorter half-lives can limit the potential of inflammatory responses (Stapleton *et al.* 2011). IgG4 is most often induced following repeat, long-term exposure to antigens in non-infectious settings such as those found on allergens (Aalberse *et al.* 1983).

1.1.3 Structural Details and Features of IgG

Igs are Y-shaped molecules, made up of a variable and a constant region which are split into six recognisable domains: The heavy variable (VH), light variable (VL), light constant (CL), heavy

constant 1 (CH1), heavy constant 2 (CH2) and heavy constant 3 (CH3) domains. These six domains are made up of the two identical light chains which make up the VL and CL domains and two identical heavy chains which make up the VH, CH1, CH2 and CH3 domains (Vidarsson, Dekkers and Rispens, 2014). The first 110 residues of the light chains as well as the first 110 residues of the heavy chains make up the VL and VH1 domains respectively which along with the next 110 residues of the light and heavy chains make up the CL and CH1 domains respectively. Together these four domains make up the Fragment Antigen Binding (FAB region) of the Igs. The final 220 residues of the heavy chain make up the CH2 and CH3 domains which make up the Fragment Crystallisable (Fc region) of the Ig and contain several binding sites for effector molecules such as C1q receptors. These multiple identical binding sites allow for greater binding affinity and strength of the antibody/antigen interaction due to the multiple contact points they provide (Potter, 1983; Vidarsson, Dekkers and Rispens, 2014). The light chains as well as the short N-terminal variable region of the heavy chains make up the antigen recognising domain of the Fragment antigen binding (FAB) region. The preceding 50 residues of the heavy chain then make up the remainder of the CH1 domain. The structure of these domains are shown in figure 1.2.

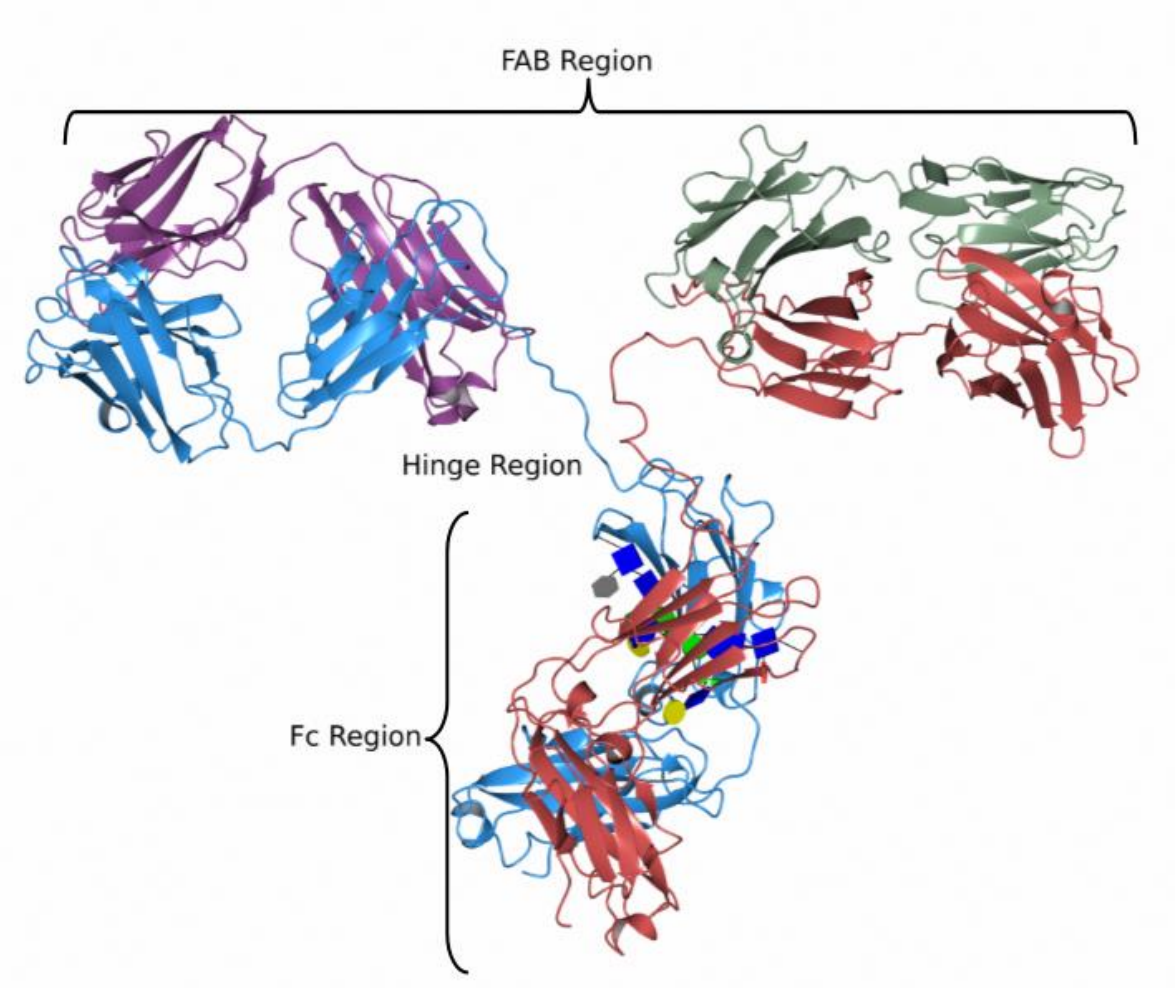


Figure 1.2 – The basic structure of an IgG antibody. The structure of a human IgG1 antibody. The light chains of the variable domains are shown in purple and green and the heavy chains of both the variable and constant domains are shown in blue and red. The N-linked glycosylation of the Fc region is shown in Glycoblocks form. Produced in CCP4mg. PDB code (1IGT).

The CH1 domains of the FAB region have properties that allow them to recognise and bind antigens. The Fc contains no antigen binding activity but crystallises readily (Schur. 1988). The two paired heavy chains within this region interact with effector molecules such as C1q, the functional differences between the different isotypes are due to differences in the Fc region (as addressed below). The receptors found on T cells and B cells are called Fcγ(Gamma) receptors. One of these receptors is the FcγRIIB-1 located on B cells which binds to the Fc region of IgG antibodies and causes the activation of the B cell once the antigens have been bound (Schur 1988).

Another important functional region of Igs is the hinge region which is located between the CH2 and CH1 domain. The hinge is a three part structure consisting of an upper, middle and lower section. The upper section is defined as a peptide bridge that stretches from the N-terminal of the CH1 domain to the first Cys (226) residue in the hinge region. This upper section is conserved between subtypes and importantly forms an interchain disulfide bond. Another disulfide bond is formed between the Cys220 residue on the hinge and one of the light chains. The upper section of the hinge region also has a very low structural stability and is very accessible to solvent. This allows the hinge region and FAB regions to flex, aiding in their antigen binding properties. The middle region of the hinge consists of residues from the C-terminal of the upper hinge to the Ala231 residue in the lower hinge. This region contains multiple Cys residues all of which are responsible for forming the intra-chain disulfide bonds which provide rigidity to the hinge. Proline residues surround these Cys residues and form a polyproline helix which provides flexibility to the hinge region as well as maintaining its 3D structure; this flexibility comes from the lack of a fixed pattern of intra-chain hydrogen bonds as this allows conformational changes within the structure. The lower hinge is encoded by a separate exon to the upper and middle regions (Brekke *et al.* 1995). X-ray crystallography and NMR studies have shown that there is a lack of significant electron density in this region, this is theorised to be due to a significant degree of movement within this region which implies that it is this region which provides the Fc with its flexibility. The lower hinge is also the site of FcγR binding and as such is involved in the interaction between Igs and the cell surface (Vidarsson *et al.* 2014). The hinge region is also known to be involved with not only the antigen binding properties of IgG by allowing them to bind to differently spaced antigens through the flexibility they provide; but they are also known to directly influence the binding of C1q and the activation of the complement system, it is assumed that the overall flexibility of the hinge region is required to allow the Fc region to adopt an optimum conformation for C1q binding. The flexibility is also important as it ensures that the Fab arms can move out of the way of the C1q binding site to prevent them from blocking access to the binding site. It has been suggested that one of the functions of the hinge region is to hold the CH2-CH3 domains of the Fc region in the optimum orientation for FcγRIIB binding rather than through direct hinge FcγR interaction. The light chain also consists of a variable N-terminal domain and a constant

C-terminal domain, which have identical sequences to N and C-terminal domains of the associated heavy chain. Structural studies of the Ig heavy and light chains have determined that each chain consists of similar sequences of about 110 amino acids long. The repeats found within these sequences correspond to discrete folded regions of the protein known as an immunoglobulin fold/ domain (whose structure is discussed below). The light chain is made of two of these domains and the heavy chains are made up of four. The amino terminal amino acids of the heavy and light chains vary between different antibodies; this variability is limited to the first 110 amino acids of the first domain (VL & VH domains), the remaining domains are conserved between Ig chains of the same isotype (CH domains). The variable region consists of the VL and VH domains. The constant region is made up of the CL and CH regions (Brekke *et al.* 1995).

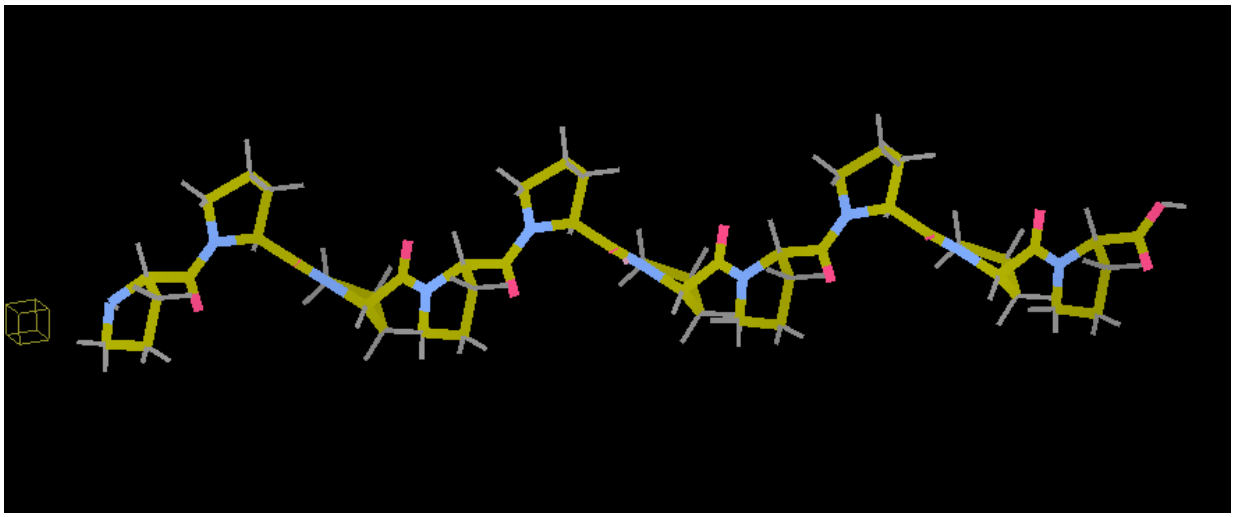


Figure 1.3 – The structure of a basic PP-I polyproline-α helix. The polyproline helices found within the hinge region of immunoglobulins give the region its characteristic flexibility. The Ig helices contain cysteine residues which have been found to produce intra-chain disulfide bonds, which are responsible for providing rigidity to the hinge region. Model produced in coot.

The two paired heavy chains within this region interact with effector molecules such as C1q, any functional differences between the different isotypes are primarily due to differences in the Fc region.

The receptors found on T cells and B cells are called Fcγ receptors. One of these receptors is the FcγRIIB-1 located on B cells (Schur, 1988).

All of the domains within IgG display immunoglobulin folds, which are made up of a pair of anti-parallel β sheets surrounding a hydrophobic core and linked by a single intra-chain disulfide bond. The second key feature of this fold is the hypervariable loops found at its N-terminus which form a binding surface. The loops in the folds of the constant regions are not variable, instead these loops are universal across all antibodies. It is the hypervariable loops which contain the complementary determining regions of the antibody and give rise to the vast diversity seen within the antibody subclasses (Janeway et al., 2001). The structure of the immunoglobulin fold is shown in figure 1.4.

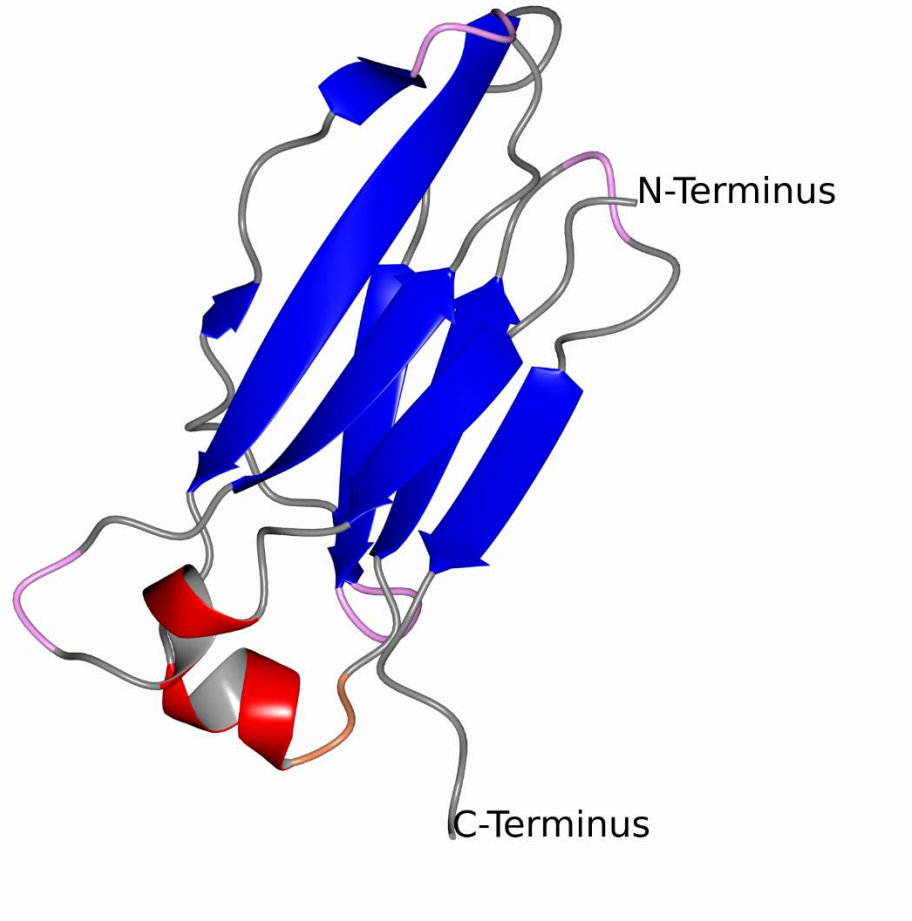


Figure 1.4 – A diagram of the immunoglobulin fold. The immunoglobulin fold is one of the most prevalent protein folds found within the human genome whose hypervariable loops are responsible for the vast diversity within each of antibodies. The anti-parallel β-sheet is shown, the α helix is shown in red and the linking regions are shown in grey. The 3 bond "β" turns are shown in pink and the 5 bond "π" turns are shown in orange. Produced in CCP4mg.

The domains within the Fc region are universally conserved, being 90% homologous between IgG isotypes and are all on average 110 amino acids in length. Structural studies performed on the Fc region have revealed that the two CH2 domains have no protein-protein contacts with each other and so only interact through their N-linked glycans either through direct hydrogen bonds or through hydrogen bonds formed using water as an intermediate. The CH3 regions do however have direct interactions due to the fact that they are closely packed together within the structure, these interactions form the base for the heavy chain dimer. These studies have also shown that the Asn297 N-linked carbohydrate is universally conserved in all IgG isotypes as well as in the CH2 domains of

IgM, IgD and IgE. The Asn297 residue is located on a loop that links the C and E chains (C'E loop). This glycan is important to the function of IgG because alteration to it causes modulation of the function of IgG and binding affinity (Prabakaran et al., 2008).

1.1.4 IgG Fc– Glycosylation

Natural glycosylation levels of IgG Fc have been found to be a reflection of the physiological conditions within an organism and therefore any changes within said organism (Wuhrer et al., 2009). There are several glycans bound to IgG, with the most important being located in the Fc region. These glycans are N-linked to the universally conserved Asn297 residue located within the first half of the CH2 domains (Sondermann et al., 2000). Through manipulation of these glycans it is possible to mediate the receptor affinity of IgG as well as its activity through their participation in Fcγ receptor binding (Sondermann et al., 2000).

These glycans are classified as "complex glycans" due to them containing some of the more rarely seen glycosidic linkages such as: β1,2 linkages and α1,3 linkages. The main structure of the glycan is that of a core heptosaccharide which contains four N-acetylglucosamine (GlcNac) residues. One of these GlcNAc residues (GlcNAc1) is bound directly to Asn297 and forms an α1,6 linkage with a fucose residue and three Mannoses which form the start of two separate arms. These arms can then be extended through the addition of two galactose residues and further again by two sialic acid residues (Krapp et al., 2003; Zauner et al., 2013). The structure of the glycans is shown in figure 1.5.

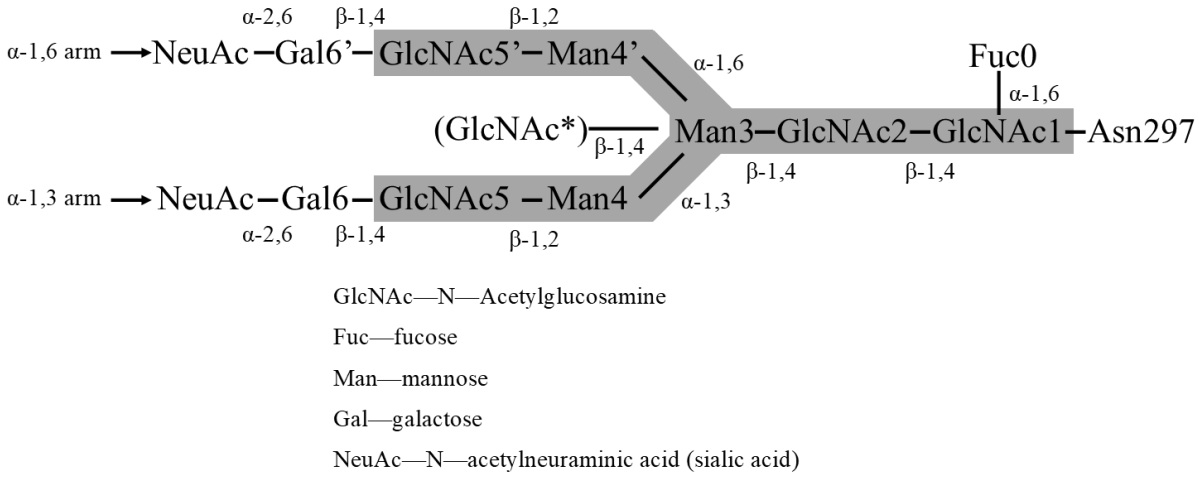


Figure 1.5 – The N-linked carbohydrate attached to the Fc region of IgG at residue Asn297. The core heptasaccharide is shown highlighted in grey. The numbers above and below the residues indicate the type of linkage between them. The bisecting GlcNAc* attached at Man4 only occurs in 5-10% of naturally occurring IgGs (Krapp et al., 2003).

Structural studies of the oligosaccharide have determined that the moiety branches into two arms at the Man3 residue; the first of these arms, the α(1-6) arm is directed towards the CH2-CH3 interface and makes several interactions with the CH2 domain through multiple non-covalent interactions with hydrophobic residues (Kiyoshi *et al.* 2017). The α(1-3) arm extends into the space between the two CH2 domains and forms a crosslink with the α(1-3) arm of the glycan associated with the other chain (Arnold *et al.* 2007). Complete deglycosylation of the Fc region results in IgG Fc losing its ability to bind to FcγRs, due to the two chains of the Fc region moving closer together effectively restricting access to the binding site, this causes IgG to fail to initiate the correct effector functions (Subedi *et al.* 2014; Barb 2015; Subedi *et al* 2015). The binding affinity of truncated glycoforms of IgG was tested and demonstrated a progressive decrease in FcγRIIb binding affinity with progressive reduction in glycan size again caused by the restriction of the binding site (Mimura *et al.* 2000). Whilst the biological effects of these different residues have been observed and documented, the specific manner in which these oligosaccharides influence the structure of the Fc region and function remains unclear (Wright *et al.* 1991). A lack of fucose, galactose and sialic acid bound to the Asp297 core N-glycan, increases the ability of IgG2 to induce antibody-dependent cell-mediated cytotoxicity

(ADCC) in mice (Quast *et al.* 2017). Deglycosylation is also associated with various diseases such as autoimmune diseases, which are associated with a reduction in galactosylation and sialylation of the Fc IgG region which is also a hallmark of the inflammation associated with these diseases. Different cancers as well as viral infections have been shown to exhibit similar low galactosylation and sialylation of the Fc N-glycans in mild inflammatory conditions (Quast *et al.* 2017). It has been proposed that low levels of galactosylation and sialylation is a biomarker of chronic weak inflammation in elderly patients and more generally as a marker of immune activation (Quast *et al.* 2017 & Bondt *et al.* 2013).

As the level glycosylation of IgG Fc is known to be altered during inflammation, several cases of the association of C-reactive protein (CRP) with several of the IgG glycoforms have been reported. It is also unsurprising because both CRP and IgG are associated with inflammation and cardiovascular diseases however there is a lack of recorded associations between glycosylation and the assorted clinical markers in the metabolic health of healthy people (Dube *et al.* 1990). One of these studies found that all non-galactosylated glycoforms of IgG were positively associated with CRP levels and di-galactosylated glycoforms were negatively associated with CRP levels. It was also found that upon the addition of CRP the half-life of the IgG antibodies was significantly increased. The different IgG subclasses show differing levels of glycosylation; with IgG2 showing higher levels of fucosylation and lower levels of bisection and galactosylation compared to other subclasses; the galactosylation and sialylation of IgG causes the IgG to display a weaker affinity for inflammation markers compared to IgG1 and IgG4 glycoforms. This corroborates the data found in other literature which indicated that IgG2 has the lowest affinity for FcγRs as well as having the lowest ADCC activity (Plomp *et al.* 2017). This lower association can be explained by the production of IgG2 being primarily triggered during T-cell independent immune reactions. All of this data points towards IgG2 having a much more limited role in the inflammatory response than other subtypes (Plomp *et al.* 2017). Due to the large potential library of these antibody glycoforms a large catalogue of antibody therapeutics is being developed. Their usefulness will be determined by their affinities to their various receptors as well as their specificity for their designated antigens. These treatments

may revolve around upregulation or downregulation of effector function depending on the treatment. Since FcγRIII is dependent on IgG glycosylation the oligosaccharide moiety must exert an influence on the conformation of the Fc fragment of IgG; it is therefore important to fully understand just how differential glycosylation affects IgG.

1.1.5 The Effector Mechanisms of IgG-Fc

Antibodies link the adaptive immune response to the innate immune system through effector mechanisms such as C1q proteins, they do this by combining antigen-binding sites with innate immune receptors. The different subclasses have different affinities for the effectors as well as differing levels of action, for example IgG1 and IgG3 are more potent activators than IgG2 and IgG4 both of which have more subtle levels of action (Bruhns *et al.* 2009). Binding to the C1q proteins causes the activation of the complement cascade through the classical pathway (IgG complement activation via the alternative pathway has not been observed). The interaction between C1q and IgG Fc occurs via the globular heads of C1q interacting with residues 285-292 and 317-340 of the heavy chain of CH2. The binding affinity of C1q for IgGs is dependent on the level of steric interference caused by the FAB arms as the C1q protein approaches the CH2 sites: IgG3>IgG1>IgG2>IgG4. Generally it has been observed that IgG4 does not interact with C1q and in fact appears to inhibit immune precipitation as well as the binding of C1q to IgG1 in mixtures of IgG4 and IgG1. It is therefore thought that IgG4 may be protective against the effects of the complement-fixing antibodies.

1.1.6 Structural Studies of Fcγ Receptors

The second effector mechanism of the IgG subclasses is the ability to interact with the cell surface Fc receptors. The Fc receptors that are specific to IgGs are referred to as FcγRs, with these structurally distinct receptors being differentiated by the superseding roman numerals (Ravetch & Lanier 2000). After they have been activated by antigens on the surface of certain cell types (usually pathogenic cells), the IgG subclasses can interact with these Fc receptors to facilitate the clearance of these cells through phagocytosis (Nimmerjahn & Revetch 2006). There are three classes of human

IgG Fc receptors on leukocytes: FcγRI, FcγRII and FcγRIII. These receptors are differentiated by their structures, cellular presence and their subclass affinities (Ravetch & Lanier 2000). The cellular expression of the receptors are as follows: FcγRI on monocytes; FcγRII on monocytes, neutrophils, eosinophils, platelets, and B cells and FcγRIII on neutrophils, eosinophils, macrophages, and natural killer cells. The FcR family of proteins are all a part of the type I transmembrane glycoprotein superfamily except for FcγRIIIB which is GPI anchored (Nimmerjahn & Revetch 2006). Whilst FcγRs primarily bind IgG they have also been known to bind pentraxins such as C-reactive protein, serum amyloid P and other innate immune proteins. However, the full importance of these interactions is as of yet unknown (Vidarsson *et al.* 2014).

1.1.7 FcγR mechanism of action

Over 100 residues in the Fc region have been found to be associated with FcγR binding to IgG1. In the first of these regions located in the lower hinge the Leu234 and Leu235 residues are thought to be critical to the binding process as a substitution of even one of these residues results in an affinity reduction of 10-100 fold (Ravetch 2003). The second region is located in the bend between two of the β strands in the CH2 domain and contributes a Pro331 residue to the binding pocket. Because of how the β strands fold during synthesis this residue is 11 Å from the hinge linker region and substitution of this residue results in an affinity reduction of Fc receptor constant by 1- fold. (Ravetch 2003), the FcγR binding pocket is shown in figure 1.6.

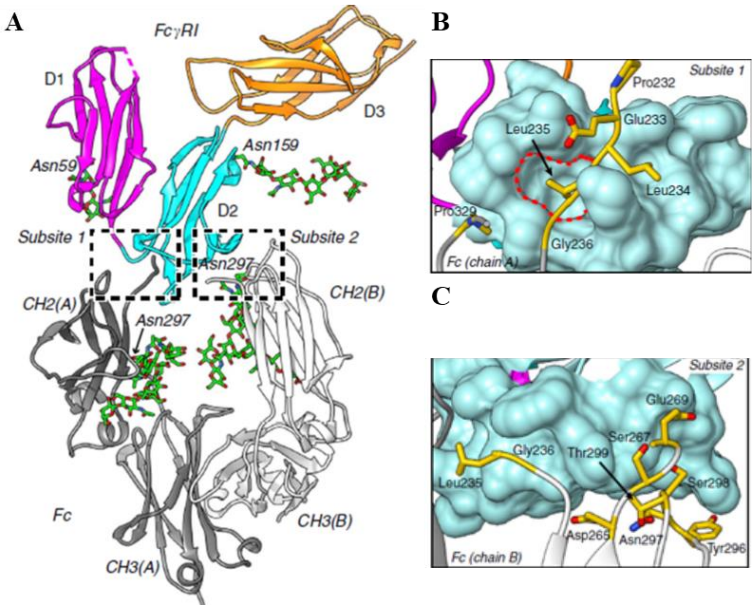


Figure 1.6 – The structure of the Fcγ receptor binding pocket. The binding pocket is located within the lower region of the hinge as well as the start of the CH2 domains of the Fc region of IgG, there is some variation within this binding pocket between the IgG subtypes as well as between the different Ig families (figures taken from Kiyoshi *et al* 2015). **A.** The structure of the IgG-Fc/ FcγR complex with the two binding pockets highlighted in boxes. **B**. Subsite 1, located on the A chain in the region between the CH2 and CH3 domains. **C.** Subsite 2, located on the B chain in the region between the CH2 and CH3 domains.

1.1.8 FcγRn

FcRns mediate the transport of maternal IgGs across the placenta to the child during pregnancy; transferring humoral immunity to the foetus/ child against antigens that the mother has already encountered. Its secondary function is to protect IgG from degradation. FcRns achieve this by acting as a receptor which is responsible for maintaining the long half-life of IgG as well as maintaining high concentrations of IgG in serum (Ghetie & Ward 2000). The Ile253 and His310 residues in between the CH2 and CH3 domains have been shown to be the key players in the interaction between FcRn and IgG Fc; these residues have been shown to be highly conserved across species (Martin *et al.* 2001). Structural studies in mice have shown that a mutation of the His433 residue to an alanine residue does not affect its activity levels however simultaneous mutation of both the His433 and Asn434 moderately reduces it (Kim *et al. 1999*). These studies have also shown that mutation of the

His435 to an alanine dramatically reduces the activity levels (Martin & Bjorkman 1999). Spatial studies have implied that the sequential differences at residue 435 in IgG1 and IgG3 (His in IgG1 and Arg in IgG3) is responsible for IgG3s reduced activity during FcRn/Fc interactions which has provided a greater level of evidence to support the importance of residue 435 in FcRn binding (Kim *et al. 1999*). There is however a discrepancy in this theory as studies performed using surface plasmon resonance analysis of a chimeric mouse/human IgG1 detected no significant difference in the pH dependence of the FcRn-IgG1 interaction when the His435 was mutated to an arginine (Medesan *et al.* 1997). Other results found in this study implied that it was in fact the His433 residue which has a role in FcRn binding. This discrepancy in the data is most likely due to the different methods used, however the data from both techniques prove that histidine residues are heavily involved in FcRn binding. The role of histidine in FcRn binding helps to explain the strict pH dependence of the interaction (Ghetie & Ward 2000). More recent studies have shown that the His310 and His435 residues become protonated between pH 6.0 and 6.5 and it is once they have become protonated that they form salt bridges with the anionic Glu117, Glu132 and/or Glu135 and Asp137 residues on the α2 -helix of FcRn.

1.2 Structural Studies of IgG

1.2.1 IgG-Fc bound to FcγRI

FcγRI is one of the major receptors found on the surface membranes of immune cells, specifically: macrophages, neutrophils and dendritic cells. Several studies have linked the FcγRI to several roles within the immune system. Despite the fact FcγRI has the highest affinity interaction of the Fc region; the structure of IgG Fc bound to FcγRI has only been recently reported. A study performed in 2015 was the first to report a high resolution (1.80 Å) structure of the interaction site of the FcγRI: Fc interaction. The x-ray crystallographic data reports that the majority of the interactions between the FcγRI binding subsite 1 and the Fc fragment involve the Pro232 – Ser239 residues of the lower hinge region of the CH2 domain. A number of conformational changes which occur during this interaction were also reported in this study, the most notable of which is the rigid-body displacement

of the D3 domain of the FcγRI, this displacement is 11 Å in magnitude with a rotation element of 19° in relation to the D1 and D2 domains (also of FcγRI). This displacement occurs away from the main binding sites and as such it is unknown whether this movement has an effect on the binding interaction or whether it is just a by-product of it (Kiyoshi *et al.* 2015).

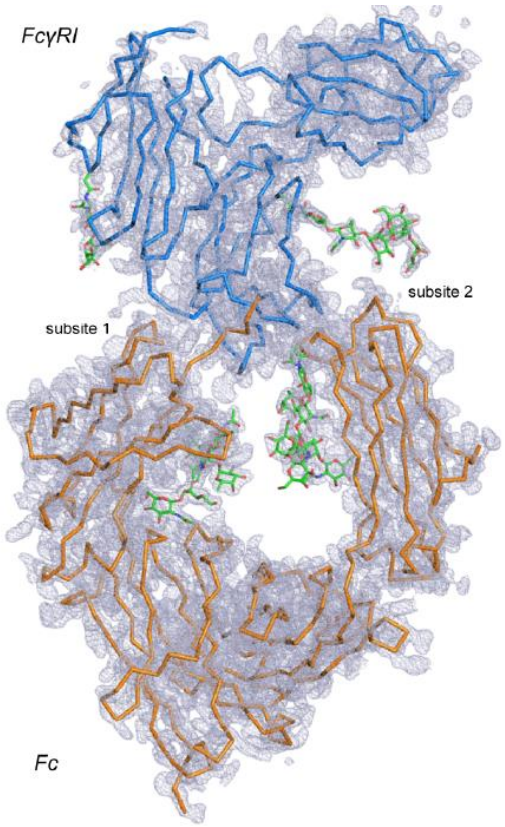


Figure 1.7 – The structure of the human IgG1 Fc region bound to the human FcγRI receptor. The electron density map is shown in grey. The Fc region is depicted in yellow, the FcγRI receptor is shown in blue and the glycosylation of both proteins is shown in green (Kiyoshi *et al.* 2015 supplementary material).

Through comparison of 1.80 Å structure of bound Fc/ FcγRI to the structure of Fc it was shown that the only change experienced by the Fc fragment is a rigid body displacement within the CH2 domains. This displacement takes the form of a pair of asymmetrical rotations within the A and B chains, with the rotation in the B chain being 10.7° with respect to the CH3 domain and a much smaller rotation in the A chain (amplitude of rotation 2 Å). The result of this shift was a 9 Å increase

in the distance between the two CH2 domains by the hinge region. There were also several features observed in FcγRI bound Fc that were not present in the structures of native Fc or Fc bound to other Fcγ receptors. The most important difference is that the Leu235 residue of the A chains CH2 domain is located within a hydrophobic pocket within the FcγR, mutational studies have implied that this Leu residue is critical to the high affinity binding of the Fc region to this receptor. This is backed up by the fact that it has been observed that the buried surface area of this residue accounts for 25% of the A chains interaction with the Fc region. This also helps to explain why the IgG2/FcγRI interaction has such a low affinity, this is due to the fact that residue 235 in IgG2-Fc is not a leucine residue and as such cannot adequately interact with the hydrophobic pocket (Kiyoshi *et al.* 2015).

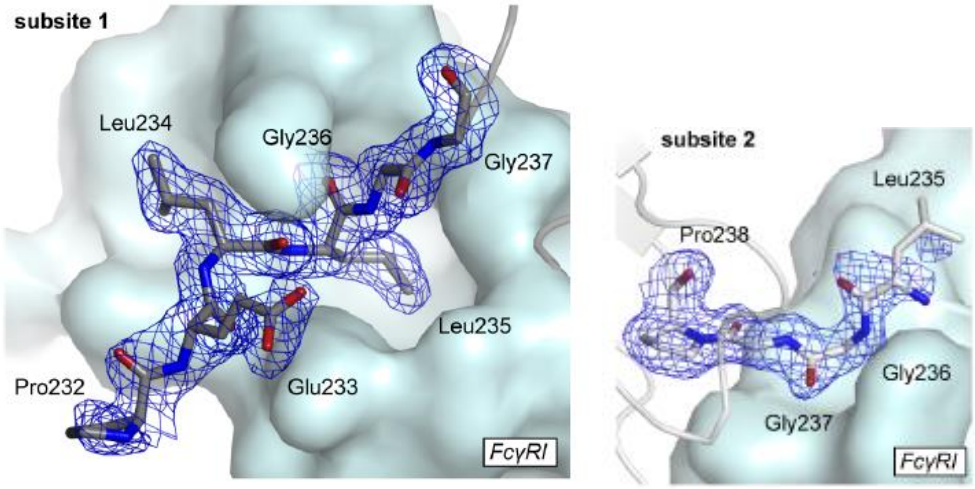


Figure 1.8 – The electron density of the hydrophobic binding pocket of the IgG1 Fc: fcγRI interaction, showing both subsites. The surface of fcγRI is represented by the cyan surface diagram. The electron density map for the A chain of the fc region is projected onto the sequence of residues 232-237 for subsite 1 and for residues 235-238 for subsite 2 (Kiyoshi *et al.* 2015 supplementary material).

1.2.2 Modification of the N-Linked Glycan

1.2.2.1 Afucosylated IgG1-Fc

It has recently been observed that the afucosylation of the N-linked glycan of the Fc region improves the antibody-dependant cellular cytotoxicity (ADCC) as shown by a 50-fold increase in the binding

affinity of FcγRIIIa. X-ray crystallographic analysis of the KM3060 (chimeric anti-CCR4) and KM3575 (chimeric anti-CD20) antibodies, which acted as the fucosylated and afucosylated antibodies respectively; determined that the removal of the fucose residue only produced a very subtle change within the secondary structure of the Fc region (RMS deviation of 0.32 Å). These antibodies were chosen due to their identical Fc region residue sequences. The structural analysis also determined that upon binding to FcγRIIIa there is a conformational change within the C'/E loop of the Fc specifically around residue Tyr296 which has been shown to have interactions with both Lys125 and Asp126 of FcγRIIIa. ^{1}H NMR data of the area around Tyr296 indicated that there was a very minute shift in the placement of Tyr296 within the afucosylated Fc region possibly due to the differing hydrogen bond profile when fucose is removed. The effect of this shift is that Tyr296 is more flexible within the afucosylated Fc and hence more readily binds and adjusts to FcγRIIIa. Another potential explanation for the increased affinity is, that the carbohydrate attached at Asn162 on FcγRIIIa can only interact with the surface of the Fc region when fucose is not present (Ferrara *et al.* 2006).

1.2.2.2 Sialylated IgG-Fc

A study by Crispin *et al* `used *in-vitro* enriched sialylated 2,6-Fc fragments (sFc) of IgG which were derived from IgG-Fc obtained from sera. The sialic acid residues were found bound to the galactose saccharide of the 1,6 arm of the biantennary N-linked glycans. Initial testing found that these Fc fragments bound to the SIGN-R1 and DC-SIGN c-type lectins whereas native IgGs would not. The sFc was also found to have a 10 fold lower affinity for FcγRIIb receptor. The reason for this new binding mechanism was first thought to be due to the 2,6-Sialic acid galactose bonds causing the Fc fragment to adopt a new conformation which allows it to bind to SIGN and DC-SIGN. However a follow up study in 2013 demonstrated using X-ray crystallography as well as CD that the structure of the sFc was typical of a native Fc region and so the mechanism of binding is uncertain at this time. This interaction is still very important however as it is the first documented case of the Fc region gaining a binding mechanism through manipulation of the N-linked glycan and has opened up the

idea that manipulation of the glycan in other ways could lead to IgG gaining other new binding mechanisms (Crispin *et al.* 2013).

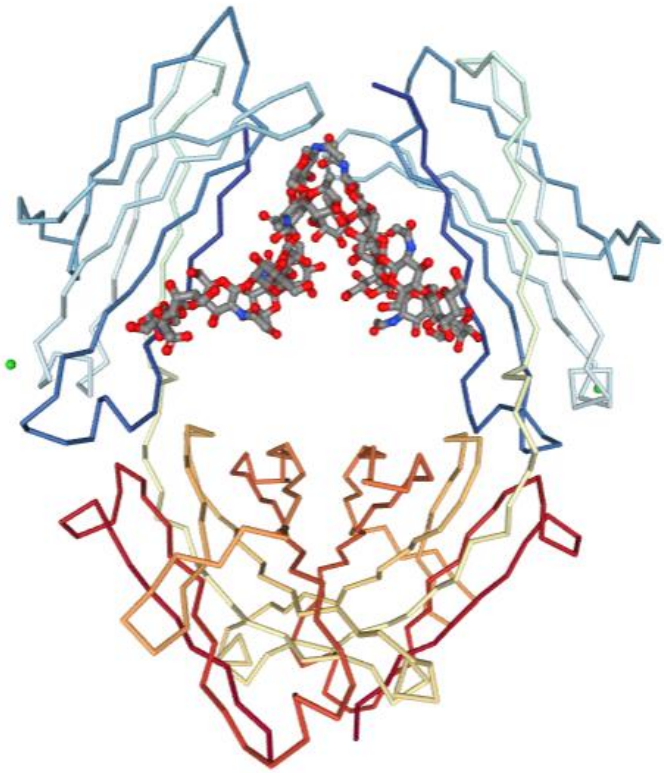


Figure 1.9 – The structure of human IgG Fc whose N-linked glycan is enriched at α2,6-sialic acid residue. The CH2 domains are shown in blue and the CH3 domains are shown in orange PDB code 4BYH (Crispin *et al.* 2013).

1.2.2.3 Mutant IgG Fc-Fragments

There have been many studies into the alteration of the Fc region through residue mutations in order to enhance or supress the binding activity of Fc to FcγRs. One study demonstrates the binding affinity for FcγRIIIa of the SDIE (S239D/I332E mutations) and the SDALIE (S239D/A330L/I332E mutations) Fc mutants. In this study these mutants showed a 60-fold and 170-fold increase in binding affinity for FcγRIIIa respectively. A separate study showed that they also had a 70-fold and 40-fold increase in affinity respectively (Lazar *et al.* 2006). The crystal structure of SDALIE produced in 2008 (Oganesyan *et al.* 2008) displayed that the Fc region is very structurally similar to the many

reported wild-type Fc fragments with the root mean square deviations being no greater than 0.6 Å for the main chain atoms. However these studies also showed that these mutant type Fc regions have a more open conformation but without any differences in the FcγR binding domain. Another feature of the structures of these mutants is that when their CH2 and CH3 domains are considered individually they show a considerable level of conservation, but with a considerable level of flexibility within the domains (Oganesyan *et al.* 2008).

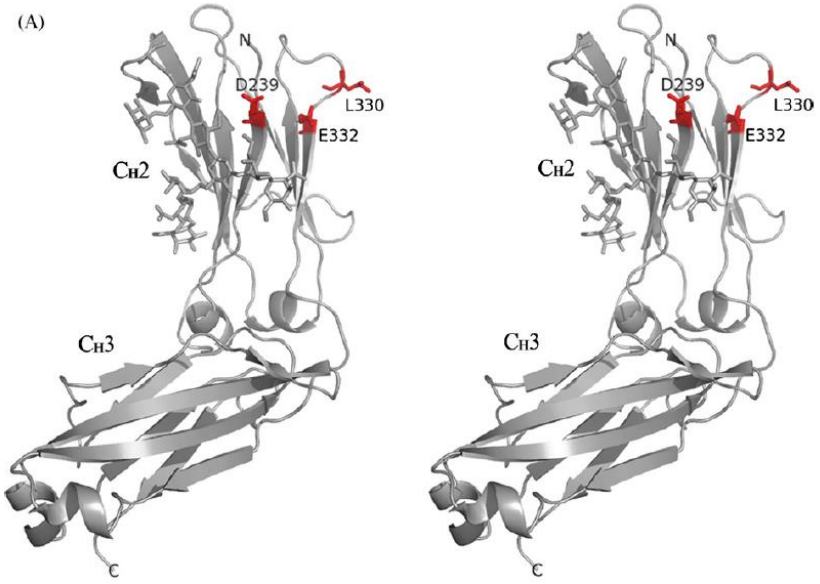


Figure 1.10 – The structures present within the asymmetric unit of the SDALIE crystal. The three mutations are shown in red with the N-linked glycan being shown as in the stick representation (Oganesyan *et al.* 2008).

Further studies into these mutated Fc fragments have revealed that the GASDALIE mutant which contained the G236A/S239D/A330L/I332E mutation pattern showed a 20 – 30 fold greater affinity for activating FcγRIIIa, with a much smaller increase in affinity for FcγRIIb (2 – 3 fold increase). The implication of which, is that GASDALIE acts as an activator rather than an inhibitor. Recent studies using a mouse model have shown that GASDALIE antibodies have protective properties against certain cancers and other infectious diseases such as HIV (Ahmed *et al*. 2016).

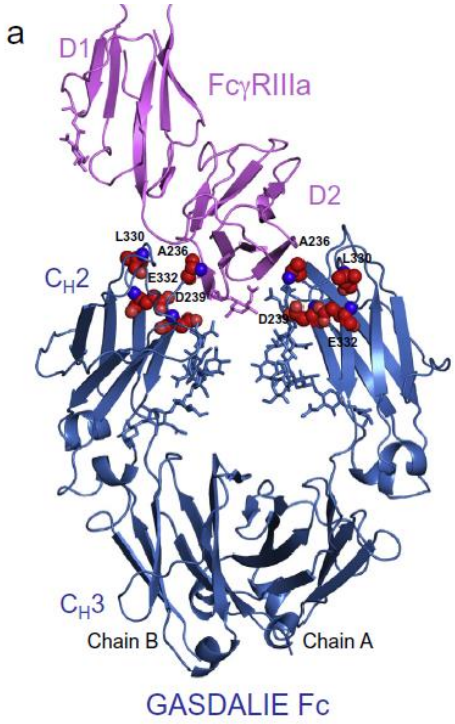


Figure 1.11 – The structure of the GASDALIE Fc region bound to FcγRIIIa. FcγRIIIa is depicted as the purple ribbon and GASDALIE is depicted as the blue ribbon. The 4 mutated residues are shown as red spheres and the N-linked glycan is depicted as a blue stick model PDB code 5D6D (Ahmed *et al*. 2016).

The mutant-Fc structure was solved through molecular replacement using the afucosylated Fc: FcγR structure as a search model before being refined to 3.1 Å. The electron density map clearly showed the position of the sites of the GASDALIE substitutions (position 236, 239, 330 and 332) on the A and B chain. The crystal structure of the mutant-Fcs showed that it had the same asymmetric 1:1 interaction with the FcγR D loops that is seen in the other IgG Fc fragments and that the dfferent mutations altered the structure and binding affinity in comparison to native Fc. The GASDALIE specific interactions of the A chain are as follows: the lys120 residue of the BC loop in FcγRIIIa was found to be positioned in-between the Asp239 and Asp265 residues, which is vastly different to the native interaction which sees the Ser239 residue forming a hydrogen bond with the lys120 residue. The AL and IE substitution were found have no effect on the interaction of the A chain with FcγRIIIa. Whilst the GA substitution of the GASDALIE Fc was also found to have no effect on the interaction, the Gly236 of the native Fc formed a hydrogen bond with the His135 residue within the FcγR. It is

thought that the increase in binding affinity in the GASDALIE Fc was due to the introduction of the salt bridge at position 239 due to the fact that a singular salt bridge is much stronger than the two hydrogen bonds which are lost due to the mutation. The B chain shows new interactions formed due to the introduction of the GASDALIE mutations and they are as follows: the IE substitution on the B chain formed an electrostatic interaction with Lys161. This interaction is not seen in the native Fc as the Ile332 cannot form an interaction with the lys161. The GA substitution makes a new hydrophobic interaction with the Phe158 residue of the FcγRIIIa, which differs to the native Fc as residue 158 is not contacted and Gly236 forms a hydrogen bond with Lys161 of FcγRIIIa. The other two mutations on the B chain were not found on the binding interface and as such had no effect on the affinity of the interaction. It is the new salt bridge and hydrophobic interactions that are thought to be responsible for contributing to the increase in affinity (Ahmed *et al*. 2016).

1.2.3 Applications of IgG-Fc in Combating the Zika Virus

The Zika virus (ZIKV), a member of the *Flaviviridea* family of viruses is the causative agent of the infection caused by bites from the *Aedes aegypti* species of mosquitoes, it was first identified in the Zika forest of Uganda in 1947 (Dick 1952). As of yet the full implications of this infection for human health is unknown, however some symptoms, such as serious birth defects (which include neurological disorders and microcephaly) have been reported in several studies focused around Brazil (Cugola *et al.* 2016). The flavivirus family is characterised by their envelope protein (ZIKV E). The ZIKV E protein is a single polypeptide that is processed into 3 mature structural, viral proteins including: a pre-membrane protein, a non-structural protein and an envelope glycoprotein. Studies of the ZIKV E have shown that it is split into three domains: DI, DII and DIII all folded into an elongated rod like structure which sits horizontal to the envelope forming antiparallel dimers. The DII is formed by an elongated finger like structure which contains a hydrophobic glycine rich fusion loop (E fusion protein) at its most distal loop. This fusion protein consists of residues 98-110 and is responsible for interacting with target cell membranes during cell fusion, it does this by inserting itself into the targets membrane during the pH-dependent conformational changes that occur during fusion (Dai *et al.* 2016). It is this fusion protein that is of particular interest to researchers due to the

fact that it is highly conserved in the flavivirus family and as such it is a tempting target for monoclonal antibodies (Mabs). To this end several mabs were tested in mice models and two in particular (ZV-54 and ZZ-67) were found to have a neutralising effect on ZIKV activity (Zhao *et al.* 2016). To date only one effective human anti-flavivirus mab has been produced; designated 2A10G6 it is a monoclonal IgG1 antibody that has been shown to be effective against type 1-4 Dengue (DENV1-4), Yellow Fever virus and West Nile virus. The method of action for 2A10G6 has been detailed through X-ray crystallographic studies, which showed that the Fab region of 2A10G6 binds to the tip of the ZIKV-E domain in a perpendicular manner which in turn embeds the fusion loop in a hydrophobic pocket formed by the heavy and light chains of the antibody as well as the W101, L107 and F108 residues of the ZIKV-E. These residues are highly conserved in the flavivirus family hence 2A10G6s broad capacity (Deng *et al.* 2011).

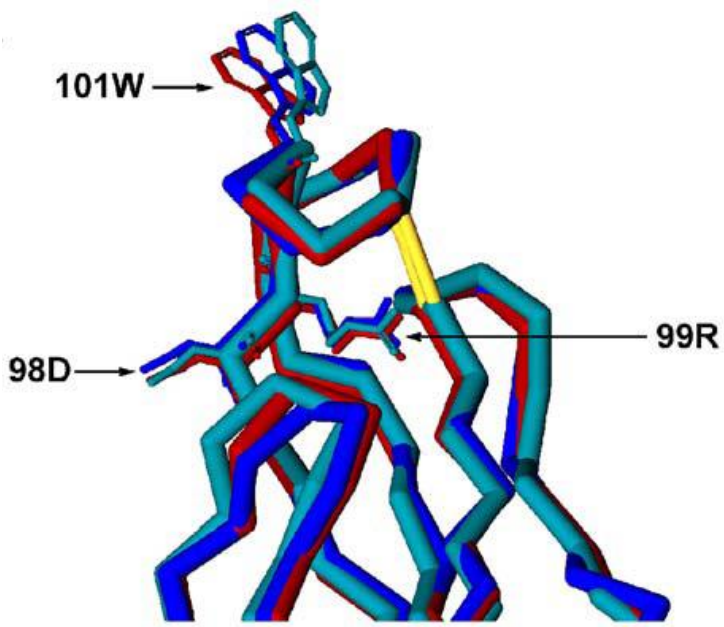


Figure 1.12 – The structure of the 2A10G6s antibody in bound to three different flavivirus E proteins, The DENV2, West Nile Virus (WNV) and Tick-borne encephalitis virus (TBEV) are shown in red, blue and cyan respectively. The three key residues which are involved with the binding of the flavivirus E fusion protein are also labelled within this structure (Deng *et al.* 2011).

1.3 An Introduction to Heparin & Heparan Sulfates

1.3.1 Heparin & Heparan Sulfates

Heparin is a linear polysaccharide which is a part of the Heparan Sulfate (HS) family of glycosaminoglycans (GAGs). The structure of Heparin is made up of uronic acid (1-4) -D-glucosamine repeating disaccharide subunits the structures of which are shown in figure 1.13. Heparin contains variable random patterns of substitution of these disaccharide subunits with N-sulfate, O-sulfate and N-acetyl groups, this substitution gives rise to a random and variable pattern of sulfation within heparin molecules, as well as a large number of complex members of the GAG polysaccharide family (Robinson *et al.* 1978).

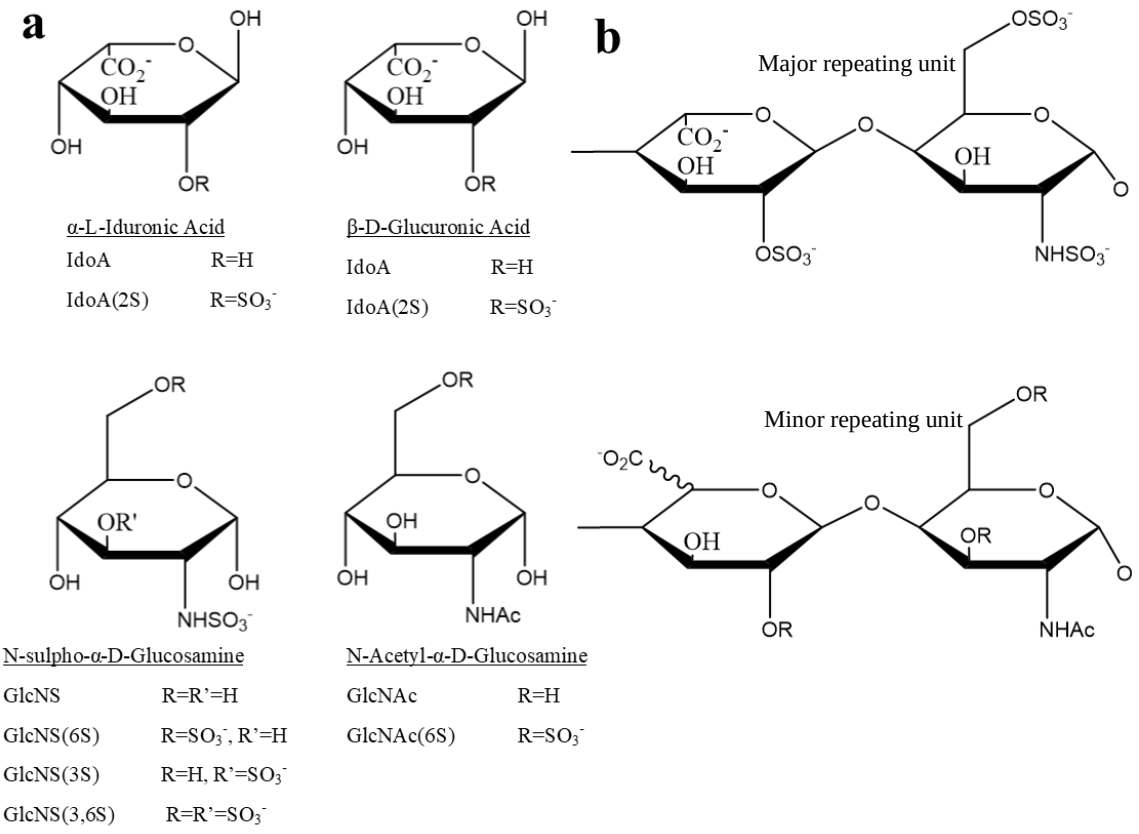


Figure 1.13. (a) The structures of the base monosaccharide units of heparin. (b) The structures of the major and minor repeating disaccharide subunits of heparin. Produced in Chemdraw.

Heparin is expressed in mast cells found in connective tissues where it is synthesised as heparin proteoglycan (75 KDa - 100 KDa) (Conrad 1998). This proteoglycan consists of a unique core protein (serglycin) and polysaccharide chains that are covalently bonded to it. Once synthesis is complete these chains are cleaved at random intervals in order to give rise to a large amount of smaller polydiverse heparin chains, these are then stored in the secretory granules of mast cells as complexes with proteases. Because heparin is produced in mast cells it is sparser within the body than HS due to the fact that mast cells are found in organs that are commonly exposed to pathogens such as the liver and the lungs and HS is found in tissues throughout the body (Salmavirta *et al* 1996). Another feature of heparin is the differing structures and sulfation patterns of heparins produced in different tissues. An example of this is pig intestine heparin which contains an ATIII binding site that contains N -acetyl D-glucosamine and cow lung heparin which contains N-sulfo D-glucosamine; these differences account for their different binding affinity. Commercial heparin was initially sourced from dog liver, Charles Best began a project to first produce bovine lung heparin and then porcine intestine heparin (Rabenstein 2002). All of the disaccharide structures found within heparin are also found within heparan sulfate but in different proportions, this helps to explain why the level of sulfation is different between the two GAGs. It also explains why their anticoagulant activity is different between the two (Rabenstein 2002).

1.3.2 Structural Details of Heparin and Heparan Sulfate

The main constituent of heparin and HS is an α1-4 linked pyranosyl uronic acid (uronic acid) and 2-amino-2-deoxyglucopyranose (D-glucosamine) repeating disaccharide unit. The uronic acid subunit ratio is 90% L-iduronic acid and 10% D-glucuronic acid. Heparin has the highest negative charge of all biological molecules making it a polyelectrolyte. There are several differences at the disaccharide levels between different forms of heparin (sequence microheterogeneity); heparin is also polydisperse with a molecular weight range of 5-40 kDa with an average of 12 kDa and an average negative charge of 75 making characterisation difficult. D-glucuronic acid makes up the majority of the structure of heparan sulfate which is polydisperse; with a molecular weight ranging from 5-50 kDa and an average of 30 kDa. Whilst the main chain structure of HS has reduced sulfation (by

comparison to heparin), the sulfation is more uniform and organised, with several of the domains contain defined extended sequences of sulfation; either being highly sulfated or less sulfated; HS also remains connected to its core protein unlike heparin. There are two types of HS proteoglycans which are found on the extracellular matrix and the cell surface; these two proteoglycans are syndecans and glypicans. The HS chains associated with these proteoglycans are responsible for mediating physiologically important processes. Heparin and HS can be distinguished by their differing affinities to the GAG degrading enzymes: heparin lyases. Neither one of them has a tertiary structure instead existing as extended α-helices; this structure plays an important role in the specificity of their interactions with proteins as these interactions depend on how the sulfo and carboxyl groups are displayed on its surface. The large range of interactions attributed to this family of GAGs is believed to be due to the flexibility of the L-iduronic acid residues within the GAGs. The level and type of flexibility depends on the position of these residues within the chain. Heparin exists largely as a linear right handed helix however they do not contain any tertiary folds or structures (Rabenstein 2002).

Complete structural analysis of the structure of heparin would require the identification and quantification of each of the uronic acid-(1,4)-glucosamine disaccharides as well as the determination of their sequence within the polysaccharide chain. It would also require the characterisation of the 3D structures of the monosaccharide residues, which is incredibly difficult due to their highly poly-diverse natures. Because of this the actual structure of heparin is unknown (Rabenstein 2002). Whilst the overall structure is unknown the disaccharide compositions of the different heparin preparations can be determined through analysis of the products formed by the exhaustive cleavage of heparin through deamination or enzymatic cleavage of the uronic acid-(1,4)-glucosamine glycosidic bonds. Of the 48 known disaccharides only 19 have been observed in heparin. Deaminative hydrolysis using nitrous acid gives selective and quantitative cleavage of the glycosidic bonds of the N-sulfated GlcN residues whilst retaining the O-sulfated substituents as well as the isomeric features of the uronic acid residues (Conrad 1998). The GlcNS, GlcNS(6S), GlcNS(3S) residues are converted to the following 2,5-anhydro-D-mannose (AM) residues: AM,

AM(6S), AM(3S) and AM(3,6S). Because the bonds of N-acetylated GlcN residues are not cleaved; the products of deamination are disaccharides, as well as tetrasaccharides. These polysaccharides contain an internal GlcNAc residue. The most common of these products is IdoA(2S)-(1,4)-AM(6S); which is formed from IdoA(2S)-(1,4)-GlcNS(6S), the most common disaccharide in heparin (Bienkowski & Conrad 1985). The N-acetylated glcN glycosidic bonds can be cleaved by removing the acetyl group through a reaction with hydrazine and then cleaving the glycosidic bonds through deamination at pH 4.0 which is much higher than the pH of 1.5 required for deamination of the glycosidic bonds of the N-sulfated GlcN residues. The AM residues can be radiolabelled using NaB_3H_4, which enables detection of the oligosaccharides after separation by gel permeation chromatography or strong anion exchange (SAX)-HPLC (Loganathan *et al.* 1990). The heparinase class of enzymes are obtained from *Flavobacterium heparinum* and cleave the uronic acid-(1,4)-glucosamine glycosidic bonds leaving the glucosamine residue and the sulfated substituents intact but cause it to produce a terminal uronate with an unsaturated 4,5 carbon bond (ΔUA) at the nonreducing end. Due to both IdoA and GlcA residues being converted into terminal ΔUA any information about the conformation of the uronic acid is lost. Heparinase I has high specificity for IdoA(2S) glycosidic bonds where the glucosamine can be GlcNS, GlcNS(6S) or GlcNS(3S) and a secondary specificity to GlcA(2S) bonds in highly sulfated regions (Conrad 1998; Venkataraman *et al.* 1999). Heparinase II is less specific in that it will cleave glycosidic linkages that contain both 2-O-sulfated and 2-OH uronic acids. Heparinase III has high specificity for 2-OH GlcA-containing glycosidic linkages and to a lesser extent 2-OH IdoA-containing linkages in unsulfated regions of heparin. The ΔUA double bond is a chromophore and is a convenient method of monitoring the progress of depolymerisation and as a way to detect the oligosaccharides which are separated by SAX-HPLC and capillary electrophoresis (CE). The disaccharides which are produced by this cleavage can be identified by comparing their retention times with those of the ΔUA containing disaccharide strands. The partial cleavage of heparin using nitrous acid and the heparinases is widely used to produce tetra, hexa, and higher oligosaccharides to allow for further study. Chemically discrete oligosaccharides can be isolated from the cleavage mixtures through a two-step separation procedure which involves size separation into small oligosaccharide fractions via gel permeation

chromatography and then separation based on charge through SAX-HPLC (Venkataraman *et al.* 1999). Each of these fractions will contain many separate components often there are more than 40 discrete components. Because of this only small quantities of each discrete oligosaccharide can be produced by heparinase depolymerisation (Rabenstein 2002).

The final structure of heparin is a product of incomplete biosynthetic modification of its precursors. This modification is mediated by substrate specificities of the modification enzymes as well as cellular regulation. The biosynthesis of the chains takes place in three stages; the first of which is chain initiation which involves the stepwise addition of a xylose residue, two galactose residues and a GlcA residue to specific serine residues at intervals along the chain. This produces the β-GlcA-(1,3)-β-Gla-(1,3)-β-Gal-(1,3)-β-Xyl- tetrasaccharide linker. The heparin chains are then constructed by the alternating addition of 1-4 linked GlcNAc and GlcA units to the linker; this forms [GlcA-(1,4)-GlcNAc]n polymer chains, these chains are then modified further through N-deacetylase/ sulfotransferase, epimerase and O-sulfotransferase catalysed reactions. These reactions occur in a specific order and the products of the reactions are used in the next ones, this produces the mature heparin. The reactions are referred to as incomplete as not all of the sites available in these reactions are used as a result the complexity of the final product is increased throughout the process (Rabenstein 2002).

1.3.3 The Structural Differences of Heparan Sulfate

HS is also synthesised as a proteoglycan, but unlike heparin it is produced and secreted by all mammalian cells. The proteoglycans associated with the structure of HS are shorter and are made up of fewer polysaccharide chains. The expression of the HS proteoglycan is tissue specific and can change during synthesis. Much like in heparin this differential expression is because of the activation tissue specific enzymes which are involved in the synthesis process; the most common tissue specific difference is the variability in chain length and sulfation patterns. Another example of the differences between heparin and HS is that the overall molecular organisation of HS remains the same during the progression of colon cells towards malignancy, however the levels of sulfation in the N-sulfated

(NS) domains and N-acetylated/N-sulfated (NA/NS) regions changes. It has also been noted that changes in the patterns of 6-O-sulfation, chain length and the number of sulfated domains with HS produced by neuronal cells occur alongside the transition from neural precursor cells to neuronal differentiation. (Venkataraman *et al.* 1999).

There are two subfamilies of the cell surface HS proteoglycans: the syndecans (four members) and the glypicans (six members) the amino acid sequences of their core proteins and the mode of interaction with cell membranes differ between the two subfamilies. The syndecans core protein is a transmembrane protein, whilst the glypicans protein is an extracellular membrane protein bound via a glycosylphosphatidylinositol tail. As such the two types of HS glycoproteins found in the extracellular matrix can be distinguished by their unique core proteins. The HS polysaccharide chains are bound to the core proteins via the same tetrasaccharide linkers found in heparin proteoglycans. The biosynthesis pathways consist of the same synthesis/ modification pathways as the heparin reactions, however the modification of the [GlcA-(1,4)-GlcNAc]n polymer chains is more localised than in heparin; The chains also contain more GlcNAc residues and the GlcNS to GlcNAc ratio is 1:1 in HS but is 4:1 in heparin. The GlcNAc disaccharides primarily occur in the unmodified NA domains and the GlcNS occur in the NS domains where there are mixed sequences of N-acetyl and N-sulfated separate the two types of domains. The mixture of the two sequences may be significant, the primary evidence of this is the antithrombin-binding domain which contains a mixture of GlcNAc and GlcNS residues and GlcA and IdoA residues (Rabenstein 2002).

HS also has fewer O-sulfate groups than heparin and the number is far more variable, the range is 0.2-0.7 sulfates per disaccharide as compared to an average of 2.4 sulfates per disaccharide in heparin. There are four isoforms of N-deacetylase/ N-sulfotransferases, three isoforms of 6-O-sulfotransferase and five isoforms of 3-O-sulfotransferases that are known; each of which have distinct substrate recognition properties. Only single isoforms of the 2-O-sulfotransferases and C5 epimerase have been identified. There is evidence to suggest that each of these isoforms act on a different precursor sequence and produce different final structures. These reactions occur in the NS,

NA and the NA/NS domains, the structures of which are shown in figure 1.14. Most of the uronate residues are converted to IdoA, which is often then 2-O-sulfonated, thus IdoA is the major repeating disaccharide of the NS domain with its variable levels of sulfonation. The primary structure of HS is tissue specific and can be different for every cell within the same tissue. There is also evidence for regulated expression of specific domains in the HS chains however the mechanism of this regulation is unknown (Rabenstein 2002; Salmavirta *et al* 1996).

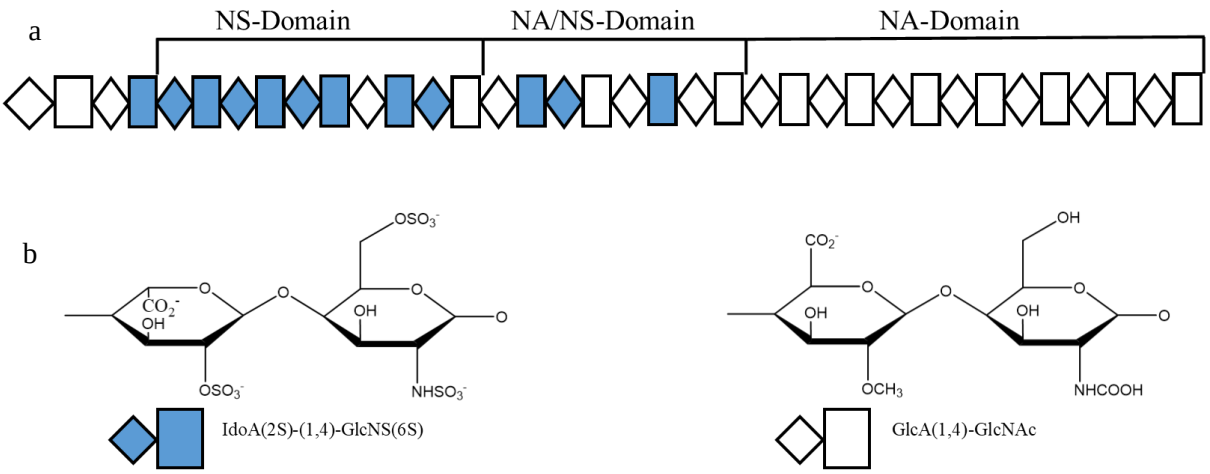


Figure 1.14. (a) An illustration of the sulfation and acetyl domains within the structure of HS. The N-acetylated residues are shown colourless shapes and the N-sulfated residues are shown as solid shapes. GlcNS and GlcNAc are represented by the rectangles and IdoA and GlcA are represented by the diamonds. **(b)** The structures of the two major repeating disaccharides of the NS and NA domains respectively. The sulfated residues are shown in blue. Figure produced in Chemdraw.

1.3.4 The Three Dimensional Structure of Heparin and Heparan Sulfate

The interactions of the heparin and HS polymer chains with proteins are determined by their sulfation pattern and the level of flexibility within their pyranose rings IdoA residues; as well as the rotations around the glycosidic linkages. Some of the information about the 3D structure has been obtained from NMR studies of native heparin and chemically modified heparin. The NMR studies of the solution structure and X-ray crystallographic studies have also been very informative in determining the compositions and structures of some of these heparin variations. The structure of one such

variation (dp30 where dp stands for degree of polymerisation) which was determined by 3D ^{1}H NMR is shown in figure 1.15 below (Khan *et al.* 2010). Heparin has a right-handed helical structure and is made up of a 1.65 - 1.73 nm repeating tetrasaccharide sequence. It is this sequence which corresponds to two-fold helices with disaccharide units in its chain which are related by a rotation of 180°. This structure is in agreement with the structure of a heparin derived hexasaccharide complexed with FGF-2 (Gatti *et al.* 1979, Mulloy *et al*, 1993, Mulloy & Forster 2000). The HS chains are typically 50-200 disaccharide units in length meaning they are about 40-160 nm in length, explaining why HS chains are a dominant feature of cell membranes. The shape of heparin and HS chains is determined by the torsion angles of the glycosidic bonds that link the monosaccharide residues.

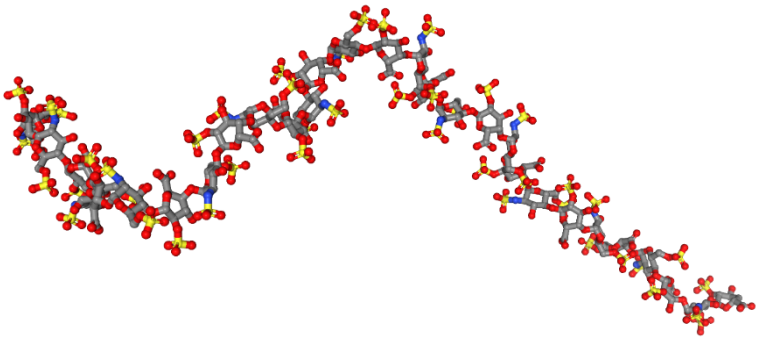


Figure 1.15 – The linear model for the dp12 heparin fragment (where dp stands for: degree of polymerisation) the structure was determined using 3D ^{1}H NMR. Figure made in COOT (PDB code 1HPN) (Khan *et al.* 2010)

1.3.5 The Ring Conformations of Heparin and Heparan Sulfate

The monosaccharide residues of heparin and HS take the $^{1}C_{4}$ and $^{4}C_{1}$ chair and $^{2}S_{0}$ skew boat conformations. NMR and X-ray structures indicate that the rings of GlcA and GlcN residues are in the $^{4}C_{1}$ chair conformation regardless of whether they are free or protein bound (Faham *et al.* 1996). The IdoA residue is more flexible, which explains how heparin and HS can bind a large range of proteins using their anionic groups to form bonds with the cationic binding sites on proteins. IdoA

exists primarily as an equilibrium of the $^{1}C_{4}$ and $^{2}S_{0}$ conformations, the position of this equilibrium depends on the position of the IdoA residue within the polysaccharide chain, as well as the substituents on adjacent GlcN residues (Chuang *et al. 2000*). The rate of interchange between the different conformations is fast enough to allow it to be observed in 1H and 13C NMR. The relative quantities of the $^{1}C_{4}$ and $^{2}S_{0}$ conformations are determined from $^{3}J_{H\text{-}H}$ coupling constants, and are roughly 40:60 for IdoA located on the reducing terminal (Chuang *et al.* 2000). This position of equilibrium is sensitive to the binding of biological molecules. The population of the $^{1}C_{4}$ conformation of both IdoA(2S) residues increases upon binding to histamine, this allows heparin to optimise its interactions with IdoA(2S) carboxylate groups. Once heparin forms a complex with FGF-2 the two IdoA(2S) conformations are locked in; ring C is in the $^{1}C_{4}$ conformation and Ring E is in the $^{2}S_{0}$ skew boat conformation (Faham *et al.* 1996). The structures of these conformations are shown in figure 1.16.

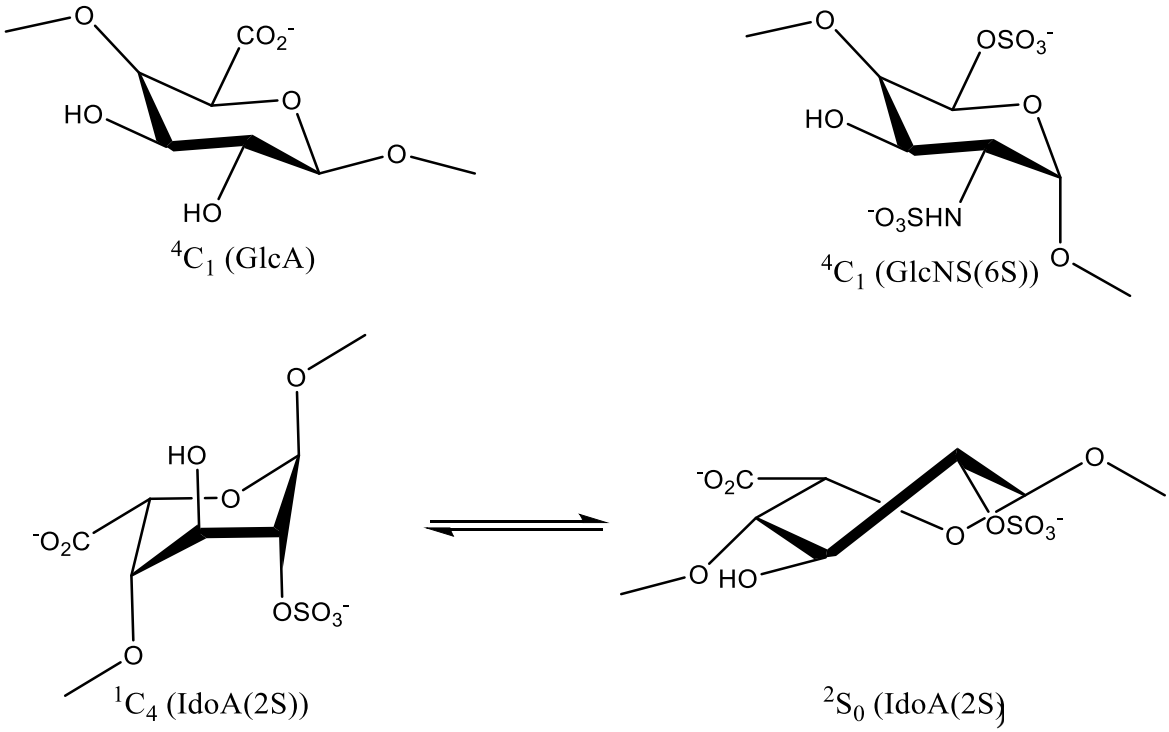


Figure 1.16 – The structures of the different conformations of the GlcA, GlcN and IdoA residues of heparin and heparan sulfate. The IdoA monosaccharide exists as an equilibrium between the $^{2}S_{0}$ and the $^{1}C_{4}$ conformations with the point of equilibrium more to the left, whereas the Glc residue exists only in the low energy $^{4}C_{1}$ conformation. Figure produced in Chemdraw.

1.3.6 The Biosynthesis of Heparin and Heparan Sulfate

Heparin and HS both have similar synthesis pathways, their core proteins are synthesised within the endoplasmic reticulum before they are transported to the Golgi apparatus, at which point a tetrasaccharide linker is attached to areas which are rich in ser-gly repeats (Lindahl 1989). This linkage is then used to synthesise a repeating α1-4 linked co-polymer of D-glucuronic acid and N-acetyl-D-glucosamine through the step-wise addition of UDP activated sugars. During the synthesis of this linear polymer it is modified through the action of N-deacetylase, C-5 epimerase and 2-,6- and 3-O-sulfotransferases. Complete modification of this polymer results in the formation of heparin which is N- and O-sulfo, L-iduronic acid rich. Incomplete modification of this polymer results in the formation of HS, which is rich in O-sulfo, N-acetyl-D-glucosamine and D-glucuronic acid (Li *et al.* 1997). Although all of the enzymes that are involved with this synthesis have been identified, cloned and expressed it is still unclear how this process is controlled. In total 22 enzymes are associated with the biosynthesis process, each of which have many isoforms that are cell and tissue specific and as such account for the tissue specific forms of HS. As most of these enzymes have transmembrane regions they readily intercalate with the proximal Golgi apparatus membranes. The first controlled step is the formation of the tetrasaccharide linker, which starts with the addition of a xylose to the serine residue of the core protein by xylosyltransferase 1 or 2. This is followed by the addition of two galactose molecules in sequence by galactosyltransferase 1 and 2, the structure is then completed by the addition of GlcA by glucuronic acid transferase 1. Following this, the HS synthesis pathway differs from the chondroitin sulfate (CS) and dermatan sulfate (DS) synthesis pathways by the addition of GlcNAc over GalNAc**.** The elongation of the linker is controlled by a complex of EXT1 and EXT2 (two enzymes of the EXT family) (Presto *et al.* 2008) the complex is responsible for the systematic addition of a GlcA followed by a GlcNAc moiety which gives rise to the alternating copolymer found in all HS/ heparin chains. Once the chain has been fully built up the modification of the chain begins with the sulfation of some of the GlcNAc residues through the removal of their acetyl group by *N*-deacetylase followed by its sulfation by one of the four isoforms of *N*-sulfotransferase enzymes. NDST2 has been shown to be specific for the GAG chains which are

bound to serglycin which makes it essential in the process of synthesising heparin (Kreuger *et al.* 2012). Once sulfation is complete further modification of the chains occurs around these regions which leaves these as sulfated domains/ S-domains. These sulfation domains only account for approximately 30% of the HS chain with the other 70% being unmodified, whilst heparin contains long stretches of modified disaccharides (Knox *et al.* 2002). After the completion of sulfation some of the GlcA residues are epimerized to iduronic acid with the vast majority of these being further sulfated at C2. The sulfation of GlcNAc and GlcNS at the C6 position is performed by one of the three isoforms of 6-O-sulfotransferase enzymes, whilst these enzymes are capable of performing both of these sulfations it has been shown that it has a preference for acting on areas with a higher degree of GlcNS flanked by 2 sulfated residues supporting the synthesis of longer S domains (Qin *et al.* 2015). The rarest type of modification is performed by 3-O-sulfotransferases which have the largest number of isoforms out of all of the HS associated enzymes. Only 3 of the 7 have a transmembrane region which suggests that the vast majority of them do not localise to the Golgi like the others (Kreuger *et al.* 2012). This sulfation occurs on the C3 position of the GlcNS which gives rise to the most highly sulfated regions of HS. Because of their similar biosynthesis pathways, all of the disaccharide subunits contained within the GAGs are the same; as well as some of the extended sequences within them such as the ATIII pentasaccharide binding sites (Lindahl *et al.* 1980). The synthesis of these binding sites needs a 3-O-sulfo group to be placed on the central glucosamine residue in the F position.

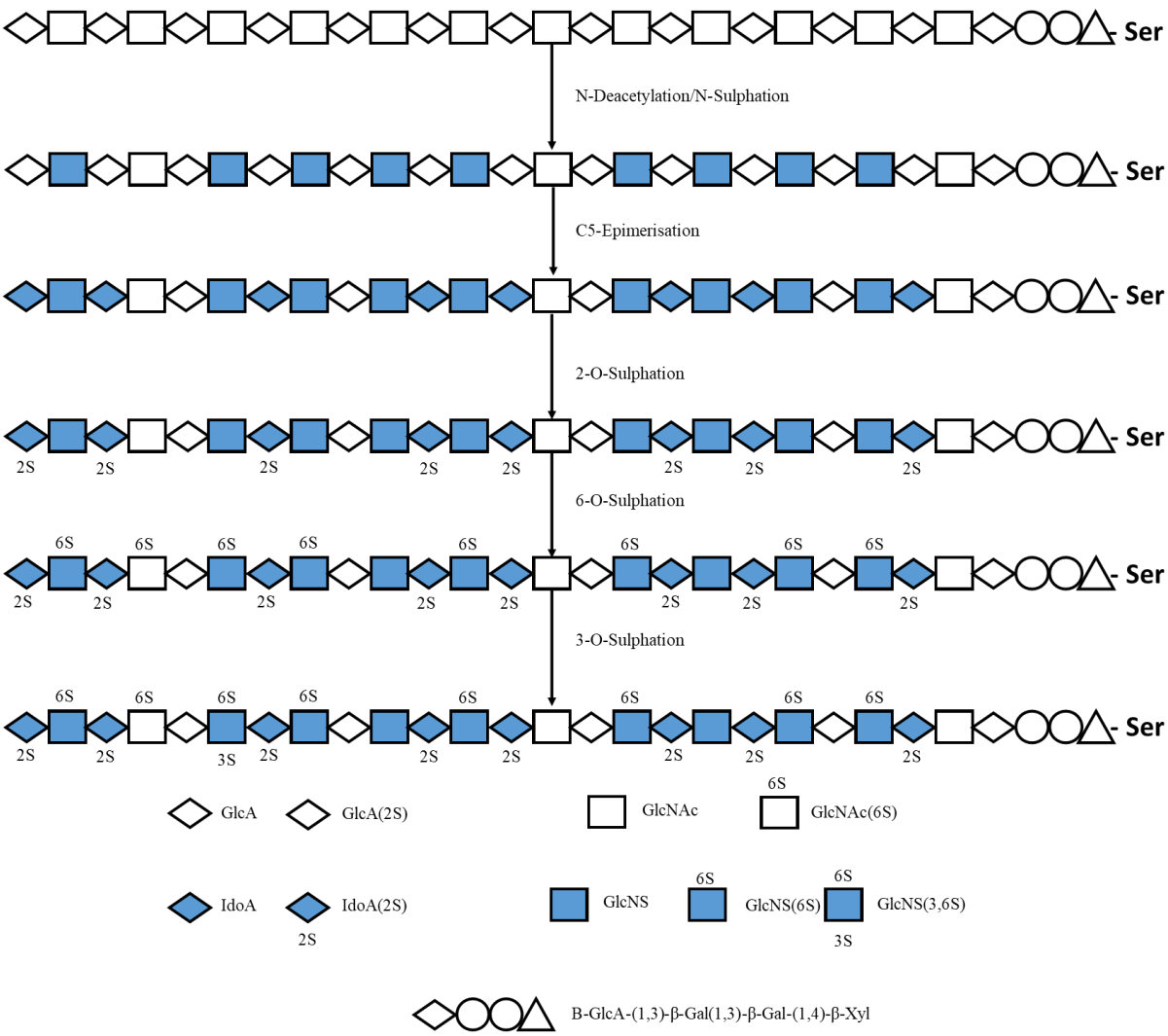


Figure 1.17 – The systematic residue modification which occurs during the biosynthesis of the heparin and HS polysaccharide chains. The polysaccharide chain is initially bound to its core protein at the serine residue via a linker tetrasaccharide. Figure produced in Chemdraw.

1.3.7 Chemical Synthesis

Artificial synthesis of heparin and HS is still the most common and reliable source of well-defined heparin and HS oligosaccharides; this is despite the considerable effort and resources required for the synthesis of heparin and HS, as well as the greatly diminished returns (Dulaney & Huang 2012). At its heart the chemical synthesis of heparin and HS involves either building up oligosaccharide sequences from scratch using disaccharide sequences or through the chemical modification of pre-existing polysaccharides to include sulfate groups. Orthogonal protecting groups are placed in key positions in order to maintain regioselectivity and stereoselectivity during the glycosylation phase.

Synthetic linkers can also be added to a pre-existing sugar chain. Due to the large range and difficulty of reactions associated with synthesis, the yields of these reactions are often very low; however several studies have focused on how to overcome these issues and increase the yields (Linhardt & Liu 2012). Performing glycosylations with the same disaccharide building block is the most typical route for generating heparin/HS oligosaccharides. Using this method many compounds of differing lengths can be synthesised leading to the formation of compounds with regular repeating patterns in them. The diversity of the heparin/HS oligosaccharides can also be emulated by changing the functional group transformations of the basic skeleton.

1.3.8 Interaction with proteins

Protein binding by heparin and by HS chains is primarily electrostatic due to their large negative charges, it involves interactions between cationic ammonium, guanidinium and imidazolium side chains of proteins and the anionic sites on the heparin or HS proteoglycans. Over the years many heparin and HS binding proteins have been identified (Conrad 1998). The best way to identify and purify these proteins is through heparin affinity chromatography. This is because of the nonselective interactions of heparin-binding domains with anionic heparin (Salmivrta *et al.* 1996). This information has led to the thought that all heparin-protein interactions are nonspecific, however there is now a key train of thought that sequences in the NS domains and NA domains of the HS chains are designed for selective interactions with specific proteins, it is these interactions which result in the regulation of heparin binding protein activities. Selectivity in the binding can be achieved through the use of GlcNS(3,6S) which is present in the antithrombin-binding sequence of heparin and essential for its anticoagulant activity (Lindahl *et al.* 1998). The pattern of N-substitution present may also play a part in a unique binding epitope as well as the combination of N- and O-substituents or even the length and positions of the NS, NA and NS/NA domains may also have an effect. Other known protein binding sequences are made up of common disaccharides; typically there are 3-9 disaccharide units are involved with protein binding although longer sequences that wrap around the surface of protein oligomers can sometimes be involved. Heparin binding domains have been identified in a large number of proteins; typically they contain large numbers of lysine and arginine

and in some cases histidine (Rabenstein 2002). The X-ray structure of the hexasaccharide complex with FGF-2 shows that the functional groups of the side chains on some of the amino acids can also interact with binding sites on heparin. Several heparin binding domains have been observed in large quantities on multiple proteins. X-ray studies have detailed that the side chains of certain other amino acids (e.g. the amide side chains of arginine and glutamine) also interact with sites on heparin. Several consensus sequences have been proposed as the heparin binding sites on the proteins. These sequences are made up of basic residues and include the sequences: XBBBXXBX, XBBXBX and XBBBXXBBBXXBBX; where B and X are basic and hydropathic residues respectively. These sequences have been proposed to adopt secondary structures where the basic residues are aligned along one exposed face and the X amino acids point into the protein core. Due to their high negative charges heparin and HS are polyelectrolytes with a fraction of their negative charge neutralised by bound counter cations. This supports the idea that the binding reaction is an ion exchange process, with the cationic sites of the bound peptides serving as counter ions (Sobel *et al.* 1992).

1.3.8.1 Interaction of a Heparin Disaccharide with Heparinase II

Native low molecular weight heparin disaccharide molecules were introduced to a solution of native heparinise II (*HepII*) in order to visualise the heparin binding site. The structure of heparin bound *HepII* was determined to a resolution of 2.30 A. The structure of the HeppII crystalised with the heparin disaccharide showed that the heparin binding site is a deep, elongated 30-35 A canyon that is formed at the top of the N-terminal domain and also includes several loops from the central domain. A central ridge divides the canyon into two distinct halves, one narrow and shallow and one narrow and deep. The nonreducing end of the disaccharide was found to bind directly to the deeper half of the binding site. The canyon is maintained by a series of hydrogen bonds produced by a series of mainchain and sidechain atoms, including the Tyr-436, Asn-405, Arg-261, Glu-205, and Arg-148 residues. The final three residues also form salt bridges with the uronic acid component of heparin. The uronic acid is held within the binding site through interactions with the His-406, Arg-261, Glu-205, and Arg-148 residues. Shaya *et al.* 2006

1.3.8.2 Interaction of heparin with a mutant heparin lyase

Heparinase I (Hep-I) is a heparin lyase that specifically depolymerises heparin molecules by cleaving the glycosidic linkage of the iduronic acid residues. The structure of Hep-I comprises of a β-jelly-roll domain containing a long, deep substrate binding canyon formed by the inner sheet of the domain. The binding site of Hep-I consists of the Glu73, Tyr75, Lys185 and Asp253 residues and is partially covered by a lid formed by two loops comprising of the Gly74-Glu79, Thr250-Asn257 and the Lys184-Val187 residues. The structural integrity of this domain is aided by a Ca^{2+} ion, bound to the hinge region of Hep-I. The canyon has been found to have a strong positive charge, favourable to forming ionic interactions with a strong negatively charged molecule such as heparin. During heparin binding, the Thumb domain of Hep-I has been noted to move towards the substrate. Han *et al.* 2009.

1.3.9 Clinical Applications of Heparin

The classical use of heparin has been as the major anticoagulant drug of the past 60 years, its main function is to inhibit the formation of clot and thrombi. This is especially important during and after invasive surgery such as open heart surgery or after the body has experienced trauma. Heparin is one of the most prolific drugs used today, second only to insulin. The reason that it is so widely used is because it interacts with many of the proteins that are associated with the blood clotting cascade such as thrombin and antithrombin. The usefulness of heparin as a clinical drug primarily stems from its ability to bind to antithrombin and accelerate its activity, which results in fast and increased inhibition of the coagulation factors Xa and IIa. In a biological sense the mode of action that heparin is involved in is more akin to a defence mechanism against exogenous pathogens and damage limitation during tissue damage; heparin accomplishes this by altering cytokine levels in response to trauma or infection. Heparin limits the release of proinflammatory cytokines such as IL-1, IL-8 and TNFα whilst promoting the release of anti-inflammatory cytokines such as IL-10 (Lima *et al.* 2017).

Exogenous heparin is clinically used as an anticoagulant and is the initial choice for an anticoagulant of use in the treatment and prevention of thromboembolic diseases. HS with its universal distribution

in animal tissues also has a wide range of functions including cell adhesion, regulation of cell growth and proliferation, development processes, an effect on blood coagulation, cell surface binding of lipoprotein lipase and other proteins, angiogenesis, viral invasion, tumour metastasis, protection from protein degradation, regulated protein transport through membranes and mediates the internalisation of proteins. HS proteoglycans on the cell surface is in a dynamic state of being constantly trafficked rapidly to the cell surface and turning over quickly, with its structure constantly changing in response to extracellular signals like growth factors. The method of action displayed by heparin and HS revolves around non-covalent interactions with other proteins to effect changes in their conformation to facilitate protein to protein interactions allowing for protein sequestration and to function as co-receptors for growth factors (Mulloy & Forster 2000, Rabenstein 2002).

1.3.10 Inhibition of Blood Coagulation

Coagulation of the blood relies on a host of precursor-serine protease reactions, which activate to produce thrombin. Thrombin then produces fibrin which is the primary crosslink in a blood clot. Coagulation is usually regulated by antithrombin which is a serine protease inhibitor, antithrombin is relatively inactive in its natural state and requires an interaction with HS in order for it to act as an inhibitor. Heparin acts as an inhibitor as it complexes with antithrombin (hence is an activator) so it accelerates the rate at which serine proteases are inactivated (Rabenstein 2002). Because of this activity it is used as a clinical anticoagulant; however because it naturally sequesters in mast cell granules it is argued that heparin is not a natural anticoagulant. The unique antithrombin binding pentasaccharide that is associated with heparin mediates the anticoagulant activity of heparin. The pentasaccharide is required for the inactivation of factor Xa by antithrombin; however a much longer sequence which includes the pentasaccharide is required for the inactivation of thrombin. Only about 50 % of all heparin is anticoagulant active (binds to antithrombin) the main way that this fraction was identified was through observing the interactions of oligosaccharides produced from the recent depolymerisation of heparin with 1H NMR. In order for thrombin to be deactivated the larger sequence needs to first bind the thrombin using a separate binding sequence and then bind to antithrombin and bring the two together so they can form a tight complex (Dowd *et al.* 1999).

1.3.11 Heparin Induced Thrombocytopenia

Heparin induced thrombocytopenia (HIT) is a severe side effect of heparin treatment for blood clots; it is caused by heparin dependent IgG antibodies which are generated against the platelet factor 4 (PF4)/ heparin complex. This complex is formed when PF4, a member of the C-X-C family of chemokines binds to heparin molecules. PF4 does this through its lysine rich moieties which follow the typical amino acid binding sites for heparin found on proteins (Ziporen *et al.* 1998). The heparin molecules "bundle" the PF4 molecules which results in a conformational change of the PF4; these changes then become the binding sites for the HIT-IgG. There are several of these binding sites on PF4, recent studies have discovered that they are located on the apposition points where PF4 tetramers are brought into proximity through charge neutralisation by heparin there is a minimum charge density and chain length required for heparin interaction with PF4. The complex of antibodies and H/PF4 activate the FcyRIIa receptors on platelets. The need for this conformational change was proved by Horsewood et al. when they performed an experiment on unbound, linear PF4 which would not bind to HIT-IgG but it would bind when it was in the presence of heparin which proved that these conformational changes were brought about by the binding of one heparin molecule to four to eight PF4 molecules (Carter *et al.* 2005).

PF4 is a small 70 residue long protein with four carboxyl-terminal positively charged lysine rich regions which are used to bind heparin. Additionally PF4 contains four cysteines that are used to form internalised disulfide bonds (Poncz *et al.* 1987). The Pro37 residue located near the surface of the tetramer on the side opposite the heparin binding sites, was found to be a key IgG binding residue when it was substituted for an N residue which was not expected to cause a structural change in the binding site. Despite this, once this substitution had occurred IgG was no longer able to bind to the H/PF4 complex which proved that P37 was important to IgG binding. Because P37 is far (relatively) away from the heparin binding rings it is thought that once the heparin binds to the complex the P37 is exposed; as well as the four amino acids which are adjacent to the P37 and the homologous residues in neutrophil activating peptide-2 (NAP-2) all of which are buried in a hydrophobic pocket created by beta sheet strands (Ziporen *et al.* 1998).

The stability of the heparin/PF4 complex depends on the relative concentrations of heparin and PF4 and can be characterised by their size and the nature of the electrostatic surface charge interactions. It has been observed that when not in the presence of heparin the PF4 particles are dispersed due to their strongly positive electrostatic charges repulsing each other. However these positive charges are neutralised by increasing the concentration of heparin leading to macromolecular aggregation. Once heparin is in molar excess the repulsive forces start to build up again which leads to the dispersion of the molecules once again (Suvarna *et al,* 2007). Other studies have found that these macro aggregates can differ in terms of size, charge and composition depending on the type of heparin that was used as well as the physiological conditions used. It was found that the larger complexes formed when a higher concentration of PF4 was used (the biggest formed with a PF4-heparin molar ratio of 13:1) (Bertini *et al,* 2017). The majority of pharmaceutical grade heparin is porcine in origin due to the fear of prion contamination in bovine and ovine sources, however in 2007 and 2008 there were several reported cases of porcine heparin from China with over O-sulfated chondroitin sulfate which resulted in fatalities. As a consequence the US is now considering supplementing the porcine heparin with bovine heparin which is still commonly used in some South American countries (Guerrini *et al.* 2008). It has been shown however that ovine heparin requires a lower [PF4] than both porcine and bovine heparin; this effect may be due to the higher average molecular weight and sulfation levels found in ovine heparin (Bertini *et al.* 2017).

1.3.12 Low Molecular Weight Heparins

The low-molecular-weight heparins (LMWH) are one of the most important forms of heparin in regards to clinical use. This class of heparin is made up of a mixture of heparin derived fragments which range from tetra to hexadecasaccharides; they are obtained via a wide array of chemical and enzymatic depolymerisation reactions (Viskov *et al.* 2009). The reason that LMWHs are thought to be superior to unfractionated heparin (UFH) is because they have a higher bioavailability and a higher anti-factor Xa/ anti-factor IIa activity by comparison to UFHs as well as a decreased haemorrhagic risk during surgery (Hoppensteadt *et al.* 2005). However LMWHs cannot be

monitored and their action cannot be reversed during procedures due to neutralisation with protamine sulfates being ineffective, so this carries an increased risk of HIT (Xu *et al.* 2012a).

LMWHs also have some potential as a treatment for cancer related haemostatic disorders through the modulation and control of FGF signalling pathways between cancerous cells (Xu *et al.* 2012b). LMWHs are once again the more desirable treatment due to their much more predictable anticoagulant response which is important in order to control any potential bleeding associated with the cancer and their extended half-life within the body, as well as having a reduced chance of causing HIT (Rickles & Falanga 2001). The prevention and treatment of these cancer-related haemostatic conditions as well as controlling the expression of tumour cell-associated clotting factors has a clear and direct impact on not only the easing of the associated symptoms but to the survival of the cancer patients. Whilst LMWH and HS have the often unwanted side effect of anticoagulation activity when being used to treat cancer, they do have the ability to alter signalling defects as well as the activity of sulf enzymes. Due to their role in the partial inhibition of the inflammation response it has been proposed that heparin could be used as a potential treatment of cystic fibrosis through the inhibition of elastase (Lima *et al. 2017*).

Due to the presence of LMWHs on all cell surfaces they are expected to act as broad-spectrum non-specific virus-binding receptors, this interaction is backed up by the fact that viral envelopes present patches of electropositive amino acids. It has been long known that viruses have the ability to bind to surface heparin and HS (in the form of herpes simplex and dengue virus). Because of its virus binding capabilities heparin has been suggested as a potential inhibitor for viral attachment; the theory of which has been proved through experiments carried out with dengue virus, however what this did not explain was how the process could be used to allow heparin to be used as efficient inhibitors (Lin et al. 2002). Heparin has also been shown to act against infection from influenza H5N1as well as acting on the fusion envelope protein of the Zika virus and preventing the programmed cell death that is associated with the virus by preventing its adhesion (Ghezzi et al. 2017).

1.4 Aims and Objectives

The main aim of this work was to determine the mechanism via which, the Fc region of human IgG binds to specific analogues of the heparin polymer. The secondary aim of this work was to determine what effect this interaction had on the structure of the two molecules, in order to observe the implications to the functionality of the two molecules. To this end, the work initially focused on a wide range of ligand binding studies of human IgG-Fc, with a specific focus on the well conditions that produced crystals which diffracted to a high resolution. This area was focused on as it would provide a starting point for the crystallisation trials, required to obtain a structure of this interaction. Once this interaction had been observed, work would then have moved on to focusing on the binding mechanisms and affinities of other heparin polymers with the IgG-Fc, as well as comparing and contrasting their binding mechanisms to each other.

In order to do this, two samples of human IgG were prepared. The first of these samples was a native Human IgG-Fc fragment derived from Human Sera and procured from Bethyl Laboratories (Cat: P80-104), from here on referred to as native IgG. The second protein sample was a recombinant Human IgG-Fc fragment procured from Sino Biological (Cat: 10702-HNAH), from here on referred to as recombinant IgG-Fc fragment. The amino acid sequence of this sample included residues from Glu99-Lys330 and a single Cys/Ser mutation was introduced at Cys103. In addition to this, the expression hosts for this sample was HEK293 cells.

Chapter 2 – Mass Spectrometry Analysis and Sequence Determination

2.1 Introduction

2.1.1 An Introduction to Proteomics

Proteomics is the large scale study of the genetic information contained within the sets of proteins produced by an organism, otherwise known as a proteome. Specifically proteomics aims to explain the structural and functional information stored within the proteomes (Fraser *et al*., 1997). Proteomes are not static and can be different between cells and can change over time. Proteomics can be used to study a wide variety of protein characteristics such as: protein expression, the life cycle of proteins and protein modification. Because of this, proteomics is often thought of as a genome-wide assay which can differentiate and study cellular states (Fraser *et al.,* 1997). The production of genomic DNA marker databases has expedited the evolution of proteomics as a technique to the point where studying hundreds of thousands of proteins is a realistic and rapid process (Fraser *et al.,* 1997).

In proteome analysis, it is expected that any data which is obtained by direct proteome analysis complements the data that is obtained by more readily available methods of proteome analysis such as high throughput gene sequencing (Aebersold and Goodlett, 2000). Despite recent advances in proteome analysis, it is still largely impossible to determine the certain properties of proteins using proteomics alone; such as their location. Mass spectrometry (MS) techniques largely focus on protein-protein interactions and post-translation modifications (Aebersold and Goodlett, 2000).

The classical approach to proteomic research has been through the purification and homogeneity of the proteome via sequential fractionation and assay cycles. However, these techniques have been superseded by more powerful modern techniques like mass spectrometry (Fraser *et al.,* 1997).

2.1.2 An Introduction to Mass Spectrometry

MS is an analytical chemical technique where the analytes are ionised and the ratio of their mass to charge is measured. MS was discovered in the early 1900s however its use was only limited to chemical sciences. The invention of Electrospray ionisation and Matrix Assisted Laser Desorption/ Ionisation (MALDI) in the 1980s allowed biological molecules to be analysed by MS (Singhal *et*

al., 2015). Recently, large protein masses of over 100,000 Da have been determined, this means that because of the ease with which proteins can be ionised, MS is considered to be an excellent complementary technique to the more classical techniques of protein analysis; such as nuclear magnetic resonance (NMR), circular dichroism and x-ray diffraction (Vertes, 2008). Since its introduction, MS has had much success in studies of the primary, secondary, tertiary and quaternary structures of proteins, as well as in studies of the structures of folded polypeptides (Chait *et al.*, 1992).

Mass spectrometers are made up of an ion source, a mass analyser which measures the mass to charge ratio (m/z) of the analytes and a detector which registers the ions which strike it. A computer then records the number of ion strikes at each m/z value and provides a readout. The two most commonly used proteomic mass spectrometric techniques are Electrospray Ionisation (ESI) and MALDI (Spengler *et al.*, 1992). There are four types of mass analysers that are commonly used in proteomics: Ion trap, time-of-flight, quadrupole and Fourier transform ion cyclotron (FT-MS) analysers (Bruno and Ruedi, 2006). Despite all having the same purpose, the ways in which these analysers function is very different and they all have their own strengths. Ion trap analysers are robust and inexpensive (relative to the other analysers which are available), however they are far less accurate in determining the mass of the peptide analytes. They work by trapping the ions for a length of time before being subjected to MS/MS analysis. The FT-MS analyser traps the ions in a high vacuum using a strong magnetic field; whilst it is much more accurate and sensitive, it is extremely expensive and has very poor peptide fragmentation efficiency and as such is not often used for proteomic evaluation. Time-of-flight (TOF) analysers are used when measuring intact peptides (Aebersold and Goodlett, 2000).

2.1.3 MALDI – TOF MS

MALDI is an MS technique which was developed by Karas and Hillenkamp in the 1980s. It produces singly charged ions either through the addition or loss of a single proton. Due to the fact that there is minimal sample fragmentation during MALDI analysis it is described as a soft ionisation technique. The sample used in MALDI is prepared by mixing the analyte with a solution of an energy absorbent, organic compound called the matrix. The matrix crystallises when dry, trapping the

analyte within the crystal complex as a co-crystal (Patterson, 1995). The matrix is then exposed to a laser beam, during this process the matrix absorbs light at the ultraviolet wavelength of 337nm and converts the light to heat energy; this causes a small amount of the matrix (100nm depth) to heat up and vaporise along with the sample (Patterson and Aebersold, 1995; Singhal *et al.*, 2015).

As with the other MS techniques the protonated ions are accelerated under a known fixed potential, this then causes the ions to separate based on their m/z and detected by the detector (Patterson and Aebersold 1992). TOF detectors are used for biological molecules such as proteins, during the MALDI-TOF analysis the m/z of an ion is determined by the length of time it takes for the ion to travel to the detector (Webster and Oxley, 2009). Some types of TOF analysers contain a mirror at the end of the flight tube which bounces the ions back towards the detector, this effectively doubles the length of the flight tube which in turn allows for smaller machines as well as correcting small energy differences within the sample (Wilm & Mann, 1996).

Identification of molecules by MALDI TOF is performed by comparing the peptide mass fingerprint (PMF), (the spectrum produced based on the TOF information of the molecule) of the molecule with PMFs contained within a database such as the mascot database. The main issue with peptide analysis using MALDI TOF is that it has an inherently reduced resolving power. This is caused by the spread of a small amount of kinetic energy during the ionisation process. This reduction in resolving power means that the natural isotope distribution of the peptides cannot be observed. Historically, two instruments have been developed in order to combat these effects; the first, a reflectron is a device which applies a decelerating voltage to the opposite end of the flight tube to the ion source (Wiley & McLaren, 1955). This decelerating voltage is higher than the accelerating voltage and as such "reflects" the ions back towards a secondary detector which is next to the ion source. Reflectrons work on the basis that higher energy isotopes will penetrate further into this voltage than lower energy isotopes before being reflected. This means that the higher energy isotopes will strike the detector at the same time as their lower energy counterparts as their flight path is longer (Wiley & McLaren, 1955). The second device is called a time-lag focuser. This device works by waiting a few second after the ions are produced to apply the accelerating voltage, this results in the energy of the

higher energy ions being reduced to a level closer to the lower energy ions resulting in a higher resolution (Vestal *et al.,* 1995).

2.1.4 An Introduction to ESI Mass Spectrometry

ESI-MS is a MS technique, developed in the late 1980s to the early 1990s which is capable of analysing molecules with a range of molecular weights and polarities within a biological sample. ESI uses electrical energy to ionise molecules and transfer them into the gas phase (Dewald, 2009). This is especially effective when the analyte is in solution. This change of phase contains three steps, the first of which is the dispersal of a fine spray of charged droplets. The electrical field strength becomes high enough where it causes the solvent drops to evaporate which causes the ions at the surface of the drop to be ejected in the gaseous phase. These ions are then sampled by a skimmer cone before being accelerated into the mass analyser (Charles, McLukey and Glish, 1994).

During the nebulisation stage the sample solution is fed into a capillary tube into a high electric field at the tip of the tube. This electric field pulls positive charge towards the front of the liquid, when the electrostatic force overcomes the surface tension of the small charged droplets. The droplets then travel through the inert gas to the counter electrode. The capillary tube is more positive than the electrode. Electrospray is defined as the dispersion of electrically charged droplets, in order to do this two steps are required, they are the droplet formation and charging. The forming of these droplets does not present an issue if the flow rate, surface tension and electrolyte concentration are low enough; if one of these variables increases it makes it more difficult for the electric field to produce the required aerosol charge (Bruins, 1998). The strength of the electric field at the tip of the capillary can be increased in order to overcome any of the adverse effects that are produced by changes in the variables. However if the electrical field is too high, an electrical discharge will occur alongside the electrospray ejection, the consequence of this is disruption of the drop formation. In the majority of ESI situations positive charge on the droplets is produced by the removal of the negative charge by discharging the negative ions against the metal wall of the capillary (King *et al., 2000*). In some conditions it is possible to directly remove electrons from molecules that have a low ionisation

energy. Because the electrons are consumed in order to neutralise the positive ions in the positively charged droplets, an ESI source is an example of an electrolysis cell (Bruins, 1998).

The ions are evaporated once the strength of the electric field overcomes the surface tension of the drop. However if the ions have more than one solvation shell it may separate from the droplet as a nanodrop which loses the solvent molecules during its flight time. Both of these processes leads to naked sample ions which are taken into the mass spectrometer. If a sample ion has not been desolvated before it enters the vacuum system it will be desolvated when it passes through the curtain gas (Ho *et al.*, 2003).

ESI mass spectrometers are usually paired with a quadrupole mass analyser which is an assembly of four equidistant metal rods which run parallel to each other. The rods are arranged so that each rod is placed in a different corner of an imaginary diamond. The pairs of opposite rods are electrically connected, with a DC voltage superimposed with an equal but opposite AC voltage. The electrical field that results from this arrangement causes the analytes to travel along the rods (in the Z plane) whilst oscillating in the X-Y plane. The amplitude of this oscillation is directly related to the m/z of the analytes and is controlled by the DC and AC voltages in order to prevent the analytes from hitting the quadrupole rods, allowing them to travel their full lengths into the detector. This means that undesirable ions can be neutralised and removed before they reach the detector (Jonscher & Yates, 1997).

During ESI-MS the signal is directly proportional to the analyte concentration and is mostly independent of the flow rate that is used for the introduction of the sample, hence ESI acts as a concentration sensitive detector. Because of the relative softness of their ionisation, both ESI and MALDI have been used to generate intact ionic proteins, with non-covalent interactions. Other ESI based studies have focused on the use of hydrogen-deuterium exchange in order to examine some higher order structural features (Akashi, Naito & Takio, 1999). An example of a structural feature which can be examined using this technique is solvent accessibility; this is based on the observation that hydrogen atoms which are exposed to solvent are exchanged with deuterium atoms at an

increased rate compared to hydrogen atoms which are shielded from solvent. This technique can also be used to detect hydrogen bonding within a protein, this is based on the observation that hydrogen atoms bound up in hydrogen bonds exchange at a slower rate than hydrogen atoms that are not involved in hydrogen bonds (Akashi, Naito & Takio, 1999). Using this concept the denaturation of an α-helix has been observed by Anderegg *et al*. Other experiments have proved without the use of H-D exchange that ESI-MS can be used to monitor the denaturation of proteins in solution (Katta & Chait, 1991). These experiments use the observation that the mass spectra produced by ESI-MS of denatured proteins show a higher charge state than those of an identical protein that has not been denatured. Other applications of ESI-MS in the study of higher order protein structure include the identification of residues which are juxtaposed residues through cross-linking with side-chain specific linkers (Konermann *et al.,* 1997; Loo *et al.,* 1991; Mirza, Cohen & Chait, 1993).

When speed and sensitivity of ESI is greatly increased when it is coupled with high performance separation techniques, such as High Performance Liquid Chromatography (HPLC) or Fast Protein Liquid Chromatography (FPLC). This is because their implementation reduces the amount of peptide required for complete sequence characterisation to a femtomolar level (Aebersold and Goodlett, 2000).

2.1.5 The Principles of Peptide Mass Mapping

Peptide mass mapping is based on the observation that proteins which undergo proteolysis with a specific protease, produce much shorter polypeptide sequences. The masses of these sequences create a unique mass finger print for that protein and that protein only (Patterson & Aebersold 1998). This means that if a protein database is searched using the selected mass fingerprint then the protein which the fingerprint originated from should be correctly identified. The basic techniques which can be used to automate peptide mass mapping are as follows: The sample protein is lysed using sequence specific cleavage enzymes. These enzymes target specific residues which means that any peptides they produce will have known amino and carboxylate terminals. The peptides are then analysed using mass spectrometry to determine their masses (Webster *et al.,* 2009). The database search is then performed by digesting the proteins it contains *in silico* using the same enzyme as was

used in the experiment. This produces a list of possible theoretical masses which are then compared to the set of experimental masses. During this comparison, an algorithm assigns a score to each of the theoretical masses depending on how well they match the set of experimental masses, where a higher score indicates a better match (Henzel *et al.*, 1998).

Peptide mass mapping does however have some limitations, for example; as protein determination requires several peptide masses, which are derived from the same protein to correlate with a set of experimental masses, it is not suitable for use when the target protein is in a mixture of several proteolysed peptides (Patterson and Aerbersold 1998). This is because it is often difficult to determine which peptides in the mixture correlate to a specific protein. Another limitation of the technique is that when a pure protein sample is digested and analysed not all of the predicted peptide peaks may be observed and some of the observed peptide masses may not be present in the database. Predicted peaks are typically not observed due to a number of issues that can arise before analysis; these include: poor sample solubility, selective ionisation, the peptides being very short and the introduction of contaminants to the sample. Unassigned peptide masses indicate a much greater problem as they can easily lead to misidentification of the sample. Unassigned masses are observed when; the mass of the target peptide is changed due to post-translational modifications, it is changed through artificial modifications which are introduced during the preparation of the sample. Unassigned peaks can also be observed when the sample has not been completely proteolysed or contaminant proteases have been introduced during preparation. The final reason for unassigned peaks is that there are multiple proteins in the sample, which is usually caused by contamination during sample preparation (Jensen *et al.* 1996).

2.2 Aims

Due to the fact that the Fc fragments that were to be used for the structural studies were purified from human serum, there was a chance that any of the four common subtypes could have been present within any crystals that were produced.

2.2.1 Methods

In order to determine which subtype was present within the protein crystals the IgG Fc sample was subjected to two different mass spectrometry (mass spec) techniques: Matrix Assisted Laser Desorption/Ionisation Time of Flight (MALDI-TOF) and Electrospray Ionisation (ESI) mass spectrometry.

2.2.2 Materials

A 100 mM solution of NH_4HCO_3 was produced by dissolving NH_4HCO_3 in filtered, deionised water, a 50mM stock of NH_4HCO_3 was then prepared using this solution. RapiGest (Waters) was suspended in 50 mM NH_4HCO_3 to produce a solution of 10mg/ml. A 1mg/ml trypsin solution was then produced by trypsin in 1mM HCl. A 100mM solution of DL-dithiothreitol (DTT) was prepared by dissolving DTT in deionised H_2O. iodoacetamide (IAA) was dissolved in deionised H_2O to prepare a 100mM solution. Finally a 30%v/v solution of trifluroacetic acid was prepared.

2.2.3 Production of Peptide Fragments for Proteomics

A 10 pmol/µl sample of the IgG-Fc from human serum was taken from a previously prepared 8mg/ml stock solution (please see section 3.2.1.1 for details), 13µl of the Rapigest solution was added to this sample followed by 7µl of the DTT solution. Once both of these solutions were added to the sample, the sample was heated to 80˚C for 15 minutes; the sample was then allowed to cool to room temperature before it was vortexed gently for 30 seconds. Next 13µl of the IAA solution was added before the sample was incubated in the dark for 60 minutes. After this, a further 18µl of the DTT solution was added to the sample, 25µl of the prepared NH_4HCO_3 solution was added to the sample to give the sample a final volume of 125µl.

The sample was then divided into two equal aliquots of 62.5µl each. The trypsin and GluC solutions were both diluted to 1:10 solutions of 10µl each. 1.2µl of the trypsin solution was added to one of the samples and 1.2µl of the GluC solution was added to the other. Both of the samples were then incubated at 30˚C for 8 hours before a further 1.2µl of trypsin digest and the GluC digest was added to their respective samples and the samples were incubated at 30˚C for a further 16 hours. In order

to stop the reaction 1µl of the TFA solution was added to each sample, the samples were then incubated at 37˚C for 45 minutes before being centrifuged at 13,000 rpm for 10 minutes.

2.2.4 Data Analysis

Data analysis was performed using the MASCOT server. The mass peaks produced by MALDI and ESI were compared with the *in silico* library of digested peptides and assigned a score based on how well they matched the library peptides. For the MALDI datasets, this library search was performed with a lower bound mass cut-off of 100 ppm and a maximum of 1 missed cleavage for both the GluC and trypsin digests. This narrowed down the library to peptides that fell under these conditions, these peptides were then reduced to only those that could have originated from human IgG-Fc. It is these peptides that the MALDI digests were compared to.

For the ESI results; the mass cut-off was increased to 200 ppm and the number of potential missed cleavages was increased to two. This was because the initial results only contained 7 hits and by increasing these parameters an additional 3 hits came back whilst still allowing a high degree of confidence that they were indeed peptides that could have originated from IgG-Fc.

2.3 Results

2.3.1 MALDI Results

Two proteolytic digests were carried out on the IgG-Fc from sera, one with the enzyme trypsin and the other with GluC. This was performed alongside the electron density analysis as a complementary technique in order to determine the sequences of the IgG-Fc isoforms that were present in the solution to aid in the sequence determination. The spectrum of the MALDI sample which was digested with trypsin showed ten clear peaks which corresponded to nine known sequences within the IgG-Fc region which are shown in figure 2.1. One of the sequences: K.DTLMISR.T was identified twice by the search. The peak at 835.4093 identified the sequence without any form of methionine oxidation and the peak at 851.4106 indicated the same sequence with oxidation of the methionine residue, thus giving it an Mr that was roughly 16 Da greater. There were also several smaller peaks which could have been associated with some of the other sequences that are typically

associated with the Fc fragment of IgG which were identified using the peptide mass program which is a part of the UniProt web server (Wilkins *et al.*, 1997*;* Gasteiger *et al*., 2005). The spectra did however contain a high degree of noise which made the identification of some of the smaller peaks difficult and contributed to the low degree of identification.

The MALDI was able to identify 10 sequences with the number of missed cleavages set to 2 and the mass cut-off set to 100 ppm, this cut-off was chosen as having a smaller cut-off would have resulted in the identification of sequences that were too short to be useful and a higher cut-off would have reduced the accuracy of the results. All of these sequences identified the subtype that was present was IgG1, however the sequence coverage was only 35% and so was not enough to tell what the full sequence was or even which of the subtypes was actually present as some of these sequences were also present in the other subtypes as well. These sequences alongside the sequence coverage are shown in figure 2.2.

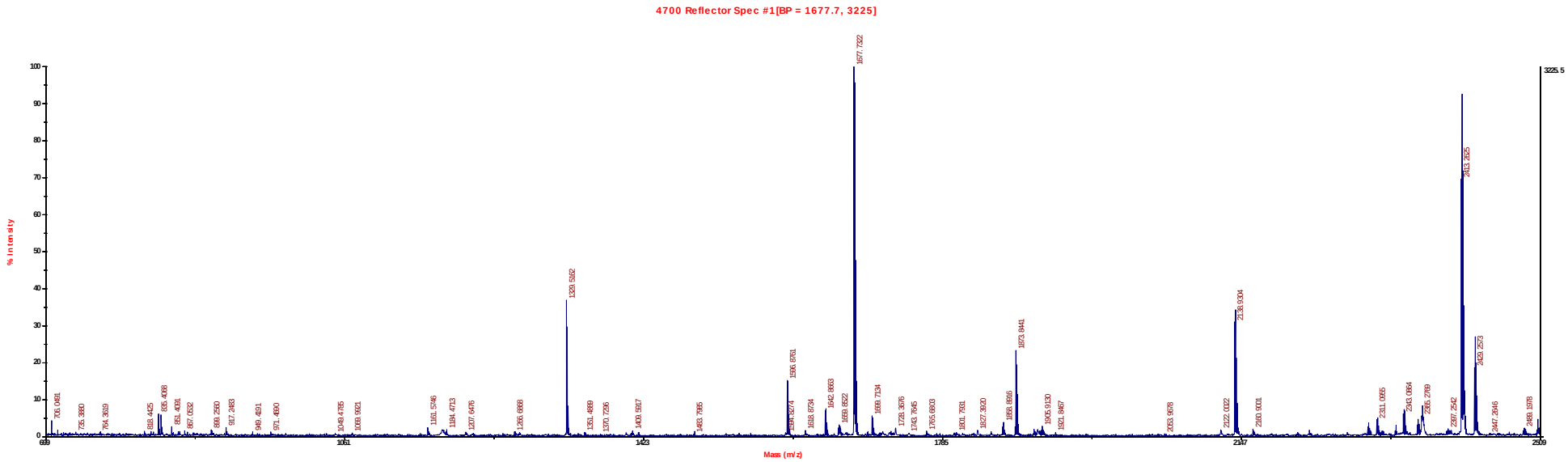


Figure 2.1 – Mass Spectrum of the Trypsin Digest of the Human IgG-Fc purified from Sera

(A)

1	ASTKGPSVFP	LAPSSKSTSG	GTAALGCLVK	DYFPEPVTVS	WNSGALTSGV
51	HTFPAVLQSS	GLYSLSSVVT	VPSSSLGTQT	YICNVNHKPS	NTKVDKKVEP
101	KSCDKTHTCP	PCPAPELLGG	PSVFLFPPKP	K**DTLMISRTP**	**EVTCVVVDVS**
151	**HEDPEVKFNW**	**YVDGVEVHNA**	**K**TKPREEQYN	STYR**VVSVLT**	**VLHQDWLNGK**
201	**EYK**CK**VSNKA**	**LPAPIEK**TIS	KAK**GQPREPQ**	**VYTLPPSRDE**	**LTKNQVSLTC**
251	**LVK**GFYPSDI	AVEWESNGQP	ENNYK**TTPPV**	**LDSDGSFFLY**	**SK**LTVDKSRW
301	QQGNVFSCSV	MHEALHNHYT	QKSLSLSPGK		

(B)

Start	–	End	Observed	Mr(expt)	Mr(calc)	ppm	M	Peptide
132	–	138	835.4093	834.4020	834.4269	-29.9	0	K.DTLMISR.T
132	–	138	851.4106	850.4033	850.4218	-21.8	0	K.DTLMISR.T + Oxidation (M)
139	–	157	2138.9311	2137.9238	2138.0202	-45.1	0	R.TPEVTCVVVDVSHEDPEVK.F
158	–	171	1677.7318	1676.7245	1676.7947	-41.9	0	K.FNWYVDGVEVHNAK.T
185	–	203	2228.1004	2227.0931	2227.2001	-48.0	1	R.VVSVLTVLHQDWLNGKEYK.C
206	–	217	1266.6960	1265.6887	1265.7343	-36.0	1	K.VSNKALPAPIEK.T
210	–	217	838.4753	837.4680	837.4960	-33.4	0	K.ALPAPIEK.T
224	–	243	2311.1040	2310.0967	2310.1968	-43.3	2	K.GQPREPQVYTLPPSRDELTK.N
244	–	253	1161.5739	1160.5666	1160.6223	-48.0	0	K.NQVSLTCLVK.G
276	–	292	1873.8429	1872.8356	1872.9146	-42.1	0	K.TTPPVLDSDGSFFLYSK.L

Figure 2.2. (A) The sequence coverage of the peptide sequences that were produced by MALDI mass spectrometry for the Trypsin digest of the native Fc fragment from sera. The start of the CH_2 domain is shown in green and the matched peptides are shown in bold red. (B) The data for the peptide sequences of the native IgG Fc from sera that were identified by the MALDI for the trypsin digest.

The MALDI spectrum for the GluC digest also had some large clear peaks as shown in figure 2.3 below, for example the peak at 1752.8345. There were also many more of the smaller peaks as well as a much higher level of noise than the trypsin digest and as such it was more difficult to identify any additional peaks usually associated with a GluC digest of IgG-Fc. The analysis of the spectra by the Mascot server was not able to detect any sequences that were associated with the Fc fragment however it was able to identify 17 sequences from the Allophycocyanin alpha subunit of *Synechococcus sp. (ATCC27264),* these sequences are shown in table 2.1. There were also 16 sequences that were associated with the Methylenetetrahydrofolate-tRNA-(uracil-5-)-methyltransferase of *Bacillus subtilis (strain 168).* These sequences are shown in table 2.2.

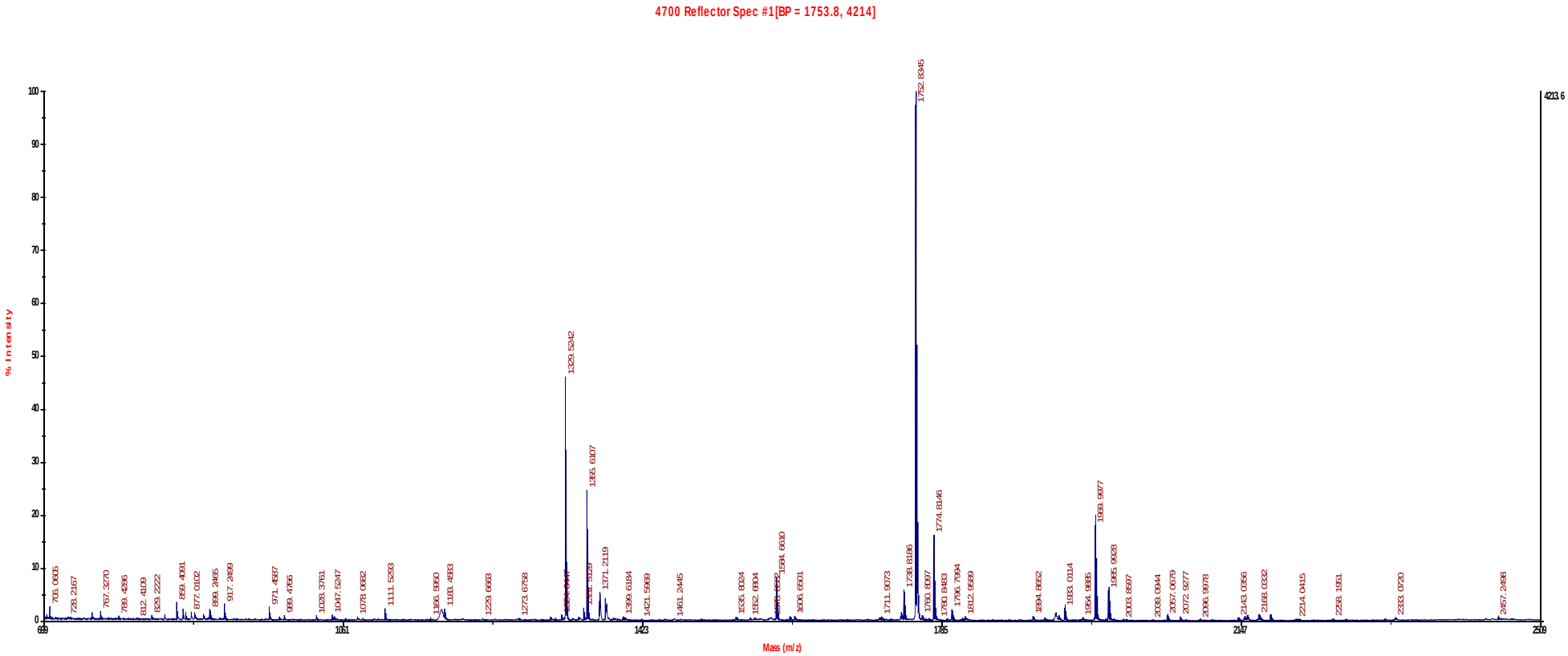


Figure 2.3. MALDI Mass Spectrum of the GluC Digest of the Human IgG-Fc purified from Sera

Table 2.1. The mass spectrometry data for the sequences of the Allophycocyanin alpha subunit of *Synechococcus sp. Strain ATCC27264* that were identified by the MALDI for the GluC digest.

Observed	Mr(expt)	Mr(calc)	ppm	Start		End	Miss	Peptide
859.4105	858.4032	858.4923	-103.79	49	-	56	0	E.RIIKSAGD.A
971.4598	970.4526	970.5447	-94.97	106	-	114	1	E.EIGLVGVRE.M
983.4484	982.4411	982.4212	20.3	77	-	84	0	E.MTATCLRD.M + Oxidation (M)
1329.525	1328.5182	1328.5411	-17.19	133	-	145	1	E.MKAVATGMMSGDD.A + Oxidation (M)
1345.5399	1344.5326	1344.5360	-2.50	133	-	145	1	E.MKAVATGMMSGDD.A + 2 Oxidation (M)
1493.8125	1492.8052	1492.7443	40.8	1	-	14	1	E-.MSIVTKSIVNADAE.A + Oxidation (M)
1584.6636	1583.6563	1583.6629	-4.18	133	-	148	2	E.MKAVATGMMSGDDAAE.A
1600.7476	1599.7403	1599.6579	51.5	133	-	148	2	E.MKAVATGMMSGDDAAE.A + Oxidation (M)
1808.8470	1807.8397	1807.7611	43.5	145	-	161	3	D.DAAEAGAYFDYVIGAME.- + Oxidation (M)
1814.7542	1813.7469	1814.0162	-148.46	49	-	64	1	E.RIIKSAGDALFQKRPD.V
1933.0139	1932.0066	1932.0905	-43.39	25	-	41	1	D.RIKAFVTSGESRLRIAE.T
2039.0393	2038.0320	2037.9646	33.1	125	-	144	1	D.AVAQAVREMKAVATGMMSGD.D + Oxidation (M)
2079.0062	2077.9989	2078.0718	-35.07	106	-	124	2	E.EIGLVGVREMYKSLGTPVD.A + Oxidation (M)
2143.0382	2142.0309	2141.9358	44.4	65	-	84	2	D.VVSPGGNAYGEEMTATCLRD.M + Oxidation (M)
2154.0058	2152.9985	2152.9915	3.26	125	-	145	2	D.AVAQAVREMKAVATGMMSGDD.A + Oxidation (M)
2257.1880	2256.1807	2255.9021	124	133	-	154	3	E.MKAVATGMMSGDDAAEAGAYFD.Y + 3 Oxidation (M)
2457.2236	2456.2163	2456.0981	48.1	125	-	148	3	D.AVAQAVREMKAVATGMMSGDDAAE.A + 3 Oxidation (M)

Table 2.2. The mass spectrometry data for the sequences of the Methylenetetrahydrofolate-tRNA-(uracil-5-)-methyltransferase of *Bacillus subtilis strain 168* that were identified by the MALDI for the GluC digest.

Observed	Mr(expt)	Mr(calc)	ppm	Start		End	Miss	Peptide
859.4105	858.4032	858.5215	-137.77	370	-	376	0	D.LLIFPKE.T
971.4598	970.4526	970.5811	-132.42	424	-	431	1	E.RALEILKE.W
983.4484	982.4411	982.3888	53.2	189	-	196	0	E.AYLNCPMD.Q
989.4794	988.4721	988.5341	-62.77	212	-	219	1	E.RAPIKEFE.K
1063.5159	1062.5086	1062.5134	-4.54	218	-	225	1	E.FEKQVHFE.G
1355.6133	1354.6060	1354.5169	65.8	189	-	199	2	E.AYLNCPMDQDE.Y
1370.7093	1369.7020	1369.7817	-58.15	129	-	142	0	E.KGIVIVASGPLTSE.S
1399.6342	1398.6269	1398.7870	-114.47	424	-	434	2	E.RALEILKEWIE.K
1535.8027	1534.7954	1534.8144	-12.35	290	-	302	1	E.QKRVFSFIPGLED.A
1600.7476	1599.7403	1599.7893	-30.59	99	-	111	2	D.REGFSRYITETLE.D
1735.8344	1734.8271	1734.8941	-38.59	290	-	304	2	E.QKRVFSFIPGLEDAE.F
1752.8375	1751.8302	1751.7647	37.4	218	-	231	2	E.FEKQVHFEGCMPVE.V + Oxidation (M)
1753.0750	1752.0677	1751.7647	173	218	-	231	2	E.FEKQVHFEGCMPVE.V + Oxidation (M)
1774.8175	1773.8102	1773.8832	-41.13	174	-	188	1	D.MTKVFRASRYGKGEE.A + Oxidation (M)
1926.0339	1925.0266	1925.0370	-5.41	113	-	128	3	D.HPLIDIHREEVTKIPE.K
2168.0496	2167.0423	2167.1861	-66.35	417	-	434	3	E.RGRAQAERALEILKEWIE.K

2.3.2 ESI Results

Due to the fact that the sequence coverage produced by MALDI analysis was only 35%, one aliquot from each of the digested samples was also used in ESI analysis of the Fc fragment. The GluC digest identified a total of 23 peptide sequences of which nine of them were not assigned to any protein hits by the mascot server, this left 14 that were all associated with IgG Fc. Of these 14 peptides, six of them were found to be associated with IgG1 Fc, six were associated with IgG3 Fc and 2 were associated with IgG2 Fc. The sequence coverage for IgG1 was 27.6%, for IgG2 it was 32.3% and for IgG3 it was 28.7%. One of the identified peptides was also present within the sequence of IgG4 Fc and gave a sequence coverage of 11.5% however it was not identified as being present in IgG4 Fc by the server. This sequence was identified as being present in the amino acid sequence of IgG4-Fc using the BLAST software on the NCBI server. Whilst there were 14 peptide hits, only 5 of these were unique peptide sequences. One of these sequences was identified by 8 of these hits, across all three of the identified subtypes (albeit with some differential oxidation). The second peptide accounted for a further three hits (being found in all three subtypes). The final three peptides were all only identified once each, with each one being unique to a different subtype. These sequences along with their corresponding data can be found in appendix 1.

When the sequences obtained from all four of the digests between these two techniques were combined, the final overall sequence coverage for IgG1-Fc was 84.5% and all but four of the identifying residues (*Tyr 291, Tyr296, Ala339 and Pro445)* had been observed. The coverage for IgG2-Fc, IgG3-Fc and IgG4 was considerably lower, at 28.2%, 12.1% and 11.3% respectively.

Table 2.3. The mass spectrometry data for the sequences of the IgG1-Fc that were identified by ESI analysis of the sample of serum IgG-Fc that was disgested by GluC. The cut-off parameters were -200 ppm and a maximum number of two missed cleavages. The sample was purified using FPLC before being inserted into the ionisation chamber of the ESI machine.

Query	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Unique	Peptide
3	792.7769	1583.5393	1583.7138	-110.16	1	67	8.2e-005	1		E.VTCVVVDVSHEDPE.V
11	685.8033	2739.1842	2739.5033	-116.49	0	(99)	7.4e-008	1		E.LLGGPSVFLFPPKPKDTLMISRTPE.V
12	914.0761	2739.2064	2739.5033	-108.41	0	158	8.1e-014	1		E.LLGGPSVFLFPPKPKDTLMISRTPE.V
13	689.8008	2755.1740	2755.4983	-117.69	0	(78)	8.9e-006	1		E.LLGGPSVFLFPPKPKDTLMISRTPE.V + Oxidation (M)
14	919.4057	2755.1952	2755.4983	-109.99	0	(130)	6.2e-011	1		E.LLGGPSVFLFPPKPKDTLMISRTPE.V + Oxidation (M)
16	972.3779	2914.1119	2914.4245	-107.26	1	92	4.8e-007	1	U	E.MTKNQVSLTCLVKGFYPSDIAVEWE.S

```
IGHG2       ASTKGPSVFPLAPCSRSTSESTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPAVLQSS    60
IGHG3       ASTKGPSVFPLAPCSRSTSGGTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPAVLQSS    60
IGHG1       ASTKGPSVFPLAPSSKSTSGGTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPAVLQSS    60
5W5M        ------------------------------------------------------------    0
IGHG4       ASTKGPSVFPLAPCSRSTSESTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPAVLQSS    60

IGHG2       GLYSLSSVVTVPSSNFGTQTYTCNVDHKPSNTKVDKTVERKCC-----------------    103
IGHG3       GLYSLSSVVTVPSSSLGTQTYTCNVNHKPSNTKVDKRVELKTPLGDTTHTCPRCPEPKSC    120
IGHG1       GLYSLSSVVTVPSSSLGTQTYICNVNHKPSNTKVDKKVE---------------------    99
5W5M        ------------------------------------------------------------    0
IGHG4       GLYSLSSVVTVPSSSLGTKTYTCNVDHKPSNTKVDKRVE---------------------    99

IGHG2       ---------------------------------VECPPCPAPPV-AGPSVFLFPPKPKDT    129
IGHG3       DTPPPCPRCPEPKSCDTPPPCPRCPEPKSCDTPPPCPRCPAPELLGGPSVFLFPPKPKDT    180
IGHG1       --------------------------PKSCDKTHTCPPCPAPELLGGPSVFLFPPKPKDT    133
5W5M        ---------------------------------TCPPCPAPEAAGASSVFLFPPKPKDT    26
IGHG4       --------------------------SKY---GPPCPSCPAPEFLGGPSVFLFPPKPKDT    130
                                          ** ****::::::************

IGHG2       LMISRTPEVTCVVVDVSHEDPEVQFNWYVDGVEVHNAKTKPREEQFNSTFRVVSVLTVVH    189
IGHG3       LMISRTPEVTCVVVDVSHEDPEVQFKWYVDGVEVHNAKTKPREEQYNSTFRVVSVLTVLH    240
IGHG1       LMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLH    193
5W5M        LMISRTPEVTCVVVDVSQEDPEVQFNWYVDGVEVHNAKTKPREEQFNSTYRVVSVLTVLH    86
IGHG4       LMISRTPEVTCVVVDVSQEDPEVQFNWYVDGVEVHNAKTKPREEQFNSTYRVVSVLTVLH    190
            ******************:*****:*:********************:***:********:*

IGHG2       QDWLNGKEYKCKVSNKGLPAPIEKTISKTKGQPREPQVYTLPPSREEMTKNQVSLTCLVK    249
IGHG3       QDWLNGKEYKCKVSNKALPAPIEKTISKTKGQPREPQVYTLPPSREEMTKNQVSLTCLVK    300
IGHG1       QDWLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSRDELTKNQVSLTCLVK    253
5W5M        QDWLNGKEYKCKVSNKGLPSSIEKTISKAKGQPREPQVYTLPPSQEEMTKNQVSLTCLVK    146
IGHG4       QDWLNGKEYKCKVSNKGLPSSIEKTISKAKGQPREPQVYTLPPSQEEMTKNQVSLTCLVK    250
            ****************:**::*******:***************::*:************

IGHG2       GFYPSDISVEWESNGQPENNYKTTPPMLDSDGSFFLYSKLTVDKSRWQQGNVFSCSVMHE    309
IGHG3       GFYPSDIAVEWESSGQPENNYNTTPPMLDSDGSFFLYSKLTVDKSRWQQGNIFSCSVMHE    360
IGHG1       GFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNVFSCSVMHE    313
5W5M        GFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSRLTVDKSRWQEGNVFSCSVMHE    206
IGHG4       GFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSRLTVDKSRWQEGNVFSCSVMHE    310
            *******:*****:*******:****:***********:*********:**:********

IGHG2       ALHNHYTQKSLSLSPGK    326
IGHG3       ALHNRFTQKSLSLSPGK    377
IGHG1       ALHNHYTQKSLSLSPGK    330
5W5M        ALHNHYTQKSLSLSLGK    223
IGHG4       ALHNHYTQKSLSLSLGK    327
            ****::*********:**
```

Figure 2.4. The Sequence coverage of all of the IgG Fc subtypes produced by the MS/MS analysis of a GluC digest and a trypsin digest performed on the Purified human serum IgG Fc Fragment (P80-204). The matched peptides are shown in bold red, the highlighted green residues indicate the start point of the CH2 domain and as a result the start of the Fc region. The * represents homologous residues within the sequences and **:** represents residues which are not homologous across the subtypes. 5W5M (Tam *et al.* 2017), was the structure that was used as the start point for the map analysis, the bold residues within this sequence indicate where its sequence differed from the sequences that were identified by the mass spectrometry analysis. The data for the sequences is located in Appendix 1. The sequences were obtained from UniProt and the alignment was produced using Clustal Ω.

2.4 Discussion

The analysis of the sequence of the IgG-Fc isoforms within the protein sample by mass spectrometry was performed alongside the x-ray crystallographic analysis of the structure to aid in the correct identification of the isoform of IgG-Fc that was present in the protein crystals. Two samples produced by digestion with two different enzymes: trypsin and GluC, were used due to their different lysing patterns. Trypsin cleaves peptide bonds at the C-terminal of lysine and arginine residues and GluC, cleaves peptide bonds at the C-terminal of glutamic acid and aspartic acid residues. The differential lysing was important because it meant that they would produce different sequence fragments and as a result give better sequence coverage. This proved to be true as the two samples that ran on the ESI identified sequences that the other missed, which improved the sequence coverage from 65.3% that came from just the GluC digest to 84.5% across the two samples.

Both ESI and MALDI were performed on both of the digest samples once again in order to improve the sequence coverage because the two techniques produce sequence fragments of different lengths with MALDI producing longer sequences than ESI mass spectrometry. This meant that there was a greater chance of producing a sequence with one of the techniques that the other failed to pick up on and as such they complement each other. This proved to be an effective technique due to the fact that the GluC sample, run on the MALDI failed to recognise any sequence fragments that were associated with IgG-Fc. As a result of this the MALDI produced a maximum of 35% sequence coverage for IgG1-Fc, which was not high enough to confirm the sequence. Once the ESI analysis was performed this improved the sequence coverage to 84.5%. However, the 49.5% sequence coverage produced by ESI was not enough to conclusively identify the sequence of any of the IgG subtypes that were in the sample. Because of this it was required to combine the results of both of the mass spectrometry techniques in order to gain enough information on the sequences to be able to definitively say which of the sequences are present in the sample.

The mass spectrometry proved to be an effective companion technique to the x-ray diffraction analysis of the protein structure. This was due to the fact that determining the sequences of the IgG-Fc isoforms IgG1-Fc, IgG2-Fc and IgG3-Fc within the sample, meant that the potential IgG-Fc

sequences were known and such made fitting the backbone into the electron density easier and made determining which isoform was in the structure more accurate. The accuracy of the mass spectrometry analysis can further be confirmed by the relative proportions of the sequences that were identified. The largest proportion of the sequences that it identified were from IgG1-Fc (84.5%), with lower sequence coverage for IgG2-Fc (28.2%) and IgG3-Fc (12.1%). These values reflected the reported abundances of the IgG isoforms in human serum: 60% IgG1, 32% IgG2, 4% IgG3 and 4% IgG4.

Chapter 3 – Protein Crystallisation and Structure Determination

3.1 Introduction

3.1.1 An Introduction to X-ray Crystallography

X-ray crystallography is the most direct method of producing models of macromolecules such as proteins. It was discovered in the early 19th century and saw huge development throughout the 19th and 20th centuries (Hunefeld, 1840). Crystallographers have produced thousands of three dimensional images of macromolecules, these images have provided detailed information on the activity of these macromolecules as well as their binding mechanisms as well as any conformational changes they may undergo (McPherson 1991, 1999).

Protein crystallography is a technique used to produce magnified images of proteins using x-rays instead of a traditional lens (Drenth, 1994). This due to the need to use light with a wavelength of no longer than 3.5 Å in order to be able to image the atomic structure of macromolecules (Helliwell, 1984). This is because covalent bonds between atoms are between 1-2Å and strong polar interactions are between 2.5-3.5Å (Cornilescu, G *et al,* 1999), hence visible light would not be able to image the atomic structure as its wavelength is between 350 nm and 750 nm (Drenth, 1994; Engh & Huber, 1991).

In a laboratory setting X-rays are produced in a vacuum by focusing a beam of electrons at a metal target, in order to do this the potential difference between the cathode which releases the electrons and the target must be in the order of 10s of kilovolts (Hammond, 1997; Gruner, 1994). This potential difference needs to be so great in order for the electrons to possess enough energy so that when they strike the target enough energy is transferred to cause the electrons within the metal atoms to transition to a higher energy state and release photons (X-rays) (Ladd & Palmer, 1993). The X-rays produced by this method have wavelengths of between 0.5 Å and 1.6 Å. These X-rays have wavelengths that are just shorter than the 'soft' X-ray region and as a result they penetrate biological materials enough to be scattered by the entire volume of any target crystals (Miyahara *et al,* 1986).

3.1.2 X-ray detectors and Sources

In X-ray crystallography the scattering by the crystal can be viewed from all possible directions, this is because the X-ray beam is incident from as many directions as possible, and as much of the scattered radiation as possible is recorded (Smyth & Martin, 2000). Because of the way that this data is collected, the images which result from crystallographic studies are 3-dimensional (Als-Nielsen & McMorrow, 2001).

X-ray detectors are devices that react to the intensity of X-ray beams that they are exposed to (Gruner *et al*, 1994). At a given wavelength, the intensity of the X-ray beam measured by these detectors is a direct measure of the number of photons the beam delivers. The classical X-ray detectors were sheets of photographic film. In which silver iodide crystals absorb X-ray photons when exposed to an X-ray beam (Su & Zang, 2014; Wiechenberger & Pozharski, 2013). This absorption releases enough energy to create an imperfection in the crystal lattice of the film; allowing a diffraction image to be built up. When the image is being developed these imperfections allow the silver iodide to be reduced to silver which causes the emulsion on the film to blacken producing spots. These spots are then used to build up the diffraction image (Su & Zang, 2014).

More modern detectors utilize the ionisation of gases, caused by X-rays passing through them to create a current which the detector can interpret (Als-Nielsen & McMorrow, 2001). The intensity of the X-ray is proportional to the rate at which this current is transmitted. In order to detect the X-ray scattering using this type of detector, an absorber with a small central hole is required (Gruner, 1994). However as the scattering can only be detected when X-rays strike this central hole, this type of detector can only observe diffraction one point at a time. Because no X-ray lenses have been discovered, modern X-ray detectors are required to perform computational transformations to mimic the glass lens of a microscope and ‘view’ the diffraction image (Drenth, 1994).

Diffractometers are instruments which manipulate the orientation of protein crystals as well as the detector position; all while keeping the crystal in an X-ray beam. However; because they measure electron energies sequentially, they are slow when experimenting with large crystals. Additionally,

they can also only be used in non-biological studies due to the complexity of the diffraction patterns that biological compounds produce (Bukowska & Gru, 2013).

In the time since the 1990s, many position sensitive detectors have been developed. These detectors are as accurate as counters but can view much larger areas of the sample, making them more effective (Gruner, 1994; Rhodes, 2000). The most widely used position sensitive detector is the image plate, in which the diffraction image is created when atoms within the plate are promoted to a higher energy state when they are struck by photons during exposure to an x-ray source. The images produced in this way are "read" by scanning a beam of light across the plate which causes the excited atoms to fluoresce. The intensity of this fluorescence is proportional to the total energy of the x-rays which struck the plate. It is this fluorescence which is used to build the image. The stored image can then be erased by "flooding" the plate with light, allowing the plates to be reused. However much like the photographic film, this method does not record the time of flight of the X-rays and a separate method is required in order to record the image (Rhodes, 2000; Su *et al, 2014*).

3.1.3 A Brief Overview of Macromolecular Crystal Point Groups

Many protein molecules are made up of multiple identical peptide chains in symmetrical arrangements. Proteins usually have the same number of symmetry operations as chains; e.g. dimers have 2-fold symmetry and trimers have 3 fold symmetry (Mazey 1990). Tetramers tend to have 2-fold symmetry around each of the 3 perpendicular directions (within 3-dimentional space). This is referred to as 222 symmetry, due to the 3, 2-fold axes (Holser, 1958). Due to the constraints of the crystal lattice, protein crystals can only have 2, 3, 4 or 6-fold symmetries limiting the number of possible point groups to 11 (International Union of Crystallography, 1995).

The final type of point group is the cubic point group. These have 3-fold symmetry and can be observed by visualising the crystal in the corner of a cube. Cubic point groups have 3 fold axes in 4 cardinal directions (Holser, 1958). The most simple cubic point group (23) also has 2-fold axes in 3 perpendicular directions. Another more complicated example (432) contains 4-fold axes in these

directions as well as 2-fold axes at 45° to these 4-fold axes (International Union of Crystallography, 1995).

3.1.4 An Introduction to Bragg's Law and Macromolecular Crystal Space-Groups

Bragg's law is a physical principal which predicts the angle of reflection for a diffracted beam, which has been diffracted by specific atomic planes and is defined by the following equation:.

$$2d\ sin\theta = n\lambda$$

This equation applies when the interplanar spacing of the planes, the wavelength of the incidence beam and the order of reflection are known (Bragg, 1975).

Bragg's law can also be defined using Laue equations, which define the following set of conditions under which diffraction occurs:

$$a \cdot S = h$$

$$b \cdot S = k$$

$$c \cdot S = l$$

Where h, k, l are the Miller indices; a set of values that define a unique plane of reflection, *a*, *b* and *c* are the lattice vectors and S is the difference between the incidence and reflected vectors for the hkl plane (Helliwell, 1984)

The unit cell of a crystal can be described as a cube whose boundaries correspond to the 3 lattice translations: **a**, **b** and **c,** with angles between the translations of α, β and γ (Als-Nielsen & McMorrow, 2001). These lattice translations describe both the length and direction of the boundaries of the unit cell and whilst they can be a range of lengths and angles, they are restricted in what these can be (Helliwell, 1984). The convention is to describe the unit cell by the smallest possible volume that can create an entire crystal lattice using only symmetry operations. These are referred to as primitive unit cells and are described by primitive lattice translations. The position of a body within a unit cell

is defined by the values *x*, *y* and *z*, these values represent the distances along the three axial directions (Bragg 1975; Helliwell, 1984; Als-Nielsen & McMorrow, 2001).

While crystal lattices can freely have translational symmetry; rotational symmetry places restrictions on the shape of the unit cell of a crystal. This is because the unit cells must stack so they fill of the space within the crystal, whilst also fulfilling the symmetry. This means that in order to have 2-fold symmetry, the unit cell must have two lattice translations that are perpendicular to the fold axis. For example; if the fold axis is in the *b* direction, then there must be a lattice translation in the *a* and *c* directions. In the case of 3-fold, 4-fold and 6-fold symmetries one of the lattice translations must be parallel to the fold axis, whilst the other two must be perpendicular to the fold axis e.g. if the fold axis is in the *c* direction then one lattice translation must also be in the *c* direction and the other two must be in the *a* and *b* directions (Bragg 1975; Helliwell, 1984; Als-Nielsen & McMorrow, 2001).

The space group of a crystal is the complete set of symmetry operations that make up a three-dimensional lattice and define its shape. The smallest unit of a crystal that can generate a complete crystal structure is called the asymmetrical unit, it is called this as it does not have or require any symmetry operations (Wukovitz & Yeates, 1995; Ramakumar & Viswamitra, 1990).

The P1 space group is the simplest as it has no symmetry except the lattice translations, the P indicates that the unit cell is primitive rather than face of body centred and that the highest rotational symmetry is 1-fold (International Union of Crystallography, 1995; Leslie, 2000). It has been demonstrated that if a tetrameric molecule has two perpendicular 2-fold axes, then it must also have a third 2-fold axis which is perpendicular to the other two (Mighell, *et al,* 1983). This observation also applies to lattices. This is because the axis directions (***a***, ***b***, ***c***) are assigned as the lattice dimensions; this means that any lattice which has three perpendicular 2-fold axes must be rectangular/ orthorhombic in shape. The most simple orthorhombic space group is the 222 space group, in this space group all three of the 2-fold axes intersect at the origin (Wukovitz & Yeates, 1995).

The P222 space group is a primitive space group with three intersecting 2-fold axes. This produces four different positions within the unit cell which means that the lattice can be built up to produce a tetramer at every lattice point. The $P2_12_12_1$ space group is very similar to the P222 space group however the three 2-fold axes are replaced by 2-fold screw axes; these are axes that rotate by 180° like a 2-fold axis but also shift half a cell length in the direction of the axis. These screw axis copy lattice features at (0, y, 0), (x, 0, 0) and (0, 0, z) to (-x, y+1/2, -x), (x+1/2, -y, -z) and (-x, -y, z+1/2) respectively (Als-Nielsen & McMorrow, 2001).

3.1.5 Diffraction by Macromolecular Crystals

In diffraction by a 3-dimensional, repetitive object like a protein crystal the incident rays are scattered in many different directions (Giacovazzo *et al.* 1992; Hammond, 1997). This means that whilst all of the unit cells cause the beam to scatter diffraction will only occur if all of the scattered rays are in phase. This is because the scattered rays will experience destructive interference if they are out of phase with the other rays that they interact with. In a crystal it is only possible for all the rays to be in phase with each other if the crystal is aligned correctly within the incident beam (Gruner, 1994; Hammond, 1997).

Fourier transformation analysis is a process where a series of waves are represented as their sum. This type of analysis is utilised in crystallography to analyse the repetitive structures within the crystal (Als-Nielsen & McMorrow, 2001). Fourier transformations are performed by determining the amplitude of the Fourier components of each individual structure from the variance of the electron density that the structures produces (Drenth, 1994). The inverse Fourier transform, is performed by using the amplitudes and phases of the waves to calculate the density of electrons within the crystal (Als-Nielsen & McMorrow, 2001). In structural studies of three-dimensional objects such as crystals, the object of interest needs to be rotated in order to build up the full image of the structure. This is because the object would only produce a 2-dimensional image if it was shot from one direction only (Hammond, 1997).

Due to their high solvent content (often as high as 50%) (McPherson & Gavira, 2014), crystals will only diffract to a limited resolution which is often between 3.0Å and 2.0 Å (Helliwell, 1984). The lower the solvent content of the crystal (and hence the higher protein content), the higher this resolution limit is (Okada *et al*, 2002; Mathews, 1968).

Protein crystals represent proteins in their most highly ordered form, it is this order that allows the protein molecules to be imaged (Hammond, 1997; McPherson, 1991). This is because all of the protein molecules within a crystal are identical, and in identical environments. Protein crystals are organised in such a way that each and every protein interacts with it neighbouring proteins in exactly the same way (Phillips *et al.* 1976). Each repeating unit inside a crystal is in an identical environment to all the other repeating units within said crystal. The way that these repeating units fit together causes the crystals to have flat faces and sharp edges (Rhodes, 2000; Wlodawer *et al.* 2013).

Crystals will only grow if the mother liquor that surrounds them is supersaturated; being polyions, the solubility of proteins is directly affected by the concentration of surrounding ions (Wlodawer *et al.* 2013). The presence of ionic compounds; or compounds that sequester water such as Polyethylene glycol 4000 (PEG), reduces protein solubility and therefore favours crystal growth (McPherson 1999; Ewald, 1963; Richmond *et al.* 1997). As a protein crystal grows, larger numbers of protein molecules are deposited on the surface of the crystal in an orientation that allows the pattern of crystal growth to continue. In an ideal scenario, if the protein was incorrectly oriented it would immediately dissociate from the lattice before any interactions are formed (Rhodes, 2000). However, this is not always the case and growth which continues in this vain often leads to imperfections within the lattice. Due to the irregular size of macromolecular crystals, the protein molecules cannot pack as tightly together and fill all of the space within the crystal. These voids are then filled by the mother liquor, usually in aqueous form. The ingress of the mother liquor results in the crystal becoming fragile with impurities getting trapped within it (McPherson, 1991; 1999; Rhodes, 2000).

3.2 Methods

Dried native human IgG Fc fragment (Bethyl laboratories) was reconstituted using 200 µl filtered, deionised water before being dialysed into 50 mM sodium acetate pH 4.5. The solution was concentrated up with a vivaspin 500 spin column to a concentration of 10 mg/ml.

Hampton crystallisation screen 2 performed on 8 mg/ml IgG Fc fragment A modified condition F10 (16% PEG 2K instead of 10K) was chosen to be the basis for a series of screens used to solve the structure of the native Fc fragment.

The Dp12 (degree of polymerisation) heparin was produced and dried by the Skidmore research group, before being made up to a final concentration of 10 mM in filtered deionised water.

Once the native structure of the Fc fragment was solved, a batch of recombinant Fc fragment (Cat: 10702-HNAH) was purchased from Sino Biological. This second protein was reconstituted in filtered deionised water before being buffer exchanged and concentrated up using a Vivaspin500 centrifuge column on a being spun at 13000 rpm.

3.2.1 Materials

3.2.1.1 Preparation of IgG–Fc from Sera

Dried native human IgG Fc fragment (Bethyl laboratories) was reconstituted using 200 µl filtered, deionised water before being dialysed into 50 mM sodium acetate pH 4.5. The solution was concentrated up with a Vivaspin 500 spin column to a concentration of 10 mg/ml, the concentration of the stock solution was monitored using a Nanodrop machine at 260/280 nm. The final concentration was obtained from the extinction coefficient that was calculated to be 1.365 by the Protparam program. This protein was used in crystallisation experiments which are detailed below. Due to the small volume of stock that was left after it was concentrated down, the stock was then further diluted to 8 mg/ml in order to maximise the number of wells which could be set down.

3.2.1.2 Preparation of Recombinant IgG–Fc

A batch of recombinant IgG–Fc fragment (Cat: 10702-HNAH) was purchased from Sino Biological. This protein was reconstituted in 1 ml of filtered deionised water to a concentration of 0.2 mg/ml before being buffer exchanged and concentrated up using a Vivaspin500 centrifuge column being spun at 13000 rpm for 12 minutes. The sample needed to be concentrated up a further two timesin order to get the solution to the minimum required concentration of 8 mg/ml. The concentration of the protein solution was monitored with a Nanodrop at 260/280 and the final concentration was obtained from the extinction coefficient that was also calculated to be 1.365 by the Protparam program.

3.2.1.3 Preparation of Heparin Ligands

The dp12 (degree of polymerisation) heparin fraction was produced, purified and dried by the Skidmore research group, before being made up to a final concentration of 10 mM in filtered deionised water. A second rough heparin fraction (mr 5.5 KDa ± 550 Da) was also produced and provided by the Skidmore research group. Finally, this fraction was taken and made up to a final concentration of 95 mM for use in crystallisation experiments.

3.2.2 Crystallisation of Native IgG-Fc From human sera

Crystallisation conditions based on previous work, both published and performed in-house were performed for the Fc fragment. The starting condition for these crystallisation trials were: 18% PEG 3350 and 0.1M MES pH 6. The suitable well conditions for the Fc region at 8mg/ml produced by these screens after optimisation were 0.1M MES pH 6.3, 18% PEG 3350 and 50 mM sodium acetate pH 4.5. Two crystal forms were produced by these conditions: short, wide rods which sometimes had pointed ends or a shower of small jagged crystals. These wells were cryoprotected by adding six lots of 2µl of glycerol aliquots to the well. These aliquots increased in concentration from 5%-30% glycerol. Once the 30% aliquot was added to the well, 12 µl of the mother liquor were exchanged with 12 µl of 30% glycerol to give a final glycerol concentration of 25%. The crystals were then frozen by submersion in liquid nitrogen. In total, eleven crystals were collected from a single well

(HE2A2). After four crystals had been collected, 2 µl of the 10 mM dp12 solution was added to the well and was left to soak for a further ten minutes.

3.2.3 Data Collection, Processing and Refinement of the Native Structure of IgG Fc

Data collection of the structure of IgG-Fc was carried out on beamline I04 at the Diamond Light Source (Oxford, England) at a wavelength of 1.000 Å and a temperature of 100 K. The diffraction images were processed using the MOSFLM program, the data was then merged and scaled using the AIMLESS program (Evans, 2006; Evans 2011) and averaged using the TRUNCATE program. The crystallographic model was produced and refined by the REFMAC program (Murshudov, 2011) and improved using molecular replacement within the COOT program (Emsley, 2010) using the coordinates of the structure of the IgG Fc region (PDB code 5W5M). This reference structure had the N-linked glycosylation at Asn297 as well as the water molecules removed from it in order to prevent bias from being introduced to the density.

3.2.3.1 Data processing and refinement of the Native Structure of IgG Fc

Fifteen datasets were collected from a total of 4 native and 7 dp12 soaked crystals of IgG Fc purified from human sera. Of these fifteen datasets, only two of them, one produced by a native crystal HE2A21 and one by a dp12 soaked crystal HE2A29 produced high resolution data. The diffraction images of HE2A21 were indexed into the primitive orthorhombic space group $P2_1\ 2_1\ 2_1$ by MOSFLM with unit cell parameters of:

$$a = 49.76\ b = 79.46\ c = 139.71 \qquad \alpha = 90.0°\ \beta = 90.0°\ \gamma = 90.0°$$

Once the images had been integrated by the MOSFLM program. The data was then scaled and merged at a resolution of 2.30 Å which gave Rmerge statistics of 0.072 and 0.368 in the outershell. The overall I/σ and CC1/2 values for the dataset were 13.8 (3.9 in the outershell) and 0.998 (0.923 in the outershell) and a completeness of 99.9% overall and 100.0% in the outershell. In Truncate an L-test was performed to determine whether or not the data was twinned. The twinning analysis determined that the data was untwinned.

The diffraction images from HE2A29 were also indexed into the primitive orthorhombic spacegroup $P2_1\ 2_1\ 2_1$ with a unit cell of:

$$a = 49.667\ b = 79.599\ c = 138.002 \qquad \alpha = 90.0°\ \beta = 90.0°\ \gamma = 90.0°$$

Once the images had been integrated by the MOSFLM program, they were scaled and merged using the Pointless-Aimless pipeline which also determined the unit cell parameters alongside the spacegroup. The data was scaled and merged at a resolution of 2.15 Å which gave Rmerge statistics of 0.064 and 0.397 in the outershell. The overall I/σ and CC1/2 values for the dataset were 12.6 (3.6 in the outershell) and 0.997 (0.795 in the outershell) and a completeness of 99.7% overall and 99.8% in the outershell. The structure factors were produced by Truncate and an L-test was performed to determine whether or not the data was twinned. The twinning analysis determined that the data was untwinned.

From these datasets the supposedly ligand bound data of HE2_A29 was chosen to be taken forward for structural analysis due to it being higher resolution.

3.2.3 Protein Crystallisation Trials of a Heparin Soaked Recombinant IgG1-Fc Fragment

The conditions set down in the dp ligand soak trials were based on the conditions set down during the native structure screen, which successfully produced the high resolution structure (HE2A2). The conditions set down were: 18% to 20% PEG 3350 and 0.1M MES pH 6.1 to pH 6.5. These wells contained the same 8mg/ml sample dried, native human IgG Fc that was used in the native structure screen.

A second set of conditions, all based off HE2A2 were set down, however the protein used in these trials was the recombinant human IgG1 Fc.

3.2.4 Ligand Soaking Trials of the Recombinant IgG1-Fc Fragment with a Rough Heparin Fraction

In the first round of soaking trials 2 µl of the 10 mM dp12 ligand was soaked into the well after cryoprotection with 30% glycerol and exchange. The crystals were then left for 10 minutes before crystals were retrieved and frozen.

The second round of soaking trials were conducted using the longer heparin fraction. During these trials two 0.5 µl additions of the heparin were added to the well, after which 0.1 µl of 1.55 M pH 4.5 sodium acetate was added to the well. This was done to emulate the conditions used within the circular dichroism experiment where the change in IgG-Fc in the presence of heparin was noted (Hadfield, 2018; A shrive 2017, personal communication, 15 November). This reduced the concentration of heparin in the well to 30.7 mM and the protein to 0.05 mM.

3.2.5 Data Collection, Processing and Refinement of the Heparin Soaked Structure of IgG Fc

Data collection of the structure of heparin soaked IgG-Fc was carried out on beamline I04 at the Diamond Light Source (Oxford, England) at a wavelength of 1.000 Å and a temperature of 100 K. The diffraction images were processed using the MOSFLM program, the data was then merged and scaled using the AIMLESS program (Evans, 2006; Evans 2011) and averaged using the TRUNCATE program. The structure produced by crystal HE2 A29 was used as a starting model and the structure was refined by the REFMAC program (Murshudov, 2011) with alternating cycles of model fitting with the COOT program (Emsley, 2010) . This reference structure had the N-linked glycosylation at Asn297 as well as the water molecules removed from it in order to prevent bias from being introduced to the density.

Four datasets were collected from a total of nine heparin soaked crystals of recombinant IgG Fc. Of these four datasets, only two of them (HE4 C18 and HE4 C19), produced any data. The diffraction images of HE4 C18 were indexed into the primitive orthorhombic space group $P2_1\ 2_1\ 2_1$ by MOSFLM with unit cell parameters of:

$$a = 49.76\ b = 79.46\ c = 139.71 \qquad \alpha = 90.0° \beta = 90.0° \gamma = 90.0°$$

Once the images had been integrated by the MOSFLM program, they were scaled and merged using the Pointless-Aimless pipeline which also determined the unit cell parameters alongside the spacegroup. The data was scaled and merged at a resolution of 3.55 Å which gave Rmerge statistics of 0.118 and 0.429 in the outershell. The overall I/σ and CC1/2 values for the dataset were 10.7 (4.7 in the outershell) and 0.980 (0.540 in the outershell) and a completeness of 95.3% overall and 89.9% in the outershell. The structure factors were produced by Truncate and an L-test was performed to determine whether or not the data was twinned. The twinning analysis determined that the data was untwinned.

The diffraction images from HE2A29 were also indexed into the primitive orthorhombic spacegroup $P2_1\ 2_1\ 2_1$ with a unit cell of:

$$a = 49.667\ b = 79.599\ c = 138.002 \qquad \alpha = 90.0° \beta = 90.0° \gamma = 90.0°$$

Once the images had been integrated by the MOSFLM program, they were scaled and merged using the Pointless-Aimless pipeline which also determined the unit cell parameters alongside the spacegroup. The data was scaled and merged at a resolution of 3.60 Å which gave Rmerge statistics of 0.118 and 0.430 in the outershell. The overall I/σ and CC1/2 values for the dataset were 11.0 (4.8 in the outershell) and 0.979 (0.520 in the outershell) and a completeness of 95.1% overall and 79.4% in the outershell. The structure factors were produced by Truncate and an L-test was performed to determine whether or not the data was twinned. The twinning analysis determined that the data was untwinned.

3.3 Results

3.3.1 Results of the crystallisation trials

Several crystallisation conditions for IgG-Fc purified from sera were put down. The well conditions were based on a previously established condition (18% PEG 3350 and 0.1M MES pH 6.5) which had produced a shower of small rod like crystals which had sharp pointed ends. This condition was then further optimised with two screens around this condition the conditions in this screen ranged from 16% - 20% PEG 3350 and 0.1M MES pH6.1 to pH6.7. These conditions produced several

crystal forms, the main form was a shower of small rod like crystals either with or without the same pointed tips seen in the initial condition. The second most common form was larger rods with flat tips.

3.3.2 Structure determination of Native IgG-Fc

Crystals from several wells were cryoprotected with 30% glycerol and used in diffraction experiments. Fifteen datasets produced by these X-ray diffraction experiments of IgG-Fc crystals were produced. Five of these datasets were produced by native IgG-Fc crystals, with the remaining ten datasets being produced by dp12 soaked crystals of IgG Fc purified from human sera. The data was collected at Diamond Light Source by the protein crystallography research group, before being processed using programs from the CCP4i software suite (Winn, 2011).

3.3.2.1 Data Processing and Selection

The datasets were integrated and refined by the MOSFLM program; after which they were scaled and merged into an average intensity using the Pointless-Aimless pipeline which also determined the unit cell parameters alongside the spacegroup. Of the original fifteen datasets only two of them were chosen to be further processed: one produced by a native crystal HE2A21 and one by a dp12 soaked crystal HE2A29, were further processed using the Truncate program. The details of the processed data statistics for the two datasets are included in table 1. The remaining thirteen datasets were not taken forward due to the data being of worse quality, in all cases the resolution was lower than the chosen datasets and the R_{merge} values were greater, meaning that the data had not merged as well. From these datasets the ligand soaked data of HE2 A29 was chosen to be taken forward for structural analysis due to it being higher resolution. In the case of the four native datasets that were not chosen, they were all at a lower resolution than the chosen dataset (2.35Å - 2.85 Å) and all of them had a much higher R_{merge} than the chosen dataset (0.087 – 0.202 overall). As R_{merge} is a measure of how well the data merges with a lower R_{merge} value indicating higher quality data.

In the case of the nine dp12 soaked datasets that were not used all of them were at a much lower resolution than the chosen dataset (2.20 Å - 2.70 Å) and all of them had worse overall R_{merge} values (0.69 – 0.145).

Table 3.1 – Initial Data Processing of the IgG1-Fc Purified from Sera Datasets.

	Native IgG-Fc purified from sera (HE2A21)	**Native heparin soaked IgG-Fc Purified from sera (HE2A29)**
Data Collection		
Synchrotron Station	I04	I04
Wavelength (Å)	0.9795	0.9795
Space Group	P 21 21 21	P 21 21 21
Cell Dimensions (Å)	a = 49.760 b = 79.460 c = 139.715	a = 49.667 b = 79.599 c = 138.002
Resolution Range (Å)	79.46 – 2.30 (2.16 – 2.30)	69.00 – 2.15 (2.21 – 2.15)
Observations	465147	540415
Unique Reflections	76108	98176
Completeness (%)	99.9 (100.0)	99.7 (99.8)
R_{merge}	0.072 (0.368)	0.064 (0.397)
CC1/2	0.998 (0.923)	0.997 (0.795)
Mean (I/σ(I))	13.8 (3.9)	12.6 (3.6)
Refinement		
Protein Atoms	7008	7008
Residues (Chain A)	208	207
Residues (Chain B)	208	208
Water Molecules	0	150
R_{work} (%)	0.2960	0.2493
R_{free} (%)	0.3558	0.3007

Figures in parentheses refer to the highest resolution bin.

[a] $R_{merge} = \Sigma_h\Sigma_j | I_{h,j} – I_h | / \Sigma_h\Sigma_j | I_{h,j}|$, where $I_{h,j}$ is the j^{th} observation of reflection *h* and I_h is the mean of the *j* measurements of reflection *h*.

[b] $R_{work} = \Sigma_h | |F_{oh}| - |F_{ch}| | / \Sigma_h |F_{oh}|$, where F_{oh} and F_{ch} are the observed and calculated structure factor amplitudes, respectively, for reflection *h*.

[c] R_{free} is equivalent to R_{work} for a randomly selected subset (5%) of reflections not used in the refinement.

[d] Determined according to MolProbity.

3.3.2.2 Density Map Generation and Analysis of Native IgG Fc

The initial density maps for the native structure of the Fc region were produced following rigid body refinement (Refmac5), using native IgG4 Fc (5W5M PDB accession code) as a starting model. Only the protein components of this structure were used for this in order to prevent any bias being introduced into the map by the glycosylation and water. The initial rigid body maps before any model fitting for the Asn297 glycosylation site residue can be seen in figure 3.1.

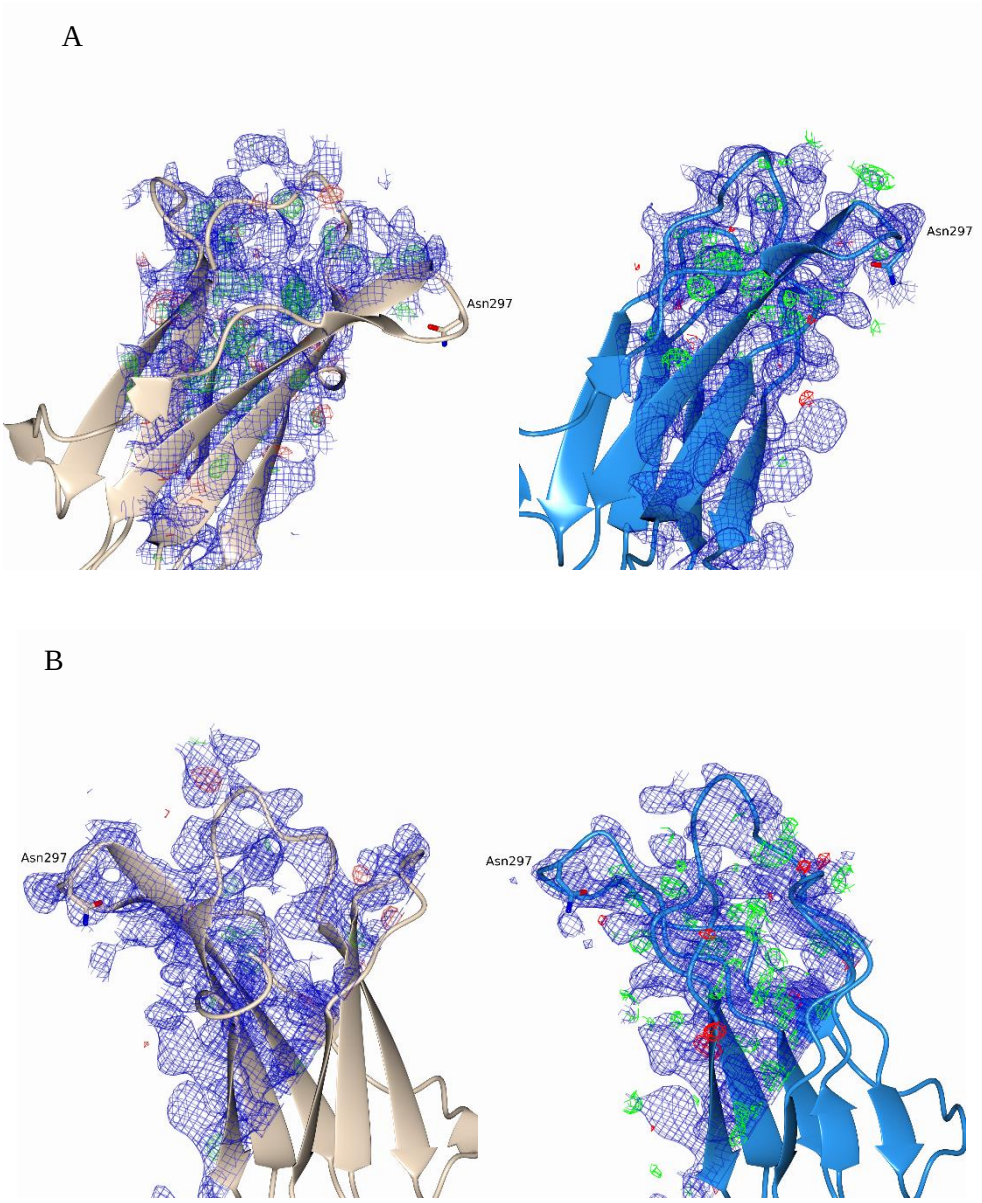


Figure 3.1 – A comparison of the A and B chains in human IgG1-Fc before and after model fitting. A: The electron density of the Asn297 residue of the A chain before and after any fitting had taken place, the observed electron density map is at 1.0σ (blue) and the difference map (green [positive density] and red [negative density]) shows areas of positive and negative density and is at 3.0σ. Asn297 was 6.26 Å away from its final position. B: The electron density of the Asn297 residue of the A chain before and after any fitting had taken place. In this chain Asn297 was not shifted over from its final position. The chain before fitting is shown in brown and the chain after fitting is shown in blue.

The electron density for the Ala237 – Thr250, Val259 – 306 and Lys326 – Lys337 loops of the CH2 domains, indicated that the CH2 domains of both of the A and B chains of the reported structure

were shifted away from their corresponding positions in the reference model (5W5M). The movement of the loops in the CH2 domains varied depending on where the residue was located within the chain with the greatest shifts being seen in the residues which were located closer to the hinge region some examples of these shifted residues are: Ala237 which was 6.66 Å from its corresponding residue, Glu269 which experienced a 6.87 Å and Phe296 which was shifted 6.98 Å. The shift was greatly reduced in the residues closer to the CH_3 domain, with the structures fully superimposing at residue 338 in both the A and B chains. The largest shifts in the structure were between Ala237 – Thr250, Val259 – 306 and Lys326 – Lys337 and can be seen in figure 3.2.

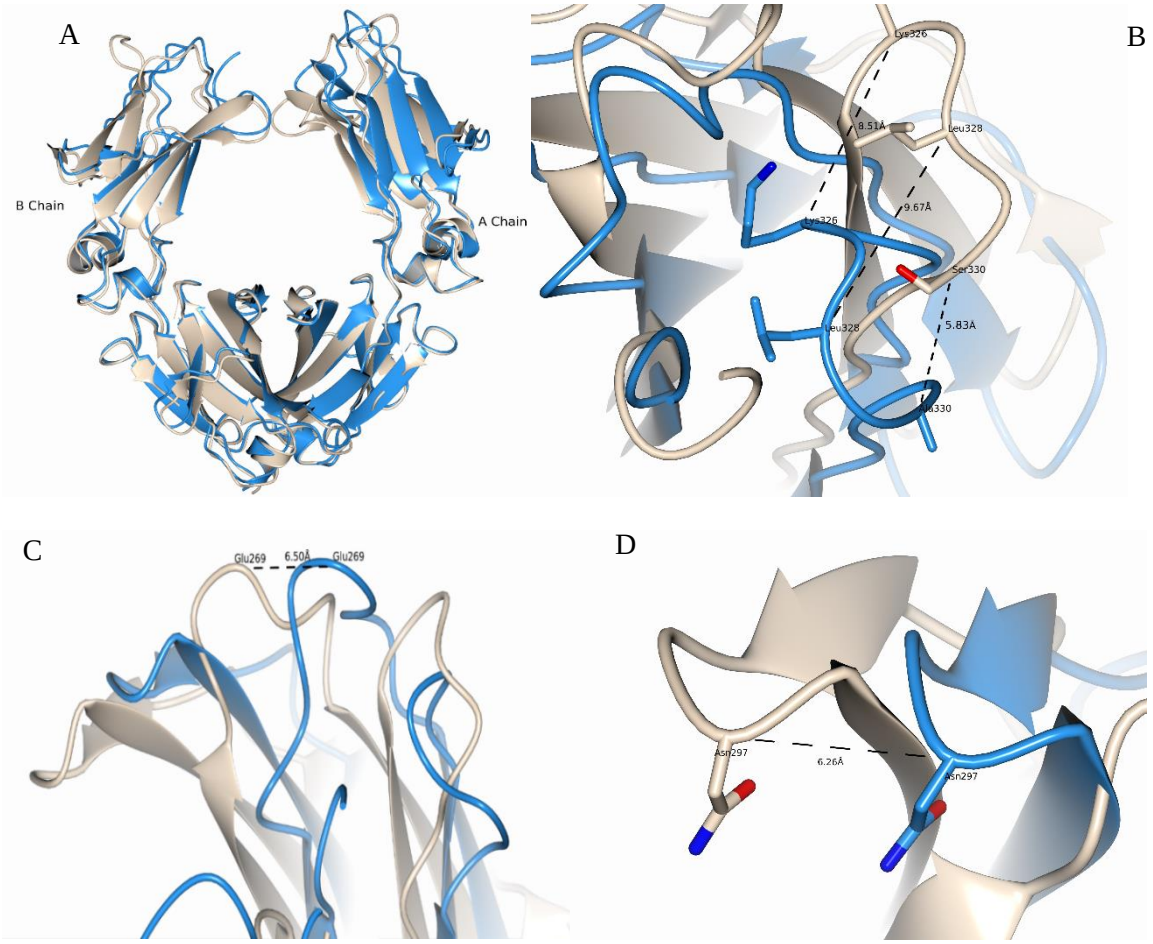


Figure 3.2 – The shift in the CH_2 domains of native IgG1-Fc and the reported structure of IgG1-Fc. A: The shift across the whole of the Fc region. The two structures superimpose after Lys338. B: the shift across residues Lys326 – Ser330 of the B chain. It is this loop which sees the greatest difference between the two structures. C: The shift at Glu269 of the B chain. D: The shift at Asn297 of the A chain. The reference structure before any fitting is shown in cream and the final structure after fitting is shown in blue.

In the initial structure of the N-termini, there was only clearly defined density visible for Gly237 in the A chain (Gly237A) and Gly236 in the B chain (Gly236B) onwards. There was some density visible for the residues which precede Gly237A and Gly236B which indicated that they were present, however the density was poorly defined and as such could not be modelled with confidence. This lack of defined density is most likely caused by the hinge region being cut with papain which allowed the N-termini to move around as they were no longer held in place by the hinge and CH_1 domains.This lack of definition within the N-terminus is found within several other native structures of IgG Fc, one example of this is the structure for a native IgG3 Fc region (5W38) which only had defined density for Lys235 onwards, another example is the structure of human IgG1 Fc (PDB code 5JII) which also only had well defined density for Gly237 onwards so this lack of density in the N-terminal appears to be a regular occurrence in structures of IgG-Fc.

The N-linked glycosylation at Asn297 was clearly visible in the density of the $2F_o$–fc (difference) map, which showed very positive density for both of the glycan chains. The glycans were modelled using the glycan add-on in the COOT program, using the complex hybrid mammal tree to build the two glycans into the density. The density for the entirety of the A chain glycan was very clearly defined and so it was very easy to model and fit into the density. Both a core fucose and a 6' galactose on the α1-6 arm of the glycan were visible in the A chain although no density for a bisecting N-Acetylglucosamine (GlcNAc) or terminal sialylation was observed for either of the glycan chains. The lack of density for these monomers implies that the the glycoform of IgG present in this structure was G1F, (where G refers to galactose and F refers to fucose) which makes up 38% of the glycoforms found within human serum. However it is also possible that the IgG-Fc sample used is of a different glycoform and that these monomers are simply not visible in the density. The structure of the glycan as well as the structure of the fitted glycan are shown in figure 3.3 and figure 3.4 respectively.

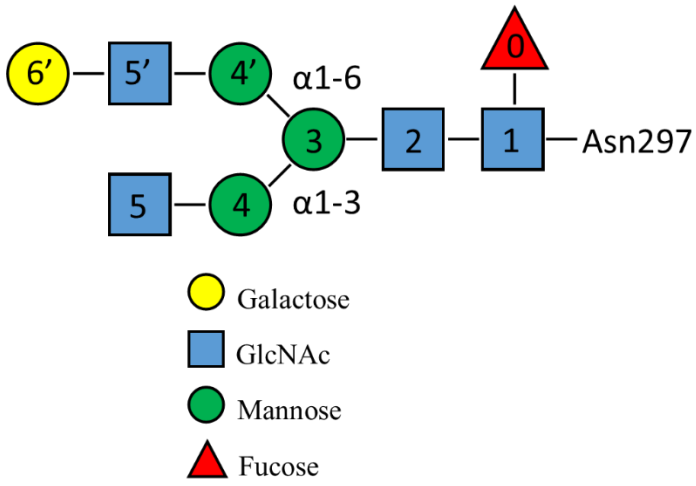


Figure 3.3 – The ASN297 N–Linked glycan of the sample of IgG-Fc purified from Sera. The association occurs through the GlcNAc at position 1 in the chain; with the GlcNAc at position 5 forming hydrogen bonds with the correlating GlcNAc of the N-linked glycan of the B chain. This hydrogen bond is the only major interaction between the two chains of the Fc fragment.

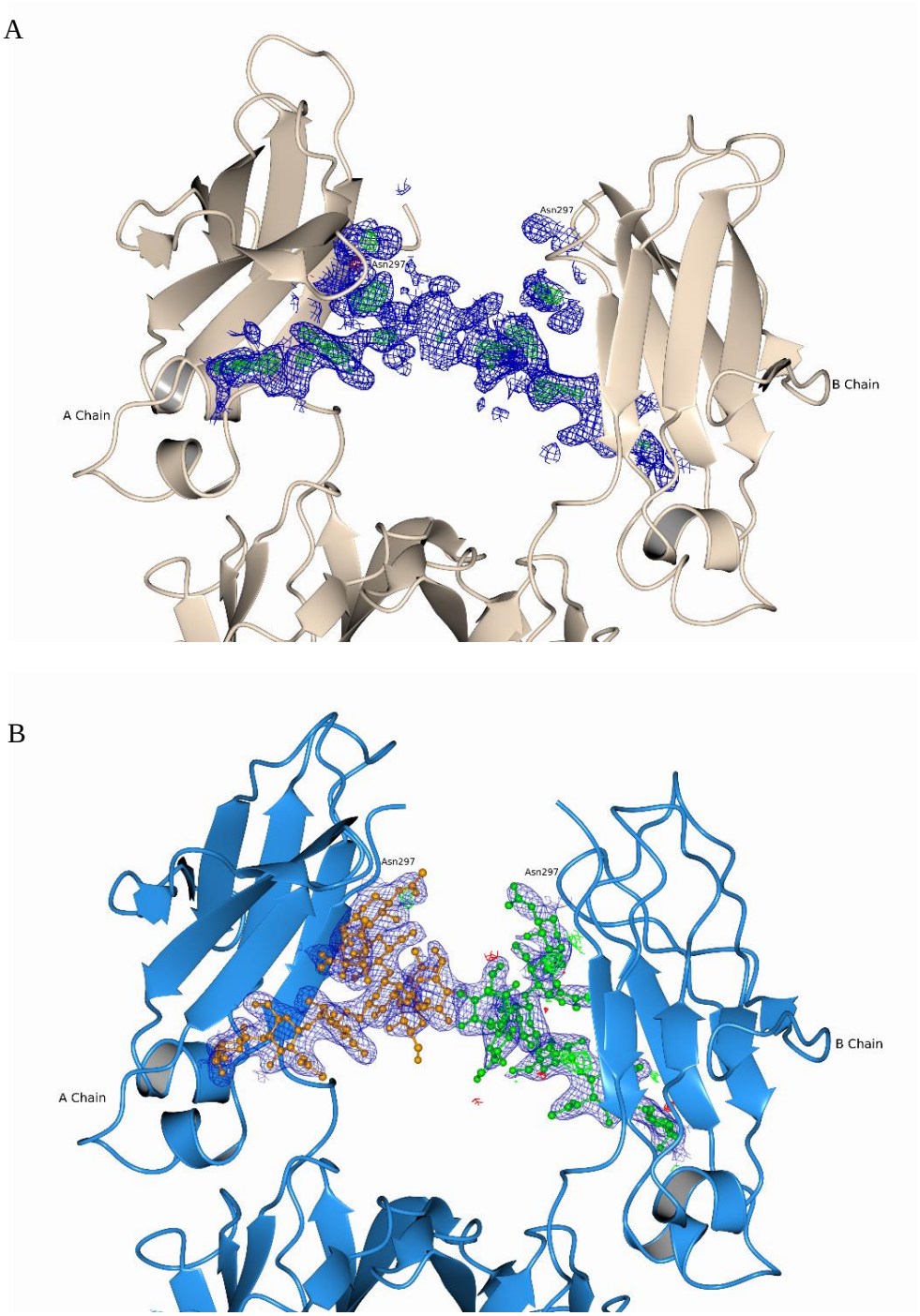


Figure 3.4 – The structure of the fitted N-linked glycan of IgG1-Fc bound to Asn297. A: The rigid body electron density map for the N-linked glycan before any fitting has been performed, the observed electron density map is at 1.0σ (blue) and the difference map is at 3.0σ. The structure is shown to elucidate on how the glycan sits within the protein chains. B: The electron density map of the N-linked glycan after 5 rounds of restrained refinement. The glycan of the A chain is shown in orange and the glycan of the B chain is shown in green. The reference structure before any fitting is shown in cream and the final structure after fitting is shown in blue.

The density for the fucose of the B chain bound glycan was much weaker than the density for the rest of the glycan, and as a result there was not enough density visible to confidently model it. There was not enough density for the 6' galactose visible in the B chain to fit it however it is still present in the glycan chain because the two glycans must be identical to each other. Once the glycans had been modelled and further refined the glycan statistics were run through the glycan validation program of the CCP4i suite, Privateer (Agirre *et al.* 2015), in order to determine whether the glycan model that had been produced had correct conformations and was not puckered. The Privateer output showed that the majority of the glycan residues were in favourable conformations. However the N-linked GlcNAc1 in the glycan of the A chain was found to be very slightly enveloped, which is a thermodynamically unfavourable conformation due to the ring being puckered. However the conformation of the ring could not be altered in COOT without causing the rest of the residues to enter unfavourable conformations and after further rounds of refinement and fitting it was determined to be acceptable. The final fitted structure for the native Fc region of IgG1 that was produced from HE2 A29 is shown in figure 3.5.

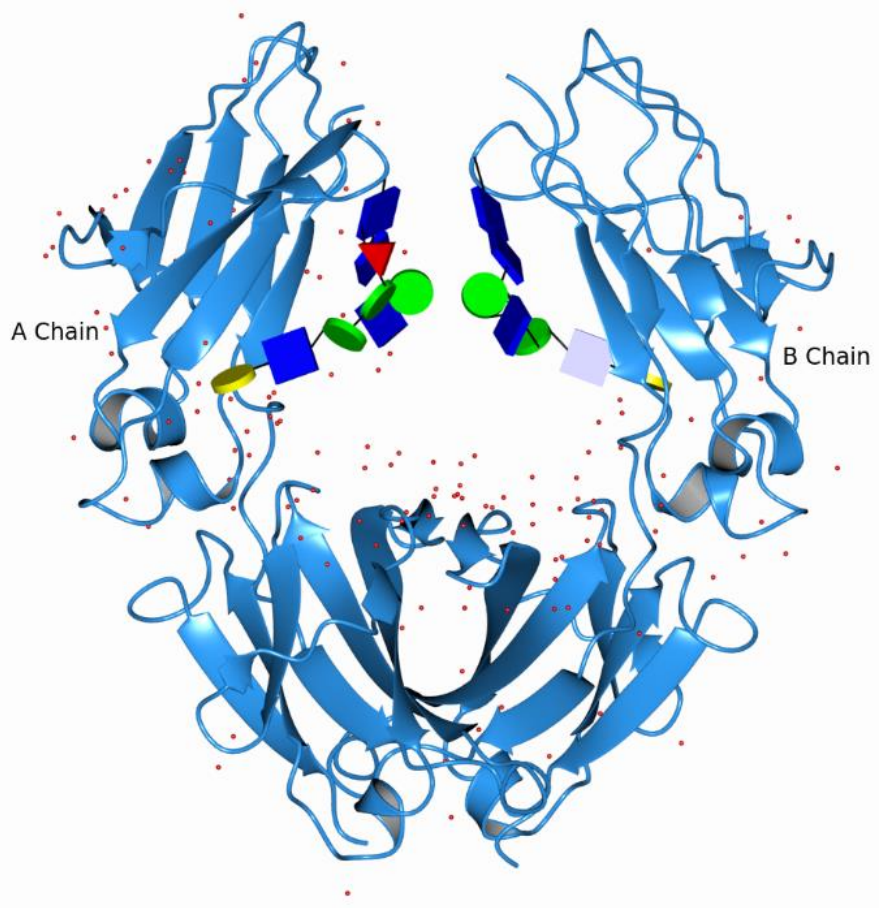


Figure 3.5 – The proposed structure of Native IgG1-Fc produced in this report. The N-linked glycan is shown in Glycoblocks form, the waters are shown as red spheres.

3.3.3 Structural Work for IgG-Fc Soaked with a Rough Heparin Fraction

3.3.3.1 Ligand Soak Trials

In order to produce a structure for heparin bound IgG-Fc a different ligand soaking strategy was produced in order to increase the molar fold difference between the heparin and the IgG-Fc from 26.0 times in the initial ligand soak to 621.6 times, as this would greatly increase the chances of the heparin binding to the IgG-Fc as any available binding sites would be saturated. The differences in this new strategy were as follows: Firstly the ligand used was a rough fraction of heparin of higher mass (5.5 KDa ± 10%) than the purified dp12 fraction used in the initial soaking trial. Secondly, the IgG-Fc was soaked with a much higher concentration of this ligand (95 mM as opposed to the original 10 mM). The final difference was the inclusion of a low level (50 mM) of the same sodium acetate buffer that was used to reconstitute the protein. The addition of the acetate was to assist the

heparin in binding to the IgG-Fc, this was performed because of the results of the preliminary circular dichroism experiment which showed a change in the secondary structure of the IgG when in the presence of the heparin and a low level of the acetate (Hadfield 2018 A shrive 2017, personal communication, 15 November). A recombinant IgG1-Fc, which was purified from human sera was used in the ligand soak trials. This is because the amino acid sequence of this recombinant protein was known (Fig. 4.3.1).

Ligand soaking trials were performed in order to determine the effect of ligand concentration on the stability of the crystals; as well as to determine an appropriate amount of time for the crystals to be soaked in the ligand without destabilising them. The crystals chosen for the soaking trials were picked due to their small size and irregular shape which made them inadequate for use in data collection. The crystals were monitored closely in the minutes following the addition. As well as being monitored twice daily in the week after the addition of the heparin and the sodium acetate with observations and pictures being taken at key times throughout the week (See Appendix 2). The key observation being that the crystals showed no major changes in their size or shape during the trial which meant that either the heparin was taken up readily by the crystals and did not disturb them; or that the heparin did not interact with them at all.

However, when this new soaking strategy was tried in the wells of interest (HE4A1, HE4A2, HE4C1 and HE5A1) the crystals started to rapidly degrade upon the addition of the 77% glycerol that was required for cryoprotection, or after the addition of 50 mM sodium acetate (which was the same buffer that the protein was initially prepared in). This degradation was most likely caused by the crystals going into shock either from the acetate having a much larger than anticipated effect on the conditions or from the heparin being taken up by the crystals. Out of these five wells crystals could only be collected from HE4C1 and HE5A1, eight crystals were collected from HE4C1. However, only one of the crystals collected from HE4C1 was exposed to the acetate. This was due to the rapid degradation of the crystals following the addition. Only one crystal was collected from HE5A1 because the well rapidly degraded after cryoprotection. The conditions for these wells are shown in table 4.

Table 3.2 – Table of well conditions for wells HE3A1, HE4A1, HE4A2, HE4C1, HE4C2 and HE5A1.

HE3A1	0.1 M HEPES pH 7.5 16% PEG 6000
HE4A1	0.1 M MES pH 6.1 16% PEG 3350
HE4A2	0.1M MES pH 6.1 18% PEG 3350
HE4C1	0.1 M MES pH 6.5 16% PEG 3350
HE4C2	0.1 M MES pH 6.5 18% PEG 3350
HE5A1	0.1 M MES pH 6.3 18% PEG 3350

3.3.3.2 Data Processing and Selection

Four datasets were collected from eight heparin soaked crystals, only one of which had been exposed to acetate. Of these four datasets only two of them were used for data analysis, because the other two datasets were not suitable for use. This was because the images produced during X-ray diffraction of the crystals did not contain enough information and as a result, they could not be processed.

The datasets were integrated and refined by the MOSFLM program; after which they were scaled and merged into an average intensity using the Pointless-Aimless pipeline which also determined the unit cell parameters alongside the spacegroup. Of the original four datasets, only two of them were able to be further processed using the Truncate program. The details of the processed data statistics for the two datasets are included in table 4.3.1. The remaining two datasets were not taken forward due to the MOSFLM program not being able to process them. Both of these datasets were taken forward in the hope of them giving some detail of the heprin binding properties of IgG-Fc.

Table 3.3 – Initial Data Processing of the recombinant IgG1-Fc Datasets

	Ligand soaked recombinant IgG-Fc (HE4_C18)	**Ligand soaked recombinant IgG-Fc (HE4_C19)**
Data Collection		
Synchrotron Station	I04	I04
Wavelength (Å)	0.9795	0.9795
Space Group	P 21 21 21	P 21 21 21
Cell Dimensions (Å)	a = 49.889 b = 79.641 c = 138.078	a = 49.249 b = 79.945 c = 138.183
Resolution Range (Å)	79.88 – 3.60 (3.60 – 3.60)	79.88 – 3.89 (3.55 – 3.55)
Observations	286360	40360
Unique Reflections	41514	6627
Completeness (%)	95.1 (79.4)	95.3 (89.9)
R_{merge}	0.113 (0.407)	0.118 (0.429)
CC1/2	0.979 (0.520)	0.980 (0.540)
Mean (I/σ(I))	11.0 (4.8)	10.7 (4.7)
Refinement		
Protein Atoms	7008	7008
Residues (Chain A)	206	204
Residues (Chain B)	207	208
Water Molecules	0	0
R_{work} (%)	0.3249	0.2493
R_{free} (%)	0.2966	0.3007

Figures in parentheses refer to the highest resolution bin.

[a] $R_{merge} = \Sigma_h\Sigma_j | I_{h,j} - I_h | / \Sigma_h\Sigma_j | I_{h,j}|$, where $I_{h,j}$ is the j^{th} observation of reflection *h* and I_h is the mean of the *j* measurements of reflection *h*.

[b] $R_{work} = \Sigma_h | |F_{oh}| - |F_{ch}| | / \Sigma_h |F_{oh}|$, where F_{oh} and F_{ch} are the observed and calculated structure factor amplitudes, respectively, for reflection *h*.

[c] R_{free} is equivalent to R_{work} for a randomly selected subset (5%) of reflections not used in the refinement.

[d] Determined according to MolProbity.

3.4 Discussion

3.4.1 Quality of the structural data

The final density map produced after several rounds of refinement had identified all the key structural components of IgG1-Fc; including the detailed structure of the N-linked glycan to the point where the glycoform of IgG-Fc could be determined.

3.4.2 Significance of the visible structural components of IgG-Fc

3.4.2.1 Analysis of the Structural Discrepancies Seen in the Reported Structure

The CH2 domain already has some degree of movement between IgG glycoforms afforded to it by the differences in the length and composition of its N-linked glycan as well as the innate flexibility of the hinge region. And many of these conformations have been well documented. However none of the conformational differences within these other structures are as large as the change seen in the reported structure.

The CH2 domain is also known to change conformation once the C1q complement protein has bound to the Fc region, with the CH2 domains changing from an open conformation (Where the two CH2 domains are at a maximal distance from each other) to a closed conformation (where the two CH2 domains are at a minimal distance to each other) as has been documented by (Mimura et al., 2001; Mimura et al., 2011; Gaboriaud et al., 2003). This could explain the change in conformation seen in the reported structure as being a product of the heparin binding to the IgG-Fc as it is feasible that this has caused the change. The reason that the heparin is not visible in this structure could be that it was not at a high enough concentration to be seen in all of the IgG-Fc molecules in the lattice, but was at a high enough concentration to have a visible effect. It could also be possible that the heparin was not held rigidly within the structure and as a result did not produce any density. The most likely explanation for the large shift in the CH2 domain is that it was caused by an artefact of the actual crystallisation process itself, potentially caused by the fact that the protein stock contained a mixture of the four IgG Isoptypes which are of different lengths.

3.4.2.2 Difficulty with producing a heparin bound structure

A heparin bound structure has not been produced by the crystals which had been soaked in either the dp12 ligand or in the rough heparin fraction. Despite the fact that the required molar excess of heparin is unknown, this is unlikely to be due to a low molar difference between the heparin and the IgG for the rough fraction because there was 621.6 times the amount of heparin in the well than IgG-Fc. This could however explain why the dp12 ligand did not bind to the IgG-Fc as there was only 26.0 times the amount of dp12 in HE2A2 than IgG-Fc and as such any potential binding sites on IgG-Fc might not be sufficiently saturated for the dp12 to be visible within the structure. It is also possible that the heparin takes longer to bind to the IgG-Fc than the 38 minutes it was given in HE2A2 or the 36 minutes it was given to soak in HE4C1. The crystals dissolving could also have been caused by the heparin entering the binding site within the native crystals and disrupting their structure from within. This could also explain why some of the crystals appeared to be stable during soaking, because in these instances the heparin could have been unable to enter the binding site and as such did not disrupt the structure.

3.4.3 The varying success of the Crystallisation and Heparin Soaking Trials

The varying degrees of success when it came to growing and cryo protecting the crystals of the IgG-Fc were well noted from the beginning of the study and were persistent throughout the study, with different crystals from the same well or even crystals with the same mother liquor conditions produced crystals which reacted differently to the chosen cryoprotectant as well as the heparin ligand. It was noted that there were two main crystal forms favoured by the IgG-Fc crystals; one of which was short, wide rods which sometimes had pointed ends, as well as a second crystal form which was a shower of small jagged crystals. These two crystal forms were not mutually exclusive to a particular condition. In the case of HE2A2, both of the crystal forms were found in the well. These two crystal forms also appeared to have different stability when the acetate was introduced: HE3A1 which contained a shower of jagged rods, was completely stable for well over three weeks after the addition of 30.7 mM Heparin and 50mM sodium acetate to the well. The details of this soak can be found in Appendix 2. However, once the acetate was added to HE4C2 which contained

several of the larger single crystals the crystals rapidly degraded with most of the smaller crystals dissolving completely. The results of this soak can be found in Appendix 3.

These differences could also be due to the way in which the wells were cryoprotected: The concentration of glycerol in HE2A2 was slowly built up with six 2 µl volumes of glycerol, whose concentrations ranged from 5% - 30% before 12 µl of the mother liquor was exchanged for the 30% solution (all solutions were made up to concentration using well conditions) to give a final glycerol concentration of 25%. Whereas HE4C1 was cryoprotected using four 0.5µl additions of 50% glycerol which gave a final glycerol concentration of 22% and caused the crystals to rapidly degrade. The 50% additions may have been too great of a shock to the crystals because of the glycerol being wildly different to the mother liquor and as such the crystals degraded.

Another explanation for the difference in crystal stability is that the well conditions of HE3A1 were different from both HE2A2 and HE4C1 as can be seen in table 5. The higher pH (7.5) of the HEPES buffer could have produced more stable crystals.

Table 3.4 – Table of well conditions for well HE3A1 and HE4C1.

HE3A1	0.1 M HEPES pH 7.5 16% PEG 6000
HE4C1	0.1 M MES pH 6.5 16% PEG 3350

During data collection on these crystals, the presence of ice-rings in the diffraction images made it apparent that the crystals had not been completely cryoprotected. This shows that ice had built up inside of the crystal causing damage to the lattice. This is also likely to have contributed to the low resolution and poor processing of these datasets as well as the poor electron density observed during refinement.

3.4.4 Quality of the Structural Data

After several rounds of refinement, all of the electron density for the protein backbone was present in the density map. However, the density for some of the side chains was not fully defined and as

such; the confidence of the fitting in these areas is not high. Much of the fine details of the structure of the N-linked glycan was also lost in this structure for example the fucose0 residue, however there was enough density to confidently model the glycan into the structure.

3.4.5 Analysis of the Key Structural Details

The structural details which can be observed in the protein backbone of this heparin soaked structure of recombinant IgG1-Fc are identical to the native structure of IgG1-Fc from human sera. This was confirmed when the model of native IgG1-Fc fit perfectly into the density map of the heparin soaked IgG1-Fc. Further refinement of this density map did not result in any changes to the density of the backbone, but did result in slight improvements to the density for the N-linked glycan. In both structures the conserved C1q binding cleft which is formed by the two CH3 domains was clearly visible. In addition to this, there was clear density for all of the identifying residues that would typically be expected to be seen within the structure of IgG1-Fc, some examples of these are shown in figure 3.6 below.

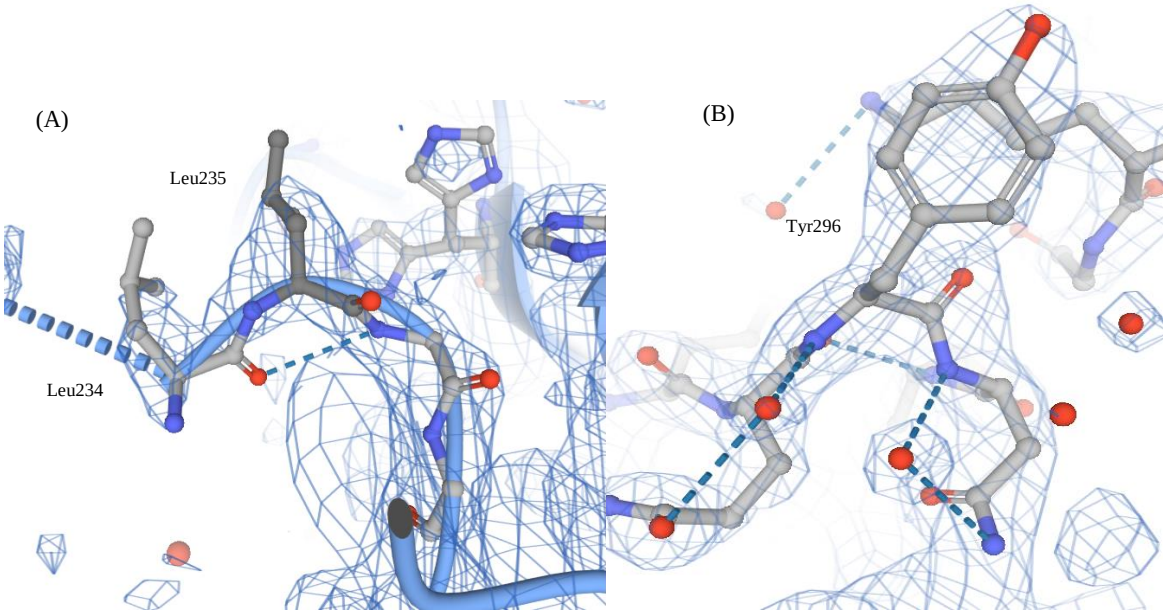


Figure 3.6 – Examples of the electron density fitting for some of the identifying residues of the native IgG-Fc. (A) The electron density fitting for Leu234 and Leu235. (B) The electron density fitting for Tyr296.

No electron density was visible for the heparin ligand in this structure. There was also no evidence present for the heparin binding activity that was reported in the initial CD experiments. This can be explained by the lack of structural differences between the structure of native IgG1-Fc and the structure of heparin soaked IgG1-Fc; because it can be expected that the reported secondary structure change would result in an obvious change to the quaternary structure which would be visible when the density of the heparin soaked IgG1-Fc was compared to the model of native IgG1-Fc. As this was not the case it can be concluded that heparin has not bound in the heparin soaked model. This could be due to there not being a great enough excess of heparin used in the soak, and as such the heparin was not at a high enough concentration to bind. Or more likely, it is because the secondary structure of IgG1-Fc cannot change once it has been locked in place in the crystal matrix and as such heparin cannot bind successfully.

3.4.6 Protein Electrostatics and Potential Heparin Binding Sites

Electrostatic analysis of the reported protein structure of the native IgG-Fc from sera revealed several highly positively and negatively charged regions (these regions can be seen in full in figure 3.7 B and C below). This analysis also revealed two grooves with an overall positive charge; the first of these grooves is made up of the CH3 domains of both the A and B chains and is made up from the Ser354 – Ser400 residues of both chains. The second groove was found to be within the empty space between the A and B chains that is occupied by the N-linked glycan. The first groove is known to be the binding site of C1q during the activation of the complement cascade. Due to the highly negative charge produced by the various S-domains within the structure of heparin; it has been observed that heparin tends to bind to highly positive grooves, such as the two examples mentioned above. This potentially implies that these two positive grooves could act as potential binding sites for heparin. However, due to the size of these grooves in comparison to dp12 heparin, it is likely that any interaction could be between IgG-Fc and a shorter heparin polymer. A potential way to test this theory would be to attempt to produce crystals of IgG-Fc bound to singular sulphate ions.

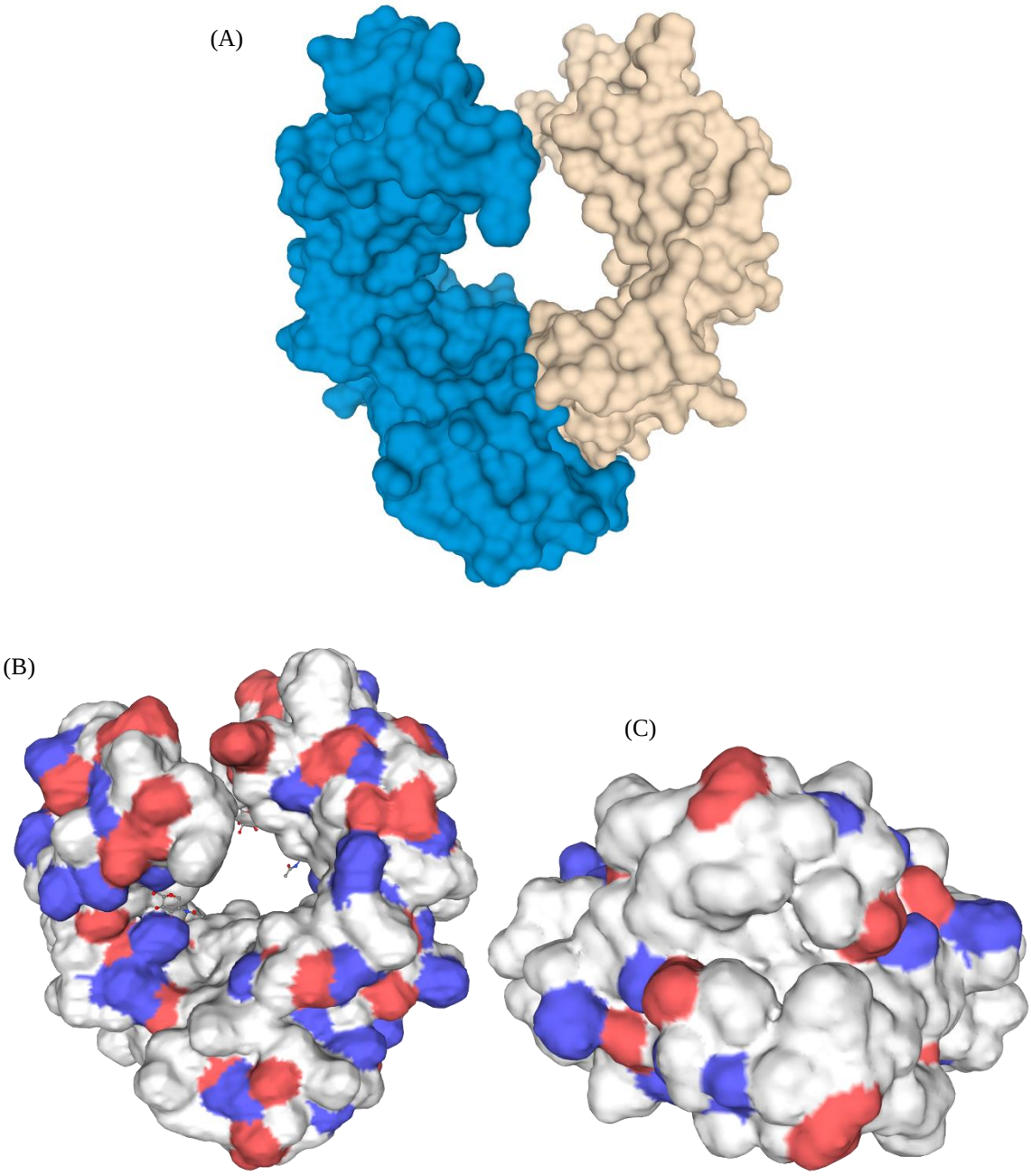


Figure 3.7 – The structure packing and electrostatics of the produced Native IgG-Fc structure. (A) The packing of the reported protein structure, the packing of the A chain is shown in light blue and the packing of the B chain is shown in light brown. (B) The electrostatics of the full structure of the reported native IgG-Fc from sera; areas of positive charge are shown in red and areas of negative charge are shown in blue. (C) The electrostatics of the groove formed by the CH3 domains of the A and B chains; areas of positive charge are shown in red and areas of negative charge are shown in blue.

3.5 Conclusion

The main aim of this work was to determine the binding mechanism for the interaction between IgG-Fc and the heparin dp12 polymer and how this interaction modifies the structures of both of these molecules. The results presented here show the findings of the initial work into the characterisation of the dp12 – IgG-Fc interaction. So far only the native structure of IgG-Fc has been produced, this was due to the dp12 ligand not successfully binding to IgG-Fc because the well conditions may have been incorrect for this interaction. As such, the ligand could not alter the crystal matrix and enter the binding site.

Immediate work to produce a ligand bound would include further optimisation of the ligand soaks and cryoprotection protocols, in order to prevent the rapid crystal degeneration seen in these trials. This work would be followed by co-crystallisation screens of the recombinant IgG-Fc fragment with the rough heparin fragment if further optimisation of the soak protocol failed to provide usable results (Clark *et al*, 1993; Nikolov *et al*, 1993). Once either the ligand soaking protocol or the co-crystallisation protocol had produced results with the rough fraction the protocol would then be repeated with the dp12 ligand.

Additionally, after a heparin bound structure had been produced, further work would look to determine whether a change in the chain length of the heparin polymer effects the binding affinity of IgG-Fc for heparin. The final goal of this work would be to look into what effect the heparin-IgG-Fc interaction has on the binding affinity of IgG-Fc to some of its immune receptors such as C1q, FcγR and FcγRn (Gaboriaud *et al.* 2003).

Chapter 4 – General Discussion, Conclusions and Future Work

The work presented in this thesis aims to increase the current knowledge of the binding mechanisms of IgG-Fc by presenting a heretofore unobserved association with analogues of heparin through structural studies of native IgG-Fc purified from human sera and a recombinant IgG1-Fc soaked with heparin. Whilst the crystals reported in this work are an advancement to understanding this interaction between heparin and IgG1-Fc as preliminary work as they form a reliable basis for further structural studies to be undertaken.

4.1 Crystal Forms and Condition Choice

Two crystal forms were observed during the crystallisation trials described in this work. The first form was large flat singular rods which either had pointed or flat ends and the second form was a large shower of much smaller rods which only had pointed ends. During the cryoprotection and soaking trials it was determined that only the larger rods with flattened ends were completely stable when exposed to both the cryoprotectant and the ligand. The larger rods with pointed ends would either completely or partially degrade when exposed and the smaller rods would consistently degrade completely. The large pointed rods degraded inconsistently between wells with repeat and similar conditions, in some wells they would degrade completely and in others they would only partially degrade. The reason for the large pointed rods degrading when exposed to the ligand may well be because of the ligand entering the crystal and interacting with the IgG-Fc molecules and disrupting the crystal lattice. This could also explain why the flattened rods did not degrade when exposed to ligand as none of the structures produced by ligand soaked crystals showed heparin binding which indicates that heparin has not entered the crystal structure. Examples of these crystal forms as well as ligand soaking and cryoprotection trials can be found in appendix 2.

4.2 Ligand Soaks and Co-Crystallisation

The reason for initially soaking the crystals, rather than co-crystallising was to preserve the limited stocks of both protein and ligand; whilst establishing conditions in which the crystals could survive when exposed to the ligand. During these ligand soaking trials many of the crystals degraded and

dissolved rapidly upon addition of the ligand. However this seemed to be largely dependent on the crystal forms that were exposed to the ligand; with the smaller, sharper crystal showers dissolving much more rapidly than the larger rods. Once a viable condition had been established by the soaking trials, a co-crystallisation approach was turned to. This method was utilised primarily because none of the ligand soaked crystals (which did not dissolve), did not produce ligand-bound structures. This was thought to because the secondary structure of IgG-Fc was require to undergo a conformational change in order to bind heparin, which could not occur once the protein was locked in place in a crystal lattice. This approach also allows for longer ligands (such as dp12 heparin) to enter into the protein, which they are unable to do in some crystal forms due to the way in which the proteins are arranged in the crystal lattice.

4.3 N–Linked Glycan

The IgG-Fc purified from human sera was reported to be of the G1F glycoform, where the G refers to the number of galactose residues and 1F refers to the number and position of fucose residues (one fucose bound to GlcNAc at position 1). This glycoform is present in roughly 38% of all IgG-Fc in human sera and whilst it is possible that the reported structure could be of a different glycoform there was not enough density to suggest this. The electron density for both chains at the Asn297 residue was very clear. However, there was no density to indicate that a bisecting GlcNAc or terminal sialic-acid residue were present. The electron density was much clearer for the α1-6 arm of the A-chain N-linked glycan and as such accurately reporting the orientation of the fucose and galactose residues in the B-chain was more difficult. However, these residues were still present as there was partial density for them and the two glycans must identical. Privateer (Agirre *et al.* 2015), reported that the ring of the GlcNAc1 in the A chain glycan was slightly puckered and as a result was not in a thermodynamically favourable orientation. However, after several rounds of refinement, this was determined to be acceptable as altering the orientation of this residue would cause the rest of the glycan to enter unfavourable orientations.

4.4 Practicality of Mass Spectrometry

Mass spectrometry techniques were utilised alongside x-ray crystallographic analysis of the structure to aid in the correct identification of the isoform of IgG-Fc that was present in the protein crystals. Trypsin and GluC were both required in order to give the greatest sequence coverage as they have differential splicing patterns and so produced different sequences. This meant that the GluC samples were able to produce sequences that the trypsin samples could not and vice versa.

Because ESI and MALDI produce fragments of different lengths (with MALDI producing longer sequences and ESI producing shorter sequences), it was possible to produce a greater range of fragments and therefore a greater sequence coverage. These techniques proved to be effective companions to x-ray diffraction as they were able to determine which isoforms of IgG1-Fc, IgG2-Fc, IgG3-Fc and IgG4-Fc were present in the sample of human sera. This was possible, because whilst the sequence of IgG-Fc is 90% conserved between the isoforms, there are a series of identifying residues which are not conserved which were observed by this technique.

4.5 Implications and Future Work

Immunoglobulins are a key component of the immune system of vertebrates due to their ability to identify pathogens and activate other components of the immune system. IgG-Fc is critical to the function of IgG as an activator of the immune system because of the interactions between it and the cell-surface receptors of immune cells. The conditions described in this thesis offer the potential to be used as a foundation for further crystallisation trials into the novel interaction of human IgG-Fc and heparin. Once a heparin bound structure is produced, the work would need to focus on fully characterising the interaction and any conformation changes within IgG-Fc. Additionally, following the production of a dp12 heparin bound structure, the next step would be to determine whether a change in the chain length of the heparin analogue effects the binding affinity of IgG-Fc for heparin. Once this has been established, the final goal would be to look into what effect, if any the heparin-IgG-Fc interaction has on the binding affinity of IgG-Fc to cell-surface receptors such as C1q, FcγR and FcγRn (Gaboriaud et al. 2003), as well as whether the interaction affects the pathogen binding properties of IgG-Fc.

References

Aalberse R.C., van der Gaag R., van Leeuwen J. “Serologic aspects of IgG4 anti- bodies. I: Prolonged immunization results in an IgG4-restricted response” **1983.** *J Immunol.* Vol 130: 722–6.

Agirre J., Iglesias-Fernandez J., Rovira C., Davies G.J., Wilson K.S., Cowtan K.D., “Privateer: software for the conformational validation of carbohydrate structures”, **2015.** *Nat. Struct. Mol. Biol.* Vol 11: 833-834.

Ahmed A.A., Keremane S.R., Vielmetter J., Bjorkman P.J. Structural characterization of GASDALIE Fc bound to the activating Fc receptor FcγRIIIa. **2016.** *J. Struct. Biol.* Vol 194: 79-89.

Akashi, S., Naito, Y. and Takio, K. (1999) ‘Observation of hydrogen-deuterium exchange of ubiquitin by direct analysis of electrospray capillary-skimmer dissociation with fourier transform ion cyclotron resonance mass spectrometry’, *Analytical Chemistry*, 71(21), pp. 4974–4980. doi: 10.1021/ac990444h.

Als-Nielsen J. and McMorrow D. (2001). *Elements of Modern X-ray Physics*, Chapters 1 and 2, Wiley, New York.

Arnold J.N., Wormald M.R., Sim R.B., Rudd P.M., Dwek RA. The Impact of Glycosylation on the Biological Function and Structure of Human Immunoglobulins. **2007.** *Annu. Rev. Immunol.* Vol 25: 21-50.

Barb A.W. Intramolecular N-glycan/polypeptide interactions observed at multiple N-glycan remodelling steps through [(13)C,(15)N]-N-acetylglucosamine labelling of immunoglobulin G1. **2015.** *Biochemistry.* Vol 54: 313.

Barrett DJ, Ayoub EM. IgG2 subclass restriction of antibody to pneumococcal polysaccharides. **1986.** *Clin Exp Immunol* 63(1):127–34. 13.

Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. and Leslie, A. G. W. (2011) ‘iMOSFLM: A new graphical interface for diffraction-image processing with MOSFLM’, Acta

Crystallographica Section D: Biological Crystallography. International Union of Crystallography, 67(4), pp. 271–281.

Bertini S., Fareed J., Madaschi L., Risi G., Torri G., Naggi A. Characterization of PF4-Heparin Complexes by Photon Correlation Spectroscopy and Zeta Potential. **2017**. *Clin. Appl. Thromb.-Hemos.* Vol: 23, Issue: 7, 725-734.

Bienkowski, M.J. and Conrad, H.E. Structural characterization of the oligosaccharides formed by depolymerisation of heparin with nitrous acid. **1985.** *J. Biol. Chem*., Vol 260: 356–365.

Bondt A, Selman M.H., Deelder A.M., Hazes J.M., Willemsen S.P., Wuhrer M., Dolhain R.J. Association between galactosylation of immunoglobulin G and improvement of rheumatoid arthritis during pregnancy is independent of sialylation. **2013.** *J Pro- teome Res.* Vol 12 Issue 10: 4522–4531.

Bragg, W.L., (1975). *The Development of X-ray Analysis* (ed. D.C. Phillips and H. Lopson). Bell, London

Brekke O.H., Michaelsen T.E., Sandlie I. The Structural Requirements for Complement Activation by IgG: Does it Hinge on the Hinge? **1995.** *Immunol. today.* Vol 16 issue 2: 85-90.

Brown, E. N. and Ramaswamy, S. (2007) ‘research papers Quality of protein crystal structures research papers’, *Acta Crystallographica Section D: Biological Crystallography*, D63, pp. 941–950. doi: 10.1107/S0907444907033847.

Bruhns P., Iannascoli B., England P., Mancardi D.A., Fernandez N., Jorieux S., Daëron M. Specificity and affinity of human Fc gamma receptors and their polymorphic variants for human IgG subclasses. **2009**. *Blood.* Vol 113 Issue 16: 3716–3725.

Bruins, AP 1998, 'Mechanistic aspects of electrospray ionization', Journal of Chromatography A, vol. 794, no. 1-2, pp. 345-357

References

Brünger, A. T. *et al.* (1998) 'Crystallography & NMR system: A new software suite for macromolecular structure determination', *Acta Crystallographica Section D: Biological Crystallography*, 54(5), pp. 905–921. doi: 10.1107/S0907444998003254.

Bruno, D. and Ruedi, A. (2006) 'Mass Spectrometry and Protein Analysis', *Science*, 312(5771), pp. 212–217. doi: 10.1126/science.1124619.

Bukowska, M. A. and Gru, M. G. (2013) 'New concepts and aids to facilitate crystallization', *Current Opinion in Structural Biology*, 23, pp. 409–416. doi: 10.1016/j.sbi.2013.03.003.

Chait, B. T. and Kent, S. B. H. (2008) 'Weighing Naked Proteins : Practical, High-Accuracy Mass Measurement of Peptides and Proteins Published by : American Association for the Advancement of Science Stable URL : http://www.jstor.org/stable/2879790', 257(5078), pp. 1885–1894.

Charles, M. J., McLuckey, S. A. and Glish, G. L. (1994) 'Competition between resonance ejection and ion dissociation during resonant excitation in a quadrupole ion trap', *Journal of the American Society for Mass Spectrometry*, 5(12), pp. 1031–1041. doi: 10.1016/1044-0305(94)85065-8.

Cherukuvada, S., Kaur, R. and Row, T. N. G. (2016) 'diversity : from pharmaceutical to materials applications', *CrystEngComm*. Royal Society of Chemistry. doi: 10.1039/C6CE01835A.

Chuang WL, Christ MD, Peng J, Rabenstein DL. An NMR and molecular modeling study of the site-specific binding of histamine by heparin, chemically modified heparin, and heparin-derived oligosaccharides. **2000.** *BioChemist.* Vol 39 Issue 13, 3542-3555.

Clark, K. L. *et al.* (1993) 'Co-crystal structure of the HNF-3 / fork head DNA-recognition motif resembles histone HS', 364(July), pp. 311–313.

Conrad H. E., *Heparin-Binding Proteins* 5th Addition, Academic Press, New York, 1998.

Cornilescu, G., Delaglio, F. and Bax, A. (1999) 'Protein backbone angle restraints from searching a database for chemical shift and sequence homology', *Journal of Biomolecular NMR*, 13, pp. 289–302.

References

Crispin, M., Yu, X. Bowden, T. A. 'Crystal structure of sialylated IgG Fc: Implications for the mechanism of intravenous immunoglobulin therapy' **2013** *Proceedings of the National Academy of Sciences,* vol 110 (38) 3544–3546

Cugola F.R., Fernandes, I.R., Russo, F.B., Freitas, B.C., Dias, J.L., Guimarães, K.P., Benazzato, C., Almeida, N., Pignatari, G.C., Romero, S., Polonio C.M., Cunha I., Freitas C.L., Brandão W.N., Rossato C., Andrade D.G., Faria D.de.P., Buchpigel C.A., Braconi C.T., Mendes E., Sall A.A., Zanotto M.de A., Peron J.P.S., Muotri A.R., Beltrão-Braga P.C.B.**.** The Brazilian Zika virus strain causes birth defects in experimental models. **2016**. *Nature* Vol 534: 267–271.

Dai L., Song J., Lu X., Deng Y.Q., Musyoki A.M., Cheng H., Zhang Y., Yuan Y., Song H., Haywood J., Xiao H., Yan J., Shi Y., Qin C.F., Qi J., Gao G.F. Structures of the Zika Virus Envelope Protein and Its Complex with a Flavivirus Broadly Protective Antibody **2016**. *Cell Host & Microbe*. Vol: 19, 696-704.

Dauter, Z. and Wlodawer, A. (2010) 'Impact of synchrotron radiation on macromolecular crystallography : a personal view', *Journal of Synchrotron Radiation*. International Union of Crystallography, 17, pp. 433–444. doi: 10.1107/S0909049510011611.

Delves P. J. Overview of the Immune System. **2017.** *MSD Man.* Vol **357:** 1777–1789.

Deng Y.Q., Dai J.X., Ji G.H., Jiang T., Wang H.J., Yang H., Tan W.L., Liu R., Yu M., Ge B.X., Zhu Q.Y., Qin E.D., Guo Y.J., Qin C.F. A Broadly Flavivirus Cross-Neutralizing Monoclonal Antibody that Recognizes a Novel Epitope within the Fusion Loop of E Protein. **2011.** *PLoS One.* Vol: 6 Issue: 1. e16059.

Dewald, H. D. (2009) 'Electrospray Ionization Mass Spectrometry: Fundamentals, Instrumentation and Applications (ed. Cole, Richard B.)', *Journal of Chemical Education*, 76(1), p. 33. doi: 10.1021/ed076p33.1.

References

Dowd C.J., Cooney C.L., Nugent M.A. Heparan sulfate mediates bFGF transport through basement membrane by diffusion with rapid reversible binding. **1999**. *J. Biol. Chem.* Vol: 274 issue: 8, 5236-5244.

Drenth J. (1994). *Principles of Protein X-ray Crystallography,* Chapter 2. Springer, New York.

Dube R., Rook G.A., Steele J., Brealey R. Dwek R., Rademacher T., Lennard-Jones J. Agalactosyl IgG in inflammatory bowel disease: Correlation with C-reactive protein. **1990.** *Gut,* Vol 31: 431–434.

Dulaney SB, Huang X: Strategies in synthesis of heparin/ heparan sulfate oligosaccharides: 2000–present. **2012.** *Adv Carbohydr Chem Biochem,* Vol 67: 95-136.

Emsley, P. and Cowtan, K. (2004) 'Coot: Model-building tools for molecular graphics', Acta Crystallographica Section D: Biological Crystallography. International Union of Crystallography, 60(12 I), pp. 2126–2132.

Emsley, P., Lohkamp, B., Scott, W. G. and Cowtan, K. (2010) 'Features and development of Coot', Acta Crystallographica Section D: Biological Crystallography. International Union of Crystallography, 66(4), pp. 486–501.

Engh, R. A. and Huber, R. (1991) 'Accurate Bond and Angle Parameters for X-ray Protein Structure Refinement', *Acta Crystallographica,* A47, pp. 392–400. doi: 10.1107/S0108767391001071.

Eriksson, J., Chait, B. T. and Fenyö, D. (2000) 'A statistical basis for testing the significance of mass spectrometric protein identification results', *Analytical Chemistry*, 72(5), pp. 999–1005. doi: 10.1021/ac990792j.

Evans, P. R. (2011) 'research papers An introduction to data reduction : space-group determination, scaling and intensity statistics research papers', *Acta Crystallographica Section D: Biological Crystallography*. International Union of Crystallography, D67, pp. 282–292. doi: 10.1107/S090744491003982X.

References

Evans, P. R. and Murshudov, G. N. (2013) 'How good are my data and what is the resolution?', Acta Crystallographica Section D: Biological Crystallography. International Union of Crystallography, 69(7), pp. 1204–1214.

Ewald, P.P. (ed.) (1963). *Fifty Years of X-ray Diffraction.* International Union of Crystallography/Oosthoek's Uitgeversmaatschappij, Urtrecht.

Faham S., Hileman R.E., Fromm J.R., Linhardt R.J., Rees D.C. Heparin Structure and Interactions with Basic Fibroblast Growth Factor. **1996.** *Science.* Vol 271 Issue 5252, 1116-1120.

Ferguson, P. L. and Smith, R. D. (2003) 'P Roteome a Nalysis By M Ass S Pectrometry ∗', *Proteome*, pp. 399–424. doi: 10.1146/annurev.biophys.32.110601.141854.

Ferrante A, Beard L J, Feldman R G. IgG subclass distribution of antibodies to bacterial and viral antigens. **1990.** *Pediatr Infect Dis J*: Vol 9 Issue 8: 16–24.

Ferrara C., Stuart, F., Sondermann, P., Brünker, P. & Umaña, P. The carbohydrate at FcγRIIIa Asn162. An element required for high affinity binding to non-fucosylated IgG glycoforms. **(2006).** *J. Biol. Chem*. 281, 5032–5036.

Fraser, C. M. and Fleischmann, R. D. (1997) 'Strategies for whole microbial genome sequencing and analysis', *Electrophoresis*, 18(8), pp. 1207–1216. doi: 10.1002/elps.1150180803.

French, B. Y. S. and Wilson, K. (1978) 'On the Treatment of Negative Intensity Observations po (Olx) ~ o px (xl O) po (O).', *Acta Crystallographica*, A34, pp. 517–525.

Gaboriaud, C. *et al.* (2003) 'The Crystal Structure of the Globular Head of Complement Protein C1q Provides a Basis for Its Versatile Recognition Properties', *Journal of Biological Chemistry*, 278(47), pp. 46974–46982. doi: 10.1074/jbc.M307764200.

Ganem, B., Li, Y. T. and Henion, J. D. (1991) 'Observation of Noncovalent Enzyme-Substrate and Enzyme-Product Complexes by Ion-Spray Mass Spectrometry', *Journal of the American Chemical Society*, 113(20), pp. 7818–7819. doi: 10.1021/ja00020a085.

References

Gatti G., Casu B., Hamer G.K., Perlin A.S., Studies on the Conformation of Heparin by ^{1}H and ^{13}C NMR Spectroscopy. 1979. *Macromolecules.* Vol 12 Issue 5: 1001-1007.

Ghetie V., Ward E.S. Multiple Roles for The Major Histocompatibility Complex Class I – Related Receptor FcRn. **2000.** *Annu. Rev. Immunol.* Vol 18: 739-766.

Ghezzi S., Cooper L., Rubio A., Pagani I., Capobianchi M.R., Ippolito G., Pelletier J., Meneghetti M.C.Z., Lima M.A., Skidmore M.A., Broccoli V., Yates E.A., Vicenzi E. Heparin prevents Zika virus induced-cytopathic effects in human neural progenitor cells. **2017.** *Antivir. Res.* Vol: 140, 13-17.

Giacovazzo C. *et al.* (1992). Fundamentals of Crystallography. International Union of Crystallography/ Oxford University Press, Oxford.

Greenfield, N. J. (2006) 'Using Circular Dichroism Spectra to Estimate Protein Secondary Structure', *Nature Protocols*, 1(6), pp. 2876–2890. doi: 10.1038/nprot.2006.202.Using.

Gruner SM. X-ray detectors for macromolecular crystallography. Curr Opin Struct Biol 1994;4:765–9.

Guerrini M., Beccati D., Shriver Z., Naggi A., Viswanathan K., Bisio A., Capila I., Lansing J.C., Guglieri S., Fraser B., Al-Hakim A., Gunay N.S., Zhang Z., Robinson L., Buhse L., Nasr M., Woodcock J., Langer R., Venkataraman G., Linhardt R.J., Casu B., Torri G., Sasisekharan R. Oversulfated chondroitin sulfate is a contaminant in heparin associated with adverse clinical events. **2008**. *Nat Biotechnol.* Vol 26 Issue 6: 669-675.

Hadfield L.C., *Heparin and heparin-like molecules inhibit the Alzheimer's β-secretase (BACE1): considerations for biological assay and future therapeutic development*. 2018. PhD Thesis. Keele University. Keele.

Hahn T. (5^{th} ed.) (1995). International Tables for Crystallography. Volume A. Space Group Symmetry. International Union of Crystallography/ Kluwer, Dordrecht.

References

Hammarstrom L, Carbonara AO, DeMarchi M, Lefranc G, Moller G, Smith CI. “Subclass restriction pattern of antigen-specific antibodies in donors with defective expression of IgG or IgA subclass heavy chain constant region Variability, biology, and function of human IgG genes”. **1987.** *Clin Immunol Immunopathol* 45(3):461–470.

Hammarstrom L, Smith CI. “IgG2 deficiency in a healthy blood donor Con-comitant lack of IgG2, IgA and IgE immunoglobulins and specific anti- carbohydrate antibodies”. **1983.** *Clin Exp Immunol* 51(3):600–604.

Hammond C. (1997). The Basics of Crystallography and diffraction, Chapters 2, 3 and 4. International union of crystallography/ oxford press, oxford.

Han Y.H., Garron M.L., Kin H.Y., Kim W.S., Zhang Z., Ryu K.S., Shaya D., Xiao A., Cheong C., Kim Y.S., Lindhardt R.J., Jeon Y.H., Cygler M., (2009). Structural Snapshots of Heparin Depolymerization by Heparin Lyase I. J. Biol. Chem. 284, 34019-34027

Hartl, F. U. and Hayer-hartl, M. (2002) ‘Molecular Chaperones in the Cytosol : from Nascent Chain to Folded Protein’, *Science*, 295(March), pp. 1852–1859.

Helliwell, J. R. (1984) ‘Synchrotron X-radiation protein crystallography : Instrumentation, methods and applications Synchrotron x-radiation protein crystallography : instrumentation, methods and applications’, *Reports on Progress in Physics*, 47, pp. 1043–1498.

Henzel W. J. *et al.* (1993) ‘Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases’, *Proceedings of the National Academy of Sciences of the United States of America*, 90(11), pp. 5011–5.

Ho, C. S. *et al.* (2003) ‘Electrospray ionisation mass spectrometry: principles and clinical applications.’ *The Clinical biochemist. Reviews*, 24(1), pp. 3–12.

Holm, L. and Sander, C. (1993) ‘Protein Structure Comparison by Alignment of Distance Matrices’, *Journal of Molecular Biology*, 233, pp. 123–138.

References

Holser, W. T. (1958) 'Point Groups and Plane Groups in a Two-Sided Plane and their Subgroups 1', *Zeitschrift f ü r Kristallographie,* 110(1–6), pp. 266–281.

Hoppensteadt D., Iqbal, O., Fareed, J. Chapter 21—Basic and clinical differences of heparin and low molecular weight heparin treatment. **2005**. *Chem. Biol. Heparin Heparan Sulfate.* Vol 1: 583–606.

Horsewood P, Warkentin TE, Hayward CP, Kelton JG. The epitope specificity of heparin-induced thrombocytopenia. **1996.** *Br. J. Haematol.* Vol 95, Issue 1: 161-167.

Hunefeld, F. L. (1840). Der Chemismus in der tierescher Organization, p. 160.

Janeway C. A., Travers, P., Walport, M., Shlomchik, M. Immunobiology 5th edition Section 33.2 The Immunoglobulin Fold Consists of a Beta-Sandwich Framework with Hypervariable Loops. W.H. Freeman and Company. 884 (2001).

Jefferis R., Kumararatne D.S. Selective IgG subclass deficiency: quantification and clinical relevance. **1990.** *Clin Exp Immunol.* Vol 81 Issue 3: 357–67.

Jensen, O., Podtelejnikov, A. and Mann, M. (1996) 'Identification of the Components of Simple Protein Mixtures by High-Accuracy Peptide Mass Mapping {…}', *Proc. Natl. Acad. Sci. USA*, 69(23), pp. 4741–4750.

Johnson, D. W. (2000) 'A rapid screening procedure for the diagnosis of peroxisomal disorders: Quantification of very long-chain fatty acids, as dimethylaminoethyl esters, in plasma and blood spots, by electrospray tandem mass spectrometry', *Journal of Inherited Metabolic Disease*, 23(5), pp. 475–486. doi: 10.1023/A:1005612214179.

Jones, T. A. and Thirup, S. (1986) 'Using known substructures in protein model building and crystallography', *The EMBO journal*, 5(4), pp. 819–822. doi: 10.1002/j.1460-2075.1986.tb04287.x.

Jonscher, K. R. and Yates, J. R. (1997) 'The quadrupole ion trap mass spectrometer - A small solution to a big challenge', *Analytical Biochemistry*, 244(1), pp. 1–15. doi: 10.1006/abio.1996.9877.

References

Katta, V.; Chait, B. T. ‘Conformational Changes in Proteins Probed by Hydrogen-exchange Electrospray-ionization’ (1991) *Rapid Cummun. Mass Spectrom.*, 5(February), pp. 214–217.

Kendrew JC, Bodo G, Dintzis HM, Parrish RG, Wyckoff H, Phillips DC. A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. Nature. 1958; 181:662–666. [PubMed: 13517261]

Khan S., Gor J., Mulloy B., Perkins S.J. Semi-Rigid Solution Structures of Heparin by Constrained X-ray Scattering Modelling: New Insight into Heparin–Protein Complexes. **2010**. *J. Mol. Biol.* Vol 395: 504-521.

Kim J-K, Firan M, Radu C.G., Kim C-H, Ghetie V., Ward E.S. Mapping the site on human IgG for binding of the MHC class I related receptor, FcRn. **1999.** Eur. *J. Immunol*. Vol 29: 2819–2825.

Kim, J. L., Nikolov, D. B. and Burely, S. K. (1993) ‘Co-Crystal Structure of TBP Recognizing the Minor Groove of a TATA Element’, *Nature*, 365, pp. 520–527.

King R., Bonfigilo, R. *et al.* (2000) ‘Mechanistic investigation of ionization suppression in electrospray ionization’, *Journal of the American Society for Mass Spectrometry*, 11(11), pp. 942–950. doi: 10.1016/s1044-0305(00)00163-x.

Kiyoshi M, Tsumoto K, Ishii-Watabe A, Caaveiro JMM. Glycosylation of IgG-Fc: a molecular perspective. **2017**. *Int. Immunol*, Vol. 29, Issue 7: 311–317

Kiyoshi M., Caaveiro J.M.M., Kawai T., Tashiro S., Ide T., Asaoka Y., Hatayama K., Tsumoto K. Structural basis for binding of human IgG1 to its high-affinity human receptor FcγRI. **2015.** *Nat. Commun.* Vol 6: 1-11.

Knox S., Merry C., Stringer S., Melrose J., Whitelock, J. Not All Perlecans Are Created Equal: Interactions with fibroblast growth factor (FGF) 2 and FGF recptors. *J. Biol. Chem.* **2002**, *277*, 14657–14665.

References

Konermann, L., Collings, B. A. and Douglas, D. J. (1997) 'Cytochrome c folding kinetics studied by time-resolved electrospray ionization mass spectrometry', *Biochemistry*, 36(18), pp. 5554–5559. doi: 10.1021/bi970046d.

Krapp S. Mimura Y. Jefferis R., Huber R., Sondermann P. Structural analysis of human IgG-Fc glycoforms reveals a correlation between glycosylation and structural integrity. **2003**. *J. Mol. Biol.* Vol 325(5): 979–89.

Kreuger J.; Kjellén, L. Heparan Sulfate Biosynthesis: Regulation and Variability. **2012.** *J. Histochem. Cytochem.* Vol 60: 898–907.

Kuhlbrandt, W., Neng Wang, D. and Fujiyoishi, Y. (1994) 'Atomic Model of Plant Light-Harvesting Complex by Electron Crystallography', *Nature*, 367, pp. 614–621.

Kuijpers T.W., Weening R.S., Out T.A. IgG subclass deficiencies and recurrent pyogenic infections unresponsiveness against bacterial polysaccharide antigens. **1992.** *Allergol Immunopathol (Madr)* Vol 20 Issue 1: 28–34.

Lacey, J. M. *et al.* (2001) 'Rapid determination of transferrin isoforms by immunoaffinity liquid chromatography and electrospray mass spectrometry', *Clinical Chemistry*, 47(3), pp. 513–518.

Ladd M.F.C. and Palmer R.A. (1993). *Structure Determination by X-ray Crystallography*. (3rd Ed.). Plenum Press, New York.

Lawrence D. A. Overview of the Immune System and Immunotoxicology. **2010**. *Compr. Toxicol.* Vol 5: 1–34.

Lazar G.A., Dang W., Karki S., Vafa O., Peng J.S., Hyun L., Chan C., Chung H.S., Eivazi A., Yoder S.C., VielmetteR J., Carmichael D.F., Hayes R.J., Dahiyat B.I. Engineered antibody Fc variants with enhanced effector function. **2006**. *Proc. Natl. Acad. Sci. U.S.A.* Vol 103: 4005–4010.

Leslie, A.G.W. (2000). In S*tructure and Dynamics of Biomolecules* (Ed. E. Fanchon *et al.*), pp. 14-20, Oxford University Press, Oxford

References

Li J.-P., Hagner-McWhirter Å., Kjellén L., Palgi J., Jalkanen M., Lindahl U., Biosynthesis of Heparin/Heparan Sulfate cDNA CLONING AND EXPRESSION OF D-GLUCURONYL C5-EPIMERASE FROM BOVINE LUNG. **1997.** *J. Biol. Chem.* Vol 272: 28158.

Lima M, Rudd T, Yates E. New Applications of Heparin and Other Glycosaminoglycans. **2017.** *Molecules.* Vol 22, 749-760.

Lin Y.-L., Lei H.-Y., Lin Y.-S., Yeh T.-M., Chen S.-H., Liu H.-S. Heparin inhibits dengue-2 virus infection of five human liver cell lines. **2002.** *Antiviral Res.*, Vol 56: 93–96.

Lindahl U., in *Heparin: Chemical and Biological Properties, Clinical Applications* **1989**, 5th ed. D. A. Lane and U. Lindahl, CRC Press, Inc., Boca Raton, FL, p. 159.

Lindahl U., Kusche-Gullberg M., Kjellen L. Regulated Diversity of Heparan Sulfate. **1998.** *J. Biol. Chem.* Vol: 273 Issue: 39, 24979-24982.

Lindahl, U.; Bäckström, G.; Thunberg, L.; Leder, I.G. Evidence for a 3-*O*-sulfated D-glucosamine residue in the antithrombin-binding sequence of heparin. **1980.** *Proc. Natl. Acad. Sci. USA*, Vol 77: 6551–6555.

Linhardt R.J., Liu J.: Synthetic heparin. **2012,** *Curr Opin Pharmacol* Vol 12: 217-219.

Loganathan D., Wang H.M., Mallis L.M., Linhardt R.J. Structural variation in the antithrombin III binding site region and its occurrence in heparin from different sources. **1990**. *Biochemistry*. Vol 29 Issue 18, 4362-4368.

Loo, J. A. *et al.* (1991) 'Solvent-induced conformational changes of polypeptides probed by electrospray-ionization mass spectrometry', *Rapid Communications in Mass Spectrometry*, 5(3), pp. 101–105. doi: 10.1002/rcm.1290050303.Fernandez-Metzler,

M. Barber R. D. Sedgwick. A. N. Tyler, R. S. B. (1981) 'Fast atom bombardmment of solids : a new ion source for mass specxtrometry', *J. Chem. Soc. Chem. Commun*, 101, pp. 325–327.

References

M.S., Rashed. *et al.* (1995) ‘Diagnosis of inborn errors of metabolism from blood spots by acylcarnitines and amino acids profiling using automated electrospray tandem mass spectrometry’, *Pediatric Research*, 38(3), pp. 324–331.

Magnusson, S., Krajcik, J. and Borko, H. (1993) ‘<Steven 2000.pdf>’, (5), pp. 95–132. doi: 10.1152/ajpheart.00773.2005.Fenn, J. B. *et al.* (no date) ‘Jllllll ~ apilar’, 246(6).

Mann, M. and Wilm, M. (1994) ‘Error-Tolerant Identification of Peptides in Sequence Databases by Peptide Sequence Tags’, *Analytical Chemistry*, 66(24), pp. 4390–4399. doi: 10.1021/ac00096a002.

Martin W.L. and Bjorkman, P.J. Characterization of the 2:1 complex between the class I MHC-related Fc receptor and its Fc and its Fc ligand in solution. **1999.** *Biochemistry*. Vol: 38: 12639–12647.

Martin WL, West AP, Gan L, Bjorkman PJ. Crystal Structure at 2.8 A° of an FcRn/ Heterodimeric Fc Complex: Mechanism of pH-Dependent Binding. **2001**. *Molecular Cell*, Vol 7: 867–877.

Mathews, B. W. (1968) ‘Solvent Content of Protein Crystals’, *Journal of Molecular Biology*, 33(August 1967), pp. 491–497.

Mazey, P. G. (1990) ‘A Global Approach to Molecular Symmetry: Theorams on Symmetry Relations Between Ground and Excited-State Configurations’, *Journal of the American Chemistry Society*, 112(6), pp. 3791–3802.

McNicholas, S., Potterton, E., Wilson, K.S. and Noble, M.E.M., 2011. Presenting your structures: the CCP4mg molecular-graphics software. Acta Cryst. D67, 67(4), pp.386-394.

McPherson A. (1991) ‘A Brief History of Protein Crystal Growth’, *Journal of Crystal Growth*, 110(1-2), pp. 1-10. doi: 10.1016/0022-0248(91)90859-4.

McPherson, A. (1999). Crystallization of Biological Macromolecules. Cold Spring Harbor: Cold Spring Harbor Laboratory Press.

References

Mcpherson, A. and Gavira, J. A. (2014) 'IYCr crystallization series Introduction to protein crystallization IYCr crystallization series', *Acta Crystallographica Section F: Structural Biology Communications*. International Union of Crystallography, F70, pp. 2–20. doi: 10.1107/S2053230X13033141.

Medesan C, Matesoi D, Radu C, Ghetie V, Ward ES. Delineation of the amino acid residues involved in transcytosis and catabolism of mouse IgG1 **1997.** *J. Immunol*. Vol 158: 2211–2217

Mighell, B. A. D. *et al.* (1983) 'Space-Group Frequencies for Organic Compounds', *Acta Crystallographica*, A39, pp. 737–740.

Mimura Y., Church S., Ghirlando R. Ashton P.R., Dong S., Goodall M., Lund J., Jefferies R. The influence of glycosylation on the thermal stability and effector function expression of human IgG1-Fc: properties of a series of truncated glycoforms. **2000.** *Mol. Immunol*. 37:697.

Mimura, Y. *et al.* (2001) 'Role of Oligosaccharide Residues of IgG1-Fc in FcγRIIb Binding', *Journal of Biological Chemistry*, 276(49), pp. 45539–45547. doi: 10.1074/jbc.M107478200.

Mimura, Y. *et al.* (2001) 'The influence of glycosylation on the thermal stability and effector function expression of human IgG1-Fc: Properties of a series of truncated glycoforms', *Molecular Immunology*, 37(12–13), pp. 697–706. doi: 10.1016/S0161-5890(00)00105-X.

Mirza, U. A., Cohen, S. L. and Chait, B. T. (1993) 'Heat-Induced Conformational Changes in Proteins Studied by Electrospray Ionization Mass Spectrometry', *Analytical Chemistry*, 65(1), pp. 1–6. doi: 10.1021/ac00049a003.

Miyahara J, Takahashi K, Amemiya Y, et al. A new type of X-ray area detector utilising LASER stimulated luminescence. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 1986;A246:572–8.

Morris, H. R. *et al.* (1996) 'High sensitivity collisionally-activated decomposition tandem mass spectrometry on a novel quadrupole/orthogonal-acceleration time-of-flight mass spectrometer',

References

Rapid Communications in Mass Spectrometry, 10(8), pp. 889–896. doi: 10.1002/(SICI)1097-0231(19960610)10:8<889::AID-RCM615>3.0.CO;2-F.Anries

Mulloy B., Forster M.J. Conformation and dynamics of heparin and heparan sulfate. **2000.** *Glycobiology.* Vol 10 Issue 11: 1147-1156.

Mulloy B., Forster M.J., Jones C., Davies D.B. NMR and molecular-modeling studies of the solution conformation of heparin. **1993.** *Biochem. J.* Vol 293 Issue 3: 849-858.

Murshudov, G. N. and Nicholls, R. A. (2011) 'research papers REFMAC 5 for the refinement of macromolecular crystal structures research papers', *Acta Crystallographica Section D: Biological Crystallography*, pp. 355–367. doi: 10.1107/S0907444911001314.

Murshudov, G. N., Vagin, A. and Dodson, E. J. (1997) 'Refinement of Macromolecular Structures by the Maximum-Likelihood Method Axj ABj Dj (s) Gc P (A)', *Acta Crystallographica Section D: Biological Crystallography*. International Union of Crystallography, D53, pp. 240–255. doi: 10.1107/S0907444996012255.

Nimmerjahn F. & Ravetch J. V. Fcγ receptors: old friends and new family members. **2006.** *Immunity* Vol 24: 19–28

Oganesyan V., Damschroder M.M., Leach W., Wu. H., Acqua W.F.D. Structural characterization of a mutated, ADCC-enhanced human Fc fragment. **2008.** *Mol. Immunol.* Vol 45: 1872-1882.

Okada, T. *et al.* (2002) 'Functional role of internal water molecules in rhodopsin revealed by x-ray crystallography', *PNAS*, 99(9), pp. 5982–5987.

P. Bruins (1998) 'Mechanistic aspects of electrospray ionization', *Journal of Chromatography A.*, 794(794), pp. 345–357.Aebersold, R. and Mann, M. (2003) '<Nature 2003 Aebersold.pdf>', 422(March).

Patterson, S.D., (1995) ‘Matrix-assisted laser-desorption/ionization mass spectrometric approaches for the identification of gel-separated proteins in the 5-50 pmol range’, *Electrophoresis.*, 16, pp. 1104–1114.

Phillips, J. C. *et al.* (1976) ‘Applications crystallography’:, *Proceedings of the National Academy of Sciences of the United States of America*, 73(1), pp. 128–132.

Plomp R, Ruhaak LR, Uh HW, Reiding KR, Selman M, Houwing-Duistermaat JJ, Slagboom PE, Beekman M, Wuhrer M. Subclass-specific IgG glycosylation is associated with markers of inflammation and metabolic health. **2017.** *Nature* Vol **7**: 1-10.

Poncz M, Surrey S, LaRocco P, Weiss M.J., Rappaport E.F., Conway T.M., Schwartz E. Cloning and characterization of platelet factor 4 cDNA derived from a human erythroleukemic cell line. **1987.** *Blood*. Vol 69: 219-223.

Potter M. Structural correlates of immunoglobulin diversity. **1983.** *Surv Immunol Res* Vol 2 Issue 1: 27–42.

Potterton, E. *et al.* (2003) ‘research papers A graphical user interface to the CCP 4 program suite research papers’, *Acta Crystallographica Section D: Biological Crystallography*, D59, pp. 1131–1137.

Prabakaran P., Vu B.K., Gan J., Feng Y., Dimitrov D.S., Ji X. Structure of an isolated unglycosylated antibody CH2 domain. **2008.** *Acta Crystallogr Sect D Biol Crystallogr.* Vol 64 (10): 1062–1067.

Presto J., Thuveson M., Carlsson P., Busse M., Wilén M., Eriksson I., Kusche-Gullberg M., Kjellén L. Heparan sulfate biosynthesis enzymes EXT1 and EXT2 affect NDST1 expression and heparan sulfate sulfation. *Proc. Natl. Acad. Sci. USA* **2008**, Vol 105: 4751–4756.

Qin Y., Ke J., Gu X., Fang J., Wang W., Cong Q., Li J. Tan, J., Brunzelle J.S., Zhang C. Structural and functional study of D-glucuronyl C5-epimerase. **2015.** *J. Biol. Chem.* Vol 290: 4620–4630.

References

Quast I., Peschke B., Lünemann J.D. Regulation of antibody effector functions through IgG Fc N-glycosylation. **2017.** *Cell Mol. Life Sci.* Vol: 74, 837-847.

Ramakumar, S. and Viswamitra, M. A. (1990) 'Space-Group Frequencies of Proteins and of Organic Compounds with More Than One Formula Unit in the Asymmetric Unit', *Acta Crystallographica Section D: Biological Crystallography*. International Union of Crystallography, A46, pp. 725–730. doi: 10.1107/S0108767390004512.

Rappsilber, J. *et al.* (2000) 'A generic strategy to analyze the spatial organization of multi-protein complexes by cross-linking and mass spectrometry', *Analytical Chemistry*, 72(2), pp. 267–275. doi: 10.1021/ac991081o.

Ravetch J. V. & Lanier, L. L. Immune inhibitory receptors. **2000.** *Science* Vol 290: 84–89

Ravetch J.V. Fc receptors. In *Fundamental Immunology* (5th ed. Paul, W. E.) 685–700 (Lippincott-Raven, Philadelphia, 2003).

Rhodes G. (2000). Crystallography made crystal clear (2nd ed.), pp. 56-57 and 173. Academic Press, San Diego.

Richmond, R. K. *et al.* (1997) 'Crystal structure of the nucleosome ˚ resolution core particle at 2 . 8 A', *Nature*, 7, pp. 251–260.

Rickles F.R., Falanga A. Molecular basis for the relationship between thrombosis and cancer. **2001.** *Thromb. Res.* Vol 102: 215–224.

Robinson H.C., Horner A.A., Höök M., Ögren S., Lindahl U. A Proteoglycan Form of Heparin and Its Degradation to Single-chain Molecules **1978.** *J. Biol. Chem.*, Vol 253: 6687.

Rowen, L., Mahalras, C., Hood, I. (1997) 'Sequencing the Human Genome', *Science*, 278(October), pp. 605. doi: 10.1126/science.278.5338.605.

References

Sabelkin, V. *et al.* (2016) 'Crack Initiation from Corrosion Pit in Three Aluminum Alloys Under Ambient and Saltwater Environments', *Journal of Materials Engineering and Performance*, 25(4), pp. 1631–1642. doi: 10.1007/s11665-016-1996-5.

Salmavirta M., Lidholt K., Lindahl U. Heparan sulfate: a piece of information. **1996.** *FASEB J.* Vol: 10, Issue: 11, 1270-1279.

Savjani, J. K. (2015) 'Co - crystallization : An approach to improve the performance characteristics of active', (September), pp. 147–151. doi: 10.4103/0973-8398.160309.

Schauer U, Stemberg F, Rieger C H, Buttner W, Borte M, Schubert S, Möllers H, Riedel F, Herz U, Renz H, Herzog W. Levels of antibodies specific to tetanus toxoid, Haemophilus influenzae type b, and pneumococcal capsular polysaccharide in healthy children and adults. **2003.** *Clin Diagn Lab Immunol* Vol 10 Issue 2: 202–207.

Schur P.H. IgG subclasses. A historical perspective. **1988** *Monogr. Allergy.* Vol 23: 1–11.

Shaya D., Tocilj A., Li Y., Myette J., Venkataraman G., Sasisekharan R., Cygler M. (2006) Crystal Structure of Heparinase II from Pedobacter heparinus and Its Complex with a Disaccharide Product. J. Biol. Chem, 281, 15525-15525

Shevchenko, A. *et al.* (1997) 'Rapid "De Novo" peptide sequencing by a combination of nanoelectrospray isotopic labelling and Q-TOF-MS', *Rapid Commun. Mass Spectrom.*, 11, pp. 1015–1024.

Siber G.R., Schur P.H., Aisenberg A.C., Weitzman S.A., Schiffman G. "Correlation between serum IgG-2 concentrations and the antibody response to bacterial polysaccharide antigens". **1980.** *N Engl J Med.* Vol 303 Issue 4: 178–182.

Singhal, N. *et al.* (2015) 'MALDI-TOF mass spectrometry: an emerging technology for microbial identification and diagnosis.', *Frontiers in microbiology*, 6(August), p. 791. doi: 10.3389/fmicb.2015.00791.

References

Sippl, M. J. (1993) 'Recognition of Errord in Three-Dimensional Structures of Proteins', *PROTEINS: Structure, Function and Genetics*, 17, pp. 355–362.

Smyth, M. S. and Martin, J. H. J. (2000) 'Review x Ray crystallography', *Journal of Clinical Pathology: Molecular Pathology*, 53, pp. 8–14.

Sobel M., Soler D.F., Kermode J.C., Harris R.B. Localization and characterization of a heparin binding domain peptide of human von Willebrand factor. **1992.** *J. Biol. Chem.* Vol: 267, 8857-8862.

Sondermann P., Huber R., Oosthuizen V., Jacob, U. The 3.2-A crystal structure of the human IgG1 Fc fragment-Fc gammaRIII complex. **2000.** *Nature* Vol 406: 267–273

Spengler, B. *et al.* (1992) 'Peptide sequencing by matrix-assisted laser-desorption mass spectrometry', *Rapid Communications in Mass Spectrometry*, 6(2), pp. 105–108. doi: 10.1002/rcm.1290060207.

Stapleton NM, Andersen JT, Stemerding AM, Bjarnarson SP, Verheul RC, Ger- ritsen J, Competition for FcRn-mediated transport gives rise to short half-life of human IgG3 and offers therapeutic potential. **2011.** *Nat Commun.* Vol 2: 599.

Su, X., Zhang, H. and Terwilliger, T. C. (2014) 'Protein crystallography from the perspective of technology developments', *Crystallography Reviews*, (December), pp. 37–41. doi: 10.1080/0889311X.2014.973868.

Subedi G.P. and Barb A.W. The structural role of antibody N-glycosylation in receptor interactions. **2015.** *Structure.* Vol 23: 1573.

Subedi, G.P., Hanson, Q.M., Barb, A.W. Restricted motion of the conserved immunoglobulin G1 N-glycan is essential for efficient FcγRIIIa binding. **2014.** *Structure.* Vol 22: 1478.

Suvarna S, Espinasse B, Qi R., Lubica R., Poncz M., Cines D.B., Wiesner M.R., Arepally G.M. Determinants of PF4/heparin immunogenicity. **2007.** *Blood.* Vol 110 Issue 13: 4253-4260.

References

Tam, S. H., McCarthy S.G, Armstrong A.A., Smani, S., Wu S.J., Liu X., Gervais A., Ernst R., Saro D., Decker R. Luo J., Gilliland G.L., Chiu M.L., Scallon B.J. Variants with Ablated Immune Functionality, **2017.** *Antibodies*, Vol 6, Issue 2 doi: 10.3390/antib6030012.

Venkataraman G., Shriver Z., Raman R., Sasisekharan R. Sequencing Complex Polysaccharides. **1999.** *Science.* Vol 286, Issue 5439, 537-542.

Vertes, A. (2008) 'Mass spectrometry in proteomics', *Medical Applications of Mass Spectrometry*, pp. 173–194. doi: 10.1016/B978-044451980-1.50010-0.

Vertes, A. (2008) 'Mass spectrometry in proteomics', *Medical Applications of Mass Spectrometry*, pp. 173–194. doi: 10.1016/B978-044451980-1.50010-0.

Vestal, M. L., Juhasz, P. and Martin, S. A. (1995) 'Delayed extraction matrix-assisted laser desorption time-of-flight mass spectrometry', *Rapid Communications in Mass Spectrometry*, 9(11), pp. 1044–1050. doi: 10.1002/rcm.1290091115.

Vidarsson G, Dekkers G, Rispens T. IgG subclasses and allotypes: from structure to effector functions. **2014**. *Front. Immunol.* Vol: 5.

Viskov C., Just M., Laux, V., Mourier P., Lorenz M. Description of the chemical and pharmacological characteristics of a new hemisynthetic ultra-low-molecular-weight heparin, **2009**, *J. Thromb. Haemost.* Vol 7: 1143–1151.

Webster, J. and Oxley, D. (2009) 'Protein Identification by Peptide Mass Fingerprinting using MALDI-TOF Mass Spectrometry', pp. 1117–1129. doi: 10.1007/978-1-59745-198-7_120.

Weichenberger, C. X. and Pozharski, E. (2013) 'laboratory communications Visualizing ligand molecules in twilight electron density laboratory communications', *Acta Crystallographica Section F: Structural Biology Communications*. International Union of Crystallography, F69, pp. 195–200. doi: 10.1107/S1744309112044387.

References

Wiley, W. C. and McLaren, I. H. (1955) 'Time-of-flight mass spectrometer with improved resolution', *Review of Scientific Instruments*, 26(12), pp. 1150–1157. doi: 10.1063/1.1715212.

Wilm, M. and Mann, M. (1996) 'Analytical properties of the nanoelectrospray ion source', *Analytical Chemistry*, 68(1), pp. 1–8. doi: 10.1021/ac9509519.

Winn, M. D. *et al.* (2011) 'Overview of the CCP 4 suite and current developments Overview of the CCP 4 suite and current developments', *Acta Crystallographica Section D: Biological Crystallography*, D67, pp. 235–242. doi: 10.1107/S0907444910045749.

Wlodawer, A. *et al.* (2013) 'Protein Crystallography for Aspiring Crystallographers or How to Aviod Pitfalls and Traps in Macromolecular Structure Determination', *The FEBs Journal*, 280(22), pp. 5705–5736. doi: 10.1111/febs.12495.Protein.

Wright A., Tao M.H., Kabat E.A., Morrison S.L. Antibody variable region glycosylation: position effects on antigen binding and carbohydrate structure. **1991**. *EMBO J.* Vol 10 Issue 10: 2717–2723.

Wuhrer M, Porcelijn L, Kapur R, Koeleman CA, Deelder A, deHaas M, Vidarsson G. Regulated glycosylation patterns of IgG during alloimmune responses against human platelet antigens. **2009.** *J ProteomeRes*) Vol 8 Issue 2: 450–456.

Wukovitz, S. W. and Yeates, T. O. (1995) 'Why Protein Crystals Favour Some Space-Groups Over Others', *Nature Structural Biology*, 2(12), pp. 1062–1067.

Xu R., Ori A., Rudd T.R., Uniewicz K.A., Ahmed Y.A., Guimond S.E., Skidmore M.A., Siligardi, G.; Yates E.A., Fernig D.G., Diversification of the structural determinants of fibroblast growth factor-heparin interactions: Implications for binding specificity **2012a**,. *J. Biol. Chem*. Vol 287: 40061–40073.

Xu Y., Pempe E.H., Liu J. Chemoenzymatic synthesis of heparin oligosaccharides with both anti-factor Xa and anti-factor IIa activities. **2012b**, *J. Biol. Chem*. Vol 287: 29054–29061.

References

Young, M. M. *et al.* (2000) ‘High throughput protein fold identification by using experimental constraints derived from intramolecular cross-links and mass spectrometry’, *Proc.Natl.Acad.Sci.U.S.A*, 97(11), pp. 5802–5806.

Yu, W.H., and Woessner, J.F., Jr. (2000). Heparan sulfate proteoglycans as extracellular docking molecules for matrilysin (matrix metalloproteinase 7). J. Biol. Chem. 275, 4183–4191.

Zauner G., Selman M.H., Bondt A., Rombouts Y., Blank D, Deelder A.M. Glycoproteomic analysis of antibodies. **2013.** *Mol. Cell. Proteomics* Vol 12 Issue 4: 856–65.

Zhao H., Fernandez E., Dowd K.A., Speer S.D., Platt D.J., Gorman M.J., Govero J., Nelson C.A., Pierson T.C., Diamond M.S., Fremont D.H. Structural Basis of Zika Virus-Specific Antibody Protection. **2016**. *Cell.* Vol: 166, 1016-1027.

Ziporen L., Li Z.Q., Park K.S., Sabnekar P., Liu W.Y., Arepally G., Shoenfeld Y., Kieber-Emmons T., Cines D.B., Poncz. M. Defining an Antigenic Epitope on Platelet Factor 4 Associated With Heparin-Induced Thrombocytopenia. **1998.** *Blood*, Vol 92, Issue 9, 3250-3259

Appendix 1 – Full ESI result Tables

Start		End	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	U	Peptide
164	–	188	685.8033	2739.1842	2739.5033	-116	1	99	1.1e-007	1		E.LLGGPSVFLFPPKPKDTLMISRTPE.V
164	–	188	914.0761	2739.2064	2739.5033	-108	1	158	1.2e-013	1		E.LLGGPSVFLFPPKPKDTLMISRTPE.V
164	–	188	689.8008	2755.1740	2755.4983	-118	1	78	1.3e-005	1		E.LLGGPSVFLFPPKPKDTLMISRTPE.V + Oxidation (M)
164	–	188	919.4057	2755.1952	2755.4983	-110	1	130	9e-011	1		E.LLGGPSVFLFPPKPKDTLMISRTPE.V + Oxidation (M)
189	–	202	792.7769	1583.5393	1583.7138	-110	2	67	0.00019	1		E.VTCVVVDVSHEDPE.V
288	–	312	972.3779	2914.1119	2914.4245	-107	2	92	6.8e-007	1	U	E.MTKNQVSLTCLVKGFYPSDIAVEWE.S

Table 1. The mass spectrometry data for the sequences of the IgG3 Fc fragment produced by the MS/MS analysis of a GluC digest performed on the Purified human serum IgG Fc Fragment (P80-204).

Start		End	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	U	Peptide
117	–	141	685.8033	2739.1842	2739.5033	-116	1	99	1.1e-007	1		E.LLGGPSVFLFPPKPKDTLMISRTPE.V
117	–	141	914.0761	2739.2064	2739.5033	-108	1	158	1.2e-013	1		E.LLGGPSVFLFPPKPKDTLMISRTPE.V
117	–	141	689.8008	2755.1740	2755.4983	-118	1	78	1.3e-005	1		E.LLGGPSVFLFPPKPKDTLMISRTPE.V + Oxidation (M)
117	–	141	919.4057	2755.1952	2755.4983	-110	1	130	9e-011	1		E.LLGGPSVFLFPPKPKDTLMISRTPE.V + Oxidation (M)
142	–	155	792.7769	1583.5393	1583.7138	-110	2	67	0.00019	1		E.VTCVVVDVSHEDPE.V
241	–	265	966.3943	2896.1610	2896.4681	-106	2	72	5.3e-005	1	U	E.LTKNQVSLTCLVKGFYPSDIAVEWE.S

Table 2. The mass spectrometry data for the sequences of the IgG1 Fc fragment produced by the MS/MS analysis of a GluC digest performed on the Purified human serum IgG Fc Fragment (P80-204).

Start		End	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	U	Peptide
106	–	137	880.6046	3518.3894	3518.7764	-110	1	66	0.00022	1	U	E.CPPCPAPPVAGPSVFLFPPKPKDTLMISRTPE. V + Oxidation (M)
138	–	151	792.7769	1583.5393	1583.7138	-110	2	67	0.00019	1		E.VTCVVVDVSHEDPE.V

Table 3. The mass spectrometry data for the sequences of the IgG2 Fc fragment produced by the MS/MS analysis of a GluC digest performed on the Purified human serum IgG Fc Fragment (P80-204).

Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Peptide
790.8323	1579.6501	1579.7334	-52.75	1	22	8.5	1	MLRSGANDLGGTLME + Oxidation (M)
876.8673	1751.7200	1751.8400	-68.50	1	16	29	1	ALDLCYQTSQLAGIAE
679.7405	1357.4665	1357.6878	-163.00	1	16	28	1	WNEGTIVVKSPE
584.9051	1751.6936	1751.8764	-104.33	2	13	57	1	SMIFQTEKQLKDLGD
844.0205	2529.0396	2529.3525	-123.71	1	12	51	1	IQRFILRSLGEKRQLYYAME + Oxidation (M)
713.6000	2137.7781	2137.9959	-101.86	1	10	99	1	SGDVLLMHLGMSGSFRVAMD + Oxidation (M)
794.3075	1586.6004	1586.8491	-156.70	1	10	1.4e+002	1	FLARVPENVVVVMD
838.6910	2513.0512	2513.2656	-85.30	2	9	1.2e+002	1	KSKNSLEDQPNAQKALAIQCRD
1162.8065	3485.3977	3485.9227	-150.62	0	8	1.7e+002	1	TRPLKHKWRPSPLLVMQRNSSVPNLKMKE + Oxidation (M)

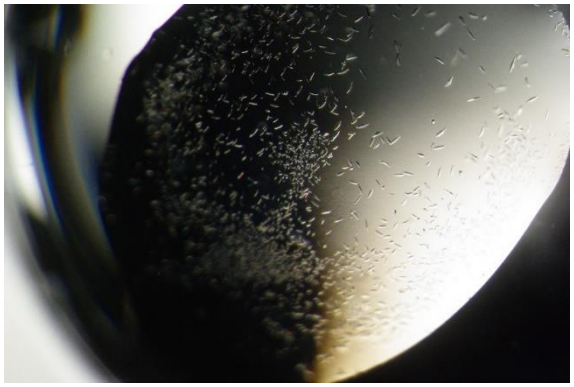
Appendix 2 – Heparin Soaking and Cryoprotection Trials

Tray Number	HE3A1

Volume Added	Time	Comments
0.5 µl, 95 mM heparin.	09:00	Heparin added- coverslip broke however no glass entered the well. 102-6762
	09:10	No change. 102-6773
	09:20	No change. 102-6776
	09:30	No change. 102-6777
	09:40	There appears to be some loss of width in some of the very small crystals. 102-6783 & 102-6785
	09:50	No change. 102-6786+6787
0.5µl, 95 mM heparin.	10:00	Second addition. 102-6793
	10:10	No change. 102-6815
	10:20	No change. 102-6816
	10:30	No change. 102-6819
	10:40	No change. 102-6834
	10:50	No change. 102-6837
	11:00	No change. 102-6852

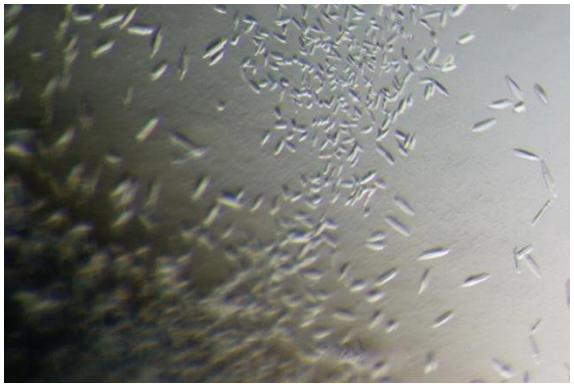
	11:20	No change. 102-6881
	11:40	No change. 102-6885
	12:00	No change. 102-6895
	12:30	No change. 102-6901
	13:00	No change. 102-6903
	13:30	No change. 102-6905
	14:00	No change. 102-6907
	15:00	No change. 102-6910
	16:00	No change. 102-6914
	17:00	No change. 102-6921
Wednesday	10:00	No change. 102-6940
	16:00	No change. 102-6950 + 102-6953
Thursday	10:00	No change. 102-6965
0.1 µl, 1.55 M Na Acetate pH 4.5 added (Final concentration 50 mM).	11:00	Crystals Stable. 102-6967
	11:10	No change. 102-6969
	11:20	No change. 102-6972
	11:30	No change. 102-6974
	11:40	No change. 102-6975

	11:50	No change. 102-6977
	12:00	No change. 102-6978
	12:20	No change. 102-6983
	12:40	No change. 102-6985
	13:00	No change. 102-6986
	13:30	No change. 102-6991
	14:00	No change. 102-6991
	15:00	No change. 102-7003
	16:00	No change. 102-7004
	17:00	No change. 102-7008
Friday	10:00	No change. 102-7013
	16:00	No change. 102-7016
Saturday	10:00	No change. 102-7017
	16:00	No change. 102-7018
Monday	10:00	No change. 102-7019
	16:00	No change. 102-7021
Tuesday	10:00	No change. 102-7022

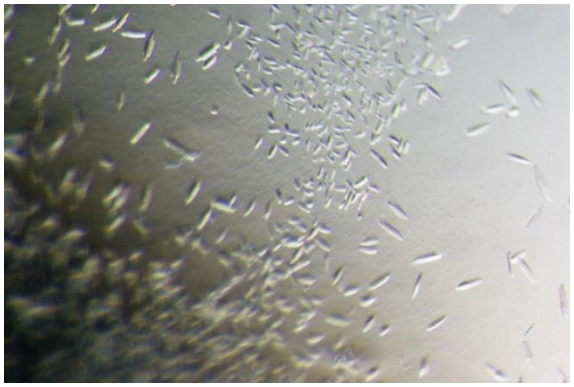


102-6762

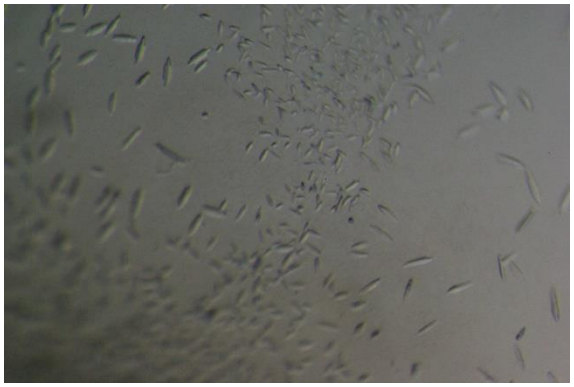
102-6773



102-6776



102-6777



102-6783

102-6785

102-6786

102-6787

102-6793

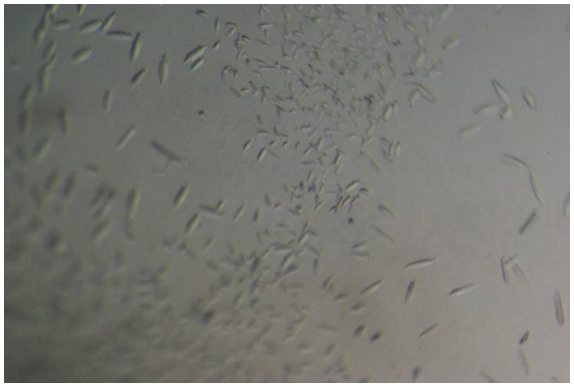
102-6815

102-6816

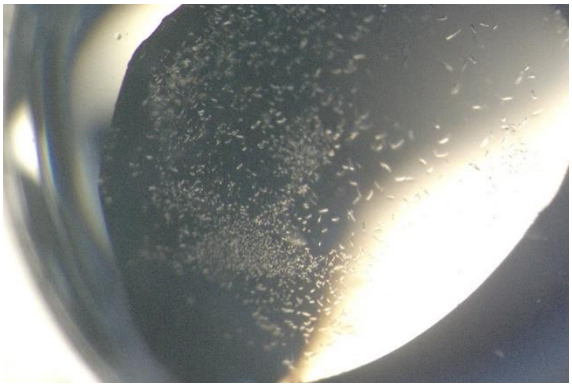
102-6819

102-6834

102-6837



102-6852



102-6881

102-6885

102-6895

102-6901

102-6903

102-6905

102-6907

102-6910

102-6914

102-6921

102-6940

102-6950

102-6953

102-6965

102-6967

102-6969

102-6972

102-6974

102-6975

102-6977

102-6978

102-6983

102-6985

102-6986

102-6991

102-6992

102-7003

102-7004

102-7008

102-7013

102-7016

102-7017

102-7018

Appendix 2

102-7019

102-7021

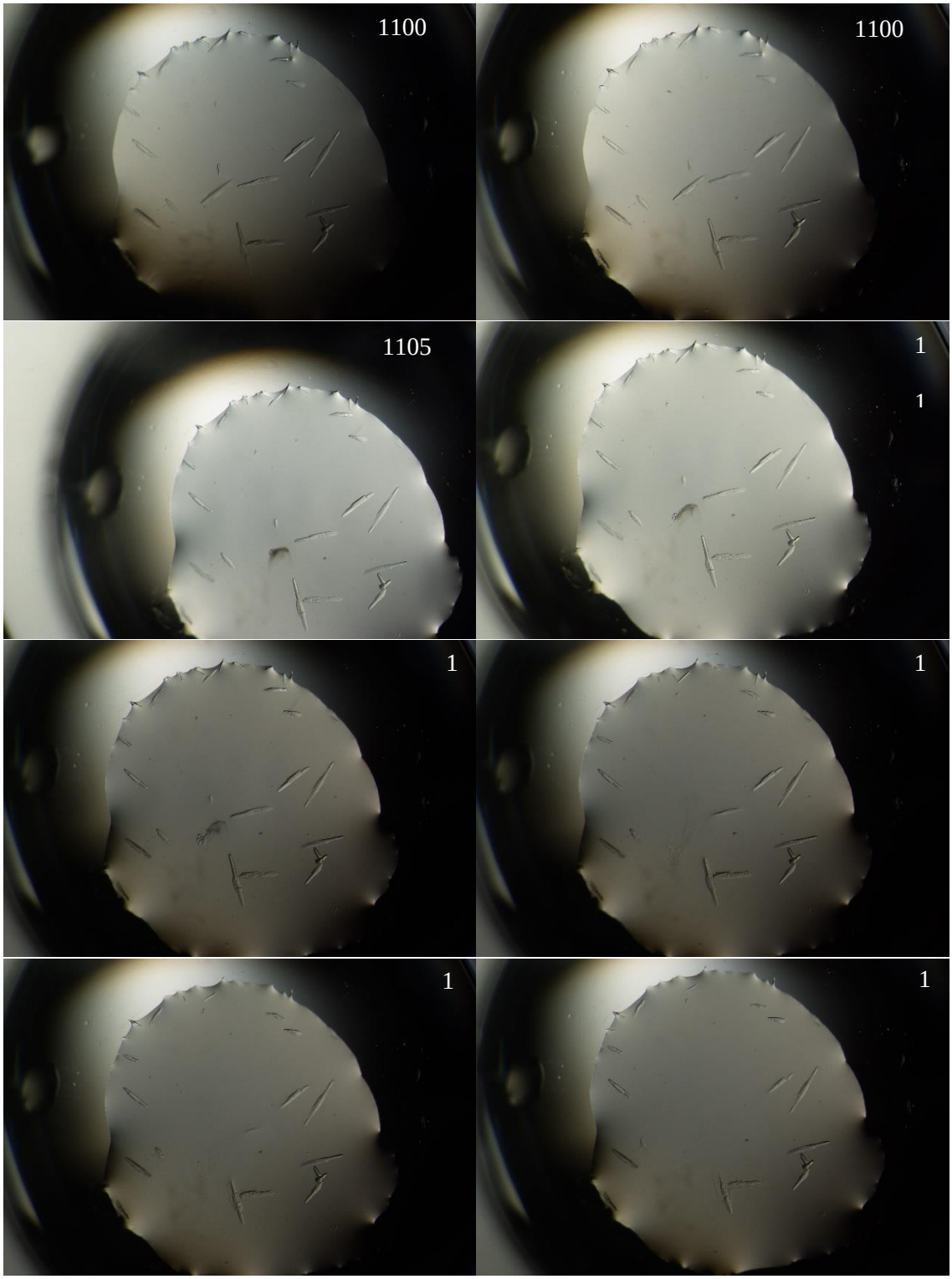
102-7022

Well Number	HE4C2

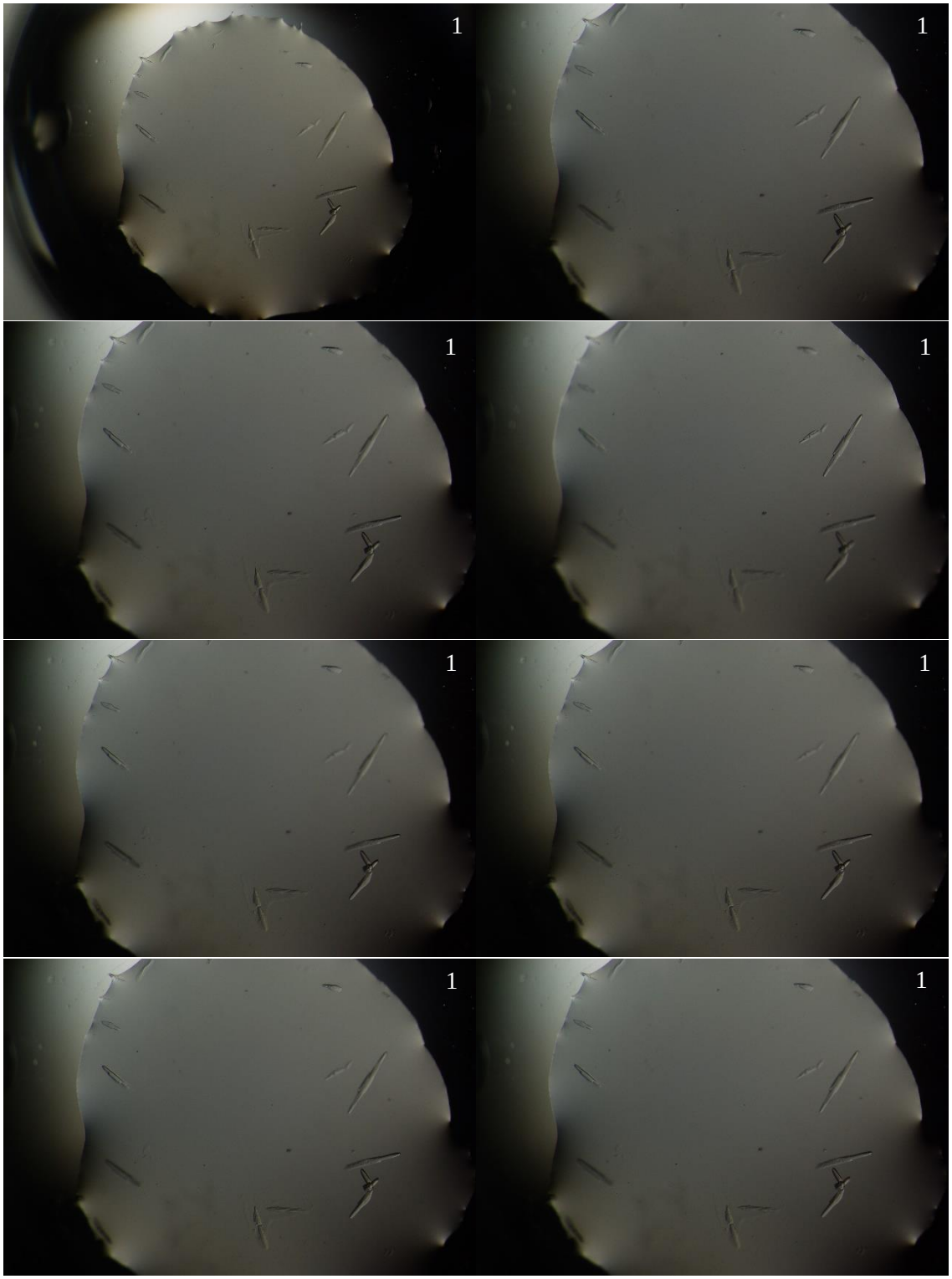
Volume Added	Time	Comments
0.5 µl 95 mM heparin	11:00	Heparin added to the well. The crystals are stable.
0.5 µl 0.30 M Na Acetate pH 4.5	11:05	Acetate added to the well. The crystal directly under the addition site dissolved almost immediately.
	11:08	Some of the crystals dissolved rapidly with some of the smaller ones dissolving completely
	11:10	There was a further reduction in crystal size.
	11:12	Most of the crystals at the top of the drop had dissolved by this point however there are some which seem unchanged.
	11:14	Even more degradation of the crystals with some of the larger ones breaking up. However some of the larger daggy one seem fine.
	11:16	No change.
	11:18	One of the crystals at 10 o'clock which looked stable has started to dissolve slightly.
	11:20	No change.
	11:22	No change.
	11:24	A slight reduction in size in some of the more shrivelled crystals.
	11:26	No change
	11:28	No change
	11:30	No change

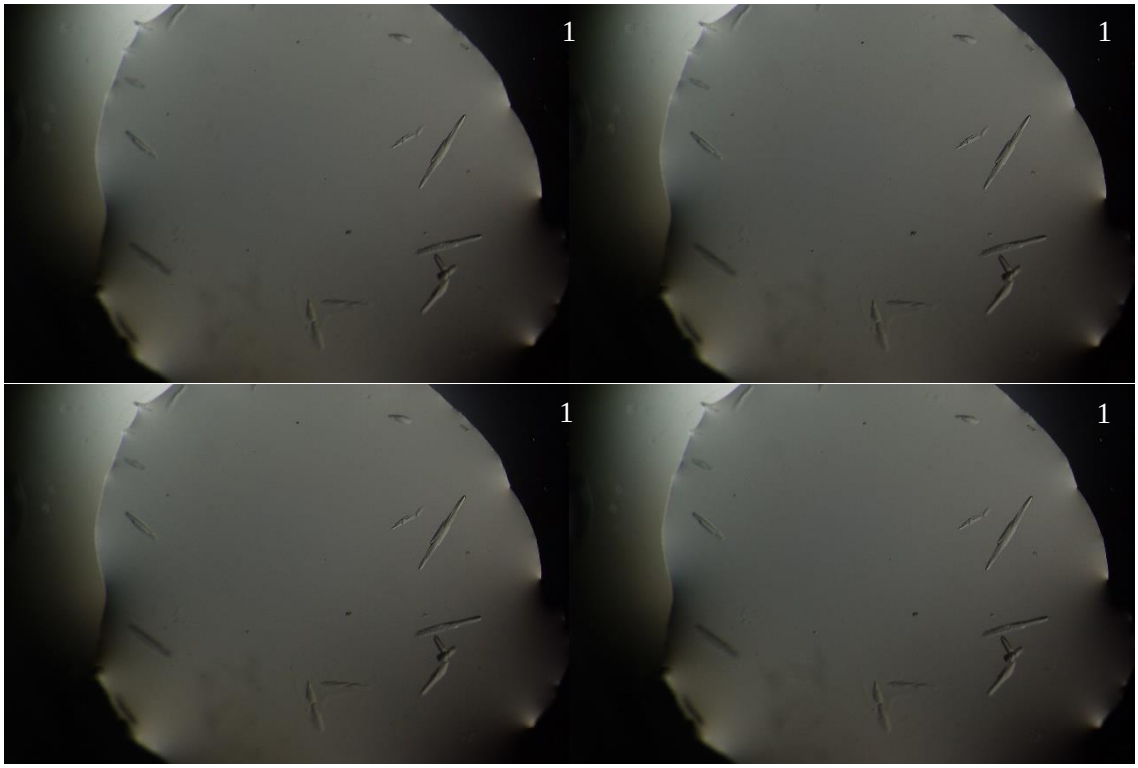
	11:32	The crystals which were stable to begin with remained so but the ones which were dissolving started to get even worse.
	11:34	No Change.
	12:30	The well seems to have stabilised. 102-7051.
0.5 µl 50 % Glycerol added	13:40	Crystals dissolved by 1347.
0.5 µl 50 % Glycerol added	13:40:30	
0.5 µl 50 % Glycerol added	13:41	
0.5 µl 50 % Glycerol added	13:41:30	

Appendix 2



Appendix 2





This trial was performed due to the dissolving of the larger xtals within wells: HE4A2 and HE4C1 during xtal freezing on 04/07/18 after a ligand soak trial on smaller xtals in well HE3A1 with the same concentrations and volumes (1 µl 95 mM heparin and 0.1 µl 1.55 M Na Acetate pH 4.5) showed that the xtals were stable after these additions.

- Make up 1 ml of 0.30 M Acetate stock using 100 µl of the 3 M stock and 900 µl of filtered deionised H_2O.
- Add 0.5 µl of 95 mM heparin stock and observe to make sure well is stable.
- Wait 5 minutes and add 0.5 µl of the 0.35 M stock of Acetate and observe whether crystals dissolve and/or how long that takes.
- Adding these two volumes would reduce the protein concentration to 0.051 mM and the heparin concentration to 15.8 mM. This would give a molar fold difference of 385.4 before the addition of any cryo buffer.