Solution Structures of Glycosaminoglycans and their Interactions with Complement Factor H

Thesis Presented for the Degree of
Doctor of Philosphy

by

Sana Ullah Khan

Department of Structural and Molecular Biology,
University College London

February 2011

I, Sana Ullah Khan, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

Factor H (FH), a complement regulatory protein comprised of 20 SCR domains, is able to discriminate host from pathogen cell surfaces by recognising glycosaminoglycans such as heparan sulphate on host cells to protect these from complement attack. Heparin is an analogue of HS. FH has heparin and HS binding sites located at SCR-6/8 and SCR-19/20. While the structure of FH is known from X-ray solution scattering and analytical ultracentrifugation (AUC), the effects of heparin and HS on intact FH are not yet known. To evaluate the interactions with FH, solution structures of heparin fragments were determined by a combination of X-ray solution scattering, AUC and constrained scattering modelling. These results revealed that heparin fragments starting from dp18 progressively show higher degrees of bending up to dp36, and resemble known crystal structures of heparin-protein complexes. Similar structural studies on similar-sized HS fragments showed that HS fragments exhibit more bent structures than heparin. This greater bending in HS fragments might be due to the reduced degree of sulphation in HS molecules. To assess the effect of heparin and HS fragments on intact FH, FH was studied in the presence of a range of purified heparin and HS fragments by scattering and AUC. The smallest heparin fragments showed little effect on the radius of gyration R_G values of FH. Dramatic increases in the R_G values were seen with the larger heparin sizes, and both AUC and scattering showed that a series of large but compact complexes were formed. In the presence of HS, smaller increases in oligomer formation were observed. The heparin- and HS-induced formation of oligomers in FH provide a first molecular picture of how FH may interact with host cell surfaces, and may facilitate an improved understanding of how extracellular deposits form on Bruch's membrane during the development of age-related macular degeneration.

Acknowledgements

I would like to thank my supervisor Professor Stephen John Perkins for providing me the opportunity to work with him in his well established research laboratory and for his guidance, encouragement and support throughout the period of my study with him. I am also thankful to my thesis committee members, Professor John Ward and Dr. Snezana Djordjevic for their encouragement and useful discussion of my results.

I would like to thank my collaborator Dr. Barbara Mulloy for her help with the preparations of heparin and heparan sulphate fragments at the National Institute of Biological Standard and Control (NIBSC), South Mimms, Potters Bar, Hertfordshire.

I thank the Higher Education Commission of Pakistan and the Mercer Fund of the Fight For Sight Charity for grant support.

I would also like to thank Dr. Azubuike I. Okemefuna for his computational help, Mr. Jayesh Gor for his help with ultracentrifugation experiments, Dr. Ruodan Nan for her help in Factor H preparation and all my laboratory colleagues for useful discussions.

I would like thank Miss Kanwal Zehra Rizvi for fruitful discussions and her encouragement at each step and finally, I am very grateful with depth of my heart to my family members, particularly my elder brother Professor Muhammad Amin Khan for their love, help and continued moral support.

Contents **Page**

Contents – Figures

Chapter One

Chapter Two

Chapter Three

Chapter Four

Chapter Five

Chapter Six

Contents – Tables **Page**

Chapter One

Chapter Four

Chapter Five

Chapter Six

Abbreviations

aHUS	atypical haemolytic uraemic syndrome
AMD	age-related macular degeneration
AUC	analytical ultracentrifugation
C1INH	C1 inhibitor
CCD	charge-coupled device
CFHR1	complement FH related protein 1
CR1	complement receptor type 1
CRP	C-reactive protein
DAF	decay accelerating factor
dp	degree of polymerisation
D_{max}	maximum dimension
DS	dermatan sulphate
EDTA	ethylenediaminetetraacetic acid
EM	electron microscopy
ESRF	European synchrotron radiation facility
FB	factor B
FD	factor D
FGF1	fibroblast growth factor 1
FGF2	fibroblast growth factor 2
FGFR1	fibroblast growth factor receptor 1
FGFR2	fibroblast growth factor receptor 2
FH	factor H
FHL-1	FH-like protein 1
FHR	FH-related protein
FI	factor I
GAG	glycosaminoglycans
Gal	galactose
GlcA	glucuronic acid
GlcN	N-glucosamine
GlcNAc	N-acetylglucosamine
GlcNS	N-sulphated glucosamine
GPI	glycosylphosphatidylinositol
GPC	gel permeation chromatography

HEPES	N-2-Hydroxyethylpiperazine-N'-2-ethanesulfonic acid
HS	heparan sulphate
IdoA	iduronic acid
IgG	immunoglobulin G
IgM	immunoglobulin M
kDa	kilodalton
L	length
LINAC	linear accelerator
M	maximum
MAC	membrane attack complex
MASP-1	MBL-associated serine protease 1
MASP-2	MBL-associated serine protease 2
MASP-3	MBL-associated serine protease 3
MBL	mannan binding lectin
MCP	membrane cofactor protein
MPGN	membranoproliferative glomerulonephritis
NMR	nuclear magnetic resonance
GlcNS	N-sulphated glucosamine
PDB	protein data bank
R_G	radius of gyration
r.m.s.d.	root mean square deviation
RPE	retinal pigment epithelium
rpm	revolutions per minute
R_{XS}	cross-sectional radius of gyration
SAS	small angle scattering
SAXS	small angle X-ray scattering
SCR	short complement regulator/ short consensus repeat
SDS	sodium dodecyl sulphate
PAGE	polyacrylamide gel electrophoresis
TEM	transmission electron microscope
UDP	uridine diphosphate
UV	ultraviolet
Xyl	xylose

Amino Acid Abbreviations

Amino Acid	Three Letter Code	One Letter Code
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic Acid	Asp	D
Cysteine	Cys	C
Glutamine	Gln	Q
Glutamic Acid	Glu	E
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

Chapter One

The complement system

(1.1) Complement overview

Complement is a biochemical cascade of the innate immune system comprising of a group of serum and cell surface proteins that interact with one another both to opsonise pathogens and to induce a series of inflammatory responses that help in defence against infection. Innate immunity refers to antigen-nonspecific defence mechanisms that a host uses immediately or within several hours after exposure to an antigen. These mechanisms include physical barriers such as skin, chemicals in the blood, and immune system cells that attack foreign cells in the body. Adaptive immunity refers to antigen-specific immune response and is more complex than the innate immunity. In adaptive immunity the antigen first needs to be processed and recognized. When once an antigen has been recognized, the adaptive immune system creates an army of immune cells specifically designed to attack that antigen. Complement has three physiological activities: defending against pyogenic bacterial infection, bridging innate and adaptive immunity, and disposing of immune complexes and the products of inflammatory injury (Walport, 2001). This system can be activated via immunoglobulins in the bloodstream once the pathogens have been recognized by antibody. In the 1890s, complement was described as a heat-labile bactericidal activity in serum which was triggered after invading microorganisms. In the 1920s, advances were made when it was found that heat-labile bactericidal activity required the presence of four serum protein fractions. At that time, due to lack of protein purification methods such as gel filtration, ion exchange chromatography, affinity chromatography, etc, very little was known about the chemical characterization of the numerous components of the complement system until the 1960s. Today more than 35 soluble and cell bound proteins have been shown to be involved in the complement system. The genetics of these proteins are becoming well understood. Most of these complement proteins circulate in the form as zymogens, the inactive or nearly inactive precursor form of the enzyme that become activated after proteolytic cleavage.

(1.2) Complement activation

The complement system protects against pathogens in three different ways:
(1). Complement system produces a large number of complement activated proteins that bind to pathogens covalently, opsonising them for their engulfment by phagocytes.

(2). The small cleavage fragments of complement proteins activate and recruit more phagocytes to the site of complement activation by acting as chemoattractants.
(3). The final components in the complement pathway forms the membrane attack complex (MAC) that destroy and eliminate pathogens by creating holes in their cell membrane.

Three pathways, namely the classical, the alternative and the lectin pathways of complement activation, trigger a sequence of biological reactions (Atkinson & Frank, 1980; Muller-Eberhard, 1988; Morgan & Harris, 1999).

(1.2.1) Classical pathway

The classical pathway of complement system was initially discovered to be the route for initiation of the proteolytic cascade, which leads to the destruction and elimination of invading pathogens. The multimolecular enzyme complex, C1, consists of a recognition subunit C1q and two associated serine proteases C1r and C1s (Arlaud *et al.,* 1987). C1q consists of six globular head domains and six collagen-like arms. The C1 complex has a well defined stoichiometry: C1q is associated with a C1s-C1r-C1r-C1s tetramer in a Ca^{+2} dependent manner (Arlaud *et al.,* 1987). The C1q molecule, which contains no enzymatic activity, initiates the activation of the classical pathway either by the binding of the globular heads of the hexameric molecule C1q to the Fc region of antibodies complexed with antigens or by direct interaction of C1q with a variety of polyanions (Figure 1.1). The physiological implication of the latter activation is not yet clear. This binding causes conformational changes in C1q, which leads to activation of the dimer of C1r, which in turn activates the coassociated serine protease C1s, and therefore enzymatic activity of C1 is derived (Law & Reid, 1995). Activated C1s cleaves the three chain C4 protein at a single point in its α-chain to yield a small fragment, C4a and a large fragment C4b. C4b binds to the target antigens through the amide bond or ester bond. C4b bound to antigen in the presence of Mg^{++} is capable of binding with the next complement component C2. Activated C1s also cleaves the single chain molecule, C2 into a larger fragment, C2a and a smaller fragment C2b. C4b binds to C2a to form the C4b2a complex (the C3-convertase). The C2b fragment is not required for the C3-convertase activity which is mediated via the C-terminal catalytic site in C2a in C4b2a. The C3-convertase cleaves native C3; a two disulphide-linked chain

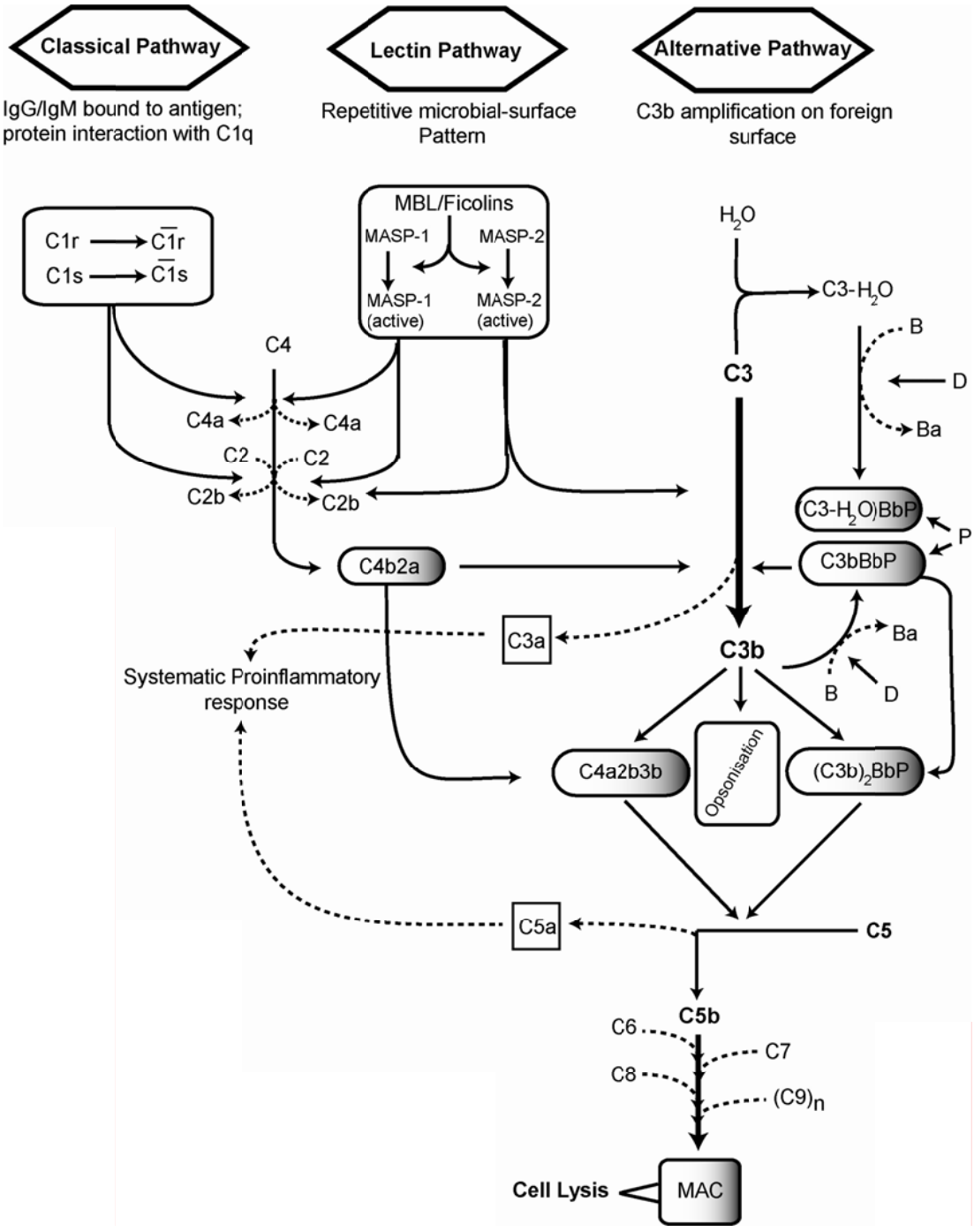


Figure 1.1

Schematic diagram of the classical, alternative, and lectin complement activation pathways. These pathways converge into a final common pathway when C3-convertase (C3 con) cleaves C3 into C3a and C3b. MASP = mannose-binding lectin-associated serine protease; MBL = mannose binding lectin; P = properdin; MAC = membrane attack complex.

complement protein, into the small fragment C3a and a large fragment C3b. The classical C5-convertase is formed from the C3-convertase by the addition of freshly activated C3b, which binds covalently via its reactive thioester group to a serine residue at position 1217 of the surface-bound C4b of C4b2a to form a C4b3b2a complex called the C5-convertase (Figure 1.2). Both C4b and C3b act as co-factor for the catalytic component C2a within the C5-convertase (Law & Reid, 1995; Rother *et al.*, 1998). It is important to mention that C4b binds to the cell surface via an amide bond. The C5-convertase cleaves C5 into C5b and C5a which then lead into the membrane attack complex (MAC) after several enzyme cascade steps.

(1.2.2) Lectin pathway

The lectin pathway resembles the classical pathway in several aspects (Figure 1.1). Lectins are sugar binding proteins, and are able to bind to a wide range of microorganisms. The initiation of this pathway is based on the recognition of certain oligosaccharide moieties on the surface of pathogen cells via macromolecular complexes in the body fluids (Matsushita *et al.*, 2000). To date, two recognition components of the lectin activation pathway have been described, i.e. mannan binding lectin (MBL) and ficolin. These components have different carbohydrate binding specificities. This pathway is activated by the association of MBL and ficolin with specific serine proteases termed MBL-associated serine protease-1 (MASP-1) MASP-2 and MASP-3 (Matsushita & Fujita, 1992; Fujita *et al.*, 2004). It is important to note that MASP-1 and MASP-2 associate with MBL in an approximate 3:1 ratio (MASP-1: 65-75%; MASP-2: 25-30%), while MASP-3 is present only in trace amounts (Hajela *et al.*, 2002). Like C1s in the classical pathway, MASP-2 also cleaves C4 and C2 into C4a/C4b and C2a/C2b respectively, which then form a C4b2a complex termed the C3-convertase (Kerr, 1980). Therefore, activated MASP-2 is considered to be the main mediator of complement activation. MASP-1 also cleaves C2 as efficiently as MASP-2 does, but it does not cleave C4 (Chen & Wallis, 2004). In addition, MASP-1 also cleaves C3 at a very slow rate compared to the C3-convertase and it has been hypothesized that this mechanism provides an initial source for amplification of surface deposited C3b by alternative pathway (Petersen *et al.*, 2001). The role of MASP-3 in the lectin pathway is not clear (Cortesio & Jiang, 2006). The formation of the C3-convertase from the cleavage of C4 by MASP-2 and the cleavage of C2 by both MASP-1 and MASP-2 are the essential steps by which

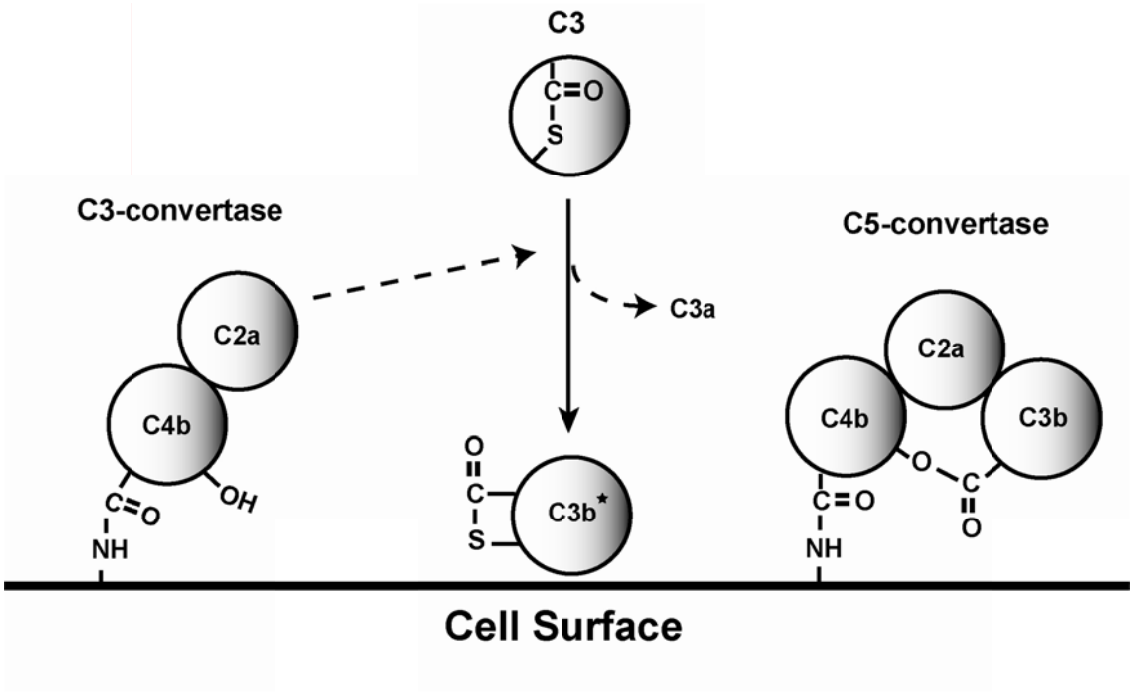


Figure 1.2

The classical pathway C5-convertase formation. Activated C3b* binds to a serine residue at position 1217 of surface-bound C4b to form a C4b-C3b covalent complex which serves as the co-factor for the catalytic component C2a in the C4b2a3b (C5-convertase). C4b binds to the cell surface via an amide bond.

the lectin as well as the classical pathway triggers the further downstream activation of complement, yielding the final assembly of the bactericidal membrane attack complex (MAC).

(1.2.3) Alternative pathway

The activation mechanism of the alternative pathway differs from those of the classical in that it operates in the absence of antibodies and hence can be considered a component of the innate immune system (Figure 1.1). C3 is the key molecule in the alternative pathway, but three other proteins namely factor B (FB), factor D (FD) and properdin, are also required. Activation of this pathway is initiated by the spontaneous hydrolysis of C3 in the fluid phase. Under physiological conditions, C3 is spontaneously hydrolyzed in plasma to C3-H_2O at a low rate leading to the covalent deposition of a small amount of molecules onto practically all surfaces in contact with plasma (Pangburn *et al.,* 1981). C3-H_2O [which is functionally similar to C3b] initiates the alternative pathway by binding to FB, a single-chain 93 kDa plasma protein. Binding of the metastable complex of C3-H_2O with FB renders it susceptible to cleavage by FD, a 26 kDa plasma serine protease. FD cleaves FB into two fragments Ba and Bb. Ba is released from the complex with C3-H_2O while the Bb remained attached and forms an active enzyme of the alternative pathway, C3bBb, called the C3-convertase (Figure 1.1). This complex has a relatively short half life, but is stabilized by the plasma protein properdin, an oligomeric glycoprotein with a monomer mass of 53 kDa (Janeway *et al.,* 2005; Thurman & Holers, 2006). Properdin increase the lifetime of the C3bBb complex three to fourfold. C3bBb cleaves more C3 to C3b and this leads to the amplification of the cascade and the efficient deposition of C3b onto target cells which binds to FB in the presence of Mg^{++} and results in the formation of more C3-convertase (Figure 1.3). Some of the C3b molecules bind covalently to C4b and C3b in the C3-convertase enzymes of the classical and alternative pathways to form the C5-convertases, C4b2a3b and C3b_2Bb respectively (Figure 1.2). The C5-convertase then cleaves C5 into the anaphylatoxin C5a and the larger fragment C5b. This cleavage is the only proteolytic event that results in the formation of MAC on the invading pathogen membrane (Walport, 2001).

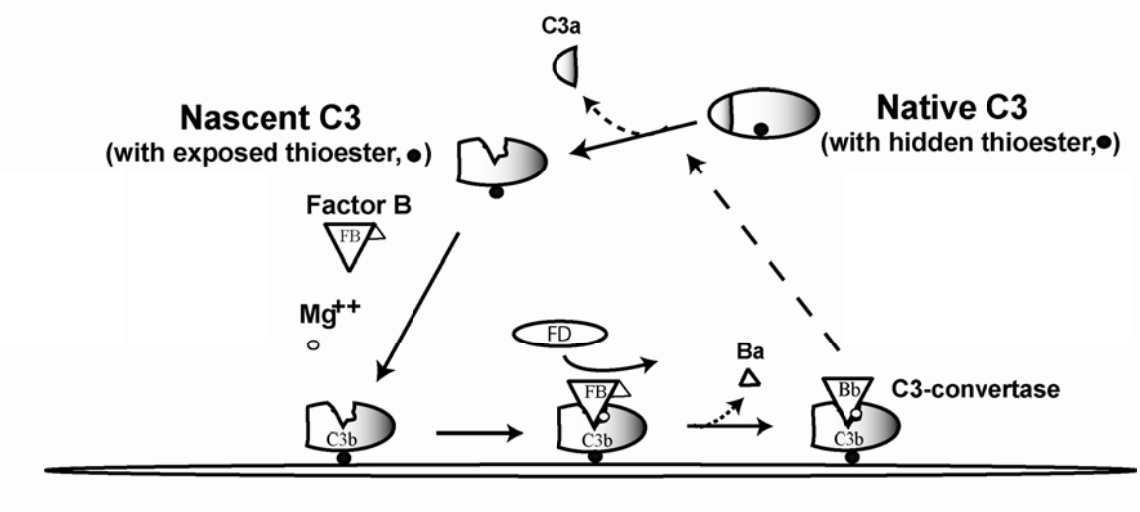


Figure 1.3

The alternative complement pathway: C3-convertase formation.

(1.3) Formation of membrane attack complex (MAC)

Although diverse mechanisms can activate complement, each activation pathway culminates in the formation of C5b, the first component of the membrane attack pathway. Once formed, C5b binds to C6 to produce a stable and soluble C5b-6 complex to initiate formation of the membrane attack complex (Figure 1.4). The C5b-6 complex binds to C7 to form C5b-7 which has a transient binding site for membrane surfaces and becomes dissociated from C3b. This C5b-7 complex, being relatively hydrophobic, attaches to the surface of cell membranes without rupturing membrane integrity. C8 then binds to a specific recognition site of C5b in the C5b-7 complex, which becomes more deeply incorporated in the membrane and causes the cell to become slightly leaky (Ramm *et al.*,1982). Once C8 is bound, then it directs the incorporation of C9 to form C5b-9. C5b-9 has two functions:

(i) It slowly lyses the red blood cells as well as certain nucleated cells.

(ii) It acts as a receptor for C9 and behaves as a catalyst in C9 polymerization to yield the C5b-9_n complex. The exact mechanism is not known but it is believed that the binding of C9 molecules to C5b8 transforms the C9 molecule from a globular, hydrophilic structure to an elongated, amphipathic structure, which inserts into and through the membrane. These conformational changes in C9 expose binding sites for additional C9 molecules to bind, unfold and insert into the target membrane (Xiong, 2003). The addition of as many as 18 copies of C9 to the C5b8 complex forms a barrel-like structure called the membrane attack complex (MAC). The stability of the MAC (C5b-8 $(9)_n$), where n lies between 1 and 18, results from the association of the hydrophic exterior of the MAC with membrane lipids. The hydrophilic interior of the MAC causes changes in the permeability of the target cell membrane by allowing loss of water and small molecules that result in the elimination of its osmotic and chemical balance and ultimately leads to osmotic lysis of target cells. It has been deduced that MAC-mediated cell injury is principally caused by the influx of calcium through the MAC channels (Campbell *et al.*, 1981).

(1.4) Difference between three pathways

Although all the three pathways of activation result in the formation of an unstable protease complex named the C3-convertase, their routes of activation are different from each other, and also different types of plasma proteins/proteins complexes are involved in their activation. Activation of classical and lectin

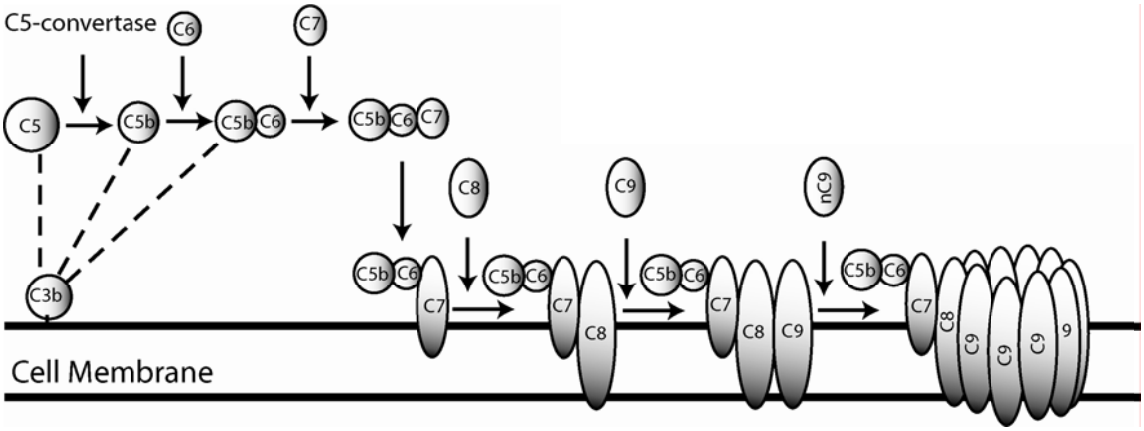


Figure 1.4

Schematic diagram of the membrane attack complex (MAC). C5b and C5b-6 are loosely bound to membrane bound C3b.

pathways requires binding of antibody or the presence of a pattern recognition molecule to the target respectively. By contrast, the alternative pathway is initiated spontaneously in plasma at low level and if left uncontrolled leads to attack against all the particles, membranes and cells which are plasma exposed but not specifically protected from it (Fearon *et al.*, 1977). The alternative pathway requires Mg^{++} ions for the assembly of the C3bBb complex, whereas the classical pathway requires both Mg^{++} ions for the assembly of C4b2a and Ca^{++} ions for the C1 complex. The alternative pathway directly activates C3 when it interacts with certain activating surfaces (e.g. zymosan, lipopolysaccharides) and this involves C3, factor B, factor D and properdin (Reid, 1986; Morgan & Harris, 1999). So, the alternative pathway provides a means of non-specific resistance against infection without the participation of antibodies (Table 1.1) and hence provides a first line of defence against a number of infectious agents. Although the lectin pathway is homologous to the classical pathway, this utilises MBL instead of C1q which interacts with repeated pattern of mannose and N-acetylglucosamine residues present in abundance on bacterial cell walls (Thiel *et al.*, 1977).

(1.5) Regulation of complement

Uncontrolled activation of the complement cascade can inflict serious damage on host tissues and therefore must be tightly regulated. Regulation of complement system occurs at several levels of the cascade and is mediated by both soluble proteins in fluid phase and membrane bound proteins.

(1.5.1) Regulation of classical pathway

The activity of C1 complex is tightly regulated by a regulatory protein known as C1 inhibitor (C1INH) which is present at a higher concentration than C1 (Davis *et al.*, 2008). C1INH is a member of the serpin family of protease inhibitors and is bound to most of the circulating C1 complex, thereby inactivating the C1r and C1s proteases (Davis *et al.*, 2008). When the C1 complex binds to antigen-antibody complex, the C1INH is released from the complex and results in the activation of classical pathway.

Table 1.1. Initiators of all three activation pathways

Pathway	Initiators
Classical	Immune complexes
	Apoptotic cells, certain viruses
	Gram-negative bacteria,
	C-reactive protein bound to ligand
Lectin	Microbes with terminal mannose groups
Alternative	Many bacteria, fungi, viruses, tumor cells

(1.5.2) Regulation of C3-convertase

The C3-convertase is a proteolytic enzyme complex that drives an amplification cascade reaction at the heart of the complement system. The C3-convertase is the main key to the activation and regulation of the complement system. The classical pathway C3-convertase consists of C4b plus C2a (C4b2a), while the alternative pathway C3-convertase consists of C3b plus Bb. Five human proteins have been identified that act to shorten the life span of the convertase enzyme or inhibit its assembly. The best well characterized examples in fluid phase are Factor H (FH) which acts as a cofactor in factor I (FI) mediated proteolysis of C3b into C3bi and decay accelerating activity of the alternative C3-convertase (C3bBb) (Jozsi & Zipfel, 2008) and C4b binding protein that dissociates the classical pathway C3-convertase (C4b2a) (Zipfel & Skerka, 2009). The membrane associating regulatory proteins are complement receptor type 1 (CR1 also known as CD35) (Khera & Das, 2009), CD46 (also known as membrane cofactor protein, MCP) and CD55 (also known as decay accelerating factor, DAF) (Spendlove *et al.*, 2006). CD35 and CD55 displace a component of the C3-convertase in the classical pathway, while CD46 acts as a cofactor in proteolytic mediated dissociation of both the classical and alternative C3-convertases (Seya & Atkinson, 1989).

(1.5.3) Regulation of membrane attack complex

Complement FH related protein 1 (CFHR1), clusterin, S-proteins (also known as vitronectin) and CD59 are the main regulatory proteins of the membrane attack complex in the terminal pathway. S-proteins, clusterin and CFHR1 are the fluid phase inhibitory proteins of the terminal pathway (Preissner & Seiffert, 1998; Schwarz *et al.*, 2008; Heinen *et al.*, 2009), while CD59 is a membrane bound regulator protein that prevents the final assembly of the terminal pathway (Kimberley *et al.*, 2007).

Chapter Two

Complement Factor H and glycosaminoglycans

(2.1) Complement Factor H (FH)

(2.1.1) Structure of FH

FH is a flexible single chain polypeptide glycoprotein, first identified by Nilson and Eberhard in 1965 and termed β1H globin. It is a major regulator of the alternative complement pathway and has a monomer molecular mass of 150-155 kDa consisting of approximately 9.3% carbohydrate which is distributed in the form of eight *N*-linked oligosaccharides at nine putative sites (Whaley & Ruddy, 1976; Sim & DiScipio, 1982; Perkins *et al*., 1991; Fenaille *et al*., 2007). The most recent sequencing by mass spectrometry showed that there are eight *N*-linked oligosaccharides (Fenaille *et al*., 2007). FH is composed of 20 domains, each consisting of approximately 60 amino acids and termed short complement regulators (SCRs) or complement control protein modules (Ripoche *et al*., 1988). These SCRs are joined by short linkers, each consisting of 3-8 amino acids residues. The SCRs contain five conserved residues including four cysteines and one tryptophan, and several partially conserved proline, glycine and other hydrophobic residues.

Intact FH has not been crystallized to date due to its large size, glycosylation, interdomain flexibility and multiple ligand binding sites. Low resolution structural studies of the intact FH have increased the understanding of its overall architecture. The first structural studies by scattering and electron microscopy showed that FH is not fully extended but is partial folded with an average length of 49.5 nm (Figure 2.1) (Perkins *et al*., 1991; Discipio, 1992). More detailed constrained X-ray solution scattering and analytical ultracentrifugation methods showed that FH is not fully extended in solution and had a maximum length of 40 nm compared to a theoretical length of 73 nm for a full extended FH molecule (Figure 2.2a) (Aslam & Perkins, 2001). These data strongly suggest that some domains within FH are folded back upon themselves in solution and this is supported by later studies using the same techniques (Figure 2.2b) (Okemefuna, *et al*., 2009). The three dimensional structure of several SCRs have been determined showing a globular structure with up to eight stranded antiparallel β sheets connected with loops and turns (Perkins *et al*., 1988; Barlow *et al*., 1992) . To date, NMR structures for SCR-5, SCR-15/16 and SCR-1/3 (Barlow *et al*., 1992; Barlow *et al*., 1993; Hocking *et al*., 2008) and crystal structures for SCR-1/4, SCR-6/8 and SCR-19/20 have been reported (Jokiranta *et al*., 2006; Prosser *et al*., 2007). For the remaining SCR domains in FH, homology models have been reported (Aslam & Perkins, 2001; Saunders *et al*., 2006).

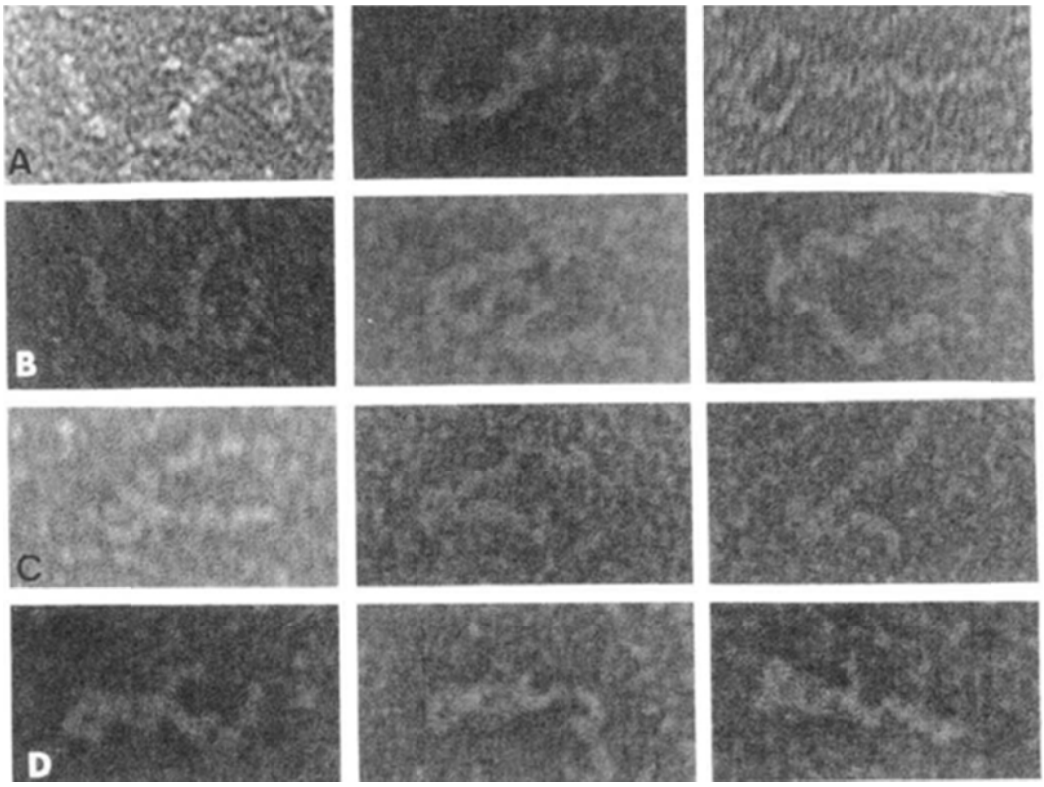
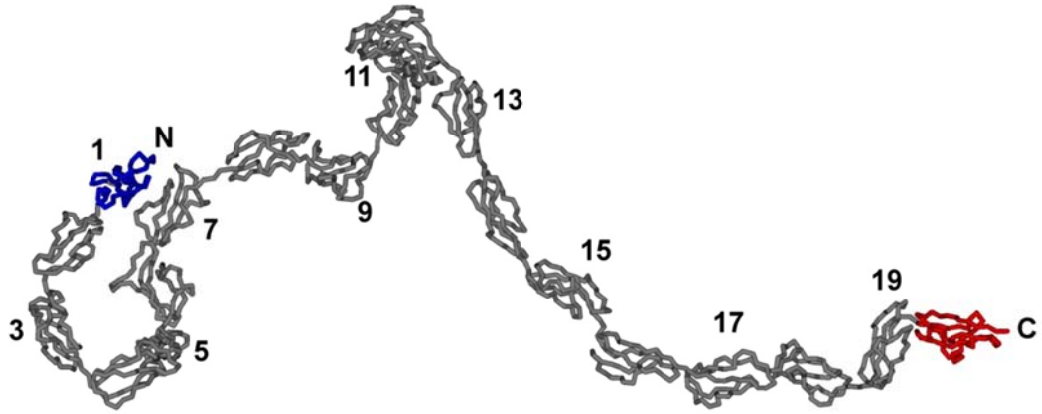


Figure 2.1

Electron microscopic images of FH showing different conformational forms. A. squiggles: B. loops: C. wishbones: D. twists (reproduced from DiScipio, 1992).

(a)



(b)

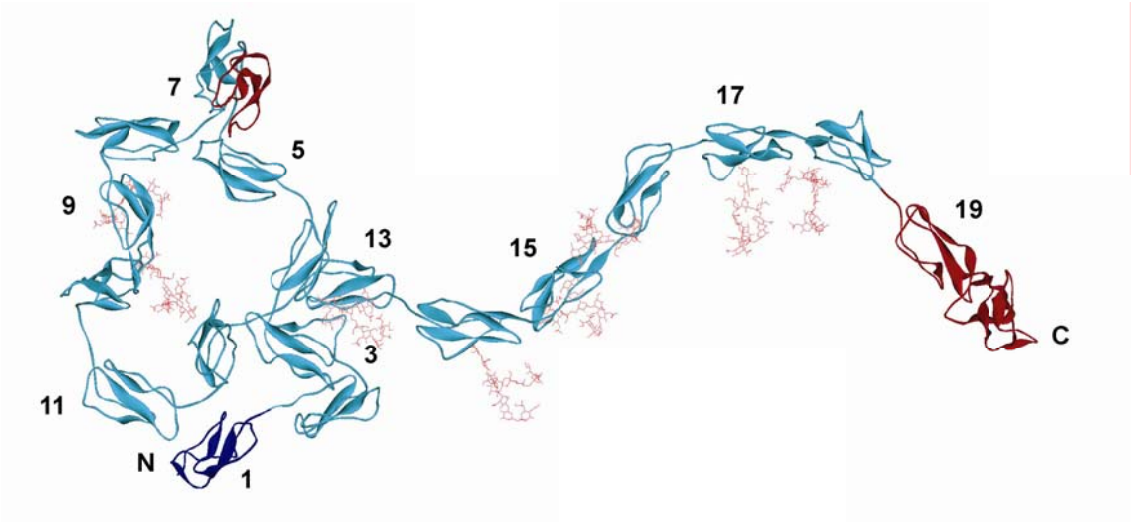


Figure 2.2

Folded back structures of FH. (a) The best fitting solution structure of native FH (PDB code: 1HAQ) by constrained solution scattering modelling (reproduced from Aslam & Perkins, 2001). (b) The updated best fit model for native FH (PDB code: 3GAV) by constrained solution scattering modelling (reproduced from Okemefuna *et al*., 2009).

(2.1.2) FH binding sites

Despite the simplicity of its domain organization, FH is functionally complex. Early studies identified that the *N*-terminus of FH possess fluid phase C3b binding property, cofactor and decay accelerating activities (Alsenz *et al*., 1984, 1985; Misasi *et al*., 1989; Kuhn & Zipfel, 1996). Under normal physiological condition, FH binds to host cell surfaces and various ligands to carry out its regulatory roles. These binding sites have been identified and characterized using different strategies based on monoclonal antibodies, enzyme digestions and deletion or point mutations. FH may have as many as three distinct binding sites for C3b that are located in SCR-1/4, in SCR-12/14 and in SCR-19/20 (Alsenz *et al*., 1985; Gordon *et al*., 1995; Kuhn *et al*., 1995; Jokiranta *et al*., 1996; Prodinger *et al*., 1998). Similarly FH has as many as three binding sites for polyanionic molecules such as heparin, heparan sulphate (HS) and sialic acids. These polyanionic binding sites are located in SCR-7, SCR-13 and SCR-19/20 (Pangburn *et al*., 1991; Blackmore *et al*., 1996, 1998; Ram *et al*., 1998). A later study suggested that the binding site located in SCR-13 should be reassigned to SCR-9. Recently, it was suggested that FH has no binding sites for these polyanionic molecules in SCR-9 and in SCR-13 (Schmidt *et al*., 2008). Because of these polyanionic and C3b binding sites, FH is believed to recognize and prevents the alternative pathway activation on the host cell surfaces bearing sialic acid and HS (Meri & Pangburn, 1990). Apart from these C3b and polyanionic binding sites, FH has binding sites for other ligands including C-reactive protein (CRP) (Okemefuna *et al*., 2010), adrenomedullin (Pio *et al*., 2001), and bacterial cell surface proteins and cell surface receptors (Horstmann *et al*., 1988; Hellwage *et al*., 1997; DiScipio *et al*., 1998). CRP binds to FH at SCR-7 that result in the inhibition of CRP-dependent alternative pathway activation induced by damaged tissues (Kaplan & Volanakis, 1974). A map of FH binding sites for these ligands is shown in Figure 2.3. Because of their potential relevance in pathology, these FH binding sites are considered to be of great interest. Some of these binding sites overlap with each other. The polyanionic binding sites overlap with C3b in SCR-19/20 (Hellwage *et al*., 2002) and with CRP in SCR-7 (Giannakis *et al*., 2003). Structural studies have shown that FH interacts with C3b, heparin and CRP through the cluster of positively charged amino acids residues located in SCR-7 and SCR-19/20 (Perkins & Goodship, 2002; Giannakis *et al*., 2003).

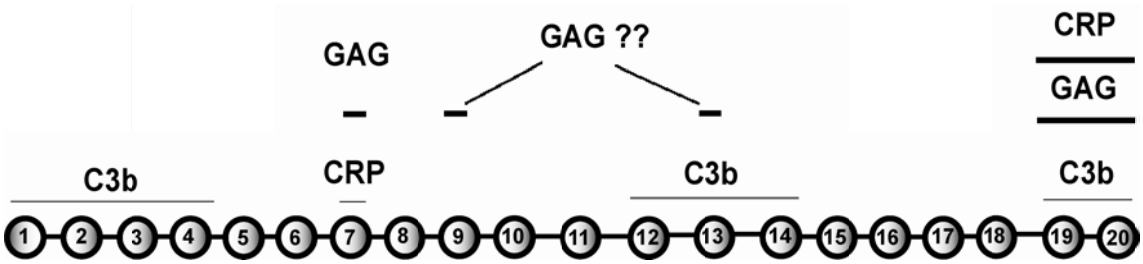


Figure 2.3

Functional domains in FH and putative binding sites. The linkers are proportional to their lengths (3-8 residues). Question marks are used where no direct proof for ligand binding is available.

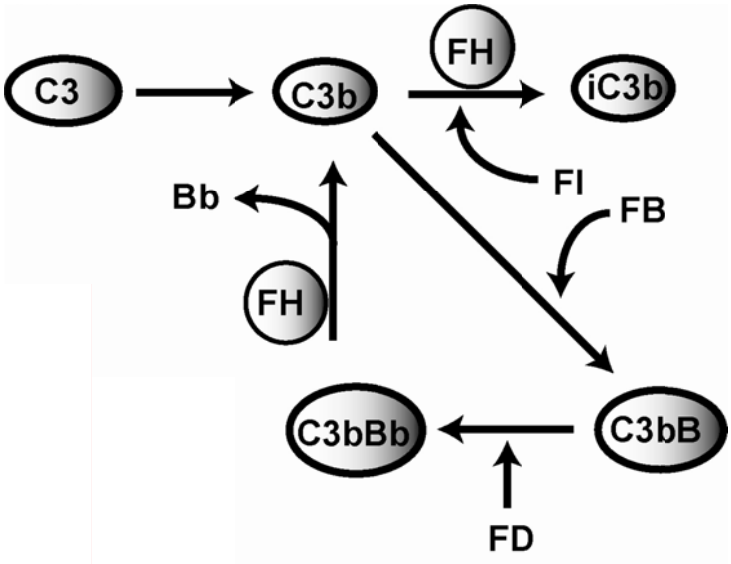
(2.1.3) Where does FH appear?

The activation and damage potential of the complement system requires that it be kept under tight control. At least 12 proteins including FH are known that do this function. FH is produced by liver cells and is found in human plasma at concentrations of 0.235-0.810 mg/ml. Extrahepatic synthesis of FH also occur in a wide variety of cell types, such as peripheral blood lymphocytes, myoblasts, fibroblast, neurons, glomerular mesangial cells, neurons and glia cells (Friese *et al*., 1999). The extrahepatic production of FH is interpreted as a mechanism to increase the local concentration of this complement regulator to protect the host cells from complement attack at sites of infection or inflammation. Recently FH has been also shown to be present in retinal blood vessels in the choroid (Klein *et al*., 2005). In addition to FH, FH like protein 1 (FHL-1), formed by an alternative transcript of the same gene that codes for FH, is also produced by the liver and is present in human plasma at concentration of 10-50 µg/ml.

(2.1.4) What does FH bind to?

FH acts selectively on C3b and C3bBb (C3-convertase) on self versus non-self surfaces (Meri & Pangburn, 1990). It binds to C3b and accelerates the decay of the alternative pathway C3-convertase (C3bBb) by irreversible dissociation of Bb from C3bBb (Figure 2.4a). FH competes for binding of complement factor B (FB) to C3b (Farries *et al*., 1990). It also acts as a cofactor for factor I (FI), a serine protease, which cleaves C3b into iC3b, an inactive form of C3b, which can no longer take part in formation of the C3-convertase (Weiler *et al*., 1976; Whaley & Ruddy, 1976; Pangburn *et al*., 1977). In the presence of FH and FI, C3b proteolysis results in the cleavage of the α-chain of C3b at two sites that generates two fragments of 68 kDa and 43 kDa respectively (Figure 2.4b). FH also binds to polyanionic molecules such as sialic acid and HS on host cell surfaces. Due to this polyanionic binding property, FH is believed to distinguish between self and non-self surfaces. It has been reported that sialic acids influence the regulation of surface bound C3b by FH (Fearon, 1978). Sialic acids induce an increase in the affinity of FH for C3b. Since FH does not bind to the cells bearing sialic acids in the absence of C3b, it is likely to be a consequence of the simultaneous recognition of both sialic acid and bound C3b by the same FH molecule. The presence of multiple binding sites for both C3b and polyanions in FH strongly support this concept. It has been also deduced that FH also binds to CRP

(a)



(b)

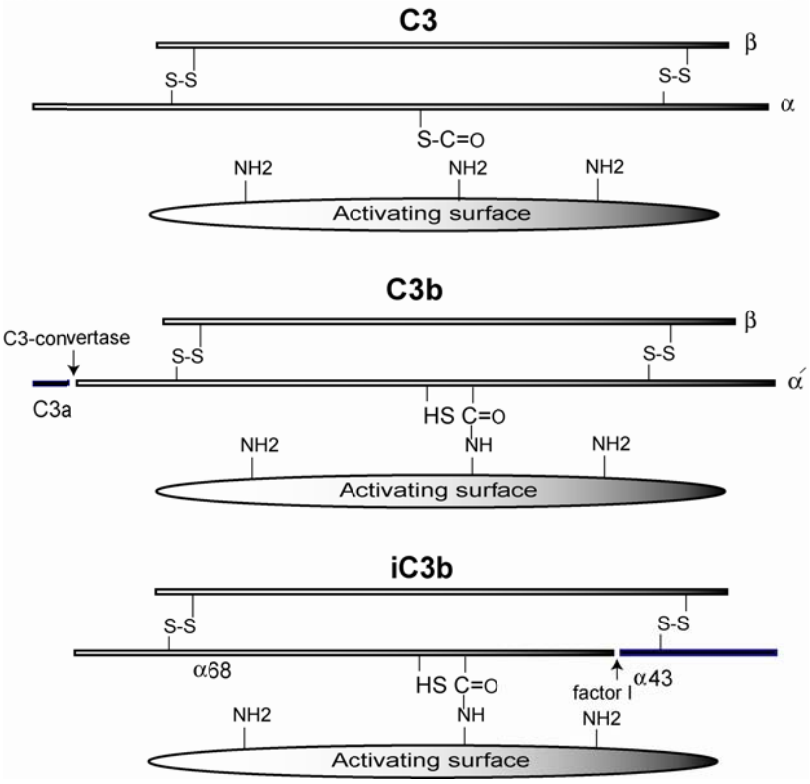


Figure 2.4

(a) Inactivation of C3b by FH and acceleration of C3-convertase (C3bBb) dissociation. (b) Cleavage of C3b by FI in the presence of FH to form iC3b.

that inhibits CRP dependent activation pathway induced by damaged tissues (Kaplan & Volanakis, 1974). FH has binding sites for some plasma proteins and microorganisms that are of great interest because of their potential relevance in pathology.

(2.1.5) What does FH do?

FH is a key regulatory protein of the complement system of innate immunity (Rodriguez de Cordoba *et al*., 2004). The most important and clearly identified function of FH is the regulation of enzymatic processing of the major complement component C3. In human serum, C3 is spontaneously converted into C3-H2O by hydrolysis. FB binds to C3-H2O and this leads to the formation of C3bBb (C3-convertase) after removing a small part, Ba, from FB by FD and C3a from C3. An immune response occurs when C3b is attached to the pathogen surface. The C3-convertase cleaves C3 to form more C3b that leads to the formation of the membrane attack complex which make a hole in the pathogen cell. On the host cell surfaces bearing sialic acid and heparan sulphate, FH competes for the binding of FB to C3b and hence prevents the formation of the C3-convertase. It also dissociates the C3bBb complex (the alternative pathway C3-convertase) by binding to the C3b subunit and displacing Bb. The interaction of FH with cell surface bound C3b decreases with removal of these polyanionic molecules such as on pathogenic surfaces. In fluid phase FH binds to and inactivates C3b in the presence of FI, a serine protease that cleaves C3b into C3bi, an inactive form of C3b (Rodriguez de Cordoba *et al*., 2004). Thus FH protects the host cells bearing polyanionic molecules such as sialic acid and HS from complement attack (Kazatchkine *et al*., 1997; Carreno, 1989).

FH has been reported to have other functions in addition to decay acceleration and co-factor activity. Many of these are probably associated with a FH receptor which was initially reported on leucocytes (Lambris & Ross, 1982). These are specific, saturable and divalent cation dependent. The function which are thought to be associated with FH receptor include the release of FI from β-lymphocytes (Erdei & Sim, 1987) and the release of prostaglandin E and thromboxane from macrophages.

(2.1.6) Role of FH in diseases

FH is encoded by a single gene (HF1) located on human chromosome 1q32. Genetic mutations in HF1 gene have been associated with two different renal diseases, membranoproliferative glomerulonephritus (MPGN) and atypical haemolytic uraemic syndrome (aHUS), and as shown recently age-related macular degeneration (AMD).

(2.1.6.1) Membranoproliferative glomerulonephritus (MPGN)

Lack of FH in plasma is a cause of MPGN. FH deficiency results in a progressive glomerulonephritus in the pig, similar to human type II membranoproliferative glomerulonephritus that leads to renal failure. However, similar to the deficient pigs, the knockout mice develop MPGN spontaneously and are hypersensitive for developing renal injury. Uncontrolled C3 activation in vivo is essential for the development of MPGN and defective regulation is explained by its association with deficiency, or functional inactivity of FH (Pickering *et al*., 2002).

(2.1.6.2) Atypical haemolytic uraemic syndrome (aHUS)

Atypical Haemolytic uraemic syndrome (aHUS) is a kidney disease, which is characterized by haemolytic anaemia, thrombocytopenia, and acute renal failure. The last five year of studies on complement revealed that aHUS is strongly associated with mutations in proteins needed either for activation or regulation of alternative pathway of complement. These mutations have been described in genes encoding five complement proteins i.e FH, membrane cofactor protein/CD46 (MCP), FI, FB, and C3 (Richards *et al*., 2001, 2993; Kavanagh *et al*., 2005; Goicoechea de Jorge, 2007; Fremeaux-Bacchi, 2006). Previously, aHUS has been associated with a reduced plasma concentration of FH (Thompson *et al*., 1981). Mutations in the FH1 gene that encodes FH became the first link between complement and aHUS. The mutations cause amino acid substitutions in the C-terminus of FH mainly within domains 19-20 of FH (Caprioli *et al*., 2001; Dragon-Durey *et al*., 2004). Since aHUS is characterized by endothelial cell damage, microthrombosis, kidney failure and alternative pathway activation, it appeared that protection of these cells and kidney glomeruli from complement attack was impaired by the malfunction of the C-terminus of FH which is crucial for FH binding with host cell surfaces.

(2.1.6.3) Age-related macular degeneration (AMD)

A recent study of the human genome project revealed the role of the alternative pathway in age-related macular degeneration (AMD), an eye-sight threatening disease. AMD, the most common cause of central vision loss in the elderly population of western countries (Klein *et al*., 2004; van Leeuwen *et al*., 2003), involves the breakdown of the macula, a small yellowish spot (~ 2 mm in diameter) on the central retina which is composed of photoreceptor cells. AMD is characterized by the appearance of extracellular deposits known as drusen, a hallmark lesion of AMD, at the interface of Bruch's membrane and the retinal pigment epithelium (Anderson *et al*., 2002). Clinically, drusen are divided into two main forms, soft and hard drusen. Soft drusen is the early indication of AMD with no visual loss, while hard drusen is an indication of late AMD with severe loss of central vision. Late AMD has two forms, dry and wet. Both these forms of AMD can be found bilaterally, meaning it occurs in both eyes of the same patient (de Jong, 2006). A variety of lipids, polysaccharides and glycosaminoglycans have been reported in drusen by histochemical and immunocytochemical studies (Abdelsalam *et al*., 1999; Hageman *et al*., 2001). To date more than 129 proteins have been identified in the drusen. Some of them are vitronectin (Hageman *et al*., 1999), amyloid A, amyloid P, α1-antitrypsin, α1-antichymotrypsin, apolipoprotein A1, apolipoprotein E, remnants of retinal pigment epithelium (RPE), complement components C3, C5, and C9, complement regulators, fibrinogen, Ig kappa, Ig lambda, and ubiquitin (Mullins *et al*., 2000; Crabb *et al*., 2002). One of these regulators is FH. A common variation in the FH1 gene that encodes FH has been associated with a significantly increased risk (50%) of AMD in human (Klein *et al*., 2005; Haines *et al*., 2005; Edwards *et al*., 2005). Thus a single polymorphism in FH (Tyr402His) is strongly associated with many AMD cases, leading to heavy deposition of complement in drusen. This residue is located in SCR-7, a domain which is present in FH and factor like protein (FHL-1). The tyrosine residue at position (FH_{YY402}) is considered as a protective variant whereas the histidine residue at this position (FH_{HH402}) is a risk variant for AMD. In addition to these major roles of FH in diseases, it also plays a role in diabetes mellitus, Alzheimer disease and atherosclerosis.

(2.2) Glycosaminoglycans (GAG)

Glycosaminoglycans are linear polyanionic polysaccharides found on the cell surfaces, in mast cells and extracellular matrix. Heparin and heparan sulphate (HS) are the most widely studied examples of this family and are biosynthesized as proteoglycans.

(2.2.1) Biosynthesis of GAG

Both heparin and HS are biosynthesized as proteoglycans through similar pathways. HS proteoglycans are expressed and secreted by most mammalian cells and are exclusively found on the cell surfaces and in the extracellular matrix (Gallagher *et al*., 1992).

The biosynthesis of GAG chains occurs in the Golgi apparatus and begins with the biosynthesis of tetrasaccharide linker (glucuronic acid (GlcA)-galactose (Gal)-galactose-xylose (Xyl)) to a serine residue in the core proteins (Salmivirta *et al*., 1996). Syndecans (an integral membrane protein) and glypicans (a GPI-anchored protein) are the two major core proteins that carry HS chain on the cell surfaces. In the extracellular matrix, especially in the basement membrane, HS chains are attached mainly to core proteins called perlecan and agrin. In heparin, this tetrasaccharide linker is attached to serglycin, a core protein, through a serine residue. The second step involves the transfer of *N*-acetylglucosamine (GlcNAc) residue from UDP-GlcNAc to form an α-linkage with C4 of the terminal GlcA of the tetrasaccharide attached to the core protein (Figure 2.5). This step is catalysed by GlcNAc transferase. This step appears to be the commitment step towards the formation of a HS/heparin chain rather than to chondroitin/dermatan sulphate. The elongation of the GAG chain takes place by the addition of alternative GlcA and GlcNAc residues from their corresponding UDP-sugar nucleotides to the non-reducing termini of the growing chain. About 300 sugar residues are added to the polysaccharide chain before the chain terminates (Lindahl *et al*., 1986). As the chain elongates, it undergoes a series of modification reactions that include *N*-deacetylation and *N*-sulphation of GlcNAc by deacetylase/*N*-sulphotransferase, C-5 epimerization of GlcA to iduronic acid (IdoA) by C-5 epimerase (Lindahl *et al*., 1972), O-sulphation of IdoA at C-2 by an iduronosyl 2-O-sulphatranferase, and O-sulphation of GlcA at C-2 by gluconosyl 2-O-sulphotransferase (Bai & Esko, 1996), O-sulphation at C-6 of GlcNAc and GlcNS by glucosamine 6-O-sulphotransferase

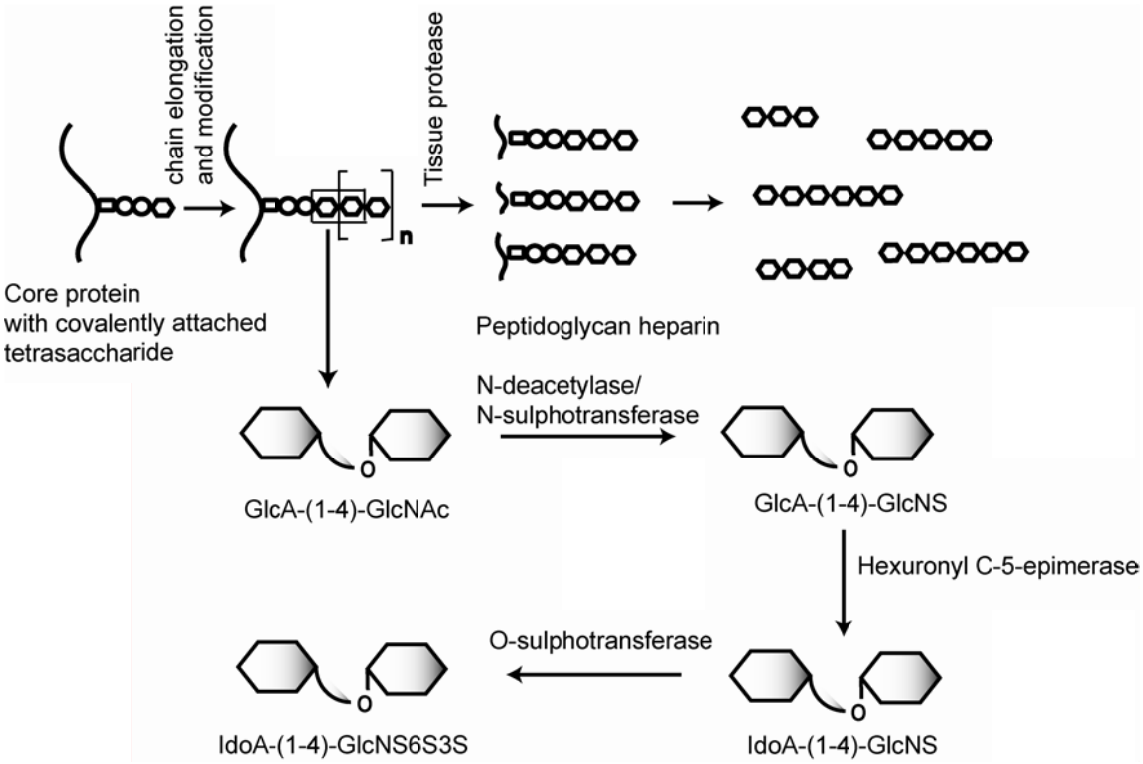


Figure 2.5

Scheme of HS/heparin proteoglycans and their degradation to peptidoglycans and glycosaminoglycans. Protein synthesis takes place in the endoplasmic reticulum, whereas linkage synthesis, chain elongation, and modification take place in the Golgi, and proteolysis and glucuronidase digestion take place in the granules or lysosomes.

(Jacobsson & Lindahl, 1980), and occasionally O-sulphation of GlcN residues at C-3 by 3-O-sulphotransferase (Rosenberg *et al*., 1997). HS/heparin possesses a high degree of structural variability which is the result of the incomplete nature of modification carried out by these biosynthetic enzymes.

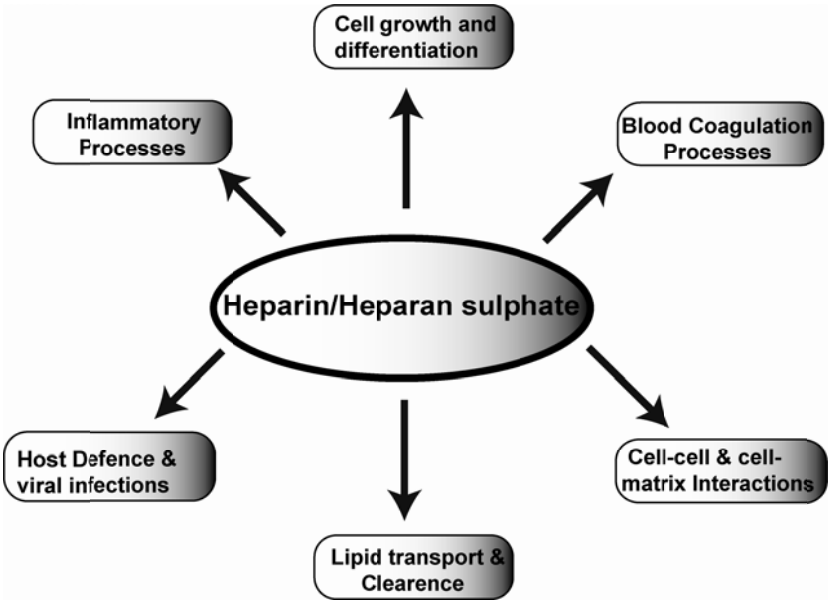
(2.2.2) Heparin

(2.2.2.1) Importance of Heparin

Heparin was discovered in 1916 and has been known for its anti-complement as well as its anti-coagulant activity (Rodén, 1989). The importance of heparin is due to its diverse biological activities. It is not only used as a natural therapeutic agent but also as analogue for HS which is present on the cell surfaces and in extracellular matrix. Direct inhibition of Clq binding to immune complexes, interaction of C1s with C4 and C2, and binding of C2 to C4b are examples of anti-complementary activity of heparin (Kazatchkine *et al*., 1981). Apart from these anti-complementary and anti-coagulant activities, heparin possesses some other biological activities such as release of lipoprotein lipase and hepatic lipase (Olivercrona & Bengtsson-Olivercrona, 1989), inhibition of angiogenesis and tumor growth (Folkman *et al*., 1983; Crum *et al*., 1985) and antiviral activity (Holondniy *et al*., 1991). These biological activities of heparin result from its interaction with proteins that play important roles in the regulation of many normal physiological processes (Figure 2.6a). The most widely studied example is its interaction with antithrombin, a serine protease inhibitor that mediates anticoagulant activity of heparin. This anticoagulant activity resides in a unique pentasaccharide sequence (Figure 2.6b) with antithrombin binding properties. Hundreds of heparin binding proteins have been identified during the past several decades. Most of this research is focused on the characterization of heparin-protein binding because heparin-protein binding serve a model for the interaction of protein with highly sulphated regions of the HS chains of cell-surface HS proteoglycans (Sasisekharan & Venkataraman, 2000).

Thus, heparin binds specifically to heparin/HS binding proteins including complement regulatory proteins, morphogens, growth factors, cytokines, chemokines and many other signalling molecules, and thereby plays an important role in fundamental biological processes (Capila & Linhart 2002). The heparin-protein interaction opens new approaches and strategies for therapeutic intervention at the cell–tissue–organ interface. At a fundamental level, the role of these complex

(a)



(b)

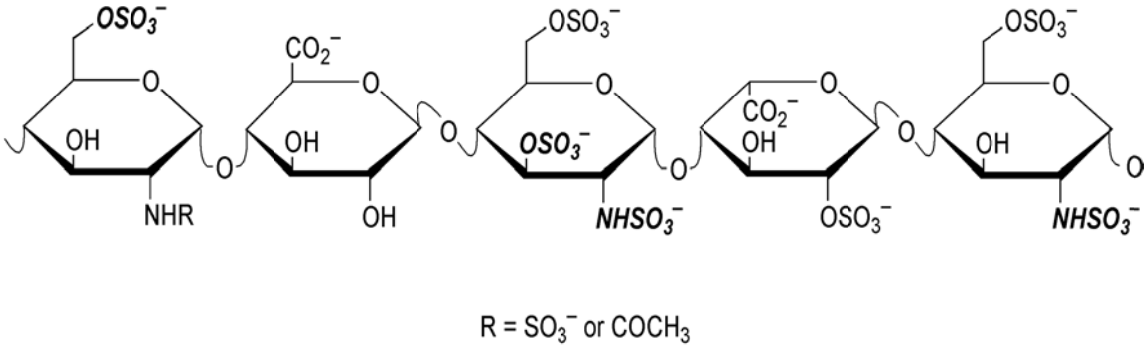


Figure 2.6

(a) Involvement of heparin and HS in various physiological processes. (b) The unique antithrombin-binding pentasaccharide sequence.

polysaccharides will assist in our understanding of biological processes in the post-genomic era. As we better understand the nature of heparin–protein interactions, there will be rapid development and expansion of the field of 'glycomics', or the study of how proteins and saccharides interact with one another and their possible role in physiological and pathophysiological processes.

(2.2.2.2) Structure of Heparin

Heparin is a linear anionic polysaccharide consisting of uronic acid (1-4)-D-glucosamine repeating disaccharide units (Comper *et al*., 1981). The uronic acid can be either α-L-iduronic acid (IdoA) which accounts for up to 90% or β-D-glucuronic acid (GlcA) which accounts for up to 10%. The repeating disaccharide unit of heparin typically contains three sulphate groups, on the 2-OH group of iduronic acid, on the NH_2 group and the 6-OH group of D-glucosamine and is considered the most common occurring repeating disaccharide (Figure 2.7). Both uronic acids may be 2-O-sulphated i.e Ido(2S), GlcA (2S). Similarly, the β-D-glucosamine (GlcN) may be either *N*-sulphated (GlcNS) or *N*-acetylated (GlcNAc), both of which may be sulphated at the 6-O or 3-O positions. Due to this high degree of sulphation and carboxylation, heparin is the most negatively charged macromolecule in nature (Capila & Linhardt, 2002). The disaccharide subunits contain a variety of groups, including *N*-sulphate, O-sulphate and *N*-acetyl groups that display extraordinary structural diversity. Heparin and HS are structurally the most complex members of the glycosaminoglycan family of anionic polysaccharides. The complete structural analysis of heparin involve identification and quantification of uronic acid-(1,4)-glucosamine disaccharide repeating units, determination of their sequence along the polysaccharide chain and characterization of the three dimensional structure of individual monosaccharide residues and polysaccharide chain. Unlike proteins, heparin is not known to display or fold into any particular tertiary structure (Capila & Linhardt, 2002). Heparin exists primarily as a helical structure where the three sulphate groups are clustered on one side of the chain, with a similar cluster forming on the other side of the chain for the next trisulphated sequence (Figure 2.8). The specificities of heparin with a wide range of biologically important proteins are due to the orientation of its sulphate and carboxyl groups in defined patterns.

Structural studies of heparin are essential for reason of the considerable importance of heparin-protein interactions and their involvement in

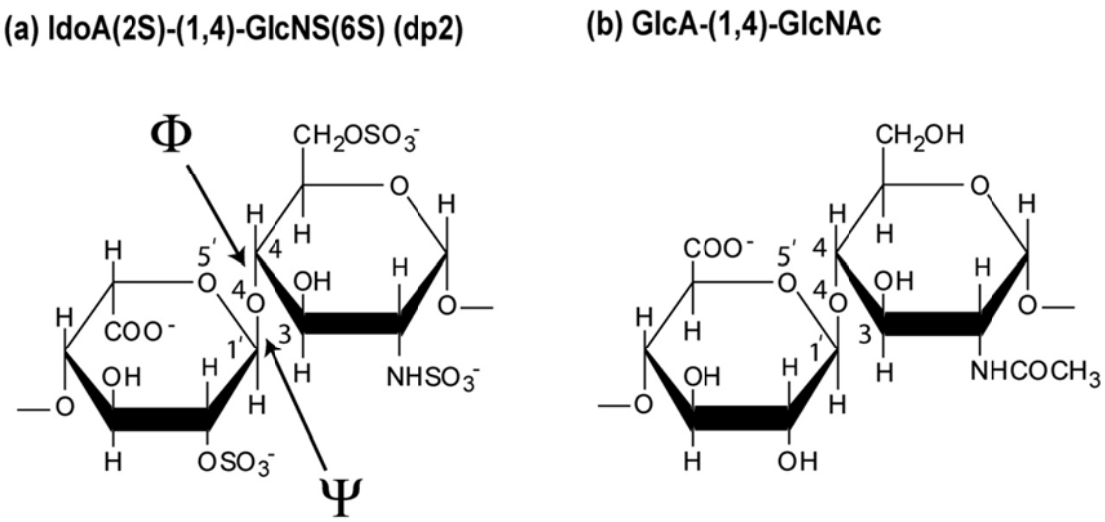


Figure 2.7

Chemical structures of the two disaccharide repeating units of heparin and HS. (a) The major repeating disaccharide unit in 90% of heparin (iduronic acid-2-sulphate -> glucosamine-2,6-disulphate). The atoms defining phi (φ) are O5′-C1′-O4-C4 atoms and for psi (Ψ) are C1′-O4-C4-C3. (b) The major disaccharide repeating unit of HS (glucuronic acid -> *N*-acetylglucosamine), which comprises 10% of heparin.

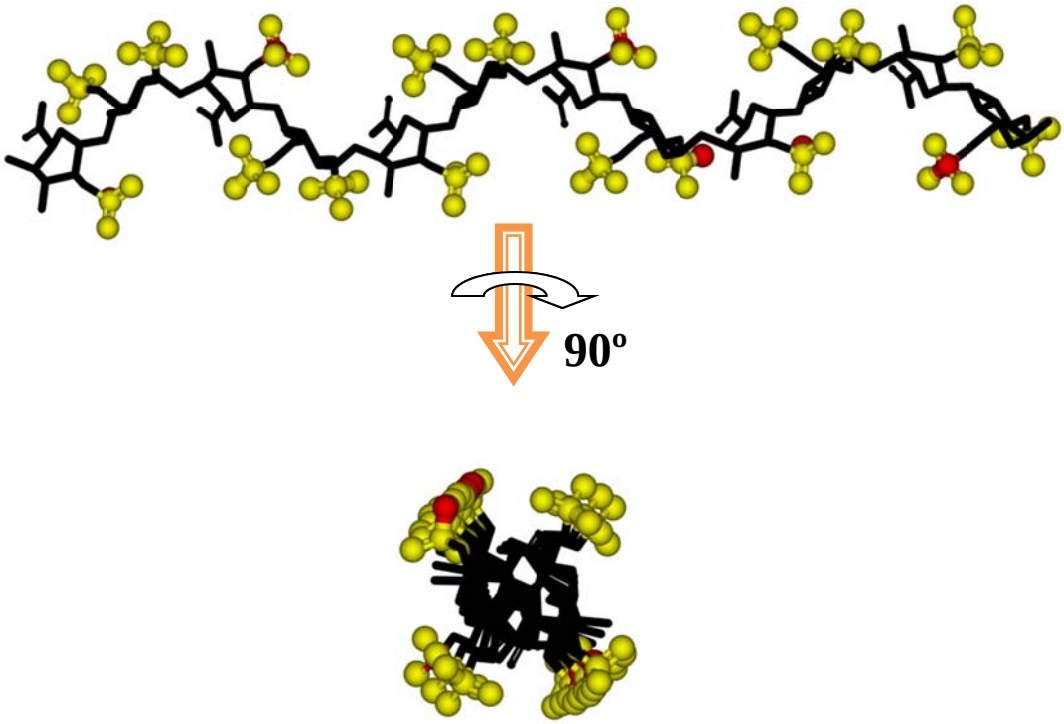


Figure 2.8

Helical structure of heparin dodecasaccharide (dp12) having three sulphate groups (yellow) clustered on one side of the chain, with a similar cluster forming on the other side of the chain for the next trisulphated sequence (reproduced from Mulloy *et al.*, 1994).

pathophysiological processes. Up to now, molecular structures for heparin have been very much limited to its small fragments. For free heparin, solution structures by NMR spectroscopy are available for heparin oligosaccharides (Mulloy *et al*., 1993) and synthetic pentasaccharides dp5 (Ragazzi *et al*., 1990). These revealed a linear structure. For heparin-protein complexes, the current availability of 19 crystal structures (July 2009) have revealed heparin structures of sizes dp4 to dp10 structures. These include crystal structures for acidic and basic fibroblast growth factors separately (Faham *et al*., 1996; DiGabriele *et al*., 1998), and the complexes of acidic and basic fibroblast growth factors with their receptors (Pellegrini *et al*., 2000; Schlessinger *et al*., 2000), thrombin (Carter *et al*., 2005), antithrombin and the complex of antithrombin–S195A and coagulation factor Xa (Jin *et al*., 1997; McCoy *et al*., 2003; Johnson *et al*., 2006; Langdown *et al*., 2009), foot and mouth disease virus (Fry *et al*., 1999), neurokinin 1 receptor (Lietha *et al*, 2001), annexins V and A2 (Capila *et al*., 2001; Shao *et al*., 2006), vaccinia complement protein (Ganesh *et al*., 2004), and protein C inhibitor (Li & Huntington, 2008). Many of these crystal structures likewise show extended heparin conformations except dp10 in the crystal structure of acidic fibroblast growth factor 1 (FGF1) with fibroblast growth factor receptor 2 (FGFR2) (PDB code: 1E0O) which revealed a kink in its structure (Figure 2.9) (Pellegrini *et al*., 2000). It has been shown that the kinked oligosaccharide provides more favourable ionic and van der Waals contacts that were calculated from the interaction energies of a kinked oligosaccharide and an oligosaccharide with a standard helical structure (Raman *et al*., 2003). Fibre diffraction studies have also been reported for heparin oligosaccharides (Nieduszynski & Atkins, 1973; Atkins & Nieduszynski, 1977). At the opposite extreme of resolution, macroscopic solution structures for large polydisperse heparin fractions have been studied by X-ray solution scattering and analytical ultracentrifugation (Pavlov *et al*., 2003; Perez Sanchez *et al*., 2006). No structural information at a molecular level was reported for these larger heparin fractions.

(2.2.3) Heparan sulphate

(2.2.3.1) Heparan sulphate structure

Structurally, heparan sulphate (HS) is closely related to heparin but with a reduced degree of sulphation than heparin. HS is also a linear anionic polysaccharide

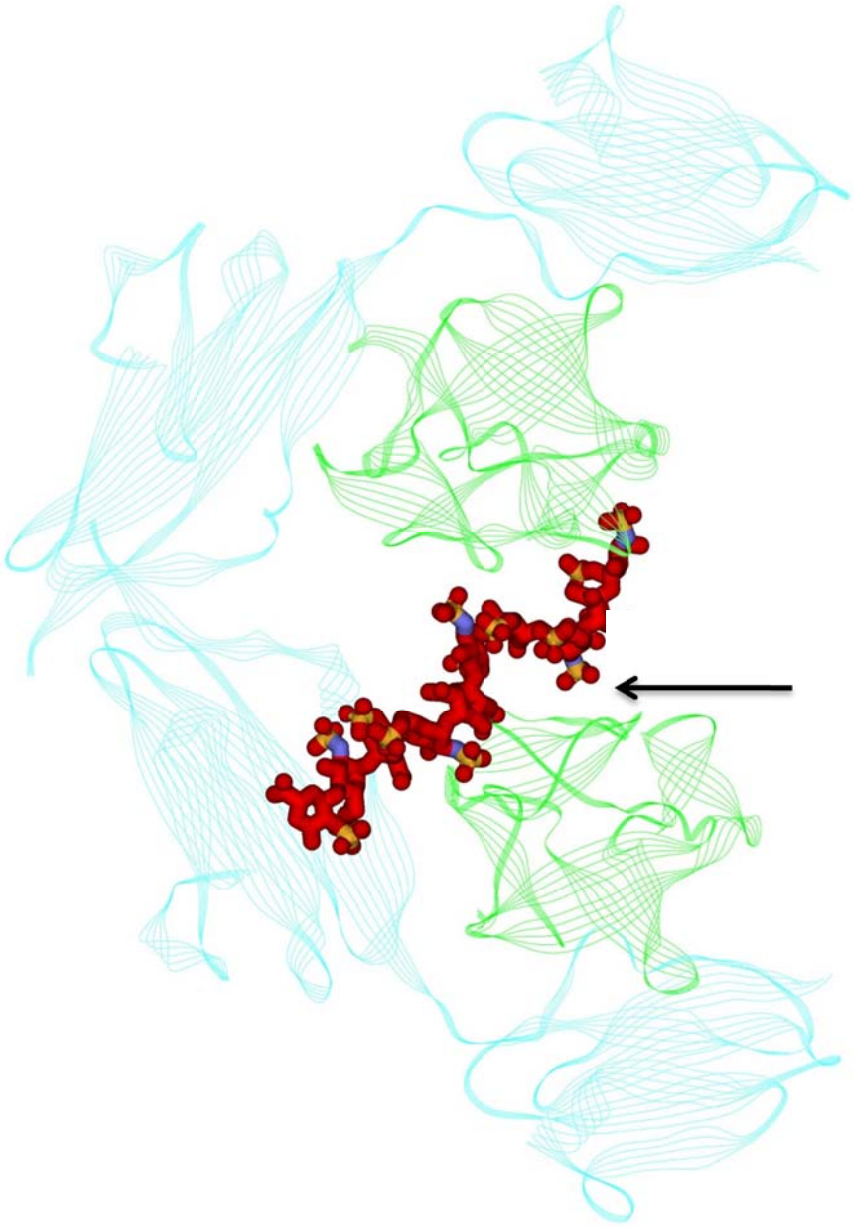
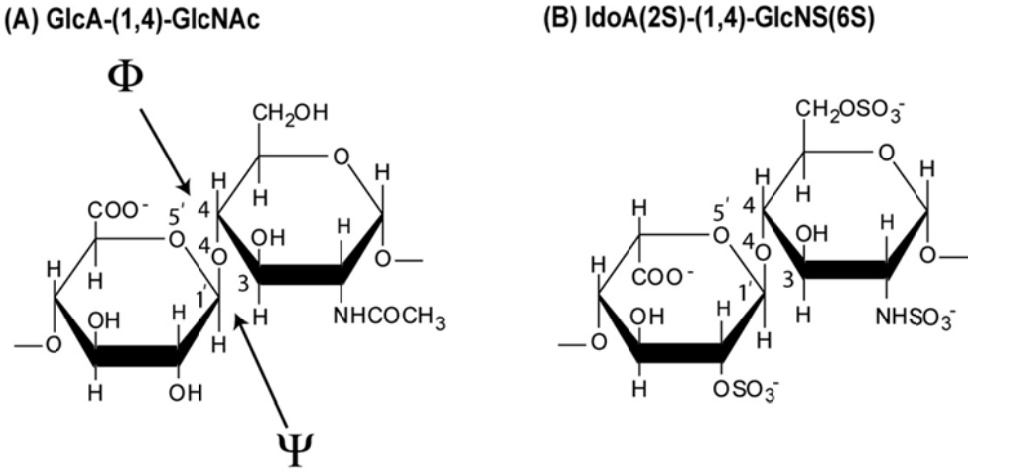


Figure 2.9

Crystal structure of the complex of FGF1 (green) with FGFR2 (cyan) and dp10. In this complex dp10 exhibits a bent structure (arrowed) (reproduced from Pellegrini *et al*., 2000).

consisting of repeating unit of uronic acid and glucosamine which are linked together by an α-(1-4) glycosidic bond (Figure 2.10A) (Lane & Lindahl, 1989). The uronic acid may be either iduronic acid or glucuronic acid. Unlike heparin, HS has a lower ratio of iduronic acid to glucuronic acid. The glucosamines are approximately equally *N*-acetylated (GlcNAc) and *N*-sulphated (GlcNS). The disaccharide unit with *N*-acetyl group (GlcA-(1-4)-GlcNAc) is the major repeating unit in HS (Lyon & Gallagher, 1998). HS has a distinct organisation of domains that is not present in heparin (Turnbull & Gallagher, 1990; Gallagher, 2001; Capila & Linhardt, 2002). The region in HS chains that is enriched in IdoA(2S) and GlcNS are called *S*-domains, while the region containing clusters of GlcA and GlcNAc residues are called *NA*-domains (Figure 2.10B). This *NA*-domain has not yet been found to directly influence protein binding; it may function as a spacer between the *S*-domains (Turnbull & Gallagher, 1990). *N*-sulphated disaccharide sequences largely occur in clusters of 2-8 repeating units that may also contain heparin-like trisulphated disaccharides as well as disulphated disaccharides. These *S* and *NA*-domains are separated by mixed sequences of *N*-acetylated and *N*-sulphated disaccharides (Bernfield, 1992). In these mixed sequences, the GlcNAc residues are often sulphated at the C-6 position. In term of interactions with various biologically active proteins, these mixed sequences of HS may have a significant biological function. The best example of such an interaction is the anti-thrombin-pentasaccharide interaction where these mixed sequences reside in the pentasaccharide sequence (Figure 2.6b). The *S*-domains along with these adjacent mixed sequences comprise the hypervariable regions that give different functional characteristics to HS from different cells. To date, no solution and crystal structural information is available for free HS in the literature. Only one crystal structure of a porcine intestinal mucosa derived HS tetrasaccharide complexed with heparinase II (PDB code: 3E7J, Figure 2.11a) is available. This HS tetrasaccharide consists of two *N*-acetyl-D-glucosamine (NAG), one D-glucuronic acid (GCU) and one 4,5-dehydro-D-glucuronic acid (GCD) residues. The arrangement of these residues is NAG-GCU-NAG-GCD (Figure 2.11b). Like heparin dp10 in a ternary complex of FGF1-FGFR2-dp10, this crystal structure also revealed a kink between the sugar rings of the bound HS tetrasaccharide.

(A)



(B)

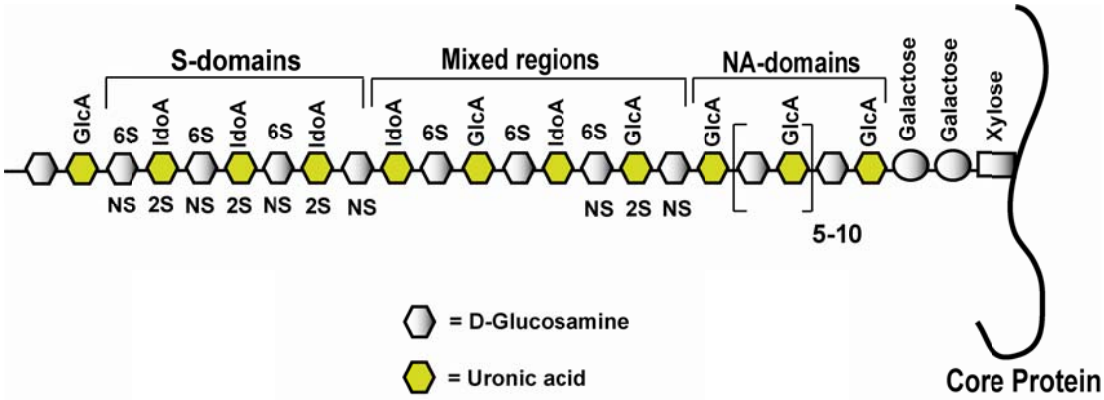
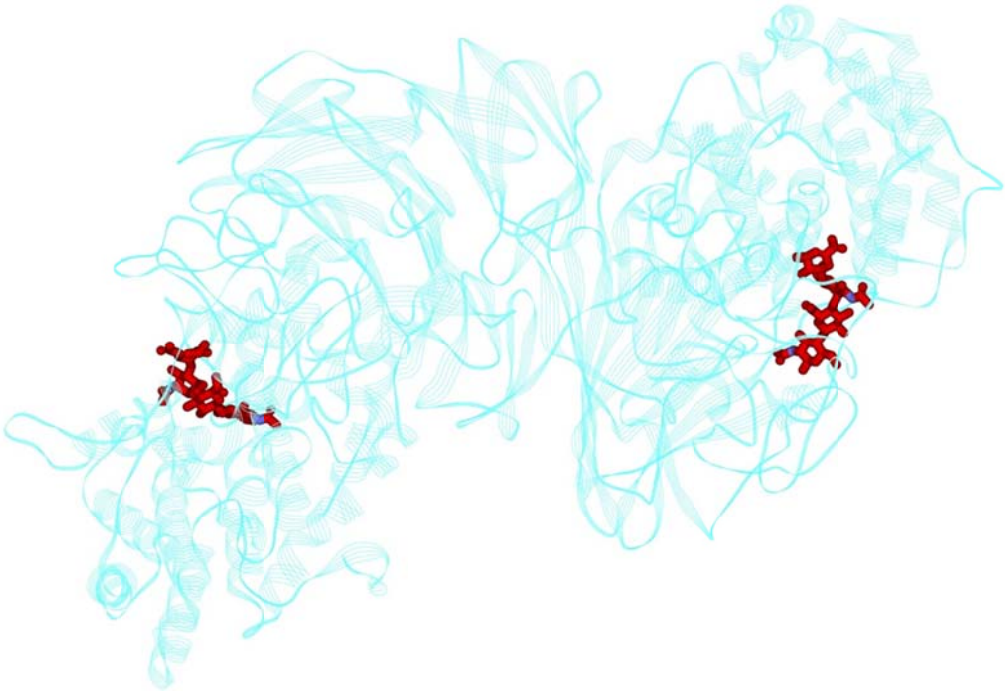


Figure 2.10

(A) Chemical structures of the two disaccharide repeating units of heparan sulphate (a) The major repeating disaccharide unit in heparan sulphate (glucuronic acid -> *N*-acetylglucosamine). **(**b) A minor disaccharide repeating unit of heparan sulphate (iduronic acid -> glucosamine, 2-6 disulphate). The atoms defining phi (Φ) are O5′-C1′-O4-C4 and psi (Ψ) are C1′-O4-C4-C3. (B) Cartoon of HS proteoglycan, showing domain organization in HS chain.

(a)



(b)

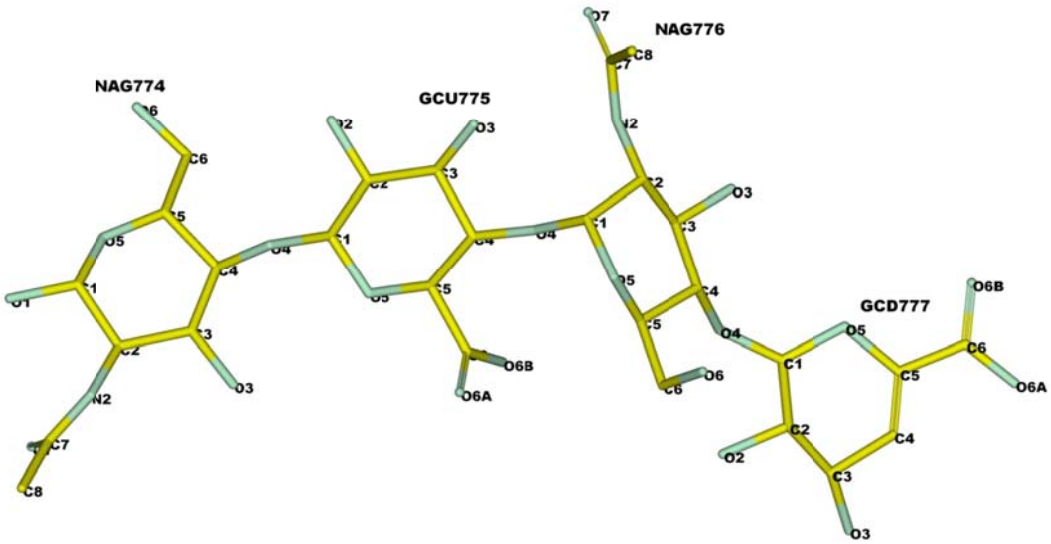


Figure 2.11

(a) Crystal structure of a dimer of Heparinase II in complex with HS tetrasaccharide (PDB code 3E7J). (b) Enlarged picture of HS tetrasaccharide shown in red in (a).

(2.2.4) Conformation of HS/Heparin

A three dimensional structure of HS/heparin is required to understand how HS/heparin carries out its normal physiological functions. The glycosidic linkage in the heparin chain is relatively stiff in solution. The GlcA and GlcNAc residues exist in the $^4C_1$ conformation within the chain, whereas the IdoA residues oscillate between the chair ($^1C_4$) conformation and skew boat ($^2S_0$) conformation without significant conformational change in the backbone (Torri *et al*., 1985) (Figure 2.12). Because the energy barrier between the $^1C_4$ and $^2S_0$ conformations is not high, the oscillation between these two conformations is rapid. IdoA favours the $^2S_0$ conformation rather than the $^1C_4$ conformation within the chain, because in the $^1C_4$ conformation the bulky carboxyl group is in the equatorial position whereas all other the substituents are in axial position (Capila & Linhardt, 2002). Similarly in the $^2S_0$ form, the three sulphate groups per disaccharide are clustered on one side with a similar arrangement on the other side for the next trisulphated disaccharide in the sequence (Mulloy & Forster, 2000; Hricovini *et al*., 2002). This cluster of sulphate groups is more dispersed in the $^1C_4$ conformation (Hricovini *et al*., 2002). These clusters of sulphate groups are separated from each other by 1.7 nm. It has been suggested that the conformational flexibility of IdoA, where the two conformations orient the 2-O-sulphate and carboxyl groups in different positions, tends to increase the number of contacts of HS/heparin chain with the protein surface (Conrad, 1998). Upon binding to protein, IdoA residues can adopt any of these two conformations or even both. For example, the crystal structure of basic fibroblast growth factor (FGF2) complexed with heparin hexasaccharide showed that IdoA residue at position 3 is in the $^1C_4$ conformation whereas IdoA at position 5 is in the $^2S_0$ conformation (Faham *et al*., 1996). The sulphate group in an IdoA residue locked in any particular conformation is not always involved in binding with positively charged amino acid residues of the interacting protein. The Ido2S ring that is involved in binding with FGF2 adopts the $^2S_0$ form, while the Ido2S ring that is not involved in binding adopts the $^1C_4$ conformation (Hricovini *et al*., 2002). This clearly illustrates how similar monosaccharide residues adopt different conformations when bound to different proteins.

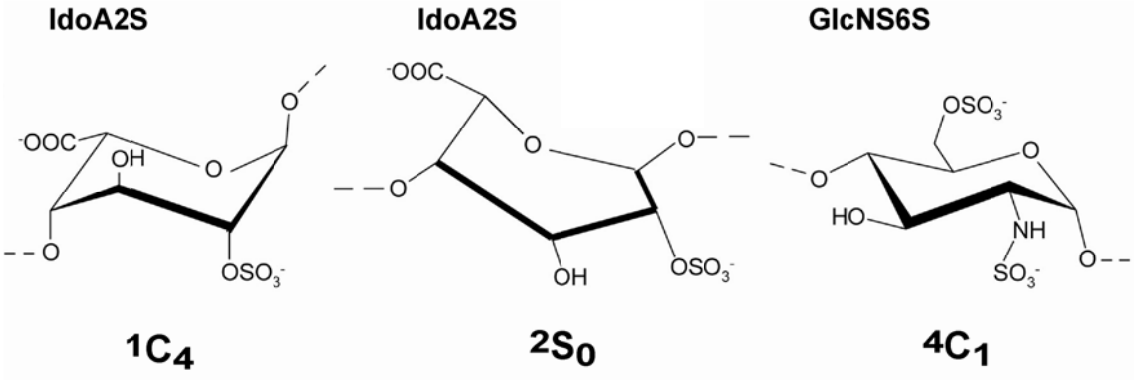


Figure 2.12

Conformations of individual sugars in HS/heparin. GlcA and GlcNS exist in the $^{4}C_{1}$ conformations, whereas IdoA exists in the $^{1}C_{4}$ and $^{2}S_{0}$ conformations (Adapted from Powell *et al*., 2004).

(2.2.5) HS/Heparin-protein interactions

Heparin and HS play essential roles in the regulation of many physiological processes. Generally, it is believed that these GAGs perform their biological functions through interactions with wide a range of proteins with diverse functions and thereby localise, stabilise, activate or inactivate these interacting proteins (Linhardt & Toida, 1997). Because of these structural similarities, heparin is often used as an analogue for the interaction of HS with proteins.

An understanding of these HS/heparin-protein interactions has been the subject of investigation for many years. Recently, X-ray crystallography and NMR studies have assisted our understanding of structural as well functional aspects of HS/heparin-protein interactions. HS has different binding specifities towards different proteins. Interleukin-8 (Spillmann, 1998), platelet factor 4 (Stringer & Gallagher, 1997) and interferon-γ (Lortat-Jacob *et al*., 1995) bind to HS through interaction of the monomer components of these proteins with *S*-domains separated by *N*-acetylated sequences. Similarly, HS oligosaccharide isolated from human skin fibroblast HS binds to basic fibroblast growth factor (bFGF, also known as FGF2) with high affinity. This binding site for FGF2 resides in the *S*-domain, which contains *N*-sulphates and iduronic acid 2-sulphates as essential residues for binding with FGF2 (Turnbull *et al*., 1992). Most of the HS binding proteins appear to interact with saccharide sequence of 5-15 units in single domains.

Like HS, heparin has also different binding sites for different proteins. Fibroblast growth factors are the most widely studied examples of HS/heparin binding proteins. FGF has three principal sulphate-binding sites within FGF-sugar binding loop. It has been shown that acidic fibroblast growth factor (aFGF, also known as FGF1) interacts with 4-5 heparin monosaccharide units through electrostatic interactions (Spivak-Kroizman, *et al*., 1994), but only 5-6 monosaccharide units are ordered in the crystal structure of heparin decasaccharide complexed with FGF1 (PDB code 2AXM). For the cross-linking of two molecules of FGF1 to form a biologically active dimer, a minimum octasaccharide is required (Ornitz *et al*., 1992). NMR and analytical ultracentrifugation studies of the interaction of FGF2 with heparin oligosaccharides showed that sugar as small as single sucrose octasulphate (SOS) residues or tetrasaccharides can induce FGF2 dimerization (Moy *et al*., 1997; Herr *et al*., 1997). In addition to this FGF dimer

formation, heparin can also induce ternary complex formation with FGF and FGF receptor (FGFR) in 1:1:1 ratios (PDB 1FQ9).

Most of these interactions of HS/heparin with heparin binding proteins are electrostatic. For example, vaccinia complement proteins (1RID), antithrombin (AT-III) and foot-and-mouth proteins form specific salt bridges between the sulphate and carboxylate groups on heparin chain and the basic residues of the proteins. In addition to these electrostatic interactions, it has also been observed that some proteins interact with HS/heparin oligosaccharide though both electrostatic interactions as well as hydrogen bonding. FGF-FGFR-heparin (PDB code 1FQ9) and thrombin-heparin oligosaccharides (PDB code 1XMN) are the best examples of such interactions. A heparin decasaccharide forms 25 hydrogen bonds within one FGF: FGFR (1:1). In the thrombin-heparin oligosaccharides complex, a heparin molecule is sandwiched between each pair of thrombin monomers (AB-GH and CD-EF). In the thrombin dimer AB-GH, the AB monomer forms six ionic interactions and four hydrogen bonds, while the GH monomer makes only three ionic interactions. In the thrombin dimer CD-EF, the CD monomer makes four ionic and two hydrogen bonds while the EF monomer makes two ionic and one hydrogen bond (Carter *et al*., 2005).

Chapter Three

Macromolecular structure and modelling

(3.1) Introduction to macromolecular structure determination

To gain an understanding of macromolecular functions at a molecular level, it is necessary to determine the three dimensional structure of the macromolecule in question. A variety of low and high resolution methods are used for macromolecular structures. High resolution methods include NMR and X-ray crystallography, while low resolution methods include electron microscopy, X-ray and neutron solution scattering and analytical ultracentrifugation.

X-ray crystallography is a high resolution technique used to determine the three dimensional structure of macromolecules, their interactions and their conformational studies at atomic resolution of 0.1-0.3 nm. Basically X-rays measure the electron density of atoms within a crystal lattice. Fitting of the electron density map into an atomic model gives the molecular structure of the macromolecule. Because the wavelength is inversely proportional to energy, the X-ray photon (a unit of light and electromagnetic radiation) has high energies (~10,000 electron volts). X-ray crystallography experiments require crystals of the molecule that contain many repeated copies of the molecules in an ordered lattice. An X-ray experiment for a single molecule is difficult due to two main practical obstacles: (1) The X-ray scattering for a single molecule is weak and contains a lot of noise from the scattering of other elements such as air and water; (2) the high energy of the X-ray photons would damage the single molecule before a useful diffraction pattern is obtained. The molecules in a crystal form reduce molecule damage. Because the X-ray diffraction pattern is obtained from the scattering of all the molecules in the crystal lattice, each individual molecule receives a smaller dose of X-rays. For example if the crystal of a macromolecule contains 10^{10} molecules and about 10^{10} photons are required to produce a diffraction pattern, then each individual molecule will diffract only one photon. When crystals of the sample are exposed to a monochromatic and collimated beam of X-rays, electrons in the molecule scatter the beam and an image of the spatial distribution of the molecule's electrons is obtained that provides a full picture of the molecule as well as the atomic coordinates of the molecule. These coordinates are obtained by the superimposition of the experimental electronic density contours on a stick drawing representing the underlying molecular structure. Computer software is used to construct the stick drawing in which the positions of individual atoms are manipulated until they fit well with the experimental image. This process of fitting of contour image to a set of atomic

coordinates is called fitting the map. The accuracy of the model is checked by calculating the diffraction pattern of the model and compared with the observed diffraction. The reliability factor (RF) is then calculated from the model that provides the objective measure of how well the model agrees with the observed data.

Nuclear magnetic resonance (NMR) is also a high resolution technique used to determine three dimensional structures of macromolecules in solution at atomic level and to investigate the dynamic features and intermolecular interactions of the macromolecules (Ferentz & Wagner, 2000; Kay, 1998; Wüthrich, 1986, 1995; Zuiderweg, 2002). Although NMR-derived three dimensional structures are of slightly lower resolution than crystallography, NMR offers an advantage over crystallography because the NMR data are recorded in solution under near physiological condition. The NMR experiment is based on the atomic nuclei that possess a spin property. The nuclei with non-zero quantum number such as ^{1}H and ^{13}C result in an NMR signal. The nuclei are oriented randomly in the absence of a magnetic field. When an external magnetic field is applied, the magnetic nuclei within the macromolecule adopt either an alpha spin or beta spin state. In an alpha spin state, the magnetic nuclei are oriented parallel to the magnetic field and have lower energy than in the absence of the external magnetic field. In contrast, the beta states in which the magnetic nuclei orientations oppose the external field are higher in energy than in the absence of the external field. When the nucleus of the molecule is irradiated with electromagnetic radiation, a nucleus with a low energy orientation can be induced to "transit" to an orientation with a higher energy. The absorption of energy during this transition is the basis of the NMR method. This absorption energy differs according to the chemical environment of different nuclei in the same molecule due to variations in the electron distribution within the molecule. NMR is used to study proteins in solution up to a size limit of ~ 35 kDa. Structural studies of large proteins and protein complexes are a challenge for NMR and the development of NMR techniques for the investigation of these larger molecular sizes is of considerable interest, for example for large proteins that cannot be crystallized, or of dynamics and molecular interactions in large complexes.

Electron microscopy is a low resolution method that operates on the same basic principle as a light microscope, but uses an electron beam instead of light. Light microscopy uses a magnifying glass lenses to focus a light beam emitted from a light source. The transmission electron microscope (TEM) uses electromagnetic

lenses to focus an electron beam on a very thin sample. This intense beam of electrons travels through the sample and creates a projection of the macromolecules on a fluorescent screen, a photographic film or a light sensitive sensor such as a charge-coupled device (CCD) camera. If many projections of different views of the molecule become available, they are computationally combined to approximate the three dimensional structure. Using TEM, structural information can also be deduced by obtaining the electron diffraction pattern of a sample. In recent years, different TEM techniques have yielded high resolution structures (Gonen *et al*., 2004, 2005). Disadvantages of this technique include radiation damage, sample preparation difficulties and requirements of extensive computational image processing.

(3.2) Analytical ultracentrifugation

Analytical ultracentrifugation (AUC) is a versatile, powerful and accurate technique for characterizing the solution state behaviour of macromolecules. Under a sufficient centrifugal force, molecules with different sizes and shapes move with different velocities and hence exhibit different sedimentation coefficient values (*s*). Many molecules have sedimentation coefficients "s" between 1 and 100 x 10^{-13} seconds. The Svedberg unit (S) is defined as 10^{-13} second. In a high centrifugal field, the molecules start sedimenting towards the bottom of the cell, resulting in a decrease in the concentration at the meniscus and an increase in concentration at the bottom. This is called a sedimentation velocity experiment. When a weak centrifugal field is applied, the process of sedimentation is opposed by the process of diffusion. After an appropriate time, equilibrium is obtained between these two opposing forces and no further changes occur with time. This is called sedimentation equilibrium. When AUC is coupled with appropriate data analysis methodologies, it can be used to determine sample purity, characterizing the assembly and disassembly mechanism of bimolecular complexes, determining subunit stoichiometries, detecting and characterising macromolecular conformational changes, and measuring equilibrium constants and thermodynamic parameters for self- and hetero-associating systems (Cole & Hansen, 1999; Harding & Winzor, 2001). This method is applicable to molecules with molecular weights ranging from several hundreds up to many millions of daltons. No other method is capable for providing such a wide range of information. In contrast to many biophysical techniques such as electron microscopy and crystallography, the analytical ultracentrifugation technique provides structural

information on macromolecules in solution under near physiological conditions. Analytical ultracentrifugation is not only applicable to proteins, nucleic acids and carbohydrates but to any molecule whose absorbance or refractive index differs from that of the solvent.

(3.2.1) Instrumentation

Historically, analytical ultracentrifugation (AUC) was a central technique in the development of biochemistry and molecular biology. It was first developed in the 1920s by Svedberg (Svedberg & Pedersen, 1940). During the last eight decades, more advances have been made in relation to the technical implementation of analytical ultracentrifuges (Schachman, 1959), the theoretical foundation of ultracentrifugation in the thermodynamics and physical chemistry of macromolecules (Tanford, 1961), the mathematical analysis of ultracentrifugation experiments (Fujita, 1975), and the application of AUC to the study of proteins (Cole *et al.*, 2008).

Beckman-Coulter Instruments has introduced two analytical ultracentrifuges, the XL-A and the XL-I. The XL-A has only UV and visible absorption optics for the detection of biopolymers (Figure 3.1) (Giebeler, 1992), while the XL-I has integrated both absorbance and Rayleigh interference optics (Furst, 1997). The absorbance optical system uses a high intensity xenon flash lamp and a scanning monochromator that detects the macromolecules at wavelengths ranging between 190 and 800 nm (Figure 3.1). The xenon flash lamp moves radially along the length of the sector to scan different positions of the solute. A slit below the sample moves to allow the light through to the cells. Thus AUC can be used for wavelength scans at a fixed position in the cell, as well as for radial scans at a fixed wavelength. The absorbance optics gives good results for macromolecules with a strong chromophore. For example it detects proteins/peptides by measuring the absorbance in the far UV (230 nm due to amide bond or 280 nm due to aromatic amino acid residues). Proteins at concentrations as low as 10-15 µg/ml can be characterized with good signal to noise ratio (Cole *et al.*, 2008). Similarly, nucleic acid can also be studied in the same concentration by measuring the absorbance at 260 nm. The beauty of this technique is that it can be applicable to samples containing two or more components with different absorption spectra i.e. proteins and nucleic acids. In such a case, data can be obtained at multiple wavelengths during the same experiment (Lewis *et al.*, 1994).

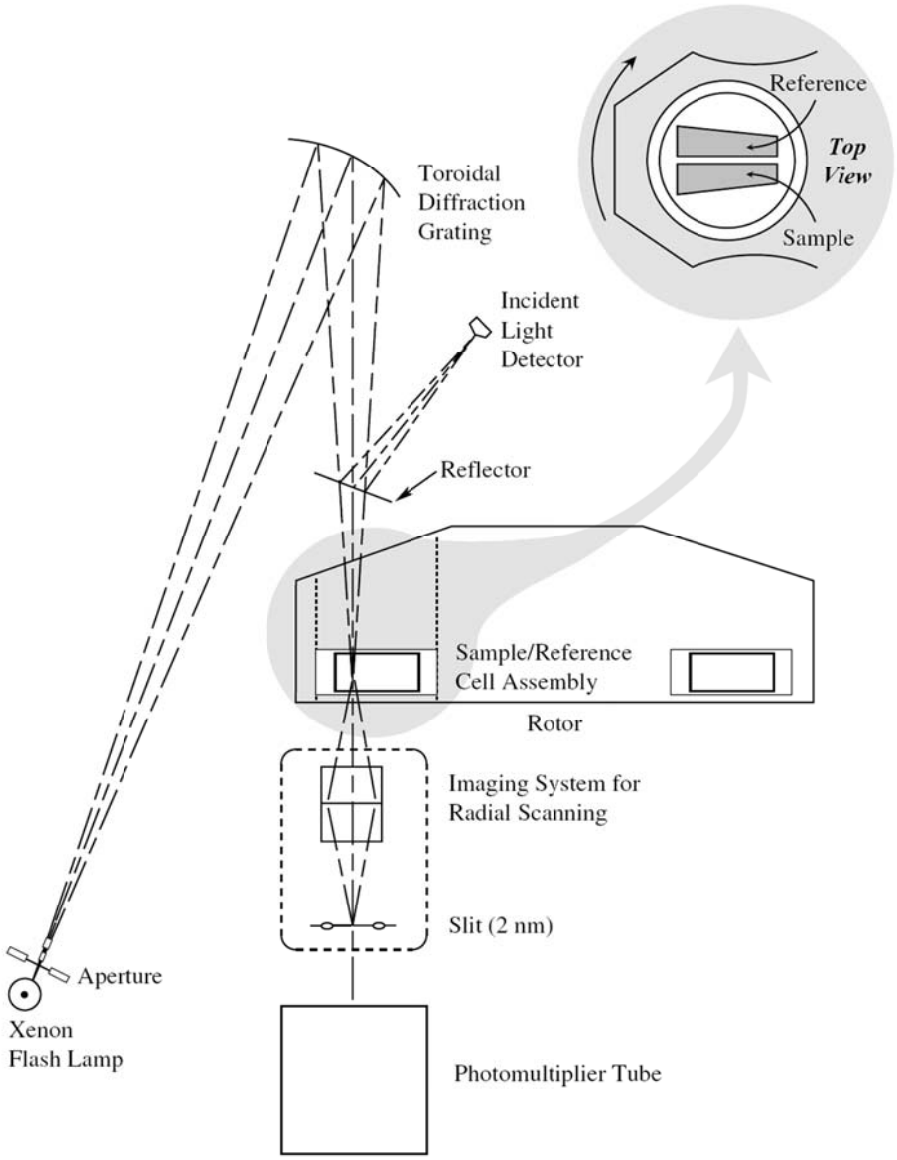


Figure 3.1

Schematic diagram of the Beckman XL-A analytical ultracentrifuge (Taken from Ralston, 1993).

The Rayleigh interference optical system is based on the refraction of light as it moves from one medium to another medium. When two beams of light from the same source are passed through different media, they refract to different extents. An interference fringe pattern is produced by superposition of the emergent beams of light from the two media. This optical system detects all biological macromolecules lacking intense chromophores through their changes in refractive index. Interference optics is useful for both sedimentation velocity and sedimentation equilibrium experi ments.

(3.2.2) Experimental applications

An analytical ultracentrifuge can be used to perform two different types of experiments, namely sedimentation velocity and sedimentation equilibrium. Both sedimentation velocity and sedimentation equilibrium measurements provide complementary information, and it is often useful to apply both techniques to a given problem. For proteins and nucleic acids to be analysed, sample concentrations ranging from 10 µg/ml to 5 mg/ml are required. It requires only a small volume of the sample, i.e. 420 µl for sedimentation velocity experiments, 110 µl for standard 3-mm column sedimentation equilibrium experiments and 15 µl for short column equilibrium experiments. The ideal ionic strength of a buffer for proteins should be 137 mM to correspond to physiological blood buffer. Before sample analysis, the sample should be equilibrated with the buffer by using dialysis bags of a specific molecular weight cut off or size exclusion gel filtration, after which the equilibrated buffer should be put in the reference cell sector.

(3.2.2.1) Sedimentation velocity

(3.2.2.1.1) Brief overview

Sedimentation velocity observes the separation of macromolecules due to their different rates of migration in the centrifugal field. Sedimentation velocity is a hydrodynamic technique which is sensitive to both the mass and shape of the macromolecules.

Sedimentation velocity cells consist of a double-sector centrepiece with a height of 12 mm and are cylindrical in shape (Figure 3.2). One sector is used for the sample, the other is the reference sector that is used for solvent. The reference sector

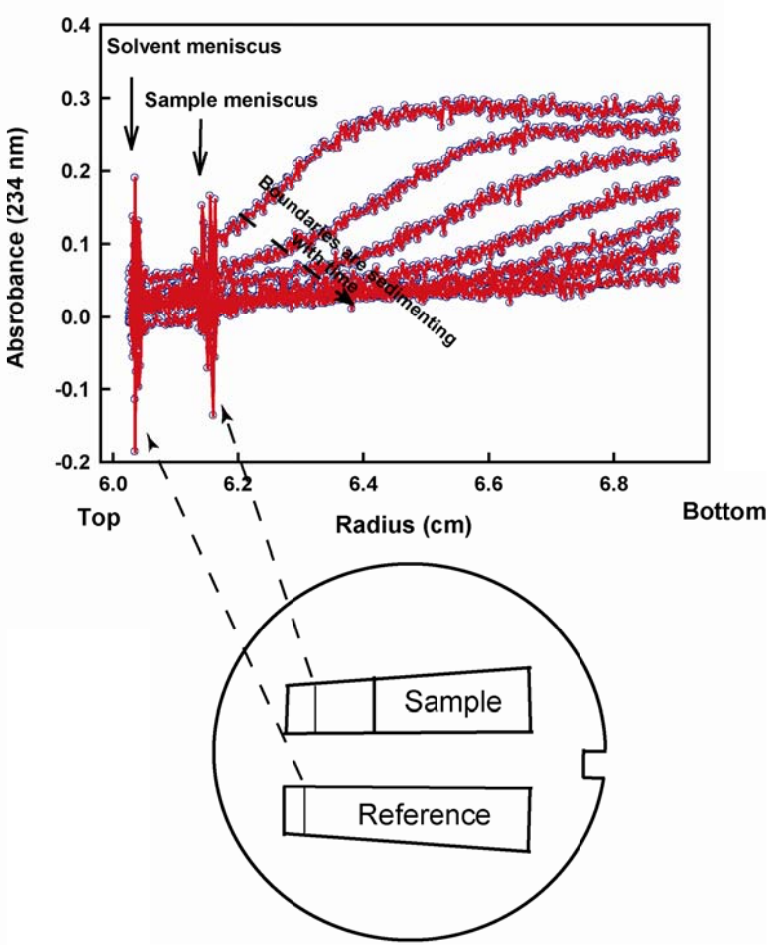


Figure 3.2

Sedimentation velocity profile of heparin dp36 showing boundary formation. Underneath is the double sector centrepiece used for sedimentation velocity experiment. The sample is placed in one sector while the solvent is placed in the other sector of the cell.

is filled slightly more than the sample sector so that the reference meniscus does not obscure the sample meniscus.

When a high centrifugal force is applied, the solute molecules start moving from the meniscus towards the bottom of the centrifuge cell in a sedimentation velocity experiment. A series of scans are recorded at regular interval to determine the rate of movement and broadening of the boundary as a function of time. The centrifugal force on the solute is partly counterbalanced by the buoyant force of the displaced solvent. The net sedimentation behaviour of macromolecules in a centrifugal field is described by the Svedberg equation:

(Eq. 3.1) $$s = v/\omega^2 r = M(1-\bar{\nu}\rho)/Nf$$

This Svedberg equation indicates that the rate of sedimentation, v, is dependent on several factors, the strength of the centrifugal field, ω^2r (where ω is angular velocity, *r* is the radial distance from the centre of rotation); the molecular mass, *M*; the molecular size and shape; the density of the solvent, ρ; the partial specific volume of the solute, $\bar{\nu}$. The frictional coefficient, *f*, is directly related to macromolecular size and shape. As the sedimentation coefficient, *s*, is dependent on shape of the macromolecules, therefore globular proteins have higher "*s*" values than elongated ones (Perkins *et al*., 2005).

A moving boundary is formed when the solute molecules start sedimenting from the meniscus towards the bottom of the centrifuge cell (Figure 3.2). Because the moving boundary sediments and diffuses with time, it begins to spread as the experiment progresses. The combination of sedimentation and diffusion in the cell is described in terms of the flow, *J*

(Eq. 3.2) $$J = s\omega^2 rc - D\,(dc/dr)$$

Where s is the sedimentation coefficient, *D* is the diffusion coefficient, *c* is the solute concentration and *d*c/*d*r is the solute concentration gradient. From this theoretical overview it can be seen that sedimentation velocity experiments can used to determine s and *D*.

Before applying a centrifugal force, the concentration of macromolecules is uniform throughout the centrifuge cell and $dc/dr = 0$. After the application of a sufficient centrifugal force, the boundary forms and moves towards the bottom of the cell at rates proportional to the sedimentation coefficients of the components in the sample. The boundary spreads due to diffusion as the experiment progresses. Sedimentation velocity experiments are usually conducted at very high rotor speed to minimize the effects of diffusion and to enhance the hydrodynamic separation. If a low rotor speed is applied, or if small molecules are studied, diffusion effects will become dominant and the concentration distributions will exhibit broad features that lead to an equilibrium state in which the sedimentation is effectively balanced by diffusion throughout the entire solution column.

(3.2.2.1.2) Data analysis methods

For the analysis of the sedimenting components in a sample, several methods have been developed. These methods include the integral sedimentation coefficient distribution $G(s)$ (van Holde & Weischet, 1978), the dc/dt method for calculating a differential apparent sedimentation coefficient distribution $g(s^*)$ (Stafford, 1992), and the sedimentation coefficient distributions $c(s)$ (Scott & Schuck, 2005).

The *G(s)* method is an integral sedimentation coefficient distribution method used for qualitative analyses of sedimentation velocity experiments. This method is implemented in ULTRASCAN and SEDFIT software. In the *G(s)* method the first step is to divide the boundary of each scan into 20 to 50 horizontal divisions that are evenly spaced between the baseline and the plateau. Extrapolating the apparent s calculated at each boundary division to infinite time, the actual s is obtained at each boundary division (Figure 3.3a). The drawback of this method is that it requires a subset of experimental scans that exhibit clear solution and solvent plateaus, hence permitting boundary divisions. This is not always used for the analyses of peptides and small proteins.

The *dc/dt* approach is the differential sedimentation velocity data analysis method used for calculating differential apparent sedimentation coefficient distribution *g(s*)*. The *g(s*)* are obtained by first subtracting pairs of scans to generate a set of $\Delta c/\Delta t$ data and then the differences are normalized to subtract any time-invariant baseline distortions. The results obtained in this method are plotted as

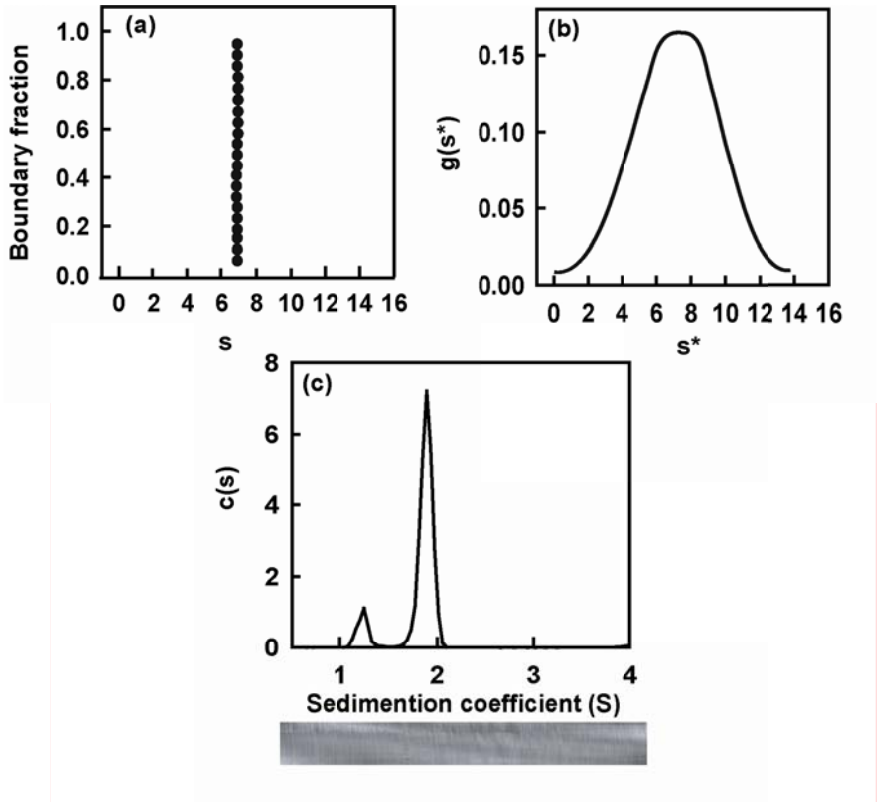


Figure 3.3

Analysis of a sedimentation velocity experiment. Graphical presentation of sedimentation velocity data analysis of an ellipsoid protein by: (a) the van Holde and Weischet method and (b) the time derivative DCDT method. (c) Sedimentation coefficient size distribution *c(s)* analyses of heparin dp36 by SEDFIT. The distribution displays two peaks at 1.24 S and 1.88 S. The major peak is heparin dp36 while the minor one is a contaminant peak.

g(s)* versus *s** (Figure 3.3b). The *g(s*)* distributions closely resemble chromatographs and can be examined visually to determine whether the sample contains a single species (single sharp peak) or many species (multiple peaks). The *g(s*)* method can be computed with DCDT+ developed by J. Philo (http://www.jphilo.mailway.com/dcdt+.htm). The main advantage of the *g(s*)* method is simplicity and the elimination of time-independent baseline components resulting in an increase in the signal to noise ratio. The main disadvantages are the limitations in the number of scans to avoid distortion of the peak shape, and the effect of diffusion in causing broadening of the peaks that can hide sample heterogeneity. The *g(s*)* distributions can be fitted to a Gaussian function for single homogenous species or a tightly associated complex. The curve from the Gaussian function gives two important results. The centre of the peak gives '$s_{20,w}$' while the width of the peak gives the molecular weight *M* (Cole & Hansen, 1999; Perkins *et al*., 2005; Scott & Schuck, 2005).

The *c(s)* analysis is also a differential sedimentation coefficient distribution method used for calculating the sedimentation coefficient of each species in the sample by a direct boundary fit approach according to the Lamm equation (Lamm, 1929). This method is computed using SEDFIT software (Dam & Schuck, 2004; Schuck, 2000). In this method, first, the program creates a grid of sedimentation coefficients values covering the expected range of interest. After creating the grid, the program produces a scaling relationship between sedimentation coefficients and diffusion coefficients by assuming a constant shape and an equal frictional ratio (f/f_O), for all species. The program then simulates the sedimentation boundaries for each point using the Lamm equation. The data are then finally fitted to a sum of these Lamm solutions using a procedure termed least-squares fitting procedure to define the concentration of each species in the grid. The resulting distribution functions obtained in this method are spiky. A regularization procedure is used to produce a smoother distribution function from this spiky distribution. This methods provides an estimated f/f_O value, an estimated molar mass distribution *c(M)*, information on the shape of the macromolecules from the size distribution *c(s)*, the residual of the fit (i.e the magnitude of deviation of the model from the data points) and the residual bitmap in order to assess if the *c(s)* model adequately describes the data (Dam & Schuck 2004, Figure 3.3c). Because the *c(s)* distribution plot is obtained from modelling the raw data, it is important to carefully inspect the goodness-of-fit

parameters (i.e. boundary fits, r.m.s.d. values, and residuals bitmaps) to obtain a model that adequately describes the experimental data. Generally, the *c(s)* analysis gives an accurate result only for dilute monodisperse samples or non-interacting protein mixtures.

The main advantage of the *c(s)* method over the *dc/dt* method is that there is no restriction and limitation on the number of scans that can be included in the analysis. The other major advantages of this method are the excellent resolution, high sensitivity and deconvolution of diffusional peak broadening from the *c(s)* distribution. For an interacting mixture of two macromolecules, it is sometimes difficult to obtain accurate sedimentation coefficient values for each species in the sample due to reaction boundaries. Reaction boundaries correspond to the co-sedimentation of both free and complexed species if the chemical equilibrium rate between these species is comparable with the sedimentation rate (Dam & Schuck, 2005).

(3.2.2.2) Sedimentation equilibrium

(3.2.2.2.1) Brief overview

Sedimentation equilibrium experiments are performed at lower rotor speeds than sedimentation velocity experiments. When solute molecules are subjected to a low centrifugal field, the molecules begin to sediment towards the bottom of the cell. The movement of these sedimenting solute molecules are significantly opposed by the process of diffusion (Figure 3.4a). Both the two opposing process balance each other after appropriate period of time and eventually an equilibrium concentration distribution of macromolecules is obtained throughout the cell (Figure 3.4b).

For an ideal noninteracting single component system, the equilibrium distribution obtained is an exponential function of the buoyant mass of the macromolecule, $M(1 - \bar{v}\rho)$ as described by the equation:

(Eq. 3.3) $$c(r) = c(a)\ \exp[M((1 - \bar{v}\rho)\omega^2(r^2 - a^2)/2RT$$

Where *c(r)* is the sample concentration at a radial position *r*, *c(a)* is the sample concentration at the meniscus, a is the radial distance of the meniscus, *M* is the molecular mass of the solute, $\bar{v}$ is the partial specific volume of the solute, ρ is the

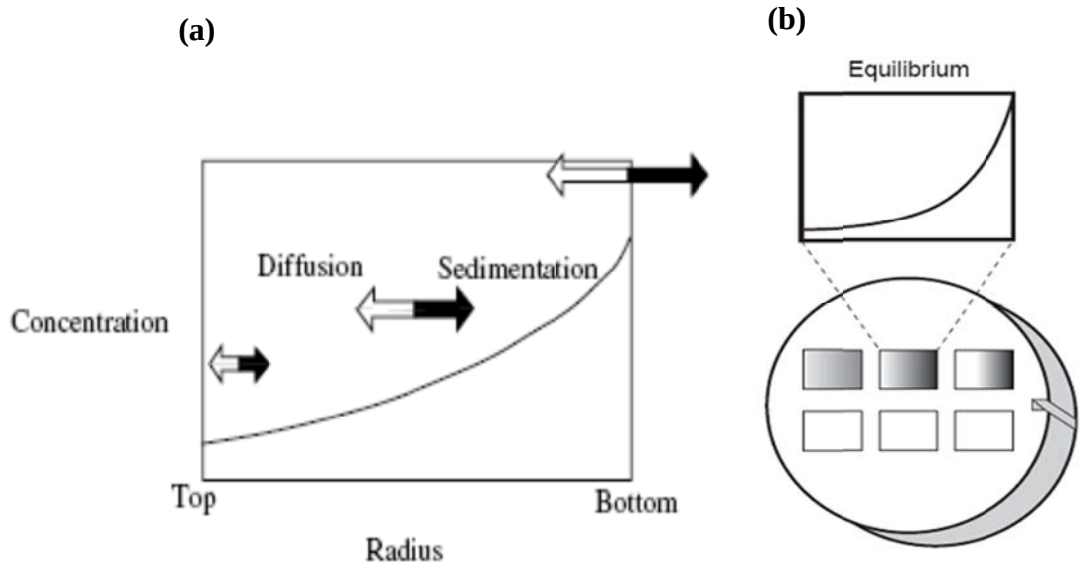


Figure 3.4

Schematic presentation of sedimentation equilibrium curves and equilibrium cell designs. (a) depicts the process of diffusion opposing the process of sedimentation (http://www.beckmancoulter.com/literature/Bioresearch/361847.pdf). (b) illustrates a six-channel sedimentation equilibrium cell (Cole & Hansen, 1999).

solvent density, ω is the angular velocity and R and T represent the gas constant and absolute temperature, respectively. The measurement of molecular mass for an ideal single component system is made by the linearization of the above equation; the logarithm of $c(r)$ plotted against r^2 yields a slope of $M(1-\bar{v}\rho)/2RT$. If the macromolecular equilibrium is more complex or if the solution is thermodynamically nonideal, these factors will also influence the concentration gradient formed in a sedimentation equilibrium experiment. In each of these cases the above equation has been modified to account for these situations (Cantor & Schimmel, 1980; van Holde, 1975). If the oligomeric species are in reversible equilibrium, the stoichiometries, equilibrium constants and the thermodynamic parameters (ΔH, ΔS) that define the interactions can be obtained.

(3.2.2.2.2) Data analysis

Generally, two main approaches are used for analysis of sedimentation equilibrium data, namely the model-independent and model-dependent approaches. The model-independent approach is most useful at the initial stages of sample analysis to determine sample behavior. This approach is also useful for comparative analysis of samples that are too complex to be fitted directly by model-dependent methods. In the model-independent method, ln c is plotted versus r^2 (Cantor & Schimmel, 1980; Ralston, 1993). According to the equation

(Eq. 3.4) $$c(r) = c(a)\exp[M((1-\bar{v}\rho)\omega^2(r^2-a^2)/2RT$$

the slope of the line is directly proportional to M. The slope of the ln c versus r^2 plot will be linear if the sample is ideal and monodisperse with respect to M. The complexity of the sample can be determined by performing non-global nonlinear least-squares of a single sedimentation equilibrium $c(r)$ gradient.

In contrast, the model-dependent analysis approach involve direct fitting of the sedimentation equilibrium concentration gradients to mathematical functions describing various physical models, such as a single ideal species, a monomer self-associating system, or an A + B → C hetero-associating system. The direct fitting method is used for detailed quantitative analysis of sedimentation equilibrium data. This approach provides the best-fit values and the associated statistical uncertainties in the fitting parameters such as molecular mass, oligomer stoichiometry and

association constants. This provides a basis to discriminate among alternative physical models (Johnson and Straume, 1994, Johnson and Faunt, 1992).

(3.3) X-ray solution scattering

Solution X-ray scattering is a low resolution structural technique that provides structural information on the size, shape, and interactions of biological macromolecules and other polymer or detergent systems in random orientations in solution (Perkins *et al*., 2008; Doniach, 2001; Svergun and Koch, 2003; Koch *et al*., 2003). It has been used to determine the structure of proteins, nucleic acids, polysaccharides and their complexes in solution under near physiological conditions without the need to crystallize the molecules. The major advantage of this technique is its capability to provide structural information on partially or completely disordered systems. Although the other structural determination technique such as X-ray crystallography and NMR produce high resolution atomic models, they have experimental requirements that limit their application to large scale structural studies: X-ray crystallography requires crystals of good diffraction quality, while NMR is limited to only small soluble proteins. In contrast, the solution scattering technique offers the possibility of varying chemical and physical conditions in solution and of exploring structural transitions that are difficult to induce and study by X-ray crystallography and NMR. This technique is applicable to molecules with molecular weight ranging from a few kilodaltons to several million daltons.

(3.3.1) X-ray solution scattering and instrumentation

Synchrotron radiation is produced by moving electrons that follow a curved path and are accelerated to speeds close to the speed of light. First, the electrons produced by a triode electron gun are accelerated to a 2000 MeV in a linear accelerator. Electrons are transmitted to a circular accelerator called a booster synchrotron (Figure 3.5a). In the booster synchrotron, the electrons are accelerated to a high energy level of 6.0 GeV using a radio frequency system. Electrons are then finally injected into a storage ring having a circumference of 844.4 m which includes both straight and curved sections. Electrons within the storage ring are forced by bending magnets, undulator magnets and focusing magnets to circulate at relativistic speeds (Figure 3.5b). The bending magnets deflect the electrons from their straight path by several degrees. The undulators are composed of small array of magnets

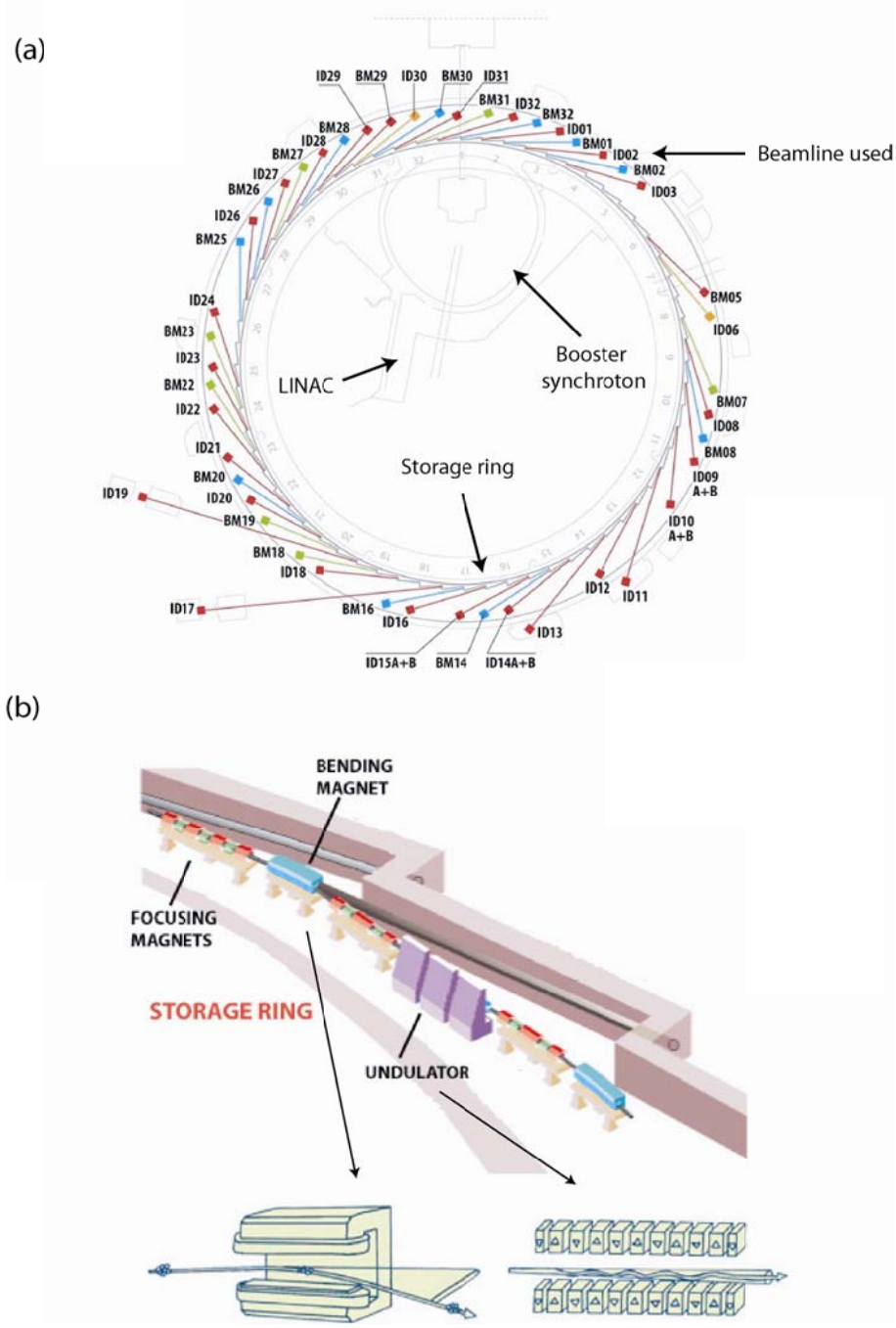


Figure 3.5

Synchrotron radiation source at ESRF. (a) The 42 beamlines of the ESRF synchrotron, the linear accelerator (LINAC), booster synchrotron and the storage ring are shown. (b) The arrangement of focusing magnets, bending magnets and undulators in the storage ring. (Adapted from ESRF website http://www.esrf.fr).

which force the electrons to follow an undulating, or wavy trajectory. The beams of radiation emitted from different bends overlap and interfere to produce even more intense X-ray beams than that produced by bending magnets. The focussing magnets are fixed in the straight section of the storage ring and are used to focus the beam of electron to keep it small and well defined. In summary, these circulating electrons emit white light comprised of all wavelengths including X-rays tangentially from the storage ring. Because the lifetime of the electron is of the order 12-14 hours, storage ring refills are performed several times a day.

Small angle X-ray solution scattering experiments are performed at Grenoble, France, using beamline ID02A. The ID02 instrument takes a white beam of X-rays from the storage ring, which is then focussed and monochromated before reaching the sample. The maximum photon flux at the sample position is of the order 3 x 10^{13} photons/second/100 mA. The ID02 camera consists of a source, optical system, cryogenic (liquid nitrogen) cooled Si-111 channel-cut monochromator, a focusing toroidal mirror, collimation and guard slits, a sample cell and a detector. A cryogenic (liquid nitrogen) cooled Si-111 channel-cut monochromator reduces the background radiation from the storage ring and the collimation and guard slits are set to define the maximum size of the beam stop and the minimum observable scattering. In order to reduce the radiation damage, fresh sample is exposed to the beam continuously during data collection through a 1.75 mm wide quartz capillary controlled by a mechanically-operated syringe. X-ray scattering measurements depend upon the sample-detector distance of 0.5 to 8 m, which is adjusted in order to get the desired Q range (Perkins *et al*., 2008). The detector is based on an image-intensified CCD device. Scattered X-ray photons fall upon on a phosphorus surface and they are then converted into visible photons by an image intensifier device. Measurements of incident and transmitted main beam intensities are important and are obtained using a diode device placed on the beam stop to protect the detector from the main beam at zero scattering angle. The detector is connected with a computer for data accumulation which in turn is connected to a work station for data storage and processing. Online data reduction using automated software is used to convert the 2D diffraction pattern into a 1D data curve for visual inspection to avoid radiation damage. The CCD takes up to 10 frames per second of 1024 x 1024 pixels (Narayanan *et al*., 2001). A schematic representation of the layout of the ID02

camera is shown in Figure 3.6. The second camera experimental hutch 2 (EH2) is used in this thesis.

(3.3.2) Basic principles

The basic principle of small angle solution scattering is to diffract X-ray photons elastically from molecules in solution and to record the scattering intensity as a function of the scattering angle. When a sample is irradiated with a highly collimated, monochromated beam of X-rays, a two dimensional circularly symmetric pattern is recorded on a two dimensional area detector placed behind the sample. The radially averaged intensities *I(Q)* are measured as a function of scattering vector *Q*

(Eq. 3.5) $$Q = 4\pi \sin \theta/\lambda$$

where 2θ = scattering angle and λ = wavelength. Scattering views structures in random orientation to a low structural resolution of about 2-4 nm from data obtained in reciprocal space in a *Q* range between 0.05 and 2 nm^{-1} (Figure 3.7). In order to get absolute scattering intensities, several important corrections are applied to the data: (1) subtraction of CCD dark current and readout noise, (2) corrections of the spatial distortion induced by image intensifier, and (3) normalization of raw intensities to absolute values. The later step can be done by Lupolen, a calibrated standard scatterer with an absolute scattering intensities in the maximum which is independent of photon energy (Kratky *et al*., 1966).

(3.3.3) Data analyses

Once the scattering curve *I(Q)* is corrected, the data is analysed by two methods, i.e. Guinier analysis and distance distribution function *P(r)*, that gives complementary information.

(3.3.3.1) Guinier analysis

Guinier method gives the radius of gyration, R_G, which measures the degree of overall structural elongation in solution and the cross-sectional structure, R_{XS}, which measures bending in the macromolecules being studied at low *Q* range (in the *Q* range between approx. 0.1 and 0.5 nm^{-1} depending on the macromolecular

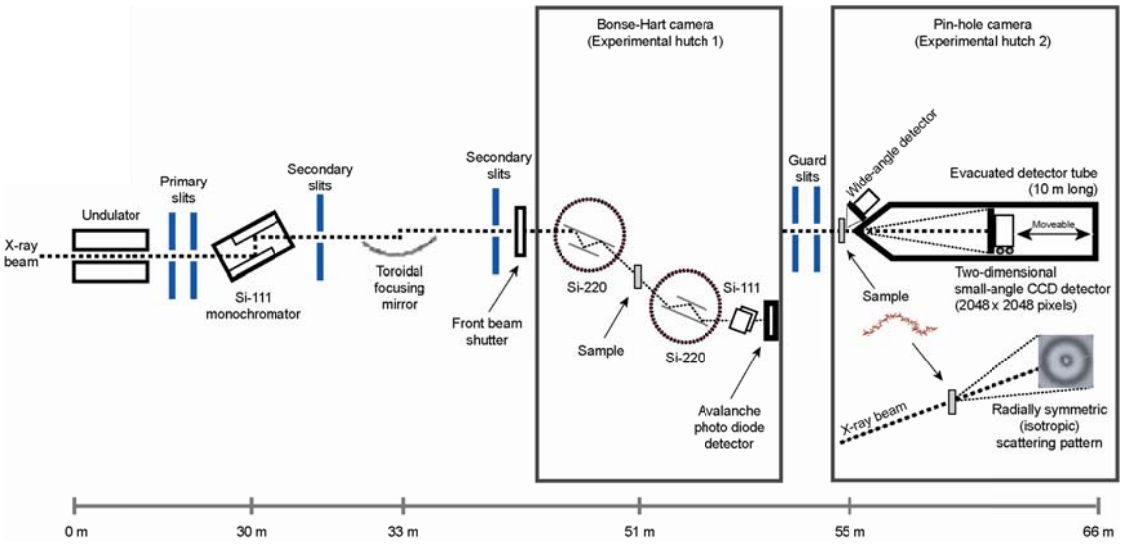


Figure 3.6

Schematic representation of the layout of the ID02 camera that is used for small angle solution scattering experiments. The undulator provides high photon flux with a low divergence. The beamline optics uses a cryogenic Si-111 monochromator and toroidal mirror optical system to produce a focused X-ray beam with wavelength 0.1 nm. The first camera experimental hutch 1 (also known as Bonse-Hart camera) is used for ultra-small angle scattering, while the second camera is used for combined wide/small angle scattering (Perkins *et al*., 2008).

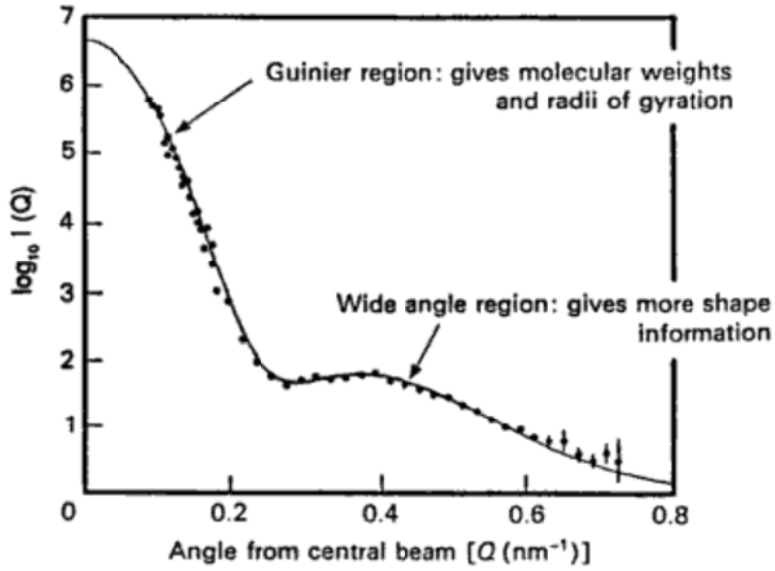


Figure 3.7

General feature of a solution scattering curve *I(Q)* for complement C1q. At low Q range the scattering curve gives R_G and R_{XS}, while at larger Q it gives more structural information (Perkins, 1986).

dimension) (Glatter and Kratky, 1982; Guinier, 1939). The Guinier plot is obtained by plotting $I(Q)$ vs Q^2:

(Eq. 3.6) $$\ln I(Q) = \ln I(0) - R_G^2 Q^2/3$$

where $I(0)$ is the intensity at zero scattering angle. The radius of gyration R_G, a measure of the structural elongation of the macromolecule, is obtained from the slope of the line if the internal inhomogeneity of scattering densities is neglected. The R_G is defined as the root-mean square distance of all the scattering centres in the macromolecules from its centre of gravity. The R_G is used to determine the anisotropy ratio, R_G/R_O, where R_O is the R_G of a sphere of equal volume to the macromolecule. The anisotropy ratio obtained for globular proteins by X-ray scattering is approximately 1.28 (Perkins, 1988). The intensity at zero scattering angle $I(0)$ is proportional to M_r^2 and this relationship can be used to calculate molecular weight and oligomeric state of macromolecule by dividing $I(0)$ by its concentration: $I(0)/c$ where c is the sample concentration.

When the structure of the macromolecule is elongated, the radius of gyration of the cross-sectional structure R_{XS} and the cross-sectional intensity at zero angle $[I(Q).Q]_{Q\to 0}$, (Hjelm, 1985), at intermediate Q values (between Q values of 0.3 to 1.0 nm^{-1}, depending on the macromolecular dimension), are obtained from the equation below:

(Eq. 3.7) $$\ln [I(Q).Q] = \ln [I(Q).Q]_{Q\to 0} - R^2_{XS} Q^2/2$$

where R_{XS} is the cross sectional radius of gyration which measures the shorter axes of an elongated molecule. The maximum length of the macromolecule, L, can be calculated from the Guinier parameters using the equation below:

(Eq. 3.8) $$L = \sqrt{12}(R_G^2 - R_{XS}^2)$$

(3.3.3.2) Distance distribution function *P(r)*

The distance distribution function $P(r)$ curve calculated from the full scattering curve provides the radius of gyration R_G value and maximum length of the macromolecule. The Fourier transformation of the full scattering curve $I(Q)$ curve

(between Q of 0.05 and 2 nm^{-1}) is applied to yield the distance pair-distribution function $P(r)$ which represent the structure in real space with units of nm. This is carried out according to the following equation:

(Eq. 3.9) $$P(r) = 1/2\pi^2 \int_0^\infty I(Q)\, Qr \sin(Qr) \mathrm{d}(Q)$$

Where $P(r)$ corresponds to the distribution of distances r between any two volume elements within macromolecule. GNOM software is used to transform the scattering data indirectly in real space which yields scattering curve $P(r)$ (Semenyuk and Svedgun, 1991).

This corresponds to the summation of all the distances r between the atoms within the macromolecule. The point at which $P(r)$ approaches zero gives the maximum length L of the macromolecule (Figure 3.8). $P(r)$ also gives an alternative calculation of the R_G and $I(0)$ values from the full scattering curve $I(Q)$ that should agree with those from Guinier analyses. The maximum length L can be imprecise because of the lower intensity at larger r values. The maxima M in the $P(r)$ curve are the most frequently occurring distances between pairs of atoms within macromolecules. This provides information on the shape and oligomeric state of macromolecules. For example if the r value at M is half of the length of the monomer macromolecule, the macromolecule is spherical. Because $P(r)$ can be calculated easily from the crystallographic atomic coordinates, it is also useful for comparing the solution structure and the crystal structure or for examining the validity of a model.

Figure 3.9a-d shows $P(r)$ functions for a monomer, and various homodimer arrangements of this monomer with a decreasing cross-section towards the ends. A parallel arrangement of the dimer gives a maximum dimension in a similar range to that of the monomer (Figure 3.9a), whereas a linear arrangement gives a maximum dimension almost twice of the monomer dimension (Figure 3.9b). The two rectangular arrangements of the dimer namely T-type (Figure 3.9c) and L-type (Figure 3.9d) lie between these two arrangements. The $P(r)$ distribution maximum is located more at larger r values for L-type than for T-type dimer arrangements (Glatter and Kratky, 1982). In Figure 3.9a-d, the hump at larger r values indicates

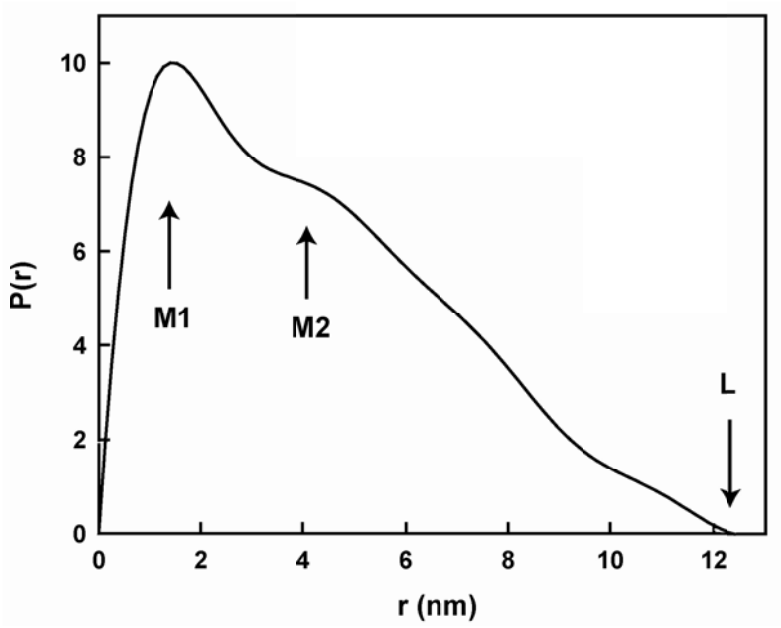


Figure 3.8

Distance distribution functions *P(r)* for heparin dp36. The maximum, *M* represents the most frequently occurring intramolecular distance. For dp36 two maxima *M1* and *M2* were observed (Chapter 4). The *P(r)* curve at an *r* value of zero gives the maximum length (*L*) of 12.3 nm.

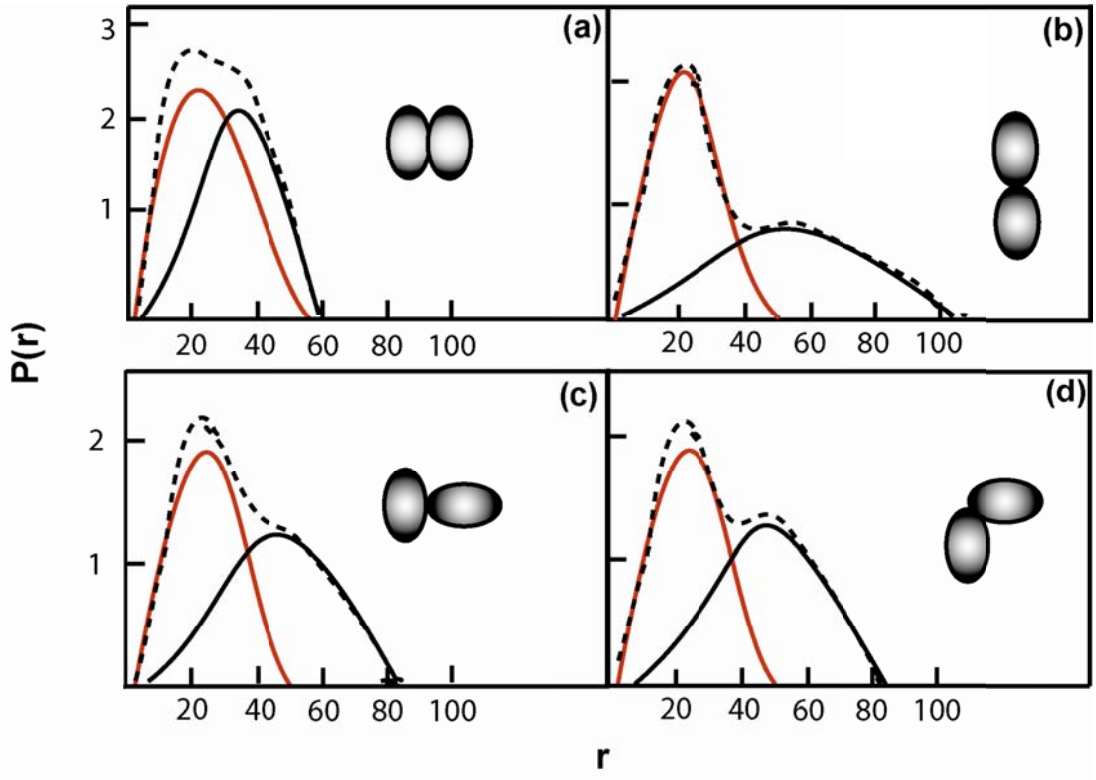


Figure 3.9

Distance distribution functions *P(r)* for different elongated macromolecular arrangements of a monomer (red solid line), dimer (dashed line) and the difference between monomer and dimer (black solid line). (a) Parallel arrangement, (b) Linear arrangement, (c) *T*-type arrangement, (d) *L*-type arrangement.

dimer formation. This hump at larger *r* values is larger for *L*-type structures than *T*-type (Figure 3.9c,d).

(3.4) Biomolecular modelling

Constrained scattering modelling is the generation of randomised structural models that best accounts for the observed scattering curve and calculated experimental parameters after filtering out the poor fit models. Since the X-ray solution scattering is a low resolution technique (2-4 nm), constrained scattering modelling is used to give medium resolution structure of the macromolecules with an improved precision of 0.5 - 1 nm compared to the use of no modelling. These models are generated based on the available atomic coordinates that represent the position of each atom in the three dimensional structure of the macromolecules determined by X-ray crystallography or NMR. If crystal or NMR structures are not available, homology models can be used to generate the models. Homology modelling is the creation of atomic-level models based on known sequences and atomic structures of homologous domains that best represents the three dimensional structure of the target domain. Homology modelling can be performed by a widely used modelling software procedure called MODELLER (Sali & Blundell, 1993). Constrained scattering modelling is able to rule out poor-fit structures and validate the scattering data by determining a family of best-fit structures that show best curve fitting with the experimental curve. Because the best fit structures are based on the observed scattering data, they can provide experimentally-derived structural and functional information on the macromolecule (Perkins *et al.*, 2008).

(3.4.1) Analysis of protein/glycoprotein composition

Before starting solution scattering modelling, it is essential to extract information from the amino acid and monosaccharide composition of the protein/glycoprotein using the program SLUV (Perkins, 1986). This program gives the partial specific volume, estimated molecular mass and molecular volume (both hydrated and unhydrated). This program also gives the extinction coefficient (at a wavelength of 280 nm) that calculated from the tryptophan, tyrosine and cysteine content of the protein (Perkins, 1986). The extinction coefficient can be used for concentration measurement. For other macromolecules these parameters can be measured experimentally.

(3.4.2) Creating atomic models

The first step is to find high resolution atomic coordinates of the macromolecule determined by X-ray crystallography or NMR. For proteins, over 65,000 crystal and NMR atomic structures are available in the PDB. If these structures are suitable for the protein or protein domains being studied, these are used as the basis for the models created by an automatic modelling procedure. If the actual high resolution structures are not available in the PDB, these individual domains are then modelled based on the known structures of homologous domains. If high resolution atomic co-ordinates for the macromolecules being studied (other than proteins) such as polysaccharides are not available for the whole macromolecule, the small oligosaccharide fragments can be joined to make a fully extended molecule. The extended molecule can then be used to as a template to generate randomized conformers.

After the construction of an initial atomic structure of the macromolecule, a set of 2000-10,000 molecular models is generated in randomised conformations. For proteins possessing fixed domains, each peptide between the fixed domains designated as linker is generated in an extended starting conformation using INSIGHT II 98.0 molecular graphic software (Accelrys, San Diego, CA, USA). For each of these linkers, a library of 5000 randomised linker conformations is generated using the same software. For proteins containing putative glycosylation sites, available three dimensional structures of N- and O-linked oligosaccharide chains are positioned properly at the glycosylation sites on the protein in the starting model. In order to obtain a wide range of conformations, the peptide structures are subjected to molecular dynamics for 300 iterations at 773K. An automated modelling procedure is then used to select linkers randomly from the linker library in order to join the domains to generate a series of complete molecular models. Typically, 2,000-10,000 trial-and-error models for the intact macromolecule are generated in one search.

For the polysaccharide being studied, conformational models in totals of 5,000-80,000 are created from the linear model by changing the phi (Φ) and psi (Ψ) angles of the glycosidic bonds. This was achieved using Discovery Studio (Accelrys, San Diego, CA, USA). The Φ and Ψ angles were randomised to take values in a maximum range of -45° or 45° from their starting values using the TorsionKick function in a PERL script that was modified from the ExtractAngle.pl script provided with Discovery Studio molecular graphics software (versions 2.1 and 2.5) (Accelrys,

San Diego, CA, USA). For example, for the heparin fragment dp36, where dp stands for degree of polymerisation, a total of 18 Φ and Ψ angles for IdoA-GlcNS and 17 Φ and Ψ angles for GlcNS-IdoA were modified in this way. In order to remove the steric overlap clashes between the atoms, minimization was carried out using a constant force field known as "Dreiding" provided with Discovery Studio. Dreiding is useful in predicting structures for molecules where there are little or no experimental data, by providing accurate geometries and reasonably accurate conformational barrier for various organic, biological and main-group inorganic molecules.

The force field refers to the functional form of the potential energy expression and the entire set of parameters used to describe the potential energy of the molecule (Ermer, 1976). Because in molecular modelling the energetic penalties are associated with the deviation of bonds and angles away from the reference or equilibrium values, the force field is used to describe the energy changes as bonds are rotated with respect to bonded and non-bonded interactions. The Dreiding force field uses general force constant and geometry parameters that are based on simple hybridisation rules rather than on individual force constants and geometric parameters that depend on the particular combination of atoms involved in the bond, angle, or torsion terms (Mayo *et al*., 1990). A concept that is more common to most force fields is that of an atom type which contains information about its hybridisation state and sometime the local environment. Atoms with the same atom type are treated identically in the molecular mechanics force field. The atom types in the Dreiding force field are denoted by five characters:

- The first two characters represent the elemental symbol (for example C_ for carbon, Sn for tin), where the elements with one letter have an underscore.
- The third character represents hybridization state or geometry (for example 1 = linear (sp1), 2 = trigonal (sp2), 3 = tetrahedral (sp3) and R = an sp2 involved in a resonance.
- The fourth character represents the number of implicit hydrogen (for example C_R2 is a resonant carbon with two implicit hydrogens, and C_32 is tetrahedral atom with two implicit hydrogens atoms.
- The fifth character indicates other special characteristic such as oxidation state.

For an arbitrary geometry of a molecule, the potential energy is expressed by valence (bonded) (E_{val}) and non-bonded interactions (E_{nb}). Valence interactions

depend on specific bonds of the structure whereas the non-bonded interactions depend on the distance between the atoms. The total potential energy is given as:

(Eq. 3.10) $$E = E_{val} + E_{nb}$$

In the Dreiding force field, the potential energy (E_{val}) of valence interactions consists of the following terms:

(Eq. 3.11) $$E_{val} = E_B + E_A + E_T + E_I$$

Where E_B, E_A, E_T and E_I are the bond energies for bond stretching, angle bending, torsion (dihedral) angle and the inversion respectively.
The potential energy of non-bonded interactions consists of the following terms:

(Eq. 3.12) $$E_{nb} = E_{vwd} + E_Q + E_{hb}$$

Where E_{vwd}, E_Q and E_{hb} are the bond energies for van der waals interactions, electrostatic interactions and hydrogen bonds. The Dreiding force field provides several functional forms for each of these energy terms which are explained in detail elsewhere (Mayo *et al*., 1990). Because randomised models created by TorsionKick function showed large deviation from the standard values (which was shown visually by steric overlap clashes between the atoms), the Dreiding force field was used to bring these deviated values to the accepted global energy values.

(3.4.3) Sphere modelling

After making molecular models, the next step is to calculate the scattering curve of all the models and fit to the experimental curve. First the atomic coordinates of models are converted into small sphere models consisting of small non-overlapping spheres of uniform density of the same total volume as the original atomic volume (Figure 3.10c). This conversion is performed by placing the atomic coordinates within a three dimensional grid of cubes. The size of the cubes and the number of atoms that each cube contains is defined by the user and are tested empirically by trial and error, so that the resulting sphere model gives a volume as

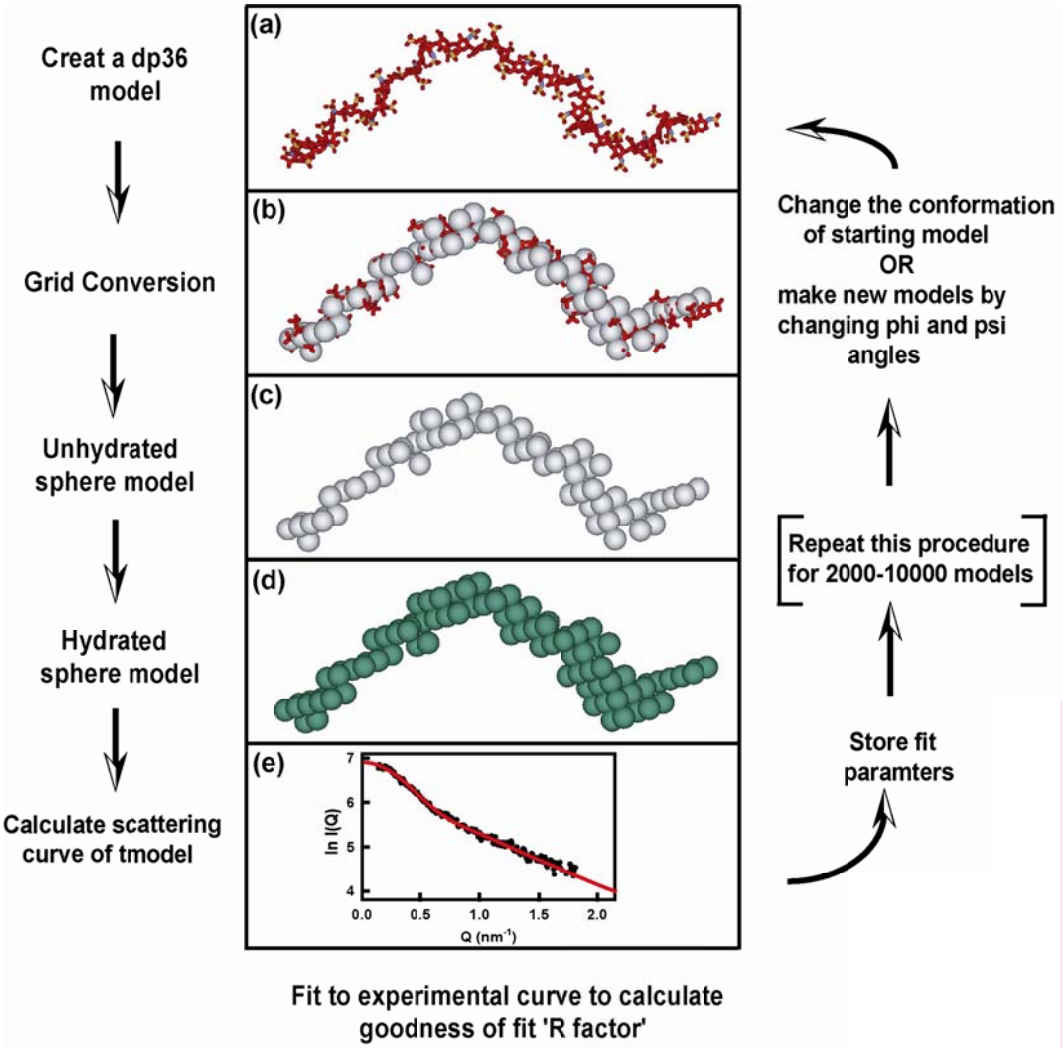


Figure 3.10

Overview of the constrained modelling procedure. (a) The best fit heparin dp36 model. (c,d) The unhydrated (grey) and hydrated (green) sphere models are shown. (d) The calculated scattering curve of the model of the best fit dp36 model (red) is compared to the experimental scattering curve (black points) to determine the goodness-of-fit *R*-factor.

close to the unhydrated volume of the macromolecule as possible. The unhydrated volume is calculated from the macromolecular composition using SLUV program (Perkins, 1986). Typically, the cube side length of 0.5 to 0.6 nm is used to produce an unhydrated sphere model. Once a grid size has been determined, it is kept constant in the automated curve fitting searches.

Because neutron experiments in ^2H_2O buffer do not detect the hydration shell, the unhydrated sphere models are directly used to calculate scattering curve. No neutron data was used in this thesis. In X-ray scattering experiment, the hydration shell is readily detected and therefore the unhydrated sphere models are hydrated before scattering curve calculation. The hydration shell is based on the usual assumption of 0.3g H_2O/g of glycoprotein and corresponds approximately to the mass of a monolayer of water molecules surrounding the macromolecule (Perkins, 2001). Spheres corresponding to this monolayer of water molecules are added evenly over the surface of the unhydrated sphere model to achieve the desired hydration volume (Figure 3.10d). The hydration shell can be seen clearly by overlaying the unhydrated and hydrated sphere models.

(3.4.4) Debye scattering curve calculation

The X-ray scattering curve *I(Q)* is calculated using the Debye equation as adapted to spheres and assuming a uniform scattering density for the spheres: (Perkins & Weiss, 1983),

(Eq. 3.13) $$\frac{I(Q)}{I(0)} = g(Q)\left(n^{-1} + 2\,n^{-2}\sum_{j=1}^{m} A_j \frac{\sin Q r_j}{Q r_j}\right)$$

(Eq. 3.14) $$g(Q) = (3(\sin QR - QR\cos QR))^2 / Q^6 R^6$$

where *g(Q)* is the squared form factor for the sphere of radius *r*, *n* is the number of spheres filling the body, A_j is the number of distances r_j for that value of *j*, r_j is the distance between the spheres, and *m* is the number of different distances r_j. Briefly the Debye equation is amended to consider the spheres of the model rather than the

atoms in the molecule. The *g(Q)* function is therefore determined by calculating the distances r from each sphere to the remaining spheres and summing the results (Equation 3.14). A single density scattering curve calculation is applicable to proteins and glycoproteins with low carbohydrate content, provided that the comparisons between the experimental data sets are consistent for the X-ray curves. If the carbohydrate content of a glycoprotein is high, differences in the scattering densities of the protein and the carbohydrate moieties may give rise to systematic deviation in X-ray curves. For such glycoproteins, models can be made using two-density sphere models. This is not related to the work presented in this thesis, as no scattering models were generated for such glycoproteins.

(3.4.5) Identification of best fit models

After scattering curve calculation, the last step is the identification of best-fit models and removal of unsatisfactory models. To do this, all these models are subjected to the following four filter parameters:

1. The automatic modelling procedure can produce models that possess unreasonable steric overlaps between domains or subunits. All those models having steric overlaps are readily detected by the grid transformation which will give too few spheres if significant overlap occurs. So all those models having less that 95% of the expected total spheres are removed, while models having at least 95% of the expected total of spheres are selected and subjected to the next filter parameter.
2. The radius of gyration, R_G and cross sectional radius of gyration, R_{XS} values are determined from the calculated scattering curves using the same Q ranges as used for experimental Guinier fits. All the models having R_G and R_{XS} values of at least 95% of the experimental data are selected for the next filter parameter.
3. The remaining good models are then filtered using a goodness-of-fit *R*-factor. An *R*-factor is calculated for a quantitative comparison of each model scattering curve $I(Q)_{cal}$ against experimental scattering data $I(0)_{exp}$ over the whole Q range, where $I(Q)_{cal}$ is set as 1000 (Figure 3.10d). The *R*-factor is given by the equation**:**

(Eq. 3.15)
$$R = \frac{\Sigma \mid I(Q)_{exp} - I(Q)_{cal} \mid}{\Sigma \mid I(Q)_{exp} \mid} \text{ x } 100\%$$

The *R*-factor is the goodness of fit between the calculated and experimental curves and is analogous to that used for the refinement of crystallographic models (Beavil *et al*., 1995). A good fit is considered to have an *R*-factor less than 10%. All these filtering parameters were carried out in a large spreadsheet with the geometrical steps used to define each model, the number of spheres in it, the R_G and R_{XS} values, and the *R*-factor values.

(3.4.6) Calculation of hydrodynamic properties

Sedimentation coefficient values monitor the degree of elongation of the macromolecule which is analogous to R_G. The calculation of $s^0_{20,w}$ values for the best fit models corroborate the outcome of the scattering data. Two approaches are used to calculate the $s^0_{20,w}$ values. The first approach is to convert the atomic co-ordinates into small spheres and to subsequently add a hydration shell. So this is the same hydrated sphere model that is used for X-ray scattering fit. These hydrated sphere models are then subjected to HYDRO software program to calculate the $s^0_{20,w}$ values (Garcia de la Torre *et al.,* 1994). The second approach is to subject the atomic co-ordinates directly into HYDROPRO software program (Garcia de la Torre, 2000). HYDROPRO represents the macromolecule as a hollow shell with equal sized spheres representing the surface. In HYDROPRO, the recommended empirical value of 0.31 nm for atomic element radius for all the atoms represents the hydration shell at the macromolecular surface. Comparison of the modelled $s^0_{20,w}$ values with the experimental $s_{20,w}$ values validate the scattering models provided that the modelled $s^0_{20,w}$ values are within the error range of ± 0.21 S to that of experimental values (Perkins *et al*., 2009).

Chapter Four

Semi-rigid solution structures of heparin by constrained X-ray scattering modelling: New insight into 19 crystal structures of heparin-protein complexes

(4.1) Introduction

Heparin is a highly sulphated linear polysaccharide that is found in the granules of mast cells and granulated cells of organs such as the liver and intestine (Lane & Lindahl, 1989). Heparin is an analogue of heparan sulphate which mediates a wide range of biological and physiological activities through its interactions with proteins. These include the inhibition of blood coagulation (Lane & Lindahl, 1989), complement activation (Kazatchkine *et al*., 1981; Sharath *et al*., 1985), angiogenesis and tumor growth (Folkman *et al*., 1983; Crum, *et al*., 1985), antiviral activity (Holondniy *et al*., 1991; Shieh & Spear, 1994; Capila & Linhardt, 2002) and the release of lipoprotein lipase and hepatic lipase (Lane & Lindahl, 1989; Liu *et al*., 1992). The breadth of heparin-protein interaction offers many potential strategies for therapeutic intervention at the cell-tissue-organ interface. The classical example is the frequent use of heparin in the clinic as an anticoagulant.

The basic subunit of heparin is a disaccharide (dp2) of uronic acid and D-glucosamine linked by a (1-4) glycosidic bond (Figure 4.1) (Capila & Linhardt, 2002). The uronic acid can be either α-L-iduronic acid (IdoA) which accounts for up to 90% of heparin or β-D-glucuronic acid (GlcA) which accounts for up to 10% of heparin. The most common occurring heparin disaccharide contains three sulphate groups, which are located on the 2-OH group of IdoA, and on the 2-NH_2 group and the 6-OH group of D-glucosamine. The high degree of sulphation and carboxylation makes heparin the most negatively charged macromolecule known in nature. Like heparin, heparan sulphate is a linear polysaccharide consisting of alternating uronic acid and α-(1-4)-D-glucosamine residues, but differs in exhibiting a reduced degree of sulphation (Figure 4.1). Unlike heparin, heparan sulphate is less modified and contains a higher proportion of *N*-acetylated saccharides (Gallagher & Walker, 1985; Hileman *et al*., 1998).

Structural studies of heparin are essential because of the extensive number of reported heparin-protein interactions and their involvement in a broad range of pathophysiological processes. Up to now, molecular structures for heparin have been limited to small fragments (Capila & Linhardt, 2002). For free heparin, solution structures by NMR spectroscopy are available for heparin oligosaccharides (Mulloy *et al*., 1993) and synthetic pentasaccharides dp5 (Ragazzi *et al*., 1990). These revealed a linear structure. For heparin-protein complexes, the current availability of 19 crystal structures (September 2009) have revealed heparin structures of sizes

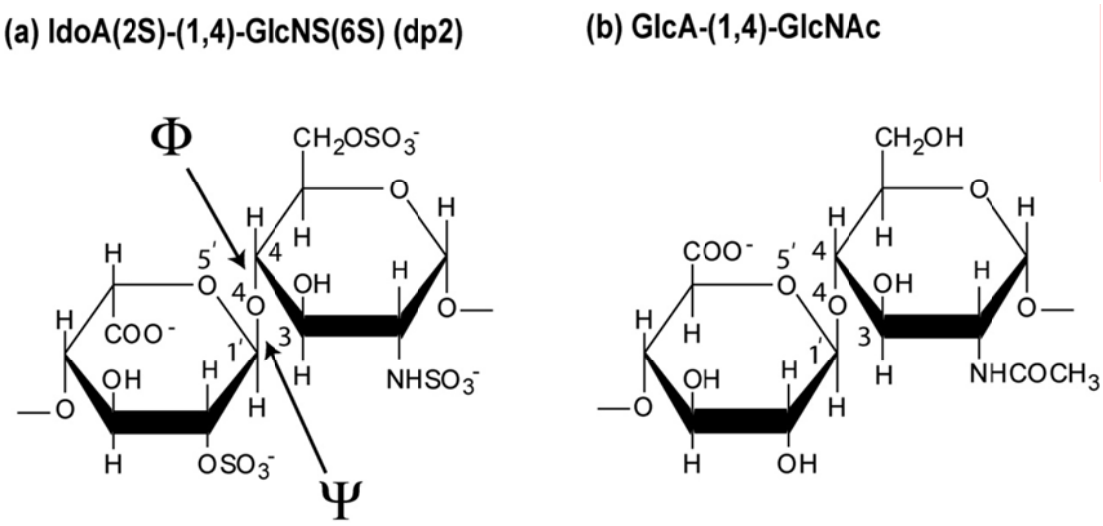


Figure 4.1

Chemical structures of the disaccharide repeats in heparin and heparan sulphate. (a) The major repeating disaccharide unit in 90% of heparin (iduronic acid-2-sulphate -> glucosamine-2,6-disulphate). The location of the bonds that are monitored to calculate phi (Φ) from the O5′-C1′-O4-C4 atoms and psi (Ψ) from the C1′-O4-C4-C3 atoms are arrowed. (b) The major disaccharide repeating unit of heparan sulphate (glucuronic acid -> N-acetylglucosamine), which comprises 10% of heparin.

dp4 to dp10. These include crystal structures for acidic and basic fibroblast growth factors separately (Faham *et al.*, 1996; DiGabriele *et al.*, 1998), the complexes of acidic and basic fibroblast growth factors with their receptors (Pellegrini *et al.*, 2000; Schlessinger *et al.*, 2000), thrombin (Carter *et al.*, 2005), antithrombin and the complex of antithrombin–S195A and coagulation factor Xa (Jin *et al.*, 1997; McCoy *et al.*, 2003; Johnson *et al.*, 2006; Langdown *et al.*, 2009), foot and mouth disease virus (Fry *et al.*, 1999) neurokinin 1 receptor (Lietha *et al.*, 2001), annexins V and A2 (Capila *et al.*, 2001; Shao *et al.*, 2006), vaccinia complement control protein (Ganesh *et al.*, 2004), and protein C inhibitor (Li & Huntington, 2008). Many of these crystal structures likewise show extended heparin conformations. At the opposite extreme of resolution, macroscopic solution structures for large polydisperse heparin fractions have been studied by X-ray solution scattering and analytical ultracentrifugation (Pavlov *et al.*, 2003; Perez Sanchez *et al.*, 2006). No structural information at a molecular level was reported for these larger heparin fractions. Fibre diffraction studies have also been reported for heparin oligosaccharides (Nieduszynski & Atkins, 1973; Atkins & Nieduszynski, 1977).

To extend our understanding of heparin structures, we require molecular structures for heparin fragments ranging from dp6 to dp36. These were obtained by applying small angle X-ray scattering and analytical ultracentrifugation in conjunction with constrained modelling as a powerful combination of three techniques to determine molecular structures (Perkins *et al.*, 2008, 2009). This multidisciplinary approach is well-established for solution structure determinations of large multidomain complement and antibody proteins (Bonner *et al.*, 2009), but is novel for oligosaccharides or polynucleotides. We show that heparin fragments starting from dp18 onwards become progressively more bent with increase in size. In solution, heparin structures are essentially extended and rigid, but possess enough flexibility in terms of bending to facilitate binding to a broad range of proteins in essentially preformed conformations. The results provide new insight on how heparin binds to proteins.

(4.2) Results and discussion

(4.2.1) Sedimentation velocity data analysis for the heparin fragments

Six oligosaccharide fragments of heparin dp6 to dp36 were prepared in collaboration with Dr. Barbara Mulloy (National Institute of Biological Standards

and Control, Blanche Lane, South Mimms, Potters Bar, Hertfordshire, UK) (Material and methods: section 4.4.1). These fragments were purified through gel permeation chromatography using Biogel P-10 column (Figure 4.2). This elution profile shows that the dp6 and dp12 fractions eluted as well-resolved peaks, while the dp18, dp24, dp30 and dp36 fractions became less resolved with increase in size. The top fractions corresponding to each individual peak from Biogel P-10 column ran as single homogenous peak in analytical high performance size exclusion chromatography, which was performed as described (Mulloy *et al*., 1997) to determine the sizes of the oligosaccharide fractions (Figure 4.3).

Analytical ultracentrifugation (AUC) studies macromolecular structures in solution by following their sedimentation behaviour under a high centrifugal force (Cole *et al*., 2008). Both the shape of the six dp6-dp36 fragments at 0.5 mg/ml and their degree of polydispersity was analysed by AUC size distribution analyses using SEDFIT. Three rotor speeds were used to obtain better resolution of size distribution analyses of these fragments with no or minimum back diffusion effects and to check the speed dependence of the size distribution analyses. Using both absorbance and interference optics, here the size distribution analyses of these fragments at rotor speed of 40,000 rpm, 50,000 rpm and 60,000 rpm were performed. The absorbance data using rotor speed of 40,000 rpm gave a single peak with sedimentation coefficient $s_{20,w}$ values of 1.03 S for dp6, 1.38 S for dp12, 1.45 S for dp18, 1.54 S for dp24, 1.61 S for dp30 and 1.83 S for dp36 was observed (Figure 4.4a). The corresponding interference analyses at 40,000 rpm for dp6-dp36 also resulted in good boundary fits and single *c(s)* peaks with similar $s_{20,w}$ values of 1.22 S for dp6, 1.60 S for dp12, 1.69 S for dp18, 1.74 S for dp24, 1.80 S for dp30 and 1.89 S for dp36 were obtained (Figure 4.4b).

Using a rotor speed of 50,000 rpm, the absorbance data analyses reproducibly resulted in good boundary fits and revealed a single peak in each case with sedimentation coefficient $s_{20,w}$ values of 1.01 S, 1.29 S, 1.43 S, 1.58 S, 1.62 S and 1.77 S for dp6-dp36 in that order (Figure 4.5a). The correspondence interference data analyses also gave good boundary fits and single *c(s)* peaks with $s_{20,w}$ values of 1.19 S, 1.34 S, 1.42 S, 1.47 S, 1.60 S and 1.89 S for dp6-dp36 (Figure 4.5b). Data analyses at 60,000 rpm, using only interference optics, resulted in single major *c(s)* peaks with $s_{20,w}$ values of 1.08 S for dp6, 1.19 S for dp12, 1.05 S for dp18, 1.19 S for dp24, 1.55 S for dp30 and 1.88 S for dp36 were obtained (Figure 4.6). One small

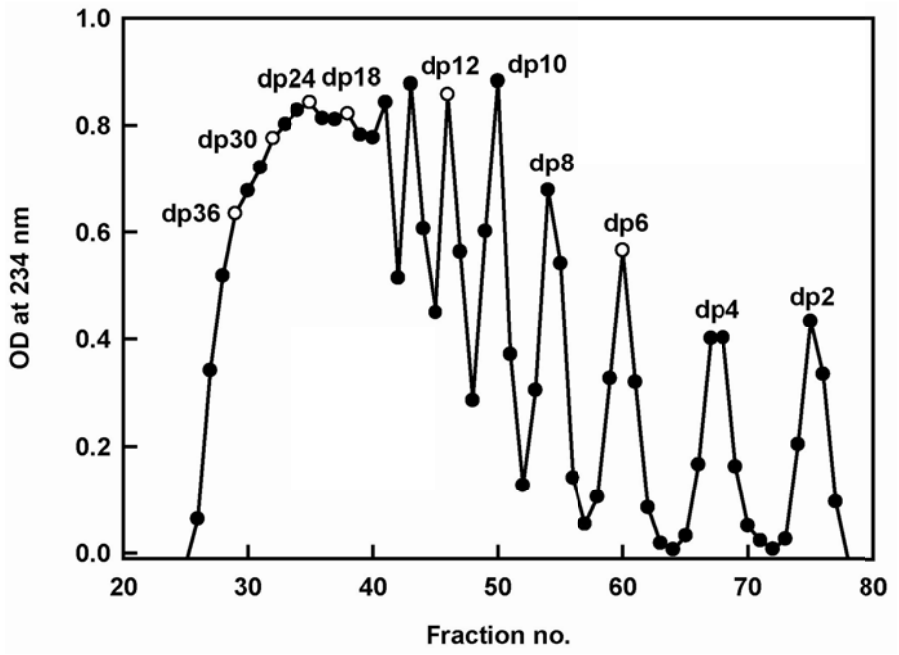


Figure 4.2

Purification profile of heparin oligosaccharide fragments dp2–dp36. The heparin fragments were eluted as peaks using a Biogel P-10 column in 2% ammonium bicarbonate solution at a flow rate of 0.2 ml (10 min). Fractions of 2 ml/10 min were collected, and their heparin concentrations were measured spectrophotometrically at 234 nm. The purified fractions used for this study are shown by open circles.

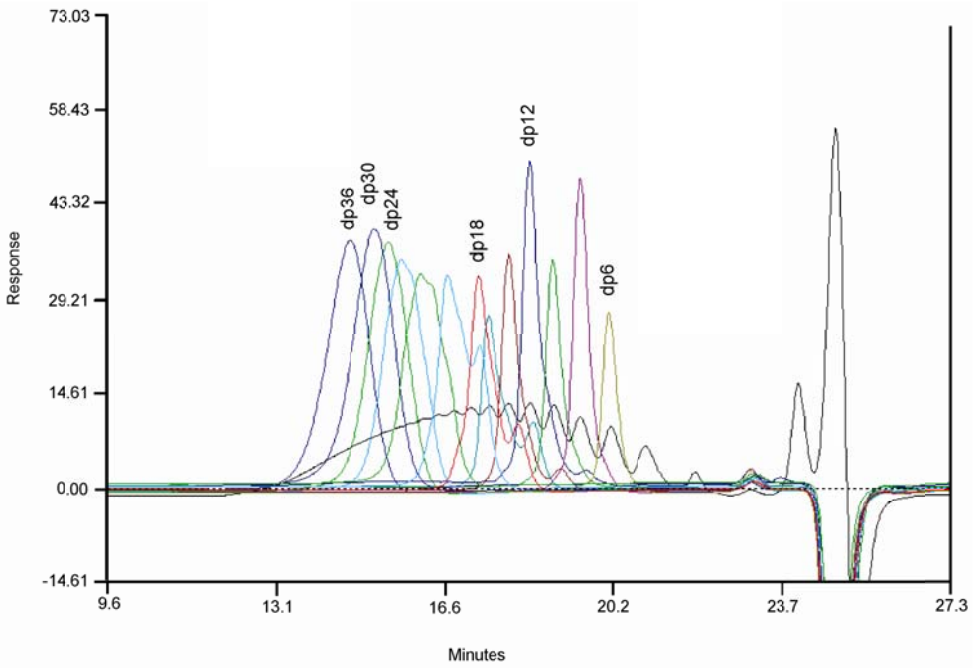


Figure 4.3

Analytical gel permeation chromatography of purified fractions of dp6 to dp36 to determine their sizes compared to a standard heparin curve (black). The heparin fragments were eluted as peaks using two columns (TSK G3000 SW-XL, 30 cm; TSK G2000 SW-XL, 30 cm; Anachem, UK) connected in series. The eluant was 0.1 M ammonium acetate solution at a flow rate of 0.5 ml/min, and heparin was detected with a refractive index detector.

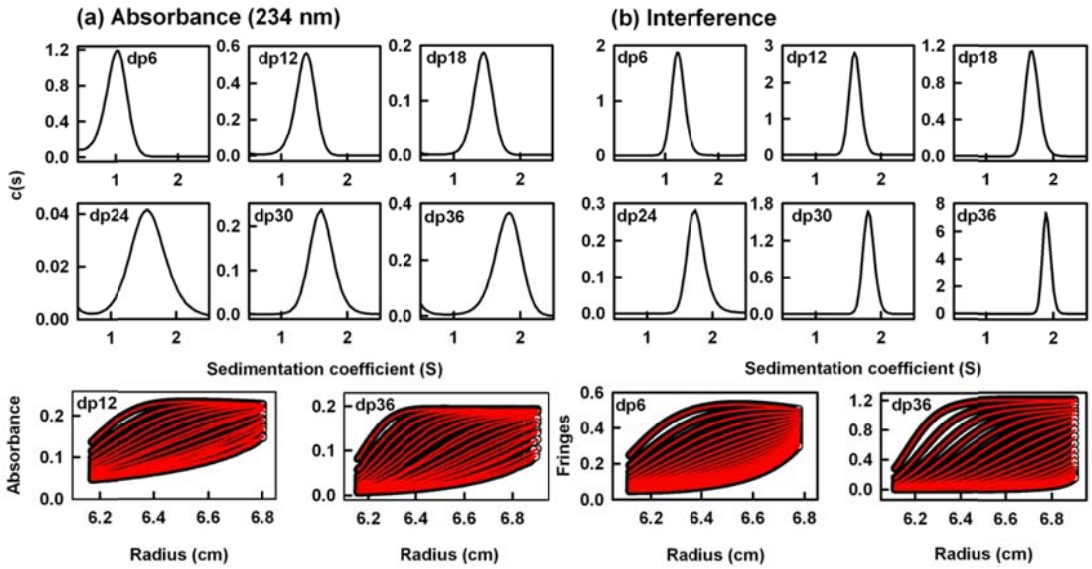


Figure 4.4

Sedimentation velocity size distribution analyses *c(s)* of heparin dp6 to dp36 at 40,000 rpm. The (a) absorbance and (b) interference boundary scans were fitted using SEDFIT software for the heparin fragments at 0.5 mg/ml.

(a) The absorbance data using a wavelength of 234 nm gave $s_{20,w}$ peaks at 1.03 S for dp6, 1.38 S for dp12, 1.45 S for dp18, 1.54 S for dp24, 1.61 S for dp30 and 1.83 S for dp36. Beneath these panels, representative boundary fits are shown for dp12 and dp36, in which only every 11th scan of the 220 fitted boundaries for dp12 and every 9th scan of the 190 fitted boundaries for dp36 are shown for clarity.

(b) The interference data gave $s_{20,w}$ peaks at 1.22 S for dp6, 1.60 S for dp12, 1.69 S for dp18, 1.74 S for dp24, 1.80 S for dp30 and 1.89 S for dp36. Beneath these panels, representative boundary fits are shown for every 14th and 13th scan of the 280 and 260 fitted boundaries for dp6 and dp36 respectively.

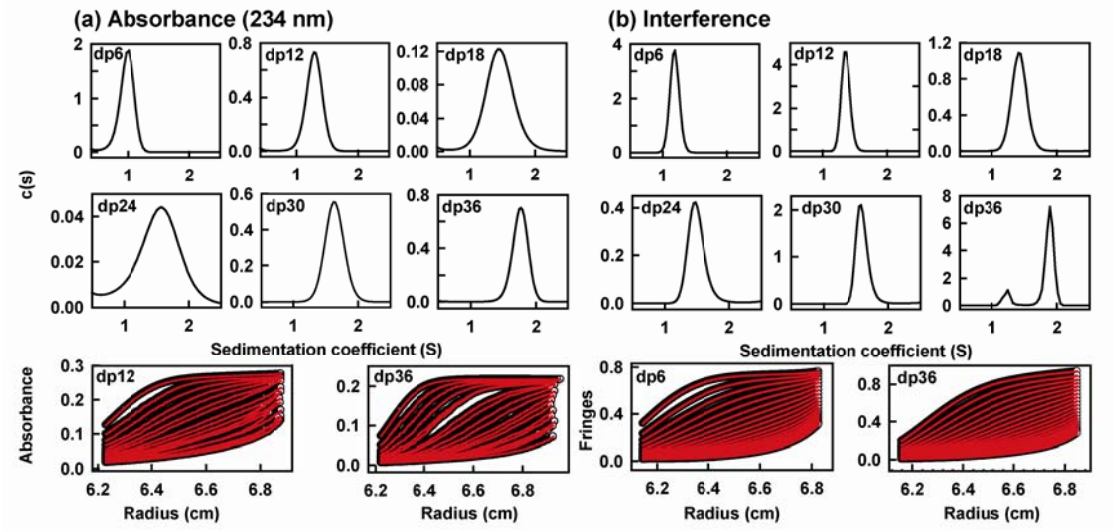


Figure 4.5

Size distribution analyses *c(s)* of heparin dp6–dp36 at rotor of 50,000 rpm. SEDFIT software was used to analyse (a) absorbance and (b) interference date for the heparin fragments at 0.5 mg/ml.

(a) The absorbance data gave $s_{20,w}$ peaks at 1.01 S for dp6, 1.29 S for dp12, 1.43 S for dp18, 1.58 S for dp24, 1.62 S for dp30, and 1.77 S for dp36. Beneath these panels, representative boundary fits are shown for dp12 and dp36, in which only every 10th scan of the 210 fitted boundaries is shown for clarity.

(b) The interference data gave $s_{20,w}$ peaks at 1.19 S for dp6, 1.34 S for dp12, 1.42 S for dp18, 1.47 S for dp24, 1.60 S for dp30, and 1.30 S and 1.89 S for dp36. Beneath these panels, representative boundary fits are shown for every 14th scan of the 280 fitted boundaries for dp6 and dp36.

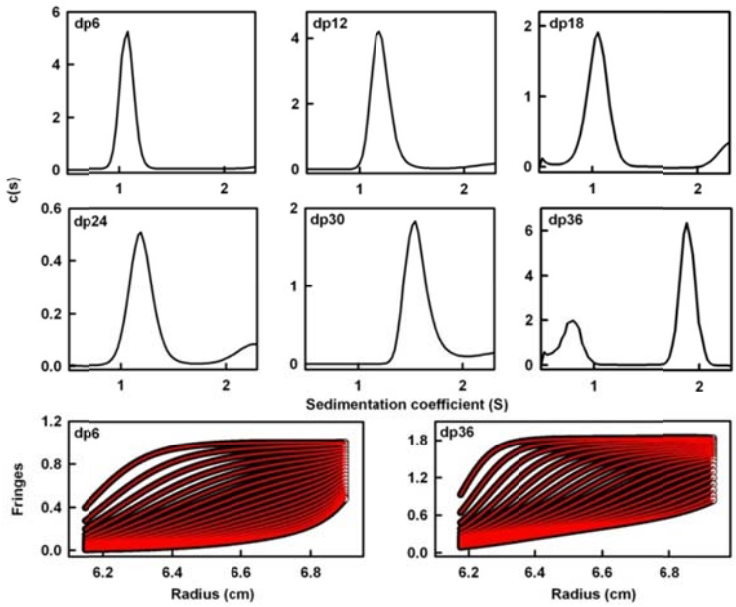


Figure 4.6

SEDFIT size distribution analyses *c(s)* of interference data for heparin dp6 to dp36 at 60,000 rpm. Single $s_{20,w}$ peaks were observed at 1.08 S for dp6, 1.19 S for dp12, 1.05 S for dp18, 1.19 S for dp24, 1.55 S for dp30. Two $s_{20,w}$ peaks; one major at 1.88 S and one minor at 0.8 S, were observed for dp36. Beneath these panels, representative boundary fits are shown for every 29th and 15th scan of the 580 and 320 fitted boundaries for dp6 and dp36 respectively.

contaminant peak at interference $s_{20,w}$ values of 1.3 S and 0.8 S for dp36 was observed both at 50,000 rpm and 60,000 rpm respectively. This contaminant peak was not observed at 40,000 rpm, indicating that this speed is comparatively lower to get better resolved $s_{20,w}$ peaks.

The peaks and their widths may be used to assess fragment polydispersity. In this regard, the interference optics showed better resolution and single narrow major peaks. The single peaks indicate that the six fragments are relatively homogenous in size. The similar peak widths from either absorbance or interference optics suggest that all six fragments showed narrow size distributions and are relatively homogenous, in agreement with the high performance size-exclusion chromatography described above. Although slight variation in the $s_{20,w}$ values of the absorbance and interference data were observed at different rotor speeds, the six averaged values (together with those for dp10 which was prepared in a similar manner (Fernando *et al.*, 2007) showed a linear increase with the heparin length (Figure 4.7a,b). Previously, $s^0_{20,w}$ values of 1.69 S, 1.88 S, 1.99 S and 2.25 S from interference optics were reported for heparin fragments of approximate sizes deduced to be dp26, dp32, dp38 and dp46 in that order in 0.2 M NaCl (Pavlov *et al.*, 2003; Perez Sanchez *et al.*, 2006). These earlier values agree well with the more extensive data presented in Figure 4.7b.

The sedimentation coefficient $s^0_{20,w}$ values of 16 heparin models corresponding to linear dp10 to dp40 structures were calculated using HYDROPRO software on the basis of the linear dp12 NMR structure (Mulloy *et al.*, 1993). The HYDROPRO software program is used to calculate the $s^0_{20,w}$ values directly from the atomic co-ordinates (Section 3.4.6). While variability was observed in their calculated values, the overall dependence of the $s^0_{20,w}$ values on length is linear (Figure 4.7) and agrees well with experiment (Figure 4.4-4.6). By AUC, there is no evidence of structural bending from the $s_{20,w}$ values for heparin lengths of up to 18 nm in dp36.

(4.2.2) X-ray solution scattering data for heparin fragments

Solution scattering is a diffraction technique that studies the overall structure of biological macromolecules in solution (Perkins *et al.*, 2008). In order to obtain more detailed structural information compared to that obtained from AUC, the solution structures of the six heparin fragments dp6–dp36 at 0.5 mg/ml were

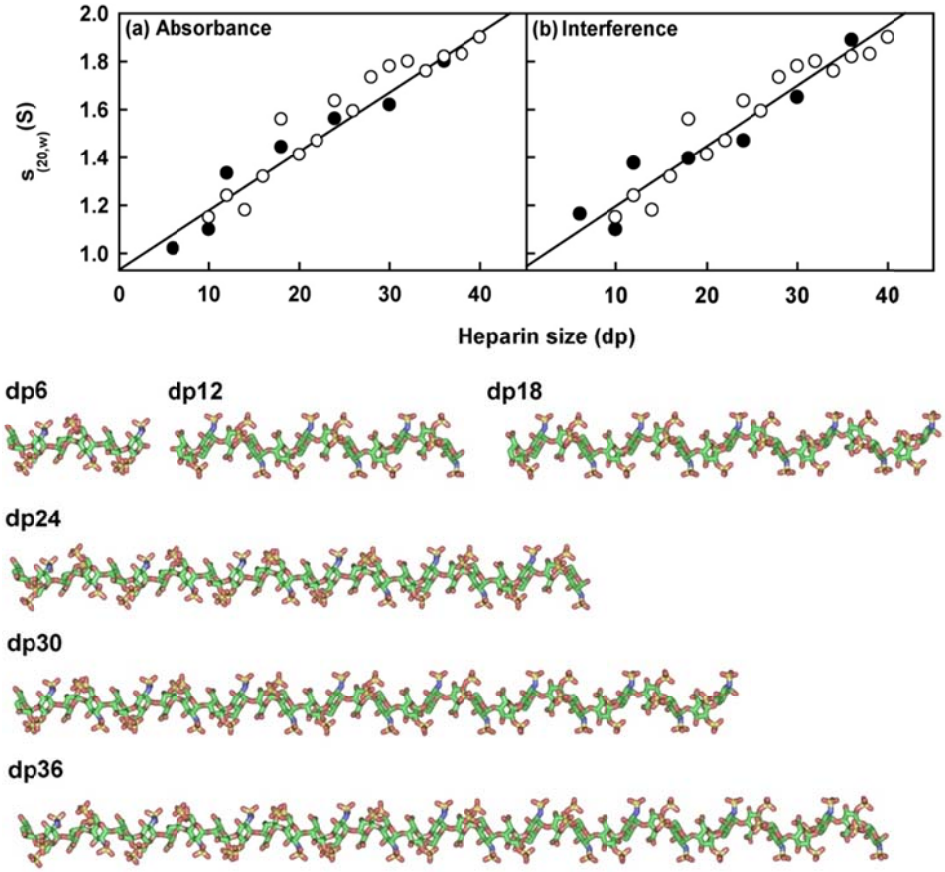


Figure 4.7

Comparison of the experimental and predicted sedimentation coefficients for the dp6–dp36 heparin fragments. The experimental values for dp6–dp36 are shown by six filled circles. The predicted values for linear dp10–dp40 models are shown by open circles. (a) Comparison with the averaged experimental sedimentation coefficients at rotor speeds of 40,000 rpm and 50,000 rpm using absorbance optics. (b) Comparison with the averaged experimental sedimentation coefficients at rotor speeds of 40,000 rpm, 50,000 rpm, and 60,000 rpm using interference optics. Beneath the graphs, the linear models for dp6–dp36 that were created from the NMR structure of dp12 (PDB code 1HPN) are shown.

characterised by synchrotron X-ray scattering. Experiments give the scattering curve $I(Q)$ as a function of Q (where $Q = 4\pi \sin\theta/\lambda$; 2θ is the scattering angle, and λ is the wavelength). Guinier analyses of $\ln I(Q)$ *versus* Q^2 at low Q values give the radius of gyration R_G, which measures the degree of macromolecular elongation. Due to differences in the lengths of the heparin fragments, different Q ranges were required for each fragment in order to work within acceptable $Q.R_G$ ranges (Figure 4.8a). Thus, the Guinier fit Q range of 0.2–0.7 nm^{-1} for dp6 was successively reduced in stages to 0.14–0.40 nm^{-1} for dp36 (Figure 4.8 legend). The mean Guinier R_G values increased from 1.33 ± 0.26 nm for dp6 to 3.12 ± 0.10 nm for dp36 (Table 4.1). The R_G values of 16 heparin models corresponding to linear dp10–dp40 structures were calculated from sphere models derived from the linear dp12 NMR structure (Mulloy *et al*., 1993; Perkins and Weiss, 1983). The calculated R_G values increased linearly with the size of heparin (Figure 4.8d). In contrast, Figure 4.8d shows that, unlike the linear increase in the $s_{20,w}$ values with size, the Guinier R_G values of the six heparin fragments do not exhibit a linear increase with heparin size, with deviations being seen for dp24, dp30, and dp36. The difference from the AUC results is attributable to the better resolution of the X-ray scattering method. The R_G values in Figure 4.8d are comparable with that of 3.2 nm for dp32 measured at higher concentrations between 6 mg/ml and 20 mg/ml in 0.2 M NaCl (Pavlov *et al*., 2003).

If the macromolecule is sufficiently elongated in shape, the cross-sectional radius of gyration R_{XS} value monitors the degree of bend within the macromolecular length. Cross-sectional Guinier analyses of $\ln I(Q)Q$ *versus* Q^2 at larger Q values showed the required linearity for dp18–dp36 in a Q range of 0.9–1.5 nm^{-1}. Fits gave values of 0.50 nm for dp18 up to 0.53 nm for dp36 (Figure 4.8b; Table 4.1). The dp6 and dp12 fragments were too small for this analysis to be carried out. The small increase in the R_{XS} values with heparin fragment size indicates that the fragments are indeed elongated but become progressively slightly more bent with increase in size. Combination of the R_G and R_{XS} values according to the relationship $L^2 = 12(R_G^2 - R_{XS}^2)$ for an elliptical cylinder (Glatter & Kratky, 1982) showed that dp18, dp24, dp30, and dp36 have approximate lengths of 7.0 nm, 9.1 nm, 9.6 nm, and 10.7 nm, in that order. In contrast, larger L values of 14 nm for dp26, 16 nm for dp32, and 20 nm for dp38 have been determined on the basis of a smaller R_{XS} value of 0.43 nm (Pavlov *et al*., 2003; Perez Sanchez *et al*., 2006). Consideration of the

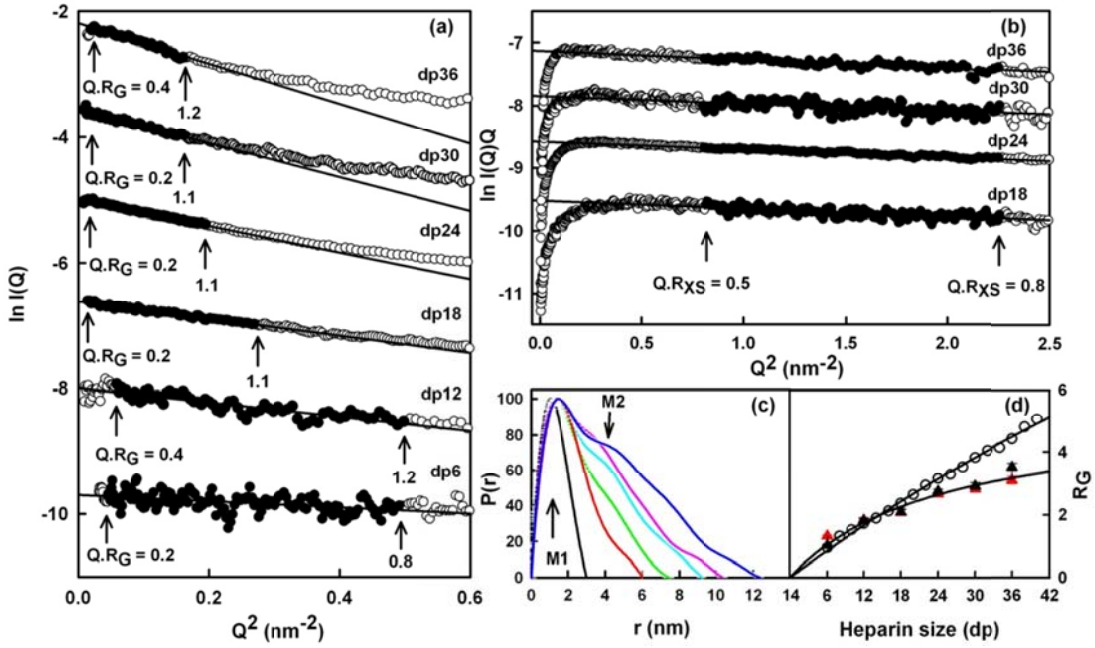


Figure 4.8

Experimental Guinier and $P(r)$ X-ray scattering analyses of heparin dp6–dp36. (a) Guinier R_G plots for dp6–dp36 at concentrations of 0.5 mg/ml. Filled circles were used to determine the radius of gyration R_G based on the best-fit lines, as shown. The Q ranges used for R_G analyses were 0.2–0.7 nm^{-1} for dp6, 0.24–0.7 nm^{-1} for dp12, 0.12–0.52 nm^{-1} for dp18, 0.08–0.44 nm^{-1} for dp24, 0.08–0.40 nm^{-1} for dp30, and 0.14–0.40 nm^{-1} for dp36. (b) Guinier cross-sectional R_{XS} plots for dp18–dp36 at concentrations of 0.5 mg/ml. The filled circles were used to determine the R_{XS} values based on the best-fit lines, as shown. The Q ranges used were 0.9–1.5 nm^{-1} in all four cases. (c) The distance distribution function $P(r)$ analyses for dp6–dp36. The r values of the maximum at M_1 were 1.05 nm (dp6: black), 1.44 nm (dp12: red), 1.36 nm (dp18: green), 1.30 nm (dp24: cyan), 1.32 nm (dp30: pink), and 1.40 nm (dp36: blue). The r values of the subsidiary maximum at M_2 were 3.7 nm (dp24), 3.4 nm (dp30), and 4.2 nm (dp36). The maximum length L values were 3.0 nm (dp6), 6.0 nm (dp12), 7.4 nm (dp18), 9.4 nm (dp24), 10.3 nm (dp30), and 12.3 nm (dp36). (d) Comparison of the experimental R_G values from Guinier plots (▲) and $P(r)$ curves (▲) with the predicted R_G values calculated from the linear models of Figure 4.7(○).

Table 4.1. X-ray scattering and sedimentation coefficient modelling fits for the solution structures of heparin fragments

Heparin fragment	Filter	Number of models	Hydrated spheres [a]	R_G (nm) [b]	R_{XS} (nm)	*R*-factor (%)	Length (nm) [c]	$s^0_{20,w}$ (S) [d]
dp6	Experimental			1.33 ± 0.26 1.03 ± 0.03			3.0	1.09 ± 0.10
dp12	Experimental			1.83 ± 0.13 1.81 ± 0.08			6.0	1.35 ± 0.03
dp18	None	5000	46-78	1.25-2.39	0.17-0.98	2.2-16.8	NA	NA
	R_G, *R*-factor	9	54-64	2.06-2.07	0.49-0.50	2.2	7.7-8.3	1.14-1.52
	Best fit	1	59	2.07	0.50	2.2	7.8, 7.0	1.32
	Experimental			2.07 ± 0.05 2.09 ± 0.09	0.50 ± 0.01		7.4	1.41 ± 0.04
dp24	None	5500	46-100	1.44-3.02	0.013-1.17	1.3-18.2	NA	NA
	R_G, *R*-factor	9	77-82	2.59-2.64	0.47-0.52	1.3-1.4	9.8-10.5	1.45-1.64
	Best fit	1	78	2.61	0.49	1.4	10.2, 9.0	1.45
	Experimental			2.67 ± 0.15 2.78 ± 0.06	0.50 ± 0.01		9.4	1.52 ± 0.07
dp30	None	6000	79-122	1.76-3.58	0.01-1.32	3.0-18.1	NA	NA
	R_G, *R*-factor	9	85-100	2.79-2.96	0.46-0.57	3.1-3.4	10.2-11.8	1.49-1.86
	Best fit	1	96	2.79	0.52	3.3	10.2, 9.6	1.86
	Experimental			2.83 ± 0.14 2.92 ± 0.11	0.52 ± 0.01		10.3	1.63 ± 0.02
dp36	None	6000	93-141	2.04-3.90	0. 10-1.46	3.5-21.7	NA	NA
	R_G, *R*-factor	9	102-120	3.18-3.28	0.47-0.63	3.5-4.0	11.1-12.9	1.72-1.95
	Best fit	1	102	3.18	0.53	3.5	12.6, 11.5	1.87
	Experimental			3.12 ± 0.10 3.52 ± 0.12	0.53 ± 0.01		12.3	1.84 ± 0.06

Table 4.1 legend

[a] The optimal totals of hydrated spheres are 53 for dp18, 71 for dp24, 88 for dp30 and 106 for dp36.

[b] The first experimental value is from the Guinier R_G analyses and the second one is from the GNOM *P(r)* analyses.

[c] The lengths *L* of the 9-15 best fit models were calculated using HYDROPRO. The length of the best-fit model length was calculated using HYDROPRO, and the second one was calculated from the modelled *P(r)* curve. The experimental L values are the model-independent values from the *P(r)* curve.

[d] The averaged experimental $s_{20,w}$ value is reported from the absorbance (234 nm) and interference data sets, with the absorbance data being recorded at 40,000 and 50,000 rpm and the interference data at 40,000, 50,000 and 60,000 rpm.

relationship $L^2 = 12(R_G^2 - R_{XS}^2)$ shows that the latter set of L values is too large to be consistent with the measured R_G values.

The distance distribution function $P(r)$ curve calculated from the full $I(Q)$ curve (Section 4.4) provides R_G values and model-independent determinations of overall lengths L following an assumption of the value of the maximum dimension D_{max} (Figure 4.8c). The mean R_G values obtained from the $P(r)$ curves increase from 1.03 ± 0.03 nm for dp6 to 3.52 ± 0.12 nm for dp36 (Table 4.1). Comparison of the six pairs of mean Guinier and $P(r)$ R_G values showed good agreement to within 0.01–0.4 nm (Table 4.1). Again, the R_G values do not increase linearly with size (Figure 4.8d). Model-independent L values are determined from the r value where the $P(r)$ curve reaches zero at large r. The experimental L values for dp6 and dp12 were 3.0 nm and 6.0 nm, respectively, in good agreement with a length of 3.2 nm for an unhydrated linear dp6 model and a length of 5.9 nm for an unhydrated linear dp12 model, provided that a hydration shell of 0.6 nm (= 2 × 0.3 nm) thickness is added to these model lengths (Perkins, 2001). The experimental L values for dp18, dp24, dp30, and dp36 were found to be 7.4 nm, 9.4 nm, 10.3 nm, and 12.3 nm, respectively (Figure 4.8c). These values show increasing deviation from the lengths of linear dp18, dp24, dp30, and dp36 models (i.e., 8.4 nm for dp18, 11.1 nm for dp24, 13.6 nm for dp30, and 16.3 nm for dp36), showing that heparin is not linear in solution. These values also deviate from the L values determined above from the R_G and R_{XS} values, and this is attributable to the assumption of an elliptical cylinder shape for this L calculation. In conclusion, both the R_G and the L values indicate a measurable degree of bending from linearity in larger heparin fragments. The $P(r)$ curves also provide the most frequently occurring interatomic distance M within the heparin structure from the r value of the peak maximum. The maximum *M1* was observed at r values that started at 1.05 nm for dp6 and slowly increased to 1.41 nm for dp36 (Figure 4.8 legend). A second subsidiary maximum *M2* may be present at about 3.4–4.2 nm for dp24, dp30, and dp36. The appearance of this second feature may correspond to a regular structure that becomes more evident in longer heparin structures. In conclusion, the *P(r)* curves indicate a measurable degree of bending from linear structures in the larger heparin fragments.

(4.2.3) Constrained modelling of the heparin fragments

Solution structures for the four larger heparin fragments dp18, dp24, dp30, and dp36 were determined by constrained scattering modelling, using the linear models created from the NMR structure of dp12 as starting constraint. Modelling was not performed for dp6 or dp12 because their sizes were too small for this procedure and poor signal-noise ratios were obtained. In the models, the bond connectivity between oligosaccharide rings was maintained, while the phi (Φ) and psi (Ψ) angles at each glycosidic bond were varied (Figure 4.1). In initial trials, around 200 random models for dp30 were made manually by changing the Φ and Ψ angles at two positions. The *R*-factors and R_G values were calculated for all the models, where the *R*-factor is a measure of goodness-of-fit, and the R_G value was calculated using the same *Q* range used for the experimental Guinier fits (Figure 4.8). The best dp30 model gave *R*-factor of 3.9 %, R_G of 2.94 nm and maximum length *L* of 10.0 nm, which are in fair agreement with the experimental R_G of 2.83 ± 0.14 nm and *L* of 10.3 nm (Figure 4.9a). The sedimentation coefficients calculated from the best-fit models dp30 model gave $s^0_{20,w}$ values of 1.66 S, in good agreement with the experimental $s_{20,w}$ value of 1.63 ± 0.02 S. The linear dp30 model showed a worsened *R*-factor of 6.3 %, R_G value of 3.41 nm, *L* value of 13.6 nm and $s^0_{20,w}$ values of 1.78 S (Figure 4.9b). This initial manual modelling for dp30 showed that the automated modelling for dp30 and other heparin fragments was possible.

In the automated modelling procedure, the bond connectivity between the oligosaccharide rings was maintained, while the Φ and Ψ angles at each glycosidic bond was varied randomly in a range of up to ± 45° from their starting values. Trial calculations showed that this range of rotations was sufficient to produce a sufficient range of bent structures that would fit the X-ray scattering curves. In all, 5,000 models for dp18, 5,500 models for dp24 and 6,000 models each for dp30 and dp36 were generated. A constant force field known as the "Dreiding minimization" protocol was applied to avoid steric clashes between the atoms (Section 3.4.2). X-ray scattering curves were calculated from these models and fitted to the experimental curves. The *R*-factors and R_G values were calculated for all the models, where the *R*-factor is a measure of goodness-of-fit and the R_G values was calculated using the same *Q* range used for the experimental Guinier fits (Figure 4.8). The comparisons of Figure 4.10a–d showed that all the models (yellow circles) spanned the experimental R_G values (broken lines), and that the best *R*-factor values were well

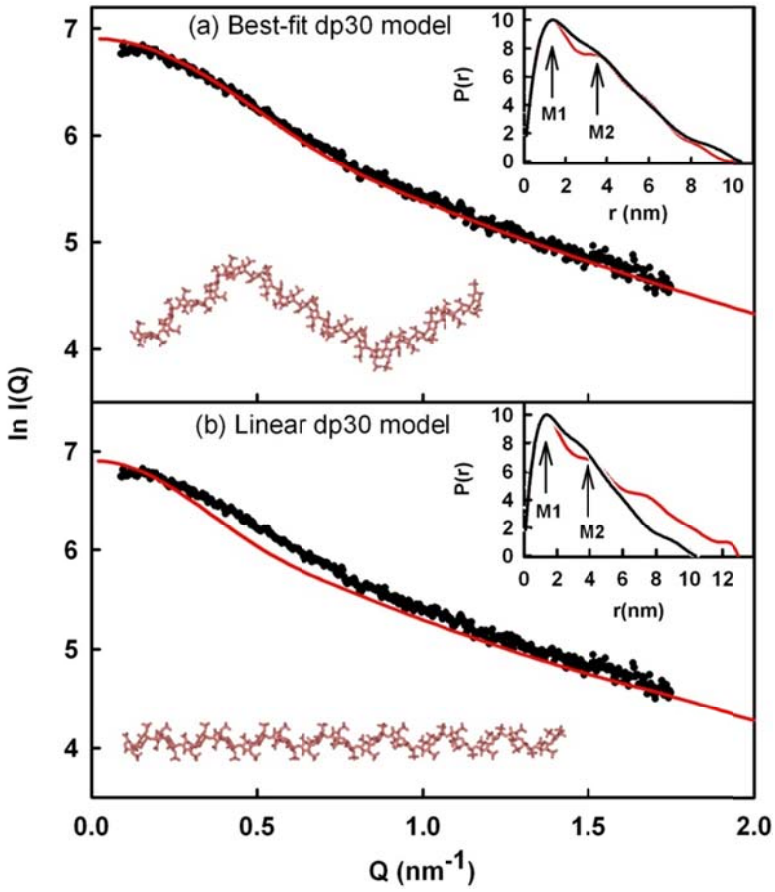


Figure 4.9

X-ray scattering curve fits for trial best-fit and linear dp30 models. The experimental scattering data are represented by points (•), and the red lines correspond to the best-fit models for dp30.The insets at the upper right show the modelled *P(r)* distance distribution functions in red compared with the experimental *P(r)* curve in black. The insets to the lower left shows the dp30 model used to calculate the curves.

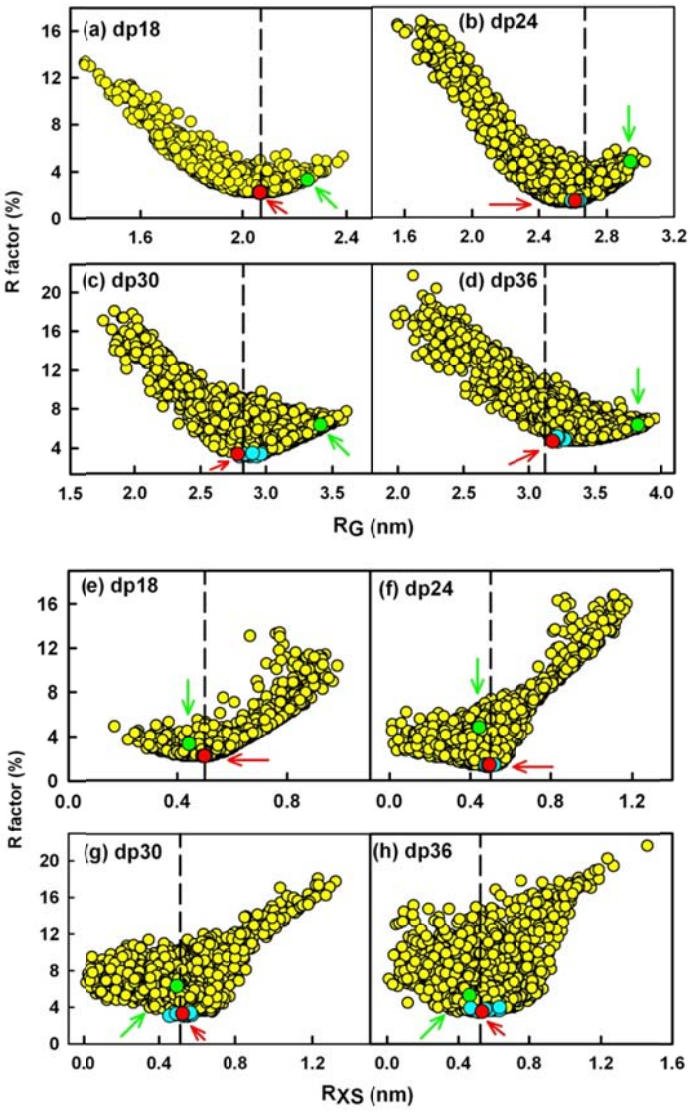


Figure 4.10

Constrained modelling analyses of heparin dp18–dp36. (a) The *R*-factor values for the 5000 trial models of heparin dp18 are compared with their R_G values. The vertical broken line corresponds to the experimental R_G value. The red circle (arrowed) denotes the best-fit model, and the green circle (arrowed) denotes the linear model from Figure 4.7. The other best-fit models, where visible, are shown in cyan close to the *R*-factor minimum. The same colour coding for heparin is used in Figure 4.11, Figure 4.12 and Figure 4.14. (b–d) The corresponding R_G analyses for the 5,500 models for dp24 and 6,000 models each for dp30 and dp36 are shown in that order. (e) The *R*-factor values for the 5,000 trial models of heparin dp18 are compared with their R_{XS} values. (f–h) The corresponding R_{XS} analyses for 5,500-6,000 models for dp24, dp30 and dp36 are shown in that order.

below the level of 5% required for excellent curve fits (Perkins *et al.*, 2008). This showed that 5,000-6,000 randomised models were sufficient to determine best fit heparin solution structures from them. The lowest *R*-factors showed the best agreement with the experimental X-ray curves, and their R_G values (red and cyan circles) agree well with the experimental R_G values. In distinction, the linear models for dp18, dp24, dp30 and dp36 showed large deviations (green circles).

For dp18, the modelling analyses show that this has a slightly bent structure that accounts for the X-ray and AUC data. All nine best-fit models gave the same *R*-factor of 2.2% and the R_G values of 2.06-2.07 nm that agree with the experimental R_G value of 2.07 ± 0.05 nm (Figure 4.10a, Table 4.1). The best-fit dp18 models show a maximum length *L* of 7.7–8.3 nm, in good agreement with the experimental maximum length of 7.4 nm. In contrast, the linear dp18 model gave a worsened *R*-factor of 3.3%, an R_G value of 2.25 nm, and an *L* value of 8.4 nm. Visual agreement between the experimental $I(Q)$ curve and the modelled $I(Q)$ curve, and that between the experimental $P(r)$ curve and the modelled $P(r)$ curve were excellent (Figure 4.11a). The calculation of sedimentation coefficients from the nine best-fit models gave $s^0_{20,w}$ values of 1.02 S to 1.47 S. The best-fit model gave a value of 1.29 S, which agrees well with the experimental $s_{20,w}$ value of 1.32 ± 0.04 S, this agreement being within the usual precision of ± 0.21 S for these calculations (Perkins *et al.*, 2009).

For dp24, modelling analyses showed similar good agreements with X-ray and AUC, but corresponded to slightly more bent structures. For the nine best-fit models, the *R*-factor and R_G values ranged from 1.3% to 1.4% and from 2.59 nm to 2.64 nm, respectively, where the latter corresponds well with the experimental R_G value of 2.67 ± 0.15 nm (Table 4.1). The maximum lengths of the models were 9.8–10.5 nm, in good accord with an *L* value of 9.4 nm. The linear model again showed a worsened *R*-factor of 4.8%, an R_G value of 2.94 nm, and an *L* value of 11.1 nm. The agreement between the experimental and the modelled $I(Q)$ and $P(r)$ curves was excellent (Figure 4.11b). The nine modelled $s^0_{20,w}$ values of 1.45–1.64 S are comparable with the experimental $s_{20,w}$ value of 1.52 ± 0.07 S, and the best-fit model gave a value of 1.45 S.

For dp30, modelling analyses showed similar good agreements with X-ray and AUC, in which deviation from a linear structure for dp30 becomes pronounced.

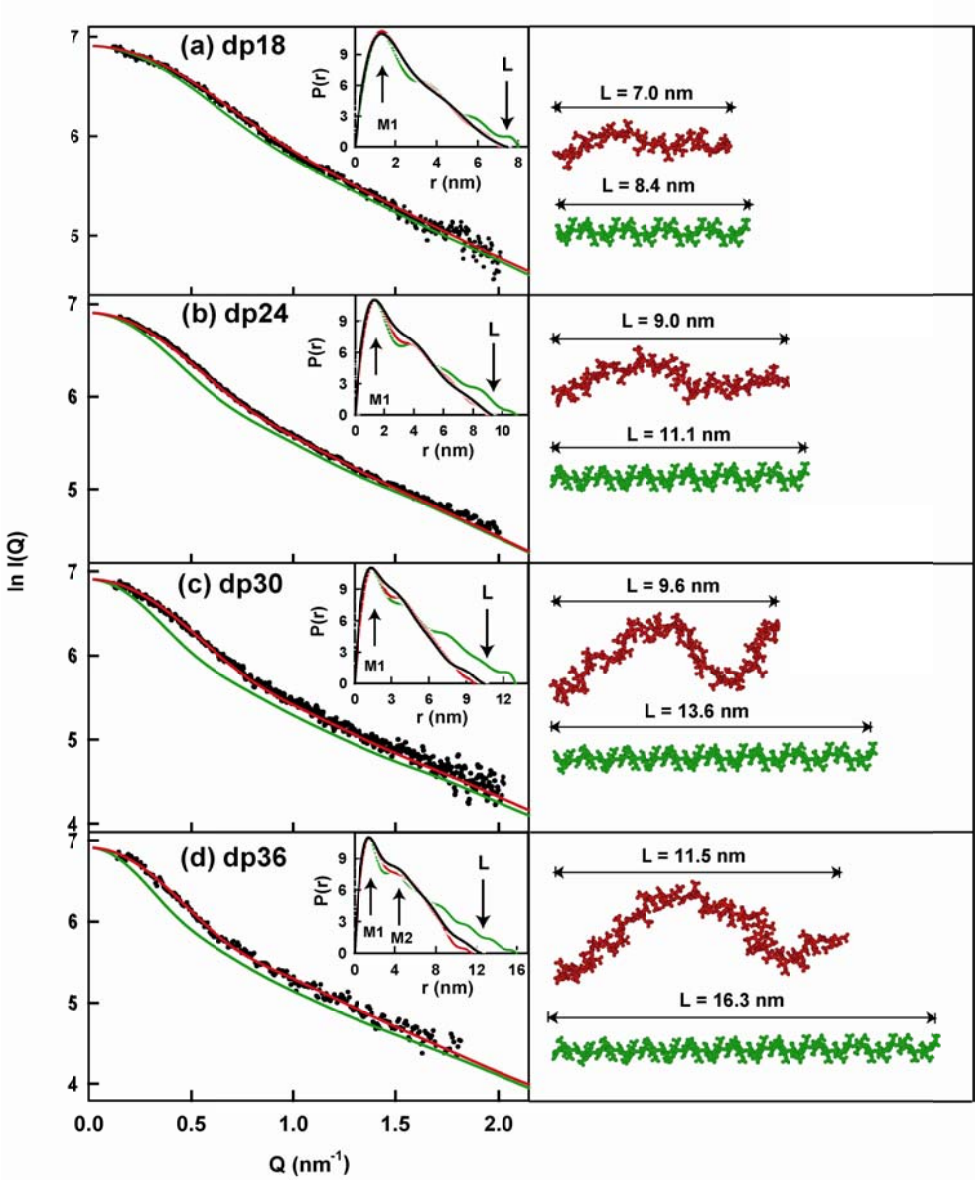


Figure 4.11

X-ray modelling curve fits for best-fit and poor-fit dp18–dp36 models. The main panels depict the $I(Q)$ curves, and insets show the $P(r)$ distance distribution functions. The experimental $I(Q)$ and $P(r)$ X-ray scattering data are represented by black circles or lines, respectively; the red lines and models correspond to the best-fit dp18–dp36 models from trial-and-error searches; and the green lines and models correspond to the linear poor-fit dp18–dp36 models from Figure 4.7. The best-fit and linear models are shown to the right, together with their maximum lengths for comparison with their L values in the $P(r)$ curves. (a–d) The four panels show the analyses for dp18, dp24, dp30 and dp36 in order.

The nine best-fit dp30 models gave *R*-factor values of 3.1–3.4%, R_G values of 2.79–2.96 nm, and *L* values of 10.2–11.8 nm. These correspond well with the experimental R_G value of 2.83 ± 0.14 nm and an *L* value of 10.3 nm (Table 4.1). The deviations from a linear dp30 model are larger, for which the *R*-factor is 6.3%, the R_G value is 3.41 nm, and the *L* value is 13.6 nm. The agreement between the experimental and the modelled *I*(*Q*) and *P*(*r*) curves was again excellent (Figure 4.11c). The nine modelled $s^0_{20,w}$ values of 1.49-1.86 S agree well with the experimental $s_{20,w}$ value of 1.63 ± 0.02 S, with the best-fit model giving a value of 1.86 S.

For dp36, similar good agreements between the models and the X-ray and AUC data were obtained, and the deviation obtained from a linear dp36 structure was larger. The nine best-fit dp36 models gave *R*-factor values of 3.5–4.0%, R_G values of 3.18–3.28 nm, and *L* values of 11.1–12.9 nm. The latter agrees well with the experimental R_G value of 3.12 ± 0.10 nm and an *L* value of 12.3 nm (Table 4.1), even though the availability of 6000 randomised models in this case reduced the quality of the fits in Figure 4.10d compared to those in Figure 4.10a and b. The best-fit models deviate from the parameters for the linear dp36 model, which have an *R*-factor of 5.3%, an R_G value of 3.82 nm, and an *L* value of 16.3 nm. The maxima *M1* and *M2* become more noticeable in the *P*(*r*) curve for the best-fit model (Figure 4.11d), and their *r* values agree with those seen experimentally. The nine modelled sedimentation coefficients calculated from the best fit dp30 models ranged between 1.72-1.95 S, in good agreement with the observed $s_{20,w}$ value of 1.84 ± 0.06 S, with the best-fit model giving a value of 1.87 S.

In conclusion, the best-fit models for dp18, dp24, dp30, and dp36 show progressively more bent structures in solution with increase in heparin size. Compared to the lengths of the linear models, these being the most extended ones possible, the four heparin fragments are reduced in length by 18%, 16%, 29%, and 29%, in that order. This overall conclusion applies to all the best-fit models in each of the four structure determinations. This is illustrated by the superimposition of the 9–15 best-fit models in each fragment (Figure 4.12), which showed that an increase

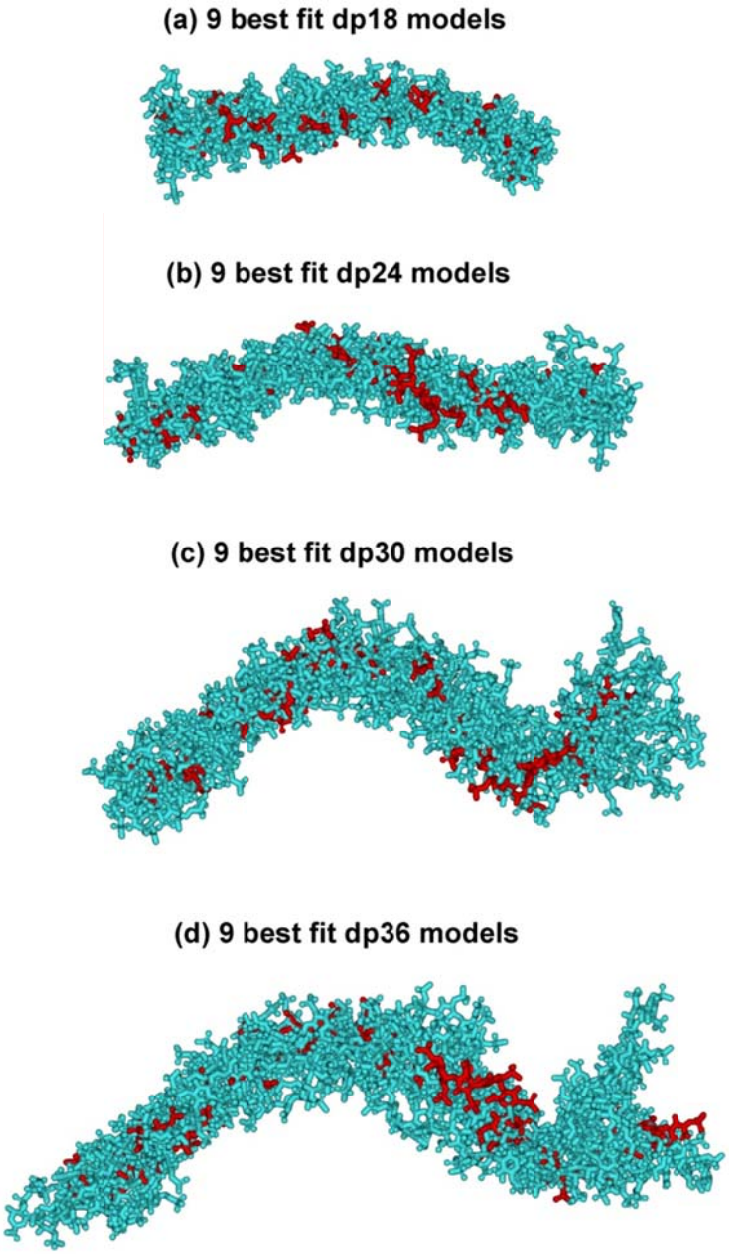


Figure 4.12

Superimposition of the 9–15 best-fit models for dp18–dp36. Each set of best-fit models for the four heparin fragments was superimposed globally using Discovery Studio VISUALISER software, then their non-hydrogen atoms were displayed as shown. The best-fit model from Figure 4.11 is shown in red, while the related best-fit structures are shown in cyan.

in bending with increase in size is common to all of them. In the case of dp30, and more so with dp36, the appearance of the minor peak *M2* in the experimental $P(r)$ curve is consistent with the appearance of more bending in the larger heparin fragments. In all four cases, the bent models showed $s^0_{20,w}$ values that are similar to those of the linear models, indicating that the AUC data are not able to distinguish between slightly bent and linear structures, unlike the X-ray scattering data.

(4.2.4) Comparison with heparin conformations in crystal structures

Given the current knowledge of 19 heparin-protein crystal structures (Section 4.4), we now survey the heparin conformations in these as well as that of the NMR dp12 solution structure. The two major conformations of six-membered iduronate rings in heparin are the $^1C_4$ chair and the $^2S_0$ skew-boat (Mulloy & Forster, 2000). The two forms can interconvert with relatively little disturbance to the glycosidic linkages to neighbouring residues, and the polysaccharide chain need not bend as a consequence, although the relative positions of sulphate groups within the ring will alter. In terms of the glycosidic linkages, information on their conformation in heparin is available from analyses of their Φ and Ψ angles (Figure 1). The first survey of five heparin-protein crystal structures (Mulloy & Forster, 2000) showed that the average Φ and Ψ angles were -75° and 135° (18 values for each) respectively for IdoA-GlcNS glycosidic bonds, whereas these were 78° and 103° (15 values each) respectively for GlcNS-IdoA glycosidic bonds (standard deviations of 10^o to 19^o). These five structures (Faham *et al.*, 1996; DiGabriele *et al.*, 1998; Fry *et al.*, 1999) thus show similar heparin geometries. Their similarity with that of the NMR structure of dp12 where the Φ and Ψ angles were -55° and 135° respectively for IdoA-GlcNS and 108° and 83° respectively for GlcNS-IdoA (Mulloy *et al.*, 1993) showed that the bound conformation is similar to that in solution (Table 4.2). The current availability of 19 heparin-protein crystal structures enabled this analysis to be extended (Figure 4.13a,b). After the removal of five pairs of outliers shown in Figure 4.13a,b, the average Φ and Ψ angles of IdoA-GlcNS glycosidic bonds were remarkably similar to previous at -79° and 132° (65 values each) respectively, whereas the average Φ and Ψ angles of GlcNS-IdoA glycosidic bonds were similar to previous at 84° and 100° (58 values each) respectively. Here, the standard deviations have increased to 19^o to 22^o. All five outliers in Figure 4.13a,b corresponded to the first or last residue of the bound

Table 4.2. Summary of the Φ and Ψ angle in heparin structures

	IdoA-GlcNS		GlcNA-IdoA	
	Φ (°)	Ψ (°)	Φ (°)	Ψ (°)
Five crystal structures (2000) [a]	-75 ± 10	135 ± 13	78 ± 16	103 ± 19
19 crystal structures (current study) [a]	-79 ± 20	132 ± 19	84 ± 22	100 ± 19
NMR structure	-55	135	108	83
dp18	-52 ± 21	135± 15	104 ± 27	96 ± 25
dp24	-61 ± 17	127 ± 20	100 ± 26	78 ± 17
dp30	-64 ± 21	136 ± 21	87± 25	82 ± 22
dp36	-63 ± 23	129 ± 27	103 ± 22	89 ± 24

[a] No outliers were included in these averages (see text).

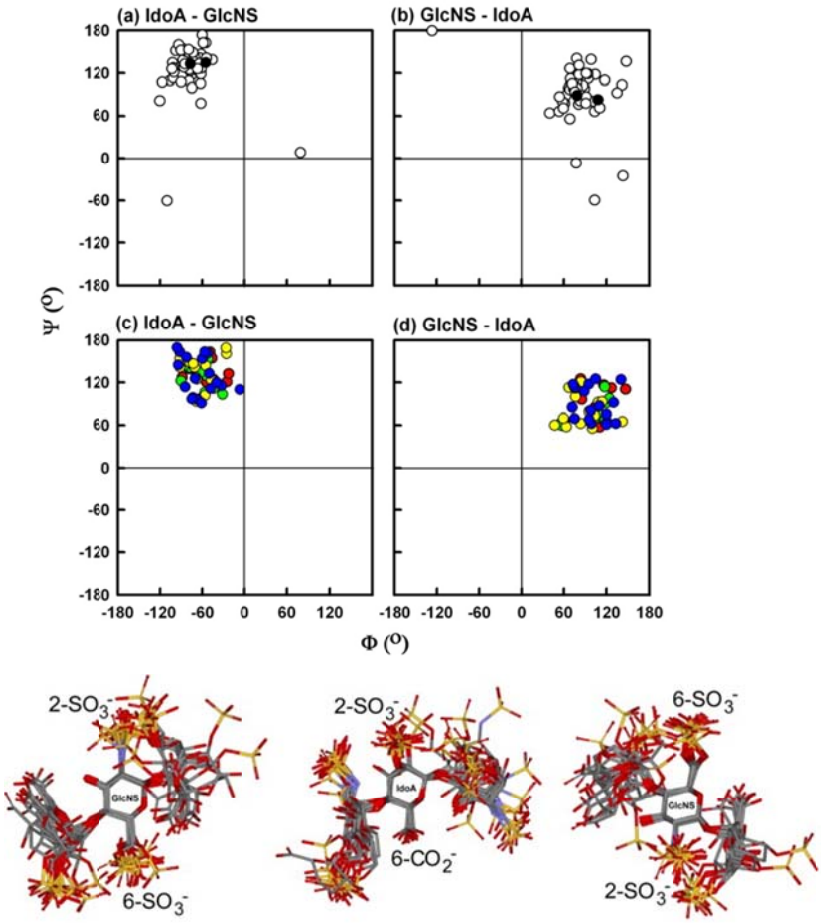


Figure 4.13

Φ and Ψ dihedral angles for heparin fragments. The atoms defining the Φ and Ψ angles are shown in Figure 4.1. (a and b) Sixty-seven pairs of IdoA–GlcNS and 62 pairs of GlcNS–IdoA Φ –Ψ angles found in 19 protein–heparin complexes are shown as open circles (PDB codes 1AZX, 1AXM, 1BFB, 1BFC, 1E0O, 1EO3, 1FQ9, 1GMN, 1GMO, 1G5N, 1RID, 1QQP, 1XMN, 2AXM, 2GD4, 2HYU, 2HYV, 3DY0, and 3EVJ). The corresponding 10 and 12 pairs of Φ–Ψ angles from two dp12 NMR solution structures are shown as filled circles (PDB code 1HPN). Note that many of the latter Φ–Ψ angles are similar, with values of 10°–10° or 12°–12°. (c and d) The corresponding Φ–Ψ angles from the best-fit model in Figure 4.10 for each of dp18 (8–9 values: red), dp24 (11–12 values: green), dp30 (14–15 values: yellow), and dp36 (17–18 values: blue) are shown in the same scale as that in (a) and (b). Beneath the graphs, the superimposition of dp5 fragments found in 12 or 13 heparin–protein crystal structures is shown. From left to right, the dp3 fragments are shown centered on the second (GlcNS), third (IdoA), and fourth (GlcNS) rings of dp5, each of which is viewed face-on. Gray, carbon; red, oxygen; yellow, sulphur; blue, nitrogen.

heparin structure. It appears likely that these terminal residues at the reducing or non-reducing ends have greater freedom to reorientate themselves in the crystal structures, and the outliers are not considered significant. Overall, it is concluded that heparin has a well-defined and similar extended conformation when bound to proteins.

Conformational variability was observed in the extended structures of bound heparin. While considerations of structural resolution and carbohydrate refinement parameters may preclude detailed analyses, such as the existence of skew boat or chair ring conformations, it should be possible to infer relative IdoA and GlcNS ring orientations. The commonly-occurring dp5 fragment in 12 or 13 of the 19 crystallographically-observed heparin structures were superimposed on the second, third and fourth rings of dp5 (bottom of Figure 4.13). The two rings flanking the central ring were in similar locations, as expected from Figure 4.13a,b, with only one instance of a 180° GlcNS ring flip (visible at the top of the central image in Figure 4.13). The sulphate and carboxylate groups in the two flanking rings were distributed in a range up to 0.5 nm, while this positional variability was reduced to less than 0.1 nm in the case of the central ring alone. The superimpositions showed that six or seven structures of the 19 exhibited clear bends in an otherwise extended heparin structure. For example, the longest-observed crystallised heparin dp10 is found in the 2:2 complex of fibroblast growth factor 1 and fibroblast growth factor receptor 2, in which dp10 links the four protein subunits together (Pelligrini *et al*., 2000). In the five IdoA-GlcNS glycosidic bonds, one Ψ angle was 78° in distinction to the average Ψ angle of 132°. In the four GlcNS-IdoA glycosidic bonds, two Ψ angles were 141° and 135° instead of the average Ψ angle of 84°. Consequently two significant bends were visible in the dp10 structure (Figure 4.14a). Heparin dp8 bound to vaccinia virus complement control protein assists the virus in its infection cycle (Ganesh *et al*., 2004) (Figure 4.14c). In the four IdoA-GlcNS glycosidic bonds, one Φ value is -106° in distinction to the mean value of -79° and one Ψ angle was 78° in distinction to the average Ψ angle of 132°. In the three GlcNS-IdoA glycosidic bonds, one Ψ angle was 135° instead of the average Ψ angle of 84°. Sometimes, the heparin conformation is not the same when binding to the same protein. For example, the crystal structure of four thrombin molecules complexed with two dp8 molecules showed that the two dp8 molecules are sandwiched between two different thrombin monomers in non-equivalent conformations (Carter *et al*., 2005).

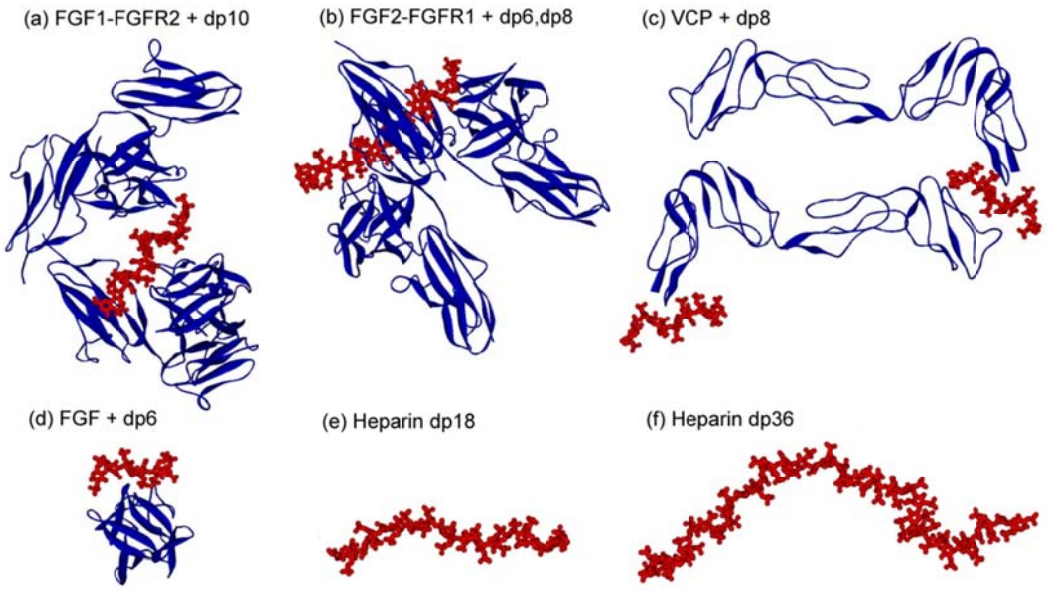


Figure 4.14

Comparison of selected crystal structures for heparin–protein complexes with the best-fit dp18 and dp36 solution structures. Proteins are shown by blue ribbons, while heparins are shown in red. The crystal structures are as follows. (a) The complex of fibroblast growth factor 1 and fibroblast growth factor receptor 2 (FGF1–FGFR2), to illustrate the longest cocrystallised heparin dp10 (PDB code 1E0O). The bending in dp10 should be noted. (b) The complex of fibroblast growth factor 2 and fibroblast growth factor receptor 1 (FGF2–FGFR1), to illustrate the binding of two heparin dp6 and dp8 molecules (PDB code 1FQ9). (c) Two molecules of VCP and two heparin dp8, to illustrate heparin cross-linking (PDB code 1RID). (d) Basic fibroblast growth factor (FGF) with dp6, to illustrate the surface location of heparin binding sites (PDB code 1BFC). (e and f) Heparin dp18 and dp36 from Figure 4.11, to illustrate the similarity of the crystal and solution structures for heparin.

The best-fit solution structures of dp18, dp24, dp30 and dp36 (Figure 4.11) can now be compared with the observed heparin conformations in crystal structures. The average Φ and Ψ angles of the IdoA-GlcNS glycosidic bonds were very similar at -52° and 135° (dp18: 9 values), -61° and 127° (dp24: 12 values), -64° and 136° (dp30: 15 values), and -63° and 129° (dp36: 18 values) (Figure 4.13c). While their standard deviations are higher at 15° to 23° (Table 4.2), no outliers were present. The Φ angles from scattering may be slightly reduced compared to the crystal structures, while being more similar to the NMR structure (Table 4.2). Likewise, the average Φ and Ψ angles of the GlcNS-IdoA glycosidic bonds were very similar at 104° and 96° (dp18: 8 values), 100° and 78° (dp24: 11 values), 87° and 82° (dp30: 14 values), and 103° and 89° (dp36: 17 values) (Figure 4.13d). No outliers were again present. The Φ angles from scattering may be slightly increased compared to the crystal structures, while being more similar to the NMR structure (Table 4.2). It is concluded that, despite the additional bending seen in dp36 compared to dp18, heparin is well-maintained in solution as a semi-rigid extended structure (Figure 4.12). Its conformations are similar to those seen in 19 heparin-protein crystal structures, while the dp12 NMR structure (Figure 4.7) may be too linear to correspond to heparin conformations in solution.

(4.3) Conclusions

In this study, we have here determined the first molecular structures in solution for heparin dp18–dp36. These provide new insight into the way that heparin (and, by analogy, heparan sulphate) interacts with its protein ligands. We show that heparin is well-represented as an extended and semi-rigid structure in solution. The dp6 and dp12 structures by both NMR (Mulloy *et al.*, 1993) and constrained scattering modelling (here) show comparatively little bending. Starting with dp18, the heparin fragments progressively show higher degrees of partial bending up to dp36, while maintaining approximately linear segments within these structures. The coordinate models have been deposited in the so-named Protein Data Bank at http://www.rcsb.org/pdb/. As far as we are aware, this study is the first successful application of constrained scattering modelling to oligosaccharides, with this method having previously been used to determine molecular structures for multidomain antibody and complement proteins (Perkins *et al.*, 2008, 2009). We envisage that this

modelling approach can be extended to structure determinations for other oligosaccharides and polynucleotides.

While solution scattering is traditionally a low-resolution method that generates shape envelopes, its utility to the determination of medium-resolution heparin solution structures is assisted by constrained modelling. Constrained modelling determines a three-dimensional molecular structure that best accounts for the observed scattering curve. It is reliable in defining the averaged extension and bending of the polysaccharide chains and in deriving information on flexibility. It cannot provide the resolution necessary to provide coordinates within the chain, even if averaged Φ–Ψ values are calculated from constrained modelling. Even though unique structure determinations are not possible due to random molecular orientations observed by scattering, this modelling is able to rule out structures that are incompatible with the scattering curves. Hence, the basic premises of constrained modelling are twofold: (i) by fixing parts of the structural analyses to what is already known about the macromolecule, scattering fits are subjected to significantly fewer modelling variables; and (ii) by identifying poor-fit models and rejecting these, good-fit models are obtained and ranked in order of their compatibility with the data. A related strategy resulted in the NMR solution structure of dp12 (Mulloy *et al.*, 1993). By NMR, the measurements of ^{1}H–^{1}H nuclear Overhauser enhancements for heparin permitted these to be tested against molecular models for heparin. Here, for the constrained scattering modelling of heparin, the only variables are the Φ and Ψ angles of each glycosidic bond. The randomised variation of these angles readily resulted in the V-shaped graphs of *R*-factor *versus* R_G values, from which the best-fit models were identified at the lowest *R*-factor values. In fact, the quality of the dp18–dp36 scattering curve fits is better than that obtained for other recent examples of multidomain proteins such as complement FH or complement receptors types 1 and 2 (Okemefuna *et al.*, 2009; Furtado *et al.*, 2008; Gilbert *et al.*, 2006). If constrained modelling is not employed, only a macroscopic view of heparin structures in solution analysed using worm-like sphere models is possible (Pavlov *et al.*, 2003; Perez Sanchez *et al.*, 2006; Ortega & Garcia de la Torre, 2007). The two approaches are conceptually different. In the worm-like model, flexibility is dynamic (fluctuating) and distributed continuously, while constrained modelling presents a family of static structures that fit the experimental X-ray data.

Our structural study of dp6–dp36 provides new insights on 19 crystal structures for heparin–protein complexes (Section 4.4). The crystal structures are notable for the observation of comparatively small heparin fragments that range in size from dp4 to dp10, with the most common being dp5. These sizes are sufficient to form well-defined heparin–protein complexes (Figure 4.14). The majority of contacts between the protein surface and heparin are ionic between basic amino acid side chains and anionic sulphate groups. The axial orientations of sulphate groups in heparin are repeated after every four oligosaccharide rings. Thus, dp4 possesses a sufficient number of at least six sulphate groups and two carboxylate groups to generate protein specificity for heparin. Within each dp4 fragment, two sulphate groups are located in similar but opposed axial orientations in GlcNS residues, while the third sulphate and the carboxylate group are located in opposed axial orientations in IdoA residues (Figure 4.7). Views of 12–13 superimposed dp3 fragments centred on either IdoA or GlcNS residues confirm that these axial orientations are largely preserved and approximately aligned with each other between IdoA and GlcNS residues (Figure 4.13). The current study shows that the solution structures of heparin in sizes of up to dp12 in size are generally linear, although not necessarily maintaining a regular helical repeat. Linearity may be influenced by the repulsion of the sulphate and carboxylate groups, which assist in this regard but are not mandatory (Mulloy *et al*., 1994). We show that, if the heparin fragments are dp18 or larger, bending of its linear structure becomes more pronounced, even though the mean Φ and Ψ angles remain unchanged (Table 4.2). In the constrained modelling, this bending appears as kinks in the heparin structure at a few positions. Thus, heparin appears as a succession of linear regions similar to dp6 or dp12 that are joined by semi-rigid kinks. The physical basis for the occurrence of kinks may result from the occasional occurrence of GlcA-GlcNAc sequences, and not IdoA–GlcNS sequences (Figure 4.1), but the crystal structures show that kinks also occur in other regions (Figure 4.14). Overall, our X-ray modelling of dp6–dp36 clarifies that large heparin segments in solution form extended stable and preformed conformations that are ideal for protein ligation, including the presence of occasional kinks in these solution structures.

This study opens the way to analysing more complex heparin–protein interactions, including the occurrence of two heparin binding sites on the same protein. Can the same heparin molecule bind simultaneously to a double heparin

binding site, or does heparin cross-link different proteins? The heparin-binding regulatory chemokine macrophage inflammatory protein 1α (MIP-1α) is one case in point (Stringer *et al*., 2002). Heparin binds strongly to dimeric MIP-1α, in distinction to monomeric or tetrameric MIP-1α. Heparinase III digests, and molecular modelling suggests that two dp14 fragments joined by a flexible dp6 fragment bind to dimeric MIP-1α. This conclusion is explained by our modelling of dp30 and dp36, which indicates the existence of linear segments joined by semi-rigid kinks. Physiologically, multiple interactions of this type will account for the packaging of heparin–protein complexes within mast cell granules. The binding of heparin dp26 to the dimer of interferon-γ represents a different case, where the interaction does not necessarily involve bent heparin structures (Perez Sanchez *et al*., 2006). In distinction to these cases, crystal studies show that heparin is able to cross-link monomers to form complexes (e.g., the fibroblast growth factor family and their receptors) (Pellegrini *et al*., 2000; Schelessinger *et al*., 2000). The complement regulator FH, which is composed of 20 short complement regulator (SCR) domains, is a distinct example altogether (Khan *et al*., 2008). The heparin–FH interaction at cell surfaces protects host cells from attack by the innate immune system, directing this instead against pathogenic bacteria, which often lack a polyanionic oligosaccharide coating. Two different heparin binding sites occur at the SCR-7 and SCR-20 domains in FH. Up to now, it is not clear whether flexibility within FH or heparin could permit SCR-7 and SCR-20 to bind simultaneously to the same heparin oligosaccharide. Our new dp18–dp36 structures indicate that heparin has insufficient flexibility to facilitate the simultaneous binding of two FH sites with one heparin. It is more likely that FH cross-links different heparin oligosaccharides to form very large oligomers (Khan *et al*., 2008). At least two different heparin sites may have evolved within FH in order to enable FH to bind selectively to host cells showing an appropriate abundance of polyanionic surfaces, in distinction to pathogens that lack these FH sites.

A further key question in relation to heparin–protein interactions is whether protein conformational changes occur on heparin binding. The availability of 19 heparin–protein crystal structures partly answers this question, being mainly limited by the relatively small sizes of the heparin fragments. Physiologically, heparin (and heparan sulphate) exists as long oligosaccharides. The largest crystallographically observed heparin to date is a decasaccharide (Pellegrini *et al*., 2000). Even when a

heparin 14-mer was cocrystallised with a protein, the largest crystallographically observed heparin was as small as a pentasaccharide dp5 (Lietha *et al.*, 2001). Heparin may induce conformational changes in proteins in order to alter activities. The best-known case is that of antithrombin, whose β-sheet structure is allosterically activated by heparin in a series of events in order to inhibit the coagulation factors IXa and Xa and thrombin (Langdown *et al.*, 2009). In most crystal structures, heparin binding sites are found on surface-exposed positions (Figure 4.14d), and conformational changes do not appear to be required for complex formation. The case of FH is distinct from these two scenarios. Each SCR domain is stabilised by two conserved disulphide bridges at each end of its β-sheet structure and a hydrophobic core that includes a conserved Trp residue (Saunders *et al.*, 2007). A single SCR domain will accommodate heparin dp8 without conformational change (Ganesh *et al.*, 2004). However the linker conformations between SCR domains are variable and not easily predicted (Perkins *et al.*, 2002). Flexibility in inter-SCR linker conformations has been demonstrated by the different solution and crystal structures of SCR-1/2 in complement receptor type 2, in which the two SCR domains are joined by a long eight-residue linker (Gilbert *et al.*, 2005). If heparin binds to two neighbouring SCR domains, heparin will need to be at least dp18 in size to do so, and its binding may alter the inter-SCR conformation. An initial study of dp10 binding to a recombinant SCR-6/8 fragment of FH showed that either a conformational change or oligomer formation resulted (Fernando *et al.*, 2007). By this present study, the improved understanding of dp18–dp36 solution structures opens the way to clarifying how FH binds to heparin, including the consideration of heparin-induced conformational change in its domain structure (Khan *et al.*, 2008).

(4.4) Materials and methods

(4.4.1) Purification of heparin fragments

Oligosaccharide fragments of heparin were prepared from both from the low molecular weight heparin Tinzaparin (Leo Pharma, Denmark), for preliminary work, and from bovine lung heparin pre-digested with heparinase I provided by my collaborator Dr. Barbara Mulloy (National Institute of Biological Standards and Control, Blanche Lane, South Mimms, Potters Bar, Hertfordshire, UK). This digestion procedure has been previously described (Pellegrini *et al.*, 2000; Mulloy *et*

al., 1997; Rice *et al*., 1985). In summary, about 80 mg of bovine lung heparin (the 2nd International Standard Heparin) was taken up in 2 ml 20 mM phosphate buffer pH 7.4, and 100 µl heparinase I solution (the kind gift of Leo Pharma, Ballerup, Denmark, used as supplied without assay) was added. The digestion was allowed to proceed for 4 min at room temperature and the reaction stopped by heating in a water bath at 100ºC. After evaporation of the solution under reduced pressure at 50ºC, the digest was taken up in 1.5 ml of 2% ammonium bicarbonate, filtered through a syringe-mounted 0.45 micron membrane, and then applied to a preparative gel-permeation chromatography column (100 cm × 1.6 cm, packed with Biogel P-10; Bio-Rad, UK). The heparin fragments were eluted using 2% ammonium bicarbonate at a flow rate of 0.2 ml/min and collected in 2-ml fractions. The absorbance of the fractions at 234 nm was measured, and the fractions corresponding to each resolved peak were pooled. Heparin oligosaccharides larger than dp12 were not completely resolved (Figure 4.2). Each fraction was evaporated under reduced pressure and lyophilised before assessment of its size by analytical gel-permeation chromatography (Mulloy *et al*., 1997). This was carried out using two columns (TSK G3000 SW-XL, 30 cm; TSK G2000 SW-XL, 30 cm; Anachem, UK) connected in series (Figure 4.3). The eluant was 0.1 M ammonium acetate solution at a flow rate of 0.5 ml/min, and heparin was detected with a refractive index detector (RI-1530; Jasco, UK). The chromatography system was calibrated using the First International Reference Reagent Low Molecular Weight Heparin for Molecular Weight Calibration (NIBSC 90/686). Heparin quantification was achieved by integration of the area under each refractive index peak and comparison with a standard curve prepared using known concentrations of low-molecular-weight standards.

(4.4.2) Analytical ultracentrifugation of heparin fragments

Sedimentation velocity experiments were performed using two Beckman XL-I analytical ultracentrifuges (Beckman Coulter, Inc., Palo Alto, CA) equipped with both absorbance and interference optics. Experiments with dp6–dp36 fragments were performed at concentrations of 0.5 mg/ml in 10 mM Hepes and 137 mM NaCl (pH 7.4). Buffer density (1.00480 g/ml) was measured at 20°C using an Anton-Paar DMA5000 density meter. A partial specific volume of 0.467 ml/g was used for heparin (Pavlov *et al*., 2003). Runs were carried out in an eight-hole AnTi50 rotor with standard double-sector cells with column heights of 12 mm at 20°C using

absorbance optics at 234 nm and interference optics. Sedimentation velocity data were collected at 40,000 rpm and 50,000 rpm using absorbance optics, and at 40,000 rpm, 50,000 rpm, and 60,000 rpm using interference optics. The continuous *c*(*s*) analysis method was used to determine the sedimentation coefficients $s_{20,w}$ of heparin fragments using SEDFIT software (version 9.4) (Dam & Schuck, 2004; Schuck, 2000). The *c*(*s*) analysis directly fits the sedimentation boundaries using the Lamm equation, the algorithm for which assumes that all species have the same frictional ratio f/f_o in each fit. The final SEDFIT analyses used a fixed resolution of 200 and optimised the *c*(*s*) fit by floating the meniscus and cell bottom when required and by holding the f/f_o value, baseline, and cell bottom fixed until the overall root-mean-square deviations and visual appearance of the fits were satisfactory (Figure 4.3, 4.4 and 4.5). The f/f_o values for the heparin fragments were calculated to be 1.33 for dp6, 1.39 for dp12, 1.64 for dp18, 1.78 for dp24, 1.82 for dp30, and 1.88 for dp36, starting from the known volume of each fragment (Perkins, 1986).

(4.4.3) X-ray scattering of heparin fragments

X-ray solution scattering of the six heparin fragments (dp6, dp12, dp18, dp24, dp30, and dp36) was performed on beamline ID02 at the European Synchrotron Radiation Facility (Grenoble, France) in two sessions with a ring energy of 6.0 GeV (Narayanan *et al*., 2001). In the first session, data were collected for the six fragments in 16-bunch mode using beam currents from 79 mA to 86 mA. In the second session, data were collected for the six fragments in four-bunch mode using beam currents from 30 mA to 43 mA. Data were acquired using an improved fibre optically coupled high-sensitivity and dynamic-range charge-coupled device detector (FReLoN) with a smaller beamstop. The sample-to-detector distance was 2.0 m. Experiments used the same heparin concentration (0.5 mg/ml) and buffers as used for the sedimentation velocity experiments. For each heparin fragment, samples were measured in a flow cell, which moved the sample continuously during beam exposure in 10 time frames with different exposure times of 0.1 s, 0.2 s, 0.25 s, and 0.5 s to check for the absence of radiation damage effects. This exposure was optimised using on-line checks for the absence of radiation damage to show that this was not detectable, after which the frames for 0.5 s were averaged in order to maximise signal-to-noise ratios.

Guinier analyses give the radius of gyration R_G, which characterises the degree of structural elongation in solution if the internal inhomogeneity of scattering within the macromolecules has no effect. Guinier plots at low Q values give the R_G and the forward scattering at zero angle $I(0)$: (Glatter & Kratky, 1982)

$$\ln I(Q) = \ln I(0) - R_G^2 Q^2/3$$

This expression is valid in a $Q.R_G$ range of up to 1.5. If the structure is elongated (i.e., rod-shaped), the radius of gyration of the cross-sectional structure R_{XS} and the mean cross-sectional intensity at zero angle $[I(Q)Q]_{Q\to 0}$ parameters are obtained from:

$$\ln [I(Q).Q] = \ln [I(Q).Q]_{Q\to 0} - R_{XS}^2 Q^2/2$$

The R_G and R_{XS} analyses were performed using an interactive PERL script program SCTPL7 (J. T. Eaton and S. J. Perkins, unpublished software) on Silicon Graphics O2 Workstations. Indirect Fourier transformation of the full scattering curve $I(Q)$ in reciprocal space gives the distance distribution function $P(r)$ in real space. This yields the maximum dimension of the macromolecule L and its most commonly occurring distance vector M in real space:

$$P(r) = \frac{1}{2\pi^2}\int_o^\infty I(Q) Qr \sin(Qr)\, dQ$$

The transformation was carried out using GNOM software (Semenyuk & Svergun, 1991). For dp6, the full X-ray $I(Q)$ curves contained 506 data points in the Q range between 0.20 nm^{-1} and 0.89 nm^{-1}. For dp12, 171 data points were used in the Q range between 0.24 nm^{-1} and 1.40 nm^{-1}. For dp18, 571 data points were used in the Q range between 0.12 nm^{-1} and 2.0 nm^{-1}. For dp24, 280 data points were used in the Q range between 0.08 nm^{-1} and 1.0 nm^{-1}. For dp30, 600 data points were used in the Q range between 0.11 nm^{-1} and 2.18 nm^{-1}. For dp36, 252 data points were used in the Q range between 0.11 nm^{-1} and 1.81 nm^{-1}.

(4.4.4) Molecular modelling of heparin oligosaccharides

To analyse the conformation of free and protein-bound heparin oligosaccharides available in the PDB, the glycosidic dihedral angles Φ and Ψ corresponding to the torsion angles O5′–C1′–O4–C4 and C1′–O4–C4–C3, respectively (Figure 4.1), were calculated for all the IdoA–GlcNS and GlcNS–IdoA disaccharides using INSIGHT II 98.0 molecular graphic software (Accelrys, San Diego, CA) on Silicon Graphics O2 and OCTANE Workstations. The solution NMR structure of free dp12 was used (PDB code 1HPN). The following 19 heparin–protein crystal structures were identified from database searches of the PDB http://www.rcsb.org/pdb/, biologically active dimer of acidic fibroblast growth factor (PDB codes 1AXM and 2AXM), basic fibroblast growth factor (PDB codes 1BFB and 1BFC), complex of fibroblast growth factor 1 and fibroblast growth factor receptor 2 (PDB code 1E0O), complex of fibroblast growth factor 2 and fibroblast growth factor receptor 1 (PDB code 1FQ9), antithrombin (PDB codes 1AZX, 1EO3, and 3EYJ), thrombin (PDB code 1XMN), antithrombin–S195A complex with factor Xa (PDB code 2GD4), neurokinin 1 (PDB codes 1GMN and 1GMO), annexin V (PDB code 1G5N), foot-and-mouth disease virus (PDB code 1QQP), VCP (PDB code 1RID), annexin A2 (PDB codes 2HYU and 2HYV), and cleaved protein C inhibitor (PDB code 3DY0). The list is not exhaustive, and other PDB files (PDB codes 1NQ9, 1SR5, 1TB6, 1XT3, 1ZA4, 2BRS, and 2B5T) also correspond to heparin–protein crystal structures. The majority of these were not analysed for this study because these involve heparin disaccharides or synthetic heparin mimetics..

Linear models of heparin dp6 to dp36 were constructed using the NMR structure of dp12 (Mulloy *et al.*, 1993) in the $^{2}S_{0}$ configuration with INSIGHT II 98.0 software. The superimposition of four dp12 molecules was used to create a dp42 oligosaccharide, from which the six heparin fragments dp6–dp36 were obtained by removing non-required disaccharides. Here, the Φ and Ψ angles were − 55° and 135°, respectively, for IdoA–GlcNS, and 108° and 83°, respectively, for GlcNS–IdoA (Figure 4.13). Totals of 5,000 conformational models for dp18, 5,500 for dp24 and 6,000 each for dp30 and dp36 were generated starting from these linear models. The Φ and Ψ angles were randomised to take values in the maximum range of − 45° or 45° from their starting values using the TorsionKick function in a PERL script that was modified from the ExtractAngle.pl script provided with Discovery Studio molecular graphics software (version 2.1) (Accelrys). For example, for dp36,

a total of 18 Φ and Ψ angles for IdoA–GlcNS and a total of 17 Φ and Ψ angles for GlcNS–IdoA were modified in this way.

(4.4.5) Constrained scattering and sedimentation coefficient modelling

Each heparin model was used to calculate X-ray scattering curves for comparison with the experimental curves using Debye sphere models (Okemefuna *et al*., 2009; Furtado *et al*., 2008; Gilbert *et al*., 2006). A cube side length of 0.513 nm, in combination with a cutoff of four atoms, was used to create the spheres in the heparin models of dp18–dp36. The hydration shell corresponding to 0.3 g/g H_2O was created by adding spheres to the unhydrated sphere models using HYPRO (Ashton *et al*., 1997), where the optimal total of hydrated spheres is listed in Table 4.1. The X-ray scattering curve $I(Q)$ was calculated using the Debye equation as adapted to spheres and assuming a uniform scattering density for the spheres: (Perkins & Weiss, 1983).

$$\frac{I(Q)}{I(0)} = g(Q)\left(n^{-1} + 2n^{-2}\sum_{j=1}^{m} A_j \frac{\sin Qr_j}{Qr_j}\right)$$

$$g(Q) = \left(3(\sin QR - QR\cos QR)\right)^2 / Q^6 R^6$$

where $g(Q)$ is the squared form factor for the sphere of radius r, n is the number of spheres filling the body, A_j is the number of distances r_j for that value of j, r_j is the distance between the spheres, and m is the number of different distances r_j. Other details are given elsewhere (Perkins, 2001; Okemefuna *et al*., 2009; Furtado *et al*., 2008; Gilbert *et al*., 2006). X-ray curves were calculated without instrumental corrections, as these are considered to be negligible for the pinhole optics used in synchrotron X-ray instruments. First, the number of spheres N in the dry and hydrated models after grid transformation was used to assess steric overlap between the heparin disaccharides, where models showing less than 95% of the optimal totals were discarded (Table 4.1). Next, the models were assessed by calculation of the X-ray R_G values in the same Q ranges used for the experimental Guinier fits in order to allow for any approximations inherent in the use of the $Q.R_G$ range of up to 1.5. Models that were filtered using the N and R_G filters were then ranked using a goodness-of-fit R-factor in order to identify the best-fit nine models.

Sedimentation coefficients $s^0_{20,w}$ for the best-fit scattering models were calculated directly from molecular structures using the HYDROPRO shell modelling program, with the default value of 0.31 nm for the atomic element radius for all atoms in order to represent the hydration shell (Garcia de la Torre, 2000).

(4.4.6) Protein Data Bank accession number

The best-fit dp18–dp36 models have been deposited in the Protein Data Bank with accession codes 3IRI (dp18), 3IRJ (dp24), 3IRK (dp30), and 3IRL (dp36).

Chapter Five

The solution structure of heparan sulphate differs from that of heparin: Implications for function

(5.1) Introduction

Heparan sulphate (HS) is a highly sulphated glycosaminoglycan molecule and is found exclusively on the host cell surfaces and in extracellular surfaces (Gallagher *et al*., 1992; Bernfield *et al*., 1992). HS has key roles in biological recognition processes at the cell-tissue-organ interface through its interactions with a wide range of proteins (Conrad, 1998; Sasisekharan *et al*., 2000). Specific interactions involving HS include in cell growth and development (Perrimon *et al*., 2000), cell adhesion (Scarpellini *et al*., 2009), inflammation and wound healing (Stringer and Gallagher, 1997), angiogenesis and cancer (Sasisekharan *et al*., 1997 and 2002; Liu *et al*., 2002), viral invasion (Chen *et al*., 1997, Shukla *et al*., 1999) and anticoagulation (Petitou *et al*., 1999). The breadth of HS-protein interaction offers many potential strategies for therapeutic intervention at the cell-tissue-organ interface.

HS is a sulphated polysaccharide consisting of uronic acid and D-glucosamine residues linked by α-(1-4) glycosidic bonds (Figure 5.1) (Lane and Lindahl, 1989; Capila & Linhardt, 2002). The uronic acid residue is either unmodified β-D-glucuronic acid (GlcA), alternating with *N*-acetylated glucosamine (Figure 5.1a); or α-L-iduronic acid (IdoA), often 2-O-sulphated, alternating with *N*-sulphated glucosamine (GlcNS) (Figure 5.1b). In the latter, sulphation often occurs at C6 and rarely also at C3 (Coombe *et al*., 2005; Lyon & Gallagher, 1998). Unlike heparin, HS has a distinct domain organisation that is comprised of small *S*-domains with IdoA2S and GlcNS residues, long *NA*-domains with GlcA and GlcNAc residues, and mixed regions with residues from both *S* and *NA*-domains (Capila and Linhardt, 2002; Coombe *et al*., 2005). The *S*-domains and mixed domain regions are termed the hypervariable regions that result in different functional characteristics for HS from different cells (Coombe *et al*., 2005).

Three-dimensional structural studies of HS are required in order to complete the understanding of the physiological significance of HS-protein interactions. Many structural studies have been carried out for heparin which is an analogue for HS but possesses a higher degree of sulphation, and for at least 19 heparin-protein co-crystal complexes. This abundance results because of the ease with which heparin is obtained and its strong binding to many of the cell surface proteins whose physiological ligand is HS. Thus an NMR structure is known for heparin (Mulloy *et al*., 1993). Solution structures are known for six heparin dp6-dp36 forms from

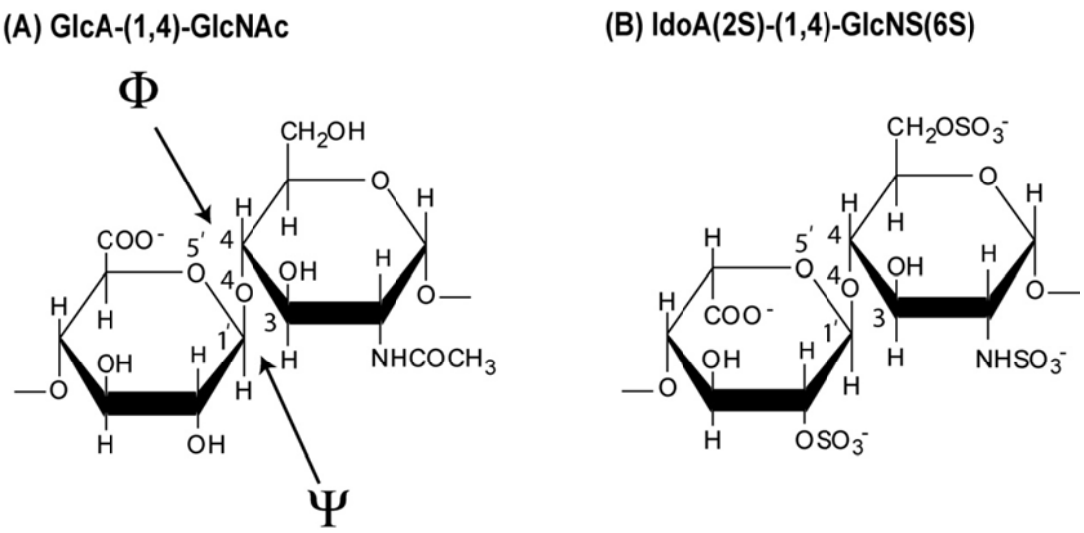


Figure 5.1

Chemical structures of two common disaccharide repeats of HS and heparin. (a) The major repeating disaccharide unit of HS (glucuronic acid -> *N*-acetylglucosamine). The NH.CO.CH3 group in the second ring is replaced by $NH.SO_3^-$ in 50% of this structure. The resulting molecular mass of this averaged disaccharide is 483 Da. (b) The minor repeating unit of HS, this being the major repeating disaccharide unit in 90% of heparin (iduronic acid-2-sulphate -> glucosamine-2,6-disulphate). For this study, heparin is considered to be 50% in the trisulphate form as shown and 50% in a disulphate form where a sulphate group is lost. The resulting molecular mass of this averaged disaccharide is 628 Da.

constrained scattering modelling; these were shown to be similar in conformation to heparin when observed in heparin-protein crystal structures (Khan *et al*., 2010). In distinction, no molecular structures for free HS are known, and only one crystal structure for a dp4 HS oligosaccharide complexed with heparinase II is available (Shaya *et al*., 2010).

Given the importance of understanding the HS solution structure, we have used a multidisciplinary approach to determine molecular structures for HS based on the combination of three methods, namely analytical ultracentrifugation, small angle X-ray scattering, and constrained scattering modelling (Perkins *et al*., 2008, 2009). This approach is well-established for solution structure determinations of large multidomain complement and antibody proteins, and was recently applied to small heparin oligosaccharide fragments (Khan *et al*., 2010; Bonner *et al*., 2009). Here, we apply this approach for the second time for oligosaccharide solution structures, this time for eight HS fragments ranging in sizes from dp6 to dp24, thus permitting detailed comparisons with heparin. The HS fragments exhibited solution structures that were distinct from those of the heparin fragments. In particular, their overall lengths are longer compared to heparin, and their structures display a greater degree of bending with increase in size compared to heparin. Our results are attributed to the difference in monosaccharide sequence between HS and heparin fragments, combined with a much reduced degree of sulphation in the HS fragments, which possessed greater structural flexibility than heparin. These results provided new insight on the potential binding modes of HS to proteins.

(5.2) Results and discussion

(5.2.1) Sedimentation velocity data analysis for the HS fragments

Eight oligosaccharide fragments of HS dp6 to dp24 were prepared in collaboration with Dr. Barbara Mulloy (National Institute of Biological Standards and Control, Blanche Lane, South Mimms, Potters Bar, Hertfordshire, UK) through Biogel P-10 column (Section 5.4). The purification profile from the Biogel P-10 column shows that the four smallest HS fragments dp6-dp8 were eluted as well resolved peaks, while the four larger fragments dp14-dp18 and dp24 were less well resolved (Figure 5.2). Analytical high performance size exclusion chromatography of the HS oligosaccharide fractions (Figure 5.3) as described (Mulloy *et al*., 1997), and polyacrylamide gel electrophoresis (PAGE) were performed to show that all eight

peak fractions showed altered sizes as expected, and were relatively homogenous in size (Figure 5.2 inset).

Analytical ultracentrifugation (AUC) studies macromolecular structures and sizes through quantitative measurements of sedimentation rates in a high centrifugal field (Cole *et al*., 2008). Sedimentation velocity experiments at three rotor speeds were performed for the eight HS fragments (dp6-dp18 and dp24) to determine their shapes and degree of polydispersity. The sedimentation coefficient distribution function *c(s)* was calculated by direct fitting of the sedimentation boundaries using SEDFIT software. The absorbance optics analyses for each HS fragment reproducibly resulted in good boundary fits that resulted in single major peaks (Figure 5.4a). The mean sedimentation coefficient $s_{20,w}$ values at three speeds ranged from 0.82 ± 0.05 S for dp6 to 1.26 ± 0.06 S for dp24. The corresponding interference optics analyses for dp6-dp24 also resulted in good boundary fits and single major *c(s)* peaks with mean $s_{20,w}$ values that ranged from 1.05 ± 0.04 S for dp6 to 1.35 ± 0.04 S for dp24 (Figure 5.4b). The number of peaks and their widths assess the polydispersity of each HS fragment. In this regard, both the absorbance and interference optics showed better resolution and single narrower major peaks when compared with the equivalent data sets for heparin (Chapter 4: Figure 4.5). The most likely explanation for this appears to be variability in the sulphate content within the heparin fragments, causing variations in mass that resulted in a broader peak width. This effect would not be present in HS because of the reduced sulphation level in HS. The similar single peak widths from either absorbance or interference optics suggest that all eight HS fragments showed narrow size distributions and are relatively homogenous, in agreement with the chromatography results above. The smaller HS fragments show slightly broader peak widths than the larger ones and this is attributed to a higher back diffusion effect. Unlike heparin, a small decrease in the $s_{20,w}$ values of the absorbance and interference data was observed with increase of rotor speed; this indicated that the $s_{20,w}$ values depend on the rotor speed (Figure 5.5a,b). Inspection of the boundary fits showed that the rotor-speed dependence resulted from a contribution from back-diffusion in the *c(s)* fits that was reduced with increase in rotor speed.

Like heparin, the HS analyses revealed $s_{20,w}$ values that increased with increase in size of the fragments (Figure 5.5a,b). A typical molecular mass of the most abundant HS dp2 structure shown in Figure 5.1a is 483 Da, while the

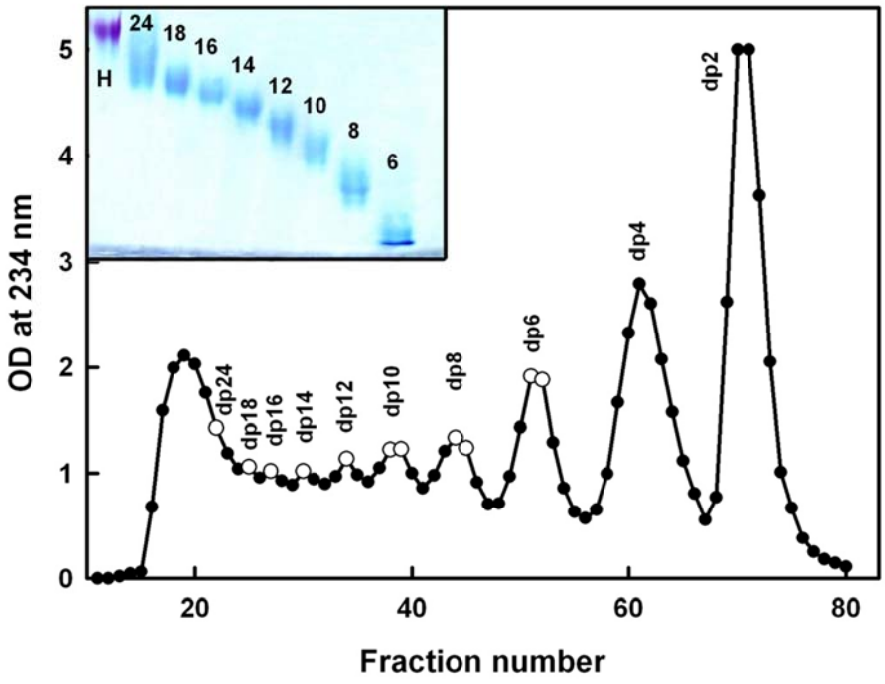


Figure 5.2

Purification profile of the HS fragments. The HS fragments were eluted with a flow rate of 0.2 ml/min using a Biogel P-10 column in 2% ammonium bicarbonate solution. Fractions of 2 ml/10 min were collected, and their HS concentrations were measured spectrophotometrically at 234 nm. The fractions taken for this study are shown by open circles. The inset shows 25% polyacrylamide gel electrophoresis (PAGE) of the HS fragments dp6-dp24 labelled as 6-24 with heparin dp24 (H) as marker. AUC (Figure 5.4) showed that these fractions are corresponded to relatively homogenous fragments of dp6-dp24, accordingly they are described as "fragments" in this chapter.

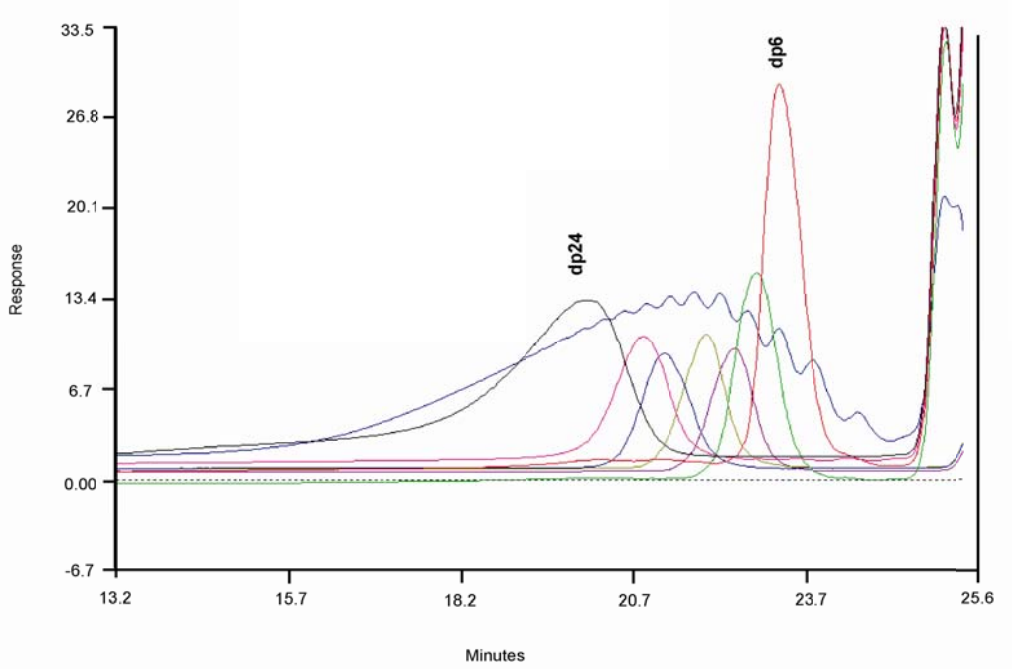


Figure 5.3

Size determination analyses of HS fragments through analytical gel permeation chromatography. The HS fragments from dp6 (right) up to dp24 (left) were eluted as peaks using two columns (TSK G3000 SW-XL, 30 cm; TSK G2000 SW-XL, 30 cm; Anachem, UK) connected in series. HS dp14 is not shown. The eluant was 0.1 M ammonium acetate solution at a flow rate of 0.5 ml/min, and HS was detected with a refractive index detector. The black curve corresponds to low molecular weight heparin standards.

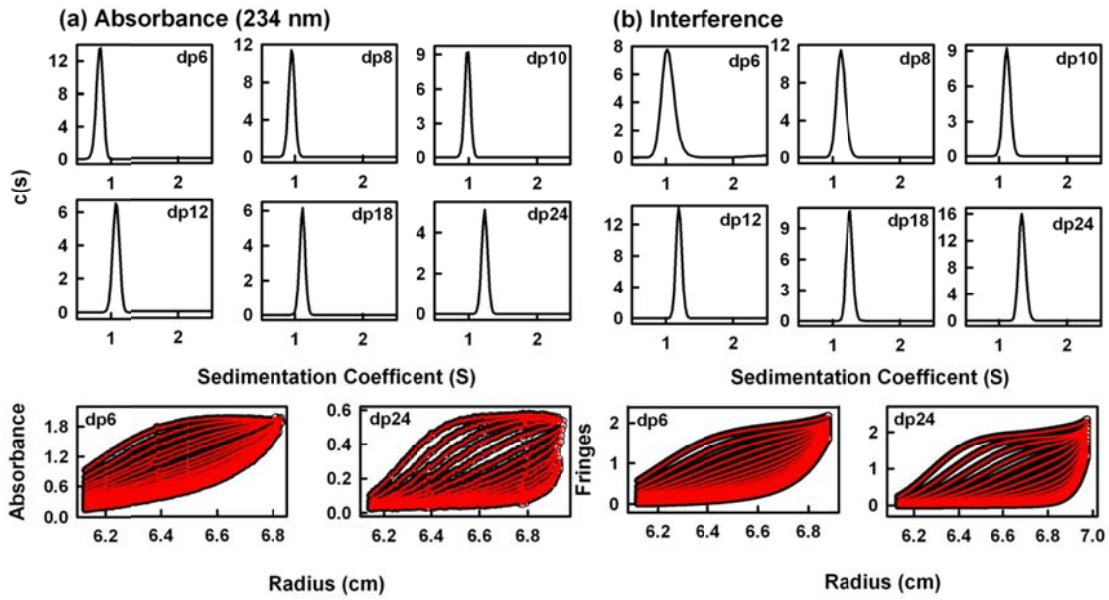


Figure 5.4

Sedimentation velocity size distribution analyses *c(s)* of six HS dp6-dp24 fragments. The (a) absorbance and (b) interference boundary scans were fitted using SEDFIT software for the HS fragments, each at 0.5 mg/ml. The mean $s_{20,w}$ and their standard deviations are reported in Table 5.1. (a) The absorbance data using a wavelength of 234 nm and a rotor speed of 50,000 rpm gave $s_{20,w}$ peaks at 0.84 S for dp6, 0.95 S for dp8, 0.98 S for dp10, 1.08 S for dp12, 1.11 S for dp18, and 1.23 S for dp24. Beneath these panels, representative boundary fits are shown for dp6 and dp24, in which only every sixth scan of the 120 fitted boundaries are shown for clarity. (b) The interference data using a rotor speed of 50k rpm gave $s_{20,w}$ peaks at 1.04 S for dp6, 1.12 S for dp8, 1.11 S for dp10, 1.19 S for dp12, 1.25 S for dp18, and 1.34 S for dp24. Beneath these panels, representative boundary fits are shown for every sixth scan of the 120 fitted boundaries for dp6 and dp24. This figure was prepared in collaboration with Miss Rima Patel under my supervision.

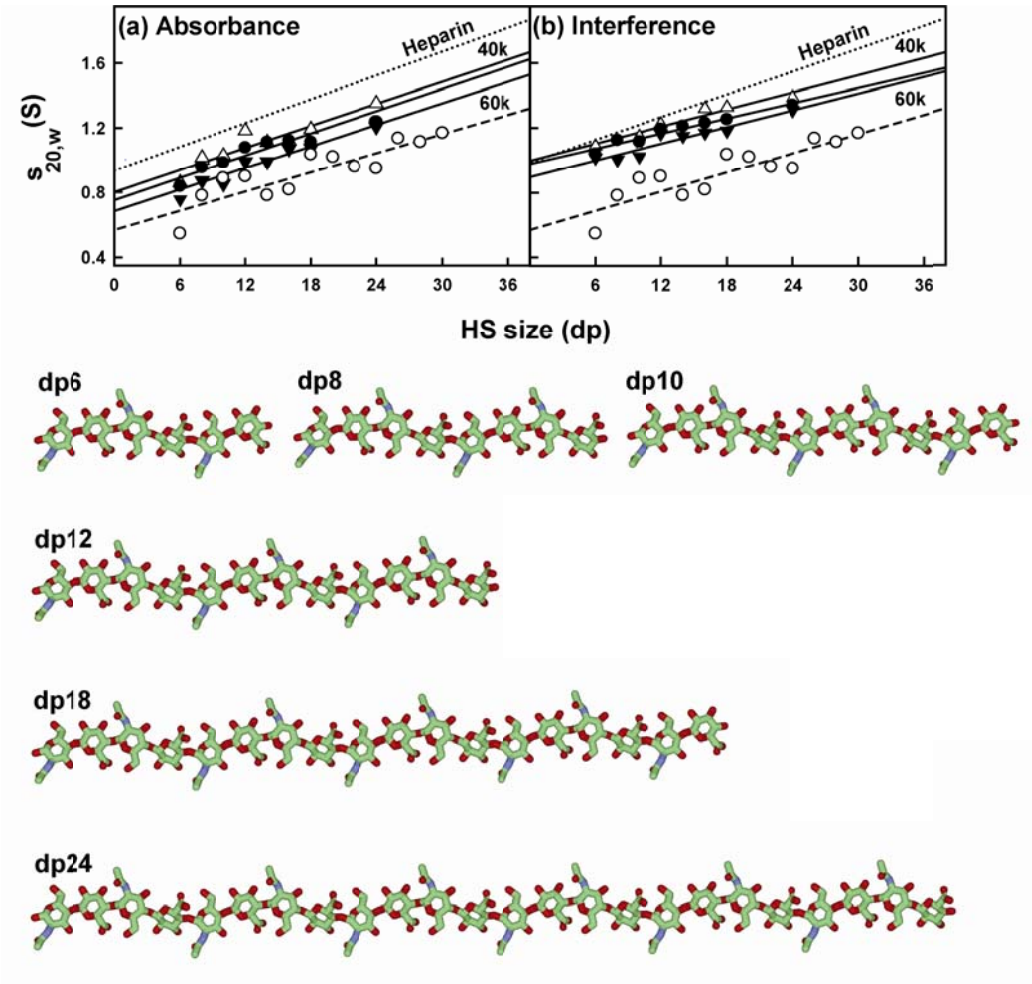


Figure 5.5

Comparison of the experimental and predicted sedimentation coefficients for eight HS dp6-dp24 fragments. The filled circles and triangles represent the 29 experimental values for dp6-dp24. The open circles represent the predicted values for linear dp6-dp30 models. (a) Comparison with the experimental sedimentation coefficients at rotor speeds of 40,000 (Δ), 50,000 (●) and 60,000 (▼) rpm using absorbance optics. (b) Comparison with the experimental sedimentation coefficients at rotor speeds of 40,000 (Δ), 50,000 (●) and 60,000 (▼) rpm using interference optics. (c) The linear models for HS dp6-dp24 that were created starting from the HS dp4 crystal structure (PDB code: 3E7J) are shown. The comparison with heparin is shown by a dotted line.

corresponding value for heparin dp2 is 628 Da (Chapter 4: Figure 4.1a). Because of the differences in molecular size, the Svedberg equation predicts that the $s_{20,w}$ values of the HS fragments will be 77% of those for the equivalent sized heparin fragments. The mean $s_{20,w}$ value for HS dp24 is 1.30 ± 0.06 S (Table 5.1), which is 86 ± 8 % of the corresponding value for heparin dp24 of 1.52 ± 0.07 S (Chapter 4: Table 4.1). This is almost within error of the mass-predicted reduction in $s_{20,w}$ value. If real, the difference between the 77% and 86% values would correspond to a 10% smaller frictional coefficient for HS compared to heparin, i.e. HS may have a slightly more compact solution structure than heparin.

The sedimentation coefficient $s^0_{20,w}$ values were calculated using HYDROPRO software from molecular models of HS. For this, 13 linear HS models (dp6 to dp30) were computed starting from the HS dp4 crystal structure seen in its complex with heparinise II (Shaya *et al.*, 2010). The HYDROPRO software program is used to calculate $s^0_{20,w}$ values directly from atomic coordinates (Chapter 3: section 3.4.6). Even though the increase of the $s^0_{20,w}$ values with size was predicted correctly, the theoretical $s^0_{20,w}$ values for HS were consistently lower than those seen experimentally (Figure 5.5a,b). This difference was not seen for heparin (Chapter 4: Figure 4.7). The lower theoretical values for HS than those seen experimentally indicate that the linear HS structures are more extended than the actual solution structures for HS. The theoretical $s^0_{20,w}$ value for HS dp24 is 1.05 S, which is 69% of the corresponding theoretical value of 1.52 S for a linear heparin dp24 structure.

(5.2.2) X-ray solution scattering data for eight HS fragment

Solution scattering is a diffraction technique that provides structural information on biological macromolecules in random orientations in solution (Perkins *et al.*, 2008, 2009). In order to complement the AUC data in more detail, the solution structures of the same eight HS fragments dp6-dp24 at 0.5 mg/ml were characterized by synchrotron X-ray scattering. The scattering experiments reports the scattering curve *I(Q)* as a function of scattering angle *Q*. Tests for possible radiation damage effects (Materials and methods) showed that these were not detectable, accordingly data were acquired with the longest exposure time of 1 sec in order to maximise signal: noise ratios. Guinier analyses of ln *I(Q)* vs. Q^2 at low *Q* value gives the radius of gyration R_G, which monitors the degree of macromolecular elongation. Because of the very different sizes of the HS fragments, different *Q* range

Table 5.1. X-ray scattering and sedimentation coefficient modelling fits for eight HS fragments

Heparin fragment	Filter	Number of models	Hydrated spheres [a]	R_G (nm) [b]	R_{XS} (nm)	*R*-factor (%)	Length *L* (nm)	$S^0_{20,w}$ (S) [c]
dp6	None	5000	9-23	0.82-1.19	0.04-0.43	4.4-7.8	n.a	
	R_G, R_{XS}, *R*-factor	8	15-20	1.01-1.03	0.30-0.31	4.4-4.5	3.0-3.5	0.52-0.79
	Best fit	1	19	1.01	0.30	4.4	3.0	0.52
	Experimental			1.03 ± 0.08	0.31 ± 0.06		3.0	0.82 ± 0.05
				0.98 ± 0.05				1.05 ± 0.04
dp8	None	5000	12-30	1.04-1.44	0.16-0.52	4.4-8.3	n.a	
	R_G, R_{XS}, *R*-factor	8	20-23	1.17-1.18	0.39-0.40	4.4	3.7-3.8	0.45-0.80
	Best fit	1	20	1.18	0.39	4.4	3.7	0.77
	Experimental			1.19 ± 0.08	0.40 ± 0.03		3.5	0.94 ± 0.06
				1.16 ± 0.02				1.06 ± 0.08
dp10	None	5000	15.37	1.17-1.63	0.24-0.63	4.3-8.0	n.a	
	R_G, R_{XS}, *R*-factor	8	25-30	1.40-1.43	0.42-0.47	4.3-4.4	4.7-4.8	0.48-0.95
	Best fit	1	26	1.42	0.42	4.3	4.8	0.91
	Experimental			1.41 ± 0.07	0.44 ± 0.04		4.5	0.95 ± 0.09
				1.37 ± 0.04				1.09 ± 0.06
dp12	None	5000	16-43	1.33-184	0.23-0.72	4.1-9.9	n.a	n.a.
	R_G, R_{XS}, *R*-factor	8	30-37	1.64-1.65	0.49-0.50	4.2	5.3-5.8	0.47-0.98
	Best fit	1	34	1.64	0.49	4.2	5.5	0.92
	Experimental			1.65 ± 0.09	0.49 ± 0.04		5.5	1.08 ± 0.09
				1.62 ± 0.03				1.16 ± 0.05
dp14	None	8000	24-50	1.52-1.94	0.18-0.76	4.2-9.0	n.a.	n.a.
	R_G, R_{XS}, *R*-factor	8	35-40	1.75-1.78	0.48-0.51	4.4	6.0-6.5	0.74-1.05
	Best fit	1	37	1.75	0.49	4.4	6.2	0.96
	Experimental			1.76 ± 0.07	0.51 ± 0.02		6.0	1.07 ± 0.07
				1.82 ± 0.09				1.18 ± 0.04
dp16	None	8000	28-54	1.57-235	0.36-0.91	6.3-19.4	n.a.	n.a.
	R_G, R_{XS}, *R*-factor	8	40-50	1.98-2.02	0.49-0.52	6.3-6.4	6.5-7.2	0.61-1.1
	Best fit	1	43	1.96	0.52	6.3	6.8	1.10
	Experimental			2.03 ± 0.07	0.52 ± 0.01		7.0	1.10 ± 0.03
				2.11 ± 0.11				1.24 ± 0.08
dp18	Experimental			2.34 ± 0.03	0.61± 0.05		8.5	1.12 ± 0.06
				2.44 ± 0.11				1.25 ± 0.07
dp24	Experimental			2.82 ± 0.10	0.65 ± 0.05		10.0	1.26 ± 0.06
				3.0 ± 0.05				1.35 ± 0.04

Table 5.1 legend

[a] The optimal totals of hydrated spheres were 16 for dp6, 21 for dp8, 26 for dp10, 31 for dp12, 37 for dp14 and 42 for dp16.

[b] The first experimental value is from the Guinier R_G analyses and the second one is from the GNOM *P(r)* analyses.

[c] The averaged experimental $s_{20,w}$ value is reported, the first value being from the absorbance (234 nm) data sets, and the second value from the interference data sets. The absorbance and interference data were recorded at rotor speeds of 40k, 50k and 60k rpm.

were required for each fragment in order to work within acceptable linear $Q.R_G$ ranges (Figure 5.6a). Thus the Guinier fit Q range of 0.4 to 0.8 nm^{-1} for dp6 was successively reduced in stages to that of 0.28 to 0.55 nm^{-1} for dp24 (Figure 5.6a). The mean Guinier R_G values increased from 1.03 ± 0.08 nm for dp6 up to 2.82 ± 0.10 nm for dp24 (Table 5.1). These R_G values for the eight HS fragments do not show a linear relationship with the size of the HS fragments, unlike the R_G values calculated from linear models (Figure 5.7a). Thus these linear HS structural models show bending in solution. In comparison with heparin, the heparin R_G values are smaller for the dp18-dp36 fragments than those seen for the HS dp18 and dp24 fragments. This shows that heparin has a more compact structure than that for HS. Like HS, heparin also showed a similar non-linear increase in the R_G values of its fragments with increase of size (Chapter 4: Figure 4.8).

Macromolecules that are sufficiently elongated in shape will show a cross-sectional radius of gyration R_{XS} value. The R_{XS} value monitors the degree of bending within the macromolecular length. As for the R_G analyses, different Q ranges were used for the R_{XS} fits for the different HS fragments depending on the size of the HS fragment, all of which were larger than the Q ranges used for the R_G analyses (Figure 5.6b). In all cases, despite the worsened signal-noise ratios of the $I(Q)$ data, linear fit ranges were identified in the plots of ln $I(Q).Q$ vs. Q^2. These gave experimental R_{XS} values of 0.31 nm for dp6 that increased up to 0.65 nm for dp24 (Figure 5.7b; Table 5.1). This increase in the R_{XS} values correlated with the deviation of the R_G values from linearity (Figure 5.7a). They were larger than the calculated R_{XS} values from 0.31 nm to 0.40 nm for the linear HS dp6-dp30 models, thus supporting the conclusion that the HS fragments become progressively more bent with increase in size. Combination of the R_G and R_{XS} values according to the relationship $L^2 = 12\ (R_G^2 - R_{XS}^2)$ for an elliptical cylinder (Glatter & Kratky, 1982) showed that HS dp6, dp8, dp10, dp12, dp14, dp16, dp18 and dp24 have approximate lengths of 3.4 nm, 3.9 nm, 4.6 nm, 5.5 nm, 5.8 nm, 6.8 nm, 7.8 nm and 9.5 nm in that order. Similar lengths of 7.0 nm, 9.1 nm, 9.6 nm and 10.7 nm were observed for the heparin dp18, dp24, dp30 and dp36 fragments (Chapter 4: Table 4.1). In conclusion, the comparison of the dp18 and dp24 lengths showed that HS has a longer structure than that of heparin, in addition to being more bent than heparin.

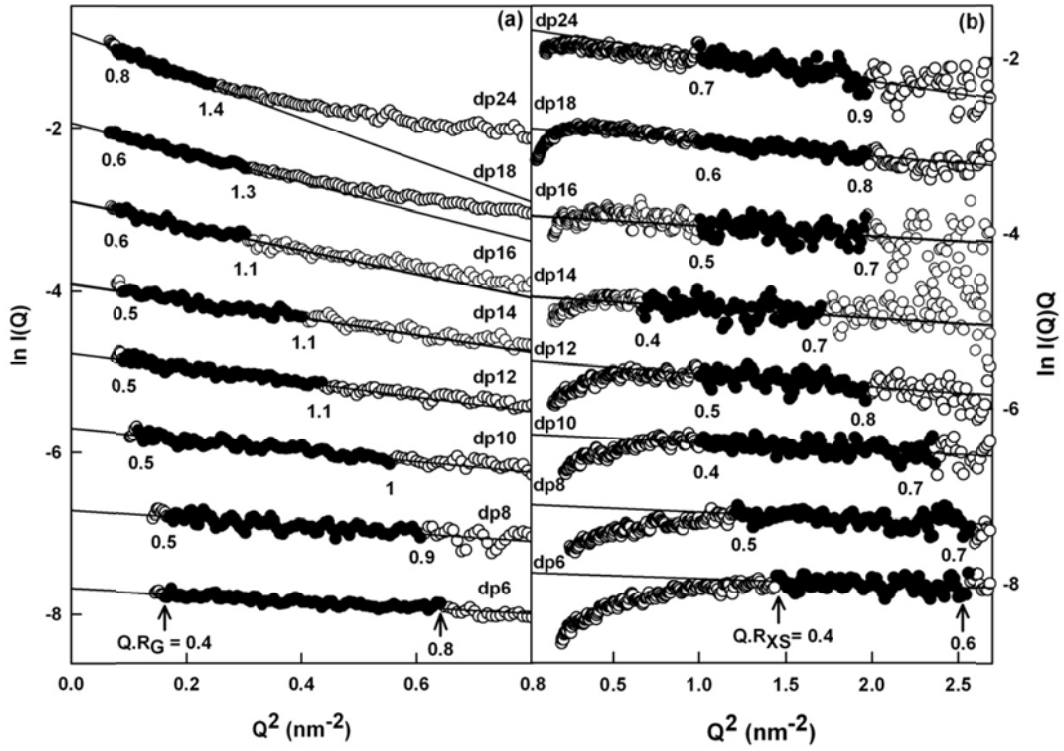


Figure 5.6

Experimental Guinier X-ray scattering analyses of eight HS dp6-dp24 fragments. (a) Guinier R_G plots for dp6-dp24 at concentrations of 0.5 mg/ml. The filled circles were used to determine the radius of gyration R_G, based on the best fit lines as shown. The Q ranges used for the R_G analyses were 0.40 to 0.8 nm^{-1} for dp6, 0.42 to 0.78 nm^{-1} for dp8, 0.34 to 0.74 nm^{-1} for dp10, 0.30 to 0.66 nm^{-1} for dp12, 0.30 to 0.64 nm^{-1} for dp14, 0.29 to 0.55 nm^{-1} for dp16, 0.28 to 0.55 nm^{-1} for dp18, and 0.28 to 0.55 nm^{-1} for dp24. (b) Guinier R_{XS} plots for dp6-dp24. The filled circles represent the Q ranges used to determine the cross sectional radius of gyration R_{XS}, based on the best fit lines as shown. The Q ranges used for R_{XS} analyses were 1.2 to 1.6 nm^{-1} for dp6, 1.1 to 1.6 nm^{-1} for dp8, 1.0 to 1.54 nm^{-1} for dp10, 1.0 to 1.44 nm^{-1} for dp12, 0.83 to 1.3 nm^{-1} for dp14, and 1.0 to 1.4 nm^{-1} for dp16, dp18 and dp24. Some parts of this Figure were prepared in collaboration with Miss Elizabeth Rodriguez under my supervision.

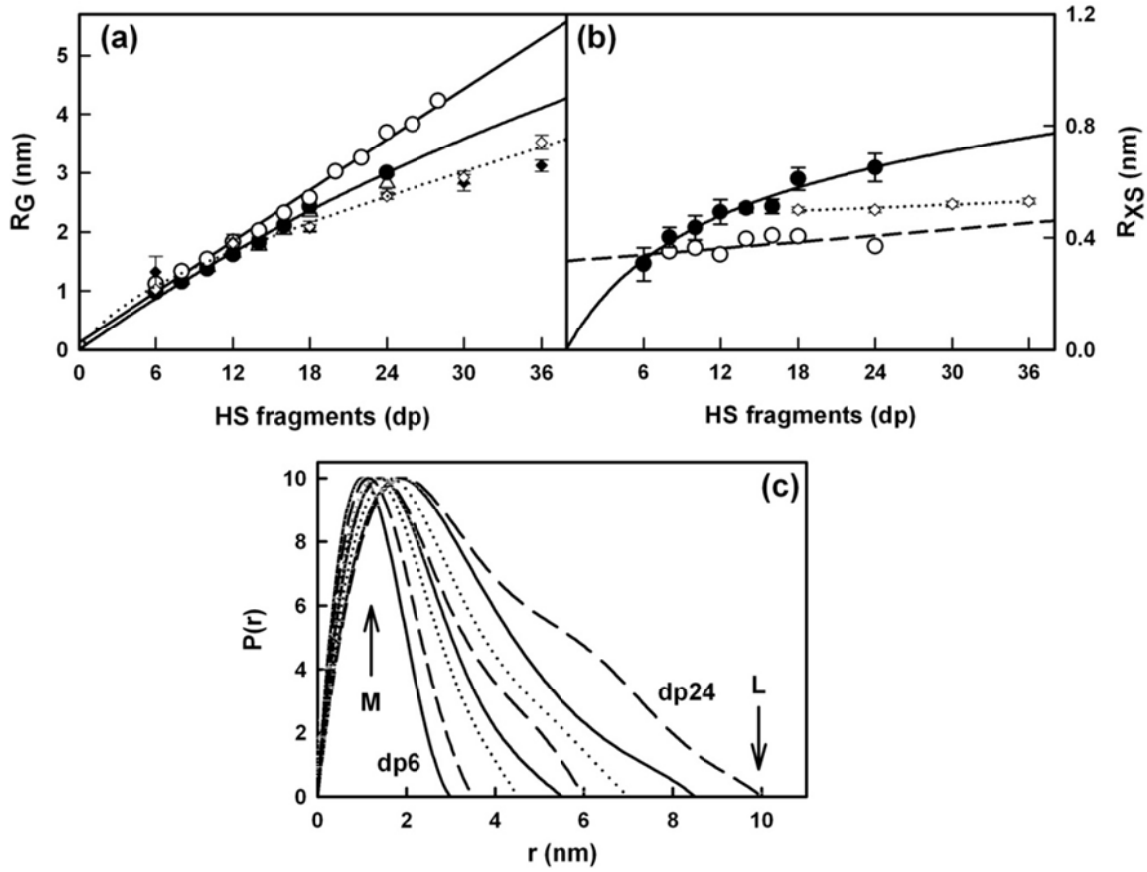


Figure 5.7

Experimental Guinier and *P(r)* X-ray data analyses of eight HS dp6-dp24 fragments. (a) Comparison of the experimental R_G values from Guinier plots (Δ) and *P(r)* curves (●) with the predicted R_G values calculated from the linear models of Figure 5.5 (○). The six corresponding Guinier and *P(r)* R_G values for heparin are denoted by filled and open diamonds respectively and fitted to a dotted line. (b) Comparison of the experimental cross-sectional R_{XS} values (●) with the predicted R_{XS} values calculated from the linear models of Figure 5.5 (○). The corresponding four R_{XS} values for heparin dp18-dp36 are denoted by open diamonds and fitted to a dotted line. (c) The distance distribution function *P(r)* analyses for dp6- dp24. The *r* values of the maximum at *M* were 1.02 nm (dp6), 1.15 nm (dp8), 1.30 nm (dp10), 1.43 nm (dp12), 1.44 30 nm (dp14), 1.61 nm (dp16), 1.87 nm (dp18) and 1.90 nm (dp24). The eight fragments are denoted by continuous, dashed and dotted lines in alternation.

The distance distribution function *P(r)* is calculated from the full *Q* range of the *I(Q)* scattering curve (Section 5.4). This provides R_G values and model-independent determinations of the overall length *L* following an assumption of the maximum dimension D_{max} (Figure 5.7c). The mean R_G values obtained from the *P(r)* curves increase from 0.98 ± 0.05 nm for dp6 to 3.0 ± 0.05 nm for dp24 (Table 5.1). These *P(r)* R_G values are in excellent accord with the corresponding Guinier R_G values from the low *Q* values and follow the same trends with size (Table 5.1; Figure 5.7a). Model-independent *L* values are determined from the *r* value where the *P(r)* curve reaches zero at large *r*. These experimental *L* values were 3.0 nm for dp6, 3.5 nm for dp8, 4.5 nm for dp10, 5.5 nm for dp12, 6.0 nm for dp14, 7.0 nm for dp16, 8.5 nm for dp18 and 10.0 nm for dp24. These values show increasing deviation with size from the longer lengths measured for the linear HS dp6-dp24 models (i.e. 3.5 nm for dp6, 4.1 nm for dp8, 5.6 nm for dp10, 6.5 nm for dp12, 7.2 nm for dp14, 7.8 nm for dp16, 9.2 nm for dp18 and 11.2 nm for dp24), noting that a hydration shell of thickness 0.6 nm (= 2 × 0.3 nm) is added to these linear model lengths (Perkins, 2001). These experimental *L* values from the *P(r)* curves are more accurate compared to the approximate *L* values calculated from the R_G and R_{XS} values that assumed an elliptical cylinder shape for HS, however these approximate *L* values show that the R_G and R_{XS} values are consistent with the *P(r)* analyses. The *P(r)* curves also provide the most frequently occurring interatomic distance *M* within the heparin structure from the *r* value of the peak maximum. *M* was observed at *r* values that started at 1.02 nm for dp6 and increased to 1.90 nm for dp24 (Figure 5.7c). In conclusion, these comparisons show that HS has a progressively more bent solution structure with increase in size.

(5.2.2) Constrained modelling of six HS fragments

The experimental X-ray R_G and R_{XS} values showed that the solution structures for HS are longer and more bent than those of heparin. Here, constrained scattering modelling was performed with HS in order to provide a molecular explanation of these scattering data. The linear models created from the HS dp4 crystal structure were the starting constraint. All eight HS fragments dp6-dp24 were subjected to modelling. As illustrated in the previous modelling of heparin, the linkage connectivity between the oligosaccharide rings was maintained (Figure 5.1a), while

the Φ and Ψ rotational angles at each glycosidic linkage was varied randomly in a range of up to ± 45° from their starting values. In all, 5,000-8,000 models for each of the eight HS fragments were generated. For each fragment, X-ray scattering curves were calculated from these randomised models and fitted to the experimental curve. The R_G, R_{XS} and *R*-factor values were calculated for each modelled curve, where the R_G and R_{XS} values were calculated using the same *Q* range used for the experimental Guinier fits (Figure 5.6), and the *R*-factor is a measurement of goodness-of-fit. The *R*-factor distributions of Figures 5.8a-f and 5.9a-f showed that all the dp6-dp16 models (yellow circles) encompassed the experimental R_G and R_{XS} values (dashed lines), and that the *R*-factor values were well below the level of 5% usually expected for excellent curve fits (Perkins *et al.*, 1998). Totals of 5000 models for dp6-dp10 and 8000 models for dp12-dp16 provided enough randomised conformers in order to be able to determine best fit HS solution structures. The lowest *R*-factors showed the best agreement with the experimental X-ray curves. At these minima, the best fit R_G values (red and cyan circles) agreed well with the experimental R_G values. The best fit R_{XS} values also showed good agreement with the experimental R_{XS} values (Figure 5.9a-f). In distinction, the linear models for the dp6-dp16 HS fragments showed more marked deviations from the experimental R_G and R_{XS} values (green circles). Of interest was that this modelling procedure could not be applied to HS dp18 or dp24 (not shown). There, no minimum *R*-factor could be identified in what turned out to be comparatively flat distributions. In addition, the curve fits showed better agreement for the linear models of dp18 or dp24 than the bent HS models, which was contrary to the analyses of Figure 5.5 and 5.7. The inability to fit the X-ray curve was attributed to the lack of a single well-defined solution structure for dp18 and dp24, unlike the cases of dp6-dp16.

For HS dp6, the modelling analyses showed that a near-linear structure with slight bending accounted for the X-ray and AUC data. The eight best fit models gave *R*-factors of 4.4-4.5%, R_G values of 1.01-1.03 nm, R_{XS} values of 0.30-0.31 nm, and maximum lengths *L* of 3.0-3.5 nm. These agree well with the experimental R_G value of 1.03 ± 0.08 nm, R_{XS} value of 0.31 ± 0.06 nm, and the *P(r)* length of 3.0 nm (Figures 5.8a and 5.9a; Table 5.1). In distinction to these agreements, the linear dp6 model gave a higher *R*-factor of 5.3 %, an R_G value of 1.10 nm and a *L* value of 3.5 nm. The visual agreement between the experimental and modelled *I(Q)* curves and

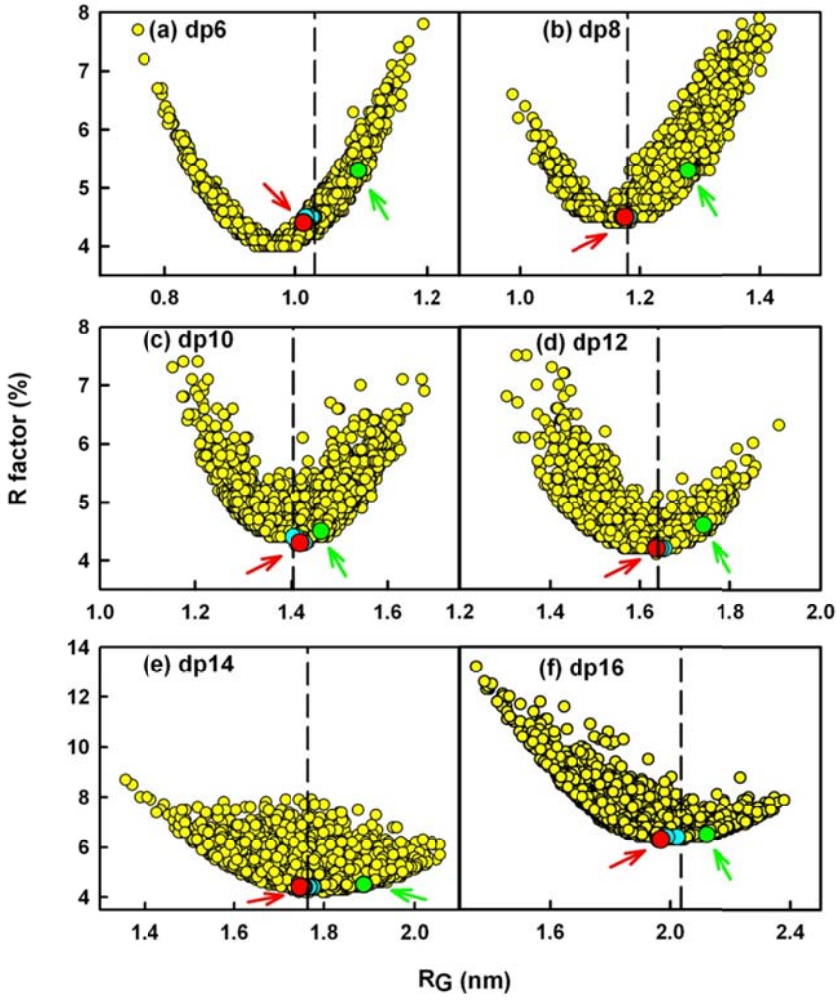


Figure 5.8

Constrained modelling analyses of six HS dp6–dp16 fragments. The vertical broken line corresponds to the experimental R_G or R_{XS} values. The red circle (arrowed) denotes the best-fit model, and the green circle (arrowed) denotes the linear model (Figure 5.5). The other best-fit models, where visible, are shown in cyan close to the *R*-factor minimum. (a–f) The *R*-factor values for the 5,000 trial models of dp6-dp12 (a-d) and 8,000 trial models of dp14-16 (e,f) are compared with their R_G values.

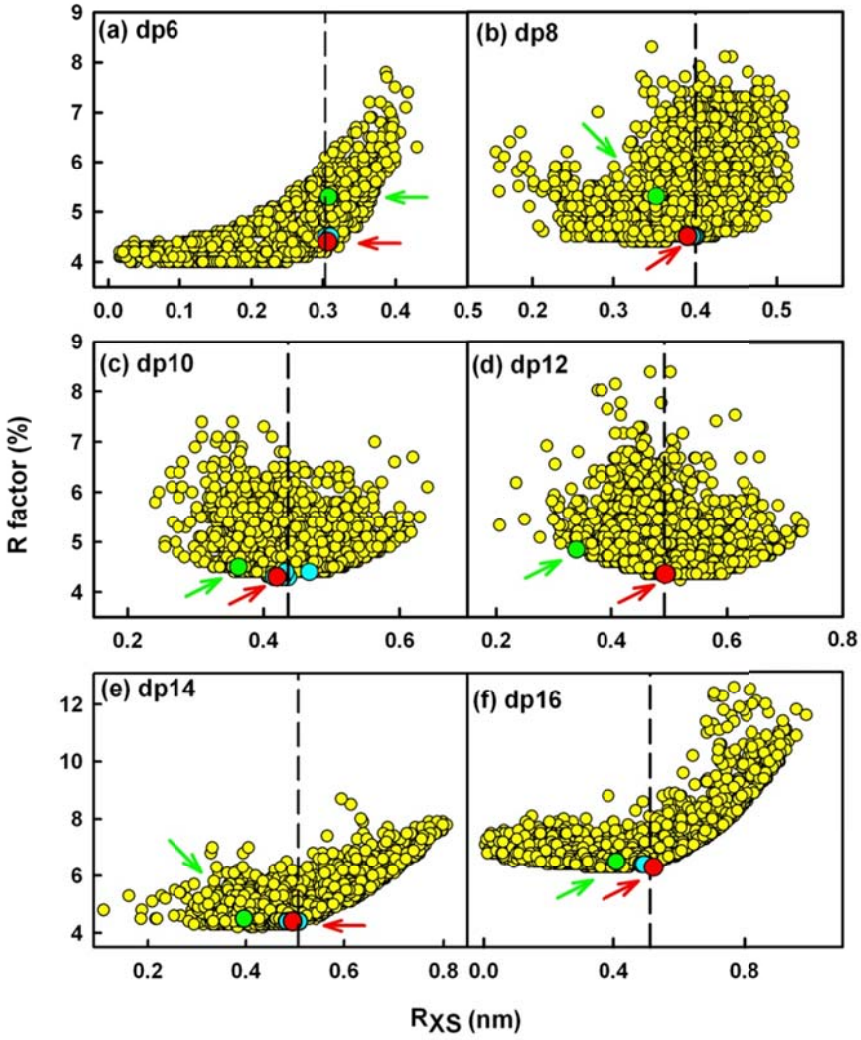


Figure 5.9

Constrained modelling analyses for six HS dp6-dp16 fragments. The vertical broken line corresponds to the experimental R_{XS} values. The red circle (arrowed) denotes the best-fit model, and the green circle (arrowed) denotes the linear model (Figure 5.5). The other best-fit models, where visible, are shown in cyan close to the *R*-factor minimum. (a–f) The *R*-factor values for the 5000 trial models of dp6-12(a-d) and 8000 trial models for HS dp14-dp16 (e,f) are compared with their R_G values.

P(r) curves was excellent (Figure 5.10a). The calculated $s^0_{20,w}$ values from the eight best-fit models gave 0.52 S to 0.79 S. These are lower but comparable with the experimental $s_{20,w}$ values of 0.82 ± 0.05 S and 1.05 ± 0.04 S, given that the typical accuracy of the $s^0_{20,w}$ calculation is ± 0.21 S (Perkins *et al*., 2009).

For HS dp8, the modelling analyses indicated slightly bent structures similar to those seen for dp6. The eight best fit models gave *R*-factors of 4.4%, R_G values of 1.17-1.18 nm, R_{XS} values of 0.39-0.49 nm, and *L* values of 3.7-3.8 nm. These values agree well with the experimental R_G value of 1.19 ± 0.08 nm, R_{XS} value of 0.40 ± 0.03 nm, and the *P(r)* length of 3.5 nm (Figures 5.8b and 5.9b; Table 5.1). In distinction to these agreements, the linear model again showed a higher *R*-factor of 5.3 % and R_G value of 1.28 nm, a lower R_{XS} value of 0.35 nm and a larger *L* value of 4.1 nm. The visual agreement of the experimental and modelled *I(Q)* and *P(r)* curves was excellent (Figure 5.10b). The eight calculated $s^0_{20,w}$ values of 0.45-0.80 S are smaller than the experimental $s_{20,w}$ values of 0.94 ± 0.06 S and 1.06 ± 0.08 S, however the best-fit model gave a calculated value of 0.77 S which is within error of experiment.

For HS dp10, the modelling analyses showed good agreements with slightly bent structures, in which the deviation from a linear dp10 structure for dp10 was slightly increased. The eight best fit dp10 models gave *R*-factors of 4.3-4.4 %, R_G values of 1.40-1.43 nm, R_{XS} values of 0.42-0.47 nm, and *L* values of 4.7-4.8 nm. These correspond well with the experimental R_G value of 1.41 ± 0.07 nm, R_{XS} value of 0.44 ± 0.04 nm, and *L* value of 4.5 nm (Figures 5.8c and 5.9c; Table 5.1). The deviations from a linear dp10 model are larger, for which the *R*-factor is 4.5 %, the R_G value is 1.46 nm, the R_{XS} value is 0.36 nm and the *L* value is 5.6 nm. The visual agreement of the experimental and modelled *I(Q)* and *P(r)* curves was again excellent (Figure 5.10c). The eight modelled $s^0_{20,w}$ values of 0.48-0.95 S are similar to that of the experimental $s_{20,w}$ values of 0.95 ± 0.09 S and 1.09 ± 0.06 S, with the best-fit model giving a modelled value of 0.91 S.

For HS dp12, the modelling analyses were also successful, in which the deviation from a linear dp12 structure was greater. The eight best fit models gave *R*-factors of 4.2%, R_G values of 1.64-1.65 nm, R_{XS} values of 0.49-0.50 nm, and *L* values of 5.3-5.8 nm. These agree well with the experimental R_G value of 1.65 ± 0.09 nm,

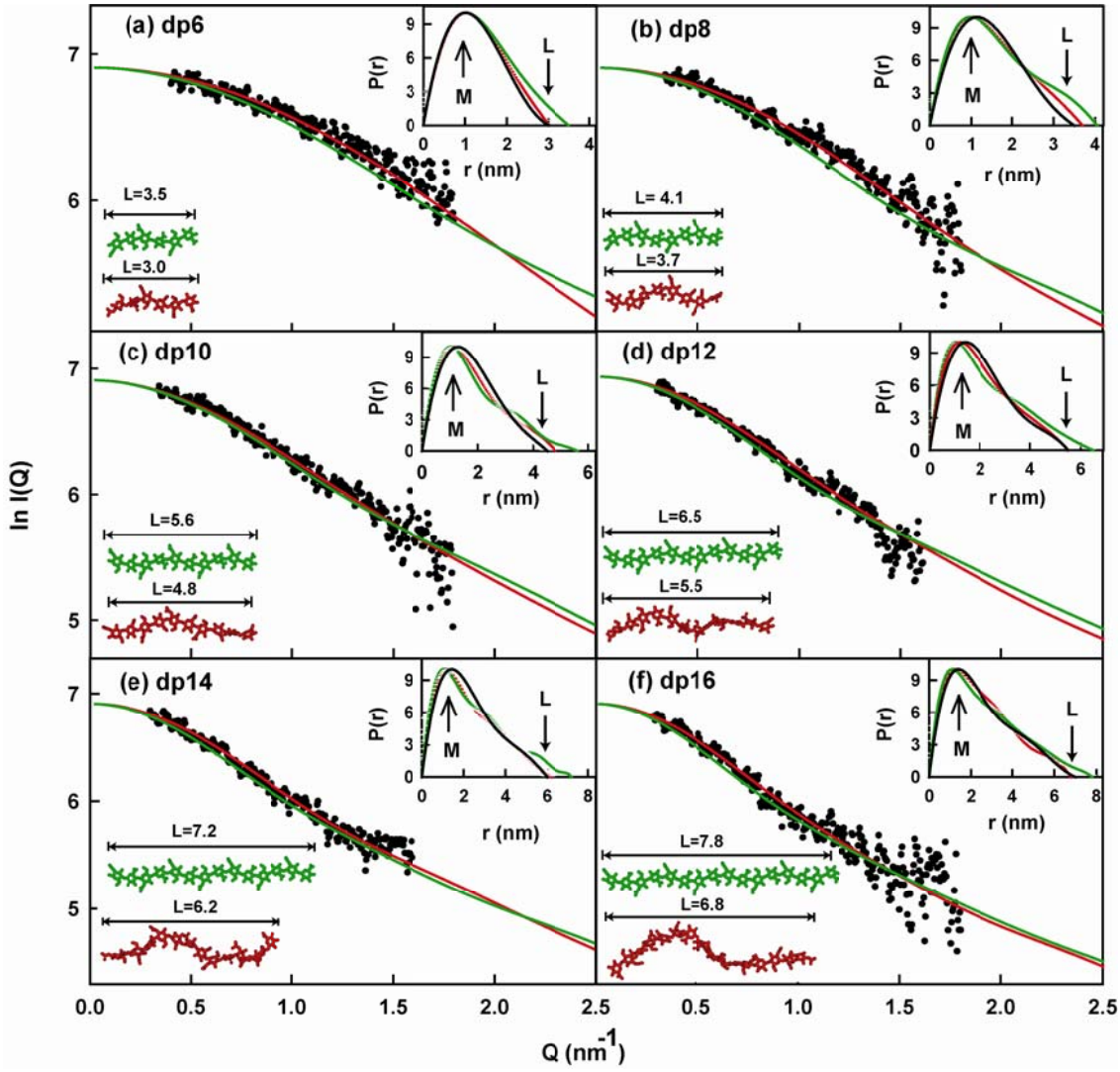


Figure 5.10

X-ray modelling curve fits for best-fit and poor-fit HS dp6-dp16 models. The main panels depicts the $I(Q)$ curve fits and the insets show the $P(r)$ distance distribution function fits. The experimental $I(Q)$ and *P(r)* scattering data are represented by black circles or lines respectively, the red lines and models correspond to the best-fit dp6-dp16 models from the trial-and-error searches, and the green lines and models correspond to the linear poor-fit dp6-dp16 models from Figure 5.5. The best-fit and linear models are shown to the left lower corner, together with their maximum lengths L in nm for comparison with the experimental L values in the *P(r)* curves.

R_{XS} value of 0.49 ± 0.04 nm and L value of 5.5 nm (Figures 5.8d and 5.9d; Table 5.1). The linear model showed a higher R-factor of 4.6 %, a higher R_G value of 1.74 nm, a reduced R_{XS} value of 0.34 nm and a longer L value of 6.5 nm. The visual agreement of the experimental and modelled *I(Q)* and *P(r)* curves was excellent (Figure 5.10(d)). The eight modelled $s^0_{20,w}$ values of 0.47-0.98 S are comparable with the experimental $s_{20,w}$ values of 1.08 ± 0.09 S and 1.16 ± 0.05 S, with the best-fit model giving a value of 0.92 S.

For HS dp14, good agreements between the models and experimental data were obtained, while the deviation from a linear dp14 structure was larger. The eight best fit dp14 models gave R-factors of 4.4 %, R_G values of 1.75-1.78 nm, R_{XS} values of 0.48-0.51 nm and L values of 6.0-6.5 nm. These agree well with the experimental R_G value of 1.76 ± 0.07 nm, R_{XS} value of 0.51 ± 0.02 nm respectively, and L value of 6.0 nm (Figures 5.8e and 5.9e; Table 5.1). These deviated from the linear dp14 model which had an R-factor of 4.5 %, a higher R_G value of 1.88 nm, a lower R_{XS} value of 0.39 nm, and an L value of 7.2 nm. Again the visual agreement of the experimental and modelled *I(Q)* and *P(r)* curves was excellent (Figure 5.10e). The eight modelled $s^0_{20,w}$ values of 0.74-1.05 S compare well with the experimental $s_{20,w}$ value of 1.07± 0.07 S and 1.18 ± 0.04 S respectively, where the best-fit model gave a value of 0.96 S.

For HS dp16, the outcome of the modelling analyses was similar to that of dp14. The eight best fit models gave R-factors of 6.3-6.4%, R_G values of 1.92-2.02 nm, R_{XS} values of 0.49-0.52 nm, and L values of 6.5-7.2 nm. These agree well with the experimental R_G value of 2.03 ± 0.07 nm, R_{XS} value of 0.52 ± 0.01 nm, and L value of 7.0 nm (Figures 5.8f and 5.9f; Table 5.1) The linear model gave a poorer fit with an R-factor of 6.5 %, R_G value of 2.12 nm, R_{XS} value of 0.41 nm and L value of 7.8 nm. The experimental and modelled *I(Q)* and *P(r)* curves showed excellent agreement (Figure 5.10f). The eight modelled $s^0_{20,w}$ values of 0.61-1.1 S are similar to the experimental $s_{20,w}$ values of 1.10 ± 0.03 S and 1.24 ± 0.08 S, with the best-fit model giving a value of 1.10 S.

In conclusion, the best fit models for HS dp6 to dp16 show progressively more bent structures in solution with increase in HS size. This is visible from the superimposition of the eight best fit models for each HS fragment (Figure 5.11). In terms of lengths, crystal structures containing HS and heparin showed that the

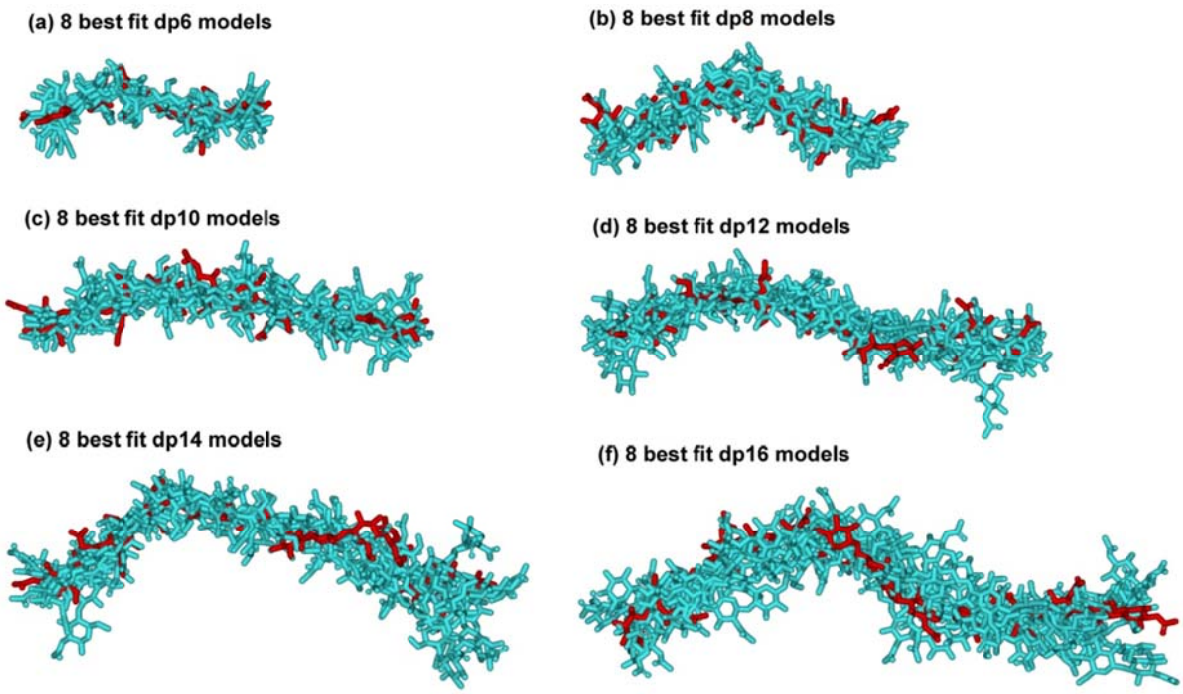


Figure 5.11

Superimposition of the 8 best-fit models for each of the HS dp6-dp16 fragments. Each set of eight best-fit models for the six HS fragments were superimposed globally using Discovery Studio VISUALISER software, then their non-hydrogen atoms were displayed as shown. The best-fit model from Figure 5.10 is shown in red, while the related best-fit structures are shown in cyan.

glycosidic bond in HS has a similar length to that in heparin (Figure 5.12b). In HS, the separation between the C1-C4 atoms of GlcA-GlcNAc is 0.237 ± 0.003 nm and that between GlcNAc-GlcA is 0.235 ± 0.002 nm. For heparin, analyses of five crystal structures containing dp6 showed that very similar separations were seen between IdoA-GlcNS of 0.241 ± 0.004 nm and between GlcNS-IdoA of 0.243 ± 0.006 nm. Thus the increased length of HS in solution compared to heparin is therefore principally the consequence of altered Φ and Ψ angles. In terms of rotational bend, it is noteworthy that the crystallographic Φ and Ψ angles for smaller HS and heparin structures are in agreement with each other within error (Table 5.2; Figure 5.13). However the comparison of the Φ and Ψ angles of HS and heparin showed that these values differed (Figure 5.13). The mean Φ and Ψ angles for the GlcA-GlcNAc glycosidic linkage were -93º and 134º in HS, and are similar to the corresponding IdoA-GlcNS values of -61º and 132º in heparin, although the standard deviations are large at ±21-36º (Table 5.2). In distinction, the mean Φ and Ψ angles for the GlcNAc-GlcA linker were 4º and 138º in HS, which are more different from the corresponding values of 98º and 86º for the GlcNS-IdoA linker in heparin, although the standard deviations are again large (Table 5.2). The distribution of the Φ and Ψ values in Figure 5.13b suggests that there is more rotational variability at this second glycosidic bond in HS when compared with those for heparin in Figure 5.13d. In conclusion, when compared with the corresponding solution structure of large heparin fragments, large fragments of HS showed a longer and more bent structure in solution than those of heparin (Figure 5.12a).

(5.3) Conclusions

The size and spacing of *S*-domains in heparan sulphate are proposed to be as important to its biologically significant interactions with proteins as are the detailed sequences of the S-domains themselves (Kreuger *et al.*, 2006). Heparin, a commonly used model compound for heparan sulphate, consists of lengthy *S*-domains, made up largely of the repeating trisulphated disaccharide shown in Figure 5.1a, separated by much smaller, unsulphated *NA*-domains. In heparan sulphate the position is reversed, and long *NA*-domain sequences (Figure 5.1b) separate short *S*-domains. Though the

Table 5.2. Summary of the Φ and Ψ angle in the solution and crystal structures of HS and heparin

	Φ (°)	Ψ (°)	Φ (°)	Ψ (°)
	GlcA-GlcNAc	GlcA-GlcNAc	GlcNAc-GlcA	GlcNAc-GlcA
HS crystal structure [a]	-90 ± 1	127 ± 4	83 ± 3	94 ± 1
HS dp6	-101 ± 38	140 ± 19	-16 ± 27	129 ± 48
HS dp8	-93 ± 26	121 ± 29	34 ± 55	152 ± 15
HS dp10	-121 ± 26	126 ± 17	18 ± 41	143 ± 23
HS dp12	-78 ± 30	114 ± 26	24 ± 35	121 ± 30
HS dp14	-64 ± 32	149 ± 21	-12 ± 53	139 ± 42
HS dp16	-112 ± 38	136 ± 45	-11 ± 88	143 ± 52
Average (27 / 35 values)	-93 ± 36	134 ± 32	4 ± 58	138 ± 37
	IdoA-GlcNS	IdoA-GlcNS	GlcNS-IdoA	GlcNS-IdoA
19 heparin crystal structures [a]	-79 ± 20	132 ± 19	84 ± 22	100 ± 19
Heparin dp18	-52 ± 22	135± 15	104 ± 27	96 ± 25
Heparin dp24	-61 ± 17	127 ± 20	100 ± 26	78 ± 17
Heparin dp30	-64 ± 21	136 ± 21	87± 25	82 ± 22
Heparin dp36	-63 ± 23	129 ± 27	103 ± 22	89 ± 24
Average (50 / 54 values)	-61 ± 21	132 ± 22	98 ± 25	86 ± 22

[a] The mean value from two HS dp4 molecules seen in the crystal structure (PDB code 3E7J).

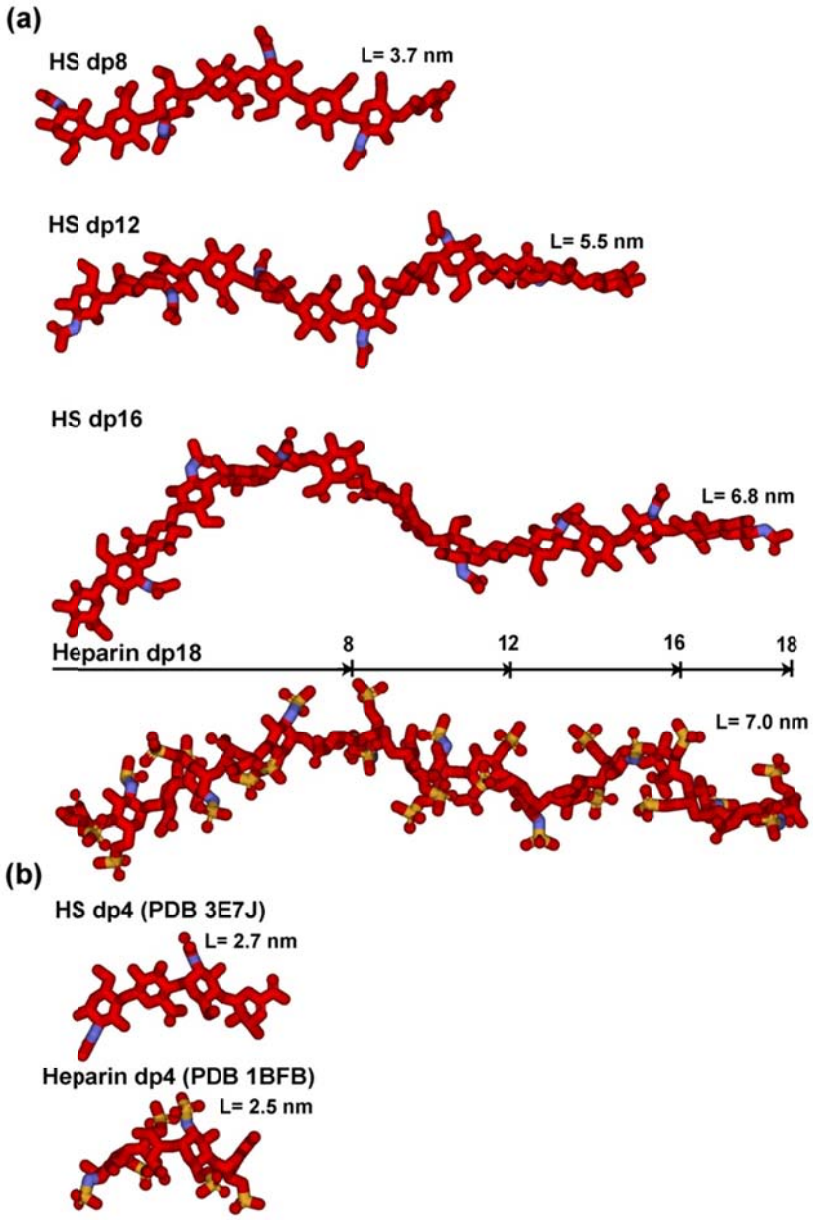


Figure 5.12

Comparison of the best-fit HS dp8-dp16 structures with the equivalent heparin dp8-dp16 structures. Red, carbon and oxygen; blue, nitrogen; sulphur, yellow. (a) Three best-fit HS models (dp8, dp12, dp16) are compared with the heparin dp18 model at the bottom all drawn to the same scale. The lengths of heparin dp8, dp12 and dp16 are indicated in the heparin dp18 structure for comparison with HS. (b) The glycosidic linkages in the crystal structure of HS dp4 complexed with heparinase II (PDB code: 3E7J)) are compared with the crystal structure of the complex of heparin dp4 with fibroblast growth factor (PDB code: 1BFB).

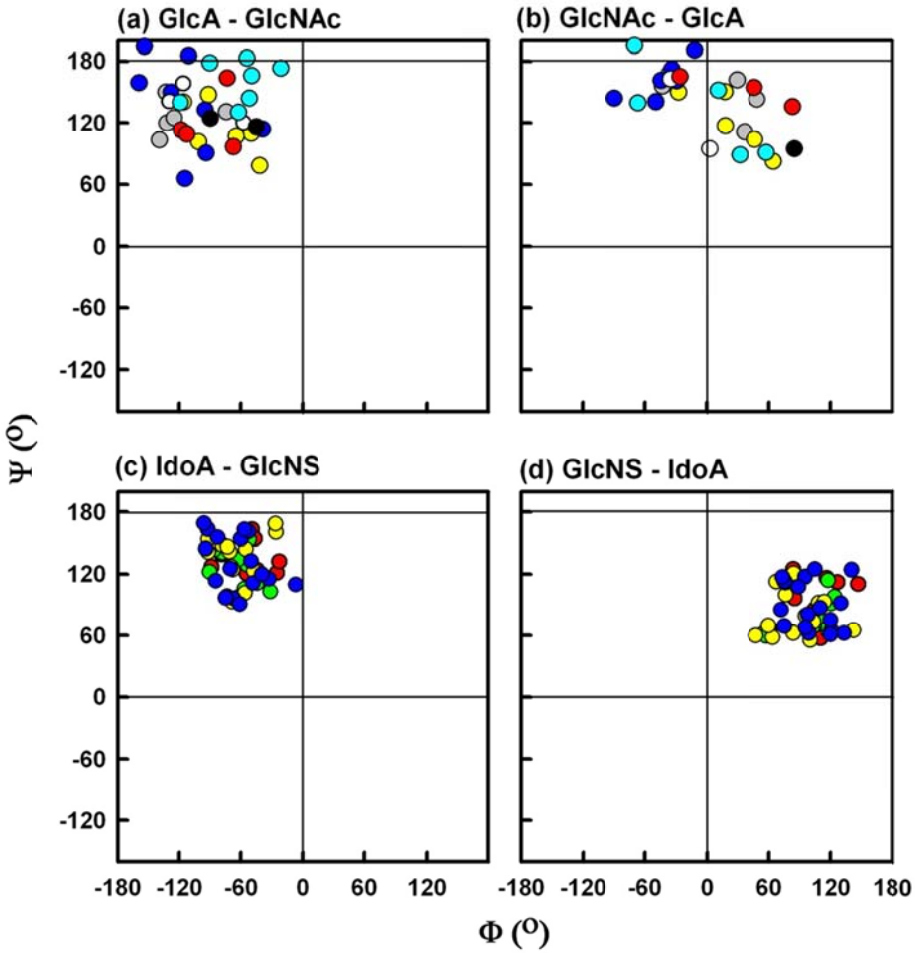


Figure 5.13

Phi (Φ) and psi (Ψ) dihedral angles for HS and heparin. The atoms defining the Φ and Ψ angles are shown in Figure 5.1. (a,b) The IdoA-GlcNS and GlcNS-IdoA Φ-Ψ angles found in the six best-fit HS solution structure models (Figure 5.10) are shown as filled circles for each of dp6 (2-3 values: white), dp8 (3-4 values: red), dp10 (4-5 values: grey), dp12 (5-6 values: yellow) dp14 (6-7 values: cyan) and dp16 (7-8 values: blue). The three Φ-Ψ angles for the HS dp4 crystal structure are shown in black. (c,d) The corresponding Φ-Ψ angles from the best fit solution structure models for heparin are shown as filled circles for each of dp18 (8-9 values: red), dp24 (11-12 values: green), dp30 (14-15 values: yellow) and dp36 (17-18 values: blue) are shown in the same scale as that in (a,b).

S-domain conformation, exemplified by heparin, has been the subject of numerous studies (Mulloy & Forster, 2000), the *NA*-domain has not. It has been proposed that the *NA*-domain repeating sequences are less flexible (Hricovini *et al.*, 1997) and more flexible (Mobli *et al.*, 2008) than the *S*-domains.

By this study, we have determined molecular solution structures for small and medium sized HS fragments dp6 to dp16, and obtained insights into the structures of HS dp18 and dp24. The fragments we used were produced by extensive depolymerisation using heparinase I, an enzyme that cleaves only within the *S*-domains (Powell *et al.*, 2010) leaving *NA*-domains untouched. It is therefore likely that some minor degree of sulphation remains at the reducing and non-reducing end of our fragments, but that internal disaccharides are unsulphated. Such fragments bear a closer resemblance to *NA*-domains of intact HS than the most commonly used model compound for this sequence, the capsular polysaccharide from *E. coli* K5 (Hricovini *et al.*, 1997; Mobli *et al.*, 2008).

These comprise extended ones with flexibility around the glycosidic linkages between monosaccharide residues and evidence of flexibility in the larger structures. The structures provide novel comparative insight into the structures of HS and heparin, and the likely manner that these two polyanionic oligosaccharides interact with their protein ligands. Previously we reported that heparin dp6-dp36 exhibited a semi-rigid structure in solution (Chapter 4). HS and heparin both share similar covalent structures (Figure 5.1). The comparison of our two sets of structures for HS and heparin becomes essentially that between *NA*-domains and *S*-domains. The greater length and bending of HS compared to heparin (Figure 5.12) may be attributed to the difference in uronic acid residue, in which GlcNAc alternates with IdoA in HS, causing HS to adopt a distinct conformation from that of fully sulphated heparin (Mulloy *et al.*, 1994). The regular repeats of sulphate groups in heparin that are absent in HS are also expected to contribute to the relative rigidity of heparin.

The application of constrained scattering modelling has proved to be as effective for the HS fragments as it was for heparin previously. Usually scattering fits are performed for protein structures with masses of 20-100 kDa and higher (Perkins *et al.*, 2008, 2009). The HS fragments dp6-dp24 and the heparin fragments dp6-dp36 are an order of magnitude smaller in size with masses of 1-7 kDa and 2-11 kDa, respectively. The ability to measure these scattering curves was attributed to the

strength of the X-ray beam at the ESRF synchrotron and the low background levels at the instrument. The constrained scattering modelling technique determines a three-dimensional molecular structure that best accounts for the observed scattering curve through trial-and-error searches that rule out structures that are incompatible with the observed scattering curves. By fixing the analyses to what is already known about the macromolecule, namely the carbohydrate rings, and varying only the flexible regions, namely the Φ and Ψ angles of each glycosidic linker, significantly fewer modelling variables are involved in the scattering fits. Through the variation of these Φ and Ψ angles, the analysis of 5,000-8,000 models provided sufficient statistical detail to result in clear V-shaped graphs of *R*-factor vs. R_G and *R*-factor vs. R_{XS} values. The best-fit models were identified by the lowest *R*-factor values, and these were verified by the agreement of the modelled and experimental R_G and R_{XS} values at this point. The quality of the HS dp6-dp16 scattering fits was similar to those of the heparin dp18-dp36 fits (Chapter 4: section 4.2.3). Interestingly, no good fits could be obtained for the HS dp18 and dp24 structures. Even though the monodispersity of these two fragments had been established by the single peaks seen in the ultracentrifugation *c(s)* analyses (Figure 5.4), it is possible that residual sample heterogeneity contribute to the lack of fit. It is more likely that flexibility, i.e. the accumulation of conformational variability at 17 and 23 glycosidic linkages in HS dp18 and dp24, compared to the variability between 5 to 15 linkages in HS dp6-dp16, has resulted in the appearance of multiple conformations for HS dp18 and dp24 that precluded the ability to fit one single conformation to these two scattering curves. This modelling outcome is distinct from that for heparin dp6-dp36 when good fits could be obtained for all six structures. This outcome suggested that the heparin structures show greater rigidity than the HS fragments.

The solution structures of the HS dp6-dp16 fragments exhibited a degree of bending (Figures 5.11 and 5.12a). In addition the HS structures are longer for reason of alterations in the glycosidic Φ and Ψ angles (Figure 5.13). The physical basis of these changes in HS is likely to arise from the GlcA-GlcNAc sequence (as opposed to the IdoA-GlcNS sequence in heparin), together with less repulsion between sulphate-sulphate, sulphate-carboxylate, and carboxylate-carboxylate groups in HS.

This study has provided fresh insight into potential HS-protein interactions and heparin-protein interactions. The outcome of 19 protein-heparin crystal

structures has been discussed previously (Chapter 4: section 4.2.4). In distinction, only one protein-HS crystal structure is known (Shaya *et al.*, 2010). It was of interest that the Φ and Ψ angles for this are in the similar range to those seen in the 19 heparin-protein crystal structures (Table 5.2). Several studies of heparan sulphate-protein interaction have noted that *SAS* sequences, in which two short *S*-domains are separated by an *NA* sequence, are preferred for optimum binding (Kreuger *et al.*, 2006). This is particularly true for oligomeric proteins such as, for example, MIP-1α (Stringer *et al.*, 2002) or platelet factor 4 (Stringer and Gallagher, 1997). In these cases the multiple heparin binding sites on the oligomer are not always arranged in a linear way, so that a single long *S*-domain cannot readily bind to more than one site on the same multimer. It has been reasonable to suppose that *NA*-domains, composed of alternating α-(1-4) and β-(1-4) linked hexopyranoses, would be more flexible than the unusually stiff heparin structure of *S*-domains, allowing *SAS* sequences to bend in order to present two *S*-domains to heparin binding sites on opposite sides of a protein complex. The conclusion that can be drawn from our study of heparan sulphate *NA* fragments supports this intuitive reasoning.

(5.4) Materials and methods

(5.4.1) Purification of HS fragments

HS oligosaccharide fragments were prepared in collaboration with Dr. Barbara Mulloy (see page 117) according to a similar method to that previously used for heparin oligosaccharides (Khan *et al.*, 2010; Pellegrini *et al.*, 2000; Mulloy *et al.*, 1997; Rice *et al.*, 1985). Exhaustive heparinase digestion was used to minimise the content of fully sulphated sequences. About 100 mg of HS (prepared from a crude glycosaminoglycan mixture, the kind gift of Laboratori Derivati Organici, Italy: a mixture of HS-I and HS-II as described in (Casu *et al.*, 1983) was weighed out and dissolved in about 2 ml phosphate buffer pH 7. An aliquot of 200 µl heparinase I stock solution was added and left to digest at room temperature for at least 2 hours, long enough for the reaction to run to completion. The reaction mixture was evaporated to dryness, using a rotary evaporator at 50°C.

To isolate the HS fragments, the dried digest was dissolved in 1.5 ml of 2% ammonium bicarbonate solution, and filtered through a 0.45 micron syringe filter before injection onto the preparative gel filtration column. The filtered digested HS

was then applied to a preparative gel permeation chromatography (GPC) column (100 × 1.6 cm; packed with Biogel P10) (Bio-Rad, UK). The HS fragments were eluted using 2% ammonium bicarbonate at a flow rate of 0.2 ml/min in 2 ml fractions. The absorbance of the fractions was measured at 234 nm and the top fractions corresponding to each individual resolved peak were pooled. The HS oligosaccharides larger than dp20 were not completely resolved to baseline resolution (Figure 5.2). The pooled fractions were evaporated under reduced pressure and lyophilised before assessment of their sizes by analytical GPC (Mulloy *et al.*, 1997). Like heparin, GPC was carried out using two columns (TSK G3000 SW-XL, 30 cm; TSK G2000 SW-XL, 30 cm; Anachem, UK) connected in series. The eluant was 0.1M ammonium acetate solution at a flow rate of 0.5 ml/min, and HS was detected with a refractive index detector (RI-1530, Jasco, UK). The chromatography system was calibrated using the First International Reference Reagent Low Molecular Weight Heparin for Molecular Weight Calibration (NIBSC 90/686). HS quantisation was achieved by integration of the area under each refractive index peak and comparison with a standard curve prepared using known concentrations of low molecular weight heparin. An absorption coefficient of 5500 M^{-1} cm^{-1} was used for HS experiments (Powell *et al.*, 2010).

(5.4.2) PAGE of HS fragments

The HS fragments were analysed by PAGE to determine the level of purity of each one according to a previously described method (Turnbull & Gallagher, 1988; Vives *et al.*, 2001). Each HS fragment (5 µg) was mixed with 20% glycerol up to a maximum volume of 10-15 µl, and then loaded into separate wells (Figure 5.2, inset). Phenol red in 20% glycerol in a maximum volume of 10 µl was also applied to a separate well as a marker. Initially samples were run through a stacking gel (5% acrylamide/0.5% bisacrylamide) at 150 V for 20-30 min until the phenol red started to enter the resolving gel. In the resolving gel (25% acrylamide/1% bisacrylamide), samples were run at a constant current of 18 mA until the phenol red reached the bottom of the gel. The discontinuous buffer system of Laemmli (Laemmli, 1970) consisted of 0.125 M Tris/HCL, pH 6.8 in the stacking gel and 0.375 M Tris/HCL, pH 8.8 in the resolving gel. The gel running buffer was 25 mM Tris/0.192 M glycine, pH 8.3. The gel was stained with 0.08% aqueous Azure A for 10 min to visualize HS

bands. The gel was then destained in water to remove excess dye and clear the gel background.

(5.4.3) Analytical ultracentrifugation of HS fragments

Sedimentation velocity data for eight HS fragments (dp6, dp8, dp10, dp12, dp14, dp16, dp18 and dp24) were obtained on two Beckman XL-I analytical ultracentrifuges (Beckman-Coulter Inc, Palo Alto, CA) using both absorbance and interference optics. Experiments with the dp6-dp24 fragments were performed at concentrations of 0.5 mg/ml in 10 mM HEPES and 137 mM NaCl (pH 7.4). The buffer density was measured at 20°C using an Anton-Paar DMA5000 density meter to be 1.00480 g/ml. A partial specific volume of 0.467 ml/g determined for heparin (Pavlov *et al*., 2003) was used for HS. An alternative higher value of 0.55 ml/g for HS has been reported elsewhere, and was used for data processing only when required to confirm that the partial specific volume has no effect on the outcome of this study (Fujiwara *et al*., 1984). Analytical ultracentrifugation runs were carried out in an eight hole AnTi50 rotor with standard double-sector cells with column heights of 12 mm at 20ºC using absorbance optics at 234 nm and interference optics. Sedimentation velocity data were collected at 40,000, 50,000 and 60,000 rpm using absorbance optics and interference optics. The continuous *c(s)* analysis method was used to determine the sedimentation coefficients $s_{20,w}$ of the eight HS fragments using SEDFIT software (version 9.4) (Dam & Schuck, 2004; Schuck, 2000). The *c(s)* analysis directly fits the experimental sedimentation boundaries using the Lamm equation, the algorithm for which assumes that all species have the same frictional ratio *f/fo* in each fit. The final SEDFIT analyses used a fixed resolution of 200, and optimised the *c(s)* fit by floating the meniscus and cell bottom when required, and holding the *f/fo* value, baseline and cell bottom fixed until the overall root mean square deviations and visual appearance of the fits were satisfactory (Figure 5.4). The individual *f/fo* values calculated previously for the heparin fragments were used for the equivalent HS fragments (Chapter 4: section 4.2.2).

(5.4.4) Synchrotron X-ray scattering of HS fragments

X-ray solution scattering of the above eight HS fragments dp6-dp24 were performed on the beamline ID02 at the European Synchrotron Radiation Facility

(ESRF) at Grenoble, France in two sessions with a ring energy of 6.0 GeV (Narayanan *et al.*, 2001). In the first session, data were collected for six HS fragments in 16-bunch mode using beam currents from 63 mA to 89 mA. In the second session, data were collected for all eight HS fragments in 16-bunch mode using beam currents from 65 mA to 78 mA. Data were acquired using an improved fibre optically-coupled high sensitivity and dynamic range CCD detector (FReLoN) with a smaller beamstop. The sample-to-detector distance was 3.0 m. Experiments used the same HS concentrations of 0.5 mg/ml and buffers used in the sedimentation velocity experiments. For each HS fragment, samples were measured in a flow cell which moved the sample continuously during beam exposure in 10 time frames with different exposure times of 0.1 s, 0.25 s, 0.5 s and 1 s to check for the absence of radiation damage effects. This exposure was optimised using on-line checks for the absence of radiation damage to show that this was not detectable.

Guinier analyses give the radius of gyration, R_G, which measures the degree of structural elongation in solution if the internal inhomogeneity of scattering within the macromolecules has no effect. Guinier plot at low Q values (where $Q = 4\pi \sin\theta/\lambda$; 2θ is the scattering angle and λ is the wavelength) gives the R_G and the forward scattering at zero angle $I(0)$: (Glatter & Kratky, 1982)

$$\ln I(Q) = \ln I(0) - R_G^2 Q^2/3$$

This expression is valid in a $Q.R_G$ range up to 1.5. If the macromolecular structure is sufficiently elongated (i.e. rod-shaped), the radius of gyration of the cross-sectional structure R_{XS} and the mean cross-sectional intensity at zero angle $[I(Q).Q]_{Q\to 0}$ parameters are obtained from fits in a higher Q range:

$$\ln [I(Q).Q] = \ln [I(Q).Q]_{Q\to 0} - R_{XS}^2 Q^2/2$$

The R_G and R_{XS} analyses were performed using an interactive PERL script program SCTPL7 (J. T. Eaton and S. J. Perkins, unpublished software) on Silicon Graphics O2 Workstations. Indirect Fourier transformation of the full scattering curve $I(Q)$ in reciprocal space gives the distance distribution function $P(r)$ in real space. This

yields the maximum dimension of the macromolecule *L* and its most commonly occurring distance vector *M* in real space:

$$P(r) = \frac{1}{2\pi^2} \int_o^\infty I(Q) Qr \sin(Qr)\, dQ$$

The transformation was carried out using GNOM software (Semenyuk & Svergun, 1991). For dp6, the full X-ray *I(Q)* curves contained 318 data points in the *Q* range between 0.40 nm^{-1} and 1.80 nm^{-1}. For dp8 and dp10, 332 data points each were used in the *Q* range between 0.34 nm^{-1} and 1.80 nm^{-1}. For dp12, 298 data points were used in the *Q* range between 0.30 nm^{-1} and 1.60 nm^{-1}. For dp14, 295 data points were used in the *Q* range between 0.30 nm^{-1} and 1.59 nm^{-1}. For dp16, 343 data points were used in the *Q* range between 0.29 nm^{-1} and 1.80 nm^{-1}.

(5.4.5) Molecular modelling of HS fragments

Linear HS models were created using the crystal structure of the HS tetrasaccharide (dp4) observed in its complex with heparinase II (PDB code: 3E7J), and using Discovery Studio (version 2.5) molecular graphics software (Accelrys, San Diego, CA). The residues in HS dp4 were *N*-acetyl-D-glucosamine (NAG), D-glucuronic acid (GCU) and 4,5-dehydro-D-glucuronic acid (GCD) in the order NAG-GCU-NAG-GCD (PDB three-letter nomenclature). The NAG-GCU disaccharide was taken to be the base structure, and these disaccharide units (NAG-GCU) were joined by glycosydic bonds to generate a fully extended linear HS dp30 structure. In this, the phi (Φ) and psi (Ψ) angles were maintained at similar values observed in the starting dp4 structure. Linear HS dp6-dp24 models were created from this extended dp30 starting model by the removal of non-required disaccharides. Totals of 5,000 conformationally-randomised models for each of dp6, dp8, dp10 and dp12, and 8,000 similar models for each of dp14 and dp16 were created starting from each linear model. In the original HS dp4 structure, the Φ and Ψ angles were -90° and 124° respectively for the GlcA-GlcNAc (GCU-NAG) disaccharide and 85° and 95° respectively for the GlcNAc-GlcA (NAG-GCU) disaccharide. These Φ and Ψ angles were randomised to take any value in a maximum range of ±45° from their starting values using the TorsionKick function in a PERL script that was modified from the

ExtractAngle.pl script provided with the Discovery Studio software. For example, in the case of dp16, a total of eight Φ and Ψ angles for GlcA-GlcNAc and seven Φ and Ψ angles for GlcNAc-GlcA were randomised in this way. To avoid steric clashes between the dp16 atoms in each randomised structure, a constant force field termed Dreiding minimization provided in Discovery Studio was used to correct this. Dreiding minimization was useful in generating structures by providing accurate geometries and reasonably accurate steric barrier for organic, biological and main group inorganic molecules (Chapter 3: section 3.4.2).

(5.4.6) Constrained scattering and sedimentation coefficient modelling

Each HS model was used to calculate the X-ray scattering curve for comparison with the experimental curve using Debye sphere models (Okemefuna *et al*., 2009; Furtado *et al*., 2008; Gilbert *et al*., 2006). A cube side length of 0.520 nm in combination with a cutoff of 4 atoms was used to create the spheres for the HS dp6-dp16 models. The hydration shell corresponding to 0.3 g/ g H_2O was created by adding spheres to the unhydrated sphere models using HYPRO (Ashton *et al*., 1997), where the optimal total of hydrated spheres is listed in Table 1. The X-ray scattering curve *I(Q)* was calculated using the Debye equation as adapted to spheres and

$$\frac{I(Q)}{I(0)} = g(Q)\left(n^{-1} + 2n^{-2} \sum_{j=1}^{m} A_j \frac{\sin Qr_j}{Qr_j} \right)$$

assuming a uniform scattering density for the spheres: (Perkins & Weiss, 1983).

where *g(Q)* is the squared form factor for the sphere of radius *r*, *n* is the number of

$$g(Q) = (3(\sin QR - QR\cos QR))^2 / Q^6 R^6$$

spheres filling the body, A_j is the number of distances r_j for that value of *j*, r_j is the distance between the spheres, and *m* is the number of different distances r_j. Other details are given elsewhere (Perkins, 2001; Okemefuna *et al*., 2009; Furtado *et al*., 2008; Gilbert *et al*., 2006). X-ray curves were calculated without instrumental corrections as these were considered to be negligible for the pinhole optics used in synchrotron X-ray instruments. First, the number of spheres *N* in the dry and

hydrated models after grid transformation was used to assess steric overlap between the HS disaccharides, where models showing less than 95% of the optimal totals (Table 5.1) were discarded. This procedure was found to be insensitive to steric overlap in the case of oligosaccharides, and was discontinued in favour of the Dreiding minimization procedure (above). Next, the models were assessed by calculation of the X-ray R_G values from Guinier fits of the modelled curves using the same Q ranges used for the experimental Guinier fits in order to allow for any approximations inherent in the use of the $Q.R_G$ range up to 1.5. Models that passed the N and R_G and R_{XS} filters were then ranked using a goodness-of-fit R-factor in order to identify the best-fit eight models for each HS fragment.

Sedimentation coefficients $s^0_{20,w}$ for the best fit scattering models were calculated directly from molecular structures using the HYDROPRO shell modelling program (Chapter 3: section 3.4.6). The default value of 0.31 nm for the atomic element radius for all atoms was used to represent the hydration shell surrounding HS.

(5.4.7) Protein Data Bank accession number

The best-fit dp6-dp16 models were deposited in the Protein Data Bank http://www.rcsb.org/pdb/ with accession codes of 3QHG (dp6), 3QHH (dp8), 3QHI (dp10), 3QHJ (dp12), 3QHK (dp14), and 3QHL (dp16).

.

Chapter Six

Complement Factor H possesses multiple independent binding sites for heparin and heparan sulphate: Implications for function and disease

(6.1) Introduction

The alternative pathway, one of the three activation pathways of the complement system of innate immunity, is a natural defence system for the recognition and destruction of foreign pathogens in vertebrate tissues (Pillemer *et al*., 1954; Pangburn & Muller-Eberhard, 1984; Muller-Eberhard, 1988; Joiner, 1988). The activation of C3, the central component of complement system, to C3b through the removal of the small anaphylatoxin C3a from C3 by cleavage initiates the alternative pathway. Complement Factor H (FH), a 154 kDa plasma glycoprotein, is the major regulator of the alternative pathway and enables host and non-host surfaces to be distinguished by inhibiting C3 activation on host cells. FH performs this regulatory function by acting on C3b in three ways. FH acts as a cofactor in the factor I-mediated proteolytic inactivation of C3b into iC3b (Weiler *et al*., 1976; Pangburn *et al*., 1977), accelerates the spontaneous decay of C3 convertase of the alternative pathway (Whaley & Ruddy, 1976), and displaces factor B from the C3 convertase (C3bBb) (Whaley & Ruddy, 1976).

FH consists of 20 short complement regulator (SCR) domains, each of which contains about 61 residues. The SCR domains constitute the most abundant domain type in complement, and are also known as short consensus repeats or complement control protein domains (Soares & Barlow, 2005). FH has three binding sites for C3b, where SCR-1/4 binds to intact C3b, SCR-6/10 binds to the C3c region of C3b, and SCR-16/20 binds to the C3d region of C3b (Sharma & Pangburn, 1996; Jokiranta *et al*., 1998). FH inhibits surface-bound C3b activity on host cell surfaces by recognising polyanionic structures such as sialic acid, heparan sulphate (HS) and dermatan sulphate, all of which are found on host cell surfaces (Kazatchkine *et al*., 1979; Carreno *et al*., 1989). Heparin has been often used as a model for such interactions for reason of its ready availability, sometimes being considered as an analogue of HS. FH has two major binding sites for heparin within SCR-7 and SCR-20. A third binding site for heparin was originally located at SCR-13, but this was reassigned to SCR-9, then this SCR-9 site has been questioned (Blackmore *et al*., 1996, 1998; Ormsby *et al*., 2006; Schmidt *et al.,* 2008). FH was the first protein to be genetically linked to age-related macular degeneration (AMD), a common disease leading to blindness in the elderly. A Tyr402His polymorphism at SCR-7 in FH has been associated with many cases of AMD (Klein *et al*., 2005; Haines *et al*., 2005; Edwards *et al*., 2005; Hageman *et al*., 2005). Drusen are extracellular deposits that

appear between the retinal pigment epithelium and Bruch's membrane, and are a biological indicator for the development of AMD (Bird *et al*., 1995). Drusen contain oxidized lipids and aggregated proteins including FH (Hageman *et al*., 2001). Glycosaminoglycans are also present in drusen (Kliffen *et al*., 2001).

To date, our studies of full-length FH and heparin and HS fragments have resulted in molecular structures at medium resolution. Although crystal structures of small FH fragments have been reported to date, the large size, glycosylation and interdomain flexibility of full-length FH have precluded its crystallisation to date. Constrained solution scattering modelling showed that FH has a folded-back SCR structure (Figure 6.1a) (Aslam & Perkins, 2001; Okemefuna *et al*., 2009). Analytical ultracentrifugation showed that FH is about 85% monomeric, together with dimeric, trimeric and higher species up to nonamers in a combination of reversible and irreversible equilibria (Okemefuna *et al*., 2009; Nan *et al*., 2008). The solution study of six heparin fragments dp6-dp36 showed that these have extended and semi-rigid structures that are pre-formed in a conformation that is ready to interact with proteins such as FH, and molecular structures were determined (Figure 6.1b) (Chapter 4: Figure 4.11d). The solution structures of HS dp6-dp16 fragments (in which the heparin-like highly-sulphated regions have been removed by digests) showed that these were longer and more bent than the corresponding heparin fragments (Figure 6.1c) (Chapter 5: Figure 5.10f). A single molecule of heparin dp36 or HS dp24 is not able to bind simultaneously to the two major binding sites in FH unless the FH domain structure becomes significantly more compact (Figure 6.1). Because the interactions of FH with heparin and HS fragments have not yet been characterised, we describe in this study the conformational and oligomeric properties of FH in the presence of heparin and HS fragments. Previously tetramer formation in FH in the presence of dextran sulphate and high molecular weight unfractionated heparin has been reported using gel filtration chromatography and analytical ultracentrifugation (Pangburn *et al*., 2009). Here analytical ultracentrifugation and X-ray scattering are used to examine the FH complexes with heparin and HS fragments, together with molecular modelling (Perkins *et al*., 2008, 2009). The results provided new insight into these FH complexes, most notably that the two major heparin sites exist independently of each other in an inflexible FH structure that is not able to bend into a more compact conformation, and that high sulphation levels are required for

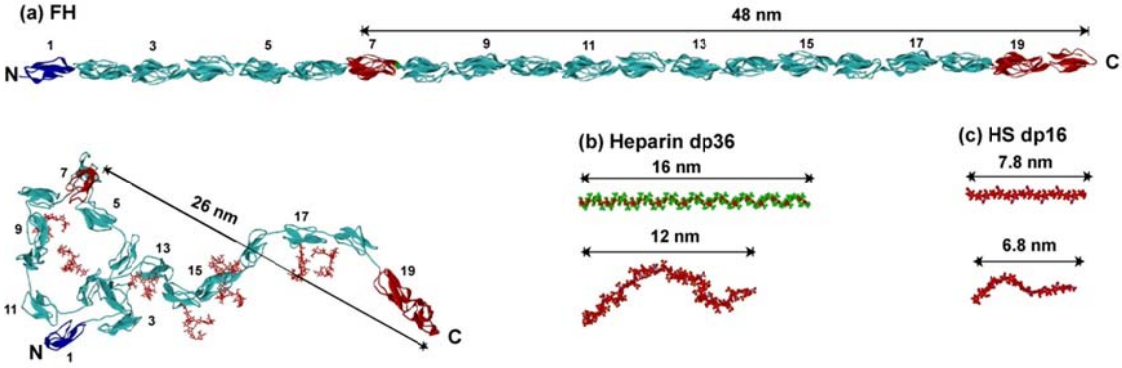


Figure 6.1

Molecular models for FH, heparin dp36, and HS dp24. (a) Linear and best fit X-ray solution scattering models for FH. The N and C-terminal domains are shown in blue and red respectively. No specific SCR proximity relationships are implied by the FH structure as shown; it illustrates the folded-back nature of the SCR domains. The heparin-binding SCR-7 and SCR-19/20 domains are shown in red, together with their separation. (b) Linear and best fit heparin dp36 structural models. (c) Linear and best fit HS dp24 structural models

strong binding. The implications of these findings for the regulatory function of FH and their importance for AMD are discussed.

(6.2) Results and discussion

(6.2.1) X-ray scattering data for FH mixtures with heparin

For X-ray scattering and analytical ultracentrifugation studies, full-length native FH was purified from the same pooled plasma stock used previously (Okemefuna *et al*., 2009; Nan *et al*., 2008a, 2008b). This gave a homogenous peak in gel-filtration chromatography and a single clean band on SDS-PAGE (Figure 6.2). All experiments were carried out at FH concentrations of 0.45-1.08 mg/ml which is comparable with physiological FH concentrations in plasma between 0.235-0.81 mg/ml (Saunders *et al*., 2006). Given the ionic nature of the FH-heparin interactions, all experiments were performed in buffers with 137 mM NaCl in order to correspond to near-physiological conditions in plasma.

Solution scattering is a diffraction technique that studies the overall structure of biological macromolecules in solution (Perkins *et al*., 2008). In order to determine the conformational and self-association properties of FH in the presence of six heparin fragments (dp6, dp12, dp18, dp24, dp30 and dp36), mixtures of FH (0.47 mg/ml) with 1:1 molar ratios of each heparin fragment were characterized by synchrotron X-ray scattering. FH concentrations of 0.47 mg/ml were advantageous in reducing the effect of large FH oligomers compared to 1 mg/ml. The experiments result in the scattering curve intensities $I(Q)$ measured as a function of the scattering vector Q. Given the small sizes of dp6-dp36 (1.88-11.30 kDa in mass) compared to FH (154 kDa), the Guinier analyses of the *I(Q)* data at low Q values will primarily observe structural and associative changes in FH when studied in the presence of heparin. These Guinier plots of In $I(Q)$ versus Q^2 gives the radius of gyration R_G, which measures the degree of macromolecular elongation. The Q range of 0.06 to 0.13 nm^{-1} gave the requisite linear R_G analyses for FH and its heparin mixtures (Figure 6.3a). The mean observed R_G values obtained for the FH mixture with dp6 and dp12 were 8.4 ± 0.5 nm and 7.3 ± 0.4 nm respectively (Table 6.1). These R_G values were slightly smaller than that of 8.9 ± 0.3 nm observed for FH alone. The R_G values increased significantly when heparin dp18-dp36 was added to FH (Figures 6.3a and 6.4a). The mean R_G values were observed to be 10.4 ± 0.3 nm, 9.9 ± 0.4 nm, 10.8 ± 0.3 nm and 11.7 ± 0.9 nm respectively for dp18, dp24, dp30 and dp36

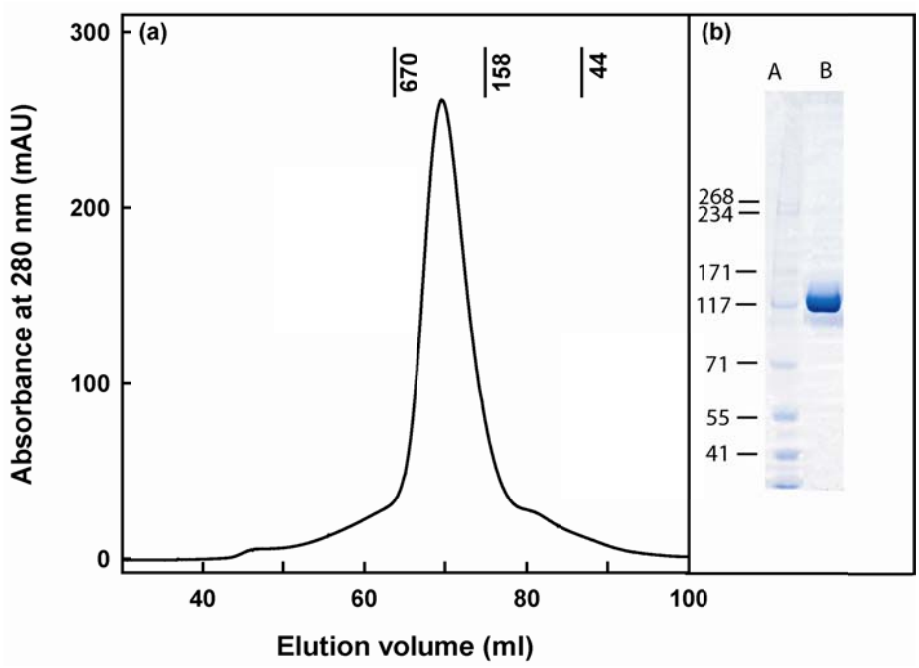


Figure 6.2

Size-exclusion gel filtration and SDS-PAGE of FH. (a) Size-exclusion gel filtration profile of FH through Superose^{TM} 6 prep grade XK 16/60 in HEPES buffer. FH eluted as a major monomeric peak. The elution positions of three molecular mass standards are shown. (b) Non-reducing SDS-PAGE analyses of FH after gel filtration. Lane A: Himark^{TM} Prestained High Molecular Weight Standards with their masses labelled to the left; Lane B: 7 μg FH.

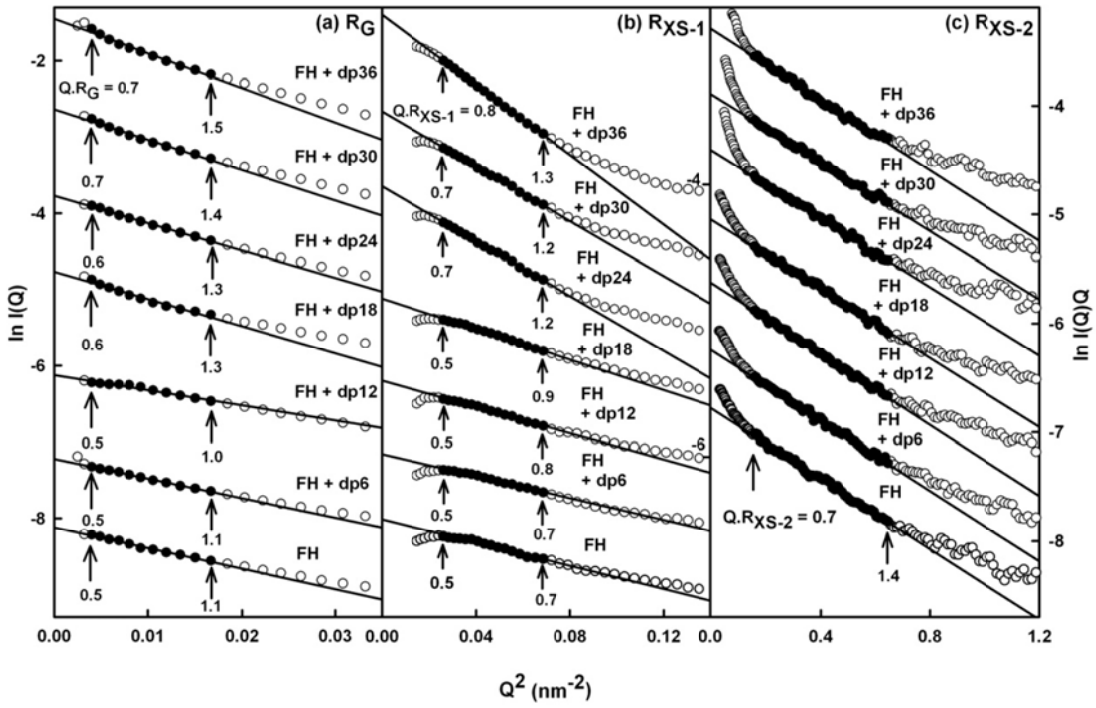


Figure 6.3

Guinier R_G and R_{XS} analyses for FH titrated with heparin dp6-dp36. The filled circles represent the $I(Q)$ values used to determine R_G and R_{XS}, while the straight lines represent the line of best fit through these points. The $Q{\cdot}R_G$ and $Q{\cdot}R_{XS}$ values for the fit ranges are arrowed. The plots are displaced successively on the vertical axis for reason of clarity. (a) R_G plot of ln $I(Q)$ versus Q^2 for the 1:1 FH-heparin mixtures at a FH concentration of 0.47 mg/ml using a Q range of 0.06–0.13 nm^{-1}. (b,c) R_{XS-1} and R_{XS-2} plots of ln $I(Q).Q$ versus Q^2 for the 1:1 FH-heparin mixtures using a Q range of 0.16–0.26 nm^{-1} and 0.4–0.8 nm^{-1} respectively.

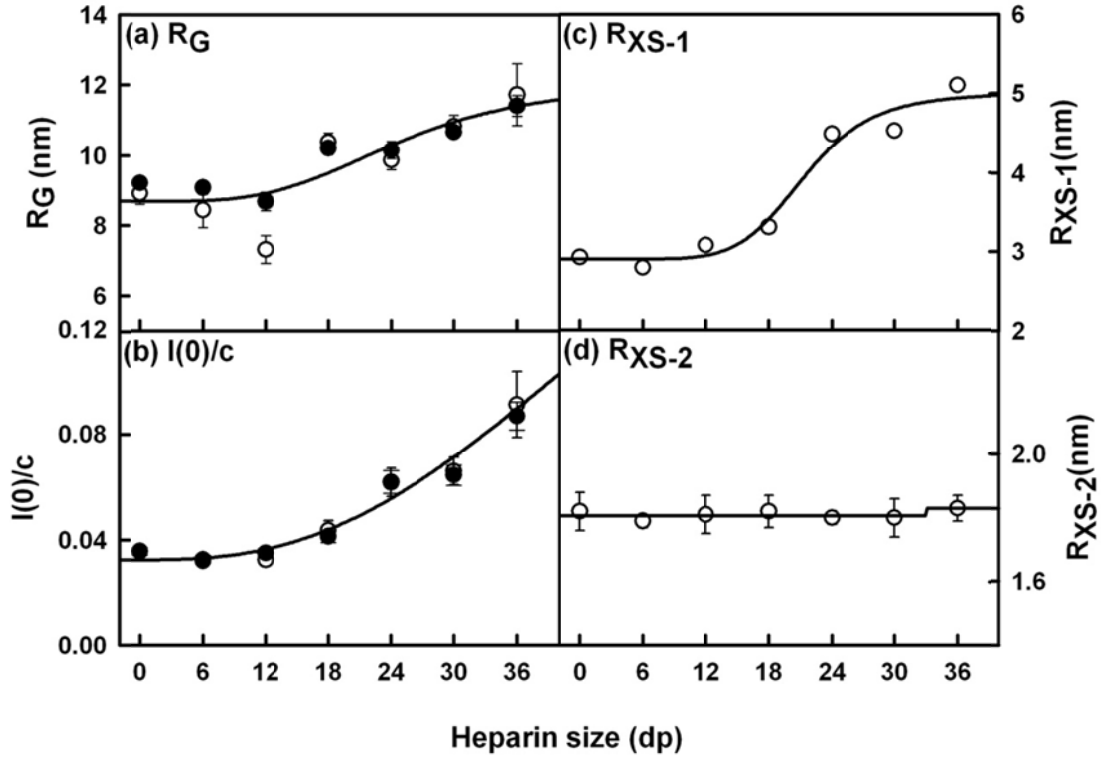


Figure 6.4

Dependence of the Guinier parameters of the 1:1 FH-heparin mixtures with heparin size. Each value was measured in quadruplicate and averaged. The data were fitted to a sigmoidal four-parameter function of the form $y = y_0 + (a /(1 + (x/x_0)^b)^{-1}$. Statistical error bars are shown where visible. (a,b) The R_G and *I(0)/c* values (Guinier values, open circles; *P(r)* values, filled circles). (c,d) The cross-sectional R_{XS-1} and R_{XS-2} values.

Table 6.1. X-ray scattering and sedimentation coefficient data for the FH complexes with six heparin fragments dp6-dp36, and the sedimentation coefficient modelling fits for the FH oligomers complexed with heparin dp36

	R_G (nm)	$R_{XS\text{-}1}$ (nm)	$R_{XS\text{-}2}$ (nm)	Length (nm)	$s_{20,w}$ (S) (monomer)	$s_{20,w}$ (S) (dimer)	$s_{20,w}$ (S) (trimer)	$s_{20,w}$ (S) (tetramer)	$s_{20,w}$ (S) (pentamer)
Experimental									
FH	8.9 ± 0.3	2.93 ± 0.03	1.82 ± 0.01	34	5.5 ± 0.1	7.6 ± 0.2	8.5	9.4 ± 0.5	10.5
FH + dp6	8.4 ± 0.5	2.80 ± 0.01	1.79 ± 0.02	35	5.6 ± 0.1	7.4	8.3 ± 0.5	9.3 ± 0.6	10.4
FH + dp12	7.3 ± 0.4	3.08 ± 0.04	1.81 ± 0.06	32	5.6 ± 0.1	7.8	8.1 ± 0.1	9.6 ± 0.5	10.3 ± 0.2
FH + dp18	10.4 ± 0.3	3.31 ± 0.07	1.82 ± 0.05	38	5.7 ± 0.1	7.1 ± 0.2	8.3 ± 0.4	9.4 ± 0.5	10.8 ± 0.1
FH + dp24	9.9 ± 0.4	4.49 ± 0.04	1.80 ± 0.02	36	5.6 ± 0.1	7.1 ± 0.2	8.5	9.1 ± 0. 1	10.4 ± 0.5
FH + dp30	10.8 ± 0.3	4.53 ± 0.04	1.80 ± 0.06	38	5.8 ± 0.1	7.2 ± 0.1	8.6 ± 0.2	9.5 ± 0.6	10.6 ± 0.3
FH + dp36	11.7 ± 0.9	5.11 ± 0.02	1.83 ± 0.02	40	5.7 ± 0.1	7.0 ± 0.2	8.2 ± 0.3	9.1 ± 0.1	10.6 ± 0.3
Modelling *									
FH + dp36 (Okemefuna *et al.*, 2009)					4.9	7.5 [a] 8.1 [b]	8.8 [c]	9.5 [d] 11.3 [f]	10.6 [e] 12.4 [g]
FH + dp36 A. I. Okemefuna & S. J. Perkins, unpublished.					5.0	8.4 [a] 8.2 [b]	9.9 [c]	11.6 [d]	13.0 [e]
FH + dp36 (Nan, 2010)					5.4	8.0 [a] 8.4 [b]	10.0 [c]	11.3 [d] 12.2 [f]	12.4 [e] 12.4 [g]

* The superscripts a-g correspond to the identification of the models in Figures 6.17, 6.18 and 6.19.

(Table 6.1). The Guinier *I(0)/c* value is proportional to molecular weight and this decreased slightly after mixture with dp6 and dp12 (Figure 6.4b). This indicated that dp6 and dp12 decreased the proportion of 8-10% of oligomers seen for free FH (Okemefuna *et al*., 2009; Nan *et al*., 2008), suggesting that these blocked the FH self-association sites. The *I(0)/c* values increased significantly with the larger heparin fragments (Figure 6.4b). That *I(0)/c* was more than doubled for the dp36 mixture compared to FH alone showed that FH oligomer formation had been strongly induced by heparin. Because the scattering curve is an average of the species present in the sample, it was not possible to state whether a single complex or multiple oligomeric species had been formed, nor was it possible to say whether conformational changes had taken place in FH on interaction with heparin.

The two cross-sectional Guinier analyses (Figure 6.3) monitor the structural proximity relationships between non-neighbouring ($R_{XS\text{-}1}$) and neighbouring ($R_{XS\text{-}2}$) SCR domains (Aslam & Perkins, 2001). The overall folded-back SCR domain structure of FH leads to the $R_{XS\text{-}1}$ value, while the localised extended structural arrangement of SCR domains leads to the $R_{XS\text{-}2}$ value. When no heparin was present, the $R_{XS\text{-}1}$ value of FH was 2.93 ± 0.03 nm, which is in agreement with the previous value of 2.7 ± 0.1 nm for free FH (Table 6.1) (Nan *et al*., 2008). The $R_{XS\text{-}1}$ values were almost unchanged at 2.80 ± 0.05 nm for the FH-heparin dp6 mixture and 3.08 ± 0.04 nm for the FH-heparin dp12 mixture. The $R_{XS\text{-}1}$ values increased significantly with heparin size for dp18-dp36 (Figure 6.4c). Thus the $R_{XS\text{-}1}$ values of 3.31 ± 0.07 nm, 4.49 ± 0.04 nm, 4.53 ± 0.04 and 5.11 ± 0.02 nm for the complexes of FH with dp18, dp24, dp30 and dp36 respectively were seen, indicating that new proximity relationships between non-interacting SCR domains had occurred in these mixtures. When no heparin was present, the $R_{XS\text{-}2}$ was 1.82 ± 0.01 nm, in good agreement with the previously reported $R_{XS\text{-}2}$ value of 1.79 ± 0.03 nm (Nan *et al*., 2008). When heparin fragments were added, no significant changes in the $R_{XS\text{-}2}$ values were observed (Figure 6.4(d)), all of which remained within the same range of 1.79 nm to 1.83 nm (Table 6.1). The $R_{XS\text{-}2}$ values indicated that the linear extended arrangement of neighbouring SCR domains remained unchanged in FH-heparin mixtures.

The distance distribution function *P(r)* represents the summation of all the distances *r* between every pair of atoms within the macromolecule. The *P*(*r*) curve, calculated from the full *I*(*Q*) curve (Materials & Methods), gives the R_G and *I(0)* values as well as the overall lengths *L* of FH (Figure 6.5). The R_G and *I(0)* values

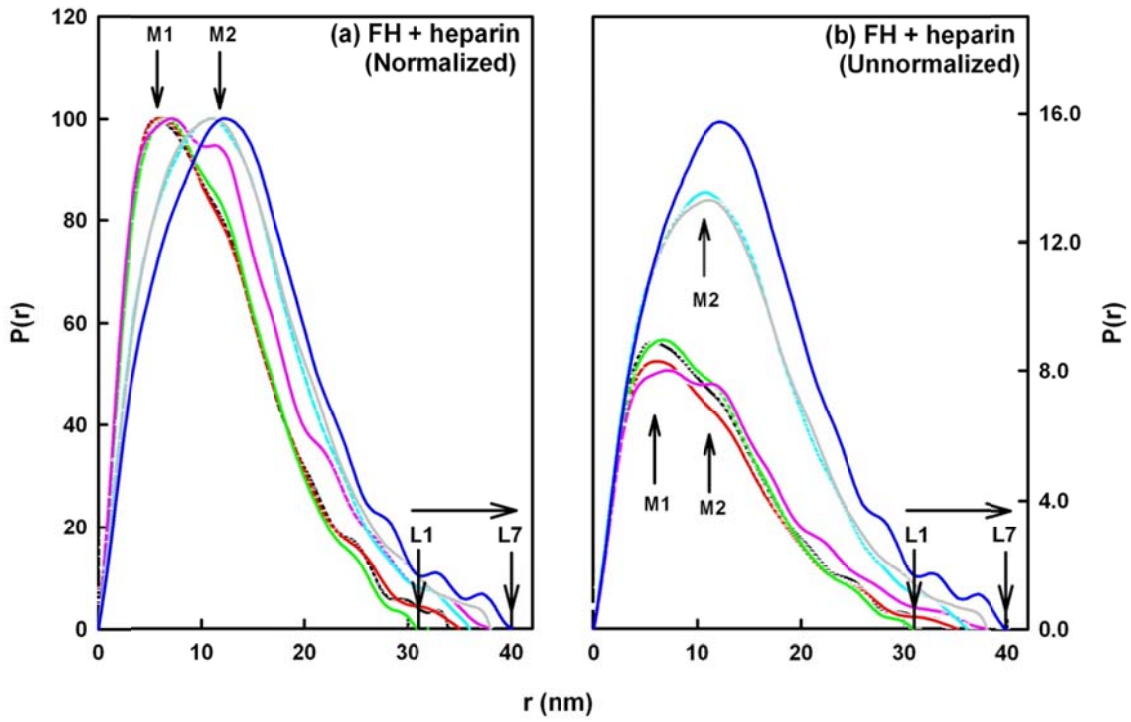


Figure 6.5

The distance distribution function $P(r)$ analyses for the 1:1 FH mixtures with heparin dp6-dp36. The *M1* maximum has *r* values of 5.8 nm (FH: black), 6.0 nm (FH + dp6: red), 6.7 nm (FH + dp12: green), 7.2 nm (FH + dp18: pink). The *M*1 peak was not observed for the complexes with dp24 (grey), dp30 (cyan) and dp36 (blue). The *M*2 maximum has *r* values of 11.6 nm, 11.6 nm, 11.5 nm, 11.4 nm, 10.8 nm, 11.4 nm and 12.4 nm in that same order. The *L* values were 34 nm, 35 nm, 32 nm, 38 nm, 36 nm, 38 nm and 40 nm in that same order. (a) The normalized $P(r)$ curves calculated from the seven scattering curves in Figure 6.3 permit comparison of the *M1* and *M2* values with heparin sizes. (**b**) The seven unnormalised *P(r)* curves permit comparison of the curve intensities.

from the *P(r)* analyses were in good agreement with the Guinier analyses (Figure 6.4a,b). The intensity of the *P*(*r*) curve increased significantly in the presence of large size of the heparin fragment. The maximum length *L* is measured when the *P*(*r*) curve reaches zero at large *r* value. In the absence of heparin, the *L* value of FH was 34 nm, in good agreement with the previously reported *L* value of 33 nm (Nan *et al*., 2008). The maximum, *M*, in the *P*(*r*) curve corresponds to the most frequently occurring interatomic distance within the structures. For free FH, the main peak *M*1 and a sub-maximum peak *M*2 were observed at *r* values of 5.8 nm and 11.6 nm respectively. When dp6 was added, the *L, M*1 and *M*2 values were unchanged at 35 nm, 6.0 nm and 11.7 nm respectively (Figure 6.5a). When dp12 was added, the *L, M*1 and *M*2 values were again unchanged at 32 nm, 6.7 nm and 11.6 nm, although there was a small increase in the size of the *M2* peak. Starting with dp18, larger changes were observed in which the *L* values were increased with heparin sizes, the *M*1 peak disappeared, and the *M*2 peak became more prominent. With dp18, dp24, dp30 and dp36, the *L* values increased to 38 nm, 36 nm, 38 nm and 40 nm respectively, indicating an increase in FH oligomer formation. There was also a large increase in the size of the *P(r)* curve for the FH mixtures with dp24, dp30 and dp36 (Figure 6.5b). The *P(r)* curves thus show that large FH oligomer had formed in the presence of 1:1 molar ratios of heparin dp24-dp36.

(6.2.2) X-ray scattering data for FH mixtures with HS

Mixtures of FH with eight HS fragments in 1:1 molar ratios that correspond to depleted sulphate contents (i.e. its *NA*-domain (Chapter 5: section 5.4.1) were studied in order to compare the effect of heparin and HS on FH. The same FH concentrations and buffer was used. In the Guinier analyses (Figure 6.6a), the R_G values measured for FH mixtures with HS dp6 and HS dp8 were 7.8 ± 0.1 nm and 7.8 ± 0.1 nm respectively, in agreement within that for free FH of 7.9 ± 0.1 nm. With HS dp10, FH showed a slightly smaller R_G value of 7.6 ± 0.1 nm. With HS dp12, HS dp14, HS dp16, HS dp18 and HS dp24, the R_G values of these mixtures with FH showed small increases to 8.0 ± 0.1 nm, 8.1 ± 0.1 nm, 8.1 ± 0.1 nm, 8.2 ± 0.1 nm and 8.3 ± 0.1 nm respectively (Figure 6.7a). The Guinier *I*(*0*)/*c* values also showed small increases with increased HS size (Figure 6.7b). The small increases in the R_G and *I*(*0*)/*c* values for FH-HS mixtures were generally smaller (with only one exception) than the corresponding R_G and *I*(*0*)/*c* values observed for the FH-heparin mixtures.

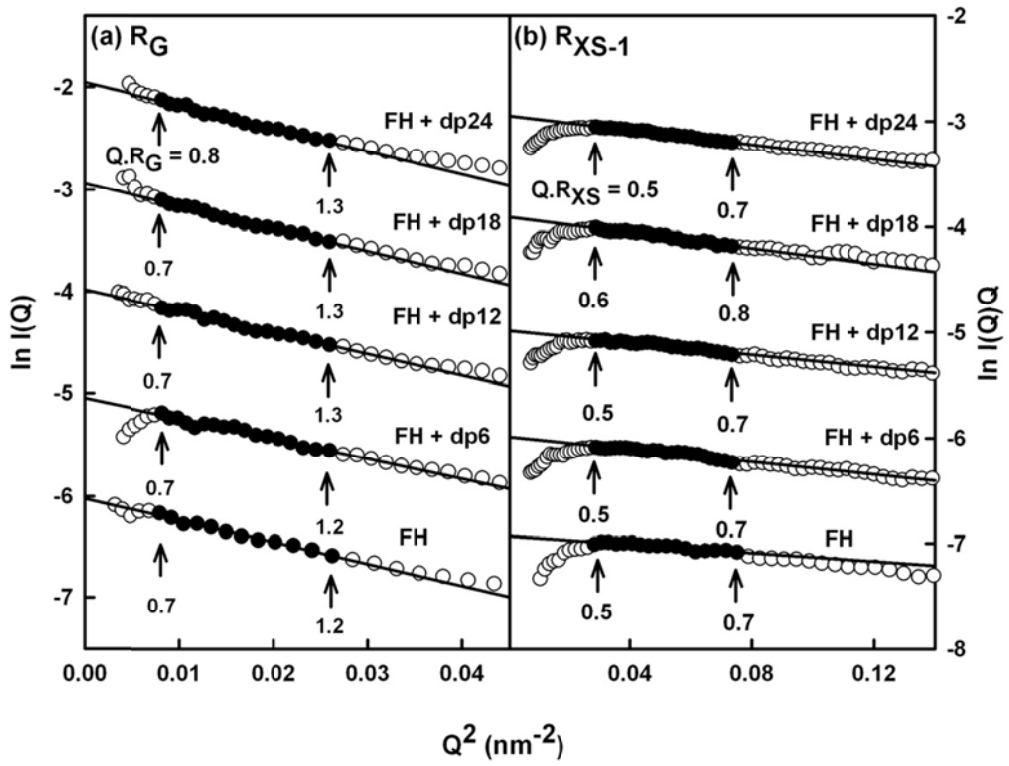


Figure 6.6

Guinier R_G and R_{XS} analyses for FH titrated with four HS dp6-dp24 fragments. Other details are given in Figure 6.3. (a) R_G plot of ln $I(Q)$ versus Q^2 for the 1:1 FH-HS mixtures at a FH concentration of 0.45 mg/ml using a Q range of 0.09–0.16 nm^{-1}. (b) R_{XS-1} plot of ln $I(Q).Q$ versus Q^2 for the 1:1 FH-HS mixtures using a Q range of 0.19–0.27 nm^{-1}.

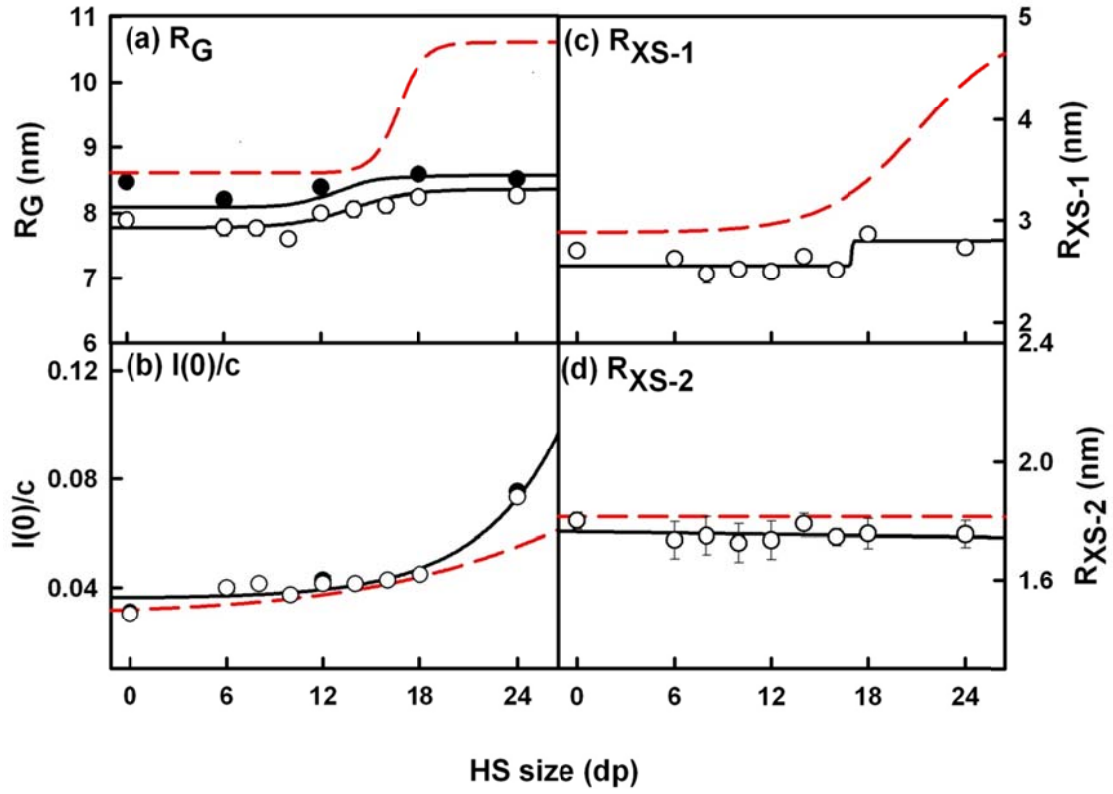


Figure 6.7

Dependence of the Guinier parameters of the 1:1 FH-HS mixtures with HS size. Other details are given in Figure 6.4. The red dashed lines correspond to the FH-heparin data of Figure 6.4. (a,b) The R_G and $I(0)/c$ values (Guinier values, open circles; *P(r)* values, filled circles). (c,d) The cross-sectional R_{XS-1} and R_{XS-2} values.

The cross-sectional radii of gyration R_{XS-1} and R_{XS-2} for the eight FH-HS mixtures were also analysed (Figure 6.6b). Compared to the R_{XS-1} value of 2.7 ± 0.1 nm for free FH, the R_{XS-1} values for the first six FH-HS mixtures were slightly smaller (Figure 6.7c). The R_{XS-1} values were 2.64 ± 0.06 nm, 2.60 ± 0.10 nm, 2.53 ± 0.04 nm, 2.50 ± 0.03 nm, 2.66 nm and 2.52 ± 0.01 nm for the mixtures with dp6, dp8, dp10, dp12, dp14 and dp16 respectively. For dp18 and dp24, the R_{XS-1} values were slightly increased at 2.87 ± 0.04 nm and 2.74 ± 0.04 nm respectively. Compared to the R_{XS-2} value of 1.79 ± 0.03 for free FH, no significant changes in the R_{XS-2} values in FH-HS mixtures were observed (Figure 6.7d).

The distance distribution function $P(r)$ curves was calculated for the FH mixtures with HS dp6, dp12, dp18 and dp24. Interactions between FH and HS were observed in all four cases, as shown by the increased intensity of the *P(r)* curves in all four cases, especially in the case of FH dp24 (Figure 6.8b). The lengths L were 30 nm for the dp6 and dp12 mixtures, similar to that for free FH (Figure 6.8a). The L values were increased to 32 nm and 31 nm for dp18 and dp24 respectively. The maximum *M1* and sub-maximum *M2* peaks were observed at r values of 4.8 nm and 10.2 nm for free FH. For the dp6, dp12, dp18 and dp24 mixtures, the *M1* peak was unchanged in position in a range between 4.2 nm to 5.1 nm. The *M2* peak was likewise also unchanged between 7.8 nm to 10.5 nm. While the *M2* peak first became visible for heparin dp12 and dp18, this change first became visible for HS only when the dp24 fragment was used (Figure 6.8a).

Compared to the effects of heparin on FH, the HS fragments showed similar effects but these were reduced in their magnitude. Thus, while interactions between FH and HS were detectable, the reduced number of sulphate groups in the HS fragments compared to those in heparin appeared to be responsible for the smaller observed effects.

(6.2.3) Sedimentation velocity data for FH mixtures with heparin

Analytical ultracentrifugation (AUC) provides structural information on macromolecules in solution by following their sedimentation behaviour under a high centrifugal force (Cole *et al*., 2008). In order to complement the X-ray scattering data for FH-heparin mixtures, sedimentation velocity data were collected for native FH at 1.08 mg/ml in the presence of six heparin fragments (dp6-dp36) in 1:1 ratios in the same buffer used for X-ray scattering. The FH concentrations of 1 mg/ml were

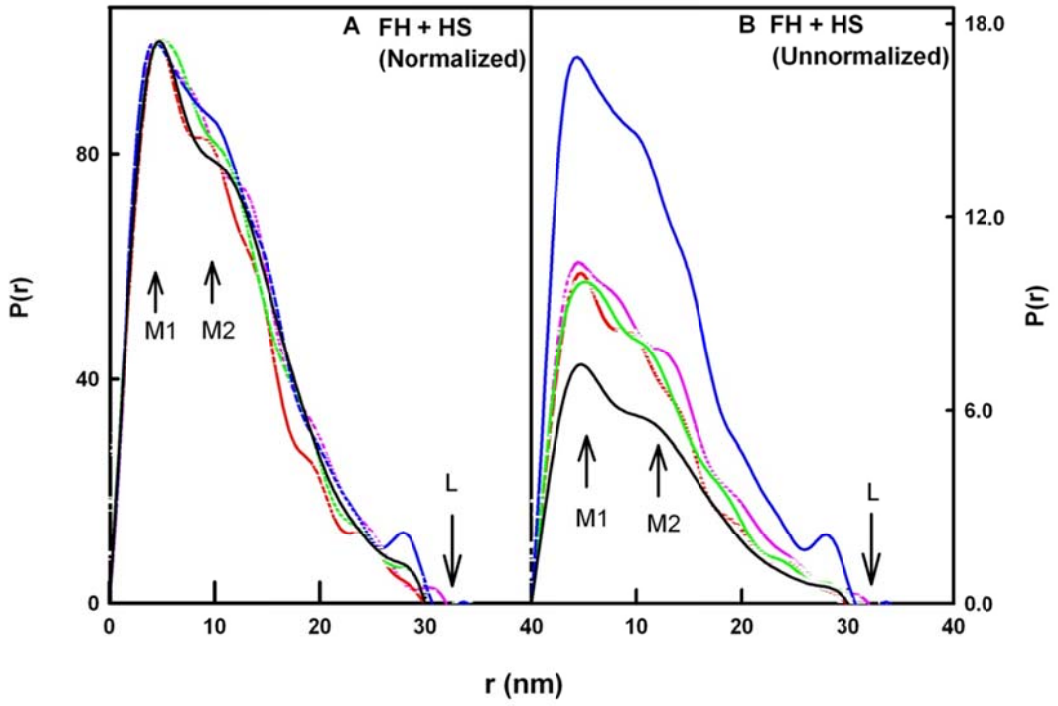


Figure 6.8

The distance distribution function $P(r)$ analyses for the four 1:1 FH mixtures with HS dp6-dp24. The *M1* maximum has an *r* value of 5.0 nm in all five *P(r)* curves, while the *M2* maximum has an *r* value of 10.2 nm for FH (black) and its mixtures with HS dp6 (red) and HS dp12 (green). The *M2* maxima were observed at *r* values of 7.6 nm and 9.5 nm for FH with dp18 (pink) and dp24 (blue). respectively. The *L* values were 30 nm for FH and its mixtures with HS dp6 and HS dp12. The *L* values were 32 nm and 31 nm for the mixtures with HS dp18 and HS dp24 Panels (a,b) correspond to the normalized and unnormalized $P(r)$ curves to follow Figure 6.5.

advantageous in making more visible the peaks that arise from its oligomers (see below). The velocity data were analysed using SEDFIT fits based on the interference optics to generate the size distribution analyses *c*(*s*) of the FH-heparin mixtures. Three rotor speeds of 40,000 rpm, 50,000 rpm and 60,000 rpm were used in order to evaluate potential reaction boundaries. Reaction boundaries are concentration-dependent and rotor speed-dependent peaks that correspond to the co-sedimentation of two different species if there is an equilibrium between themselves during sedimentation. They exhibit well-defined $s_{20,w}$ values and masses that are intermediate between the actual $s_{20,w}$ values and masses of the two different species, and are distinct from undisturbed boundaries which correspond to the actual sedimentation species themselves (Okemefuna *et al*., 2010; Li *et al*., 2010).

The analyses of interference scans showed that FH behaves differently in the presence of different fragments of heparin (Figures 6.9A, 6.9B and 6.9C). Because the $s_{20,w}$ values of free heparin dp6-dp36 sediment at 1.2 S to 1.9 S (Chapter 4: Figure 4.7), peak overlap of the FH peaks between 5 S to 22 S with the heparin peaks does not occur. In all the FH-heparin mixtures, the *c(s)* plots showed that monomeric FH was observed to sediment at an $s_{20,w}$ value of 5.66 ± 0.09 S, in good agreement with an $s_{20,w}$ value of 5.5 ± 0.1 S for free FH at 0.87 mg/ml. In the mass distribution analyses *c*(*M*), this peak corresponded to a molecular mass value of 132 ± 4 kDa in the FH-heparin mixtures, compared to one of 134 ± 4 kDa for free FH at 0.87 mg/ml. These values were in good agreement with the previously reported $s_{20,w}$ value of 5.65 ± 0.05 S and *c(M)* value of 142 ± 2 kDa for free FH at concentration between 0.17 to 5.92 mg/ml (Nan *et al*., 2008).

About 12-15% of FH oligomers were previously reported for free FH at different pH and salt concentrations (Okemefuna *et al.,* 2009; Nan *et al*., 2008). These were seen as small peaks numbered 2 to 10 in the *c(s)* plots (Figures 6.9A, 6.9B and 6.9C). No significant changes were observed in the presence of small size of heparin dp6 and dp12. Thus the FH-heparin dp6 mixture showed the same oligomeric peaks ranging from dimer to heptamers at $s_{20,w}$ values of 7.95 S, 9.77 S, 11.39 S, 14.09 S, 16.63 S and 19.60 S at 50,000 rpm (Figure 6.9Bd). Similar peaks were seen for the FH-heparin dp12 mixture at $s_{20,w}$ values of 8.07 S, 9.96 S, 11.86 S, 13.44 S, 16.09 S and 19.50 S (Figure 6.9Bd). The mean proportion of all the oligomers at 40,000 rpm, 50,000 rpm and 60,000 rpm were 9 ± 3% and 9 ± 2% in the presence of dp6 and dp12 respectively, which is slightly less 12 ± 3% observed

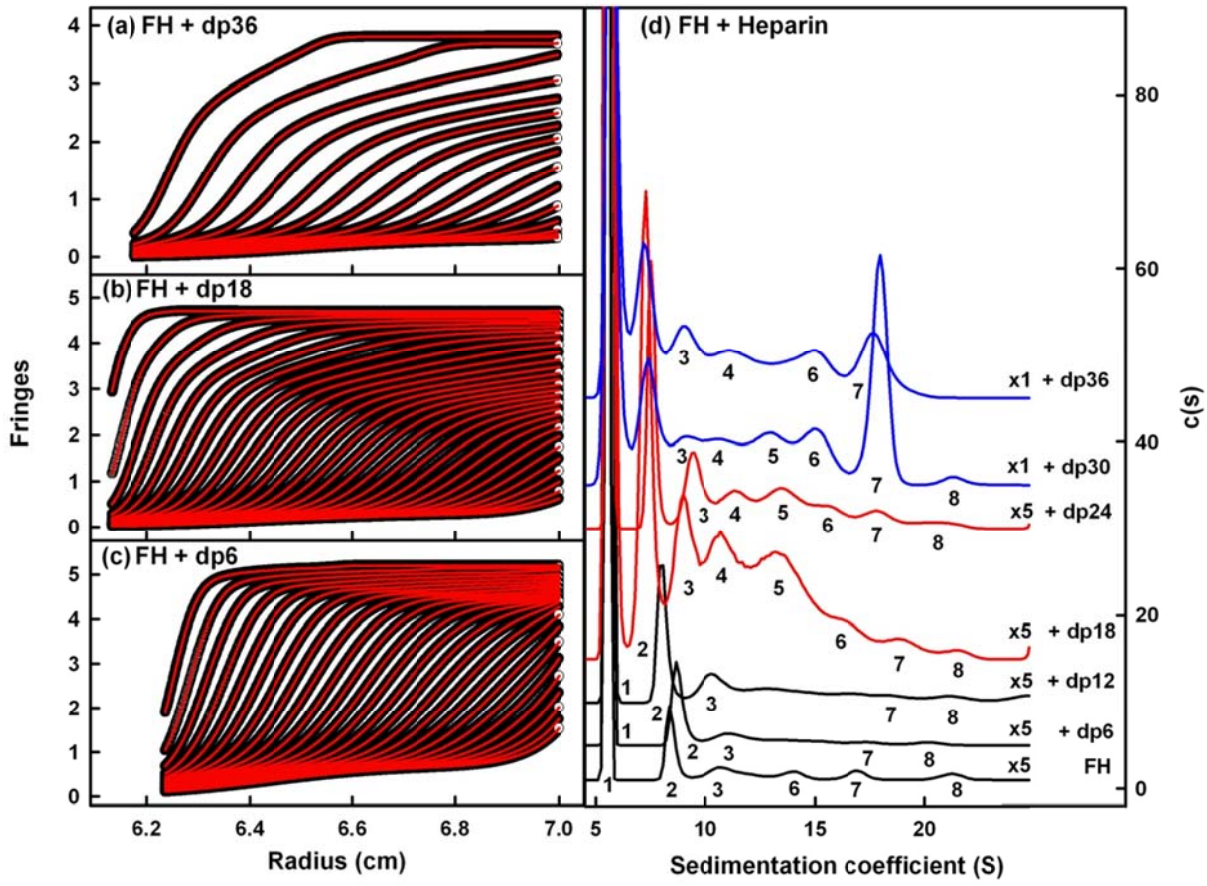


Figure 6.9A

Ultracentrifugation size distribution analyses *c(s)* of FH titrated with six dp6-dp36 heparin fragments. Parts A, B and C correspond to 40,000 rpm, 50,000 rpm and 60,000 rpm. Thus in part B, the sedimentation velocity experiments were performed at a rotor speed of 50,000 rpm with 1.08 mg/ml FH in the presence of 1:1 molar ratios of dp6, dp12, dp18, dp24, dp30 and dp36. The experimental boundaries using interference optics are shown in black, and the fits are shown in red.

(a) Every fifth boundary scan of 100 scans was fitted for the FH-dp36 mixture.

(b) Every sixth boundary scan of 120 scans was fitted for the FH-dp18 mixture.

(c) Every fifth boundary scan of 100 scans was fitted for the FH-dp6 mixture.

(d) The *c*(*s*) analyses for the FH-heparin mixtures are shown. For greater visibility of the FH oligomer peaks, the intensities were multiplied by a factor of 5 for FH and the dp6-dp24 mixtures. The FH monomer peak is denoted as 1, and the FH oligomer peaks are denoted by 2-12 in increasing order of sedimentation coefficient values.

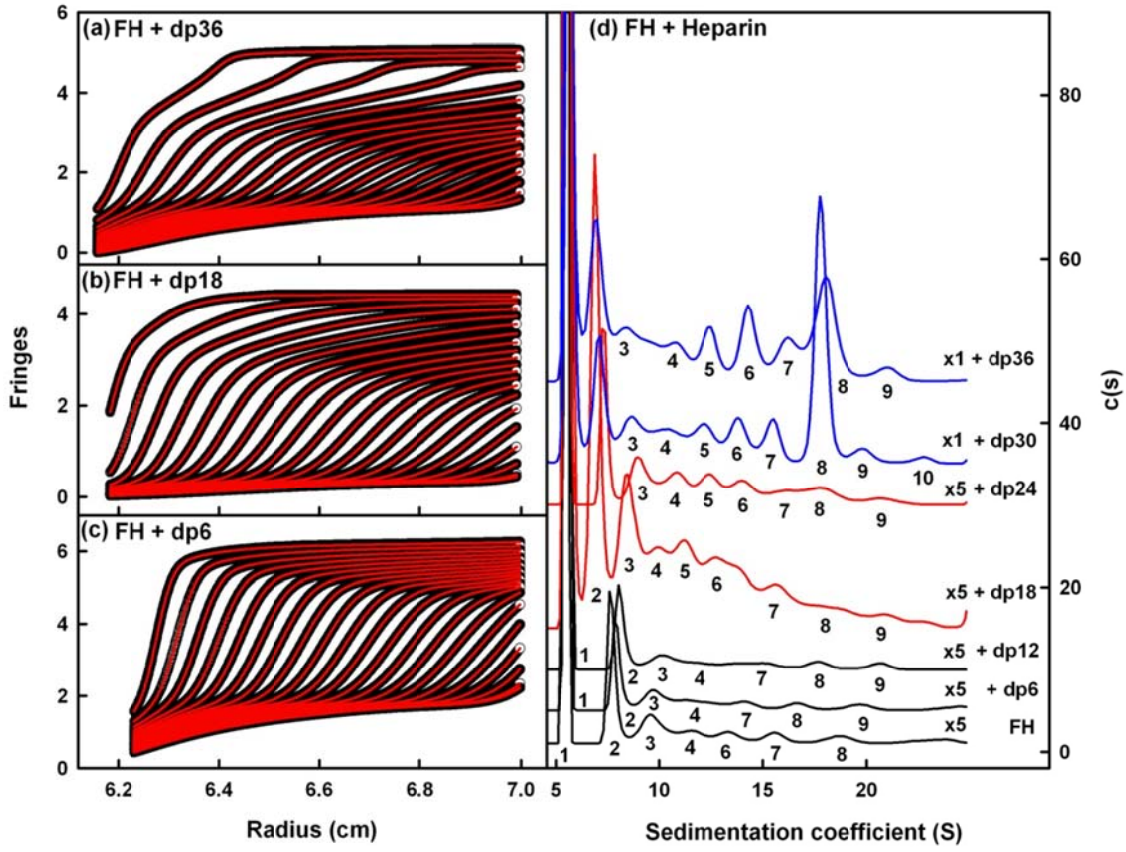


Figure 6.9B

Details are given in Figure 6.9A.

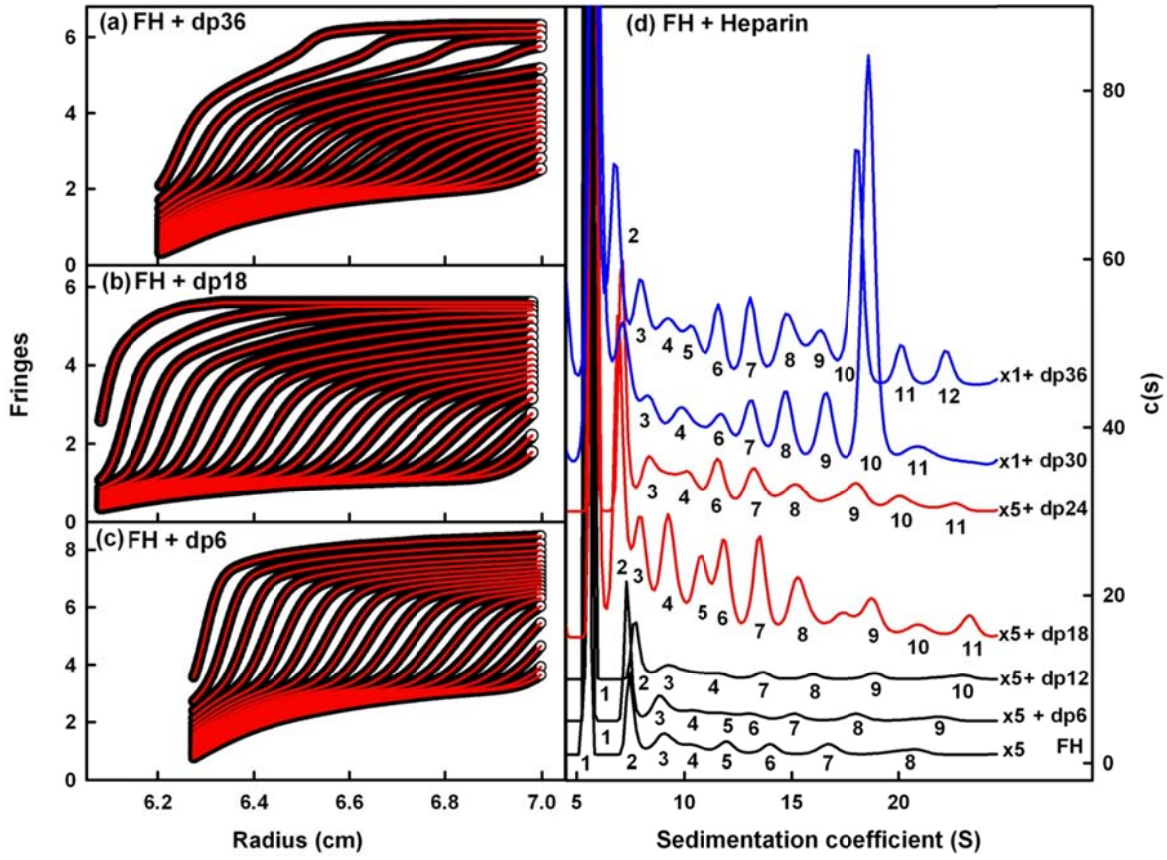


Figure 6.9C

Details are given in Figure 6.9A.

for free FH (Figure 6.10a). Ultracentrifugation indicated that slight decrease in the oligomerisation of FH occurred in the presence of heparin dp6 and dp12 in 1:1 ratios, and this complements the observation of small reductions in the R_G values seen by X-ray scattering. Accordingly, heparin dp6 slightly inhibit FH oligomer formation.

In distinction, the FH mixtures with the larger heparin fragments (dp18-dp36) showed major increases in the oligomer peak intensities between $s_{20,w}$ values of 7 S to 22 S (Figures 6.9Bd and 6.10a). These oligomers amounted to 42 ± 2 %, 29 ± 1 %, 61 ± 2 %, and 63 ± 4% of FH in the dp18, dp24, dp30 and dp36 mixtures respectively (Figure 6.10a). A greater FH oligomerization in the presence of heparin dp30 and dp36 was shown by the exceptional increase in size of peak 8 at an $s_{20,w}$ value of 18.6 S and 18.0 S, which accounted for 49% and 28% of the FH oligomers (Figure 6.9Bd). Peak 8 correspond to higher molecular masses of around 765 kDa and 714 kDa respectively in the $c(M)$ plots for FH-dp30 and FH-dp36, although this calculation assumes that the frictional ratio has not changed, which is unlikely for this large FH oligomer. Moreover, the proportion of dimer also increased with increase in heparin size. In the presence of dp6 and dp12, the dimer proportion was 4 ± 1 % and 5 ± 2 % respectively, which is similar to the free FH value of 4 ± 1 %. With the larger heparin fragments, the proportion of dimer was increased to 11 ± 2 % (dp18), 7 ± 1 % (dp24), 9 ± 1 % (dp30) and 14 ± 1% (dp36) (Figure 6.10a). In summary, these results suggest that heparin dp6 and dp12 are not large enough to crosslink FH, but may inhibit the formation of those observed in free FH. Starting with dp18, heparin becomes long enough to crosslink FH molecules to form a different set of large oligomers.

A further examination of the *c(s)* analyses extended these conclusions. The FH monomer peak was found at the same position for all the FH-heparin mixtures (Figure 6.9Bd; Table 6.1), which is as expected because the FH monomer will be almost unaffected by the binding of a single heparin molecule (1.88-11.30 kDa in mass, compared to 154 kDa for FH). The lack of change in the S values shows that no major conformational change occurs when heparin binds to monomeric FH. The FH dimer peaks in the dp18-dp36 complexes shifted towards lower S values by about 0.7 S compared to the peaks seen with dp6 and dp12 (Figure 6.9Bd; Table 6.1). This is explained by the formation of a new type of dimer for FH-dp18 onwards with a

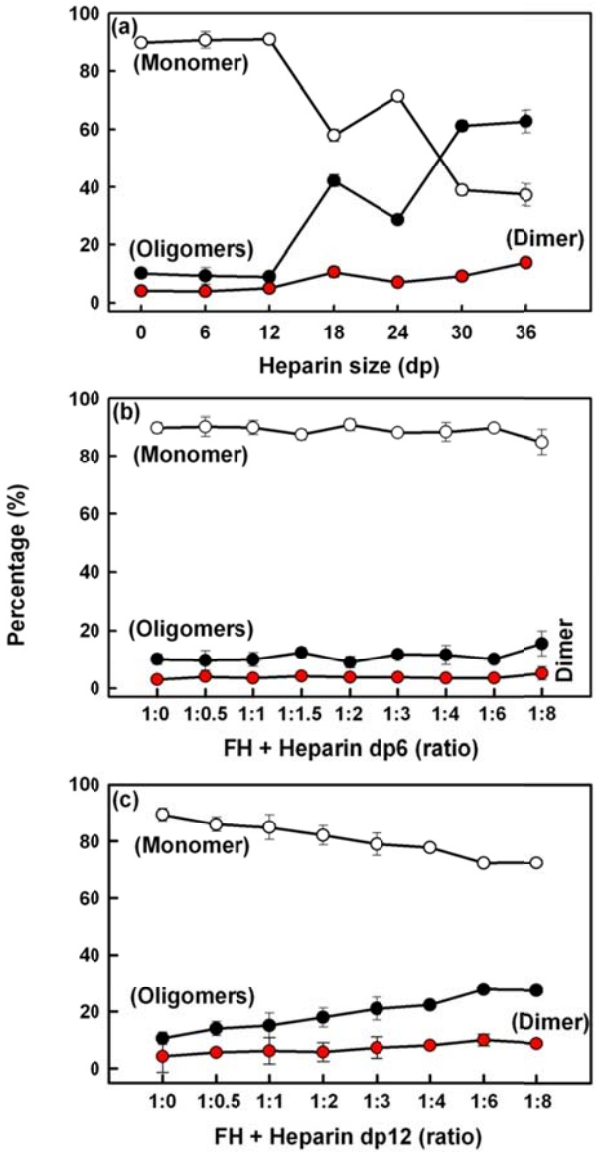


Figure 6.10

Summary of the *c(s)* plots of FH-heparin mixtures. The percentages of monomer (open circles), dimer (red circles) and oligomer (black circles) were derived by integration of the *c(s)* analyses and correspond to the mean value obtained from experiments performed at 40,000 rpm, 50,000 rpm and 60,000 rpm. (a) The relative proportions of FH monomer, dimer and oligomer for 1:1 molar ratios of FH with each of heparin dp6 to dp36. The experimental data are represented in Figure 6.9. (b) The relative proportions of FH monomer, dimer and oligomer for molar ratios of FH-dp6 that ranged from 1:0.5 to 1:8. The experimental data are represented in Figure 6.11. (c) The relative proportions of FH monomer, dimer and oligomer for molar ratios of FH-dp12 that ranged from 1:0.5 to 1:8. The experimental data are represented in Figure 6.12.

more extended conformation than that for the free FH dimer. The comparison of three rotor speeds (Figures 6.9A, 6.9B and 6.9C) showed no change in the peak positions, although there was significant oligomer peak broadening at 40,000 rpm and peak narrowing at 60,000 rpm. The absence of peak shifts between Figures 6.9A, 6.9B and 6.9C indicate that the FH-heparin complexes showed slow association and dissociation rates compared to sedimentation rates, i.e. those reaction boundaries do not appear to be present.

Sedimentation velocity experiments for FH-heparin dp6 mixtures at 50,000 rpm were performed in order to assess whether dp6 was bound to FH. These employed molar ratios that ranged from 1:0.5 to 1:8 at a FH concentration of 1.12 mg/ml, and using the same buffer (Figure 6.11). In the *c(s)* analyses, peaks 1 and 2 at 5.5 S and 7.5 S correspond to the FH monomer and dimer, while peak A at 1.1 S correspond to unbound dp6. For the molar ratios of 1:0.5 to 1:4, no unbound dp6 peak was observed at $s_{20,w}$ values of 1.02 S to 1.1 S, indicating that dp6 was fully bound to FH and was incorporated within peak 1. The unbound heparin dp6 peak only appeared at the high molar ratios of 1:6 and 1:8, indicating that there were multiple binding sites for heparin dp6 within FH. Because two major heparin binding sites are known at SCR-7 and SCR-20 in FH, this would explain the absence of the dp6 peak A at low molar ratios of 1:2. The absence of the dp6 peak at molar ratios of 1:3 and 1:4 also indicate an interaction between dp6 and monomeric FH. Here, the two small peaks B and C that become visible at high molar ratios may correspond to reaction boundaries that involve the formation of 1:1 and 1:2 complexes of FH with dp6. This interpretation is complicated by the appearance of peaks B and C in free FH, suggesting that they originate from the presence of unknown buffer components. Integration of the *c(s)* plots showed no changes in the proportions of monomer, dimer and oligomers proportions at all the FH-dp6 ratios (Figure 6.10b). In conclusion, even though no increased FH oligomer formation took place, the changes in the heparin peaks A, B and C are consistent with a FH-heparin dp6 interaction.

Similar sedimentation velocity experiments for FH-heparin dp12 mixtures at 50,000 rpm were performed in order to assess whether an interaction between FH and dp12 was detectable. These used the same range of molar ratios used for dp6 above, based on a FH concentration of 1.0 mg/ml. In distinction to the dp6 experiments, the *c(s)* analyses showed that FH oligomers were promoted with

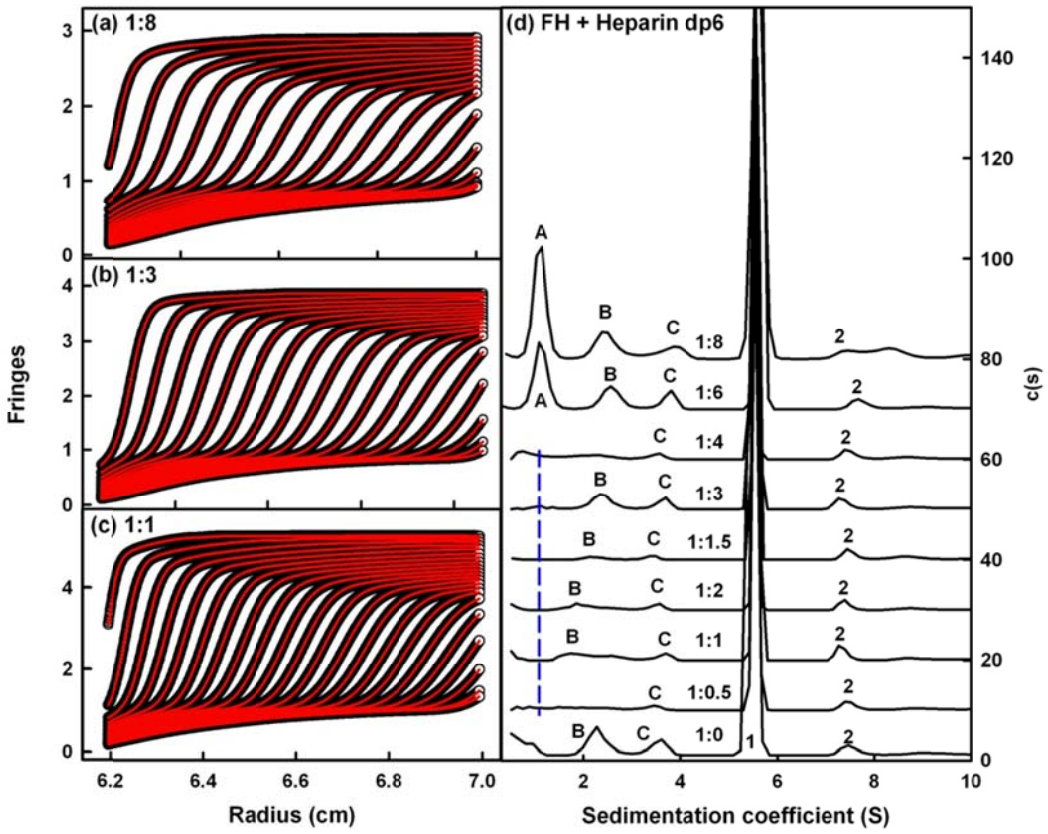


Figure 6.11

The *c(s)* analyses for FH-heparin dp6 mixtures. The sedimentation velocity experiments were performed at 60,000 rpm with 1.12 mg/ml of FH in the presence of dp6 in molar ratios that ranged from 1:0.5 to 1:8. Other details are given in Figure 6.9. (a-c) Every fifth boundary was fitted for the 1:1 and 1:3 molar ratios, and every thirteenth boundary was fitted for the 1:8 molar ratio. (d) In the *c(s)* analyses, the FH monomer peak (labelled as 1) was observed at 5.54 ± 0.04 S, with the dimer peak labelled as 2. Unbound dp6 (labelled as A) was observed at 1.1 ± 0.02 S at the highest molar ratios, and its putative position at lower molar ratios is shown by the vertical blue dashed line. Peaks B and C are discussed in the text.

increase in the dp12 molar ratios (Figure 6.12). Integration showed that no changes were observed for the 1:0.5 and 1:1 ratios, while increased FH oligomerization was observed for the higher ratios up to 28 ± 2 % and 28 ± 1 % for the 1:6 and 1:8 ratios respectively (Figure 6.10c). Interestingly, the proportion of the FH dimer and trimer was increased for the 1:6 and 1:8 ratios, being 10 ± 1 % and 9 ± 1 % of dimer, and 7 ± 1 % and 7 ± 1 % of trimer respectively (Figure 6.10c). In comparison, these proportions were more than halved in free FH, being 4 ± 1 % and 2 ± 1 % respectively. These results indicate that heparin dp12 interacts with FH in a concentration-dependent manner.

Sedimentation velocity experiments for FH-heparin dp36 showed the extent of the peak shifts that occur with increased oligomer formation (Figure 6.13). Figure 6.13 is derived from the FH-dp36 *c(s)* analyses shown in Figures 6.9A, 6.9B and 6.9C. The significant peak broadening at 40,000 rpm and peak narrowing at 60,000 rpm for the FH-dp36 mixture was already noted above. The increased number of peaks seen at 60,000 rpm compared to 40,000 rpm suggested that more than one type of FH-heparin oligomer may be structurally formed for each FH-heparin oligomer.

(6.2.4) Sedimentation velocity data for FH mixtures with HS

To evaluate the FH-HS interaction, sedimentation velocity experiments were also performed for 1:1 FH mixtures with HS dp6, dp12, dp18 and dp24 at two rotor speeds of 40,000 rpm and 50,000 rpm in HEPES buffer. The FH concentration was 0.45 mg/ml. In all the FH-HS mixtures at 50,000 rpm, monomeric FH gave an $s_{20,w}$ value of 5.52 ± 0.04 S and a molecular mass of 132 ± 4 kDa (Figure 6.14). This is in good agreement with the monomer $s_{20,w}$ value of 5.48 ± 0.02 S and mass of 134 ± 4 kDa for free FH at 0.87 mg/ml, and a monomer $s_{20,w}$ value of 5.67 ± 0.07 S and mass of 132 ± 4 kDa observed for the FH-heparin complexes. Even though the FH concentration was halved compared to that used with heparin (above), the oligomer peaks remained visible in the *c(s)* plots. In the presence of HS dp6, five oligomeric species were observed at $s_{20,w}$ values of 8.0 S, 9.8 S, 12.4 S, 14.9 S and 18.5 S for dimer to pentamer respectively (Figure 6.14). In the presence of HS dp12, six oligomer peaks were observed at slightly altered $s_{20,w}$ values of 7.7 S, 9.2 S, 10.5 S, 12.7 S, 14.8 S and 17.4 S. The proportion of the FH oligomers in the presence of HS dp6 and HS dp12 at 50,000 rpm were 8% and 11% respectively, which is slightly

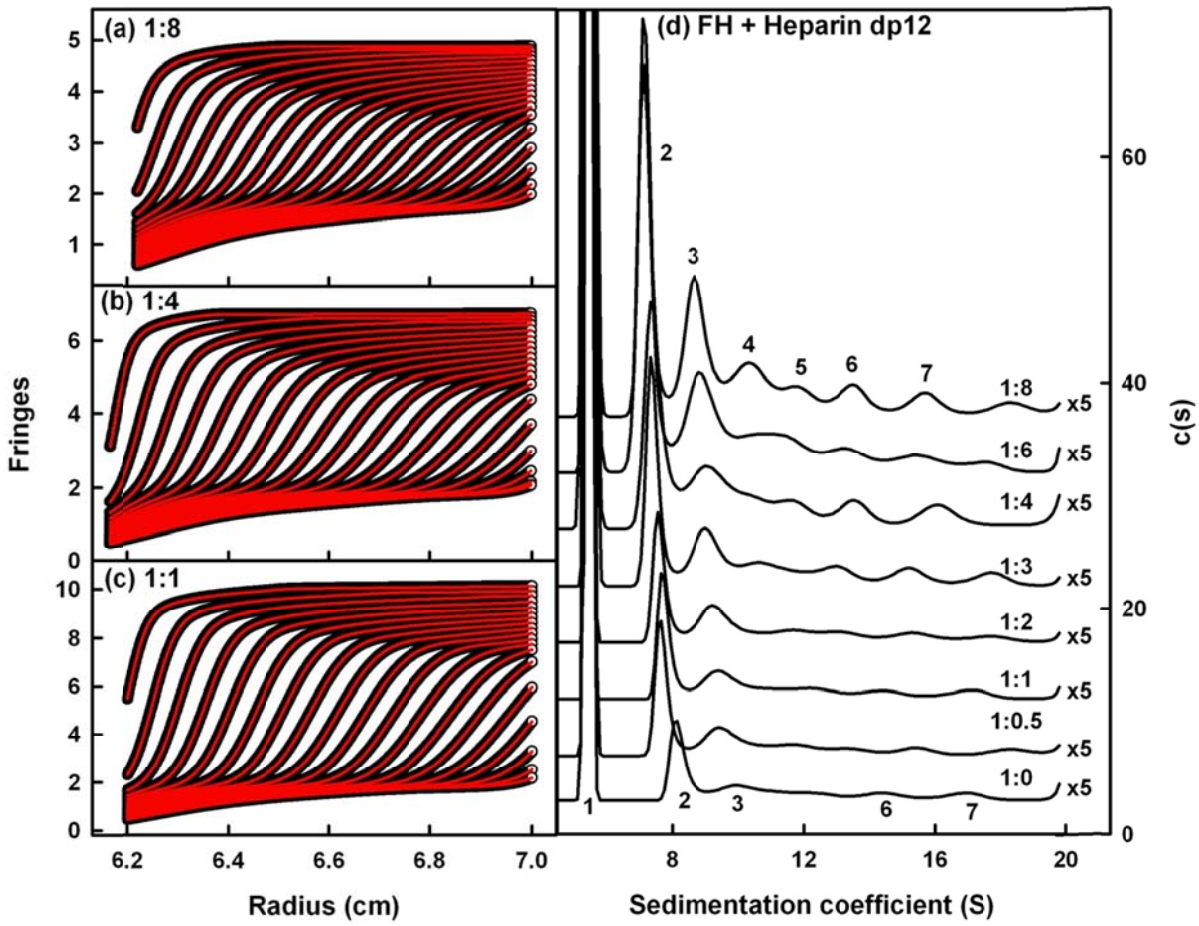


Figure 6.12

The *c(s)* analyses for FH-heparin dp12 mixtures. The sedimentation velocity experiments were performed at 50,000 rpm with 1.0 mg/ml of FH in the presence of dp12 in molar ratios that ranged from 1:0.5 to 1:8. Other details are given in Figure 6.9. (a-c) Every ninth boundary was fitted for the 1:1 and 1:4 ratios, and every eight boundary was fitted for the 1:8 ratio. (d) In the *c(s)* analyses, the intensities were multiplied by a factor of 5 in all cases. The FH monomer peak (labelled as 1) was observed at 5.46 ± 0.03 S. The successive FH oligomer peaks are numbered from 2-7 with increase of their sedimentation coefficients.

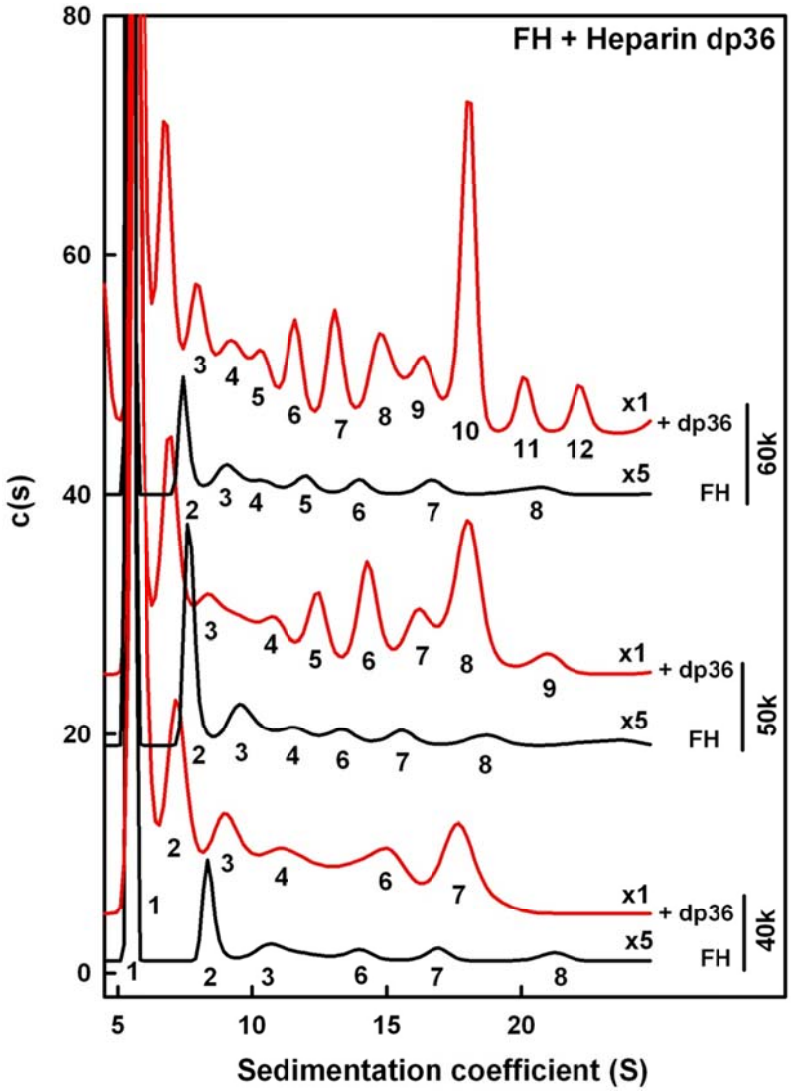


Figure 6.13

The *c(s)* analyses for FH-heparin dp36 mixtures at rotor speeds of 40,000 rpm, 50,000 rpm and 60,000 rpm. The FH analyses are show in black, and its mixtures with dp36 are shown in blue. To improve the visibility of the FH oligomer peaks, the intensities were multiplied by a factor of 5 for the **F**H plots only. The successive FH oligomer peaks are numbered from 2-12 with increase of their sedimentation coefficients.

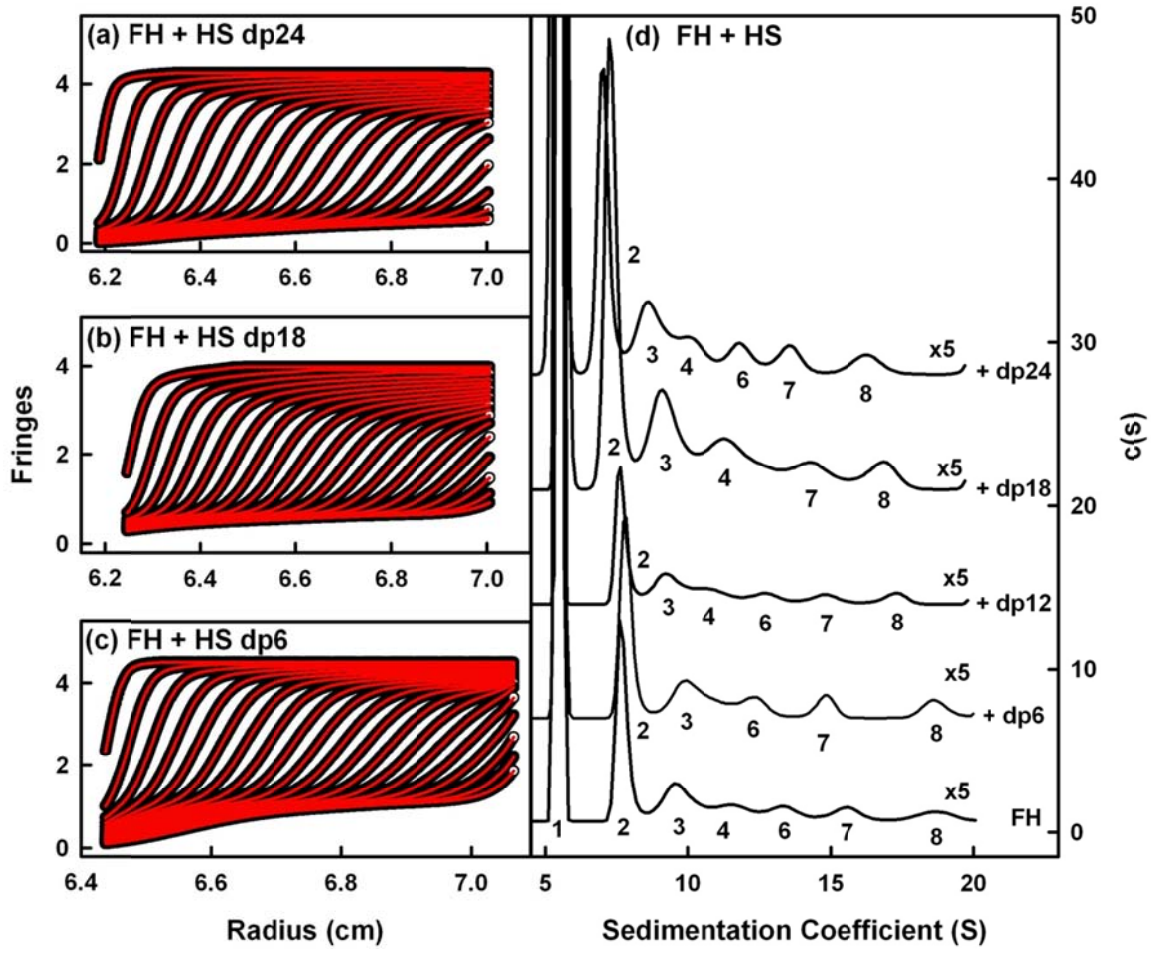


Figure 6.14

The *c(s)* analyses for FH titrated with four HS dp6-dp24 fragments mixture. The sedimentation velocity experiments were performed at 50,000 rpm with 0.45 mg/ml FH in the presence of 1:1 molar ratios of HS dp6 – HS dp24. Other details are given in Figure 6.9. (a) Every sixteenth boundary scan of 350 scans was fitted for the FH-HS dp24 mixture. (b) Every fifteenth boundary scan of 300 scans was fitted for the FH-HS dp18 mixture. (c) Every twelfth boundary scan of 240 scans was fitted for the FH-HS dp6 mixture. (d) The *c*(*s*) analyses for the FH-HS mixtures are shown. For greater visibility of the FH oligomer peaks, the intensities were multiplied by a factor of 5 in all cases. The FH monomer peak is denoted as 1, and the FH oligomer peaks are denoted by 2-8 in increasing order of sedimentation coefficient values.

lower than the 12% proportion of oligomers in free FH (Figure 6.15). However the oligomer proportion of FH in the presence of HS dp18 and dp24 was higher at 21% and 14% respectively. As for the FH-heparin mixtures, these results suggest that the small HS fragments dp6 and dp12 slightly inhibit FH oligomerisation, while the larger HS fragments increase FH oligomerisation (Figure 6.15). Given that the proportion of FH oligomer in the presence of heparin dp18 was larger than that for FH-heparin dp24 (Figure 6.10), it was of interest that the same has occurred with HS dp18 and HS dp24 (Figure 6.15). For both FH-heparin and FH-HS, the dimer peak is more pronounced in the mixtures of FH with dp18/dp24 and less so for dp6/dp12 (Figures 6.9 and 6.14). As for the FH-heparin mixtures, the FH oligomers in the presence of HS dp18 and HS dp24 were also shifted towards lower $s_{20,w}$ values, indicating changes in the oligomer shapes to more extended conformations when these incorporated the larger HS fragments (Figures 6.13 and 6.14). In summary, similar results were obtained for FH-HS mixtures when these were compared to those for FH-heparin mixtures, although the magnitudes of the oligomerisation changes were reduced for HS when these were compared with heparin (Figures 6.10 and 6.15).

(6.2.5) Modelling of FH-heparin dp36 complexes

Molecular modelling of the scattering and ultracentrifugation data provides a validation of the experimental data and their interpretation (Perkins *et al*., 2008, 2009). In this instance, scattering modelling for the FH-heparin and FH-HS mixtures were not feasible because of insufficient monodispersity in the samples that were mixtures of free and complexed protein, the latter of which formed a series of oligomers. Ultracentrifugation modelling of the experimental $s_{20,w}$ values had previously explained the peaks 2-7 to arise from dimeric to heptameric forms of FH, starting from the 2001 X-ray scattering solution structure (Aslam & Perkins, 2001; Nan *et al*., 2008). Here, the SCR-7 and SCR-20 domains in FH were joined up with heparin dp36 to create models for dimeric, trimeric, tetrameric and pentameric oligomeric complexes (Figure 6.16). It should be noted that scattering modelling was not able to identify specific SCR domain arrangements; only that the SCR domains are folded back. Accordingly the relative locations of SCR-7 and SCR-20 are putative only. While ring-like structures were readily modelled, alternative dimeric structures and tetramer/pentamer structures forms could also be modelled. Starting

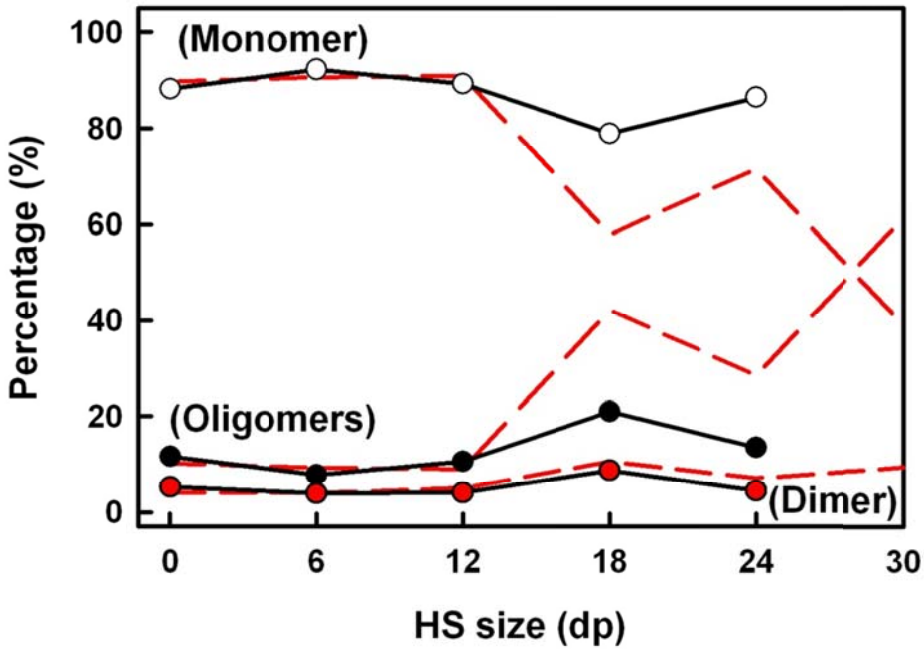


Figure 6.15

Summary of the *c(s)* plots of FH-HS mixtures. The details follow those in Figure 6.13, except that only a single rotor speed at 50,000 rpm was used. The relative proportions of FH monomer, dimer and oligomer for 1:1 molar ratios of FH with each of HS dp6 to dp24. The experimental data are from Figure 6.14. The red dashed lines correspond to the FH-heparin data of Figure 6.10a.

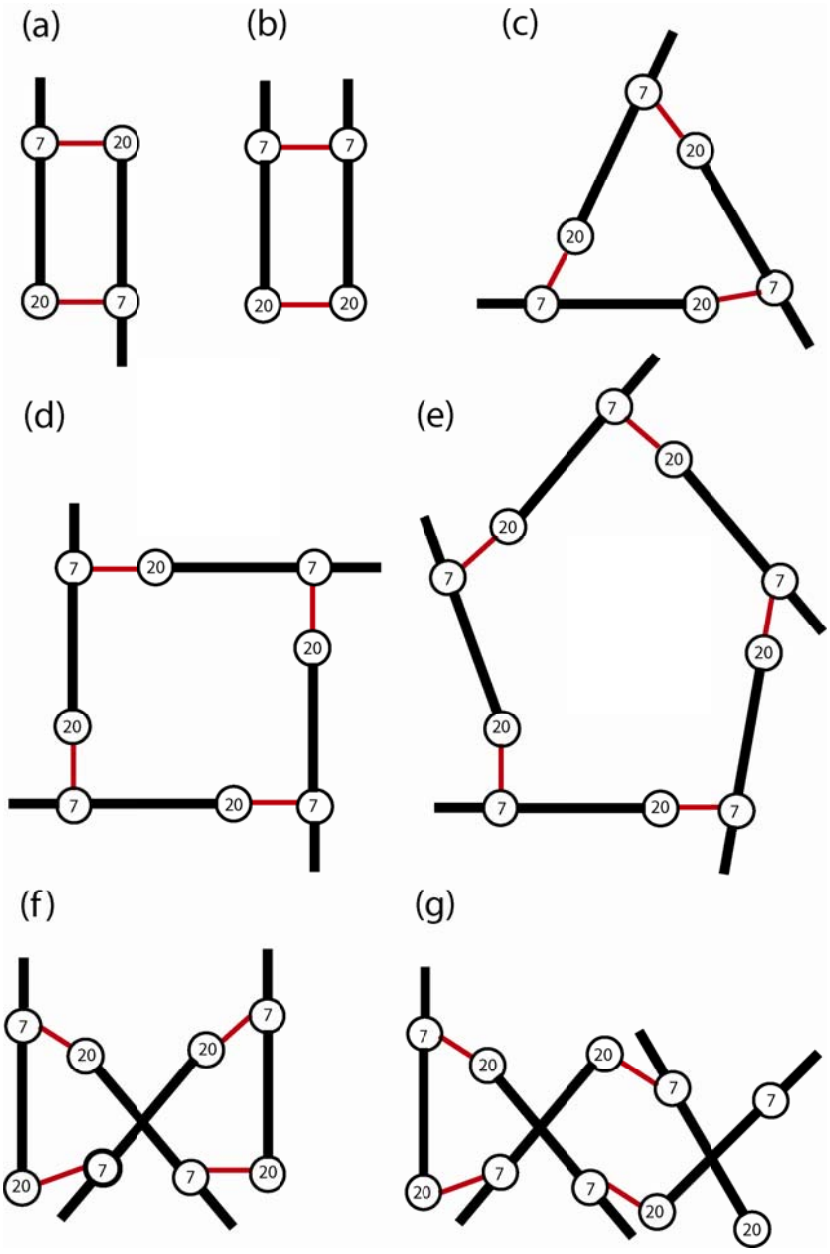


Figure 6.16

Cartoon of seven possible models for FH-heparin dp36 complexes. FH is shown in a thick black outline, while dp36 is shown in a thin red outline. (a,b) The dimer models were created by the cross linking of two FH molecules by two dp36 molecules. Two alternative schemes are represented by contacts between SCR-7↔SCR-20 and SCR-20↔SCR-7 (antiparallel) and between SCR7↔SCR7 and SCR20↔SCR20 (parallel). (c-g) Higher oligomer models were created though contacts based on alternating arrangements of SCR-7↔SCR-20 and SCR-20↔SCR-7 for the trimer, tetramer and pentamer complexes. The first three comprise ring-like structures, while the last pair comprises more compact arrangements.

from a recent FH model based on fitting the 20 SCR domains with 19 flexible linkers between the domains (PDB code 3GAV) (Okemefuna *et al*., 2009) in combination with the heparin dp36 model (PDB code 3IRL) (Chapter 4: Figure 4.11d), the FH-heparin complexes were created manually from the best-fit solution structures of FH and dp36 using Discovery Studio molecular graphics. In this FH model, the SCR-7 and SCR-20 domains were separated by 27 nm.

The outcome of the modelling is summarised:

(a) The predicted $s^0_{20,w}$ value for an antiparallel dimer model created from two dp36 and two FH molecules was 7.5 S. This agreed well with the observed values of 7.0-7.2 S for dp18-dp36, given that the method is accurate to ± 0.21 S (Perkins *et al*., 2009) (Figures 6.16a and 6.17a; Table 6.1). A parallel dimer model that was created using these four molecules was more compact in its shape and gave a larger predicted $s^0_{20,w}$ value of 8.1 S (Figures 6.16b and 6.17b, Table 6.1). An antiparallel dimer was favoured by this calculation.

(b) The predicted $s^0_{20,w}$ values for the trimeric, tetrameric and pentameric ring-like complexes were 8.8 S, 9.5 S and 10.6 S, in good agreement with the experimental peaks observed at 8.2-8.6 S, 9.1-9.5 S and 10.4-10.8 S for dp18-dp36 (Figures 6.16(c-e) and 6.17(c-e); Table 6.1). In distinction to this agreement, the more compact crossed-over tetrameric and pentameric gave predicted $s^0_{20,w}$ values of 11.3 S and 12.4 S that were almost 2 S higher than the experimental values (Figures 6.16(f,g) and 6.17(f,g); Table 6.1).

(c) Control calculations were performed using other scattering best-fit FH models. One was derived from X-ray scattering curve fits for homozygous FH, based on fitting five rigid SCR fragments with four flexible linkers between the domains (PDB code 3N8O) (A. I. Okemefuna & S. J. Perkins, unpublished modelling). Here SCR-7 and SCR-20 were positioned more closely together with a separation of 10 nm. The other was derived from X-ray scattering curve fits for homozygous FH, based on fitting the 20 SCR domains with 19 flexible linkers between the domains (PDB code 3N0J) (Nan, *et al*., 2010). Here SCR-7 and SCR-20 were separated by 13 nm. Similar FH-heparin models were created (Figures 6.18 and 6.19). The outcome of the modelling (Table 6.1) showed that the predicted $s^0_{20,w}$ values were higher by about 1 S compared with the 2009 model. This difference is attributed to the closer separations of 10 nm and 13 nm between SCR-7 and SCR-20 in these two extra FH models, compared to 27 nm in the first model. The two extra FH models were too

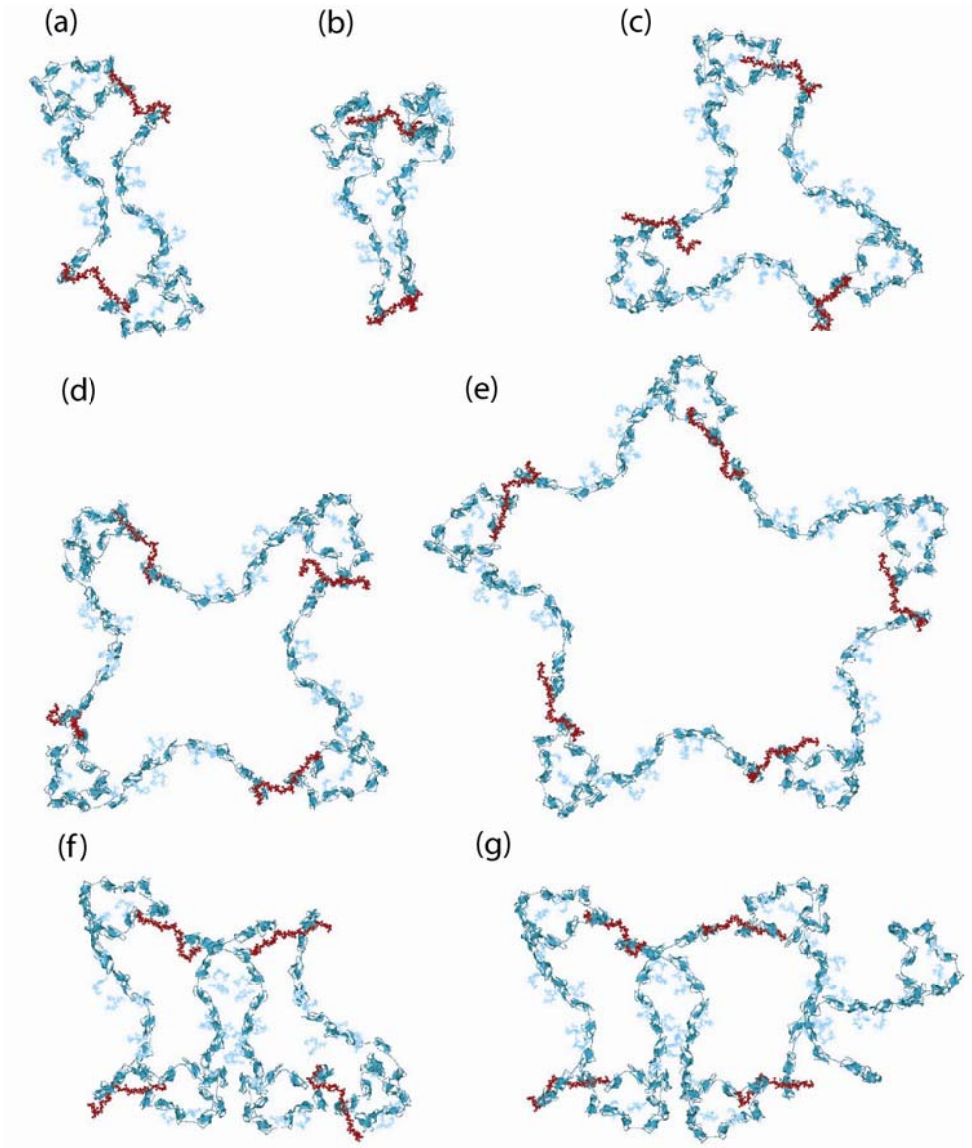


Figure 6.17

Seven possible molecular models for the FH-heparin dp36 complex. The models were created from the best-fit X-ray scattering structure for FH (PDB code 3GAV; (Okemefuna et al., 2009) and that for dp36 (PDB code 3IRL). The predicted $s^0_{20,w}$ values for these models are reported in Table 6.1. FH is shown in cyan, with its eight oligosaccharide chains shown in light blue, while heparin dp36 is shown in red. (a,b) Antiparallel and parallel dimer models of FH and heparin dp36 complexes. (c-e) Trimer, tetramer and pentamer ring-like arrangements of FH and heparin dp36 complexes. (f,g) Tetramer and pentamer compact arrangements of FH and heparin dp36 complexes.

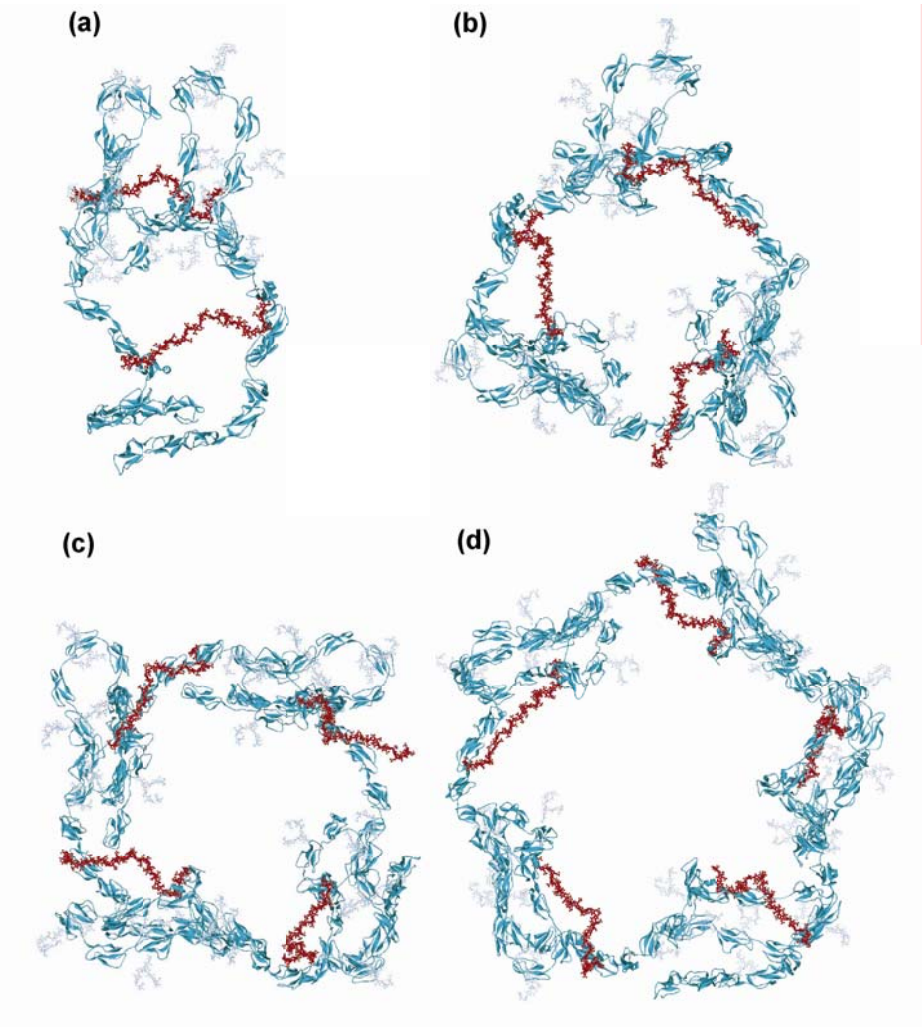


Figure 6.18

Four possible molecular models for the FH-heparin dp36 complex. The models were created from the best-fit X-ray scattering structure for FH (PDB code 3N8O; (Okemefuna *et al*., 2011, in preparation) and that for dp36 (PDB code 3IRL). The predicted $s^0_{20,w}$ values for these models are reported in Table 6.1. FH is shown in cyan, with its eight oligosaccharide chains shown in light blue, while heparin dp36 is shown in red. (a) Parallel dimer models of FH and heparin dp36 complexes. (b-d) Trimer, tetramer and pentamer ring-like arrangements of FH and heparin dp36 complexes.

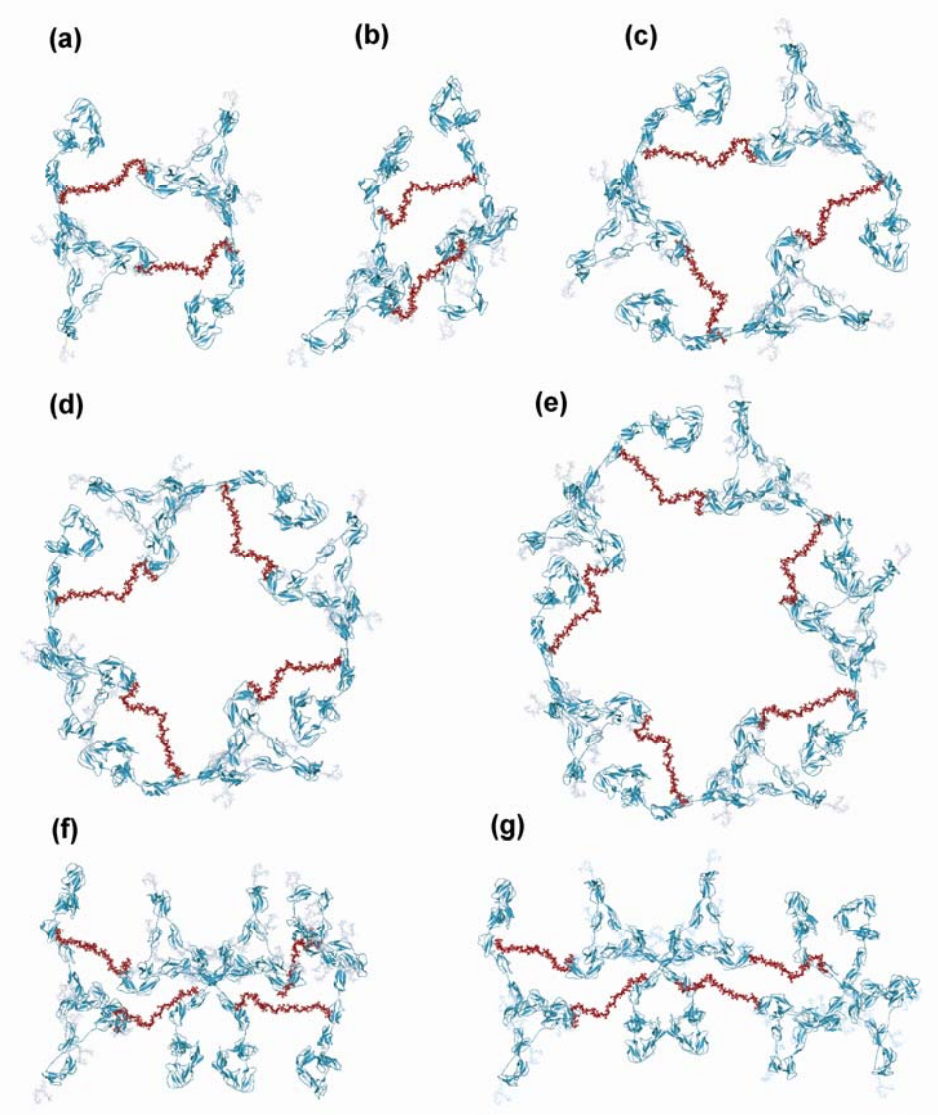


Figure 6.19

Seven possible molecular models for the FH-heparin dp36 complex. The models were created from the best-fit X-ray scattering structure for FH (PDB code 3N0J; (Nan, 2010) and that for dp36 (PDB code 3IRL). The predicted $s^0_{20,w}$ values for these models are reported in Table 6.1. FH is shown in cyan, with its eight oligosaccharide chains shown in light blue, while heparin dp36 is shown in red. (a,b) Antiparallel and parallel dimer models of FH and heparin dp36 complexes. (c-e) Trimer, tetramer and pentamer ring-like arrangements of FH and heparin dp36 complexes. (f,g) Tetramer and pentamer compact arrangements of FH and heparin dp36 complexes.

compact to explain the experimental $s_{20,w}$ values.

(6.3) Conclusions

Heparin and HS interact with a wide range of proteins that are involved in biological and physiological processes (Capila & Linhardt, 2002). The most prominent type of interaction between heparin or HS and proteins is ionic, being provided by the sulphate and carboxyl groups (Capila & Linhardt, 2002). FH performs its regulatory function by recognising polyanionic molecules present on the host cell surfaces that are absent on bacterial surfaces, thereby positioning itself for the control of activated C3b that would otherwise damage the host. Despite the biological importance of FH, there has been no structural understanding up to now of how full-length FH interacts with heparin and HS fragments. Polyanionic molecules such as HS and sialic acids on the host cell surface enhance the regulatory effectiveness of FH by 10-fold through its inhibition of complement activation (Pangburn, 2000; Meri & Pangburn, 1990; Pangburn & Müller-Eberhard, 1978). In this study, we have studied the heparin and HS interactions with FH as a function of oligosaccharide size. Our results provide four novel insights into the FH interaction with heparin and HS in near-physiological salt by showing that (i) both polyanionic oligosaccharides promote strong FH oligomer formation; (ii) at least two heparin binding sites on FH are involved; (iii) little structural changes in FH occur after heparin or HS binding; and (iv) the higher sulphation of heparin leads to strengthened FH binding when compared to HS.

FH is a major complement regulator found in plasma at concentrations of 0.235 mg/ml to 0.81 mg/ml (Saunders *et al*., 2006). Free FH self associates to form 12% oligomers at physiological concentrations (Nan *et al*., 2008), and these oligomers persist over a range of salt concentrations and pH (Okemefuna *et al*., 2009). The SCR domains responsible for FH self-association sites are not yet known, but are found within at least the SCR-6/8 and SCR-16/20 domains, both of which also bind to heparin (Fernando *et al*., 2007; Okemefuna *et al*., 2008). An understanding of FH oligomers was central to interpret the interaction between FH and C-reactive protein, given that C-reactive protein inhibited FH oligomer formation (Okemefuna *et al*., 2010). Here, our solution scattering and analytical ultracentrifugation data showed that small fragments of heparin dp6 and dp12 showed a slight decrease in the R_G and *I(0)/c* values of FH (Figure 6.4) and a slight

decrease in the proportion of the FH oligomers seen in the *c(s)* plots (Figure 6.10). The small HS fragments dp6 and dp12 also resulted in slight decreases in FH oligomer formation. Even though these changes were at the level of experimental error, they were consistently seen in all four cases. Further experiments with higher molar ratios of heparin dp6 and dp12 showed that both interacted with FH (Figures 6.11 and 6.12). This was corroborated by comparative ultracentrifugation experiments with the SCR-6/8 fragment in the presence of 1:0.3 and 1:2 molar ratios of heparin dp10 that showed these interacted (Fernando *et al*., 2007). Overall, these data demonstrate that the small fragments of heparin/HS in 1:1 molar ratios were not long enough to cross-link FH molecules to promote FH oligomer formation, and that the FH oligomer sites within SCR-6/8 and SCR-16/20 may overlap with heparin/HS binding sites.

Here, in the presence of large heparin and HS oligosaccharides, different set of FH oligomers were identified that were strongly promoted by the use of 1:1 molar ratios of heparin dp18-dp36 and HS dp18-dp24 (Figures 6.4, 6.7, 6.10 and 6.15). These are consistent with the reports of FH dimer and tetramer formed in the presence of dextran sulphate and high molecular weight unfractionated heparin (12-14 kDa; approximately dp38-dp46) (Pangburn, *et al*., 2009). That 2009 study was based on sedimentation equilibrium data, which averages the ultracentrifugation data; this approach was less informative that our present use of *c(s)* analyses in sedimentation velocity experiments which directly observed FH-heparin oligomers as individual peaks ranging from dimer to octamer. The higher FH-heparin and FH-HS oligomers are distinct from those seen with FH and >10 μM zinc, when large polydisperse FH oligomers were formed through the presence of weak zinc binding sites on the surface of FH that cross-link FH monomers (Nan *et al*., 2008; 2011). No discrete FH-zinc oligomers were observed, unlike those seen for FH-heparin and FH-HS (Figures 6.9 and 6.14), suggesting that the interaction between FH and heparin/HS is a well defined one at the molecular level. This is explained by the occurrence of two separate heparin binding sites within SCR-7 and SCR-19/20 (Pangburn *et al*., 1991; Blackmore *et al*., 1996, 1998). Provided that the heparin structure is long enough (Figure 6.1), these two sites can be joined together by single heparin molecules in order to form cross-linked oligomers (Figure 6.17; Table 6.1). The larger heparin fragments dp18-dp36 (5.65-11.30 kDa) resulted in a dramatic increase in the oligomerisation of FH. The proportion of FH oligomers increased to

29-63% (Figure 6.10), which is as much as a five-fold increase in FH oligomerisation observed with heparin dp36 compared to the 12% of oligomers observed for free FH. In addition to FH oligomer formation, the movement of the *c(s)* peak positions towards lower S values showed that the FH oligomers became more extended in their shapes. The scattering data also showed that the larger heparin fragments increase FH oligomer formation dramatically (Figure 6.4).

The formation of large FH oligomers by heparin and HS provides new insight into the structural properties of FH. Even though there are two heparin binding sites in FH, there were no indications that heparin dp18-dp36 caused a large conformational change in FH to bring these two sites together and result in a 1:1 complex of FH and heparin. The spatial separation of the SCR-7 and SCR-19/20 sites would need to be about 6-12 nm apart for such a 1: 1 complex to form with dp18-dp36. Such a conformational change to a more compact FH structure would have caused visible increases in the $s_{20,w}$ values of the FH monomer peak. These were never observed in the *c(s)* analyses (Figures 6.9 and 6.14; Table 6.1). The molecular modelling of FH oligomer structures confirmed this reasoning. Good agreements between the experimental and predicted $s_{20,w}$ values for the FH-heparin dimer to pentamer were obtained when the heparin dp36 models were combined with FH models in which the SCR-7 and SCR-20 domains were separated by 26 nm (Figures 6.1 and 6.17; Table 6.1). If the SCR-7 and SCR-20 domains were positioned closer together at 10-13 nm, the predicted $s_{20,w}$ values of the FH-heparin oligomers became about 1-2 S larger than the experimental values (Figures 6.18 and 6.19; Table 6.1). This outcome indicates that the inter-SCR domain linkers in the FH SCR domain structure are much less flexible than anticipated in earlier structural studies, where its folded-back structure shows little variation as a function of salt, pH or heparin ligand (Okemefuna *et al*., 2009). If the two heparin binding sites in FH are independent of each other, this implies that the role of the heavily glycosylated and smaller SCR-12/15 domains at the centre of FH is to maintain the functional independence of the N-terminal and C-terminal ends of FH.

The comparative use of heparin and HS in the study of FH oligomers clarifies the binding of anionic oligosaccharides to FH. Constrained scattering modelling for the heparin fragments showed semi-rigid and partially bent structures, and this becomes more evident when the heparin sizes increase beyond dp18 (Chapter 4: Figure 4.11). The bends arise through kinks at several positions in heparin dp18-

dp36, and kinks have also been seen in protein crystal structures such as those between FGF1-FGFR2 and heparin dp10 (Pelligrini *et al*., 2000). The kink in heparin provided optimal electrostatic and hydrogen-bond contacts with the protein (Raman *et al*., 2003), and these kinks may facilitate optimal contact formation between FH and heparin in their complexes. By ultracentrifugation, it was of interest that the oligomer peak 8 was much increased in size with heparin dp30 and dp36 (Figure 6.9 and 6.13). Given the existence of an inflexible FH domain structure and a semi-rigid heparin structure, it follows that some oligomer arrangements with heparin dp30 and dp36 may be energetically more favoured and their relative proportions will increase. In distinction to heparin, the HS fragments used in the present study corresponds to the long *NA*-domains with GlcA and GlcNAc residues, in which the highly-sulphated *S*-domains with IdoA2S and GlcNS residues have been removed by heparinase I digests. For HS, constrained scattering modelling for dp6-dp16 showed a less rigid structure that was longer and more bent than that of heparin, and such a structure became more evident for the largest HS dp18 and dp24 fragments which could not be modelled as a single structure using the X-ray scattering curve (Chapter 5). The scattering and ultracentrifugation data for the FH-HS mixtures are fully consistent with the FH-heparin data in terms of FH oligomer formation. Two major differences from the FH-heparin complexes were observed. Firstly, HS caused reduced levels of FH oligomers to form (Figures 6.10 and 6.15). This is attributed to the reduced number of sulphate groups in HS compared to those in heparin. Secondly, the relative proportions of the peak intensities for the individual FH-HS oligomers decreased uniformly with increased size of the HS fragment (Figure 6.14), compared to the less uniform decrease observed with the heparin fragments (Figure 6.9). This is attributed to the greater flexibility of the HS fragments, which are better able to optimise their conformations to match those presented by the SCR-7 and SCR-20 sites in a comparatively inflexible FH SCR domain structure.

The identification of at least two independently-binding heparin sites on FH by the observation of FH-heparin oligomers has functional implications. The need for heparin binding to two different FH sites will confer a greater selectivity of FH for host cells that have a high enough density of heparin-like or HS-like surface oligosaccharides, unlike that found at bacterial surfaces. Two separate weak binding events with μM affinities will become an overall interaction with a high affinity of nM if both sites on FH bind to a single location at the host cell surface (Jencks,

1981). This has implications for disease. Common polymorphisms contribute to genetic disease in a manner that is distinct from rare mutations. Disease-risk polymorphisms will exert their effect over a period of decades, while disease-causing mutations will show a much earlier effect in life. Mechanisms for FH-mediated disease involve either the aggregation of FH to form pathogenic deposits or the reduced control of inflammatory events by FH (Perkins *et al*., 2010). Age-related macular degeneration is common in the aged population, being responsible for over 50% of blindness in the elderly in the Western world (Bird, 2010), while atypical haemolytic uraemic syndrome affects young individuals, typically after an immunological insult to the kidney (Goodship & Kavanagh, 2010). FH is found in the drusen deposits of the retina that are a hall-mark of age-related macular degeneration (Hageman *et al*., 2005). A single FH polymorphism at Tyr402His in SCR-7 has been associated with many cases of age-related macular degeneration (Klein *et al*., 2005; Haines *et al*., 2005; Edwards *et al*., 2005; Hageman *et al*., 2005). An investigation of glycosaminoglycans within Bruch's membrane found little evidence of chondroitin sulphate and HS in drusen, and the size and structure of glycosaminoglycans in drusen is not yet known (Kliffen *et al*., 1996). If the drusen deposits contains glycosaminoglycans of size dp18 or larger, their presence would enable FH to aggregate continuously to form larger aggregates resembling those seen when zinc was added to intact FH (Nan *et al*., 2008). It is alternatively possible that the weaker binding of the Tyr402His FH allotype to host cell surfaces may trigger a higher level of immune attack and inflammation if the His402 allotype binds more weakly to heparin or HS. The evidence for this is currently mixed, where some studies state that there is weaker heparin and HS binding with His402, while others state that there is no difference (Clark *et al*., 2006, 2010). The loss of either the SCR-7 or the SCR-19/20 heparin binding function in FH will compromise its regulatory function, and will lead to disease. The present ultracentrifugation and scattering studies now open the way for more detailed studies of the FH-heparin and FH-HS interactions that will enable these molecular mechanisms to be deciphered.

(6.4) Materials and methods

(6.4.1) Purification of FH and the heparin/HS fragments

Native FH was purified in collaboration with Dr. Ruodan Nan (Department of Structural and Molecular Biology, UCL) from pooled frozen human plasma using

monoclonal affinity chromatography with MRC-OX23 column (Sim *et al*., 1993). To prevent FH from proteolytic cleavage, 5 mM EDTA and 0.5 mM protease inhibitor Pefabloc-SC[70] were added into 50 ml of this anonymised human plasma. The mixed plasma was then dialysed against Tris buffer (25 mM Tris-HCl, 140 mM NaCl, 0.5 mM EDTA, pH 7.4). After dialysis, the plasma was centrifuged at 10,000 rpm at 4ºC. In order to remove IgG binding proteins, the plasma was passed through a column of non-immune IgG immobilized on Sepharose. The plasma was then passed through a lysine-sepharose column to remove plasminogen/plasmin. After that, the plasma was passed through MRC-OX23-Sepharose, a monoclonal antibody column specific for FH, at a rate less than 1 ml/min. Bound FH was eluted with 3 M $MgCl_2$, pH 6.9. To remove Mg^{2+}, the eluant was dialysed against 4 litres of HEPES buffer (10 mM HEPES, 137 mM NaCl, pH 7.4) with 0.5 mM EDTA and then with 4 litres of HEPES buffer without EDTA. After dialysis the plasma was passed through a Hitrap Protein G HP column to remove contaminant IgG. The volume of purified FH was reduced using Amicon® Ultra-15 centrifugal filter with a molecular weight cut-off of 50 kDa at 2,000 rpm. Non-specific aggregates and human serum albumin were removed using Superose™ 6 gel-filtration column in HEPES buffer at flow rate of 1 ml/min (Figure 6.2(a)). Fractions corresponding to the major FH peak were pooled. The purified sample was reconcentrated by centrifugation. The purity of FH was checked by sodium dodecyl sulphate polyacrylamide gel elctrophorsis (SDS-PAGE) (Figure 6.2(b)). SDS-PAGE was performed by using NuPAGE® Novex 3-8% Tris-Acetate mini gel and the standard protocol. The FH concentration was determined from an absorption coefficient (1%, 1 cm) of 16.7 (Okemefuna *et al*., 2009). FH was dialysed into HEPES buffer for scattering and ultracentrifugation experiments, and its integrity was routinely checked by SDS-PAGE before and after experiments.

The dp6-dp36 fragments of heparin were prepared from bovine lung heparin as described in Chapter 4: section 4.4.1. In summary, about 80 mg of heparinase-I-digested heparin was subjected to chromatography in order to isolate the dp2-dp16 fragments as well-resolved peaks and the dp18-dp36 fragments as slightly less well-resolved species. The high monodispersity of the six dp6-dp36 fragments was shown by the single peaks observed by analytical ultracentrifugation *c(s)* size distribution analyses (Chapter 4: Figures 4.4, 4.5 and 4.6). Heparin concentrations were measured by weighing the lyophilised product after evaporation of 2% ammonium bicarbonate under reduced pressure (Chapter 4: section 4.4.1).

HS is comprised of short *S*-domains (IdoA2S and GlcNS residues), long *NA*-domains with GlcA and GlcNAc residues, and mixed domain regions at the junctions between the *S*-domains and *NA*-domains (Capila & Linhardt, 2002; Coombe & Kett, 2005). The dp4-dp24 fragments of HS were prepared following exhaustive heparinase I digestion of 100 mg of a HS stock to leave only the *NA*-domains by minimizing the presence of fully sulphated *S*-domain sequences (Chapter 5: section 5.4.1). The four dp6-dp12 fragments were eluted as well-resolved peaks, while the four larger fragments dp14-dp18 and dp24 were slightly less well resolved. Monodispersity was demonstrated by PAGE and *c(s)* analyses (Chapter 5: Figures 5.2 and 5.4). HS concentrations were measured using an absorption coefficient of 5500 M^{-1} cm^{-1} (Powell *et al*., 2010).

(6.4.2) X-ray scattering data for FH mixtures with heparin and HS

X-ray scattering data for mixtures of FH with either six heparin or four HS fragments were fully acquired in one of two sessions on the ID02 beamline at the European Synchrotron Radiation Facility (ESRF) at Grenoble, France, operating in four-bunch mode with a ring energy of 6.0 GeV (Narayanan *et al*., 2001). These two sessions provided the final data presented here, and followed trial experiments performed in three earlier sessions with 0.4-1.95 mg/ml FH concentrations and fewer heparin fragments. Storage ring currents ranged from 40 mA to 43 mA (heparin) and from 69 mA to 76 mA (HS). Data were acquired using an improved fibre optically-coupled high-sensitivity and dynamic-range CCD detector (FReLoN) with a smaller beamstop. The sample-to-detector distance was 2 m. Data were collected for FH in mixtures with a 1:1 molar ratio of heparin or HS in 10 mM HEPES buffer and 137 mM NaCl (pH 7.4). The FH concentration was 0.47 mg/ml (heparin mixtures) and 0.45 mg/ml (HS mixtures). Samples were measured in a flow cell which moved the sample continuously during beam exposure in 10 time frames with different exposure times between 0.1 sec to 0.5 sec to avoid radiation damage effects. Exposure times were optimised using on-line checks during acquisitions to verify the absence of radiation damage, after which the frames were averaged.

Guinier analyses give the radius of gyration, R_G, which characterizes the degree of structural elongation in solution if the internal inhomogeneity of scattering within the macromolecules has no effect. Guinier plot at low Q values (where $Q = 4$

$\pi \sin \theta/\lambda$; 2θ is the scattering angle; λ is the wavelength) gives the R_G and the forward scattering at zero angle $I(0)$: (Glatter & Kratky, 1982)

$$\ln I(Q) = \ln I(0) - R_G^2 Q^2/3$$

This expression is valid in a $Q.R_G$ range up to 1.5. If the structure is elongated (i.e. rod-shaped), the radius of gyration of the cross-sectional structure R_{XS} and the mean cross-sectional intensity at zero angle $[I(Q).Q]_{Q\to 0}$ parameters are obtained from:

$$\ln [I(Q).Q] = \ln [I(Q).Q]_{Q\to 0} - R_{XS}^2 Q^2/2$$

The R_G and R_{XS} analyses were performed using an interactive PERL script program SCTPL7 (J. T. Eaton and S. J. Perkins, unpublished software) on Silicon Graphics OCTANE workstations. Indirect Fourier transformation of the full scattering curve $I(Q)$ in reciprocal space gives the distance distribution function $P(r)$ in real space. This yields the maximum dimension of the macromolecule L and its most commonly occurring distance vector M in real space:

$$P(r) = \frac{1}{2\pi^2} \int_o^\infty I(Q)\, Qr \sin(Qr)\, dQ$$

The transformation was carried out using GNOM software (Semenyuk & Svergun, 1982). For this, the X-ray curves contained 192-215 intensity data points in the Q range between 0.08 nm^{-1} extending up to between 1.34 nm^{-1} and 2.15 nm^{-1} for the FH mixtures with heparin dp6-dp36. Similar X-ray curves were used for the FH mixtures with HS dp6-dp24.

(6.4.3) Sedimentation velocity data for FH mixtures with heparin and HS

Sedimentation velocity data were obtained using two Beckman XL-I analytical ultracentrifuges (Beckman-Coulter Inc, Palo Alto, CA, U.S.A.) equipped with both absorbance and interference optics (Cole *et al*., 2008). Experiments were carried out in an eight-hole AnTi50 rotor with standard double-sector cells with column heights of 12 mm at 20ºC using interference optics. Sedimentation velocity

data were collected with rotor speeds of 40,000 rpm, 50,000 rpm and 60,000 rpm using interference optics. Experiments with FH and six 1:1 mixtures with heparin dp6-dp36 were performed at 1.08 mg/ml FH in 10 mM HEPES, 137 mM NaCl, pH 7.4. Experiments with FH mixtures were performed with heparin dp6 and dp12 in molar ratios that ranged from 1:0.5 to 1:8 at 1.0-1.12 mg/ml FH. The buffer density of 1.00480 g/ml was measured at 20 ^{o}C using an Anton-Paar DMA5000 density meter. Partial specific volumes were calculated for FH and its 1:1 heparin mixtures to be 0.717 ml/g for FH, 0.714 ml/g for FH and dp6, 0.711 ml/g for FH and dp12, 0.708 ml/g for FH and dp18, 0.705 ml/g for FH and dp24, 0.702 ml/g for FH and dp30 and 0.699 ml/g for FH and dp36 (Perkins, 1986). The continuous *c(s)* analysis method was used to determine the sedimentation coefficients $s_{20,w}$ using SEDFIT software (version 9.4) (Dam & Schuck, 2004; Schuck, 2000). In the *c(s)* analysis, the sedimentation boundaries are fitted using the Lamm equation, the algorithm for which assumes that all species have the same frictional ratio f/f_o in each fit. The final SEDFIT analyses used a fixed resolution of 200 and a fixed frictional ratio f/f_o, of 1.78 for FH and all the heparin mixtures. The $c(s)$ fits were optimised by floating the meniscus and cell bottom when required, and the baseline and cell bottom fixed until the overall root mean square deviations and visual appearance of the fits were satisfactory. The percentage of FH oligomers was derived using the $c(s)$ integration function.

The sedimentation velocity data for four 1:1 FH-HS mixtures were performed as above, using 0.45 mg/ml FH. The partial specific volumes for the 1:1 HS mixtures were 0.716 ml/g for FH and dp6, 0.715 ml/g for FH and dp12, 0.714 ml/g for FH and dp18, and 0.705 ml/g for FH and dp24 (Perkins, 1986).

(6.4.4) FH-heparin dp36 oligomer modelling

Models to account for the $s_{20,w}$ of the FH oligomers formed with heparin dp36 were created using the best-fit X-ray solution scattering model for FH (PDB code: 3GAV) and heparin dp36 (PDB code: 3IRL) (Okemefuna *et al*., 2009; Khan *et al*., 2010). To check for consistency, two other best-fit FH structures were used that were based on alternative X-ray scattering modelling (R. Nan, A. I. Okemefuna and S. J. Perkins, unpublished work). Using the Discovery Studio molecular graphics software (version 2.1) (Accelrys, San Diego, CA, U.S.A.), FH-heparin models were created manually by positioning one end of the dp36 structure close to FH SCR-7 or SCR-20

in the first FH structure and the other end of the dp36 structure close to FH SCR-7 or SCR-20 in the next FH structure. This procedure was used to create FH-heparin dp36 structures based on FH dimer, trimer, tetramer and pentamer. The sedimentation coefficients $s^0_{20,w}$ for these oligomeric models were calculated directly from the molecular structures using the HYDROPRO shell modelling program (Garcia de la Torre *et al*, 2000). The default value of 0.31 nm for the atomic element radius for all atoms was used to represent the hydration shell surrounding FH, heparin and HS.

Chapter Seven
Conclusions

(7.1) Prologue

FH is a major regulatory protein of the alternative complement pathway and is able to discriminate host from pathogen cell surfaces by recognising glycosaminoglycans, heparan sulphate (HS) and dermatan sulphate (DS), on the host and thus inhibits complement activation (Clark *et al*., 2006). FH was the first protein to be genetically linked to AMD, which further underlines its importance to the host innate immune response. A single polymorphism in FH at position 402 (Tyr402His) in SCR-7 domain increases the risk of AMD (Hageman *et al*., 2005; Haines *et al*., 2005; Klein *et al*., 2005). FH has two major binding sites for heparin/HS in domains SCR-6/8 and SCR-19/20. The extent and degree of the effects of heparin/HS fragments on the native FH and the potential relevance of these interactions to AMD had not been elucidated prior to this thesis. Prior to performing interaction studies of FH with heparin/HS fragments, it was necessary to determine the molecular structures of a range of heparin and HS fragments in solution. Heparin is readily available and is often used as an analogue for HS. Up to now, molecular structures for heparin had been determined only for small fragments. For free heparin, solution structures by NMR spectroscopy were available for heparin dodecasaccharide (dp12) and synthetic pentasaccharides (dp5). Unlike heparin, no structural data were currently available for free HS fragments except for a crystal structure in which HS dp4 was complexed with heparinase II. This thesis presents the molecular structures of a range of heparin and HS fragments and their effects on native FH in solution. A multidisciplinary approach involving X-ray scattering, AUC and constrained modelling was used to determine the molecular structures of heparin and HS fragments, and their effects on native FH.

(7.2) Solution structure of Heparin

Prior to this thesis, no molecular structures had been reported for larger heparin fragments. In this thesis, the molecular structures of large heparin fragments in solution were determined by the joint use of X-ray solution scattering, AUC and constrained scattering modelling. AUC revealed that purified heparin fragments of sizes dp6-dp36 were monodisperse from the single peaks seen in *c(s)* analyses and analytical size exclusion chromatography. The sedimentation coefficients $s_{20,w}$ increased linearly from 1.09 S to 1.84 S with heparin size. The corresponding scattering studies on the same fragments resulted in R_G values that also increased

with heparin size from 1.33 nm to 3.12 nm and maximum lengths *L* that increased from 3.0 nm to 12.3 nm. The greater sensitivity of the R_G values to the conformation revealed that the heparin solution structures became progressively more bent with increased size. The constrained modelling of 5,000 randomised heparin conformers starting from the NMR structure resulted in best-fit structures for dp18, dp24, dp30 and dp36 that indicated moderate flexibility in mildly bent structures (Chapter 4: section 4.2.3). The comparison of these solution structures with the 19 crystal structures of heparin-protein complexes revealed similar angular orientations between the iduronate and glucosamine rings, both in the solution and the crystal states. It was concluded that heparin has a semi-rigid and extended conformation that is preformed in a conformation that is well defined for binding to proteins such as FH without major conformational changes (chapter 4: section 4.3).

(7.3) Solution structure of HS

Prior to this study, no molecular structures had been reported for free HS fragments. In this study, the molecular structures of purified HS fragments ranged from dp6 to dp24 in solution were determined following the same method used above for the heparin structure determinations. AUC, analytical size exclusion chromatography and PAGE revealed that the HS fragments were monodisperse from the single peaks seen in *c(s)* analyses, the observation of single sharp peaks in analytical size exclusion chromatography, and single bands by PAGE analysis. The $s_{20,w}$ values increased from 0.82 S to 1.26 S and 1.05 S to 1.35 S for dp6 to dp24 using absorbance and interference optics respectively (Chapter 5; Table 5.1). The X-ray scattering studies on the same fragments showed that the R_G values increased from 1.03 nm to 2.82 nm, the cross-sectional radii of gyration R_{XS} ranged from 0.30 nm to 0.65 nm and the maximum lengths *L* ranged from 3.0 nm to 10.0 nm (Chapter 5; Table 5.1). Deviations of these R_G, R_{XS} and *L* values from linearity revealed that the HS solution structures became more bent with increased size. This bending was confirmed by constrained molecular modelling of 5,000-8,000 randomised HS conformers that resulted in 8 best fit structures for each of dp6 to dp16. Like heparin, this bending is revealed by the appearance of kinks in the HS structure at a few positions. Comparisons of these solution structures with the solution structures of heparin fragments revealed that HS structures exhibit a more bent conformation than heparin. The physical basis of greater bending in HS is likely to arise from the GlcA-

GlcNAc sequence (as opposed to the IdoA-GlcNS sequence in heparin), together with the absence or reduction of repulsion between sulphate-sulphate, sulphate-carboxylate, and carboxylate-carboxylate groups in HS. Thus it was concluded that HS exhibits more flexibility around the glycosidic linkages than the semi-rigid structures of heparin.

(7.4) Effects of heparin and HS fragments on the native FH structure

Prior to this study, the effects of heparin and HS fragments on the native FH were not known. This thesis provides a novel comparative insight into the interactions of these two anionic polysaccharides with their protein ligands. Oligomer formation and the conformational properties of FH in the presence of a range of heparin and HS fragments were studied by X-ray solution scattering and AUC. The AUC data showed that the small heparin fragments (dp6, dp12) slightly decreased the tendency of FH to form a series of oligomers, while the large heparin fragments from dp18 to dp36 showed a dramatic increase in the oligomerisation of FH up to between 42% to 63%. These results revealed that small heparin fragments dp6-dp12 were not long enough to cross-link the FH molecules, while the large fragments dp18-dp36 were long enough to cross-link the FH to form higher oligomers. X-ray scattering data for FH in the presence of large heparin fragments (dp18-dp36) also revealed a dramatic increase in the R_G values that ranged from 11.4 to 11.73 nm, R_{XS} values from 3.31 to 5.11 nm, $I(0)/c$ values from 0.044 to 0.092 and maximum length L from 35 to 40 nm, indicating higher oligomers formation as well as conformational changes in FH. Binding studies of the dp6 molecules to FH in various ratios demonstrated at least two heparin binding sites on the FH molecule.

Like heparin, the effects of HS fragments (dp6-dp24) on the native FH were also studied. Both X-ray and AUC data for FH titrated with small HS fragments revealed slight decreases in the oligomerization of FH. In the presence of larger HS fragments (dp14-dp24), FH showed increases in the R_G values from 8.05 to 8.27 nm and $I(0)/c$ values from 0.042 to 0.073. Only slight changes in the R_{XS} values were observed. An increase in the maximum length L of FH was seen only in the presence of dp18 and dp24. Like heparin, AUC data for FH in the presence of large HS fragments revealed increase in the oligomerisation of FH up to between 16% to 24%. As for heparin, these results also suggested that the smaller HS fragments slightly inhibit the tendency of FH to form oligomers, while the larger HS fragments increase

oligomerisation in FH. The overall effects of HS fragments on the structure of FH is less than observed with heparin fragments, and this is attributed to the reduced degree of sulphation of HS fragments.

The results from this thesis provided four novel insights into the FH interaction with heparin and HS in near-physiological salt conditions. First, both polyanionic oligosaccharides promote strong FH oligomer formation. At least two independent heparin binding sites on FH were involved in these interactions. Little structural changes in FH were observed after heparin or HS binding. Finally the higher sulphation of heparin leads to strengthened FH binding when compared to HS. Taken together, the identification of at least two independent heparin/HS sites on FH through the observation of FH-heparin oligomers has functional implications. The principal one is that the weakening of the FH interaction at either of the two heparin/HS sites on host cell surfaces through the genetic alteration of FH at either site provides an avenue for higher levels of inflammatory attack, given that FH will not bind so well to host cells and will be less able to protect them. This thesis opens the way for more detailed studies of the FH-heparin and FH-HS interactions to enable these molecular mechanisms to be deciphered in more detail.

Future Work: To improve further the understanding of the solution structures of the HS fragments and their effects on intact FH, the HS parent material will be digested with heparinase III. Such heparinase III digests will produce sulphated HS fragments in distinction to heparinase I digests which produce HS fragments with a few sulphate groups at its reducing and non-reducing ends. For structural studies of these different sulphated HS fragments, the same techniques used for heparin and less-sulphated HS fragments will be used. Following solution structural determinations of these fragments, the effects of these fragments on intact FH will be studied for comparison with our previous results. By the same token, the solution structures of two other classes of glycosaminoglycans, namely dermatan sulphate and chondroitin sulphate, will be studied using the same techniques used in this thesis. These common glycosaminoglycans consist of *N*-acetylgalactosamine (GalNAc) residues that alternate at their glycosidic linkages with iduronic acid (IdoA) residues or glucuronic acid (GlcA) (1-3). All these studies will however need to acknowledge some uncertainty in the oligosaccharide composition and the intrinsic inability of the modelling process to account for flexibility leading to conformational diversity.

References

Abdelsalam, A., Del Priore, L. & Zarbin, M. A. (1999). Drusen in age-related macular degeneration: pathogenesis, natural course, and laser photocoagulation-induced regression. *Surv. Ophthalmol*. **44**, 1–28.

Alsenz, J., Lambris, J. D., Schulz, T. F. & Dierich, M. P. (1984). Localization of the complement-component-C3b-binding site and the cofactor activity for factor I in the 38 kDa tryptic fragment of factor H. *Biochem. J*. **224**, 389–398.

Alsenz, J., Schulz, T. F., Lambris, J. D., Sim, R. B. & Dierich, M. P. (1985). Structural and functional analysis of the complement component factor H with the use of different enzymes and monoclonal antibodies to factor H. *Biochem. J.* **232**, 841–850.

Anderson, D. H., Mullins, R. F., Hageman, G. S. & Johnson, L. V. (2002). A role for local inflammation in the formation of drusen in the aging eye. *Am. J. Ophthalmol*. **134**, 411–431.

Arlaud, G., Colomb, M. & Gagnon J. (1987). A functional model of human C1 complex: Emergence of a functional model. *Immunol. Today*, **8**, 106–111.

Ashton, A. W., Boehm, M. K., Gallimore, J. R., Pepys, M. B. & Perkins, S. J. (1997). Pentameric and decameric structures in solution of the serum amyloid P component by X-ray and neutron scattering and molecular modelling analyses. *J. Mol. Biol*. **272**, 408–422.

Aslam, M. & Perkins, S. J. (2001). Folded-back solution structure of monomeric factor H of human complement by synchrotron X-ray and neutron scattering, analytical ultracentrifugation and constrained molecular modelling. *J. Mol. Biol*. **309**, 1117-1138.

Atkins, E. D. & Nieduszynski, I. A. (1977). Conformation of the mucopolysaccharides: X-ray fibre diffraction of heparin. *Fed. Proc.* **1**, 78–83.

Atkinson, J. P. & Frank, M. M. (1980). *Complement*. In: Parker, C.W. (Ed.), Clinical Immunology. WB Saunders, Philadelphia, pp. 219–271.

Bai, X. & Esko, J. D. (1996). Turnover of heparan sulfate depends on 2-O-sulfation of uronic acids. *J. Biol. Chem*. **271**, 17711–17717.

Barlow, P. N., Norman, D. G., Steinkasserer, A., Horne, T. J., Pearce, J., Driscoll, P. C., Sim, R. B. & Campbell, I. D. (1992). Solution structure of the fifth repeat

of factor H: a second example of the complement control protein module. *Biochemistry*, **31**, 3626–3634.

Barlow, P. N., Steinkasserer, A., Norman, D. G., Kieffer, B., Wiles, A. P., Sim, R. B. & Campbell, I. D. (1993). Solution structure of a pair of complement modules by nuclear magnetic resonance. *J. Mol. Biol*. **232**, 268–284.

Beavil, A. J, Young, R. J., Sutton, B. J. & Perkins, S. J. (1995). Bent domain structure of recombinant human IgE-Fc in solution by X-ray and neutron scattering in conjunction with an automated curve fitting procedure. *Biochemistry,* **34**, 14449–14461.

Bernfield, M., Kokenyesi, R., Kato, M., Hinkes, M. T, Spring, J, Gallo, R. L. & Lose, E. J. (1992). Biology of the syndecans: a family of transmembrane heparan sulfate proteoglycans. *Annu. Rev. Cell Biol*. **8**, 365–393.

Bird, A. C. (2010). Therapeutic targets in age-related macular disease. *J. Clin. Invest.* **120**, 3033–3041.

Bird, A. C., Bressler, N. M., Bressler, S. B., Chisholm, I. H., Coscas, G., Davis, M. D. et al. (1995). An international classification and grading system for age related maculopathy and age-related macular degeneration. The International ARM Epidemiological Study Group. *Surv. Ophthalmol*. **39**, 367–374.

Blackmore, T. K., Sadlon, T. A., Wardm, H. M., Lublin, D. M. & Gordon, D. L. (1996). Identification of a heparin binding domain in the seventh short consensus repeat of complement factor H. *J. Immunol*. **157**, 5422–5427.

Blackmore, T. K, Hellwage, J., Sadlon, T. A., Higgs, N., Zipfel, P. F., Ward, H. M. & Gordon, D. L. (1998). Identification of the second heparin-binding domain in human complement factor H. *J. Immunol*. **160**, 3342–3348.

Bonner, A., Almogren, A., Furtado, P. B., Kerr, M. A. & Perkins, S. J. (2009). Location of secretory component on the Fc edge of dimeric IgA1 reveals insight into the role of secretory IgA1 in mucosal immunity. *Mucosal Immunology (Nat. Publ. Group)*, **2**, 74–84.

Campbell, A. K., Daw, R. A., Hallett, M. B. & Luzio, J. P. (1981). Direct measurement of the increase in intracellular free calcium ion concentration in response to the action of complement. *Biochem. J*. **194**, 551–560.

Cantor, C. R. & Schimmel, P. R. (1980). *Biophysical Chemistry,* part II: Techniques for the Study of Biological Structure and Function. pp 1369, San Francisco, W. H. Freeman.

Capila, I. & Linhardt, R. J. (2002). Heparin-protein interactions. *Angew. Chem. Int. Ed. Engl.* **41**, 391–412.

Capila, I., Hernáiz, M. J., Mo, Y. D., Mealy, T. R., Campos, B., Dedman, J. R., Linhardt, R. J. & Seaton, B. A. (2001). Annexin V-heparin oligosaccharide complex suggests heparan sulfate-mediated assembly on cell surfaces. *Structure*, **9**, 57–64.

Caprioli, J., Bettinaglio, P., Zipfel, P. F., Amadei, B., Daina, E., Gamba, S., Skerka, C., Marziliano, N., Remuzzi, G. & Noris, M. (2001). The molecular basis of familial hemolytic uremic syndrome: mutation analysis of factor H gene reveals a hot spot in short consensus repeat 20. *J. Am. Soc. Nephrol*. **12**, 297–307.

Carreno, M. P., Labarre, D., Maillet, F., Jozefowicz, M. & Kazatchkine, M. D. (1989). Regulation of the human alternative complement pathway: formation of a ternary complex between factor H, surface-bound C3b and chemical groups on nonactivating surfaces. *Eur. J. Immunol.* **19**, 2145–2150.

Carter, W. J., Cama, E. & Huntington, J. A. (2005). Crystal structure of thrombin bound to heparin. *J. Biol. Chem*. **280**, 2745–2749.

Chen, C. B. & Wallis, R. (2004). Two mechanisms for mannose-binding protein modulation of the activity of its associated serine proteases. *J. Biol. Chem*. **279**, 26058–26065.

Chen, Y., Maguire, T., Hileman, R. E., Fromm, J. R, Esko, J. D., Linhardt, R. J. & Marks, R. M. (1997). Dengue virus infectivity depends on envelope protein binding to target cell heparan sulphate. *Nat. Med.* **3**, 866–871.

Clark, S. J., Higman, V. A., Mulloy, B., Perkins, S. J., Lea, S. M., Sim, R. B. & Day, A. J. (2006) H384 allotypic variant of factor H associated with age-related macular degeneration has different heparin-binding properties from the non-disease associated form. *J. Biol. Chem.* **281**, 24713–24720.

Clark, S. J., Perveen, R., Hakobyan, S., Morgan, B. P., Sim, R. B., Bishop, P. N. & Day, A. J. (2010). Impaired binding of the AMD-associated complement factor H 402H allotype to Bruch's membrane in human retina. *J. Biol. Chem*. **285**, 30192–30202.

Conrad, H. E. (1998). Heparin-binding proteins. San Diego: Academic Press.

Cole, J. L., Lary, J. W., P Moody, T. & Laue, T. M. (2008). Analytical ultracentrifugation: sedimentation velocity and sedimentation equilibrium. *Meth. Cell Biol*. **84**, 143–179.

Cole, J. L. & Hansen, J. C. (1999). Analytical ultracentrifugation as a contemporary bimolecular research tool. *J . Biomolec. Techniques*, **10**, 163–176.

Comper, W. D. (1981). *Heparin and Related Polysaccharides*, 7, Gordon and Breach.

Coombe, D. R. & Kett, W. C. (2005). Heparan sulfate-protein interactions: therapeutic potential through structure-function insights. *Cell. Mol. Life Sci*. **62**, 410–424.

Cortesio, C. L. & Jiang, W. (2006). Mannan-binding lectin-associated serine protease 3 cleaves synthetic peptides and insulin-like growth factor-binding protein 5. *Arch. Biochem. Biophys*. **449**, 164–170.

Crabb, J. W., Miyagi, M., Gu, S., Shadrach, K., West, K. A., Sakaguchi, H., Kamei, M., Hasan, A., Yan, L. & Rayborn, M. E. (2002). Drusen proteome analysis: an approach to the etiology of age-related macular degeneration. *Proc. Natl. Acad. Sci. U.S.A.* **99,** 14682–14687.

Crum, R., Szabo, S. & Folkman, J. (1985). A new class of steroids inhibits angiogenesis in the presence of heparin or a heparin fragment. *Science*, **230**, 1375–1378.

Dam, J. & Schuck, P. (2005). Sedimentation velocity analysis of heterogeneous protein-protein interactions: sedimentation coefficient distributions *c(s)* and asymptotic boundary profiles from Gilbert-Jenkins theory. *Biophys. J.* **89**, 651–666.

Dam J. & Schuck, P. (2004). Calculating sedimentation coefficient distributions by direct modeling of sedimentation velocity profiles. *Methods Enzymol*. **384**, 185–212.

Davis, A E. 3rd., Mejia, P. & Lu, F. (2008). Biological activities of C1 inhibitor. *Molecular Immunology*, **45**, 4057–4063.

de Jong, P. T. (2006). Age-Related Macular Degeneration, *N. Engl. J. Med*. **355**, 1474–1485.

DiGabriele, A. D., Lax, I., Chen, D. I., Svahn, C. M., Jaye, M., Schlessinger, J. & Hendrickson, W. A. (1998). Structure of a heparin-linked biologically active dimer of fibroblast growth factor. *Nature*, **393**, 812–817.

DiScipio, R. G. (1992). Ultrastructures and interactions of complement factors H and I. *J. Immunol*. **149**, 2592–2599.

DiScipio, R. G., Daffern, P. J., Schraufstatter, I. U. & Sriramarao, P. (1998). Human polymorphonuclear leukocytes adhere to complement factor H through an interaction that involves alphaMbeta2 (CD11b/CD18). *J. Immunol*. **160**, 4057–4066.

Doniach, S. (2001). Changes in biomolecular conformation seen by small angle x-ray scattering. *Chem. Rev*. **101**, 1763–1778.

Dragon-Durey, M. A., Fremeaux-Bacchi, V., Loirat, C., Blouin, J., Niaudet, P., Deschenes, G., Coppo, P., Hermab Fridman, W. & Weiss, L. (2004). Hetrozygous and homozygous factor H deficiencies associated with haemolytic uremic sundrome or membranoproliferative glomerulonephritis: report and genetic analysis of 16 cases. *J. Am. Soc. Nephrol*. **15**, 787–795.

Edwards, A. O., Ritter III, R., Abel, K. J., Manning, A., Panhuysen, C. & Farrer, L. A. (2005). Complement factor H polymorphism and age-related macular degeneration. *Science*. **308**, 421–424.

Erdei, A. & Sim, R. B. (1987). Complement factor H-binding protein of Raji cells and tonsil B lymphocytes. *Biochem. J*. **246**, 149–156.

Ermer, O. (1976). *Struct. Bond.* **27**, p. 161-208.

Esko, J. D. & Lindahl, U. (2001). Molecular diversity of heparan sulphate. *J. Clin. Invest.* **108**, 169–173.

Faham, S., Hileman, R. E., Fromm J. R., Linhardt, R. J. & Rees, D. C. (1996). Heparin structure and interactions with basic fibroblast growth factor. *Science*, **271,** 1116–1120.

Farries, T. C., Seya, T., Harrison, R. A. & Atkinson, J. P. (1990). Competition for binding sites on C3b by CR1, CR2, MCP, factor B and Factor H. *Complement Inflamm*. **7**, 30–41.

Fearon, D. T. & Austen, K. F. (1977). Activation of alternative pathway with rabbit erythrocytes by circumvention of the regulatory action of endogenous control proteins. *J. Exp. Med*. **146**, 22–23.

Fearon, D. T. (1978). Regulation by membrane sialic acid of beta1H-dependent decay-dissociation of amplification C3 convertase of the alternative complement pathway. *Proc. Natl. Acad. Sci. U.S.A.* **75**, 1971–1975.

Fenaille, F., Le Mignon, M., Groseil, C., Ramon, C., Riandé, S., Siret, L. & Bihoreau, N. (2007). Site-specific *N*-glycan characterization of human complement factor H. *Glycobiology*, **17**, 932–944.

Fernando, A. N., Furtado, P. B., Clark, S. J., Gilbert, H. E., Day, A. J., Sim, R. B. & Perkins, S. J. (2007). Associative and structural properties of the region of complement factor H encompassing the Tyr402His disease-related polymorphism and its interactions with heparin. *J. Mol. Biol*. **368**, 564–581.

Ferentz, A. E. & Wagner, G. (2000). NMR spectroscopy: a multifaceted approach to macromolecular structure. *Q. Rev. Biophys*. **33**, 29–65.

Folkman, J., Langer, R., Linhardt, R. J., Handeschild, C. & Taylor, S. (1983). Angiogenesis inhibition and tumor regression caused by heparin or a heparin fragment in the presence of cortisone. *Science*, **221**, 719–725.

Fremeaux-Bacchi, V., Regnier, C., Blouin, J., Drageon-Durey, M. A., Fridman, W. H., Janssen B. & Loirat, C. (2007). Protective or aggressive: paradoxical role of C3 in atypical hemolytic uremic syndrome. *Mol. Immunol.* **44**, 172.

Friese, M. A., Hellwage, J., Jokiranta, T. S., Meri, S., Peter, H. H., Eibel H. & Zipfel, P. F. (1999). FHL-1 and factor H: two human complement regulators which are encoded by the same gene are differently expressed and regulated. *Mol. Immunol*. **36**, 809–818.

Fry, E. E., Lea, S. M., Jackson, T., Newman, J. W., Ellard, F. M., Blakemore, W. E., Abu-Ghazaleh, R., Samuel, A., King, A. M. & Stuart, D. I. (1999). The structure and function of a foot-and-mouth disease virus-oligosaccharide receptor complex. *EMBO J.* **18**, 543–554.

Fujita, T., Matsushita, M. & Endo, Y. (2004). The lectin-complement pathway – its role in innate immunity and evolution. *Immunological Reviews,* **198**,185–202

Fujita, H. (1975). *Foundations of ultracentrifugal analysis.* New York: John Wiley and Sons.

Fujiwara, S., Wiedemann, H., Timpl, R., Lustig, A. & Engel, J. (1984). Structure and interactions of heparan sulphate proteoglycans from a mouse tumor basement membrane. *Eur. J. Biochem*. **143**, 145–157.

Furst, A. (1997). The XL-I analytical ultracentrifuge with Rayleigh interference optics. *Eur. J. Biochem*. **35**, 307–310.

Furtado, P. B., Huang, C. Y., Ihyembe, D., Hammond, R. A., Marsh, H. C. & Perkins, S. J. (2008). The partly folded back solution structure arrangement of

the 30 SCR domains in human complement receptor type 1 (CR1) permits access to its C3b and C4b ligands. *J. Mol. Biol*. **375**, 102–118.

Gallagher, J. T. (2001). Heparan sulfate: growth control with a restricted sequence menu. *J. Clin. Invest*. **108**, 357–361.

Gallagher, J. T., Turnbull, J. E. & Lyon, M. (1992). Patterns of sulphation in heparan sulphate: polymorphism based on a common structural theme. *Adv. Exp. Med. Biol*. **313**, 49–57.

Gallagher, J. T. & Walker, A. (1985). Molecular distinctions between heparan sulphate and heparin. Analysis of sulphation patterns indicates that heparan sulphate and heparin are separate families of N-sulphated polysaccharides. *Biochem. J*. **230**, 665–674.

Ganesh, V. K., Smith, S. A., Kotwal, G. J. & Murthy, K. H. (2004). Structure of vaccinia complement protein in complex with heparin and potential implications for complement regulation. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 8924–8929.

Garcia de la Torre, J., Huertas, M. L., & Carrasco, B. (2000). Calculation of hydrodynamic properties of globular proteins from their atomic-level structure. *Biophys. J*. **78**, 719–730.

Garcia de la Torre, J., Navarro, S., Martinez, M. C. L., Diaz, F. G., & Cascales, J. L. (1994). HYDRO: A computer program for the prediction of hydrodynamic properties of macromolecules. *Biophys. J*. **67**, 530–531.

Giannakis, E., Jokiranta, T. S., Male, D. A., Ranganathan, S., Ormsby, R. J., Fischetti, V.A., Mold, C. & Gordon, D. L. (2003). A common site within factor H SCR 7 responsible for binding heparin, C-reactive protein and streptococcal M protein. *Eur. J. Immunol*. **33**, 962–969.

Giebeler, R. (1992). The Optima XL-A: a new analytical ultracentrifuge with a novel precision absorption optical system. In Harding, S. E., Rowe, A. J., Horton, J. C. (ed): *Analytical Ultracentrifugation in Biochemistry and Polymer Science.* Cambridge, UK Royal Society of Chemistry, 16–25.

Gilbert, H. E., Asokan, R., Holers, V. M. & Perkins, S. J. (2006). The flexible 15 SCR extracellular domains of human complement receptor type 2 can mediate multiple ligand and antigen interactions. *J. Mol. Biol*. **362**, 1132–1147.

Gilbert, H. E., Eaton, J. T., Hannan, J. P., Holers, V. M., & Perkins, S. J. (2005). Solution structure of the complex between CR2 SCR 1-2 and C3d of human

complement: an X-ray scattering and sedimentation modeling study. *J. Mol. Biol.* **346**, 859–873.

Glatter, O. & Kratky, O. (1982). *Editors of Small-Angle X-ray Scattering.* Academic Press, New York.

Goodship, T. H. J. & Kavanagh, D. (2010). Pulling the trigger in atypical hemolytic uremic syndrome: The role of pregnancy. *J. Am. Soc. Nephrol.* **21**, 731–732.

Goicoechea de Jorge, E., C. L. Harris, J. Esparza-Gordillo, L. Carreras, E. A. Arranz, C.A. Garrido, M. Lopez-Trascasa, P. Sanchez-Corral, B. P. Morgan, & S. Rodriguez de Cordoba. (2007). *Proc. Natl. Acad. Sci. U.S.A.* **104**, 240–245.

Gonen, T., Cheng, Y., Sliz, P., Hiroaki, Y., Fujiyoshi, Y., Harrison, S. C. & Walz, T. (2005). Lipid-protein interactions in double-layered two-dimensional AQP0 crystals. *Nature*, **438**, 633–638.

Gonen, T., Sliz, P., Kistler, J., Cheng, Y. & Walz, T. (2004). Aquaporin-0 membrane junctions reveal the structure of a closed water pore. *Nature*, **429**, 193-197.

Gordon, D. L., Kaufman, R. M., Blackmore, T. K., Kwong, J. & Lublin, D. M. (1995). Identification of complement regulatory domains in human factor H. *J. Immunol*. **155**, 348–356.

Guinier, A. (1939). La diffraction des rayons X aux tres petits angles; application a l'etude de phenomenes ultramicroscopiques.. *Ann. Phys. (Paris)*, **12**,161–238.

Hageman, G. S., Anderson, D. H., Johnson, L. V., Hancox, L. S., Taiber, A. J., Hardisty, L. I., Hageman, J. L., Stockman, H. A., Borchardt, J. D., Gehrs, K. M., Smith, R. J., Silvestri, G., Russell, S. R., Klaver, C. C., Barbazetto, I., Chang, S., Yannuzzi, L. A., Barile, G. R., Merriam, J. C., Smith, R. T., Olsh, A. K., Bergeron, J., Zernant, J., Merriam, J. E., Gold, B., Dean, M. & Allikmets, R. (2005). A common haplotype in the complement regulatory gene factor H (HF1/CFH) predisposes individuals to age-related macular degeneration. *Proc. Natl Acad. Sci. U.S.A.* **102**, 7227–7232.

Hageman, G. S., Luthert, P. J., Chong, N. H. V., Johnson, L. V., Anderson, D. H. & Mullins, R. F. (2001). An integrated hypothesis that considers drusen as biomarkers of immune-mediated processes at the RPE-Bruch's membrane interface in aging and age-related macular degeneration. *Prog. Retinal Eye Res.* **20**, 705–732.

Hageman, G. S, Mullins, R. F., Russell, S. R., Johnson, L. V. & Anderson, D. H. (1999). Vitronectin is a constituent of ocular drusen and the vitronectin gene is expressed in human retinal pigmented epithelial cells. *FASEB J*. **13**, 477–484.

Haines, J. L., Hauser, M. A., Schmidt, S., Scott, W. K., Olson, L. M., Gallins, P., Spencer, K. L., Kwan, S. Y., Noureddine, M., Gilbert, J. R., Schnetz-Boutaud, N., Agarwal, A., Postel, E. A. & Pericak-Vance, M. A. (2005). Complement factor H variant increases the risk of age-related macular degeneration. *Science*, **308**, 419–421.

Hajela, K., Kojima, M., Ambrus, G., Wong, K. H., Moffatt, B. E., Ferluga, J., Hajela, S., Gál, P. & Sim, R. B. (2002). The biological functions of MBL-associated serine proteases (MASPs). *Immunobiology*, **205**, 467–475.

Heinen, S., Hartmann, A., Lauer, N., Wiehl, U., Dahse, H. M., Schirmer, S., Gropp, K., Enghardt, T., Wallich, R., Hälbich, S., Mihlan, M., Schlötzer-Schrehardt, U., Zipfel, P. F. & Skerka, C. (2009). Factor H related protein 1 (CFHR-1) inhibits complement C5 convertase activity and terminal complex formation. *Blood,* **114**, 2439–2447.

Hellwage, J., Jokiranta, T. S., Friese, M. A., Wolk, T. U., Kampen, E., Zipfel, P. F. & Meri, S. (2002). Complement C3b/C3d and cell surface polyanions are recognized by overlapping binding sites on the most carboxyl-terminal domain of complement factor H. *J. Immunol.* **169**, 6935–6944.

Hellwage, J., Kühn, S. & Zipfel, P. F. (1997). The human complement regulatory factor-H-like protein 1, which represents a truncated form of factor H, displays cell-attachment activity. *Biochem. J*. **326**, 321-327.

Harding, S. E. & Winzor, D. J. (2001). Sedimentation velocity analytical ultracentrifugation. *In "Protein-ligand Interaction: hydrodynamics and calorimetry" (eds.)* Harding, S. E. and Chowdhry B. Z. Oxford University Press, New York. Chapter **4**, 75–103.

Herr, A. B., Ornitz, D. M., Sasisekharan, R., Venkataraman, G. & Waksman, G. (1997). Heparin-induced self-association of fibroblast growth factor-2: evidence for two oligomerization processes. *J. Biol. Chem.* **272**, 16382–16389.

Hileman, R. E., Fromm, J. R., Weiler, J. M. & Linhardt, R. J. (1998). Glycosaminoglycan–protein interactions: definition of consensus sites in glycosaminoglycan binding proteins. *BioEssay*s, **20**, 156–167.

Hjelm, R. P. (1985). The small angle approximation of X-ray and neutron scatter from rigid rods of non-uniform cross section and finite length. *J. Appl. Cryst.* **18**, 452–460.

Hocking, H. G., Herbert, A P., Kavanagh, D., Soares, D. C., Ferreira, V. P., Pangburn, M. K., Uhrín, D., & Barlow P. N. (2008). Structure of the N-terminal region of complement factor H and conformational implications of disease-linked sequence variations. *J. Biol. Chem*. **283**, 9475–9487.

Holondniy, M., Kim, S., Katzenstein, D., Konrad, M., Groves, E. T. & Merigan, T. C. (1991). Inhibition of human immunodeficiency virus gene amplification by heparin. *J. Clin. Microbiol*. **29**, 676–679.

Horstmann, R. D., Sievertsen, H. J., Knobloch, J. & Fischetti, V. A. (1988). Antiphagocytic activity of streptococcal M protein: selective binding of complement control protein factor H. *Proc. Natl. Acad. Sci. U. S. A.* **85**, 1657–1661.

Hricovini, M., Guerrini, M., Bisio, A., Torri, G., Naggi, A. & Casu, B. (2002). Active conformations of glycosaminoglycans. NMR determination of the conformation of heparin sequences complexed with antithrombin and fibroblast growth factors in solution. *Semin. Thromb. Hemost*. **28,** 325–334.

Hricovini, M., Guerrini, M., Torri, G., & Casu, B. (1997). Motional properties of E. coli polysaccharide K5 in aqueous solution analyzed by NMR relaxation measurements. *Carbohydr. Res*. **300**, 69–76.

Jacobsson, I. & Lindahl, U. (1980). Biosynthesis of heparin. Concerted action of late polymer-modification reactions. *J. Biol. Chem*. **255**, 5094–5100.

Janeway, C. A., Travers, P., Walport, M. & Shlomchik, M. J. (2005). Immunobiology, the immune system in health and disease. *Garland Science Publishing*. (6th Edition).

Jencks WP. (1981). On the attribution and additivity of binding energies. *Proc. Natl. Acad .Sci. U.S.A.* **78**, 4046–4050.

Jin, L., Abrahams, J. P., Skinner, R., Petitou, M., Pike, R. N. & Carrell, R. W. (1997). The anticoagulant activation of antithrombin by heparin. *Proc. Natl. Acad .Sci. U.S.A.* **94**, 14683–14688.

Johnson, D. J., Li, W., Adams, T. E. & Huntington, J. A. (2006). Antithrombin-S195A factor Xa-heparin structure reveals the allosteric mechanism of antithrombin activation. *EMBO J.* **25**, 2029–2037.

Johnson, M. L. & Straume, M. (1994). Comments on the analysis of sedimentation equilibrium experiments In Schuster, T. M, Laue, T. M. (ed): *Modern Analytical Ultracentrifugation.* Boston: Birkhauser, 37–65.

Johnson, M. L. & Faunt, L. M. (1992). Parameter estimation by leastsquares methods. *Meth. Enzymol*. **210**, 1–37.

Joiner, K. A. (1988). Complement evasion by bacteria and parasites. *Annu. Rev. Microbiol.* **42**, 201–230.

Jokiranta, T. S., Jaakola, V. P., Lehtinen, M. J., Parepalo, M., Meri, S. & Goldman, A. (2006). Structure of complement factor H carboxyl-terminus reveals molecular basis of atypical haemolytic uremic syndrome. *EMBO J*. **25**, 1784–1794.

Jokiranta, T. S., Hellwage, J., Koistinen, V., Zipfel, P. F. & Meri, S. (1998). Each of the three binding sites of factor H interacts with a distinct site on C3b. *Mol. Immunol*. **35**, 360.

Jokiranta, T. S., Zipfel, P. F., Hakulinen, J., Kuhn, S., Pangburn, M. K., Tamerius, J. D. & Meri, S. (1996). Analysis of the recognition mechanism of the alternative pathway of complement by monoclonal anti-factor H antibodies: evidence for multiple interactions between H and surface bound C3b. *FEBS. Lett*. **393**, 297–302.

Jozsi, M. & Zipfel, P. F. (2008). Factor H family proteins and human diseases. *Trends Immunol*. **29**, 380–387

Kaplan, M. H. & Volanakis, J. E. (1974). Interaction of C-reactive protein complexes with the complement system. I. Consumption of human complement associated with the reaction of C-reactive protein with pneumococcal C-polysaccharide and with the choline phosphatides, lecithin and sphingomyelin. *J. Immunol*. **112**, 2135–2147.

Kavanagh, D., Kemp, E. J., Mayland, E., Winney, R. J., Duffield, J. S., Warwick, G., Richards, A., Ward, R., Goodship, J. A. & Goodship, T. H. (2005). Mutations in complement factor I predispose to the development of atypical hemolytic uremic syndrome. *J. Am. Soc. Nephrol.* **16,** 2150–2155.

Kay, L. E. (1998). Protein dynamics from NMR. *Biochem. Cell Biol*. **76**, 145–152.

Kazatchkine, M. D., Fearon, D. T. & Austen, K. F. (1979). Surface-associated heparin inhibits zymosan-induced activation of the human alternative

complement pathway by augmenting the regulatory action of the control proteins on particle-bound C3b. *J. Immunol.* **122**, 75–81.

Kazatchkine, M. D., Fearon, D. T., Metcalfe, D. D., Rosenberg, D. D. & Austen, F. K. (1981). Structural determinants of the capacity of heparin to inhibit the formation of the human amplification C3 convertase *J. Clin. Invest.* **67**, 223–228.

Kerr, M. (1980). The human complement system: assembly of the classical pathway C3 convertase. *Biochem. J*. **189**, 173–181.

Khan, S., Gor, J., Mulloy, B. & Perkins, S. J. (2010). Semi-rigid solution structures of heparin by constrained X-ray scattering modelling: New insight into heparin-protein complexes. *J. Mol. Biol*. **395**, 22, 504–521.

Khan, S., Gor, J., Mulloy, B. & Perkins, S. J. (2008). Solution structure of heparin and its complexes with the factor H SCR-6/8 domains and factor H: implications for disease. *Mol. Immunol*. **45**, 4126; (Abstract).

Khera, R. & Das, N. (2009). Complement receptor 1: disease associations and therapeutic implications. *Mol. Immunol*. **46**, 761–772.

Kimberley, F. C., Sivasankar, B. & Morgan, B. P. (2007). Alternative roles for CD59. *Mol. Immunol*. **44**, 73–81.

Klein, R. J., Zeiss, C., Chew, E. Y., Tsai, J. Y., Sackler, R. S., Haynes, C., Henning, A. K., Sangiovanni, J. P., Mane, S. M., Mayne, S. T., Bracken, M. B., Ferris, F. L., Ott, J., Barnstable, C. & Hoh, J. (2005). Complement factor H polymorphism in age-related macular degeneration. *Science*, **308**, 385–389.

Klein, R., Peto, T., Bird, A. & Vannewkirk, M. R. (2004). The epidemiology of age-related macular degeneration. *Am. J. Ophthalmol*. **137**, 486–495.

Kliffen, M., Mooy, C. M., Luider, T. M., Huijmans, J. G. M., Kerkvliet, S. & de Jong, P. T. V. M. (1996). Identification of glycosaminoglycans in age-related macular deposits. *Arch. Ophthalmol.* **114**, 1009–1014.

Koch, M. H., Vachette, P. & Svergun, D. I. (2003). Small-angle scattering: a view on the properties, structures and structural changes of biological macromolecules in solution. *Q. Rev. Biophys*. **36**, 147–227.

Kratky, O., Pilz, I. & Schmitz, P. J. (1966). *J. Colloid Interface Sci*. **21**, 24–34.

Kreuger, J., Spillmann, D., Li, J. P., & Lindahl, U. (2006). Interactions between heparan sulfate and proteins: the concept of specificity. *J. Cell Biol*. **174**, 323–327.

Kuhn, S. & Zipfel, P. F. (1996). Mapping of the domains required for decay acceleration activity of the human factor H-like protein 1 and factor H. *Eur. J. Immunol*. **26**, 2383–2387.

Kuhn, S., Skerka, C. & Zipfel, P. F. (1995). Mapping of the complement regulatory domains in the human factor H-like protein 1 and in factor H. *J. Immunol*. **155**, 5663–5670.

Lambris, J. D. & Ross, G. D. (1982). Characterization of the lymphocyte membrane receptor for factor H (beta 1H-globulin) with an antibody to anti-factor H idiotype. *J. Exp. Med*. **155**, 1400–1411.

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London), **227**, 680-685.

Lamm, O. (1929). Die Differentialgleichung der Ultrazentrifugierung. *Ark. Mat. Astr. Fys.* **21B**, 1–4.

Lane, D. A. & Lindahl, U. Eds. (1989). Heparin, Chemical and Biological Properties, Clinical Applications, CRC Press, Boca Raton, FL.

Langdown, J., Belzar, K. J., Savory, W. J., Baglin, T. P. & Huntington, J. A. (2009). The critical role of hinge-region expulsion in the induced-fit heparin binding mechanism of antithrombin. *J. Mol. Biol.* **386**, 1278-1289.

Law, S. K. A. & Reid, K. B. M. (1995). Complement. *Oxford University Press*. (2nd Edition).

Lewis, M. S., Shrager, R. I. & Kim S. J. (1994). Analysis of protein nucleic acid and protein-protein interactions using multi-wavelength scans from the XL-A analytical centrifuge. In Schuster, T. M, Laue, T. M. (eds): *Modern Analytical Ultracentrifugation,* Boston: Birkhauser, 94–115.

Li, K., Gor, J. & Perkins, S. J. (2010). Self-association and domain rearrangements between complement C3 and C3u provide insight into the activation mechanism of C3. *Biochem. J.* **431**, 63–72.

Li, W. & Huntington, J. A. (2008). The heparin binding site of Protein C inhibitor is protease-dependent. *J. Biol. Chem.* **51**, 36039–36045.

Lietha, D., Chirgadze, D. Y., Mulloy, B., Blundell, T. L. & Gherardi, E. (2001). Crystal structures of NK1-heparin complexes reveal the basis for NK1 activity and enable engineering of potent agonists of the MET receptor. *EMBO J.* **20**, 5543–5555.

Lindahl, U., Feingold, D. S. & Roden, L. (1986). Biosynthesis of heparin. *Trends Biochem. Sci*. **11**, 221–225.

Lindahl, U., Bäckström, G., Malmström, A. & Fransson, L. A. (1972). Biosynthesis of L-iduronic acid in heparin: epimerization of D-glucuronic acid on the polymer level. *Biochem. Biophys. Res. Commun*. **46**, 985–991.

Linhardt, R. J. & Toida, T. (1997). Heparin analogs-development and application [review]. In carbohydrate as drug (eds. Z.B. Witczak and K.A Neiforth) Marcel Dekker, New York, chapter 7, pp.277–341.

Liu, D., Shriver, Z., Venkataraman, G., El Shabrawi, Y. & Sasisekharan, R. (2002). Tumor cell surface heparan sulphate as cryptic promoters or inhibitors of tumor growth and metastasis. *Proc. Natl. Acad. Sci. USA*, **99**, 568–573.

Liu, G., Hultin, M., Ostergaard, P. & Olivercrona, T. (1992). Interaction of size-fractionated heparins with lipoprotein lipase and hepatic lipase in the rat. *Biochem. J*. **285**, 731–736.

Lortat-Jacob, H., Turnbull, J. E. & Grimaud, J.-A. (1995). Molecular organization of the interferon gamma-binding domain in heparan sulphate. *Biochem. J.* **310**, 497–505.

Lyon, M. & Gallagher, J. T. (1998). Bio-specific sequences and domains in heparan sulphate and the regulation of cell growth and adhesion. *Matrix Biol.* **17**, 485–493.

Matsushita, M., Endo, Y. & Fujita, T. (2000). Cutting edge: complement-activating complex of ficolin and mannose-binding lectin-associated serine protease. *J. Immunol*. **164**, 2281–2284.

Matsushita, M. & Fujita, T. (1992). Activation of the classical complement pathway by mannose-binding protein in association with a novel C1s-like serine protease. *J. Exp. Med*. **176**, 1497–1502.

Mayo, S. L., Olafson, B. D. & Goddard, W. A. (1990). DREIDING: a generic force field for molecular simulations. *J. Phys. Chem.* **94**, 8897–8909.

McCoy, A. J., Pei, X. Y., Skinner, R., Abrahams, J. P. & Carrell, R. W. (2003). Structure of beta-antithrombin and the effect of glycosylation on antithrombin's heparin affinity and activity. *J. Mol. Biol.* **326**, 823–833.

Meri, S. & Pangburn, M. K. (1990). Discrimination between activators and nonactivators of the alternative pathway of complement: regulation via a sialic

acid/polyanion binding site on factor H. *Proc. Natl. Acad. Sci. U.S.A*, **87**, 3982–3986.

Misasi, R., Huemer, H. P., Schwaeble, W., Solder E., Larcher, C. & Dierich, M. (1989). Human complement Factor H: an additional gene product of 43 kDa isolated from human plasma shows cofactor activity for cleavage of the third component of complement. *Eur. J. Immun.* **19***,* 1765–1768.

Mobli, M., Nilsson, M., & Almond, A. (2008). The structural plasticity of heparan sulfate NA-domains and hence their role in mediating multivalent interactions is confirmed by high-accuracy 15N-NMR relaxation studies. *Glycoconj. J*. **25**, 401–414.

Morgan, B. (1999). Regulation of the complement membrane attack pathway, *Crit. Rev. Immunol*. **19**, 173–198.

Morgan, B. P. & Harris, C. L. (1999). *The complement system.* In: Morgan, B. P., Harris, C. L. (Eds.) Complement Regulatory Proteins. Academic Press, San Diego, pp.1-13.

Moy, F. J., Safran, M., Seddon, A. P., Kitchen, D., Böhlen, P., Aviezer, D. & Yayon, A. (1997). Properly oriented heparin-decasaccharide-induced dimers are the biologically active form of basic fibroblast growth factor. *Biochemistry*, **36,** 4782–4791.

Muller-Eberhard, H. J. (1988). Molecular organization and function of the complement system. *Annu. Rev. Biochem*. **57**, 321–347.

Mullins, R. F., Russel, S. R., Anderson, D. H. & Hageman, G. S. (2000). Drusen associated with aging and age-related macular degeneration contain proteins common to extracellular deposits associated with atherosclerosis, elastosis, amyloidosis, and dense deposit disease. *FASEB J*. **14**, 835–846.

Mulloy, B. & Forster, M. J. (2000). Conformation and dynamics of heparin and heparan sulphate. *Glycobiology*, **10**, 1147–1156.

Mulloy, B., Gee, C., Wheeler, S. F., Wait, R., Gray, E. & Barrowcliffe, T. W. (1997). Molecular weight measurements of low molecular weight heparins by gel permeation chromatography. *Thromb. Haemost*. **77**, 668–674.

Mulloy, B., Forster, M. J., Jones, C., Drake, A. F., Johnson, E. A. & Davies, D. B. (1994). The effect of variation of substitution on the solution conformation of heparin: a spectroscopic and molecular modelling study. *Carbohydr. Res*. **255**, 1–26.

Mulloy, B., Forster, M. J., Jones, C. & Davies, D. B. (1993). N.m.r. and molecular-modelling studies of the solution conformation of heparin. *Biochem. J*. **293**, 849–858.

Nan, R., Farabella, I., Schumacher, F., Miller, A., Gor, J., Martin, A.C.R., Jones, D.T., Lengyel, I. & Perkins, S. J. (2011). Localisation of the major zinc-binding site of complement Factor H to the SCR-6/8 domains: possible implications for age-related macular degeneration. *J. Mol. Biol*. **In press**.

Nan, R. (2010). Self-association of complement Factor H in the presence and absence of metals. PhD Thesis, University College London

Nan, R., Gor, J., Lengyel, I & Perkins, S. J. (2008a). Uncontrolled zinc- and copper-induced oligomerisation of the human complement regulator Factor H and its possible implications for function and disease. *J. Mol. Biol*. **384**, 1341–1352.

Nan, R., Gor, J. & Perkins, S. J. (2008b). Implications of the progressive self-association of wild-type human factor H for complement regulation and disease. *J. Mol. Biol.* **375**, 891–900.

Narayanan, T., Diat, O. & Bosecke, P. (2001). SAXS and USAXS on the high brilliance beamline at the ESRF. *Nucl. Instrum. Methods Phys. Res. A*, **467-468**, 1005–1009.

Nieduszynski, I. A. & Atkins, E. D. (1973). Conformation of the mucopolysaccharides. X-ray fibre diffraction of heparin. *Biochem. J*. **135**,729–733.

Okemefuna, A. I., Nan, R., Miller, A., Gor, J. & Perkins S. J. (2010). Complement factor H binds at two independent sites to C-reactive protein in acute phase concentrations. *J. Biol. Chem*. **285**, 1053–1065.

Okemefuna, A. I., Nan, R., Gor J. & Perkins, S. J. (2009). Electrostatic interactions contribute to the folded-back conformation of wild type human factor H. *J. Mol. Biol*. **391**, 98–118.

Okemefuna, A. I., Gilbert, H. E., Griggs, K. M., Ormsby, R. J., Gordon, D. L. & Perkins, S. J. (2008). The regulatory SCR-1/5 and cell-surface-binding SCR-16/20 fragments of Factor H reveal partially folded-back solution structures and different self-associative properties. *J. Mol. Biol*. **375**, 80–101.

Olivercrona, T. & Bengtsson-Olivercrona, G. (1989). In *Heparin: Chemical and Biological Properties, Clinical Applications*, ed. D. A. Lane & U. Lindahl, CRC Press, Inc., Boca Raton, FL, pp. 335.

Ormsby, R. J., Jokiranta, T. S., Griggs, K., Giannakis, E., Sadlon, T. & Gordon, D. L. (2006). Localisation of the third heparin-binding domain in human complement factor H. *Molec. Immunol.* **43**, 1624–1632.

Ornitz, D. M., Yayon, A., Flanagan, J. G., Svahn, C. M., Levi, E. & Leder, P. (1992). Heparin is required for cell-free binding of basic fibroblast growth factor to a soluble receptor and for mitogenesis in whole cells. *Mol. Cell. Biol.* **12**, 240–247.

Ortega, A. & García de la Torre, J. (2007). Equivalent radii and ratios of radii from solution properties as indicators of macromolecular conformation, shape, and flexibility. *Biomacromolecules*, **8**, 2464–2475.

Pangburn, M. K., Rawal, N., Cortes, C., Alam, M. N., Ferreira, V. P. & Atkinson, M. A. (2009). Polyanion-induced self-association of complement factor H. *J. Immunol*. **182**, 1061–1068.

Pangburn, M. K. (2000). Host recognition and target differentiation by factor H, a regulator of the alternative pathway of complement. *Immunopharmacology*, **49**, 149–157.

Pangburn, M. K., Atkinson, M. A. & Meri, S. (1991). Localization of the heparin-binding site on complement factor H. *J. Biol. Chem*. **266**, 16847–16853.

Pangburn, M. K. & Muller-Eberhard, H. J. (1984). The alternative pathway of complement. *Springer Semin. Immunopathol*, **7**, 163–192.

Pangburn, M. K., Schriber, R. D. & Muller-Eberhard, H. J. (1981). Formation of initial convertase of the alternative complement pathway. Acquisition of C3b like activities by spontaneous hydrolysis of the putative thioester in native C3. *J. Exp. Med*. **154**, 856–867.

Pangburn, M. K. & Müller-Eberhard, H. J. (1978). Complement C3 convertase: cell surface restriction of βIH control and generation of restriction on neuraminidase treated cells. *Proc. Natl. Acad. Sci. U.S.A.* **75**, 2416–2420.

Pangburn, M. K., Schreiber, R. D. & Müller-Eberhard, H. J. (1977). Human complement C3b inactivator: isolation, characterization, and demonstration of an absolute requirement for the serum protein beta1H for cleavage of C3b and C4b in solution. *J. Exp. Med*. **146**, 257–270.

Pavlov, G., Finet, S., Tatarenko, K., Korneeva, E. & Ebel, C. (2003). Conformation of heparin studied with macromolecular hydrodynamic methods and X-ray scattering. *Eur. Biophys. J.* **32**, 437–449.

Pellegrini, L., Burke, D. F., von Delft, F., Mulloy, B. & Blundell, T. L. (2000). Crystal structure of fibroblast growth factor receptor ectodomain bound to ligand and heparin. *Nature*, **407**, 1029–1034.

Perkins, S. J., Nan, R., Okemefuna, A. I., Li, K., Khan, S. & Miller, A. (2010). Multiple interactions of complement factor H with its ligands in solution: a progress report. Current Topics on Complement and Eye Diseases (Ed. J.D. Lambris and A. Adamis). *Adv. Exp. Med. Biol*. **703**, 25–47.

Perkins, S. J., Okemefuna, A. I., Nan, R., Li, K. & Bonner, A. (2009). Constrained solution scattering modelling of human antibodies and complement proteins reveals novel biological insights. *J. R. Soc. Interface*, **6**, S679–S696.

Perkins, S. J., Okemefuna, A. I., Fernando, A. N., Bonner, A., Gilbert, H. E. & Furtado, P. B. (2008). X-ray and neutron scattering data and their constrained molecular modelling. *Methods Cell Biol*. **84**, 375–423.

Perkins, S. J., Gilbert, H. E., Lee, Y. C., Sun, Z. & Furtado, P. B. (2005). Relating small angle scattering and analytical ultracentrifugation in multidomain proteins. *In "Modern Analytical Ultracentrifugation: Techniques and Methods"* (eds.) Scott, D.J., Harding, S. E. & Rowe, A. J. Royal Society of Chemistry, London, U.K. Chapter 15,pp 291–319.

Perkins, S. J. & Goodship, T. H. J. (2002). Molecular modelling of the C-terminal domains of factor H of human complement: a correlation between haemolytic uremic syndrome and a predicted heparin binding site. *J. Mol. Biol*. **316**, 217–224.

Perkins, S. J. (2001). X-ray and neutron scattering analyses of hydration shells: a molecular interpretation based on sequence predictions and modeling fits. *Biophys. Chem.* **93**, 129–139.

Perkins, S. J., Ashton, A. W., Boehm, M. K. & Chamberlain, D. (1998). Molecular structures from low angle X-ray and neutron scattering studies. *Int. J. Biolog. Macromol*. **22**, 1–16.

Perkins, S. J., Nealis, A. S. & Sim, R. B. (1991). Oligomeric domain structure of human complement factor H by X-ray and neutron solution scattering. *Biochemistry*, **30**, 2847–2857.

Perkins, S. J., Haris, P. I., Sim, R. B. & Chapman, D. (1988). A study of the structure of human complement component factor H by Fourier transform infrared

spectroscopy and secondary structure averaging methods. *Biochemistry*, **27**, 4004–4012.

Perkins, S. J. (1988). X-ray and neutron solution scattering. *New Comp. Biochem*. **18B (Part II)**, 143–265.

Perkins, S. J. (1986). Protein volumes and hydration effects: the calculation of partial specific volumes, neutron scattering matchpoints and 280 nm absorption coefficients for proteins and glycoproteins from amino acid sequences. *Eur. J. Biochem.* **157**, 169–180.

Perkins, S. J. & Weiss, H. (1983). Low resolution structural studies of mitochondrial ubiquinol-cytochrome c reductase in detergent solutions by neutron scattering. *J. Mol. Biol*. **168**, 847–866.

Perez Sanchez, H., Tatarenko, K., Nigen, M., Pavlov, G., Imberty, A., Lortat-Jacob, H., Garcia de la Torre, J. & Ebel, C. (2006). Organization of human interferon gamma-heparin complexes from solution properties and hydrodynamics. *Biochemistry*, **45**, 13227–13238.

Perrimon, N. & Bernfield, M. (2000). Specificities of heparan sulphate proteoglycans in developmental processes. *Nature*, **404**, 725–728.

Petersen, S. V., Thiel, S. & Jensenius, J. C. (2001). The mannan-binding lectin pathway of complement activation: biology and disease association. *Mol. Immunol*. **38**, 133–149.

Petitou, M., Herault, J. P, Bernat, A., Driguez, P. A, Duchaussoy, P., Lormeau, J. C. & Herbert, J. M. (1999). Synthesis of thrombin-inhibiting heparin mimetics without side effects. *Nature*, **398**, 417–422.

Pickering, M. C., Cook, H. T., Warren, J., Bygrave, A.E., Moss, J., Walport, M. J. & Botto, M. (2002). Uncontrolled C3 activation causes membranoproliferative glomerulonephritis in mice deficient in complement factor H. *Nat. Genet*. **31**, 424–428.

Pillemer, L., Blum, L., Lepow, I. H., Ross, O. A., Todd, E. W. & Wardlaw, A. C. (1954). The properdin system and immunity. I. Demonstration and isolation of a new serum protein, properdin, and its role in immune phenomena. *Science*, **120**, 279–285.

Pio, R., Martinez, A., Unsworth, E. J., Kowalak, J. A., Bengoechea, J. A., Zipfel, P. F., Elsasser, T. H. & Cuttitta, F. (2001). Complement factor H is a

serumbinding protein for adrenomedullin, and the resulting complex modulates the bioactivities of both partners. *J. Biol. Chem*. **276**, 12292–12300.

Powell, A. K., Ahmed, Y. A., Yates, E. A. & Turnbull, J. E. (2010). Generating heparan sulphate saccharide libraries for glycomics applications. *Nat. Protoc*. **5**, 821-833.

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

Preissner, K. T. & Seiffert, D. (1998). Role of vitronectin and its receptors in haemostasis and vascular remodeling. *Thromb. Res*. **89**, 1–21.

Prodinger, W. M., Hellwage, J., Spruth, M., Dierich, M. P. and Zipfel, P. F. (1998). The C-terminus of factor H: monoclonal antibodies inhibit heparin binding and identify epitopes common to factor H and factor H-related proteins. *Biochem. J*. **331**, 41–47.

Prosser, B. E., Johnson, S., Roversi, P., Clark, S. J., Tarelli, E., Sim, R. B, Day, A. J. & Lea, S. M. (2007). Expression, purification, cocrystallization and preliminary crystallographic analysis of sucrose octasulphate/human complement regulator factor H SCRs 6-8. *Acta Crystallogr. Sect. F: Struct. Biol. Cryst. Commun*. **63**, 480–483.

Ragazzi, M., Ferro, D. R, Perly, B., Sinaÿ, P., Petitou, M. & Choay, J. (1990). Conformation of the pentasaccharide corresponding to the binding site of heparin for antithrombin III. *Carbohydr. Res*. **195**, 169–185.

Ralston, G. (1993). *Introduction to Analytical Ultracentrifugation.* Fullerton, CA: Beckman Instruments.

Ram, S., Sharma, A. K., Simpson, S. D., Gulati, S., McQuillén, D. P., Pangburn, M. K. & Rice, P. A. (1998). A novel sialic acid binding site on factor H mediates serum resistance of sialylated *Neisseria gonorrhoeae*. *J. Exp. Med.* **187**, 743–752.

Raman R., Venkataraman G., Ernst S., Sasisekharan V. & Sasisekharan, R. (2003). Structural specificity of heparin binding in the fibroblast growth factor family of proteins. *Proc. Natl. Acad. Sci. U.S.A*. **100**, 2357–2362.

Ramm, L., Whitlow, M. & Mayer, M. (1982). Size of transmembrane channels produced by complement proteins C5b-8. *J. Immunol*. **129**, 1143–1146.

Reid, K. M. (1986). Activation and control of complement system. *Essyas biochem*. **22**, 27–68.

Rice, K. G, Kim, Y. S, Grant, A. C, Merchant, Z. M. & Linhardt, R. J. (1985). High-performance liquid chromatographic separation of heparin-derived oligosaccharides. *Anal. Biochem*. **150**, 325–331.

Richards, A., Buddles, M. R., Donne, R .L., Kaplan, B .S., Kirk, E., Venning, M. C., Tielemans, C. L., Goodship, J. A. & Goodship, T. H. (2001). Factor H mutations in hemolytic uremic syndrome cluster in exons 18-20, a domain important for host cell recognition. *Am. J. Hum. Genet*. **68**, 485–490.

Richards, A., Kemp, E. J., Liszewski, M. K., Goodship, J. A., Lampe, A. K., Decorte, R., Muslumanolu, M. H., Kavukcu, S., Filler, G., Y. Pirson, *et al*. (2003). Mutations in human complement regulator, membrane cofactor protein (CD46), predispose to development of familial hemolytic uremic syndrome. *Proc. Natl. Acad. Sci. U.S.A*. **100**, 12966–12971.

Ripoche, J., Day, A. J., Harris, T. J. R. & Sim, R. B. (1988). The complete amino acid sequence of human complement factor H. *Biochem. J*. **249**, 593–602.

Rodén, L. (1989). In Heparin: Chemical and Biological Properties, Clinical Applications, ed. D. A. Lane and U. Lindahl, *CRC Press, Inc., Boca Raton, FL*, p. 1.

Rodríguez de Cordoba, S. R., Esparza-Gordillo, J., Goicoechea, de Jorge, E., Lopez-Trascasa, M. P. & Sanchez-Corral. (2004). The human complement factor H: functional roles, genetic variations and disease associations. *Mol. Immunol.* **41**, 355–367.

Rosenberg, R. D., Shworak, N. W., Liu, J., Schwartz, J. J. & Zhang, L. J. (1997). Heparan sulfate proteoglycans of the cardiovascular system. Specific structures emerge but how is synthesis regulated? *Clin. Invest*. **100**, 567–575.

Rother, K., Till, G. O. & Hansch, G. M. (Editors). (1998). The complement system. *Springer-Verlag Berlin Heidelberg New York*. (2nd Edition).

Salmivirta, M., Lindholt, K. & Lindahl U. (1996). Heparan sulphate: a piece of information. *FASEB J*. **10**, 1270–1279.

Sasisekharan, R., Shriver, Z., Venkataraman, G. & Narayanasami, U. (2002). Roles of heparan-sulphate glycosaminoglycan in cancer. *Nat. Rev. Cancer,* **2**, 521–528.

Sasisekharan, R. & Venkataraman, G. (2000). Heparin and heparan sulphate: biosynthesis, structure and function. *Curr. Opin. Chem. Biol*. **4**, 626–631.

Sasisekharan, R., Ernst, S. & Venkataraman, G. (1997). On the regulation of fibroblast growth factor activity by heparin-like glycosaminoglycans. *Angiogenesis*. **1**, 45–54.

Saunders, R. E., Abarrategui-Garrido, C., Frémeaux-Bacchi, V., Goicoechea de Jorge, E., Goodship, T. H. J., López Trascasa, M., Noris, M., Ponce Castro, I. M., Remuzzi, G., Rodríguez de Córdoba, S., Sánchez-Corral, P., Skerka, C., Zipfel, P. F. & Perkins, S. J. (2007). The interactive factor H—atypical haemolytic uraemic syndrome mutation database and Website: update and integration of membrane cofactor protein and factor I mutations with structural models. *Human Mutation*. **28**, 222–234.

Saunders, R. E., Goodship, T. H. J., Zipfel, P. F. & Perkins, S. J. (2006). Factor H associated haemolytic uraemic syndrome: a web database of the structural consequences of disease-associated mutations. *Human Mutation*. **27**, 21–30.

Scarpellini, A., Germack, R., Lortat-Jacob, H., Muramatsu T., Billett E., Johnson T. & Verderio, E. A. (2009). Heparan sulphate proteoglycans are receptors for the cell-surface trafficking and biological activity of transglutaminase-2. *J. Biol. Chem.* **284**, 18411-18423.

Schachman, H. K. (1959). *Ultracentrifugation in Biochemistry*. New York: Academic Press.

Schlessinger, J., Plotnikov, A. N., Ibrahimi, O. A., Eliseenkova, A. V, Yeh, B. K, Yayon, A., Linhardt, R. J. & Mohammadi, M. (2000). Crystal structure of a ternary FGF-FGFR-heparin complex reveals a dual role for heparin in FGFR binding and dimerization. *Mol. Cell*. **6**, 743-450.

Schmidt, C. Q., Herbert, A. P., Kavanagh, D., Gandy, C., Fenton, C. J., Blaum, B. S., Lyon, M., Uhrin, D. & Barlow, P. N. (2008). A new map of glycosaminoglycan and C3b binding sites on Factor H. *J. Immunol*. **181**, 2610–2619.

Schwarz, M., Spath L., Lux, C. A., Paprotka, K., Torzewski, M., Dersch, K,. Koch-Brandt, C., Husmann, M. & Bhakdi, S. (2008). Potential protective role of apoprotein J (clusterin) in atherogenesis: binding to enzymatically modified low-density lipoprotein reduces fatty acid-mediated cytotoxicity. *Thromb. Haemost*. **100**, 110–118.

Schuck, P. (2000). Size distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. *Biophys. J.* **78**, 1606–1619.

Scott, D. J. & Schuck, P. (2005). A brief introduction to the analytical ultracentrifugation of proteins for beginners. *In "Modern Analytical Ultracentrifugation: Techniques and Methods"* (eds.) Scott, D. J., Harding, S. E. & Rowe, A. J. Royal Society of Chemistry, London, U.K. Chapter **1**,pp 1-25.

Shao, C., Zhang, F., Kemp, M. M., Linhardt, R. J., Waisman, D. M., Head, J. F. & Seaton, B. A. (2006). Crystallographic analysis of calcium-dependent heparin binding to annexin A2. *J. Biol. Chem.* **281**, 31689–31695.

Sharma, A. K. & Pangburn, M. K. (1996). Identification of three physically and functionally distinct binding sites for C3b in human complement factor H by deletion mutagenesis. *Proc. Natl Acad. Sci. U.S.A.* **93**, 10996–11001.

Sharath, M. D., Merchant, Z. M., Kim, Y. S., Rice, K. G., Linhardt, R. J. & Weiler, J. M. (1985). Small heparin fragments regulate the amplification pathway of complement. *Immunopharmacology*, **9**, 73–80.

Shaya, D., Zhao, W., Garron, M. L, Xiao, Z., Cui, Q., Zhang, Z., Sulea, T., Linhardt, R. J. & Cygler, M. (2010). Catalytic mechanism of heparinase II investigated by site-directed mutagenesis and the crystal structure with its substrate. *J. Biol. Chem*. **85**, 20051–20061.

Shieh, M. -T. & Spear, P. G. (1994). Herpes virus-induced cell fusion that is dependent on cell surface heparan sulfate or soluble heparin. *J. Virol*. **68**, 1224–1228.

Shukla, D., Liu, J., Blaiklock, P., Shworak, N. W., Bai, X., Esko, J. D, Cohen, G. H, Eisenberg, R. J, Rosenberg, R. D. & Spear, P. G. (1999). A novel role for 3-O-sulfated heparan sulfate in herpes simplex virus 1 entry. *Cell*, **99**, 13-22.

Semenyuk, A. V. & Svergun, D. I. (1991). GNOM-a program package for small angle scattering data processing. *J. Appl. Crystallogr*. **24**, 537–540.

Seya, T. & Atkinson, J. P. (1989). Functional properties of membrane cofactor protein of complement. *Biochem. J*. **264**, 581–538.

Sim, R. B., Day, A. J., Moffatt, B. E. & Fontaine, M. (1993). Complement factor I and cofactors in control of complement system convertase enzymes. *Methods Enzymol*. **223**, 13–35.

Sim, R. B. & DiScipio, R. G. (1982). Purification and structural studies on the complement-system control protein beta 1H (Factor H). *Biochem. J.* **205**, 285–293.

Soares, D. & Barlow, P. N. (2005). Complement control protein modules in the regulators of complement activators. In "Structural Biology of the Complement System" (Eds. Morikis, D & Lambris, J. D.) pp 19-62. Taylor & Francis, Boca Raton, U.S.A.

Spendlove, I., Ramage, J. M., Bradley, R., Harris, C. & Durrant, L. G. (2006). Complement decay accelerating factor (DAF)/CD55 in cancer. Cancer. *Immunol. Immunother*. **55**, 987–995.

Spillmann, D., Witt, D. & Lindahl, U. (1998). Defining the interleukin-8-binding domain of heparan sulphate, *J. Biol. Chem*. **273**, 15487–15493.

Spivak-Kroizman, T, Lemmon, M. A, Dikic, I., Ladbury, J. E, Pinchasi, D., Huang, J., Jaye, M., Crumley, G., Schlessinger, J. & Lax, I. (1994). Heparin-induced oligomerization of FGF molecules is responsible for FGF receptor dimerization, activation, and cell proliferation. *Cell*, **79,** 1015–1024.

Stafford, W. F. (1992). Boundary analysis in sedimentation transport experiments: A procedure for obtaining sedimentation coefficient distributions using the time derivative of the concentration profile. *Anal. Biochem*. **203,** 295–301.

Stringer, S. E., & Gallagher, J. T. (1997). Specific binding of the chemokine platelet factor 4 to heparan sulphate. *J. Biol. Chem*. **272**, 20508–20514.

Stringer, S. E., Forster, M. J., Mulloy, B., Bishop, C. R., Graham, G. J. & Gallagher, J. T. (2002). Characterization of the binding site on heparan sulfate for macrophage inflammatory protein 1α. *Blood*, **100**, 1543–1550.

Svedberg, T. & Pedersen, K. O. (1940). *The ultracentrifuge*. London: Oxford University Press.

Svergun, D. I. & Koch, M. H. (2003). Small angle scattering studies of biological macromolecules in solution. *Rep. Prog. Phys*. **66**, 1735–1782.

Tanford, C. (1961). *Physical chemistry of macromolecules*. New York: Wiley.

Thiel, S., Vorup-jensen, T., Schaeble, W., laursen, S. B., Poulsen, K., Willis, A. C., Eggleton, P., Hansen, S., Holmskov, U., Reid, K. B. & Jensenius, J. C. (1977). A second serine protease associated with mannan binding lectin that activates complement. *Nature*, **386**, 506–510.

Thompson, R. A. & Winterborn, M. H. (1981). Hypocomplementaemia due to a genetic deficiency of beta 1H globulin. *Clin. Exp. Immunol*. **46**,110–119.

Thurman, J. & Holers, V. M. (2006). The central role of the alternative complement pathway in human disease. *J. Immunol.* **176**, 1305–1310.

Torri, G., Casu, B., Gatti, G., Petitou, M., Choay, J., Jacquinet, J. C. & Sinay, P. (1985). Mono- and bidimensional 500 MHz 1H-NMR spectra of a synthetic pentasaccharide corresponding to the binding sequence of heparin to antithrombin-III: evidence for conformational peculiarity of the sulfated iduronate residue. *Biochem. Biophys. Res. Commun*. **128**, 134–140.

Turnbull, J. E, Fernig, D. G, Ke, Y, Wilkinson, M. C. & Gallagher, J. T. (1992). identification of the basic FGF binding sequence in fibroblast heparan sulphate. *J. Biol. Chem*. **267**, 10337–10341.

Turnbull, J.E & Gallagher, J. T. (1990). Molecular organization of heparan sulphate from human skin fibroblasts. *Biochem. J*. **265**, 715–724.

Turnbull, J.E & Gallagher, J. T. (1988). Oligosaccharide mapping of heparan sulphate by polyacrylamide-gradient-gel electrophoresis and electrotransfer to nylon membrane. *Biochem. J*. **251**, 597–608.

van Holde, K. E. (1975). Sedimentation analysis of proteins. In Neurath H, Hill RH (ed): *The Proteins,* I. New York: Academic Press. pp 225–291.

van Holde, K. E. & Weischet, W. O. (1978). Boundary analysis of sedimentation velocity experiments with monodisperse and paucidisperse solutes. *Biopolymers.* **17**, 1387–1403.

van Leeuwen, R., Klaver, C. C., Vingerling, J. R., Hofman, A. & de Jong, P. T. (2003). Epidemiology of age-related maculopathy. *Eur. J. Epidemiol.* **18**, 845–854.

Vives, R.R., Goodger, S. & Pye, D.A. (2001). Combined strong anion exchange HPLC and PAGE approach for the purification of heparan sulphate oligosaccharides. *Biochem. J*. **354**, 141-147.

Venkataraman, G., Sasisekharan, V., Herr, A. B, Ornitz, D.M., Waksman, G., Cooney, C. L, Langer, R. & Sasisekharan, R. (1996). Preferential self-association of basic fibroblast growth factor is stabilized by heparin during receptor dimerization and activation. *Proc. Natl. Acad. Sci. USA*. **93**, 845-850.

Walport, M. J. (2001). Complement. Second of two parts. *N. Engl. J. Med*. **344**, 1140–1144.

Weiler, J. M., Daha, M. R., Austen, K. F. & Fearon, D. T. (1976). Control of the amplification convertase of complement by the plasma protein beta1H. *Proc. Natl. Acad. Sci. U.S.A.* **73**, 3268–3272.

Whaley, K. & Ruddy, S. (1976). Modulation of the alternative complement pathway by beta 1 H globulin. *J. Exp. Med*. **144**, 1147–1163.

Whaley, K. & Ruddy, S. (1976). Modulation of C3b hemolytic activity by a plasma protein distinct from C3b inactivator. *Science*, **193**, 1011–1013.

Wüthrich, K. (1995). Nmr - This Other Method for Protein and Nucleic-Acid Structure Determination. *Acta Cryst.* **51**, 249-270.

Wüthrich, K. (1986). NMR of proteins and nucleic acids. New York, Wiley.

Xiong, Z. Q., Qian, W., Suzuki, K. & McNamara, J. O. (2003). Formation of complement membrane attack complex in mammalian cerebral cortex evokes seizures and neurodegeneration. *J. Neurosci.* **23**, 955-960.

Zipfel, P. F. & Skerka C. (2009). Complement regulators and inhibitory proteins. *Nature Reviews Immunology*, **9**, 729–740.

Zuiderweg, E. R. (2002). Mapping protein-protein interactions in solution by NMR spectroscopy. *Biochemistry*, **41**, 1–7.

Publications

Khan, S., Rodriguez, E., Patel, R., Gor, J., Mulloy, B. & Perkins, S. J. (2011). The solution structure of heparan sulphate differs from that of heparin: Implications for function. *J. Biol. Chem*. **submitted**.

Khan, S., Gor, J., Mulloy, B. & Perkins, S. J. (2010). Semi-rigid solution structures of heparin by constrained X-ray scattering modelling: new insight into heparin-protein complexes. *J. Mol. Biol*. **395**, 504-521.

Perkins, S. J., Nan, R, Okemefuna, A. I., Li, K., Khan, S. & Miller, A. (2010). Multiple interactions of complement factor H with its ligands in solution: A progress report. *Adv. Exp. Med. Biol*. **703**, 25-47.

Perkins, S. J., Nan, R., Li, K., Khan, S. & Abe, Y. (2011). Analytical ultracentrifugation combined with X-ray and neutron scattering: experiment and modeling. *Methods,* **in press**.

Perkins, S. J., Nan, R., Li, K. & Khan, S. (2011). Complement factor H and its ligands–a multidisciplinary approach to interactions and modelling. *Encyclopedia of Biophysics: Molecular shape and hydrodynamics,* **in press**.

Publications in preparation

Khan, S., Nan, R, Gor, J., Mulloy, B. & Perkins, S. J. Complement factor H possesses multiple independent binding sites for heparin and heparan sulphate: Implications for function and disease

Abstracts

Khan, S., Miller, A,. Gor, J., Mulloy, B. & Perkins, S. J. (2010). Oligomeric solution structures of factor H in the presence of heparin fragments. *Molecular Immunology*, **47**, 2264, (Poster).

Mulloy, B., Khan, S., Perkins, S. J. & Forster, M. (2009). Heparin-protein interactions: a three dimensional approach, *Glycobiology* (Conference abstract), 1296.

Khan, S., Mulloy, B. & Perkins, S. J. (2008). The solution structure of the factor H SCR-6/8 domains differs from its crystal structure: Implications for mechanism**.** *Molecular Immunology*, **45**, 4125-4126, (Poster).

Khan, S., Mulloy, B. & Perkins, S. J. (2008). Solution structure of heparin and its complexes with factor H SCR-6/8 domains and factor H: Implications for disease, *Molecular Immunology*, **45**, 4126, (Poster).