Glycosaminoglycan-protein interactions and human complement factor H

Bärbel Blaum

PhD Thesis

University of Edinburgh

2010

Abstract

Glycosaminoglycans (GAGs) are linear polysaccharides expressed ubiquitously on animal cell surfaces and within extracellular matrices. GAGs usually occur as parts of proteoglycans and often accomplish their biological functions through their interactions with proteins. GAG oligosaccharides for this work were produced via enzymatic digest of heparin, followed by gel filtration and ion exchange chromatography. Two tetrasaccharide species obtained from this digest were characterised using ^{1}H and ^{13}C NMR spectroscopy.

Complement factor H (fH) is a regulatory protein of the alternative pathway of the complement system, a major component of human innate immunity. Acting as a cofactor to factor I, fH inhibits C3b-initiated complement activation on host cells, protecting cells from auto immune attack. This study focused on the interaction of factor H with GAGs, which are thought to be among the markers allowing factor H to distinguish between self and non self surfaces. Binding studies of two heparin-binding sites in fH are presented. These include the C-terminal modules 19 and 20 (fH∼19-20) and fH∼7-8. FH∼7, fH∼7-8 and fH∼19-20 were produced recombinantly in various isotope forms. The techniques used to study the protein-GAG interactions in this work encompass NMR spectroscopy, mass spectrometry, gel mobility shift assays (GMSA) and chemical cross linking.

Several genetic studies suggest that a common polymorphism in the heparin-binding module fH∼7, Y402H, plays a role in the development of age-related macular degeneration (AMD). The work presented here included preparation and backbone resonance assignment of a ^{13}C, ^{15}N- labelled sample of fH∼ 7-8 via triple resonance NMR experiments. Further NMR experiments were employed to investigate the role of the lysine and arginine sidechains of fH∼7 in GAG binding. These studies were combined with the preparation and characterisation of a covalently cross linked GAG-protein complex using NMR and mass spectrometry.

A range of fH∼19-20 mutations that are linked to a severe kidney disease, atypical haemolytic uraemic syndrome (aHUS), were characterised using GMSA. No correlation between the disease and the heparin binding properties of the aHUS mutants was observed. The mutant proteins were also characterised with respect to their ability to compete with full-length fH in a physiological complement assay. Simultaneous binding of WT fH∼19-20 to GAGs and C3d, the relevant fragment of C3b, was assessed using NMR.

NMR experiments were also conducted with NK1, which comprises the two N-terminal heparin-binding modules of hepatocyte growth factor/scatter factor (HGF/SF), and heparin as well as dermatan sulfate-derived GAGs. Relaxation studies on a human defensin, HBD2, were performed to assess the role of GAGs in HBD2 self-association.

Acknowledgements

I am deeply indebted to my supervisor, Dr. Dušan Uhrín, for his continuous support, patience and encouragement, and for his practical help in learning how to use an NMR spectrometer.

Edinburgh University was an enjoyable place to work, largely due to the enormous friendliness which I encountered in the Uhrín and Barlow group offices and the shared laboratory. I am grateful to Andy, Conny, Emily, Lan, Juraj and Dušan for making our office a pleasant place to work. Many thanks to Juraj for happily answering all my technical NMR questions and being such an incredibly friendly person, to Andy for teaching me many things from fermentation to spectra processing (and English), and also for being a good listener; and to Conny for convincing me that mass spec is useful, explaining what att palla äpplen means and demonstrating that big Swedes are also big friends.

I am grateful for the great company I had in lab 120, i.e. Isa, Christoph, Carla, Marie, Patience, Maria, Mara, Nicky, Elisa and many more. Thanks for putting up with my moodiness and good luck to all of you!

Thanks to John White for his friendliness and interest which made the fermentation suite a better place.

Malcolm Lyon and Jonathan Deakin is thanked for a fruitful collaboration. Their expertise for GAGs and their willingness to share it are much appreciated.

I also wish to thank Janice Bramham, Dave Kavanagh and Dinesh Soares who encouraged me on many occasions and whom I wish the stamina to remain in science.

Especially in the first phase of my PhD, Prof. Paul Barlow helped to guide this project through unexpected turns and twists. I gratefully acknowledge his help as well as his continuing interest and valuable opinions during the years which I spent in Edinburgh.

Despite the distance to mainland Europe, my family continued to support me in many ways, as they have done all my life. Without their affection, motivation and understanding, I would likely not be in this privileged situation of completing a scientific PhD abroad. I therefore thank, last but not least, my parents, my grandma and my brother, Wolf.

And thanks for the bike! It reminds me of a good place up North.

Unless stated in the text, the work described in this thesis is my own work and has not been submitted in whole or in part for a degree or other qualification at this or any other university. Articles in which the work described in this thesis has been published can be found in the bibliography.

Contents

0.1 Abbrevations

Table 0.1: Abbrevations

1D	one-dimensional
2D	two-dimensional
3D	three-dimensional
A_{XYZ}	absorbance at XYZ nm
AGE	advanced glycation endproducts
AMAC	2-aminoacridone
AMD	age-related macular degeneration
AOX	alcohol oxidase
AP	alternative pathway of complement activation
aHUS	atypical haemolytic uraemic syndrome
BMG	buffered minimal glycerol
BMM	Buffered minimal methanol
C4bBP	C4b-binding protein
CCP	complement control protein
CD	cluster of differentiation
CHCA	α-cyano-4-hydroxy-cinnamic acid
COSY	correlation spectroscopy
CRINEPT	cross relaxation-enhanced polarisation transfer
CRIPT	cross relaxation-induced polarisation transfer
CRn	complement receptor type n
CRP	C-reactive protein
CSA	chemical shift anisotropy
CW	continuous wave
DAF	decay accelerating factor
DD	dipole-dipole
ΔUA2S	4, 5-dehydrated uronic acid-2-sulfate
dH_2O	destilled water
DMSO	diemthylsulfoxide
DO	dissolved oxygen
DOSY	diffusion ordered spectroscopy
dpn	oligosaccharide with degree of polymerisation n
dpnC	fully sulfated oligosaccharide with degree of polymerisation n
DS	dermatan sulfate

Table 0.1: *continued*

DTT	dithiothreitol
ϵ_{XYZ}	extinction coefficient at XYZ nm
ECM	extracellular matrix
EDC	1-Ethyl-3-[3-dimethylaminopropyl]-carbodiimide hydrochloride
EDTA	ethylenediamine-N, N, N', N'-tetraacetic acid
EGTA	ethylene glycol-bis(2-aminoethyl)-N, N, N', N'-tetraacetic acid
fB	complement factor B
FGF	fibroblast growth factor
FGFR	fibroblast growth factor receptor
fH	complement factor H
fH∼n	complement factor H module n
fH∼7Z	complement factor H module 7 with amino acid type Z in position 402
fH∼7Yxdp4C	complement factor H module 7 cross linked to dp4C
fHL-1	factor H-like protein 1
fI	complement factor I
FID	free induction decay
FPLC	fast protein liquid chromatography
FT	Fourier transform
GABA	γ-aminobutyric acid
GAG	glycosaminoglycan
GalNAc	N-acetylated α-D-galactosamine
GalNAc4S	N-acetylated α-D-galactosamine-4-sulfate
GlcA	β-D-glucuronic acid
GlcN	α-D-glucosamine
GlcNS6S	N-sulfated α-D-glucosamine-6-sulfate
GMSA	gel mobility shift assay
h	hour(s)
HBD	human β-defensin
HBV	hepatitis B virus
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HGF/SF	hepatocyte growth factor/scatter factor
HISCQ	heteronuclear in-phase single quantum coherence
HPLC	high performance liquid chromatography
HS	heparan sulfate
HS-NA	non-sulfated domain in heparan sulfate

Table 0.1: *continued*

HS-NS	sulfated domain in heparan sulfate
HMQC	heteronuclear multi quantum coherence
HSQC	heteronuclear single quantum coherence
IdoA	α-L-iduronic acid
IdoA2S	α-L-iduronic acid-2-sulfate
IL	interleukin
INEPT	insensitive nuclei enhanced by polarization transfer
ITC	isothermal titration calorimetry
K1	kringle 1 domain of HGF/SF
K_d	equilibrium dissociation constant
kDa	kilodalton
KOH	potassium hydroxide
MAAH	microwave-assisted acid hydrolysis
MAC	membrane attack complex
MALDI-TOF	matrix-assisted laser desorption/ionization-time of flight
MBL	mannose binding lectin
MCP	membrane cofactor protein
MES	2-(N-morpholino)ethanesulfonic acid
MET	mesenchymal-epithelial transition factor
MG	macroglobulin
min	minutes
MPGN	Membrano-proliferative glomerulonephritis
MS	mass spectrometry
MS-MS	tandem mass spectrometry
MWCO	molecular weight cutoff
NaCl	sodium chloride
NaOH	sodium hydroxide
NHS	normal human serum
NK1	N-terminal domain and kringle 1 domain of HGF/SF
NK2	NK1 and kringle 2 domain of HGF/SF
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect
OD_{600}	optical density at 600 nm
P. Pastoris	*Pichia pastoris*
PBS	phospate buffered saline

Table 0.1: *continued*

PDB	protein database
PMSF	phenylmethyl sulfonyl fluoride
ppm	parts per million
R_1	longitudinal or spin-lattice relaxation rate
R_2	transverse or spin-spin relaxation rate
RBC	red blood cell
RCA	regulators of complement activation
ROESY	rotational nuclear overhauser effect spectroscopy
RPE	retinal pigment epithelium
rpm	revolutions per minute
s	seconds
SAX	strong anion exchange
SCR	short consensus repeat
SDS	sodium dodecyl sulfate
SDS-PAGE	sodium dodecyl sulfate - polyacrylamide gel
SNP	single nucleotide polymorphism
SOS	sucrose octasulfate
SPR	surface plasmon resonance
sulfo-NHS	sulfo-N-hydroxysuccinimide
T_1	longitudinal or spin-lattice relaxation time
T_2	transverse or spin-spin relaxation time
τ_c	rotational correlation time
TED	thioester-containing domain (of C3)
TEMPO	2,2,6,6-tetramethylpiperidine-1-oxy
TFA	trifluoroacetic acid
TOCSY	total correlation spectroscopy
TROSY	transverse relaxation-optimized spectroscopy
TNF	tumor necrosis factor
Tris	tris(hydroxymethyl)aminomethane
UV/vis	ultraviolet/visible
WT	wilde type
YNB	yeast nitrogen base medium
YPD	yeast extract peptone dextrose

1 Introduction

1.1 The Complement System

The human complement system owes its name to Jules Bordet's discovery that fresh serum complements antibody activity (though the actual name complement was coined by Paul Ehrlich). In a historic experiment, Bordet found that heated blood could no longer lyse bacteria and that it regained this ability after the addition of fresh serum which did not contain antibodies [1]. He concluded that the heated antibodies where still active when complemented by a heat-sensitive serum component - the complement system (more specifically its classical pathway). It was later found that the complement system does also comprise an antibody-independent pathway (the alternative pathway) that protects its host against microorganisms which have not previously been encountered, and for which the host lacks antibodies. Since this part of the complement system - unlike the acquired immune system - does not require exposure to antigens prior to functioning it is classified as part of the innate immune system. As such it poses an early line of defence against invading microorganisms, including many bacteria, fungi and viruses [2].

Defence mechanisms by which the complement system act include lysis of foreign cells via the membrane attack complex, chemotaxis, increase of blood flow at infected sites, and activation of leukocytes [3, 4, 5], including opsonisation for phagocytosis [6, 7]. The complement system clears the host of immune complexes [8, 9, 10] and apoptopic cells [11, 12, 13]. Through its link to the acquired immune system via the classical pathway the complement system can augment the adaptive immune system's response [14, 15].

1.1.1 The Three Pathways of Complement

The complement system can be activated via three different pathways: the classical, lectin and alternative pathway. The *classical pathway* is activated by binding of C1q to IgG or IgM complexed with antigens or by directly binding to pathogen surfaces [16, 17, 18]. The initiating complex of the *lectin pathway* is formed when mannose-binding lectin (MBL) or ficolin bind to mannose residues on pathogen surfaces, making the lectin pathway part of the antibody-independent innate immune system [19]. The *alternative pathway* (AP) is permanently active on a low level, needs to be actively controlled and can be boosted through a positive feedback loop upon encounter with pathogens [16]. Of the three pathways of complement, this work is mostly concerned with the alternative pathway.

1.1.2 Components of the Complement System

Complement proteins make up about 5% of the blood serum globulin fraction. Additionally, the complement system contains a number of membrane bound immune receptors, some of them integrins [7, 20]. Most of the serum proteins are present as zymogens, with only a small portion being permanently activated (the so called complement "tick-over" of the AP [21]). Many of the fluid phase proteins follow a simple nomenclature: one group (proteins of the classical pathway and the membrane attack complex) comprises C1q, C1s, C1r, C2, C3 etc. up to C9 (where C stands for *C*omplement component). A second group (belonging to the alternative pathway) contains Factor B, Factor D, Factor I (fI) and Factor H (fH). C3 is a central member of all three complement pathways. Some proteins have individual names, such as properdin, decay accelerating factor (DAF) or C4b binding protein (C4bBP). Complement receptors (CR) follow the cluster of differentiation (CD) nomenclature but have often several names - DAF, for example, does also have the systematic name CD55, and the integrins CR3 and CR4 have additional names which follow the integrin nomenclature. The numbering of the C-type proteins reflects the order of discovery rather than functional aspects.

Upon proteolytic activation, C2-C5 and Factor B split into a (usually smaller) subunit, named C3a, C5a etc. and a bigger subunit C3b, C5b etc (with the exception of C2, which is composed of a smaller subunit C2b and a bigger subunit C2a). After cleavage, the a and b subunits fulfil separate functions: while the bigger subunits form large complexes like C3bBb or C3BbBb the smaller fragments fulfil signalling functions. C3a and C5a, for example, are anaphylotoxins [22]. The large complexes are again proteolytic and go on to activate further molecules, leading to a dramatic amplification and making the complement system a triggered-enzyme cascade like the blood coagulation system.

1.1.3 The central role of C3b

All three pathways converge at the level of C3, which is comprised of an α chain (111 kDa) and a β chain (75 kDa), held together by a disulfide bond. It is cleaved into the opsonin C3b (176 kDa) and the anaphylotoxin C3a (9 kDa) by either C4b2a (the classical and lectin pathway C3 convertase) or the alternative pathway C3 convertase C3bBb. Nascent C3b exhibits an activated thioester which can covalently bind to cell surfaces, making the attached C3b molecule the centre of more C3b production (after formation of C3bBb) and thus initiating a positive feedback loop. C3bBb has a short life time ($t_{1/2}$ of only ca. 90 sec [23]) unless stabilized by properdin. After cleaving another C3 molecule C3bBb forms $C3b_2Bb$, the AP C5-convertase. C4b2a can also cleave C3 to yield the classical pathway 5-convertase C4b2a3b. Proteolytic cleavage of C5 by C5-convertase leads to formation of C5b, the starting point of the membrane attack complex (MAC). C5b assembles C6, C7, C8 and multiple copies of C9. The C5bC6C7 complex is able to insert into the phospholipid bilayer of cell membranes via a hydrophobic patch on C7 that becomes exposed during the C5bC6C7 formation. After association of C8 and up to 16 C9 molecules the complex forms a large pore in the bilayer that results in cell lysis [24, 25, 26, 27]. C3a, on the other hand, acts as a signalling peptide, increasing vasolidation, up-regulating complement receptors and serving as a chemo attractant for leukocytes [28].

The activation of the thioester bond in C3b - which originates from an unusual post translational modification of Cys988 and Gln991 [29] - has been described as a two-step mechanism [30, 31]: in C3, the thioester lies well-protected in a hydrophobic pocket between the TED (*T*hio*E*ster-containing *D*omain) and MG8 (*M*acro*G*lobulin 8) domains, but cleavage of C3 into C3a and C3b triggers a conformational change which translocates the TED domain and thioester by 85 Å, exposing the latter and bringing it into close proximity with His1104 [32]. Following the domain rearrangement, an acyl-imidazole intermediate is formed with the His1104 side chain, releasing Cys988 from the ester. This deprotonated thiolate now acts as a base, conferring higher nucleophilicity to hydroxyl groups on the substrate surface, which in turn displaces the His1104 side chain from the acyl-imidazole intermediate. The His-intermediate is extremely short lived with a half life of less than 100 µs [33] and can be hydrolysed by solvent molecules like water. Formation of nascent C3b is, however, likely to take place near target cell surfaces or immune complexes because C3 convertases are generated on such sites. Thus, the His-intermediate is likely to react with a cell surface hydroxyl group instead of a water molecule, leading to covalent attachment of C3b through an ester bond with the Gln991 sidechain. It has been suggested that the presence of negatively charged cell surfaces might facilitate the dislocation of the TED domain and exposure of the thioester through electrostatic interactions between a positively charged patch on the TED domain and the complementary charged cell surface [31]. Additionally, the exact chemical nature of the binding nucleophile on the target surface has been shown to modulate binding affinity of the thioester, providing some direct specificity for C3b deposition [34]. This specificity, however, is not sufficient to discriminate between friend and foe, and in order to protect autologous cells from deposition of C3b a range of complement regulators are needed.

1.1.4 Fluid-phase activation of C3

The AP is permanently "switched on" due to a spontaneous albeit slow hydrolysis of an internal thioester bond in C3 ($t_{1/2} > 6$ days, [35]). After hydrolysis, C3 is denoted

$C3(H_2O)$, which has been shown to structurally resemble C3b [36], despite the fact that it still contains the ANA domain which, after cleavage, forms the *ana*phylotoxin C3a (Fig.1.1). The conformational change triggered by the hydrolysis reaction allows $C3(H_2O)$ to bind factor B, which is then cleaved by factor D to yield the fluid-phase C3 convertase $C3(H_2O)Bb$. This convertase can now catalyse the production of C3b through proteolysis of C3. Nascent C3b can either be deactivated by hydrolysis of the critical thioester or follow the above outlined mechanism to opsonise target surfaces, thereby triggering complement activation on cells.

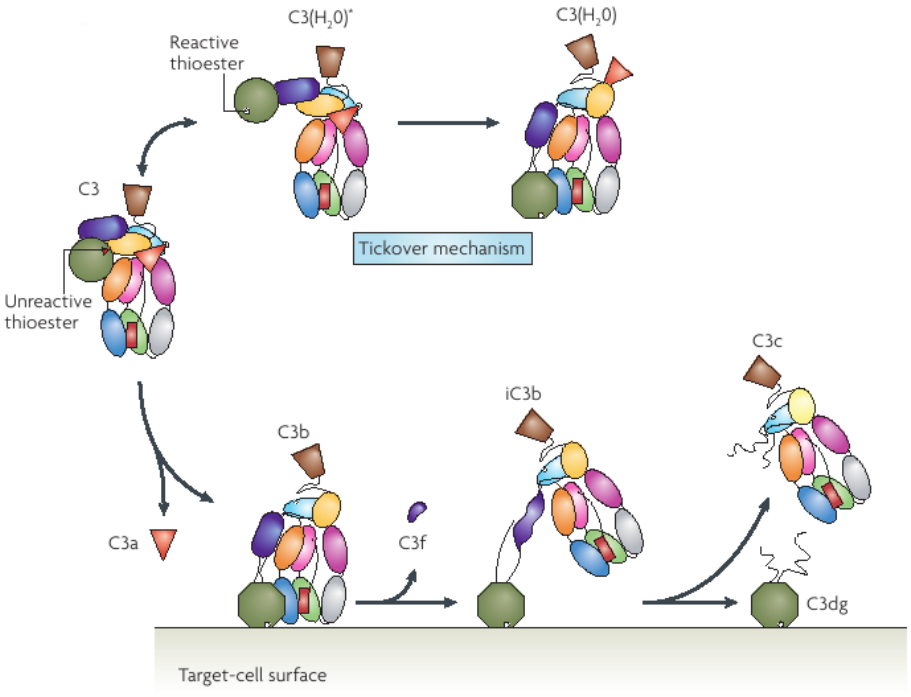

Figure 1.1: C3 activation by slow hydrolysis (top) or enzymatic cleavage (bottom) lead to conformational changes which displace the TED domain (dark green). Deactivation of C3b by fI and co-factors leads to loss of the small C3f fragment, followed by the cleavage of the TED domain, now dubbed C3dg. Figure taken from [37].

1.2 Regulation of the Complement System

In order to protect host cells from the complement system it needs to be actively controlled by a group of regulatory proteins, many of which are structurally homologous. These are fH, C4b-binding protein (C4bBP), decay acceleration factor (DAF, or CD55), membrane co-factor protein (MCP, also called CD46) and complement receptor type 1 (CR1, also called CD35). Their common structural basis are the so called *C*omplement *C*ontrol *P*rotein (CCP) modules or short consensus repeats (SCR). A CCP module is typically made of approximately 60 residues which adopt an ellipsoid shape of ca. 30 Å length and ca. 15 Å diameter. A hydrophobic core is wrapped in β-sheets and stabilized by two disulfide bonds (cys^1-cys^3 and cys^2-cys^4). Conserved amino acids are - along with the four cysteine residues involved in disulfide bridging - an almost invariant buried tryptophan between cys^3 and cys^4, prolines and glycines as well as other hydrophobic residues (Fig. 1.2). The CCP-domain boundaries are usually defined relative to the N- and C-terminal cysteine residues cys^1 and cys^4, respectively, and connected by linker regions of three to seven residues length and variable sequence [38]. N- and C-terminus of a CCP module lie at opposite ends of the ellipsoid shape, conferring extended structures to proteins with several CCP domains. FH, C4bBP and their viral analogue vaccinia virus complement control protein (VCP), which all contain a high percentage of CCP modules, have been shown to be extremely stable under adverse conditions such as high temperature and extreme pH values [39, 40]. The high level of structural homology of CCP domains suggests that this stability is a general feature of CCP domains. Presumably the hydrophobic core and high number of proline residues - both characteristics of thermophilic proteins - contribute to the stability of CCP modules as well as the conserved disulfide bonds. Charged patches on the protein surfaces, as present in many CCP modules, could additionally confer stability as they counteract protein aggregation. Interestingly, the CCP domain structure has also been found in proteins outside the complement system. One example is the GABA receptor GABA(B)R1a which occurs in three splice variants in humans, two of which contain at least one N-terminal CCP

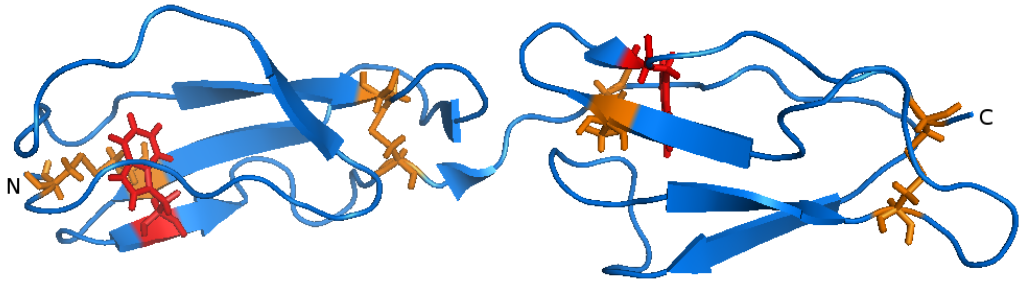

Figure 1.2: NMR structure of the two N-terminal CCP-domains of factor H (CCP 1 and CCP 2) with highlighted cysteine (orange) and conserved tryptophane residues (red). PDB entry 1RLP [41].

module [42, 43]. The b subunit of the coagulation factor XIII is also made of 10 CCP modules [44].

Most of the complement regulators are essentially co-factors to a central protease that diffuses through blood in its activated form and that does not belong to the CCP structural family: complement factor I (fI) [45]. The regulators also accelerate the decay of C3-convertase and act as inhibitors of the convertase formation, consequently their activity is usually described as two-fold: decay acceleration and fI co-factor activity. FI can cleave both C3b and its classical pathway analogue C4b only after co-factor assisted dissociation of the convertase complexes. On cell surfaces, the decay accelerating proteins comprise the fluid phase proteins factor H (for decay of C3bBb) and C4BP (for decay of C4b2a) as well as the membrane bound co-factors CR1 (CD35) and DAF (CD55) (both for decay of the classical and the alternative pathway convertase). After dissociation from the convertase complex, C3b binds to factor H, CR1 or MCP and is subsequently degraded to iC3b and C3f by fI [46]. C4b suffers the same fate after binding of C4BP, CR1 or MCP, leaving C4c and C4d behind. MCP does not have decay accelerating activity and DAF lacks co-factor activity. The dissociation and deactivation of fluid phase C3bBb and C3b, which can arise from the activity of $C3(H_2O)Bb$ in solution, is only assisted by fI and factor H, making factor H the main complement control protein in

the fluid phase and on cell surfaces that lack membrane bound regulators [45]. In the presence of its co-factor CR1, fI further cleaves iC3b into C3c (140 kDa) and C3dg (35 kDa) which remains bound to the cell surface (Fig.1.1) [47].

It has been noted that the fate of the central complement component C3b can be described as a competition between Factor B and fH: binding of Factor B (in the presence of factor D) initiates formation of C3bBb and the activation of the complement cascade, while binding of fH (in the presence of fI) results in degradation and deactivation of C3b. It appears that the immediate environment of C3b determines which of these two ways is chosen, and subtle contributions of many (de)stabilizing factors could contribute to the decision.

All regulators of complement activation (RCA) are encoded in a single gene cluster on chromosome 1q32 [48, 49, 50] (along with CR2 which is also made of CCP modules but does not share the regulatory functions of its genetic neighbours). Despite individual activity profiles and some varying structural characteristics like membrane anchors, all RCAs are entirely or almost entirely composed of CCP modules and form a homologous protein family which could have arisen from gene duplication and exon shuffling.

Jules Bordet found as early as 1898 that exposure of serum to foreign erythrocytes leads to cell lysis, easily detectable through the liberation of haemoglobin. It is now known that *some* foreign erythrocytes act as AP activators (such as rabbit erythrocytes), while others are non-activators of the AP (such as red blood cells from sheep).

1.3 Complement Factor H

1.3.1 Biological significance and overall structure of Factor H

FH is the principal complement regulator in the fluid phase, but its significance extends beyond the fluid phase. Some host structures lack membrane-bound regulators, among them basement membranes and kidney glomeruli [51, 52]. For these cells, fH is the only known down-regulator of the AP, and, through the central role of C3b, of the late complement cascade. FH also seems to contribute significantly to the protection of at least some cell types that do possess membrane-bound regulators, as has been shown for human erythrocytes [53]. It is an abundant (100-500 µg/mL [54]), single chain plasma glycoprotein of about 155 kDa (1213 residues) which is - like most complement proteins - expressed mainly in hepatocytes [55]. Different opinions have been voiced with respect to the question if fH undergoes self-association. Transmission electron microscopy has led to the conclusion that fH is a monomer [46], while x-ray and neutron solution scattering suggested dimerisation in the same concentration range (1-10 mg/ml) [56]. It is possible that fH has a small tendency to multimerise and that this tendency increases in the presence of metal ions or polyanions, the latter of which which are thought to be an important ligand to fH [57, 58].

FH has nine potential N-glycosylation sites, eight of which exhibit glycosylation *in vivo* [59]. The physiological glycosylation sites are located in fH∼9, fH∼12, fH∼13, fH∼14, fH∼15 (twice), fH∼17 and fH∼18. Its protein moiety consists of a single polypeptide chain of 20 CCP modules, while the carbohydrate moiety is inhomogeneous and not required for fI co-factor activity [59, 60]. The 20 domains are connected through short linker regions of 3-8 residues and fH is often schematically depicted as 20 “beads on a string”. Transmission electron microscopy has shown that it is indeed elongated and flexible, albeit not necessarily linear. Instead, the structural flexibility of at least some of the linkers lead to a number of different conformations with differing cross sections [46]. Sedimentation velocity experiments along with x-ray and neutron solution scatter-

ing and molecular dynamics simulation led to the conclusion that full-length fH bends back upon itself [61]. Recently, it has been found in a NMR structural study that not all linker regions do confine flexibility. Long linkers in the middle region of fH might well be rigid and responsible for significant back-bending of the protein, possibly bringing binding sites for different interaction partners closer in space than a linear arrangement would permit [62].

A number of so called factor H-related proteins (fHR1 - fHR5) are encoded in close proximity of the fH encoding gene on 1q32. There is also a truncated splice variant of fH (factor H-like protein, fHL-1) which maintains fI co-factor activity [63, 64, 65].

1.3.2 Interaction partners of fH

Two fH binding sites for C3b have been reported and confirmed by different studies, located in fH∼1-4 and fH∼19-20 [66, 67, 68]. However, there is no consensus about a possible third binding site which has been proposed for fH∼6-10 [66] and alternatively near or within fH∼12-14 [67]. Only fH∼1-3 is essential for the fI-co-factor activity, though fH∼4 additionally contributes significantly to it [41]. The four N-terminal domains are therefore usually referred to as the regulatory domains. A C-terminal C3b/polyanion binding site in fH∼19-20 is necessary for the prevention of AP activation on host cells [69] and therefore has been said to contain the central host recognition element. Cell surface polyanions and sialic acid are known to play a central role in the recognition process. Enzymatic removal of the latter from cell surfaces averts fH binding and turns non-complement-activating sheep erythrocytes into complement activators [70]. Sialic acid residues often form the head group of glycans, extracellular carbohydrate chains that are attached to glycoproteins and gangliosides, which are both part of cell membranes. Even a slight chemical alteration of cell surface sialic acid residues results in loss of fH-mediated protection from the AP on sheep erythrocytes [71]. Synthetic liposomes that contain physiological glycolipids become subject to lysis through the AP

upon removal of sialic acid residues [72]. Sialation of the complement activator *Neisseria gonorrhoeae* results in C3 conversion to iC3b, signalling binding and co-factor activity of fH. Mutant studies located the responsible binding site for this interaction in fH∼16-20 [73]. Heparin, a close relative of heparan sulfate which is found on cell surfaces as carbohydrate moiety of proteoglycans (so-called heparan sulfate proteoglycans, HSPGs), can convert activators of the AP into non-activators when associated with the cell surface [74]. Several polyanions enhance the affinity of fH to activator-bound C3b, among them dextran sulfate, heparin and DNA. Interestingly, polysialic acid did not show this effect, provoking the hypothesis that the sialic acid head group might be part of a larger binding epitope which possibly contains further (negatively charged) carbohydrate residues [75]. Since heparan sulfate does naturally occur on cell surfaces and in the extracellular matrix it is among the possible ligands for fH, and heparin/HS are often employed as mimics for the yet unidentified *in vivo* ligand. The quest for the physiological ligand is complicated by the natural heterogeneity of cell surface polyanions and the fact that fH protects a variety of cell surfaces, leaving the possibility of several structurally diverse ligands instead of a single specific interaction.

Binding of fH to C3b-coated erythrocytes can be significantly inhibited by recombinant fH∼19-20 for both human and sheep cells, while the fluid-phase regulation of C3b is unaffected by the addition of fH∼19-20. It has been shown, however, that on a broader range of cells contributions to the recognition process are made by different regions of fH, and that the weight of each of these regions for a specific recognition process varies among different AP activators [76]. While for C3b-coated sheep erythrocytes fH∼16-20 contribute most to the binding affinity of fH, fH∼1-10 seem most important for C3b-coated human erythrocytes and fH∼6-10 as well as fH∼16-20 are central for C3b-coated rabbit erythrocytes [76]. Cultured endothelial cells have been shown to bind full-length fH and a truncated version of the protein encompassing fH∼ 15-20 with comparable strength [77], but generally C3b deposition on cells seems to be a requirement for fH binding to the cells [78].

A cluster of mutations in the two C-terminal modules has been found to be related to a rare but severe renal condition, atypical hemolytic uremic syndrome (aHUS), which might be caused by a change in the polyanion recognition abilities of fH∼19-20 [69]. Indeed, several disease-linked mutations are found in the heparin binding site in fH∼20 [79]. Several genetic studies suggest that another mutation in fH, Y402H in fH∼7, plays a role in the development of age-related macular degeneration (AMD) [80, 81, 82], the leading cause of blindness in the elderly Western population. Although many factors may contribute to this disease, including features of lifestyle such as smoking [83], the existence of heparin binding sites in these modules point to the possibility of polyanion involvement in the pathophysiology of the diseases. The connection of these conditions with mutations in different fH modules supports the hypothesis that individual sites in this multi-modular protein contribute differently to the recognition of various tissues, relying on mediation by cell specific polyanion patterns.

FI needs fH or another co-factor to bind to C3b (or C4b) in order to fulfill its enzymatic task of cleaving C3b into iC3b (at positions 1281 and 1299). The third cleavage, assisted by CR1, produces C3c (140 kDa) and C3dg (35 kDa). FI activity can be tested via the breakdown of fluid phase C3b in the presence of fH or another co-factor. After disulfide reduction two fragments of the α'-chain of iC3b can be observed, one of 68 kDa and one of 43 kDa size (Fig.1.3). This assay can also be employed to assess the fH co-factor activity [45]. FI cleaves a range of bonds in both C3b and C4b and daughter products and its specificity is relatively low. The amount and exact nature of all FI proteolysis products in the presence of the co-factors depends on salt concentration and pH and non-physiological co-factor activity of complement regulators has been described under favourable conditions [47].

C-reactive protein (CRP), an acute phase protein whose concentration in the blood plasma rises dramatically upon inflammation, has been reported to bind to fH∼7 and fH∼8-11 [85, 86]. This interaction has been much disputed: *in vivo* CRP adopts a

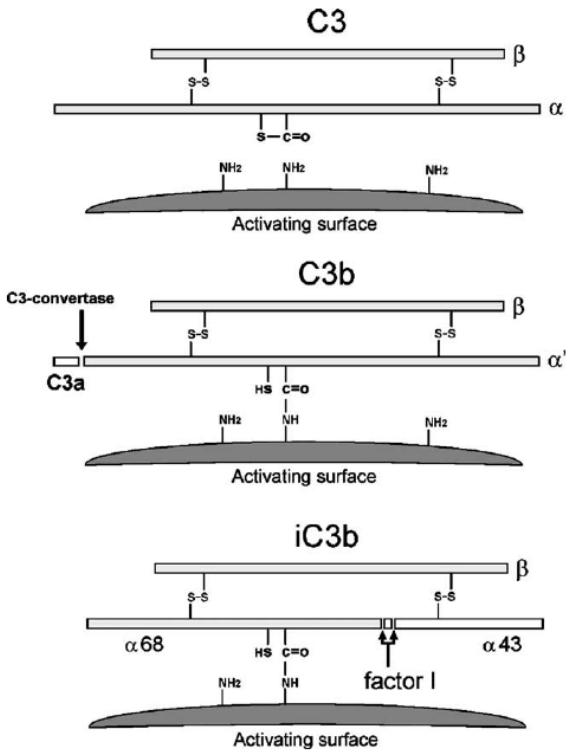

Figure 1.3: Schematic Figure of C3, C3b and iC3b with chains and cleavage positions of fI in the presence of fH or another co-factor. Taken from [84].

relatively fragile pentameric structure which is only stable in the presence of Ca^{2+}, and degradation or unfolding of this complex structure have been made responsible for the binding of fH [87]. To date, no study has proved an interaction of the two proteins *in vivo*.

1.3.3 FH-related disease

A number of diseases have been associated with deficiencies or genetic modifications in the fH gene and/or other complement proteins. Deficiency of fH can, through high turnover and consequently low C3 serum levels, lead to membranoproliferative glomerulonephritis (MPGN or dense deposit disease) and, in case of homozygous deficiency, atypical haemolytic-uraemic syndrome (aHUS) [88]. A cluster of mutations in the two C-terminal modules has also been found to be linked to this rare but severe renal condition which might be caused by a change in the polyanion recognition abilities of fH∼19-20 [69]. Indeed, several disease-linked mutations are found in the heparin binding site in fH∼20 [79]. Several genetic studies suggest that another mutation in fH, Y402H in fH∼7, plays a role in the development of age-related macular degeneration (AMD) [80, 81, 82], the leading cause of blindness in the elderly Western population. Although many factors may contribute to this disease, including features of lifestyle such as smoking [83], the existence of heparin binding sites in these modules point to the possibility of polyanion involvement in the pathophysiology of the diseases. The connection of these conditions with mutations in different fH modules supports the hypothesis that individual sites in this multi-modular protein contribute differently to the recognition of various tissues, relying on mediation by cell specific polyanion patterns.

1.3.4 Atypical haemolytic uraemic syndrome

AHUS is characterised by thrombocytopenia, renal impairment and microangiopathic haemolytic anaemia. A possible mechanism for aHUS assumes a badly controlled AP

which leads to excess deposition of C3b on renal cells and, simultaneously, to increased inflammatory cell recruitment through release of the anaphylotoxin C3a. MAC formation and non-lytic membrane perturbation as consequences of uncontrolled C3b deposition could be responsible for thrombosis in aHUS [89] and for a reduced flow across the glomerular membrane, which could be blocked with immune complexes. Mutations in fH which have been linked to aHUS cluster in the polyanion/C3b binding site in fH∼19-20 but do also occur in other regions of fH [90]. Some of these mutations simply lead to low expression levels of fH. Studies of aHUS associated mutations, along with NMR- [79] and x-ray [91] structures of the two C-terminal domains of fH, have come to different conclusions as to whether the C3b or polyanion binding site on fH∼19-20 is disturbed by the mutations [92, 93, 94, 95]. Very recently, it was shown that neither of both interaction partners alone is a marker for aHUS but that very likely the formation of a ternary complex is impaired [96]. It seems evident that a slight misregulation of the AP in some of the characterised aHUS mutants predisposes to aHUS. The hypothesis that slight misregulations of the complement system are at the heart of diseases such as aHUS and AMD is supported by the link of mutations in complement proteins other than fH to these diseases. In particular, gain-in-function mutations in Factor B which lead to increased stability of the C3bBb complex fit into this picture, underlining how increased stability of C3-convertase could result in similar consequences as decreased activity of fH or other complement regulators [97].

1.3.5 Age related macular degeneration

Age related macular degeneration (AMD), the leading cause of blindness in the elderly Western population, is strongly linked to a common SNP in fH's module 7, Y402H [80, 81, 82, 98, 99]. Unlike aHUS, whose first outbreak is acute, AMD is a chronic disease in which protein rich drusen is slowly accumulated between the basal lamina of the retinal pigment epithelium (RPE) and the innermost layer of Bruch's membrane, which forms the layer between the retina and the choroid. Characterisation of the de-

posit's composition led to the discovery of an accumulation of complement proteins (among other components) in the drusen and to the conclusion that AMD is histologically related to other extracellular deposit diseases, such as atherosclerosis, amyloidosis (among which are Alzheimer disease and Prion diseases) and dense deposit disease (also called Membranoproliferative glomerulonephritis, MPGN, affecting the kidney glomerulus) [100]. AMD is further subdivided into wet and dry forms of the disease. Wet AMD is associated with abnormal neovascularization, while dry AMD results in atrophy of RPE cells. The formation of drusen occurs in dry and wet AMD, both of which eventually lead to degeneration of the RPE and loss of macular vision. Among the complement components found in drusen are C3, C5, C5b-C9 (the components of the membrane attack complex), CRP and fH [100, 101, 102]. The link between the fH variant Y402H and the disease has prompted the hypothesis that an altered ability to recognize RPE cells in this variant could be responsible for the formation of drusen through diminished protection from the complement system. A binding study using both protein variants and primary RPE cells did not find different affinities for the two variants [103], while another one did [104]. The controversy around potential reasons for the fH-linked genetic predisposition for AMD generally overlaps with the controversy concerning the question whether CRP and fH interact *in vivo* [103] since both proteins have been detected in the deposits. The hypothesis that an altered recognition process is at the centre of AMD is supported by age-related changes in the retinal cell surface charges: an altered polyanion pattern on these surfaces could trigger the disease through impaired protection by factor H. Diabetis and smoking have been identified as risk factors for AMD, both of which are associated with increased formation of so-called advanced glycation endproducts (AGEs). The accumulation of AGEs in the retina has been linked to both diabetic retinopathy and AMD and is a process that is generally enhanced in ageing eyes [105, 106, 107].

1.3.6 Complement evaders

It is noteworthy that while fH aims to prevent host cells from attacks by the complement system, some pathogens have developed ways to acquire fH and other complement regulatory proteins in order to overcome the innate immune defence by "disguising" as host cells [108]. Among these complement evaders is *Neisseria menigitidis*, the bacteria responsible for meningitis, as well as *Neisseria gonorrhoeae* [109].

1.4 Human β-defensin HBD2

Defensins are small (3.5-6 kDa) antimicrobial peptides which are rich in cysteine and the cationic residues arginine and lysine. They act as bactericidal agents against Gram-positive and -negative bacteria, fungi and some enveloped viruses, which they kill through the formation of channels in lipid bilayers or disruption of these structures. They are also cytotoxic and thus usually stored in granules, to be released only into the vacuoles of phagocytic cells, so that no harm is done to the host organism [110]. The destabilization of lipid bilayers might be conferred by the presence of aliphatic patches on their surfaces, allowing the interaction with hydrophobic structures such as lipids. Among their targets are *Escherichia coli*, *Staphylococcus aureus* (human beta defensin 3) *Candida* and the *Herpes simplex* virus [111]. Just like many proteins of the Complement AP, defensins are among the effector molecules which are mobilised or activated by the innate immune system in response to a breech of the epithelial barrier by pathogenic microorganisms. Defensins also act as chemokines and thus also play a role as mediators of the host defence. Defensins contain a large proportion of cysteine residues and are classified according to their disulfide pattern. In β-defensins, the disulfide scaffold follows the pattern cys^1-cys^5, cys^2-cys^4, cys^3-cys^6 (in α-defensins, cys^6 and cys^5 are swapped). All human β-defensins that have been structurally analysed so far - that is, HBD1-3 - display pronounced positively charged patches. HBD2, in addition, bears a large hydrophobic patch on the face opposite the positive patch (Fig 1.4).

HBD2 is produced and released by kerotinocytes following stimulation through contact with the lipopolysaccharides on the surface of Gram-negative bacteria or through cytokines such as interleukin (IL)-1β and tumor necrosis factor (TNF)-α. It possesses antimicrobial activity against Gram-negative bacteria and the yeast *Candida* but not the Gram-positive *S. aureus*, which is antagonised by HBD3 [113]. The permanent expression level of HBD-2 is low, but pro-inflammatory induction leads to a large increase in expression. The main tissue types which produce HBD-2 are skin, lung and trachea [113].

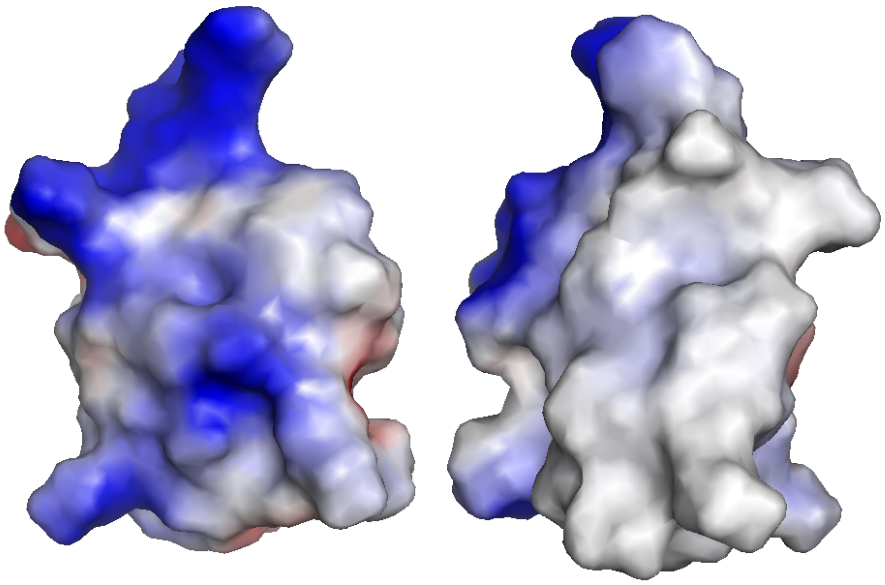

Figure 1.4: Electropositive and aliphatic faces of monomeric solution structure of HBD2, calculated in PyMOL. Red and blue refer to negative and positive electropotentials, respectively (PDB entry 1FD3 [112]).

Due to their relative recent discovery, the amount of data on the biochemical nature and mode of action of β-defensins (and defensins in general) in the literature is still limited. However, one x-ray and two NMR structures of HBD2 have been published [112, 114, 115]. Uniformly, the peptide has been described as containing a three-stranded β-sheet which is stabilized by two disulfide bonds. Additionally, there is a small α-helix which passes opposite the sheet and forms a disulfide bond with strand β2 (Fig. 1.5). In the x-ray crystals, HBD2 forms dimers which - through interactions of the two β1 strands - give rise to a six-stranded β-sheet [112]. Four such dimers are arranged in a symmetric octameric assembly, with no evidence of formation of a pore between the monomers (Fig. 1.6). The mode of action by which HBD2 interrupts cellular integrity of invaders is thus, according to the crystallographers, not to be described by membrane pore formation, a mechanism that had been proposed elsewhere for the anti-microbiol activity of defensins [116]. Instead, it was postulated that a uniformly positively charged site of the octamer disrupts membranes via electrostatic interactions with the polar heads of the membrane-

forming lipids. Both NMR structures describe HBD2 as a monomer in solution, assuming the multimeric assembly found in the x-ray structure was mainly reflective of high protein concentrations, but not entirely dismissing the proposed mechanism for possible high local concentrations [115]. The absence of a hydrophobic core in HBD2 was noted [115], stressing the importance of the conserved disulfides for the overall fold.

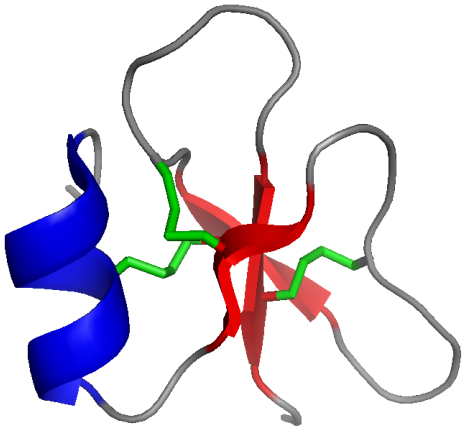

Figure 1.5: Secondary structure elements and disulfide bridges in HBD2: α-helix (blue) from residues 4-11, β-strands (red) from residues 14-16, 25-28 and 36-39, and three conserved disulfide bridges (green). PDB entry 1FD3 [112].

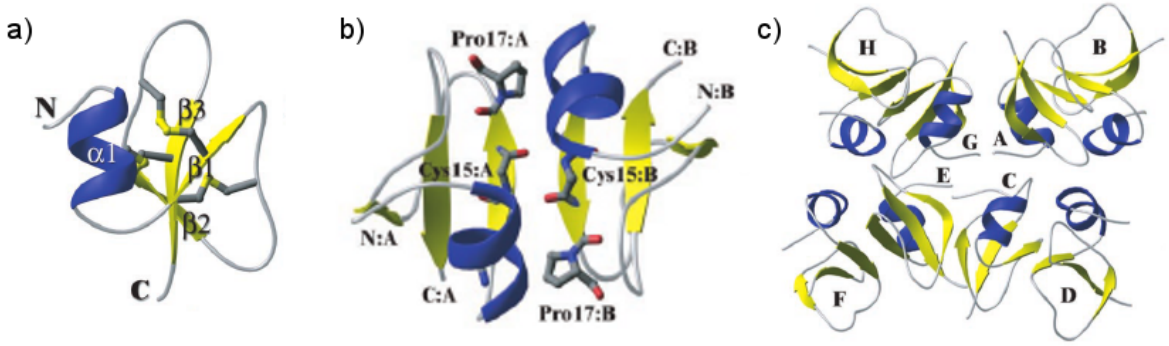

Figure 1.6: HBD2 a) monomer, b) dimer and c) octamer as found in the protein crystal. Taken from [117].

1.5 Hepatocyte growth factor / scatter factor

Growth factors are proteins which stimulate cell proliferation and differentiation and are therefore critical during foetal development and for repair mechanisms following injury. The hepatocyte growth factor / scatter factor (HGF/SF) was simultaneously discovered in two different functions: as a mitogenic factor in hepatocytes and as a scattering factor in epithelial cells. HGF activates the receptor tyrosine kinase mesenchymal-epithelial transition factor (MET, also called HGF-receptor) [118], which in turn initiates different signal transduction pathways, all of which are related to developmental processes. The mechanism by which HGF activates MET involves dimerisation of the membrane-bound receptor upon binding of HGF to the extracellular portion. This dimerisation activates MET for subsequent crosswise autophosphorylation of two tyrosine residues, Y1234 and Y1235 [119]. Misregulation of the proto-oncogene *met*, which encodes the transmembrane receptor, leads to tumorigenesis and metastasis [120]. Consequently, the HGF-MET interaction is a target for cancer research [121].

The HGF precursor is a single polypeptide chain which is cleaved by HGF-activator (HGFA) at position R494 to yield the so-called heavy chain (ca. 60 kDa) and light (ca. 34 kDa) chain. Both chains remain attached to each other through a disulfide bond [122]. The heavy chain is comprised of an N-terminal domain (N) and 4 kringle domains (K1-K4), each containing three conserved disulfide bonds. The light chain is structurally classified as a serine protease-like domain (and thus abbreviated SP-domain) but it does not show actual protease activity (Fig. 1.7).

The N-domain displays a largely positively charged surface as often seen for heparin binding domains, and its interaction with heparin has been characterised by chemical shift mapping and NMR dynamics measurements [124, 125] as well as by x-ray crystallography [126]. Heparin rigidifies the otherwise relatively loose N-domain as indicated by relaxation time measurements, additionally, the glycosaminoglycan increased the stability of the N-domain against disulfide bond reduction with dithiothreitol (DTT) [125]. The N-terminus of NK1 gave only weak electron density in a x-ray crystallography study

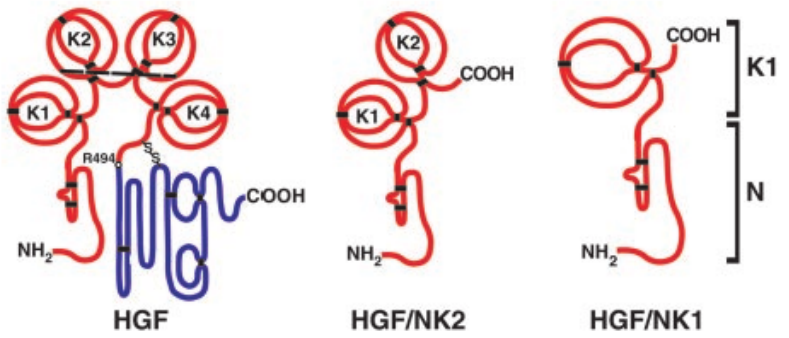

Figure 1.7: Schematic overview of naturally occurring isoforms of HGF. In full-length HGF, the heavy chain (red, comprised of the N-terminal and the four kringle domains) and the light chain (blue, containing the SP domain) stay attached through a disulfide bond after proteolytic cleavage at R494, while NK2 and NK1 are only comprised of heavy chain modules. Taken from [123].

of the free protein [127] and no NH crosspeaks [124] in the isolated N-domain. A mutagenesis study suggested that these residues are important for heparin binding [128], yet most of the residues which were identified in this study are not part of the heparin binding site as described in the x-ray study of the heparin-NK1 complex.

Heparin, heparan sulfate and dermatan sulfate have been shown to act as co-factors for the HGF-MET interaction [129, 130, 131, 132, 133]. However, the question of whether glycosaminoglycans (GAGs) are actually necessary for activation of MET by HGF is somewhat controversial [134, 135, 136, 137]. It has been reported that MET, just like HGF, has a high affinity for HS and that, in the presence of HS, MET activation can take place even in the absence of full-length HGF but in the presence of its first kringle domain, K1 [123]. The GAG-binding site in HGF, although mainly contained in the N-terminal domain, extends into the first of the four kringle domains, together termed NK1 [138, 125]. A smaller contribution to GAG binding is made by the kringle domain 2, and the protein comprised of these three domains was termed NK2 [139]. NK1 and NK2 are both splice variants of HGF [140]. Their physiological role is not precisely

known. NK1 has been characterised both as an HGF antagonist which competes with HGF for binding of the MET receptor but does not trigger its autophosphorylation and downstream signalling cascade [141] as well as a "partial agonist" which stimulates cell proliferation, migration and branching morphogenesis through activation of MET [140, 142, 143]. GAG binding to NK1 has been shown to be necessary at least for some of HGF's physiological functions that can be assumed by NK1 [134]. The crystal study of the heparin-NK1 complex (using a heparin 14mer) reported two different types of crystals but many conserved heparin-GAG interactions in all protomers present in these two crystal types [126]. A network made of hydrogen bonds and ionic interactions is observed in the N-domains of both crystals and involves residues K60 (sidechain), T61 (backbone), K63 (backbone) and R73 (sidechain). In one crystal form, K58 (sidechain) and K62 (sidechain) provide additional hydrogen bonds. As often observed in protein-GAG complexes the binding site was located on the surface of the protein and a relatively short stretch of four pyranose residues was found to provide the key interactions in the N-domain in both crystal types (Fig. 1.8). The longest heparin oligosaccharide which was found to have a well-defined electron density in this study does additionally maintain contacts to the K-domain of a different protomer.

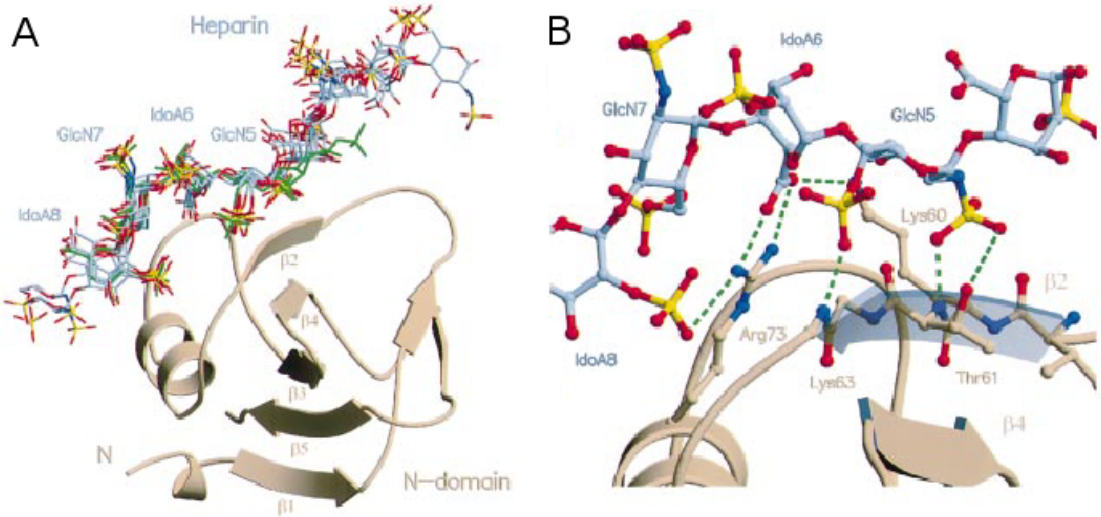

Figure 1.8: Details from a NK1:heparin X-ray structure: A) N-domain with heparin binding site B) enlargement of the binding site. Taken from [126].

1.6 Glycosaminoglycans

Glycosaminoglycans (GAGs) are linear polysaccharides of widely varying molecular weights which are present on all animal cell surfaces and in the extracellular matrix. GAGs are usually linked to protein side chains, forming the macromolecular superfamily of proteoglycans. They are classified according to repeating disaccharide units which make up their primary structure and which can be sulfated or acetylated to different degrees. A common classification of GAGs uses four different groups: the hyaluronic acid type, the heparin/heparan sulfate (HS) type, the chondroitin/dermatan sulfate (DS) type and the keratan sulfate type (Fig. 1.9). This work focuses mostly on heparin/HS and dermatan sulfate.

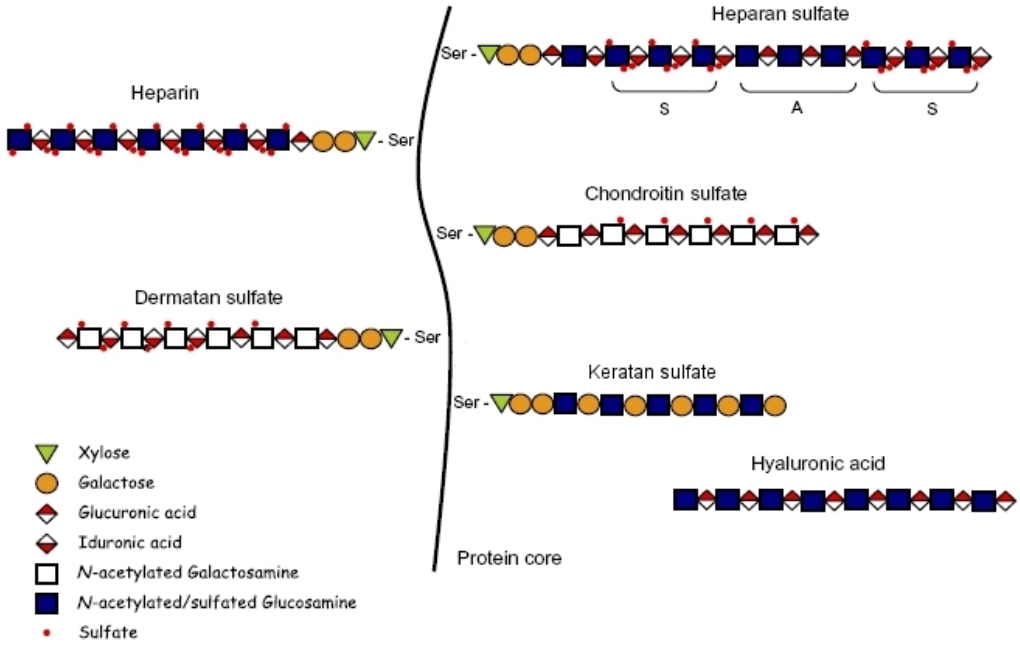

Figure 1.9: Schematic overview of the different classes of GAGs. S and A labels in heparan sulfate denote sulfated and non-sulfated domains, respectively. Taken from [144].

1.6.1 Heparin and Heparan sulfate

Heparin is different from other GAGs - including HS - because it is synthesized as proteoglycan but subsequently cleaved from the protein core to give a heterogeneous mixture, comprising linear chains of molecular weights between 5,000 and 25,0000 Da [145]. It is then stored in secretory granules of mast cells where it was expressed until the granules are released upon external stimuli such as IgE-antigen complexes or complement anaphylotoxins [146, 147]. Heparin is best known for its anti-inflammatory and anticoagulating activities. HS, on the other hand, is expressed in a variety of different cells. It is, however, structurally closely related to heparin and made from the same building blocks. Both species share identical glycosidic linkages. Disaccharides of the heparin/HS type are comprised of glucosamine and hexuronic acid residues, both of which can be sulfated in different positions and, additionally, acetylated on the amino group of the glucosamine. One structural difference between heparin and HS lies in the overall degree of sulfation (Fig. 1.9): heparin exhibits a relatively homogeneous and high level of sulfation, while in HS sulfated domains (so called NS-domains) alternate with non-sulfated domains (NA-domains), linked by transition domains with an intermediate degree of sulfation. The maximum length of one NS-domain is around 8-9 disaccharides [148, 149]. Small oligosaccharides from HS-NS-domains are identical to those isolated from heparin, while less sulfated species can only be obtain from HS or from heparin after subsequent chemical desulfation. Besides the higher degree of sulfation in heparin/HS-NS-domains these structures differ from HS-NA-domains by the relative amounts of glucuronic acid (GlcA) versus iduronic acid (IdoA), the latter of which is more abundant in heparin/HS-NS-domains.

In vivo, HS is synthezised as linear chains of alternating α-D-GlcNAc-(1$\rightarrow$4) and β-D-GlcA-(1$\rightarrow$4). Subsequently, both residues are modified by a set of specific enzymes which catalyse deacetylation, sulfation and epimerisation (of D-GlcA to L-IdoA). However, these enzymes do not *quantitatively* modify the heparin/HS raw structure, eventually leading to a complex mixture of slightly differing stretches. Some of the process-

ing steps are inter-dependent, i.e. deacetylation/sulfation of GlcNAc residues (yielding GlcNS residues) allows epimerisation of the GlcA residues at the non-reducing end of the GlcNS residues, giving rise to IdoA-GlcNS disaccharides. This stretch can further be subject to O-sulfation on the C2 position in IdoA and the C-6 position in GlcNS [150]. O-3 sulfation of the GlcN residue is less frequent but essential for the HS-antithrombin interaction. Antithrombin, a serine protease, acts as an inhibitor of the coagulation cascade. Through formation of a complex with HS chains of proteoglycans on the lumen of vascular endothelial cells it is brought into proximity of its substrate thrombin, which is generated at the blood-surface interface and which also binds to HS [151]. In heparin and HS-NS-domains the most abundant disaccharide is α-L-IdoA2S-(1→4)-α-D-GlcNS6S (Fig. 1.10). GlcA residues occur prevalently instead of IdoA residues in the HS-NA-domains, leading to a smaller degree of overall sulfation as GlcA is less easily sulfated than IdoA by the 2-*O*-sulfating enzyme of the heparin building pathway, urosonyl-2-*O*-sulfotransferase [152]. The glycosidic linkages in the heparin/HS group of GAGs are 1→4 and the sugar rings adopt α, α(GlcN, IdoA)- or α, β(GlcN, GlcA)-configuration, though an equilibrium between α- and β-configurations occurs in the reducing end residue in isolated oligosaccharides. This is heavily shifted towards the α-form.

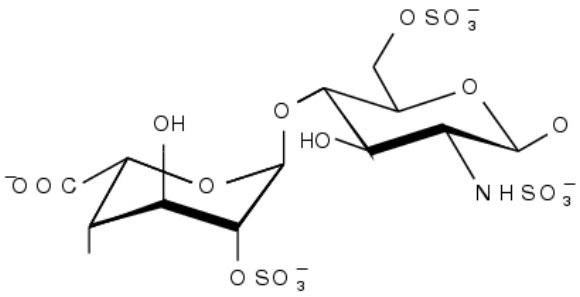

Figure 1.10: The most abundant disaccharide in heparin, α-L-IdoA2S-(1→4)-α-D-GlcNS6S.

HS is a regulator involved in a variety of processes, ranging from blood coagulation to

angiogenesis. In HS, the sulfation density of the 2-position of IdoA and of the 6-position as well as the amino group of GlcN vary depending on tissue type and organism and are determined by the originating cell type and organism rather than the protein which carries the HS chain [145, 153]. Taken together with the rare but biologically important sulfation of the 3-position in GlcN and the alternation of GlcA and IdoA residues, the incomplete sulfation of the IdoA and GlcN residues leads to an outstanding feature of heparin/HS and other GAGs: pronounced chemical heterogeneity. The disaccharide composition of HSPGs in different tissues has been identified as a possible bio-marker, especially in the field of oncology [154, 155, 156]. Since the HS-fingerprint also changes with age it has implications for age-related diseases [157, 158]. The great variety in composition and sulfation levels makes it difficult to investigate GAG-protein interactions in detail. The question of the exact GAG epitope for a specific interaction is not trivial and gains in complexity with increasing minimum GAG binding length - a heparin/HS disaccharide can accommodate 23 different primary structures out of 48 theoretically possible structures, some of which do not occur *in vivo* due to enzymatic restrictions in the heparin/HS pathway [152]. Heparin obtained from different sources, if not further processed and thoroughly characterised, may well lead to different results in a series of analogous experiments due to different prevalences of sulfation composition in the individual preparations.

The inherent flexibility of the glycosidic linkage and the coexistence of different conformations of the IdoA residues lead to large potential flexibility in the structure of heparin. While GlcNS and GlcA adopt the usual $^4C_1$ chair conformation, internal IdoA residues in heparin, HS and DS exist in an equilibrium between the $^1C_4$ chair and the $^2S_0$ skew-boat conformations in the liquid state. In other scenarios, such as non-reducing terminal residues, in monosaccharides or synthetic oligosaccharides, IdoA can also adopt the $^4C_1$ chair conformation [159] (Fig. 1.11). The main conformation in a given polysaccharide is influenced by other structural factors such as the preceding residue and the positions of sulfate groups [160]. Coordination of counterions could also be an important factor as it

requires a specific orientation of the ligating groups - copper, for example, is coordinated by the IdoA residues in heparin and requires the planar orientation of the coordinating groups that is often observed in copper complexes [161].

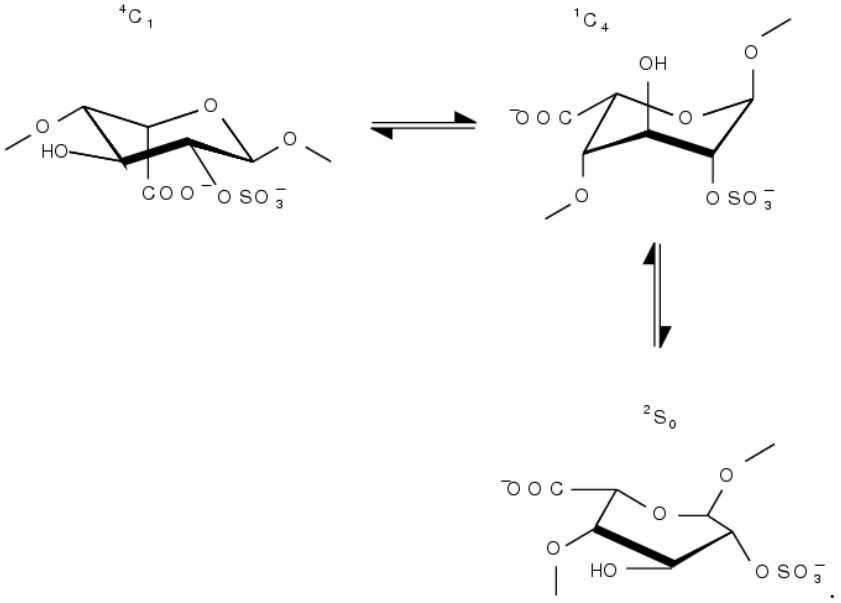

Figure 1.11: Conformations of IdoA. Internal IdoA residues in heparin, HS and DS inter convert between the $^1C_4$ chair and the $^2S_0$ skew-boat. In other contexts, the $^4C_1$ chair has also been observed.

Complexes with some proteins have been found to favour one specific conformation of the IdoA residues - the $^2S_0$ skew-boat, for example, in the antithrombin III - heparin pentasaccharide complex [162], the $^1C_4$ chair in the histamine-heparin tetrasaccharide complex [163] - while other proteins do not seem to lock the IdoA ring in any specific conformation [164]. Both heparin with all IdoA residues in the $^2S_0$ conformation and heparin with all IdoA residues in the $^1C_4$ conformation result in overall similar 3D structures, as elucidated by NMR and molecular modelling [165]. In this study, two dodecasaccharides were generated based on NMR NOE data and molecular modelling. Both forms adopt an α-helical shape with a pitch of ca. 17 Å, corresponding to ca. 4 residues per turn. This helical 3D structure has also been found by a recent ^{1}H-^{1}H NOE

and molecular dynamics study on ^{13}C labelled heparin and its precursors. The transition from N-acetylheparan to heparin was accompanied by rigidification of the helical chain [166]. It is noteworthy that the exact orientation of the sulfate groups with regard to the helical axis changes when the IdoA residues undergo conformational changes despite the preservation of the helical shape. Though the interconversion between the $^1C_4$ and $^2S_0$ ring forms provoke relatively small changes in the geometry of the glycosidic linkages to neighbouring rings - hence the preservation of the helical shape - the spatial orientation of sulfate groups in the IdoA residues experiences marked changes [167]. The investigation of a fully sulfated tetrasaccharide via NMR-measurements of residual dipolar couplings also concluded that the flexibility of the IdoA residues had a limited effect on the molecular shape of heparin [168]. The differences in the orientation of the sulfate groups depending on the IdoA conformation might explain the preference of some proteins to bind to specific conformations of the IdoA residues. Additional modifications in the spacing and exact orientation of the sulfate groups can be accommodated by changes in the glycosidic dihedral angles, leading to an improved fit of heparin into binding sites of different proteins. The resulting oligosaccharide structures can be described as containing kinks or helical overwinds when compared to the structure of free heparin [169].

During this project small oligosaccharides were used which were obtained from heparin via enzymatic digestion with heparinase I. Heparinase III, which cleaves HS instead of heparin, is less robust and therefore less suited for *in vitro* work. Because sulfated and non-sulfated domains alternate in HS but not in the more heavily sulfated heparin (see Fig. 1.9), enzymatic digestion of HS leads to a larger variety of oligosaccharides. In the binding experiments conducted as part of this project, however, only fully sulfated oligosaccharides were used as ligands (with the exception of the fH∼19-20 titration with size-fractionated octasaccharide). These species can be obtained from both HS and heparin, but yields are higher when using heparin. Since heparin and HS protein interactions are very often dominated by charge-charge interactions [170, 171, 172] the question of the raison d'être of non-sulfated domains in HS is at hand. It is possible

that the reason for their existence is of more general nature than the specific interactions with certain proteins. In their physiological environment, HS chains are attached to protein cores of proteoglycans such as syndecans or collagen XVIII. Especially in an extracellular environment proteoglycans often contribute to the physical properties of their environment, conferring strength, elasticity or, in the case of hyaluronic acid in the joints, viscosity. These physical properties are likely connected to the overall charge and charge distribution of proteoglyan sidechains, for example via their degree of hydration. Larger flexibility of non-sulfated regions in HS [166] may well be required for proper positioning of sulfated domains near membrane bound receptors. Heparin, on the other hand, is confined to intracellular vascules in mast cells and not found in the extra cellular matrix. Being the physiological molecule with the largest negative charge density it might be used to neutralize positively charged biomolecules (i.e. histamine and proteases) stored in secretory granules [173]. In the light of the different physiological distribution of heparin and HS it is possible that the dual function of HS - binding of specific proteins as well as contributing to the physical properties of extracellular environments - is represented in its irregular domain-based structure.

The physiological HS synthesis, unlike DNA and protein production, is a non-template driven process. The exact nature of a HS molecule is a consequence of different enzymatic modifications and is therefore, through expression and activity levels of the modifying enzymes, linked to regulation processes which take place on a different cellular level, making the field of "heperonomics" a subject matter for system biologists [174].

1.6.2 Dermatan sulfate

Dermatan sulfate (DS), originally termed chondroitin sulfate B, contains *N*-acetyl-galactosamine residues instead of *N*-glucosamine, with β-D-GalNAc-(1→4) and α-L-IdoA-(1→3) and α-L-IdoA-(1→4) glycosidic linkages. Unlike HS, DS exhibits relatively uniform 4-*O*-sulfation of the GalNAc residues, while 2-*O*-sulfation of IdoA is less frequent than in HS. Additionally, 6-*O*-sulfation can occur in the GalNAc residues. The most

abundant disaccharide unit in DS is α-L-IdoA-(1→3)-β-D-GalNAc4S (Fig. 1.12) leading to a smaller overall sulfation density in DS when compared to heparin. Like HS, DS occurs in the ECM and promotes angiogenesis and other developmental processes as well as wound healing through its interaction with various growth factors [132, 175, 176]. DS interacts - albeit comparably weakly - with many HS/heparin-binding proteins [177, 178, 179], suggesting an overall functional overlap for both GAGs.

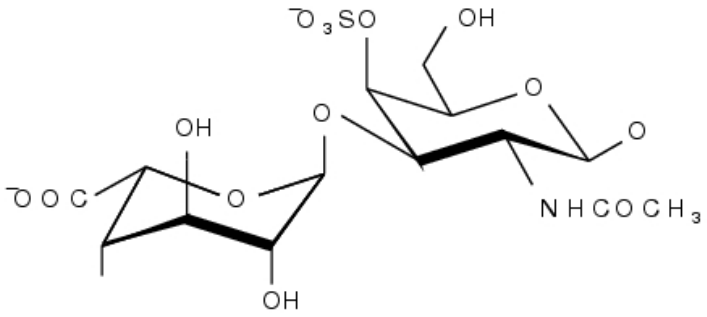

Figure 1.12: The most abundant disaccharide in dermatan sulfate, α-L-IdoA-(1→3)-β-D-GalNAc4S.

1.7 Protein-heparin/HS complexes

HS-protein interactions play a key role in a wide range of biological contexts, many of which are developmental processes. Among the interacting proteins are promoters of angiogenesis (fibroblast growth factors), cell adhesion (fibronectin) and cell growth (hepatocyte growth factor/scatter factor), regulators of cellular communication (interferons) as well as different enzymes of the blood coagulation cascade. HS chains make up a part of the extracellular matrix where they fulfil scaffolding purposes. Many of the HS/protein interactions are transient, serving the purpose of recruiting proteins into proximity of other proteins or cell surfaces where they subsequently execute their enzymatic functions. As a consequence, many HS-binding proteins lack a pronounced binding groove and do not undergo large conformational rearrangements upon GAG binding. This particular characteristic, along with the structural flexibility of the HS-chains, make HS-protein interactions a challenging target for docking studies.

One of the best-studied protein-heparin interactions is the complex formed by antithrombin and a heparin or HS pentasaccharide. Antithrombin is a serine protease inhibitor which inactivates several enzymes of the coagulation cascade, among them thrombin (hence its name) and Factor Xa. The interacting HS chains line the inner walls of the micro vascular system. The interaction of antithrombin with thrombin (Factor IIa) and Factor Xa is greatly enhanced in the presence of heparin, albeit to different degrees. It has been shown that antithrombin is activated by a specific pentasaccharide motif with the sequence α-D-GlcNAc/NS(6S)-(1→4)-β-D-GlcA-(1→4)-α-D-GlcNS(3S,6S)-(1→4)-α-L-IdoA(2S)-(1→4)-α-D-GlcNS(6S), with the rare 3-*O*-sulfation being of special importance [180]. Binding to this oligosaccharide induces a conformational change in antithrombin's active site [181, 182]. Factor Xa does itself not bind to HS, but for the inactivation of thrombin by antithrombin HS chains act as a scaffold for both proteins, bringing them close enough in space for the enzymatic deactivation of thrombin. The structural requirements of antithrombin and thrombin with respect to heparin/HS

differ: while for antithrombin the above mentioned pentasaccharide binds with much higher affinity than other heparin-derived oligosccharides, thrombin interacts with heparin hexasaccharides without a pronounced preference for any primary structure [152]. In the tertiary complex both binding sites on the GAG chain need to be separated by an octasaccharide, resulting in a minimum length of the binding HS oligosaccharide of nineteen sugar units. Thus, although thrombin itself does not require a specific binding motif there is a necessity for two protein binding sites in a defined distance on the HS chain. Because one of these protein binding sites has specific requirements (the one for antithrombin) the overall frequency of this arrangement is rare, imposing more specificity on the thrombin-deactivation than its own binding requirements for HS chains suggest.

Another role of HS outside the coagulation cascade which has been the subject of much attention is its interaction with growth factors, in particular the fibroblast growth factor (FGF) family and, to a lesser extent, hepatocyte growth factor/scatter factor (HGF/SF). The FGF family contains several isoforms, two of which have been crystallized in the presence of heparin oligosaccharides (FGF-1 [183] and FGF-2 [184]). Additionally, structures of these two proteins in complex with their respective receptor (FGFR2 and FGFR1) and heparin decasaccharides exist [185, 186]. Interestingly, two different ways of preparation were used for these two ternary complexes, yielding strikingly different geometries (for a comparison see [187]). It remains to be seen if these differences are due to preparative differences or if FGF-heparin-FGFR complexes really lack a common architecture. For both the FGF and HGF growth factor-receptor interactions heparin as a third component increases the tendency to form the otherwise low-affinity protein-protein complexes in the required stoichiometry. In the case of FGF heparin cross links growth factors and receptors in a 2:2:1 (protein:protein:heparin) stoichiometry. For the HGF-Met interaction, only the growth factor contains a heparin binding domain (NK1, see above). However, because heparin increases the tendency of NK1/HGF to dimerise it likely also promotes the formation of the signalling complex with the HGF-receptor: for efficient signalling, two receptor molecules in close proximity might have to be activated

by binding to the HGF-dimer. It has been suggested that the necessity of a third interaction partner (heparin) for both the HGF and FGF signalling complex helps overcome the otherwise low protein-protein affinities, which in turn guarantees a low noise level in the respective signalling pathways [188]. In the case of the HGF/SF interaction with heparin, charge density rather than a specific binding epitope seems to be the dominant force [189, 172].

Another field in which charge-charge interactions are at the centre of GAG-protein interactions are chemokines. These small (8-10 kDa) peptides are, among other processes, involved in the recruitment of leukocytes from the vasculature into surrounding tissue. The chemokines‘ role in this process consists in the adhesion of leukocytes on the endothelial surface by means of interaction with G-protein-coupled receptors which are expressed on the leukocyte surface. The chemokine gradient which directs this leukocyte recruitment does in turn depend on a further interaction partner which protects the gradient against shear forces and diffusion, and there is evidence suggesting that this function is carried out by GAGs on the endothelial cell surfaces [190]. Mutagenesis studies for a range of chemokines - among them MCP-1, RANTES and IL-8 - have found that basic residues such as arginine, lysine and, to a lesser extent, histidine residues are most important for GAG-binding (for an overview see [191]). The majority of GAG-binding chemokines posses pIs in the basic range, pointing to a high percentage of positively charged residues. There are, however, also chemokines with acidic pIs which have nevertheless been implicated in GAG binding [190]. One of these acidic chemokines, MCP-I (pI $\sim$4.5), has been shown to form tetramers in the presence of a heparin oligosaccharide, thereby bringing the GAG-binding residues close in space and forming one continuous basic patch on the surface of the tetramer despite its acidic pI [192].

In summary, in the field of GAG-protein interactions the lack of specificity for some interactions seems to co-exists with strict structural requirements in other complexes, making it difficult to combine the different types of interactions in a unifying model [152].

It should be noted that HS also plays a role in infections as an "anchor" for invading organisms [193]. Recently, it has been shown that sporozoites of the malaria-parasite *Plasmodium berghei* migrate through cells expressing low-sulfated HSPGs in the skin and endothelium before the highly sulfated HSPGs of hepatocytes activate the microorganism for invasion [194]. Similarly, the hepatitis B virus (HBV) attaches to hepatocyte-HS chains in the initial phase of its entry process into liver cells [195]. During the internalization of the transactivator protein of the human HI-virus, Tat, HSPGs might also act as receptor molecules [196].

Methods to determine binding epitopes and affinities usually employ enzymatic or chemical digestion and subsequent fractionation of heparin/HS according to length and chemical composition/charge in an attempt to isolate homogeneous species from the complex starting material. Following the separation of these smaller oligosaccharides different biophysical methods can be employed to determine binding affinities and binding sites of the well-defined oligosaccharides, among them ITC, SPR and NMR titrations. Other methods do not attempt separation of the oligosaccharides prior to binding studies but use the differential affinities to "pull out" the tightest binder from a mixture of oligosaccharides by means of chromatography or mass spectrometry [197]. The disadvantage of the first approach is the loss of rarely occurring oligosaccharide sequences during the separation/purification process.

1.7.1 Preparation of heparin-derived oligosaccharides

In a medical context fractions of low molecular weight heparins (4-6 kDa) are often used as anticoagulants. This fractionating does, however, not lead to sufficient homogeneity for the investigation of protein-heparin interactions by NMR, especially if a distinct binding epitope is of interest. A combination of enzymatic digestion, size fractionation and

ion-exchange chromatography can be used in order to obtain species of defined length, composition and sulfation from heparin sources. The enzyme of choice is heparinase I. Digests of heparin from different sources with heparinase I lead to a characteristic pattern of distinct oligosaccharide species [198], reflecting the different compositions of individual heparin types. The enzymatic cleavage performed by heparinase I is a β-elimination which produces an unsaturated uronic acid residue (ΔUA2S) at the non-reducing end of the oligosaccharides [199] (Fig. 1.13).

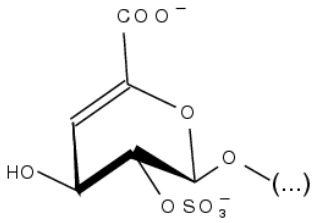

Figure 1.13: ΔUA2S forms the non-reducing end of all heparinase I-produced oligosaccharides, giving rise to an absorbance at 230 nm.

The double bond in conjugation with the uronic acid carboxyl group absorbs in the UV region with a maximum near 230 nm and can be used for detecting heparin-derived oligosaccharides and for concentration measurements. The extinction coefficient at 230 nm (ϵ_{230}) is almost unaffected by the presence of other functional groups found in heparin [198], allowing the usage of the same coefficient for all heparinase derived oligosaccharides. Following heparinase digestion, oligosaccharides can be size-fractionated using size exclusion chromatography. Further purification by anion exchange allows separation of chemically well-defined species with uniform sulfation and GlcA/IdoA content [172].

1.8 Biomolecular NMR spectroscopy

Nuclear magnetic resonance spectroscopy (NMR) makes use of the microscopic magnetic moments of atomic nuclei in order to gain information about the structure and dynamics of molecules. Not all chemical elements and not all isotopes of one given element have the same - or indeed any - magnetic moment. Some nuclei are not magnetic because this quality is coupled to the nuclear spin, which is zero for some elements/isotopes. Nuclei with a spin quantum number of 1/2 are most commonly used as magnetic probes in biomolecules because their magnetic moments are conveniently described as dipoles, while spin quantum numbers $> 1/2$ are technically demanding because for many such nuclei electric quadrupole moments have to be taken into account additionally to the magnetic moments. Coincidently, three of the most abundant nuclei in biomolecules (proton, carbon and nitrogen nuclei) have isotopes with spin 1/2, these are ^{1}H, ^{13}C and ^{15}N.

Like other types of spectroscopy, NMR spectroscopy creates non-equilibrium situations in a sample which absorbs electromagnetic radiation. In NMR, the basis of this absorption is the interaction of the small nuclei magnetic moments with the electromagnetic radiation from an external source. In thermal equilibrium, the sample is under the influence of a large (up to around 20 T) static magnetic field, which causes the nuclei magnetic moments to align parallel or antiparallel to the external field (depending on the sign of their gyromagnetic ratios) according to the Boltzmann distribution. At room temperature, this leads to a very small population difference between the two energy levels of spin 1/2 nuclei. The population difference is very small because the energy levels of the nuclear magnetic moments are close to each other: at room temperature, the energy differences between parallel and antiparallel orientations are in the range of kT even at large field strengths, due to the small magnetic moments of even the most sensitive nuclei. Whether the alignment parallel or antiparallel to the external field is energetically more favourable depends on the sign of the gyromagnetic ratio of the nucleus, the pro-

portionality constant that links magnetic moment to spin. A perturbation of the thermal equilibrium state is achieved through additional short-lived radiofrequency fields which are applied perpendicular to the external field, permitting the magnetic moments to absorb energy at discrete frequencies in order to undergo energetic transitions. Generally, one distinguishes between experimental setups in which this resonance process creates magnetisation perpendicular to the static field (so-called transverse magnetisation) used in Fourier transform (FT) NMR or (anti)parallel to it (longitudinal magnetisation) used in continuous wave (CW) NMR. FT NMR is superior to CW NMR and only the former is described here. Following the perturbing radiofrequency pulse in FT NMR experiments the system starts to relax back to its equilibrium state while precessing around the external magnetic field. Electromagnetic coils in close proximity to the NMR sample record this process. The precessing magnetic spins induce small currents in the detection coil which is oriented perpendicular to the static magnetic field.

The decay of the non-equilibrium components of the magnetisation is described by oscillating exponential functions. The oscillation results from the constant precession of all spins around the z-axis (the direction of the external magnetic field, B_0), a movement that produces a frequency modulated, exponentially decaying electric current in the detecting coil. The permanent precession of nuclei around B_0 is a consequence of the nuclei having spin and magnetic moment at the same time and is often compared to the precession of a child's top. By means of Fourier transform, the precession frequency for individual groups of spins can be extracted from the oscillating currents that are detected for the spin ensemble, yielding a distribution of resonance frequencies referred to as the NMR spectrum.

The central position that NMR spectroscopy occupies in analytical chemistry and, alongside x-ray crystallography, in structural biology, is rooted in the fact that resonance frequencies are extremely sensitive to a spin's chemical environment. NMR active chemical elements/isotopes resonate within certain frequency ranges determined by the gyromagnetic ratio, an inherent physical property of the nucleus. The resonance frequency (ν) of a nucleus is given as

$$\nu = \gamma B_0 / 2\pi \tag{1.1}$$

where γ is the gyromagnetic ratio and B_0 is the magnetic field in Tesla. However, the small magnetic moments which nearby electrons create in proximity to a given spin lead to small changes in the experienced external field, slightly shifting the resonance frequency. The parameter used to describe this phenomenon is the so-called shielding constant σ, changing the resonance condition to

$$\nu = \gamma B_0 (1 - \sigma) / 2\pi. \tag{1.2}$$

Rather than expressing the resonance frequency of each nucleus in absolute numbers, a relative scale is introduced, the so-called chemical shift

$$\delta = \frac{\nu - \nu_{ref}}{\nu_{ref}} * 10^6 \tag{1.3}$$

where ν_{ref} is the frequency of a reference compound. This dimensionless parameter is field independent, making spectra acquired at different field strengths directly comparable.

Contributions to chemical shielding do not only arise from bonding electrons. Through space effects such as magnetic fields produced by aromatic ring currents affect the overall shielding. The resulting individual local environments experienced by chemical building blocks of biomolecules (such as amino acids or nucleotides) allow the distinction between identical residue types at different positions within a compound. Individual nuclei are assigned their respective chemical shifts which function as “name tags”. This procedure allows individual residues to be distinguished and subsequently allows determination of individual spin systems and their interaction with other spin systems. In unfolded proteins - where all amino acids experience very similar chemical environments - this distinction becomes difficult as identical nuclei of amino acids of the same type resonate close to a single frequency, termed the random coil frequency.

Two of the most abundant biological nuclei, ^{14}N and ^{12}C, are not favourable for NMR spectroscopy since they have spin quantum numbers of 1 and 0, respectively. Proteins are therefore often isotopically labelled with ^{15}N and/or ^{13}C if the information gained from protons is not sufficient. This technique allows for extended "name tags" in which the chemical shift of several nuclei are observed at the same time, adding additional resolution. Labelling a proton with the chemical shifts of a directly attached ^{15}N or ^{13}C atom gives rise to so-called two- and three-dimensional NMR spectroscopy. Additional dimensions can be used by adding frequencies of further neighbouring groups, but such high-dimensional NMR spectroscopy usually requires extended measurement time, large compound concentrations and non-Fourier transform processing. Alternatively, pseudo nD spectroscopy methods have been developed [200, 201, 202].

Because ^{1}H, ^{15}N and ^{13}C have different gyromagnetic ratios, their resonance frequencies in biomolecules are well-separated from each other, and modern NMR spectrometers have several channels for separate irradation of different nuclei. This design makes it possible to excite only one isotope while leaving the other istopes in their equilibrium state. Additionally, the bandwidth, shape and centre frequency of a pulse can be set to selectively excite narrow frequency ranges within one channel (so-called selective pulses). Following the excitation, the nuclei (usually starting on proton for sensitivity reasons) are given time to dephase according to their different chemical shifts and scalar couplings. After chemical shift rephasing, achieved through central 180° pulses and an aedequate delay (so-called spin echos), magnetisation is transferred from the proton spins to neighbouring heteronuclei (in homonuclear experiments the magnetisation is transferred between different types of protons). The transfer is accomplished via simultaneous pulsing on both the proton and the heteronucleus. The polarization transfer can be mediated through scalar couplings (J-couplings) between covalently linked nuclei. It is also possible to transfer the magnetisation via dipole-dipole (DD) interactions through space. In this case the interaction is based on the so-called Nuclear Overhauser Effect (NOE) and decays rapidly with the distance r between the two dipoles, proportionally to r^{-6}. The NOE is essential for the use of NMR spectroscopy in structural biology because

it allows the proximity of protons in space to be established. Thus, NOE experiments are essential to establish tertiary and quaternary structures of biomolecules.

In recent years, efforts have been made in protein structure determination to replace the traditional, NOE-based methodology by structural information gained entirely on the basis of the chemical shifts of backbone atoms, which in principle reflect protein dihedral backbone angles - the parameters which dictate the secondary structure of proteins [203, 204]. These methods rely on secondary chemical shifts (the deviation of a chemical shift from its random coil value) and on the bulk of chemical shifts and protein structures that are deposited in databases. Methods based on chemical shifts require a larger contribution of modelling techniques in order to determine likely orientations of the sidechains since only backbone angles are found on an experimental basis.

For a rigorous description of the spin manipulations that take place during a NMR experiment it is possible to describe the magnetisation of a given spin using its x-, y- and z-components. For protons, the three Cartesian components are usually denoted I_x, I_y and I_z, for heteronuclei (i.e. ^{15}N or ^{13}C) S_x, S_y and S_z are often used. Due to the evolution of couplings mixed states arise, such as I_xI_y (a multiple quantum state) or I_zS_x (an antiphase single quantum state). In weakly coupled systems, the single components of these mixed states can be manipulated individually, i.e. a selective 90_{-z} degree pulse on the heteronuclear channel transforms I_zS_x into I_zS_y, leaving the proton component unaffected. (Selective excitation of single nucleus types is possible because the Larmor frequencies of different nuclei are sufficiently separated.) The so-called spin operator formalism is used here to describe the INEPT experiment.

1.8.1 Multi-dimensional NMR

NMR experiments follow a common scheme of four essential steps: preparation, evolution, mixing and detection. The preparation does often simply consist of a delay that is long enough for the spins to return to the thermal equilibrium before a new scan. This

is important because each pulse sequence needs to be repeated many times in order to obtain a good signal to noise ratio. Because many scans are added together it is essential that they all start from the same well-defined spin state. The preparation can contain a strong radio frequency field for one of the channels, leading to saturation of this nucleus. Other nuclei can, at the same time, be in thermal equilibrium. In the last step of the preparation period, spins are often placed at 90° with regard to the static field. This provides a basis for the evolution period, in which chemical shift or coupling evolution are allowed to take place. After a delay t_1 each spin is in a distinct orientation as a result of different precession frequencies, and only certain orientations are carried on through to the next steps of the experiment. Between the repetitions of the single experiments, however, the evolution time t_1 increases in regular intervals from scan to scan. In this way the magnetisation is labelled by the chemical shifts or couplings of spins. During the following mixing period the magnetisation is transferred between spins. This could be a simple delay (NOESY) or a repetitive train of pulses (TOCSY) or a single 90° pulse on one (COSY) or two (HSQC) channels. The mixing period is typically followed by the acquisition period.

1.8.2 The INEPT and HSQC sequences

The INEPT (*I*nsensitive *N*uclei *E*nhanced by *P*olarization *T*ransfer) sequence (Fig. 1.14) is a central building block in many bioNMR applications.

The pulses and delays used on the proton channel form a spin echo as described above. The effect of the first pulse on the proton channel is the creation of $-I_y$ magnetization from the initial I_z state. The J-coupling that evolves during the entire period τ transforms the $-I_y$ term into $2I_xS_z$ if τ equals $1/(2J_{IS})$. Two centred 90° pulses on both channels transform this proton antiphase term into $2I_zS_y$. This term corresponds to S-antiphase magnetization, i.e. a doublet with a 180° phase shift between the two peaks. The centre of the doublet corresponds to the S resonance frequency and its separation in Hz equals J_{IS}.

In order to acquire decoupled signals refocussing must be applied via an additional spin

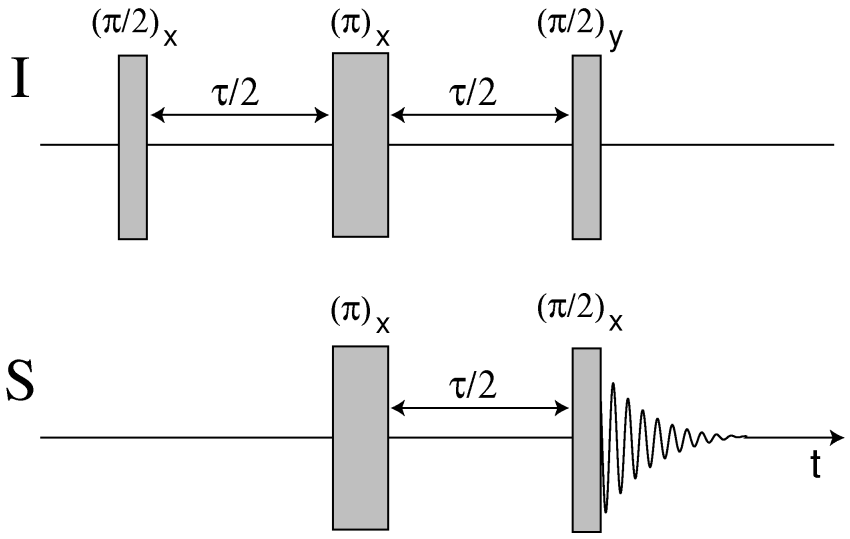

Figure 1.14: INEPT pulse sequence. I denotes the proton channel, S the insensitive nucleus. Taken from [205].

echo (Fig. 1.15). This sequence yields a decoupled in-phase signal that is enhanced in intensity by a factor of γ_I/γ_S for IS moieties in comparison to direct S detection.

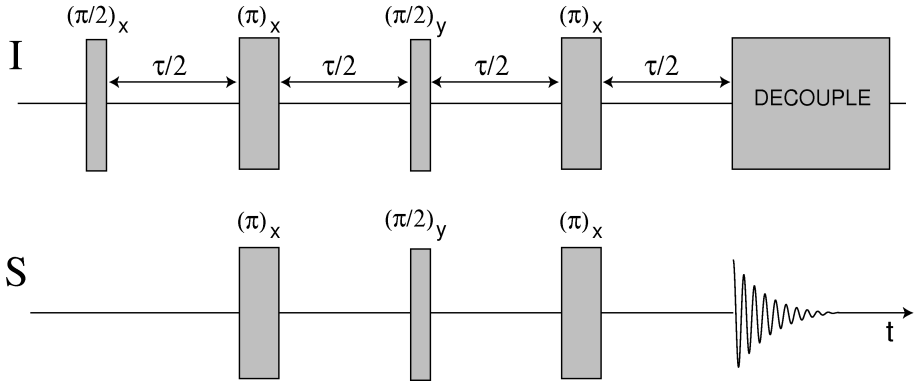

Figure 1.15: Refocussed INEPT pulse sequence. I denotes the proton channel, S the insensitive nucleus. Taken from [205].

The main use of INEPT is the detection of insensitive nuclei in an indirect yet more sensitive manner, namely via protons which are directly bound to insensitive nuclei (i.e. ^{15}N or ^{13}C) [206]. In ^{15}N-labelled proteins, the scalar coupling between protons and adjacent nitrogen nuclei can be used for *H*eteronuclear *S*ingle *Q*uantum *C*oherence

(^{1}H, ^{15}N-HSQC) experiments. In this experiment, magnetisation is transferred from the proton to the adjacent nitrogen nucleus using the INEPT pulse sequence. The scalar coupling constant between amide ^{1}H and ^{15}N, J_{NH} is around 90 Hz, and the INEPT transfer is optimised when $\tau=1/(2J_{NH})$. The magnetisation is then labelled with the nitrogen chemical shift and, in a reverse INEPT step, transferred back to the proton before detection takes place. Thus, the directly detected dimension contains the proton chemical shifts while the nitrogen shifts are displayed in the indirectly detected dimension (Fig. 1.16). Likewise, this experiment can be conduced on ^{13}C-labelled proteins to transfer magnetisation between protons and adjacent carbons, giving rise to the ^{1}H, ^{13}C-HSQC experiment.

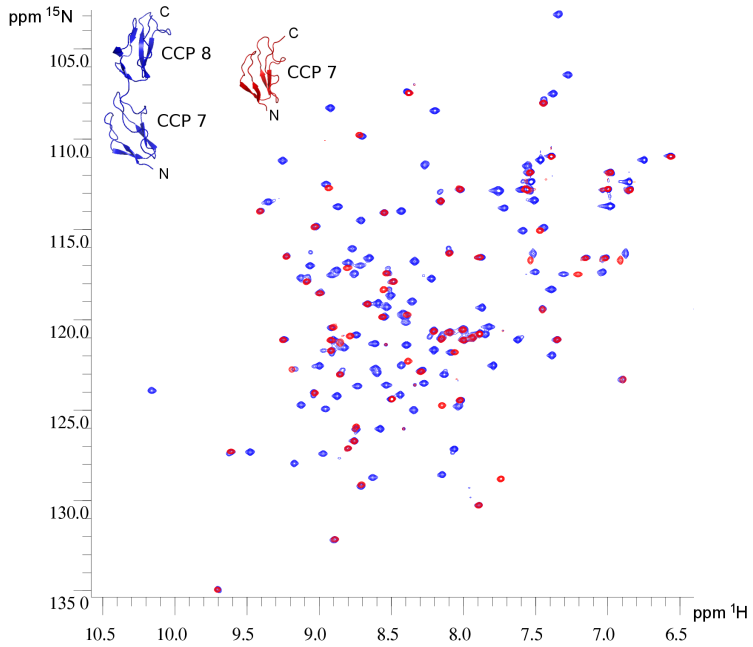

Figure 1.16: Overlay of the ^{1}H, ^{15}N-HSQC spectra of ^{15}N-fH∼7,8Y (blue) and ^{15}N-fH∼7Y (red) with inlayed structures of the two proteins (same colour coding, PDB 2JGX for fH∼7 and derived from PDB 2UWN for fH∼7-8).

While in a ^{1}H, ^{15}N-HSQC experiment amide NH resonances are monitored, ^{1}H, ^{13}C-HSQC spectra contain in principle all CH correlations. In practice, separate versions of the ^{1}H, ^{13}C-HSQC experiment are typically recorded for aliphatic and aromatic reso-

nances of proteins. ^{1}H, ^{15}N-HSQC experiments are usually optimised for protein backbone amides but also contain the glutamine and asparagine sidechains. Additionally, the arginine ϵ and, in exceptional cases, lysine ζ sidechain amine resonances can be observed at favourable pH values.

The 2D HSQC experiments form the basis of many techniques used in protein NMR spectroscopy. The ^{1}H, ^{15}N-HSQC spectrum is often the first experiment that is collected for a ^{15}N-labelled protein and used to judge if the protein is folded and monomeric. If the ^{1}H, ^{15}N-crosspeaks are well dispersed and deviate significantly from the random coil values, and if the number of crosspeaks corresponds well with the number that is expected for the protein sequence, the chosen buffer and temperature conditions can be maintained for further NMR studies. For backbone assignment, the 2D ^{1}H, ^{15}N-HSQC experiment is used to link a number of different 3D spectra which each other: each 3D spectrum uses the ^{1}H, ^{15}N-HSQC plane for the first two dimensions and adds a third dimension which varies between experiments. In this manner, further carbon and proton chemical shifts are linked to the backbone ^{1}H, ^{15}N-resonances of the HSQC-plane for each amino acid.

1.8.3 Sequential backbone assignment of proteins

For protein structure determination, ^{15}N, ^{13}C-labelled protein samples are routinely used in order to obtain almost complete chemical shift information of the individual amino acids, allowing a large number of NOE-restraints to be used. Structures of small proteins and peptides with molecular weights < 6 kDa can be determined without isotope labelling using homonuclear TOCSY, COSY and NOESY experiments. These techniques are especially useful in the case of proteins which can not readily be produced by micro-organisms, for example in the case of bactericidal activity - isotope labelled amino acids for peptide synthesis are expensive. For the range of approximately 6-12 kDa ^{15}N labelling is sufficient for structure determination. During the sequential assignment,

chemical shift values of individual amino acids are used to identify residues as belonging to a certain amino acid type.

For $^{13}C/^{15}N$ labelled proteins so-called triple resonance experiments link the resonance frequencies of different atoms to the backbone amide proton and nitrogen chemical shifts that are displayed in the ^{1}H, ^{15}N-HSQC experiment. In the third dimension, chemical shift information from the C^{α} and C^{β} atoms can be gained as well as the resonance frequencies of the backbone carbonyl groups and those of H^{α} and H^{β} atoms. Triple resonance experiments are usually acquired in pairs: in one of the pairwise collected spectra the magnetisation transfer is directed through the carbonyl group, linking the amide NH frequencies exclusively to the chemical shift information of the preceding amino acid. In the second spectrum the magnetisation transfer occurs via C^{α} atoms of the n and n-1 residue, linking each NH crosspeak to both the chemical shift of its own and the preceding C^{α}. Using both spectra simultaneously allows assignment of each peak in the third dimension as belonging to the nth or $(n-1)$th amino acid, allowing unambiguous sequential assignment of the polypeptide chain. In practice, several pairs of triple resonance experiments are combined to determine the sequential connectivity in order to overcome difficulties arising from spectral overlap or low signal-to-noise ratios. Depending on the atom type in the third dimension, these pairs are denoted CBCA(CO)NH/HNCACB [207, 208], HAHB(CO)NH/HAHBNH [209] and CONH/(CA)CONH [210] (Fig. 1.17). When all expected signals are present, the spectra pairs can be used independently to establish the connectivity, otherwise they can complement each other. Programs for automated backbone assignment facilitate the assignment process [211]. Due to the lack of a secondary amide in the proline backbone the sequential assignment of backbone resonances is interrupted since these residues are not visible in the ^{1}H, ^{15}N-HSQC experiment. The carbon resonances of proline residues can, however, be determined on the following residue during the sequential assignment.

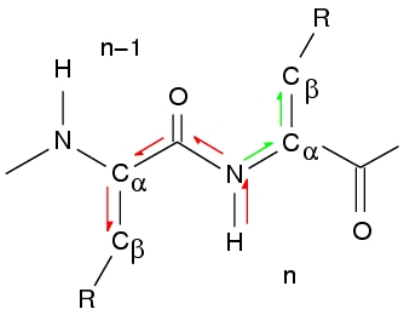

Figure 1.17: Magnetisation transfer during the CBCA(CO)NH (red) and CBCANH (red and green) experiments. The CBCA(CO)NH experiment transfers resonances from the n-1 residues to the backbone NH of the n residue only, the CBCANH experiment additionally transfers resonances from the n residue. Excitation takes place on the aliphatic protons and is detected on the amine NH, or both excitation and detection take place on NH-protons (out-and-back experiments).

1.8.4 Relaxation and correlation times

After bringing the nuclear magnetic moments out of their thermal equilibrium orientations in the static magnetic field, one can measure the time constants with which they return to the thermal equilibrium, the so-called relaxation rates. Because different processes are responsible for the evolution of the magnetic components that are perpendicular and parallel to the external field, the time constants with which these components return to equilibrium are determined separately. In order to return to the thermal equilibrium, the spins interact in multiple ways with the surrounding environment. The longitudinal relaxation time constant T_1 is used to describe the return to thermal equilibrium populations along the z-axis, while processes leading to the decay of single-quantum coherences in the xy-plane are described by the transverse relaxation time constant T_2. Small fluctuating magnetic fields caused by molecular tumbling are the source for relaxation processes. For spin 1/2 the most prominent sources of fluctuating fields are through space dipole-dipole (DD) interactions and chemical shift anisotropy (CSA) of nuclei. When several relaxation mechanisms are present these can cross-correlate because they have a common origin: thermal motion. Pulse sequences

which remove the cross-correlation between dipolar and CSA relaxation must be used to determine the DD relaxation. The suppression of cross-correlation can be achieved by a train of 180° degree pulses during the ^{15}N relaxation period [212]. Additionally, selective inversion of the NH protons is applied in the centre of this period which has the double advantage of suppressing both cross-correlation and exchange broadening caused by water. The relaxation then becomes monoexponential and the decay of magnetisation can be expressed as

$$M(t) = M_0 * exp(-R * t), \tag{1.4}$$

where R is either the longitudinal or the transverse relaxation rate, R_1 or R_2 respectively, depending on the experiment. Relaxation is often reported by using longitudinal (or spin-lattice) and transverse (or spin-spin) relaxation times T_1 and T_2 ($T_1 = 1/R_1$ and $T_2 = 1/R_2$).

In proteins, the measurement of T_1 and T_2 is commonly performed in a residue-dependent manner. Because relaxation times reflect local as well as global mobility, relaxation measurements are used to investigate site-specific flexibility and dynamics but also overall molecular size. For example, backbone amide groups are often used for relaxation measurements to determine flexible loops and rigid structure elements. T_1 and T_2 measurements can also be used to estimate a protein's overall rotational correlation time (τ_c) because both T_1 and T_2 depend on the overall frequency with which the protein tumbles in solution. For macromolecules in solution, T_1 increases beyond a certain molecular weight with increasing molecular size, while T_2 decreases monotonically (Fig. 1.18). The correlation time estimated from the R_2/R_1 ratio are proportional to the protein size, and thus to the molecular weight, making relaxation measurements a convenient tool for investigation of self-association processes [213].

2D ^{1}H, ^{15}N-HSQC like relaxation spectra, acquired using different ^{15}N relaxation delays, yield accurate relaxation times through fitting the decay of intensity for each residue as a function of the delay time. In small peptides where little overlap between the amide signals exists 1D spectra recorded with different ^{15}N delays can be used to estimate

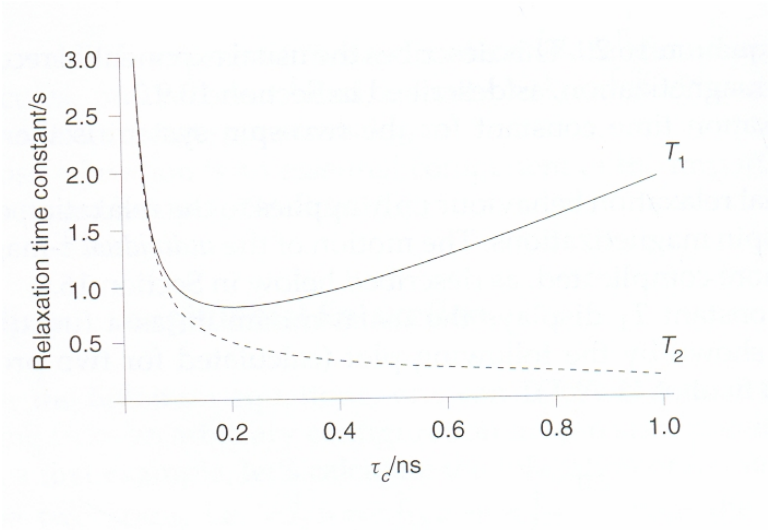

Figure 1.18: Schematic representation of the change in relaxation time constants with increasing correlation time τ_c. Taken from [205].

relaxation times. The peak intensity I decays exponentially with the relaxation time constant T, thus with the intensities I_x and I_y belonging to the two different delay times the time constant can be estimated from just two measurements (though precision increases with more points):

$$I_x = I_0 exp\left(\frac{-t_x}{T}\right) \tag{1.5}$$

$$I_y = I_0 exp\left(\frac{-t_y}{T}\right)$$

$$T = \frac{t_x - t_y}{log(I_y/I_x)} \tag{1.6}$$

this is equally valid for T_1 and T_2, depending on which NMR experiment is chosen. For T_1, the so-called inversion recovery sequence is used which inverts the ^{15}N magnetization to $-I_z$. After this point increasing the relaxation delay allows sampling of the exponential decay through recording the magnetisation after the transfer to protons.

To determine T_2 the ^{15}N spin echo pulse sequence 90°-τ-180°-τ with $\tau << 1/2\ ^1J_{NH}$ is applied. This effectively removes the $^1J_{NH}$ coupling evolution. However, in order to suppress the effects of cross correlation between dipolar and CSA relaxation mechanisms every 5-10 ms the multiplet components are exchanged by the action of a ^{1}H 180° pulse [212]. Plotting the intensities as a function of τ and fitting an exponential function yields the decay constant T_2.

Because signal detection can only take place in the xy-plane, every NMR signal is inevitably affected by T_2. Since T_2 gets shorter with increasing τ_c (and thus with increasing molecular weight) (Fig. 1.18), and because the linewidth is proportional to $1/T_2$, NMR spectroscopy generally suffers from increased line broadening with increasing molecular weight, restricting routine INEPT-based NMR spectroscopy to molecules of up to around 30 kDa. For larger proteins, the different relaxation properties of different spin states are used. For a given spin, the doublet in which its resonance is split by the scalar coupling with a neighbouring spin is non-symmetric, i.e. for an amide proton spin the $^{15}N^{\alpha}$ and $^{15}N^{\beta}$ states of the neighbouring amide nitrogen have different relaxation properties. This effect becomes more pronounced with large molecular weights. Instead of mixing the different spin states and yielding one relative broad signal, decoupling is avoided and the spectral analysis focussed on the sharper signal of the in-phase doublet (Fig. 1.19). This concept is used in *T*ransverse *R*elaxation-*O*ptimised *S*pectroscop*Y* (TROSY) and Cross Relaxation-Enhanced Polarization Transfer (CRINEPT) Spectroscopy [214]. TROSY is based on constructive interference between DD coupling and CSA in the absence of decoupling [215]. While DD coupling is field-independent, CSA increases proportionally to the field strength, and at 900 MHz both effects almost cancel for ^{15}NH-groups. In TROSY, the compensation of DD couplings and CSA is active during the ^{15}N chemical shift evolution and the ^{1}H acquisition. For large proteins with molecular weights > 100 kDa for which TROSY is already used, transverse relaxation during the magnetisation transfers becomes limiting. The CRINEPT sequence combines the cross-relaxation transfer with the coherent J-transfer. The implementation used in this

work, CRINEPT-HMQC-TROSY, removes the effective coupling evolution in F1 but relies on relaxation to remove the fast relaxing component prior to the t_2 acquisition period [216]. For our system this process was not fully effective and asymmetric ^{1}H-^{15}N-doublets were observed in the F2 dimension. Both TROSY and CRINEPT, however, do not eliminate DD couplings between NH groups and remote protons - therefore, both techniques achieve their maximum potential only for deuterated proteins.

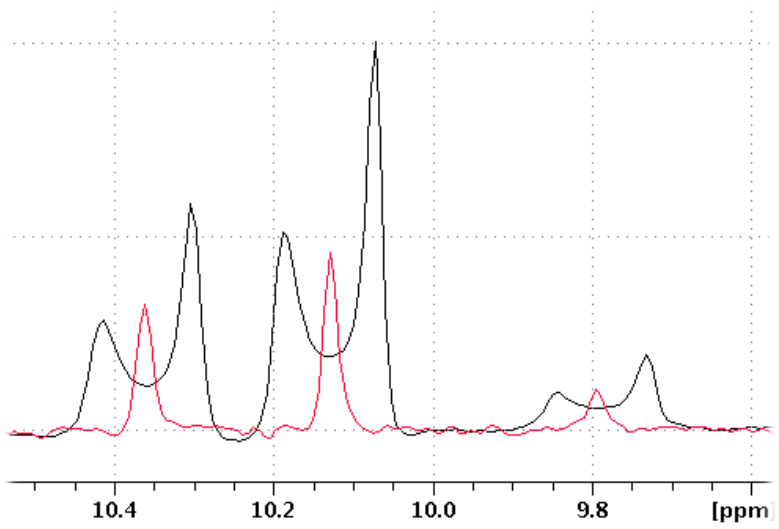

Figure 1.19: Comparison of ^{1}H proton traces extracted from 2D amide spectra with and without decoupling (red: decoupled HSQC, black:CRINEPT-HMQC-TROSY without decoupling). The different intensities of the doublet components in the spectrum acquired without decoupling are averaged in the decoupled HSQC spectrum. The red trace was acquired on a ca. 15 kDa ^{15}N labelled protein, the black trace after addition of a second protein of ca. 185 kDa, forming a protein-protein complex of ca. 200 kDa with the smaller labelled protein.

1.8.5 ^{1}H, ^{15}N-HSQC monitored titrations

Following the backbone assignment, binding studies can be conducted using the ^{1}H, ^{15}N-HSQC, TROSY or CRINEPT experiments. The chemical shift changes that are observed during a titration with a small ligand are used to determine the ligand binding

site on the protein. This so-called chemical shift mapping is a common application of bio-NMR spectroscopy and can, if the kinetics of the binding are favourable, be used to determine the dissociation constant of the complex which forms during the titration. A gradual chemical shift change can be observed for residues in the fast chemical exchange regime, i.e. when the interconversion between free and bound protein is fast compared to the difference between the Larmor frequencies of the free and bound state. In this case, the obtained spectrum represent the population-weighted average of the resonance frequencies of the bound and unbound species (Fig. 1.20), and with higher ligand concentration the averaged peak moves further towards the chemical shift of the bound species. In the case of slow exchange, both the bound and free protein are observed simultaneously, giving rise to two peaks for a single amide group. The most inconvenient case is, however, exchange on the intermediate timescale: in this exchange regime, intermediate broadening of peaks is observed which can lead to complete "bleaching" and disappearance of cross peaks during the titration. Because the difference in Larmor frequencies depends on the field strength and temperature, varying these two parameters can improve the spectral quality by changing the exchange regime.

Figure 1.20: Schematic representation of ^{15}N-HSQC shift changes during fast exchange: V445 in ^{15}N-fH∼7,8Y upon titration with fully sulfated heparin tetrasaccharide. Molar ratios are protein:ligand 1:0 (red), 1:0.4 (green), 1:0.8 (orange) and 1:1.5 (blue). Vertical dimension: ppm(^{15}N), horizontal dimension: ppm(^{1}H).

In ^{1}H, ^{15}N-HSQC titrations the effect of ligand binding on the protein backbone is monitored, whilst in the ^{1}H, ^{13}C-HSQC experiment usually the sidechains can be investigated. It is noteworthy that one interaction can exhibit different exchange regimes

in both HSQC-experiments because the difference between the bound and free state resonance frequencies (in Hz) can be different for NH and CH protons. In the fast exchange regime, the chemical shift differences during a titration are usually reported as combined chemical shift differences after scaling of the heteronucleus frequency. In the case of ^{15}N-labelling,

$$\sqrt{(\Delta\delta_{\mathrm{H}})^2 + \left(\frac{\Delta\delta_{\mathrm{N}}}{5}\right)^2} \tag{1.7}$$

is used, where $\Delta\delta_{\mathrm{H}}$ and $\Delta\delta_{\mathrm{N}}$ are the chemical shift differences between each titration point and the ligand-free reference spectrum for ^{1}H and ^{15}N, respectively.

1.8.6 NMR studies of glycosaminoglycans

Because no method for homogeneous isotope labelling in carbohydrates exists to date, homonuclear NMR is used for glycosaminoglycan characterisation. A large emphasis is on *J*-couplings and chemical shifts and common NMR experiments are ^{1}H-COSY (*CO*rrelation *S*pectroscopy), TOCSY (*TO*tal *C*orrelation *S*pectroscopy) and NOESY experiments. With the COSY and TOCSY experiments single hexoses are identified, and NOEs between the rings are used to establish the primary sequence. Natural abundance ^{13}C NMR can also be used if sufficient quantities of the GAG can be produced.

2 Materials and Methods

2.1 Protein production and purification

2.1.1 Fermentation for production of isotope labelled proteins

For production of ^{15}N fH∼7 (Y402), ^{15}N fH∼7-8 (H402), ^{15}N fH∼7-8 (Y402) and ^{15}N fH∼19-20 fermentation of the respective *Pichia pastoris* (*P. pastoris*) clones was conducted. The clones were provided by Dr. Christoph Schmidt and Dr. Andrew Herbert and contained the expression vector pPICZαB with the human fH residues 386-446 (fH∼7), 386-507 (fH∼7-8) and 1107-1231 (fH∼19-20) where the amino acid in position 402 is a tyrosine residue in fH∼7Y and fH∼7-8Y and a histidine in fH∼7-8H. (The complete sequences can be found in the Appendix.) With respect to residue 402 both fH variants are commonly found but Y402 is the prevalent form and H402 has been suggested as a risk factor for age related macular degeneration. The vector conveys resistance against the antibiotic Zeocin. The protein genes had been placed upstream of the *Saccharomyces cervisiae* α-factor secretion sequence, directing the proteins to the secretory pathway and thereby facilitating purification. Short cloning artefacts between the protein sequence and the secretion signal where present in these clones, facilitating the cleavage of the secretion signal by Kex2 endopeptidase. These artefacts are subsequently subject to partial proteolysis, resulting in a mixture of protein with and without the cloning artefact. Ion exchange or heparin affinity chromatography at mildly acidic conditions (pH 5.0-6.0) allowed separation of the two species. For the fH∼7Y and fH∼7-8Y/H constructs, the cloning artefact comprised of the additional sequence EAAG

at the N-terminus. For fH∼19,20, the cloning artefact was EAEF. Mutant proteins of fH∼19,20 were provided by Dr. David Kavanagh and Dr. Andrew Herbert and have, in general, been produced in a similar manner.

Previously, the vectors had been transformed into the *P. pastoris* strain KM71H. Clones were streaked from glycerol stocks on YPD-agar containing Zeocin (100 μg/mL) and incubated at 30°C for 48 h. Single colonies were used to inoculate 10 mL of buffered minimal glycerol (BMG). After shaking at 200 rpm for 48 h at 30°C the 10 mL cultures were combined with 200 mL of fresh BMG and left shaking at 30°C for another 48 h, after which the inoculant (ca. 300 mL, OD_{600} of 20) was spun down (10 min, 1,500 g) to separate the cells from the ^{14}N-containing media. Pellets were resuspended in 20 mM sterile potassium phosphate, pH 6.0 and used to inoculate 600 mL of Minimal Media (MM) for fermentation at 30°C.

A New Brunswick Bioflow3000 system fitted with a pH probe (Mettler Toledo), a dissolved oxygen (DO) probe (Mettler Toledo), a stainless steel stirrer and a sterile filter (0.2 μm) air inlet were used for the fermentation. The fermentation was controlled and recorded using an AFS Bio Command Interface. 2.6 mL PTM1 trace salts and 0.5 mL antifoam were added to the MM prior to inoculation. The temperature was set to 30°C until the glycerol from the initial media and additional 10 mL of glycerol had been metabolised. The DO level served as a general measure of metabolic activity. The pH of the culture was regulated with 2 M KOH to stay above 4.5, and air was pumped through the culture at a constant flow rate of 0.5 units. The pH probe was calibrated prior to autoclaving using reference solutions of 7.0 and 4.0 pH units. For DO probe calibration, a value of 100% was set to equal the DO level in the MM before inoculation, at 30°C, 200 rpm and 0.5 air flow, and a DO level of 0% was set after deconnecting the probe for 30 s. The agitation was coupled to the DO value via a control circuit in order to keep the DO level at 40 % during the fermentation. Additionally, oxygen air enrichment was used if at top agitation a DO value of 40% was still not reached, as a constant DO level is believed to be critical for protein expression. The lower and upper value for the agitation

were 200 and 1000 rpm, respectively. A second control circuit was used to maintain a pH of around 5.0 by adding small amounts of 2 M KOH. Before induction with methanol (enriched with 4.4 PTM1 salts/mL methanol) the culture was cooled down to 15 °C to limit proteolysis. Methanol feeding periods were manually adjusted to match the level of adaptation to methanol, leading to ever increasing amounts of injected methanol. The estimated total amount of methanol in the cell culture was kept below 1.5 % at any time to avoid intoxication. After 4 days of induction the cells were removed from the fermenter, spun at 5,000 rpm for 10 min (centrifuge Sorvall RC3B plus) and discarded. The supernatant was spun at 10,000 g for 30 min (centrifuge Sorvall RC26 plus) and filtered through a 0.2 µm filter to cut off any remaining cells. 5 mM ethylendiamin tetraacetic acid (EDTA), 0.5 mM phenylmethyl sulfonyl fluoride (PMSF) and 0.002% (v/v) sodium azide were added to the filtered supernatant to inhibit proteolysis and bacterial growth.

For production of ^{13}C,^{15}N-fH∼7-8Y, the initial fermentation media did not contain glycerol but 15 g of ^{13}C-D-glucose as the sole carbon source. An additional feed before induction consisted of 1 g of ^{13}C-glycerol. For induction, overall 20 g of ^{13}C-methanol was used.

The production of ^{15}N-NK1 followed a generally similar procedure. The clone was provided by Jon Deakin, Paterson Institute for Cancer Research, Manchester University, and contained the expression vector pPIC-9K with the human HGF/SF residues 28-210 and the mutation A29V. An initial colony was grown on YPD agar containing the antibiotic G418 (10 µg/mL) against which pPIC-9K conveys resistance. The overall amount of methanol used during the induction phase was much higher than for the fH-producing clones to account for the intact Alcohol Oxidase 1 (AOX1) promotor in the NK1-producing clone. This gene is disrupted in the pPICZα vector, leaving only the AOX2 gene intact and resulting in a slower methanol metabolism in the latter case.

2.1.2 Shaker flask expression of unlabelled proteins

For the expression of unlabelled fH-constructs, shaker flasks were used instead of fermentors. Cultures were inocculated using BMG, 30 °C and 250 rpm and transferred to BMM for induction at 15-20°C. 0.5-1.5% (v/v) methanol was added twice a day and the cells were harvested 4 days after transfer to BMM.

2.1.3 Ion Exchange and Heparin Affinity Chromotography

Following the addition of protease inhibitors and sodium azide, the supernatant was diluted with the 4 fold volume of distilled water and an initial purification step using a self-packed 5 mL column of SP-Sepharose (Sigma) or Heparin-Sepharose (Amersham) was conducted at gravity flow and 4°C. The bound protein was washed with 20 mM potassium phosphate pH 6.0 until the absorbance at 280 nm (A_{280}) had reached a value below 0.1, then step-eluted with NaCl. For factor H constructs 1 M NaCl, 20 mM potassium phosphate pH 6.0 were used for elution, for NK1 1.5 M NaCl, 20 mM potassium phosphate pH 6.0. 0.5 -1 mL fractions were collected until the A_{280}-value levelled close to 0. The fractions were screened for the protein of interest using SDS-Polyacrylamide gelelectrophoresis (SDS-PAGE), combined accordingly and desalted by buffer exchange or dialysis.

For the second purification step another cation exchange or heparin affinity chromatography was applied using either a 1 mL MonoS column (General Electrics) or a Poros 20HE (0.46 x 10 cm, Applied Biosystems) heparin-affinity column and Äkta FPLC system (Amersham) for both factor H modules and NK1. The column was equilibrated with 20 mM potassium phosphate (plus 100 mM NaCl for NK1), pH 5.0-6.0. Desalted protein was injected and a linear gradient to 1 M or 1.5 M NaCl (NK1) was applied. The elution of protein was followed by the A_{280}-value and fractions with high absorbance values were subject to SDS-PAGE. Clean fractions were combined and - depending on their further usage - concentrated, buffer exchanged or frozen in liquid nitrogen and stored at -80°C.

2.1.4 SDS-Polyacrylamide Gel Electrophoresis

Protein molecular weights were estimated and the purity of protein samples assessed using sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). Protein amounts ranged between ca. 1-5 µg per lane. Protein samples were mixed with 0.5 equivalents (v/v) of 1:2 reducing or non-reducing loading buffer (2x sample buffer according to Lämmeli, Fluka) and heated at 80°C for 3 min. Total sample volumes were around 20 µL. Each sample was loaded into an individual well on a prepacked polyacrylamide gradient gel (NuPAGE 4-12% (w/v) Bis-Tris, Invitrogen). For small molecular weights MES buffer was used, for large molecular weights MOPS buffer (20x NuPAGE MES or MOPS SDS running buffer, Invitrogen). For referencing, a protein standard mix (Precision Plus Unstained Standard, BioRad) was loaded into one of the wells. Electrophoresis was persued for about one hour at 150 V. Afterwards, the gel was washed several times with dH_2O for 30 min, then Coomassie-stained (EZblue, Sigma) for about one hour. For destaining, the gel was left in dH_2O over night.

2.1.5 Mass spectrometry

Spectra were acquired on a Voyager-DE STR MALDI-TOF (Applied Biosystems) instrument with a nitrogen LASER in negative mode. Matrices used were sinapinic acid for protein samples and α-cyano-4-hydroxy-cinnamic acid (CHCA) for peptides. Matrices were obtained by mixing 15 mg sinapinic acid or 10 mg CHCA, respectively, with 400 µL ddH_2O, 100 µL 3% (v/v) TFA and 500 µL acetonitrile, followed by sonication for 3 min. Generally, 0.5 µL protein or peptide samples were mixed with 0.5 µL of the appropriate matrix directly on the MALDI-TOF plate. The LASER intensity was adjusted manually for each sample. For analysis of peptides obtained during the microwave assisted TFA digestion the web server ProSight PTM was used.

2.2 Oligosaccharide purification

2.2.1 Size exclusion chromatography

Heparinase I digested heparin from bovine intestines was provided by Jon Deakin. 50 mg of digested heparin was dissolved in 1 mL 250 mM ammonium bicarbonate and applied to a Bio Gel P-10 gel filtration column (120 x 1.5 cm, Bio-Rad) in 250 mM ammonium bicarbonate under gravity flow conditions. The A_{232} of each fraction was measured and the major species, disaccharide (dp2), tetrasaccharide (dp4), hexasaccharide (dp6) and octasaccharide (dp8) were isolated. For dp2, dp4 and dp6 the size-separated oligosaccharides were subsequently subject to anion exchange chromatography in order to achieve separation according to sulfation levels/overall charge.

2.2.2 Ion exchange chromatography

A strong anion exchange chromatography (SAX) column (AS-17 Ionpac, Dionex) on a HPLC system (Waters, 600 controller and 486 detector) was used to further purify the size-fractionised oligosaccharides. Distilled water, titrated with HCl to a pH of 3.5, was used as running buffer and a salt gradient to 1 M NaCl was applied. The elution was followed by A_{232} and fractions were collected manually. Purified oligosacchrides were freeze-dried and concentrated.

2.2.3 Desalting

Desalting of dp4 and dp6 was accomplished on the same Waters HPLC system using a Superdex Peptide 10/300 GL column (General Electrics) with 250 mM ammonium bicarbonate as running buffer and following the A_{232}-trace or, for dp8, using a PD-10 desalting column (General Electrics) and dH_2O. For the PD-10 columns, 0.5 mL fractions were collected and the A_{232}-value was measured for each fraction on a UV/vis spectrometer (Lambda, HP) using quartz cuvettes. Fractions with low A_{232}-values on the later end of the A_{232} peak were omitted due to overlap with the salt elution. This

overlap was found to be too large for oligosaccharides smaller than dp8, hence the two different desalting methods.

2.3 Concentration measurements for proteins and GAGs

Protein concentrations were determined by using the A_{280} values. Concentrations were calculated using the Beer-Lambert equation with the following values for extinction coefficients at 280 nm (ϵ_{280}):

Table 2.1: Calculated extinction coefficients at 280 nm

protein	ϵ_{280} / $M^{-1}cm^{-1}$
fH∼7Y	13200
fH∼7H	11710
fH∼7-8Y	24910
fH∼7-8H	23420
fH∼19,20 WT	27430
fH∼19,20 mutant proteins containing an additional tryptophan	21930
NK1	26630
C3b	176700

as obtained from the Swiss-Prot database for the respective amino acid sequences. HBD2 samples were obtained from Dr. Emily Seo who used the Bradford assay for concentration measurements. All oligosaccharide concentrations were determined using A_{230} values and ϵ_{230}=5200 $M^{-1}cm^{-1}$ [198]. The absorbtion at 230 nm is due to the double bond and neighboring carboxyl group at the non-reducting end of the oligosaccharide and largely independent of the exact chemical composition of the GAG.

2.4 NMR

2.4.1 Preparation of protein and GAG samples for NMR

Purified protein was buffer exchanged using Vivaspin concentrators (Satorius, with PES membranes) or overnight dialysis (Tube-O-DIALYZERS, GBiosciences). The molecular weight cutoff for concentrators (MWCO) was 3, 5 or 30 kDa, depending on the protein size, and 1 kDa for overnight dialysis. For 5 mm (outer diameter, O.D.) NMR tubes, the concentration of the sample and its volume ($\leq$ 550 µL) were adjusted and 5-10% (v/v) of D_2O as well as 0.01% (w/v) sodium azide were added. Samples for lysine sidechain experiments were not directly supplemented with D_2O. Instead, the sample volume was adjusted to 400 µL and a capillary (Norell, 2 mm O.D.) was filled with D_2O and inserted into the sample in the 5 mm O.D. tube. For 3 mm NMR tubes, the sample volume was 200 µL and the D_2O concentration was 10% (v/v). Protein concentrations varied between 50 µM (for HSQC-titrations) and 0.8 mM (for triple resonance experiments). Oligosaccharides were, after desalting, repeatedly freeze-dried and dissolved in 500 µL ultrapure D_2O (Deuterium oxide "100", Sigma). Prior to NMR spectra collection the sample pH was measured using a Hamilton Minitrode and, if necessary, adjusted with small amounts of the buffer's conjugated acid or base (0.1 M stocks). NMR samples were stored at 4°C.

2.4.2 NMR spectrometers

A Bruker AVANCE 14.1 T spectrometer (600 MHz proton Larmor frequency) and a Bruker AVANCE 18.8 T spectrometer (800 MHz proton Larmor frequency), both equipped with a 5-mm triple resonance cryoprobe, were used throughout the project. For the C3b binding study, a 3-mm sample tube was used which was inserted in the 800 MHz 5-mm probe using a 3-mm spinner. Spectra were acquired using the Bruker TOPSPIN software.

2.4.3 Backbone assignment of fH∼7-8Y

Spectra for the backbone assignment of fH∼7-8Y were aquired at 600 MHz field strength using 5 mm sample tubes at 310 K on a 0.8 mM ^{15}N-protein sample in 20 mM sodium acetate pH 5.0 containing 5% (v/v) D_2O. For backbone assignment, the CBCA(CO)NH/CBCANH [217, 218, 219] pair as well as the NH(CA)CO/NHCO [220, 221, 222, 223] and HBHA(CO)NH/HBHANH [217, 218] pairs were acquired. Chemical shift labelling periods were typically set to 8-12 ms for ^{1}H, 4-6 ms for ^{13}C and 15-20 ms for ^{15}N. 1D spectra were acquired in between 3D spectra to confirm the stability of the sample. Generally, the focus during the sequential assignment was on the CBCA(CO)NH/CBCANH spectra; the other two spectra pairs were used to complete the assignment where severe overlap or sparse data made the assignment of the CBCA(CO)NH/CBCANH pair ambigous. Spectra were processed using the *process* program from the AZARA program suite (v2.7, copyright (C) 1993-2002 Wayne Boucher and Department of Biochemistry, University of Cambridge) and assigned using CcpNmr Analysis [224]. The AZARA code can be obtained from http://www.ccpn.ac.uk/azara/. After grouping resonances in the three experiments as C^α, C^β, H^α and CO-resonances and identifying glycine residues as far as possible, the peaks were exported into the semi-automated backbone assignment program MARS [211]. Further information required by MARS were the primary sequence and a secondary structure prediction, which was obtained using PSIRED (Position Specific Iterated Prediction) [225]. Glycine residues needed to be identified prior to MARS because a missing C^β entry in the MARS input list is treated as a potentially larger amino acid whose C^β resonance could not be identified. Glycine residues therefore have to be denoted as such in the MARS input file. The assignment predictions obtained by MARS were imported into Analysis via a Format Converter (CcpNmr FormatConverter, Wim Vranken, European Bioinformatics Institute).

2.4.4 Relaxation experiments

^{15}N relaxation measurements were performed at 283 K or 298 K on 600 MHz (Bruker AVANCE) or 800 MHz (Bruker AVANCE) spectrometers equipped with cryprobes as 1D or 2D experiments. 1D measurements were found to be sufficient in the case of HBD2 (283 K, 600 MHz AVANCE) due to little overlap in this relatively small peptide. 1D spectra with different delay times for T_1 and T_2 measurements were collected, the intensity decrease was measured for amide signals and the relaxation times were calculated using eq. 1.6. Delay times for the free HBD2 peptide were 11.1, 201.1, 301.1, 401.1, 501.1, 601.1, 701.1 and 801.1 ms for T_1 and 16.0, 48.0, 80.0, 96.0, 112.0, 128.0, 160.0 and 192.0 ms for T_2 measurements. For a sample of HBD2 with a 4 fold excess of DS-dp6, 50.1, 201.1, 301.1, 401.1, 501.1, 601.1 and 701.1 ms were used for T_1 and 16.0, 48.0, 80.0, 112.0, 144.0, 160.0 ms for T_2 measurements. For a sample of HBD2 with a 16 fold excess of the pentasaccharide Fondaparinux (a commercial name for the thrombin/antithrombin binding sequence in HS). The following delays were used: 11.1, 201.1, 401.1, 601.1, 701.1 and 801.1 ms for T_1 and 16.0, 48.0, 80.0, 96.0, 112.0, 128.0 ms for T_2 measurements.

Six delay times for T_1 (51.0, 301.0, 501.0, 601.0, 751.0 and 851.0 ms) and six delay times for T_2 (16.0, 32.0, 48.0, 64.0, 80.0 and 96.0 ms) were used for 2D experiments on factor H constructs. Relaxation rates for each of the chosen signals were obtained by exponential decay fitting in gnuplot (1D, after deconvolution using TOPSPIN) or CcpNmr Analysis (2D). Correlation times τ_c were calculated from the obtained R_1- and R_2-values and the spectrometer frequency using the program R2R1_tm (A. G. Palmer III, Columbia University). In the case of the 1D spectra, only 8-10 peaks in the amide region were selected and the values obtained for T_1 and T_2 were averaged before calculating the correlation times using R2R1_tm. Sidechain amides were excluded from both 1D and 2D relaxation analysis as they might be very flexible. For analysis of 2D spectra the relaxation times of all backbone amides were calculated but poorly fitted peaks were omitted prior to averaging.

2.4.5 ^{1}H, ^{15}N-HSQC titration experiments

Following the backbone assignment of fH∼7-8Y several ^{1}H, ^{15}N-HSQC spectra were measured between 310 and 298 K and a pH-titration was conducted at 298 K between pH 5.0 and 7.4. Using both temperature and pH dependency, the backbone assignments were transferred from 310 K, pH 5.0 to 298 K, pH 7.4 with the help of these titrations using CcpNmr Analysis. The oligosaccharide titration experiments on fH∼7-8Y and fH∼7-8H were done at the latter conditions with protein concentrations around 50 µM and varying ligand concentrations. Overlays of ^{1}H, ^{15}N-HSQC spectra of ^{15}N-fH∼7-8Y and ^{15}N-fH∼7-8H showed only very few differences in chemical shifts, implying only limited local changes around the SNP. The assignment of the fH∼7-8H-^{1}H, ^{15}N-HSQC spectrum was therefore done based on chemical shift similarities with the Y-form. The largest chemical shift difference between both isoforms was found to be experienced by the SNP-neighbouring residue 403.

Titration conditions for NK1 were 303 K, 20 mM sodium acetate pH 6.0, 100 mM NaCl. No double labelled sample was prepared and assignments were not available for this protein. A few crosspeaks of residues that are involved in GAG binding could, however, be assigned on the basis of previously published spectra [125].

Analysis of chemical shift changes during titrations was facilitated by CcpNmr Analysis, which includes macros for assignment transfer during titrations and the calculation of chemical shift differences between different spectra.

2.4.6 CRINEPT-HMQC-TROSY experiments on the fH19-20-C3b complex

CRINEPT-HMQC-TROSY spectra were acquired on samples containing ^{15}N fH19-20 and C3b. The samples were prepared in 3 mm tubes to yield higher concentrations and to minimize sensitivity losses due to relatively high salt content of the samples (100 mM

potassium chloride and 50 mM potassium phosphate). 10% (v/v) D_2O was added to the NMR samples. The relaxation delay was shortened to 0.9 s (instead of 1.5 s used for smaller proteins) and acquisition times were 46 ms in the directly detected dimension and 12.5 ms in the indirectly detected dimension. The CRIPT transfer delay was varied and the intensities of the first acquired FIDs were compared in order to determine the optimal delay, which was found to be 5 ms (for the dependence of CRIPT transfers on correlation times see Fig. 5 in [214]).

2.4.7 NMR experiments for structure determination of GAG oligosaccharides

1D proton spectra of oligosaccharides were acquired at 298 K, at different pH values and with different degrees of water suppression, depending on the size of the residual water signal after repeated freeze-drying and resuspension in D_2O.

The general approach for the determination of oligosaccharide sequences was largely based on proton 1D spectroscopy because these substances were only available at natural abundance of ^{13}C. In 1D spectra soft water supression was used to suppress the residual HDO signal without affecting oligosaccharide signals. The resonance frequencies obtained from this experiment were then used for selective excitation of protons in 1D TOCSY and 1D COSY experiments [226]. Since the hexose rings in GAG-oligosaccharides are separated by glycosidic linkages, which interrupt the J-coupling based TOCSY transfer, this strategy allowed the characterisation of one hexose at a time; occasionally peak overlap would result in simultaneous excitation of two hexose rings. By varying the TOCSY mixing times between 60 and 120 ms and thereby modulating the transfer efficiency, the proton assignments of individual hexose rings were obtained. The length and power of the selective pulse were varied in order to enhance its selectivity where peak overlap was a problem. Comparison of chemical shifts and J-couplings with literature values [227] aided the proton assignment and the analysis of the sulfation pattern. In the second step of the GAG characterisation, 1D ROESY and

NOESY experiments [228, 229, 230] were used to establish the connectivity of individual hexose rings. This step was facilitated by the fact that heparin is an unbranched oligosaccharide, limiting the number of NOEs between different rings. In cases of severe overlap in the proton dimension or inefficient TOCSY transfer, natural abundance 2D ^{13}C,H-HSQC spectra were collected to help the assignment by the addition of ^{13}C chemical shift information. In case of the fully sulfated hexasaccharide, dp6C, a 2D homonuclear TOCSY spectrum was also acquired [231].

2.5 Binding studies

2.5.1 Preparation of fluorescent GAGs for gel mobility shift assay

Fluorescent labelling of purified oligosaccharides was achieved using 2-aminoacridone (AMAC) essentially as described in [232]. Freeze-dried oligosaccharides were dissolved in 40 µL of a solution of 85% (v/v) dimethylsulfoxide (DMSO) and 15% (v/v) acetic acid containing 0.1 M AMAC. The mixture was incubated at room temperature for 15 min before 40 µL of 1 M aqueous sodium cyanoborhydride was added. Further incubation at 37°C for 16 hrs followed after which the reaction mixture was freeze-dried. In order to separate labelled oligosaccharides from unreacted AMAC, the dried mixture was re-dissolved in 5-10 mL pre-chilled (-20 °C) ethanol and left at -20°C for 30 min before centrifugation at 12,000 rpm for 5 min. The ethanol was discarded and the procedure repeated until the superantant was colourless and did not emit under UV-excitation, signalling that all excess AMAC had been removed.

2.5.2 Gel mobility shift assay

The Gel Mobility Shift Assay (GMSA) was conducted as described in [233]. Protein was buffer exchanged to phosphate buffered saline (PBS) and mixed with fluorescently labelled oligosaccharides at different concentrations to a total volume of 10 µL PBS containing 25% (v/v) glycerol and a trace of phenol red. The samples were incubated

at room temperature for 15 min and subsequently loaded on a 1% (w/v) agarose gel in 10 mM tris(hydroxymethyl)-aminomethane-HCl, pH 7.4, and 1 mM EDTA. Electrophoresis was performed for 8 - 15 min at 200 V in a horizontal agarose electrophoresis system using 40 mM Tris-HCl, pH 8.0, and 1 mM EDTA acid as electrophoresis buffer. The fluorescent oligosaccharides were visualised under UV-light using an excitation wavelength of 230 nm. The AMAC emission maximum is in the visible region, at 550 nm.

2.5.3 Heparin affinity chromatography

Buffer exchange to PBS was conducted as described for NMR samples. A sample containing both isoforms of fH∼7-8 (0.05 mM in 1 mL) was prepared and loaded onto a Poros 20HE heparin-affinity column (0.46 x 10 cm, Applied Biosystems) that had been equilibrated with PBS. A gradient reaching to 1 M NaCl was applied and the elution was followed by means of the A_{280}-value .

2.5.4 Isolation of sheep erythrocytes

Sheep erythrocytes were washed from defibrinated sheep blood (TCS Biosciences). 1 mL blood was transferred to a 15 mL falcon tube and diluted with 10 mL of buffer containing 20 mM HEPES, 145 mM NaCl, 0.1 % (w/v) gelatin (from pork skin, Fluka) and 10 mM EDTA, pH 7.3 at 25°C. The cell suspension was mixed gently and spun for 10 min at 500 g and 4°C. The supernatant and a slim white layer of leucocytes were discarded and the procedure was repeated two times. Subsequently, the washing buffer was replaced by the same buffer but without EDTA. Three more washing steps were done with the EDTA-free buffer and at 1,000 g. The washed cells were stored on ice up to one week. Prior to use the cells were washed once more in EDTA-free buffer to remove lysed cells and haemoglobin. For concentration measurements cells were lysed with distilled water and the absorbance at 412 nm (A_{412}) was measured. An A_{412} value of 0.87 at 1 cm pathlength corresponds to a cell concentration of $5*10^8$ cells/mL (personal information from Dr. Kerry Pangburn, Complement Technology). EDTA in the first washing buffer

had the function of removing calcium ions from the cell suspension in order to inhibit the calcium-dependent classical pathway during the haemolytic assay.

2.5.5 Haemolytic complement assay

This assay was originally outlined in [95] and applied to investigations on the recombinantly expressed C-terminus of fH by [69]. 20 µL of normal human serum (NHS) were mixed with 1.25 µL 0.1 M MgEGTA (1:1 $MgCl_2$ and ethylene glycol-bis(2-aminoethyl)-N,N,N',N'-tetraacetic acid), 2.5 µL sheep erythrocytes stock solution and 1.25 µL haemolytic assay buffer (20 mM HEPES, 141 mM NaCl, 5 mM MgEGTA, 0.1% (w/v) gelatin, pH 7.3). EGTA had the function to selectively bind calcium, not magnesium (a cofactor to factor B) from NHS. The concentration of sheep erythrocytes in the cell stock was adjusted to give an A_{412} of 0.5-0.6 units at the conditions used for the assay. In practice, 2.5 µL of different concentration of cell solutions were lysed through addition of 197.5 µL distilled water. 100 µL of the lysed solutions were transferred to separate wells of a 96 well plate and the A_{412} was measured for each well. A cell stock yielding a A_{412} of 0.5-0.6 units in this setup was then chosen for the actual assay. To the 25 µL solution containing buffer, NHS and erythrocytes 25 µL of different concentrations of fH19-20 WT in haemolytic assay buffer was added. For comparison of several mutant fH19-20 proteins, a range of concentrations of fH19-20 WT (0.2-10 µM) was first used in order to determine the concentration at which 50% lysis was observed. This concentration was subsequently used with each of the mutant proteins for comparison with the WT. In the case of the fH19-20 mutant D1119G and proteins encompassing fH7 a range of protein concentrations was used (0.2-10 µM for D1119G and 0.2-67 µM for fH7 and encompassing constructs) because at the 50%-lysis value for fH19-20 WT no significant lysis was observed for any of these proteins. The 50 µL reaction mixture was incubated at 37°C for 20 min and quenched by addition of 150 µL cold quenching buffer (20 mM HEPES, 145 mM NaCl, 5 mM EDTA, pH 7.3). Remaining cells were spun down for 5 min at 1,500 g and 4°C in a round bottom 96 well plate and 100 µL of supernatant was transferred from each well to a fresh 96 well plate, which was then used for A_{412} readings.

Each experiment was repeated at least four times and a standard error was calculated for each sample. As a negative control a well was filled with the same reaction mixture but serum which had been heat-inactivated at 56°C for 30 min [234]. In practice, stock solutions containing buffer, cells and NHS were made up for a whole plate at a time and supplemented with the protein solutions in order to minimize pipetting errors.

2.6 GAG-protein cross linking

For cross linking of dp4C to fH∼7Y, a method generally based on [235] was used. 0.2 mg desalted dp4C was added to a mixture of 22 mM 1-Ethyl-3-[3-dimethylaminopropyl]-carbodiimide hydrochloride (EDC) and 30 mM sulfo-NHS (N-Hydroxysuccinimide) in 50 mM 2-(N-morpholino)ethanesulfonic acid (MES), pH 6.0. The total reaction volume was 1 mL. The mixture was incubated for 15 min at 25°C and excess reagent was removed by rapid passing through a PD-10 desalting column using the same buffer. Recovered activated dp4C was combined with 400 µL of 750 µM fH∼7Y (2.3 mg) for further incubation at 25°C for 3 hrs. Subsequently, the reaction mixture was passed over a 1 mL MonoS column equilibrated with 20 mM potassium phosphate pH 5.0 and eluted with a linear gradient to 20 mM potassium phosphate, 400 mM NaCl pH 5.0. Fractions were checked for the cross linked complex using matrix-assisted laser desorption/ionization-time of flight mass spectrometry (MALDI-TOF). Fractions which contained a species with the expected mass were found in the flow through, signalling changed electrostatic characteristics of the protein after cross linking. The fractions containing fH∼7Yxdp4C were combined, freeze-dried and desalted using two 1 mL HiTrap desalting columns (Amersham) in a row. 20 mM potassium phosphate, pH 6.0 and a flowrate of 3 mL/min was used in this step. The injection volume was 0.5 mL, and the A_{280} was monitored to follow the protein elution. Fractions containing the cross linked complex (fH∼7Yxdp4C) were identified using MALDI-TOF mass spectrometry and subjet to dialysis into 50 mM ammonium acetate for tandem mass spectrometry (MS-MS).

2.6.1 Trypsin digestion

Prior to digestion with trypsin, fH∼7Y (87 µL, 105 µM) and fH∼7Yxdp4C (100 µL, 91 µM) were incubated with 8 M urea and 250 mM DTT at 75°C for one hour and subsequently with 125 mM iodoacetamide for 30 min in the dark at room temperature. To remove the denaturing reagents the samples were subject to overnight dialysis into 50 mM ammonium acetate, pH 7.4. For the digest, trypsin (Promega, sequencing grade) was pre-incubated at 30 °C for 15 min and 5 µg added to each sample. A small amount of trypsin was subject to autodigestion for comparison. All three samples were incubated for 6 hours at 37 °C and subject to mass spectrometry.

2.6.2 Microwave assisted TFA digestion

Since conventional trypsin digestion did not prove successful for the cross linked complex, a more radical approach was followed to obtain peptides from the complex [236]. 30 µL of 80 µM fH∼7Yxdp4C was mixed with 100 mM Dithiothreitol (DTT) and 20% (v/v) trifluoro acetic acid (TFA) in a 1.5 mL Eppendorf container. The sample was placed within a glass beaker in a domestic microwave along with a second glass beaker containing roughly 100 mL water (for absorbance of excess radiation). The sample was heated at 950 Watt for 6 min in 2 min steps. After each step, the sample was opened to release pressure and avoid violent opening of the container during microwaving. The sample was then subject to MALDI-TOF mass spectrometry. For peptide analysis the web server ProSight PTM (https://prosightptm.scs.uiuc.edu/) was used.

2.7 Media and buffers

YPD-Agar for growth of *P. pastoris*:
1% (w/v) yeast extract
2% (w/v) peptone
2% (w/v) *D*-glucose

1.5% (w/v) agar

BMG (buffered minimal glycerol) for *P. pastoris* growth:
100 mM potassium phosphate pH 6.0
1.34% (w/v) yeast nitrogen base (Sigma)
4 x 10^{-5}% (w/v) biotin
1% (v/v) glycerol

BMM (buffered minimal methanol) for *P. pastoris* expression:
as BMG but no glyerol and 0.5% methanol

PTM1 trace salts for isotope labelled *P. pastoris* expression:
6.0 g cupric sulfate pentahydrate
0.08 g sodium iodide
3.0 g manganese sulfate monohydrate
0.2 g sodium molybdate dihydrate
0.02 g boric acid
20.0 g zinc chloride
65.0 g ferrous sulfate heptahydrate
0.5 g cobalt chloride
0.2 g biotin
5.0 mL sulfuric acid
in 1L distilled water

Buffers for erythrocyte isolation and haemolytic assay:

1st erythrocyte washing buffer: 20 mM HEPES
145 mM sodium chloride
10 mM EDTA

0.1 % (w/v) gelatin (from pork skin, Fluka)

pH 7.3 at 25°C.

2nd erythrocyte washing buffer:

20 mM HEPES

141 mM NaCl

10 mM EDTA

0.1% gelatin (w/v), pH 7.3

Haemolytic assay buffer:

20 mM HEPES

141 mM NaCl

5 mM MgEGTA

0.1% (w/v) gelatin

pH 7.3

Quenching buffer:

20 mM HEPES

145 mM NaCl

5 mM EDTA

pH 7.3

3 Oligosaccharide preparation and characterisation

3.1 Preparation of heparin-derived oligosaccharides

3.1.1 Differences between heparin and heparan sulfate and their relevance for this study

Heparin binding to fH was investigated because several studies conclude that the interaction of fH and heparan sulfate (HS) could be at the center of the host-recognition process of fH [69, 74]. HS is more heterogeneous and therefore isolation of pure species is more difficult than for heparin, making the latter a common replacement for HS. This approach seems reasonable on the basis of the common carbon skeleton. The main chemical difference between heparin and HS is the level of sulfation which is higher in heparin. Sulfated domains in HS are, however, more often involved in protein binding than unsulfated domains, lending justification for the use of heparin in place of the physiological HS.

It should be noted that the oligosaccharides used in this study were all desalted in 0.25 M ammonium bicarbonate buffer after ion exchange chromatography. This buffer was chosen due to the voltile nature of its components, most of which were lost in subsequent freezdrying steps with D_2O. However, due to this procedure the counterion for carboxyl and sulfate groups in the NMR samples was likely ammonium. The available literature values which are quoted in this chapter, however, were obtained mostly on

samples containing potassium or sodium counterions. This difference might explain a few resonances for which the chemical shifts were notably different from literature values.

3.1.2 Purification of heparin-derived oligosaccharides

Heparinase I-digested heparin was supplied by Jon Deakin (Paterson Institute, University of Manchester). An initial gel filtration yielded separation of disaccharide, tetrasaccharide etc. up to decasaccharide fractions (Fig. 3.1). The different size oligosaccharides are labelled as dp2, dp4, dp6 etc. throughout this work, where dp stands for degree of polymerisation.

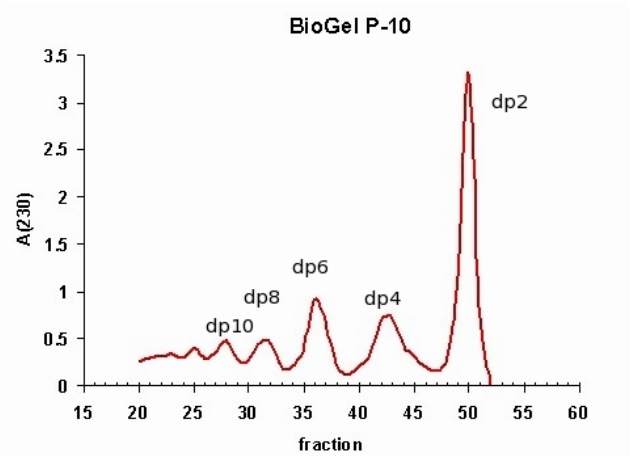

Figure 3.1: Gel filtration chromatogram of 50 mg bovine intestinal heparin digested with heparinase I. The labelling refers to disaccharide (dp2), tetrasaccharide (dp4) etc.

The subsequent strong anion exchange (SAX) chromatograms of different size fractions became more complex with increasing molecular weight as a consequence of the heterogeneous nature of heparin (Fig. 3.2, 3.3). Pure species were obtained for dp2, which is almost pure after the gel filtration, and dp4, but no attempt was made to isolate chemically pure species from crude dp8 and dp10 fractions. Fully sulfated dp6 was also prepared but no further hexasaccharide species were isolated. Differently charged oligosaccharides of defined length were labelled alphabetically according to the order in which they eluted from the anion exchange column, starting with the lowest salt con-

centration. The fully sulfated tetrasaccharide containing six sulfate and two carboxyl groups is therefore labelled dp4C throughout this report.

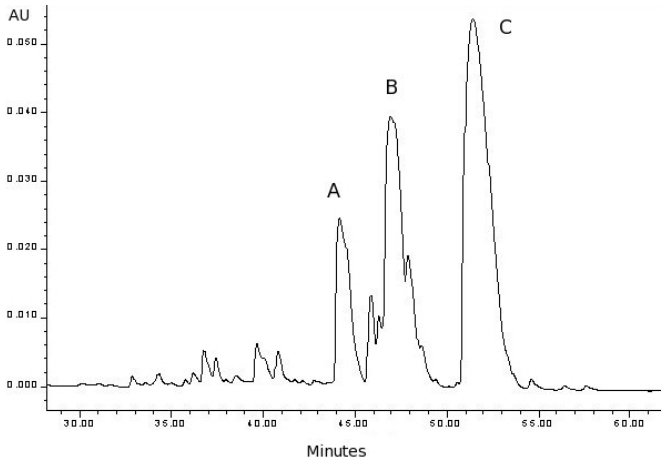

Figure 3.2: SAX chromatogram of heparin tetrasaccharides from heparinase I digestion of bovine intestinal heparin. The absorbance (AU) was measured at 230 nm. A, B and C refer to the three prevalent tetrasaccharide species with C being the fully sulfated tetrasaccharide.

For dp2 and dp4 fractions PD-10 desalting columns did not yield sufficient separation of the oligosaccharides from the salt as determined by the length of the 90° pulse on a 800 MHz NMR spectrometer (14.8 μs at a power level of -0.5 dB, when the default pulse length at this power level for a salt free sample was 8 μs). The dp2-dp6 fractions were therefore desalted again using a Sephadex column which yielded better separation. The desalted samples were freeze-dried and re-dissolved in D_2O several times before NMR spectra were collected for primary structure determination. It was found that the use of high-grade D_2O and a repetition of at least three freeze-drying steps were necessary for sufficient exchange of water with D_2O. The primary structures of dp4A-C obtained by the digestion of bovine intestinal heparin had been established prior to this work [237, 238]. The three tetrasaccharides in Fig. 3.4 are representative of the oligosaccharide variety obtained by enzymatic digestion of heparin, though percentages vary depending on the heparin source, i.e. organism and tissue. dp4C is fully sulfated

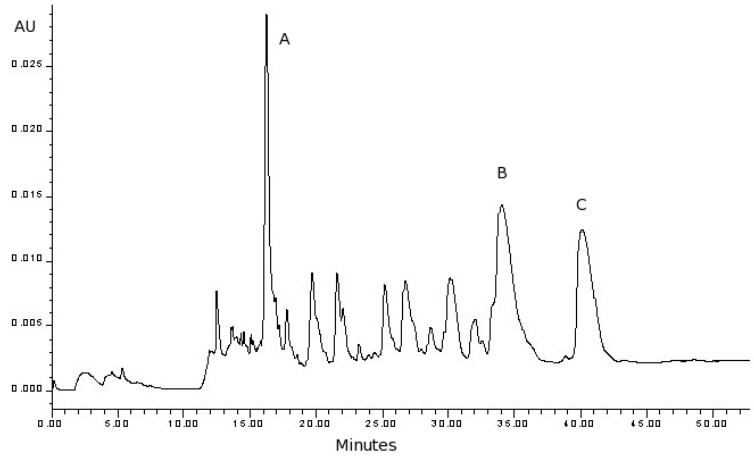

Figure 3.3: SAX chromatogram of heparin hexasaccharides from heparinase I digestion of bovine intestinal heparin. The absorbance (AU) was measured at 230 nm. In analogy to the tetrasaccharide nomenclature C refers to the fully sulfated hexasaccharide dp6C.

heparin tetrasccharide while dp4A and dp4B each contain one sulfate less. In dp4A, the sulfate group at position 6 of the reducing end glucosamine is missing, while in dp4B the internal IdoA2S is replaced by unsulfated GlcA.

Following purification, 1D ^{1}H NMR spectra were recorded to verify the primary structures of the obtained oligosaccharides. A detailed ^{1}H NMR analysis of the three main tetrasaccharide species obtained by heparinase I digestion of heparin had been carried out previously in our group and served as a control for fully sulfated tetrasaccharide dp4C, the most frequently used species (Fig. 3.5).

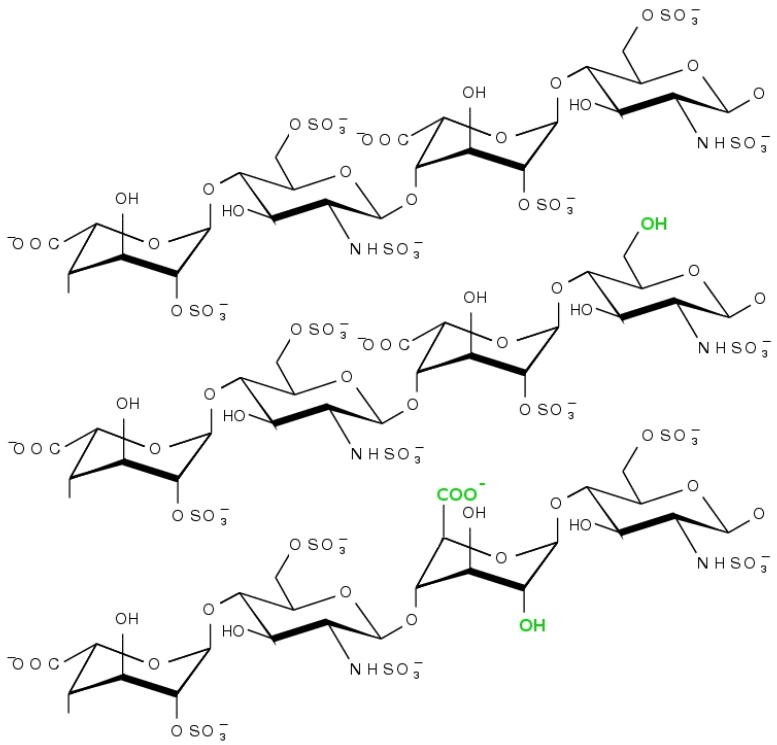

Figure 3.4: Different tetrasaccharides obtained from heparinase I digestion of bovine intestinal heparin. From top to the bottom dp4C (ΔUA2S-(1→4)-α-D-GlcNS6S-(1→4)-α-L-IdoA2S-(1→4)-α-D-GlcNS6S), dp4A (ΔUA2S-(1→4)-α-D-GlcNS6S-(1→4)-α-L-IdoA2S-(1→4)-α-D-GlcNS) and dp4B (ΔUA2S-(1→4)-α-D-GlcNS6S-(1→4)-β-D-GlcA-(1→4)-α-D-GlcNS6S). Varying structural elements are highlighted in green.

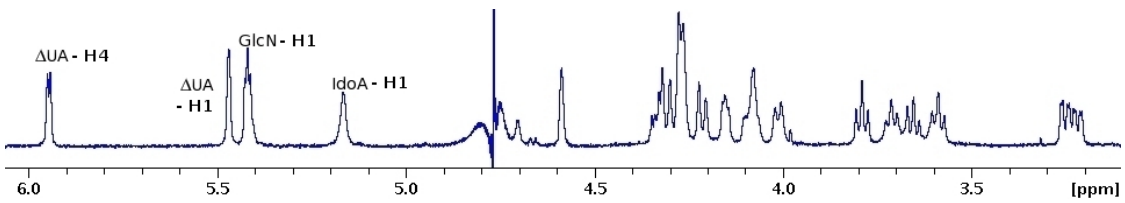

Figure 3.5: ^{1}H NMR spectrum of tetrasaccharide dp4C at 298 K and 600 MHz.

3.2 Characterisation of two new heparin tetrasaccharides

During a heparin digestion carried out by Liu Chong in our group (using heparin from porcine intestinal mucosa and heparinase I from Grampian enzymes) two new tetrasaccharide species were isolated that had not been present in previous heparin digestions, or only as very minor components (compare Fig. 3.6 with Fig. 3.2). The characterisation of these compounds became part of the presented work. The two tetrasaccharides, which eluted from a SAX column at lower salt concentrations than dp4A-C, were desalted, repeatedly freeze-dried and dissolved in D_2O, and subjected to NMR analysis at 800 MHz. The two species were labelled dp4D and dp4E.

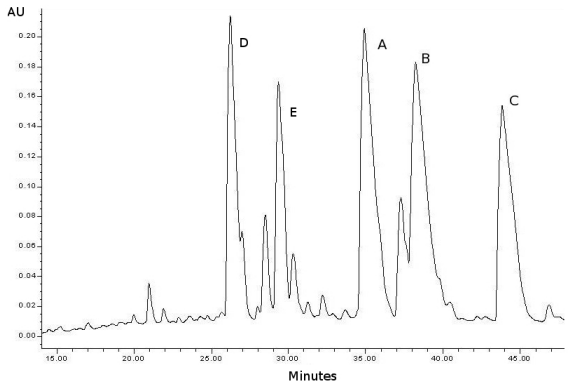

Figure 3.6: The SAX chromatogram of the tetrasaccharide fraction from porcine intestinal mucosa heparin digested by heparinase I showed two additional species D and E. The absorbance (AU) was measured at 230 nm. A higher flowrate was used than in Fig. 3.2.

3.2.1 Characterization of dp4D

The primary sequence of dp4D was determined by NMR to be ΔUA2S-(1→4)-α-D-GlcNS-(1→4)-α-L-IdoA2S-(1→4)-α-D-GlcNS (Fig. 3.7). It is presented at the start of this chapter to facilitate understanding of the subsequently presented spectra.

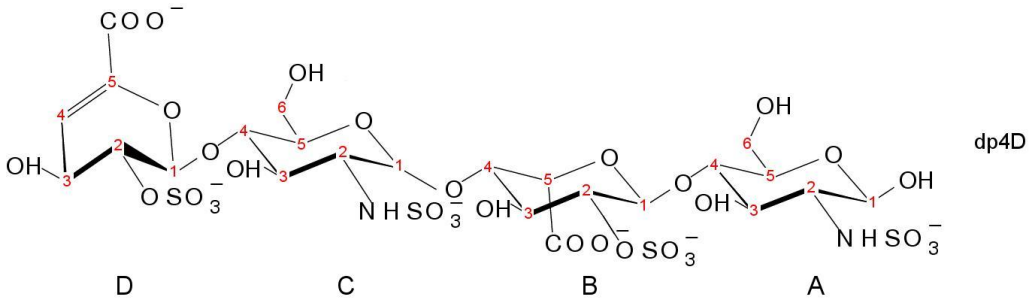

Figure 3.7: Scheme of ΔUA2S-(1→4)-α-D-GlcNS-(1→4)-α-L-IdoA2S-(1→4)-α-D-GlcNS (dp4D). Carbon positions are numbered in red.

Sufficient separation of the anomeric signals allowed efficient assignment of proton resonances of individual rings in this tetrasaccharide using 1D TOCSY experiments. Selective excitation of individual anomeric protons was followed by varying the length of the mixing time. Signals of all four rings were assigned in this way (Fig. 3.8).

Signals from the unsaturated D-ring appeared at positions corresponding to documented ΔUA2S frequencies. The small chemical shift of the B-ring anomeric proton indicated that this ring was IdoA rather than GlcA. The signals of protons H2-H4 in the B-ring were characteristic for IdoA2S, with the H5 peak likely suppressed along with the water signal. The chemical shifts and the coupling constants of the two remaining rings generally exhibited features that were in agreement with GlcNS. Significantly smaller chemical shifts of H5 and H6 protons in comparison with the protons in fully sulfated GlcNS6S rings suggested that both rings were missing sulfate groups at the C6 postion. Thus, analysis of 1D TOCSY spectra of dp4D yielded the primary sequence ΔUA2S-α-D-GlcNS-α-L-IdoA2S-α-D-GlcNS (Fig. 3.7). In order to establish which of the GlcNS rings was ring A and which one was ring C 1D ROESY spectra were recorded. The anomeric proton of the D- and B-rings were excited, respectively, and ROE transfers to protons of ring C (from D-H1) or ring A (from B-H1) were observed (Fig. 3.9). From the anomeric proton of the D-ring ROE transfers were observed to H2 of the same ring as well as to H4 and H6 in GlcNS-1. No transfer to protons from GlcNS-2 was observed in this spectrum, strongly indicating that GlcNS-1 is ring C. Similarly, ROE transfer

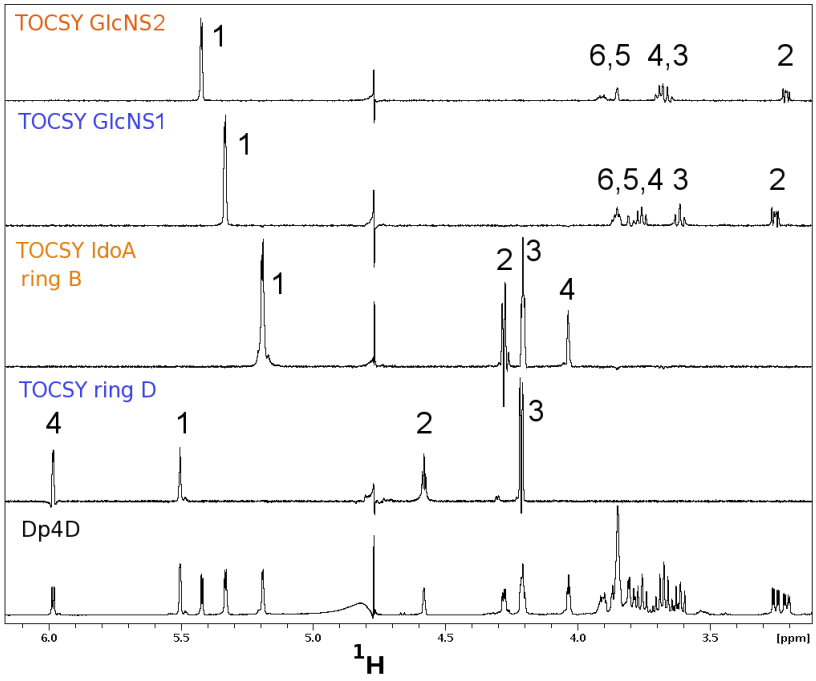

Figure 3.8: 1D ^{1}H NMR spectrum of dp4D (bottom) and 1D TOCSY spectra of the individual rings (α-D-GlcNS1, α-D-GlcNS2, B=α-L-IdoA, D=ΔUA from top to bottom, 120 ms mixing time). The B-H5 proton frequency is suppressed along with the water. However, TOCSY excitation at 4.77 ppm lead to transfer to all remaining protons of the B ring (not shown), indicating the position of B-H5.

from H1 in the B-ring to H2 and H3 of the same ring were observed and to H4 as well as H6 in GlcSN-2, confirming GlcNS-2 as the reducing end ring A.

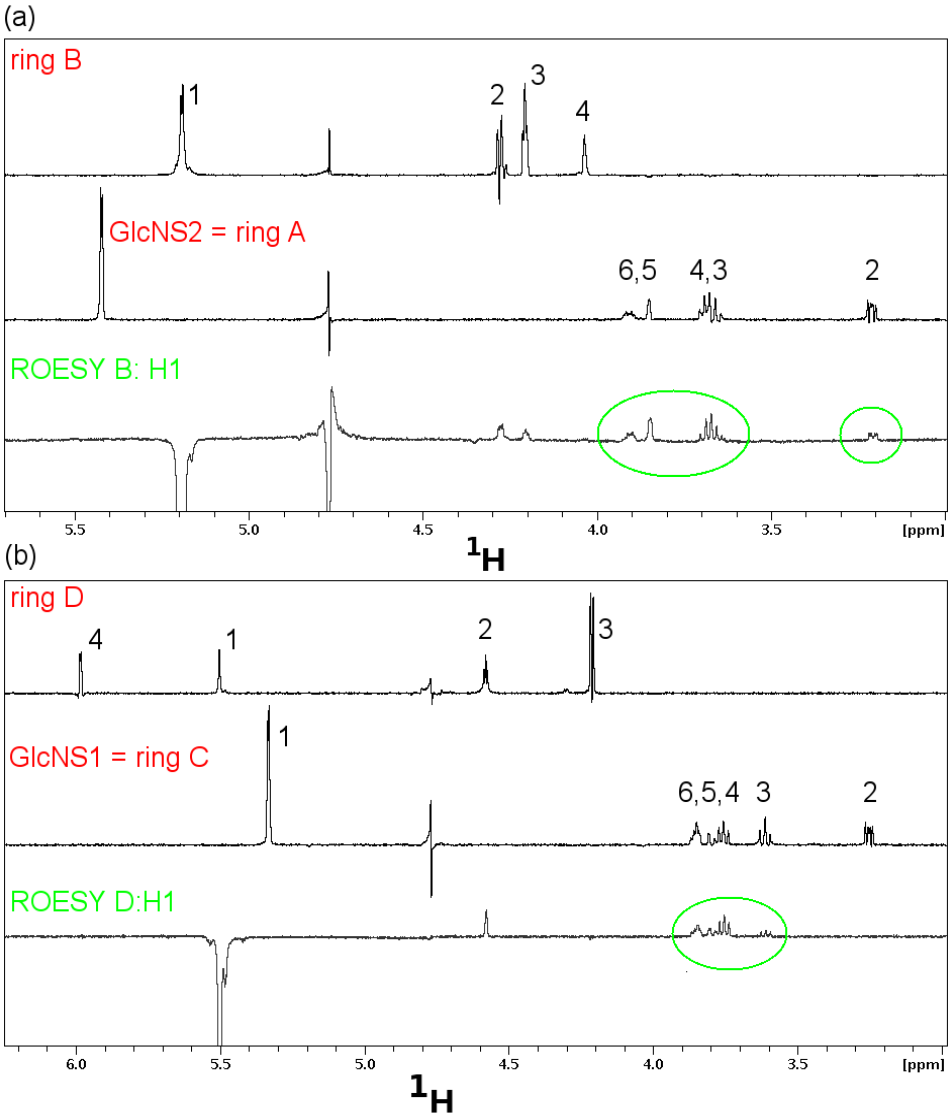

Figure 3.9: (a) ROE analysis of the linkage between ring B and A. Bottom: ROESY spectrum with excitation at 5.19 ppm (H1 of ring B) and 60 ms mixing time yielded ROE transfer to GlcNS2 (green circles), middle: TOCSY spectrum of GlcNS2, top: TOCSY spectrum of ring B. (b) ROE analysis of the linkage between ring D and C. Bottom: ROESY spectrum with excitation at 5.50 ppm (H1 of ring D) and 60 ms mixing time yielded ROE transfer to GlcNS1 (green circle), middle: TOCSY spectrum of GlcNS1, top: TOCSY spectrum of ring D.

To complete the assignment of dp4D through addition of ^{13}C chemical shift information, a ^{13}C natural abundance 2D ^{1}H, ^{13}C-HSQC spectrum was recorded. Cross peaks of the anomeric CH groups and the allylic CH group D4 were well separated from the remaining signals due to their high ^{1}H and ^{13}C chemical shifts. A1 is the only anomeric CH group which does not participate in a glycosidic linkage, resulting in an upfield shift of its ^{13}C resonance (Fig. 3.10).

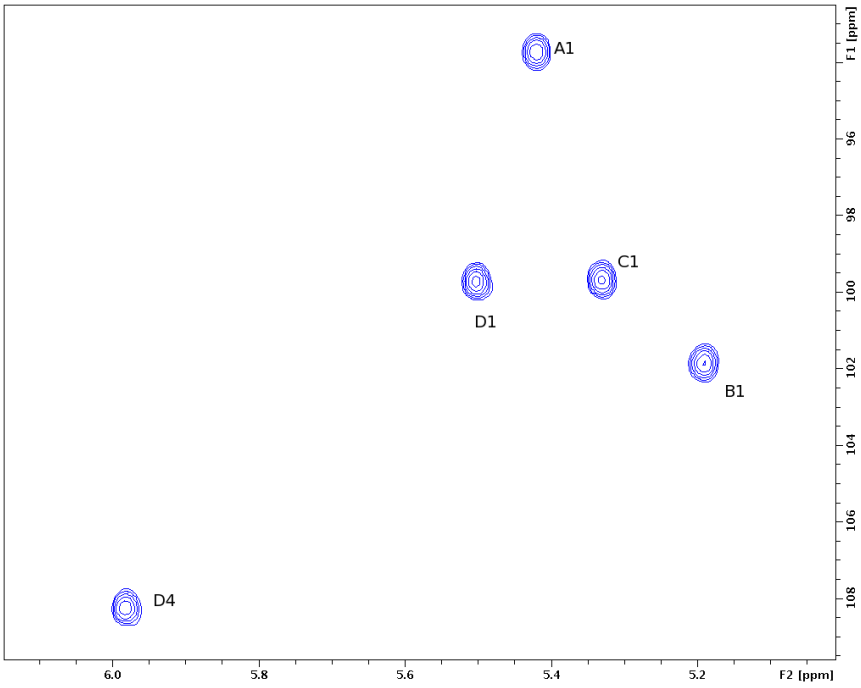

Figure 3.10: ^{1}H, ^{13}C-HSQC spectrum of dp4D (anomeric region). F2 (the directly detected dimension) corresponds to ^{1}H, F1 (the indirectly detected dimension) to ^{13}C.

The cross peaks of the remaining CH and CH_2 groups resonate between 3.0 and 5.0 ppm in the ^{1}H dimension and between 60 and 85 ppm in the ^{13}C dimension (Fig. 3.11). As expected for two GlcNS residues, the CH groups attributed to the A and C rings resonate at very similar frequencies in both dimensions, with the exception of the above mentioned anomeric groups. The low ^{1}H and ^{13}C chemical shifts of the A6 and C6 groups are additional indicators for the absence of sulfate groups at position C6 in both GlcNS rings. The HSQC spectrum also confirmed the existence of a strong coupling between protons H5 and H6 in both the A- and the C-ring which distorted the shape of

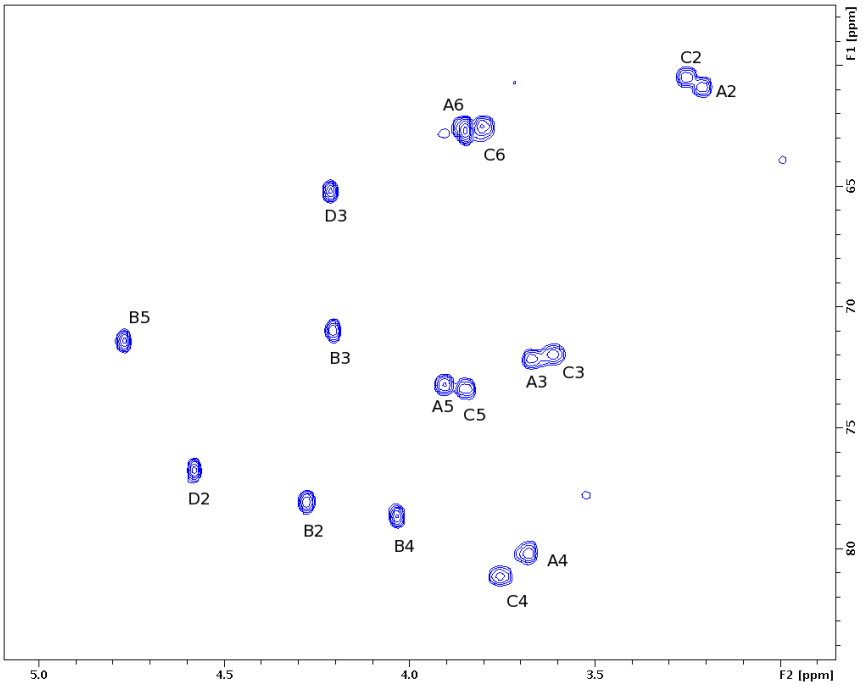

Figure 3.11: ^{1}H, ^{13}C-HSQC spectrum of dp4D (non-anomeric region). F2 (the directly detected dimension) corresponds to ^{1}H, F1 (the indirectly detected dimension) to ^{13}C.

proton multiplets seen in ^{1}H TOCSY spectra. As part of glycosidic linkages A4, B4 and C4 all show relatively high ^{13}C chemical shifts of around 80 ppm. This supports the H3 and H4 assignment in GlcNS rings A and C as deduced from the analysis of 1D TOCSY spectra (see above).

A comparison of the measured ^{1}H chemical shifts together with literature values is presented in Table 3.1. The data are in good agreement, supporting the analysis of the dp4D primary structure.

Table 3.1: Measured ^{1}H and ^{13}C chemical shifts in ppm for dp4D (dp4D (exp)) and ^{1}H literature values for the same pyranoses in similar chemical environments (Lit).

Resonance	^{1}H Lit [237]	^{1}H dp4D (exp)	Resonance	^{13}C dp4D (exp)
ΔUA2S-H1	5.50	5.50	ΔUA2S-C1	99.70
ΔUA2S-H2	4.61	4.58	ΔUA2S-C2	76.73
ΔUA2S-H3	4.31	4.21	ΔUA2S-C3	65.19
ΔUA2S-H4	5.98	5.98	ΔUA2S-C4	108.23
αGlcNS-2-H1	5.43	5.33	αGlcNS-2-C1	99.67
αGlcNS-2-H2	3.22	3.25	αGlcNS-2-C2	60.46
αGlcNS-2-H3	3.67	3.61	αGlcNS-2-C3	71.95
αGlcNS-2-H4	3.70	3.76	αGlcNS-2-C4	81.12
αGlcNS-2-H5	3.92	3.85	αGlcNS-2-C5	73.35
αGlcNS-2-H6	3.86	3.86	αGlcNS-2-C6	62.50
αGlcNS-2-H6’	3.86	3.85		
αIdoA2S-H1	5.20	5.19	αIdoA2S-C1	101.83
αIdoA2S-H2	4.29	4.28	αIdoA2S-C2	78.04
αIdoA2S-H3	4.22	4.21	αIdoA2S-C3	70.96
αIdoA2S-H4	4.07	4.03	αIdoA2S-C4	78.65
αIdoA2S-H5	4.80	4.77	αIdoA2S-C5	71.39
αGlcNS-1-H1	5.43	5.40	αGlcNS-1-C1	93.73
αGlcNS-1-H2	3.22	3.21	αGlcNS-1-C2	60.86
αGlcNS-1-H3	3.67	3.67	αGlcNS-1-C3	72.13
αGlcNS-1-H4	3.70	3.68	αGlcNS-1-C4	80.19
αGlcNS-1-H5	3.92	3.91	αGlcNS-1-C5	73.20
αGlcNS-1-H6	3.86	3.85	αGlcNS-1-C6	62.81
αGlcNS-1-H6’	3.86	3.85		

3.2.2 Characterization of dp4E

The primary sequence of dp4E was found to be ΔUA2S-(1→4)-α-D-GlcNS-(1→4)-β-D-GlcA-(1→4)-α-D-GlcNS6S (Fig. 3.12). It is presented at the start of this chapter to facilitate understanding of the subsequently presented spectra.

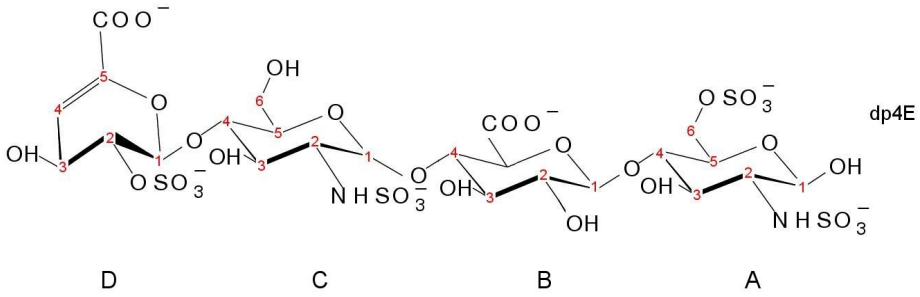

Figure 3.12: Scheme of ΔUA2S-(1→4)-α-D-GlcNS-(1→4)-β-D-GlcA-(1→4)-α-D-GlcNS6S (dp4E). Carbon positions are numbered in red.

The strategy employed for the characterisation of dp4E was generally the same as for dp4D. First, a 1D ^{1}H NMR spectrum and a series of 1D TOCSY spectra for the individual rings were recorded (Fig. 3.13), followed by a more detailed analysis of the individual rings using 1D TOCSY spectra with different mixing times and different excitation frequencies as explained for dp4D. The linkages were again established using 1D ROESY experiments. The absence of a peak at 5.2 ppm indicated that this compound did not contain an IdoA ring but the IdoA C5-epimer, GlcA, the anomeric proton of which resonates at a lower frequency. In GlcA the anomeric proton resonates at around 4.6 ppm, partially overlapping with the H2 of the unsaturated ring D (Fig. 3.13).

The 1D proton spectrum of dp4E displays the β-anomer of the reducing end ring as a minor species (Fig. 3.14). It is characterised by the high field shift of the anomeric proton in this ring (from 5.45 ppm in the α-anomer to 4.7 ppm in the β-anomer). The presence of weak signals from H6 and H6' at relatively high frequencies for the β-anomer suggest that the reducing ring is a GlcNS6S residue, but ROESY spectra were recorded to further test this hypothesis.

Analysis of 1D TOCSY spectra of dp4E confirmed the assumption that the B-ring was

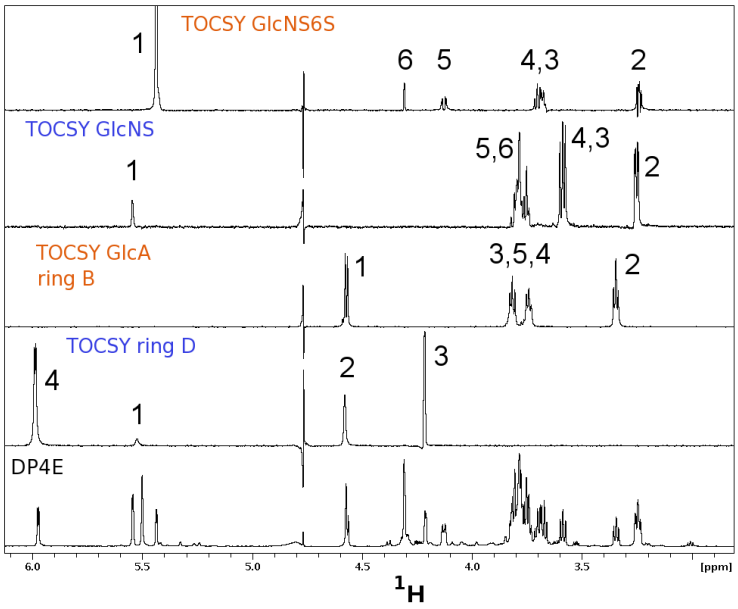

Figure 3.13: 1D ^{1}H spectrum of dp4E (bottom) and 1D TOCSY spectra of the individual rings (A, C, B, D from top to bottom) acquired using a 120 ms mixing time. The B5 proton signal is suppressed along with the water.

a GlcA, not a IdoA residue. Furthermore, due to the low chemical shifts of the H5 and H6 protons in one of the GlcN rings, it became clear that either ring A or ring C would have to be a GlcNS residue, the other one a GlcNS6S residue. The D-ring was found to be the expected ΔUA2S residue that is inherent to oligosaccharides produced by enzymatic digestion with heparinase. In order to establish which of the rings was the GlcNS residue 1D ROESY spectra were recorded. The anomeric proton of the D- and B-rings were excited, respectively, and ROE transfers to protons of ring C (from D-H1) or ring A (from B-H1) were observed (Fig. 3.15). From the anomeric proton of the D-ring ROE transfers were observed to H5 and H6 in the GlcNS residue, indicating that GlcNS is ring C. ROE transfers from H1 in the B-ring were observed to H4 and H3 as well as H6 in GlcSN6S, confirming it as the reducing end ring A. Thus, the sequence of dpE was found to be ΔUA2S-(1→4)-α-D-GlcNS-(1→4)-β-D-GlcA-(1→4)-α-D-GlcNS6S.

A detailed comparison of the resonance frequencies with literature values is presented in Table 3.2. Good agreement between experimental and literature values was observed.

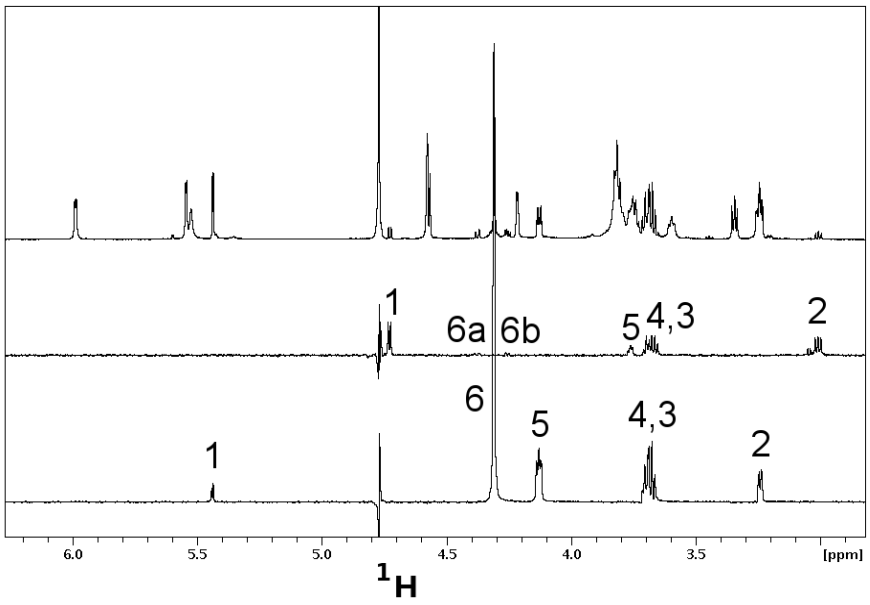

Figure 3.14: 1D ^{1}H TOCSY spectra of ring A of dp4E. Excitation at 5.435 ppm, 4.307 ppm and 3.007 ppm from bottom, mixing time 120 ms. The proton at 3.007 ppm belongs to the β-anomer of ring A. Top: 1D spectrum at slightly lower pH than Fig. 3.13 displays clearer peaks of the β-anomer.

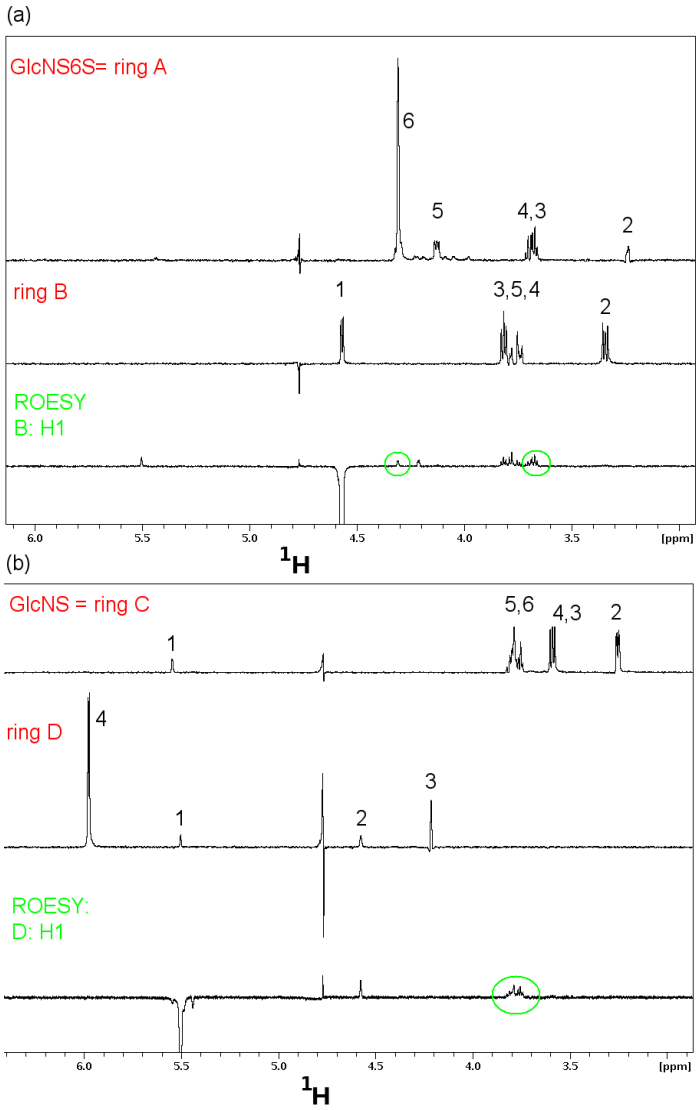

Figure 3.15: (a) ROE analysis of the linkage between rings B and A. Top: TOCSY spectrum of the GlcNS6S residue, middle: TOCSY spectrum of ring B, bottom: ROESY spectrum with excitation at 4.57 ppm (H1 of ring B) and 60 ms mixing time yielded ROE transfer to GlcNS6S (green circles), establishing this residue as ring A. (b) ROE analysis of the linkage between rings D and C. Top: TOCSY spectrum of the GlcA residue, middle: TOCSY spectrum of ring D, bottom: ROESY spectrum with excitation at 5.49 ppm (H1 of ring D) and 60 ms mixing time yielded ROE transfer to GlcA (green circles), establishing this residue as ring C.

Table 3.2: Measured ^{1}H and ^{13}C chemical shifts in ppm for dp4E (dp4E (exp)) and ^{1}H literature values for the same pyranoses in similar chemical environments (Lit).

Resonance	^{1}H Lit [237]	^{1}H dp4E (exp)	Resonance	^{13}C dp4E (exp)
ΔUA2S-H1	5.50	5.50	ΔUA2S-C1	99.60
ΔUA2S-H2	4.61	4.57	ΔUA2S-C2	76.74
ΔUA2S-H3	4.31	4.21	ΔUA2S-C3	65.15
ΔUA2S-H4	5.98	5.97	ΔUA2S-C4	108.18
αGlcNS-2-H1	5.43	5.54	αGlcNS-2-C1	100.26
αGlcNS-2-H2	3.22	3.25	αGlcNS-2-C2	60.54
αGlcNS-2-H3	3.67	3.59	αGlcNS-2-C3	72.01
αGlcNS-2-H4	3.70	3.75	αGlcNS-2-C4	80.86
αGlcNS-2-H5	3.92	3.78	αGlcNS-2-C5	73.29
αGlcNS-2-H6	3.86	3.79	αGlcNS-2-C6	62.39
αGlcNS-2-H6’	3.86	3.81		
βGlcA-H1	4.59	4.57	βGlcA-C1	105.77
βGlcA-H2	3.37	3.35	βGlcA-C2	75.55
βGlcA-H3	3.84	3.81	βGlcA-C3	78.60
βGlcA-H4	3.77	3.74	βGlcA-C4	79.86
βGlcA-H5	3.81	3.78	βGlcA-C5	79.16
αGlcNS6S-1-H1	5.43	5.44	αGlcNS6S-1-C1	93.65
αGlcNS6S-1-H2	3.25	3.24	αGlcNS6S-1-C2	60.13
αGlcNS6S-1-H3	3.68	3.70	αGlcNS6S-1-C3	72.02
αGlcNS6S-1-H4	3.73	3.68	αGlcNS6S-1-C4	81.38
αGlcNS6S-1-H5	4.12	4.13	αGlcNS6S-1-C5	70.76
αGlcNS6S-1-H6	4.29	4.31	αGlcNS6S-1-C6	69.33
αGlcNS6S-1-H6’	4.35	4.31		

The here presented NMR studies led to the conclusion that both dp4D and E contained four sulfate groups. As in the case of the two tetrasaccharides with 5 sulfates, dp4A and B, the later eluting compound was found to contain a GlcA rather than an IdoA ring: dp4B binds with higher affinity to the SAX column than dp4A, and dp4E binds stronger than dp4D. A possible reason for the sudden occurance of these two less sulfated species in this particular heparin digestion could be a sulfatase impurity in the heparinase batch used for this digestion. The heparin which was used in this particular digestion originated from porcine, not bovine like the previously used heparin. The different source is, however, unlikely to be the reason for the observed differences as both porcine and bovine heparin are common sources for heparin oligosaccharide preparations and such a difference has not been described in the literature.

3.3 Characterization of fully sulfated hexasaccharide

Fully sulfated hexasaccharide was isolated by purifying the hexasaccharide fraction from the initial gel filtration (Fig. 3.1) using SAX chromatography (Fig. 3.3). Only the tightest binding fraction, dp6C, was kept since it was assumed that it would be the fully sulfated species. Homonuclear 1D and 2D NMR was employed to verify this assumption. The 1D ^{1}H spectrum showed the expected anomeric protons for three GlcN residues and the unsaturated non-reducing end ring (Fig. 3.16). Only one anomeric signal was found at the resonance frequency of the IdoA, but a 1D TOCSY (not shown) clarified that this anomeric signal was indeed made up of the signals of two different IdoA rings. The unsaturated ring and the two IdoA rings were identified in a 2D TOCSY spectrum (Fig. 3.17).

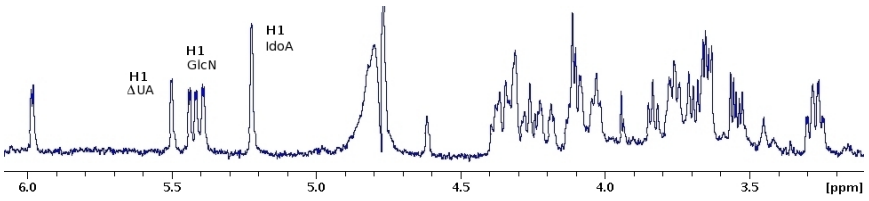

Figure 3.16: 1D ^{1}H NMR spectrum of hexasaccharide dp6C at 298 K and 600 MHz.

The sequence in which the rings are linked was established by comparison of 2D TOCSY and 2D ROESY spectra. It was found that the anomeric protons of the GlcN residues posed a useful starting point because two of the GlcN H1 strips in the ROESY showed cross peaks with the H3 and H4 resonances of the IdoA residues. This link also allowed grouping of the resonances of the two IdoA residues, whose H3, H4 and H2 protons resonate at different frequencies. The GlcN H1 which resonates as the middle of the three GlcN H1 signals was thus linked to the IdoA with the higher resonance frequency for H4 and the lower resonance frequency for H3 (Fig. 3.18). The H2 of this IdoA (IdoA1) was found to overlap with the H3 signal of the unsaturated ring. Analogously, the GlcN H1 with the lowest resonance frequency among the three anomeric GlcN protons showed cross peaks with the H3 and H4 resonances of the second IdoA ring (IdoA2) (Fig. 3.18). In this IdoA residue the H5 proton could also be assigned.

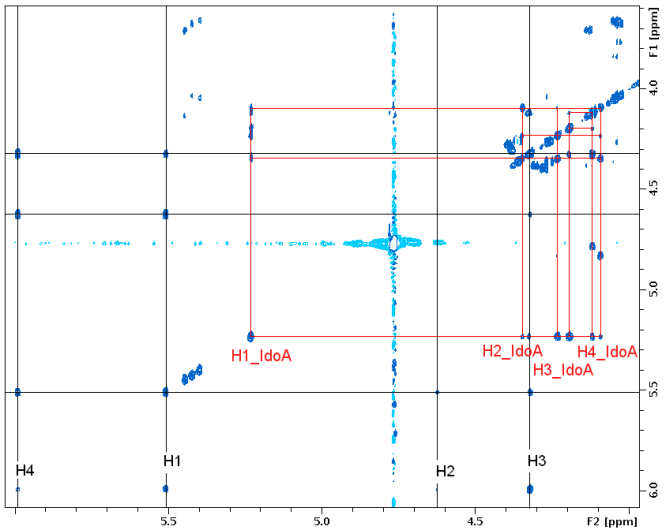

Figure 3.17: Homonuclear ^{1}H 2D TOCSY spectrum of fully sulfated heparin derived hexasaccharide. Proton strips of the unsaturated ring D are labelled black. The peaks that are linked by red lines belong to two IdoA residues with overlapping anomeric proton shifts. The IdoA H5 protons are supressed along with the water signal. F2 corresponds to the ^{1}H, F1 to the ^{13}C dimension, respectively.

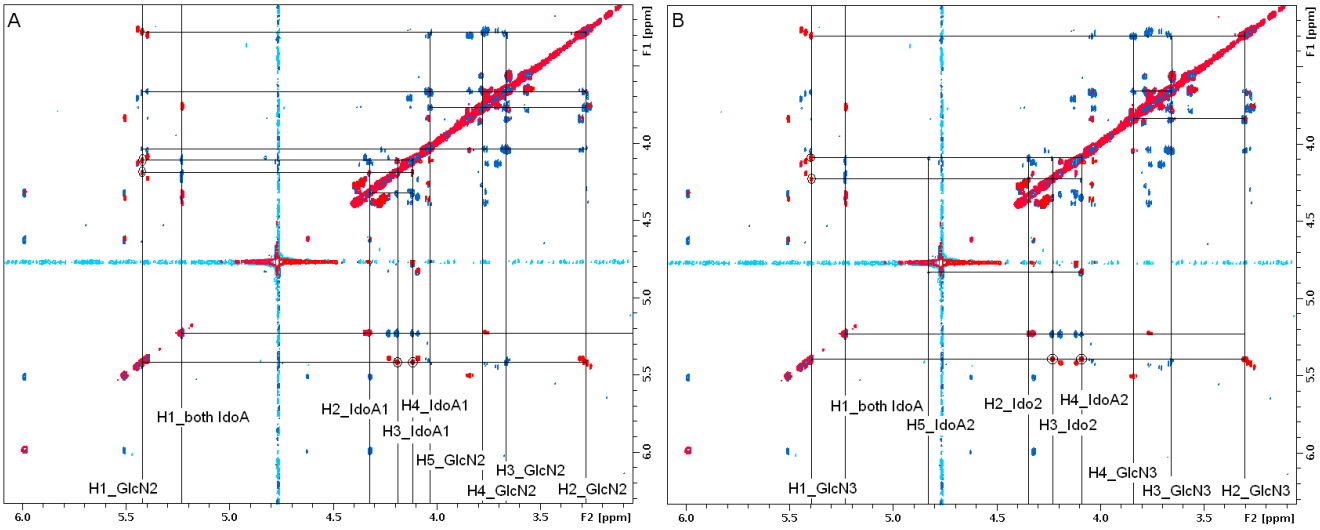

Figure 3.18: Homonuclear ^{1}H 2D TOCSY (blue) and ROESY (red) of fully sulfated hexasaccharide. The ROESY peaks which were used to establish the link between IdoA1 and GlcN2 (A) and between IdoA2 and GlcN3 (B) are highlighted (circles). F2 corresponds to the ^{1}H, F1 to the ^{13}C dimension, respectively.

The remaining GlcN residue was identified as the reducing monosaccharide because its anomeric proton (the one with the highest frequency) did not show cross peaks with protons from other rings (Fig. 3.19). The anomeric proton of the unsaturated non-reducing end ring showed cross peaks with the H4 resonance of the GlcN3 ring (circled cross peaks in Fig. 3.19).

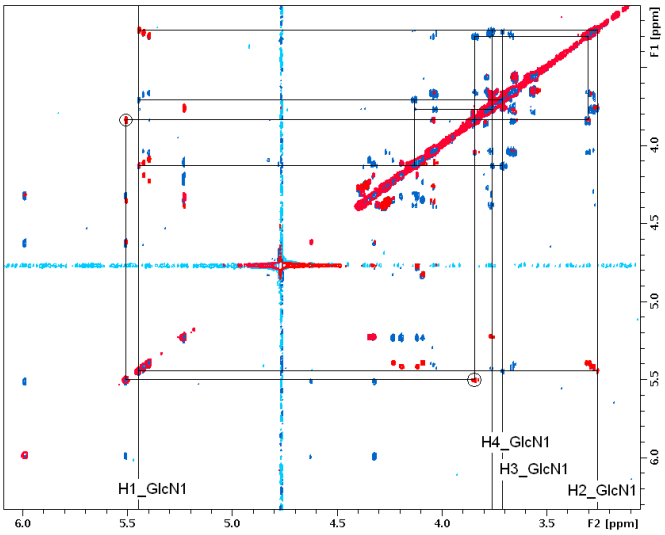

Figure 3.19: Homonuclear ^{1}H 2D TOCSY (blue) and ROESY (red) of fully sulfated hexasaccharide. H1 of GlcN1 shows ROESY cross peaks (red) only with protons which belong to the same ring and are present in the TOCSY spectrum (blue). The anomeric proton of the unsaturated ring is linked to the H4 of GlcN3 in the ROESY spectrum (circles, for assignment of GlcN3 see Fig. 3.18). F2 corresponds to the ^{1}H, F1 to the ^{13}C dimension, respectively.

Comparison of the chemical shifts of the H2-IdoA groups with literature values [239] confirmed that both IdoA1 and IdoA2 were IdoA2S residues. The six H6-GlcNS resonances cluster between 4.2 and 4.4 ppm, which confirmed sulfation in the 6 position for all three rings. The H2-ΔUA resonance frequency was above 4.6 ppm, identifying the unsaturated non-reducing end ring as ΔUA2S. N-sulfation of the GlcN-rings is more difficult to judge from NMR spectra because of the larger distance from the CH skeleton. The high salt concentration which was necessary for elution of the hexasaccharide does, however, suggest that these positions in the GlcN rings were sulfated as well, so that

the hexasaccharide was found to be fully sulfated (ΔUA2S-(1→4)-α-D-GlcNS6S-(1→4)-α-L-IdoA2S-(1→4)-α-D-GlcNS6S-(1→4)-α-L-IdoA2S-(1→4)-α-D-GlcNS6S). This compound was further titrated into fH∼7-8Y and monitored by ^{1}H, ^{15}N-HSQC spectra (see below). A detailed comparison of ^{1}H chemical shifts of dp6C with literature values is presented in Table 3.3.

Generally, the obtained resonances are in good agreement with the literature. An exception are the H5 resonances in both IdoA2S rings. Another literature source, however, reports the IdoA2S H5 resonance in dp4C to be 4.765 [237], which is much closer to the value obtained for the two IdoA2S H5 resonances in dp6C in this work. ^{1}H resonances do not seem to be very sensitive to the length of the oligosaccharide under observation but rather to the degree of sulfation. Thus, it is reasonable to assume that specific resonances found in dp4C and dp6C residues do not differ significantly and that the dp4C resonance frequencies listed in [237] are relevant for this comparison.

Table 3.3: Measured dp6C (ΔUA2S-(1→4)-α-D-GlcNS6S-(1→4)-α-L-IdoA2S-(1→4)-α-D-GlcNS6S-(1→4)-α-L-IdoA2S-(1→4)-α-D-GlcNS6S) resonance frequencies in ppm (dp6C (exp)) and literature values (Lit).

Resonance	Lit [239]	dp6C (exp)	Resonance	Lit [239]	dp6C (exp)
ΔUA2S-H1	5.519	5.501	αGlcNS6S-2-H1	5.439	5.423
ΔUA2S-H2	4.612	4.616	αGlcNS6S-2-H2	3.289	3.30
ΔUA2S-H3	4.317	4.315	αGlcNS6S-2-H3	3.645	3.563
ΔUA2S-H4	6.053	5.982	αGlcNS6S-2-H4	3.715	3.672
αGlcNS6S-3-H1	5.402	5.399	αGlcNS6S-2-H5	3.956	4.844
αGlcNS6S-3-H2	3.273	3.301	αGlcNS6S-2-H6	n.d	4.045
αGlcNS6S-3-H3	3.661	3.657	αGlcNS6S-2-H6'	n.d	4.262
αGlcNS6S-3-H4	3.840	3.835	αIdoA2S-1-H1	5.23	5.222
αGlcNS6S-3-H5	3.976	4.050	αIdoA2S-1-H2	4.30	4.344
αGlcNS6S-3-H6	n.d.	4.263	αIdoAS2-1-H3	4.27	4.230
αGlcNS6S-3-H6'	n.d.	4.371	αIdoAS2-1-H4	4.14	4.089
αIdoA2S-2-H1	5.17	5.222	αIdoAS2-1-H5	5.12	4.829
αIdoA2S-2-H2	4.33	4.325	αGlcNS6S-1-H1	5.446	5.446
αIdoAS2-2-H3	4.33	4.194	αGlcNS6S-1-H2	3.257	3.262
αIdoAS2-2-H4	4.10	4.116	αGlcNS6S-1-H3	3.735	3.705
αIdoAS2-2-H5	5.17	4.778	αGlcNS6S-1-H4	3.699	3.763
			αGlcNS6S-1-H5	4.15	4.132
			αGlcNS6S-1-H6	4.28	4.303
			αGlcNS6S-1-H6'	4.28	4.376

4 Studies on fH∼7 and fH∼7-8

4.1 Protein production

4.1.1 Fermentation and purification of ^{13}C, ^{15}N-fH∼7-8Y

In order to collect triple-resonance NMR experiments for the backbone assignment experiments a 0.7 mM sample of ^{13}C,^{15}N-fH∼7-8Y NMR sample was produced. The *P. pastoris* clone (strain KM71H, expression vector pPICZα) containing the DNA sequence coding for fH∼7-8Y was grown and induced in a Bioflow3000 fermentor (Fig. 4.1). The fermentation temperature was set to 30 °C before the induction phase but due to technical problems with the water supply the temperature was fluctuating until the induction phase, when the water regulation was altered and the temperature was set to 15 °C. Small increases in temperature after each methanol feed reflected the increasing metabolic activity of the cells. The pH was initially set to 5.0 and was maintained by a control circuit. This circuit could, however, only act by adding 2 M KOH to correct for the metabolic production of acid. As can be seen from the fermentation record (Fig. 4.1), injection of ^{13}C-glucose (injection 1) and ^{13}C-methanol (injections 2-8) led to production of acid which was compensated for by the automated addition of 2 M KOH. As a result of a relatively slow pH control circuit the pH had a tendency to overshoot through the addition of too much KOH. Occasionally a small amount of concentrated phosphoric acid was added manually to the batch in order to decrease the pH to the setpoint before induction. As a positive side effect the addition of phosphoric acid increased the buffer capacity of the batch.

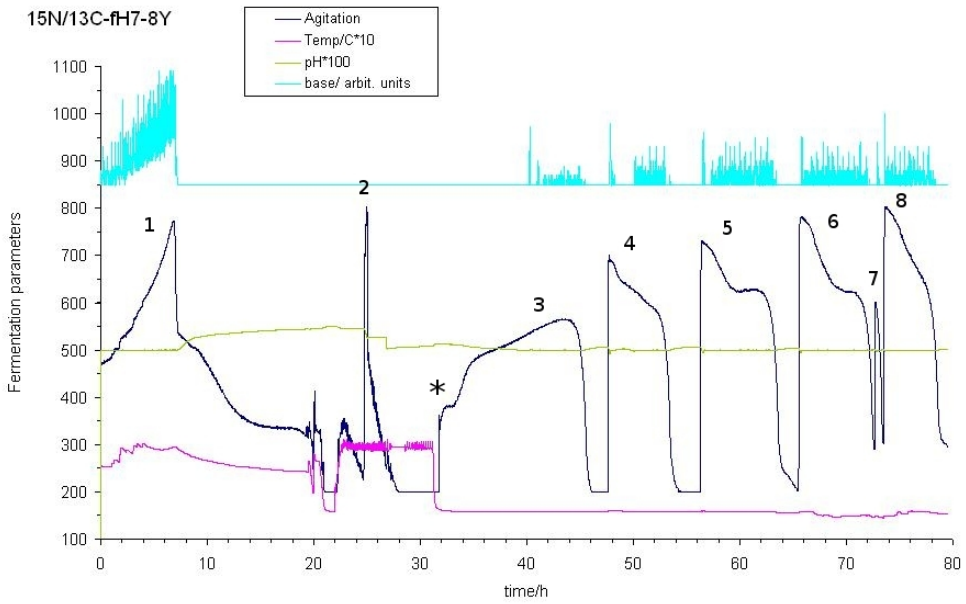

Figure 4.1: Fermentation of ^{13}C,^{15}N-fH∼7-8Y. The spikes in the agitation curve (dark blue) reflect metabolic activity, i.e. oxidation of initial ^{13}C-glycerol (1) and ^{13}C-glucose (2) and, during induction, of ^{13}C-methanol (3-8). The starting point of the induction period is marked with an asterisk.

The reaction of the cells to ^{13}C-methanol feeds reflects an adaptation process: the first injection of 4 mL (injection 3 in 4.1) was slowly oxidised, as can be seen by both the low absolute value to which the agitation rose and the long time it took for it to drop back to the baseline. The second injection of another 4 mL of methanol was oxidised faster (injection 4), and the two following injections of 6 mL each were again digested quickly (injections 5 and 6). The small agitation peak before the last methanol injection is due to a small volume of methanol that had been gained from washing the empty methanol vials (injection 7). The last injection consisted of 5 mL of methanol (injection 8). Thus, a total volume of 31 mL methanol was used.

The overall yield was 60 g cells/800 mL fermentation volume. After separation of the

supernatant and addition of protease inhibitors 2.5 mL of the supernatant were diluted with 10 mL of dH_2O to perform a small-scale initial purification test on 0.7 mL SP-Sepharose (Amersham) and gravity flow conditions. The bound protein was washed with 20 mM potassium phosphate and eluted with 1 M NaCl in 20 mM potassium phosphate. A SDS-gel of the flow through, washing and elution fractions showed a single band of ca. 15 kDa in the elution fractions and no concentrated protein in the flow through or washing fractions (Fig. 4.3). The test was repeated using 1.5 M NaCl for elution to obtain a sharper elution profile (Fig. 4.2). The purification was then scaled up to 7 mL of SP-Sepharose and the whole supernatant was processed in one step, giving similar results.

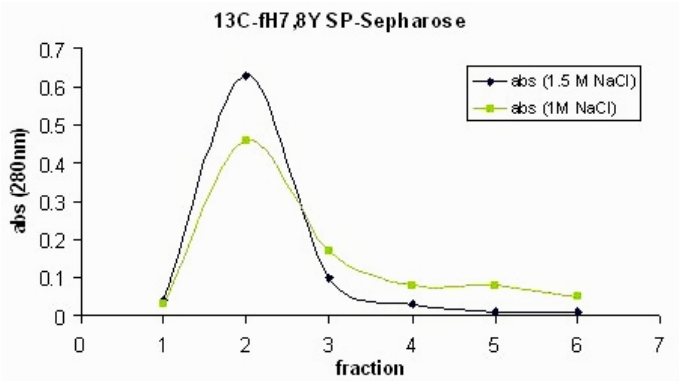

Figure 4.2: Small scale test purifications of ^{13}C,^{15}N-fH∼7-8Y on SP-Sepharose. Shown are step elutions with 1.5 M (blue) and 1 M NaCl (green). Fractions were collected manually and the fraction size was 0.5 mL.

After elution the protein was frozen in liquid nitrogen and stored at -80 °C until further purification. During the next purifcation step, MonoS cation exchange chromatography at pH 6.0, two peaks eluted at similar salt concentrations (Fig. 4.4).

Since no mass difference was observed for these two peaks on a SDS-PAGE gel (Fig. 4.5) it was assumed that the acidic cloning artefact EAEAAG had partially been removed from the protein (corresponding to the later peak) and that the resolution of SDS-PAGE was not sufficient to resolve the two species. The second species was more abundant and could be separated from the first species using small protein loads and shallower

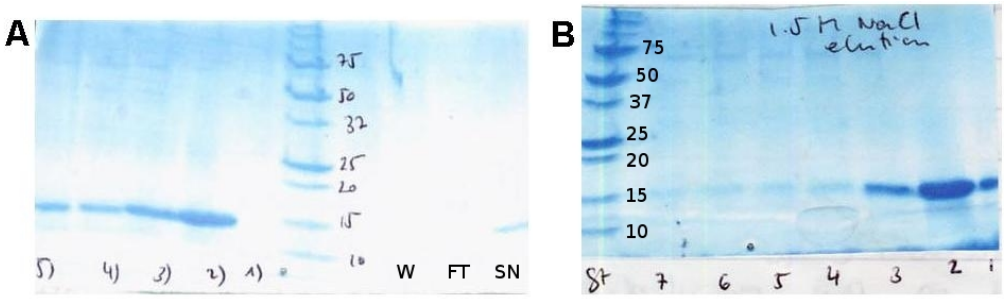

Figure 4.3: SDS-gels for small scale test purifications of ^{13}C,^{15}N-fH∼7-8Y. A) elution with 1 M NaCl, B) elution with 1.5 M NaCl. SN: supernatant, FT: flow through, W: wash, 1) - 7) fractions corresponding to Fig. 4.2. A protein ladder with individual bands in kDa for comparison is included in each gel. The molecular weight of fH∼7-8Y is 13 kDa.

salt gradients. Partial proteolysis of similar cloning artefacts had been observed when using the same cloning method in the past. ^{1}H, ^{15}N-HSQC spectra of both proteins displayed two additional ^{15}NH cross peaks for the earlier eluting species (Fig. 4.6). These additional cross peaks and the backbone assignment of fH∼7-8Y accomplished on the later eluting species led to the conclusion that the protein belonging to the later peak contained the slightly truncated cloning artefact EAAG, while the earlier eluting peak contained the full-length cloning artefact EAEAAG. One additional negatively charged sidechain in the N-terminus seems thus sufficient to lower the affinity of the 13 kDa protein for a cation exchange column.

Despite the lower theoretical pI of fH∼7-8Y (8.76 with the artefact EAAG and 8.61 with the artefact EAEAAG) when compared to fH∼7Y (9.24 with the artefact EAAG, all values obtained from SwissProt) an overlay of the purification chromatograms of both proteins at pH 6.0 shows that fH∼7-8Y elutes at a higher salt concentration than fH∼7Y (Fig. 4.4).

Comparison of the ^{1}H, ^{15}N-HSQC spectra of fH∼7Y and fH∼7-8Y showed that fH∼7-8Y was well folded and that most residues in module 7 experience no or only relatively small chemical shift changes in the presence of module 8 (Fig 4.7).

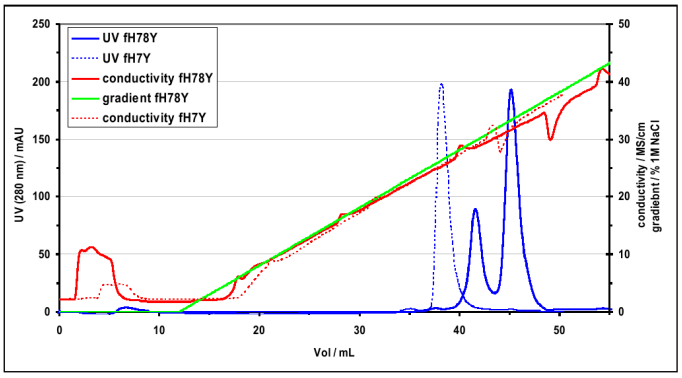

Figure 4.4: Purification of ^{13}C,^{15}N-fH∼7-8Y (solid line) on a MonoS cation exchange column. The two peaks correspond to proteins with the artefact EAEAAG (earlier peak) and EAAG (later peak). For comparison the purification of fH∼7Y is overlayed (dashed lines, cloning artefact EAAG).

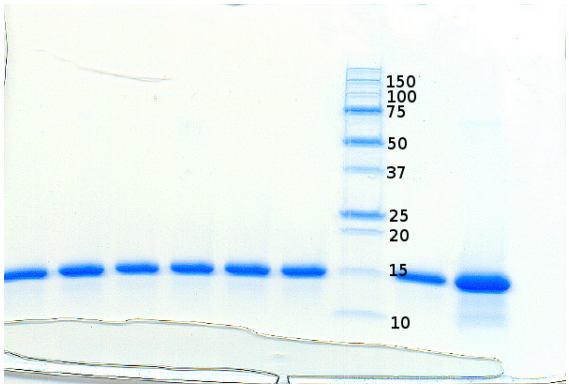

Figure 4.5: SDS-PAGE of purified ^{13}C,^{15}N-fH∼7-8Y after cation exchange on a MonoS column. Shown are fractions between 40 and 50 mL of the chromatogram shown in Fig. 4.4 as well as a protein ladder with individual bands in kDa for comparison. No mass difference can be seen between the two peaks in Fig. 4.4 signalling only a small difference in mass or different oligomeric states or different conformations for both peaks.

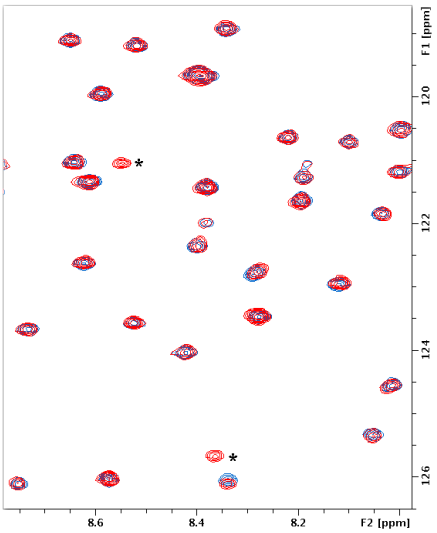

Figure 4.6: Overlay of ^{1}H, ^{15}N-HSQC spectra of the first (red) and second (blue) peak of fH∼7-8Y-elution from a MonoS column (Fig. 4.4). The asterisks denote the two ^{15}NH-cross peaks that are only present in the earlier eluting protein. F1 refers to ^{15}N, F2 to the ^{1}H dimension.

4.1.2 Fermentation and purification of unlabelled and ^{15}N-labelled proteins

In principle, fermentations of ^{15}N-fH∼7Y, ^{15}N-fH∼7-8Y/H, ^{15}N-fH∼19-20 and ^{15}N-NK1 were conducted in a similar manner but unlabelled glycerol was used instead of ^{13}C-glycerol and ^{13}C-glucose. 30 mL glycerol was present in the minimal media and further 10 mL of glycerol were added for the additional feeding step. Unlike the clones containing the fH∼7-8 and fH∼19-20 sequences the NK1-clone belonged to a Mut^{+}-strain with an intact AOX1 gene. As a result, this clone oxidised methanol at a much higher rate. Ca. 110 g of cells were obtained from a 960 mL fermentation of ^{15}N-NK1 and yields were similar for the other ^{15}N-labelled protein expressions. The supernatants of different expressions were processed in the same way as the ^{13}C,^{15}N-fH∼7-8Y-containing supernatant with the exception of ^{15}N-NK1 which was purified on Heparin-Sepharose during the initial purification step since a small scale test showed better purification results than when using SP-Sepharose. ^{15}N-fH∼7-8H, ^{15}N-fH∼7-8Y and ^{15}N-fH∼19-20 were

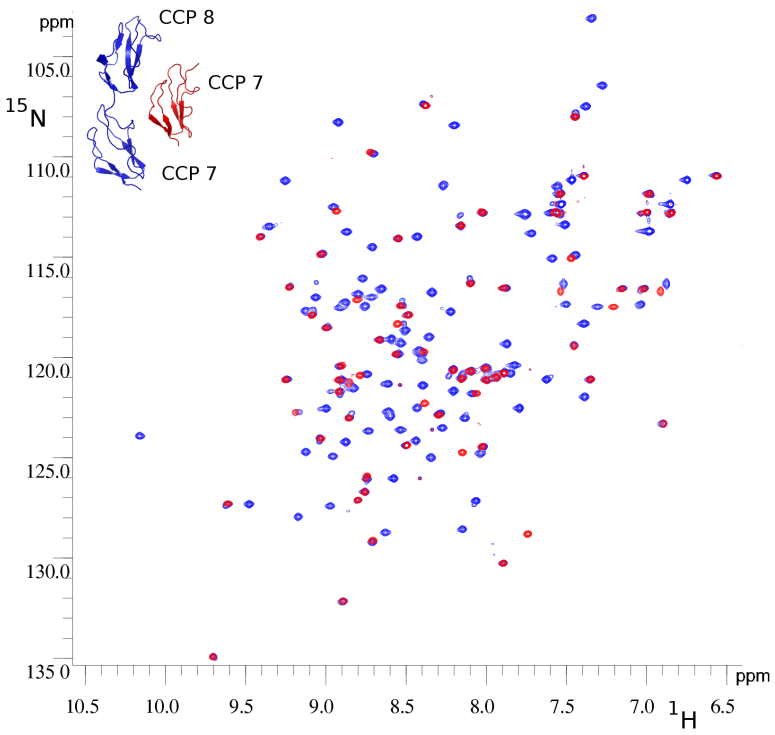

Figure 4.7: Overlay of the ^{1}H, ^{15}N-HSQC spectra of ^{15}N-fH∼7,8Y(blue, later eluting peak from MonoS) and ^{15}N-fH∼7Y (red) with inlayed structures of the two proteins (same colour coding, PDB 2JGX for fH∼7 and derived from PDB 2UWN for fH∼7-8).

purified like ^{13}C, ^{15}N-fH∼7-8Y and gave similar results with minor peaks corresponding to complete cloning artefacts and major peaks corresponding to truncated cloning artefacts. ^{15}N-NK1 was produced as one uniform protein but, unlike the fH modules, suffered from degradation if stored at room temperature or 4 °C. For production of unlabelled fH constructs (fH∼7Y and fH∼19-20) shaker flask expressions were sufficient. Yields were lower when compared to the fermentor expressions but still several mg/1L culture. Similar distributions of cloning artefacts were observed for shaker flask and fermentor expressions.

4.2 Backbone assignment of $^{13}C,^{15}N$-fH∼7-8Y

Binding experiments with fH∼7 and dp4C showed significant chemical shift changes for K446, the C-terminal residue of fH∼7. Consequently, in order to assess possible binding contributions from module 8 (which had not previously been implicated in heparin binding), the ^{13}C, ^{15}N labelled double module fH∼7-8 was prepared and its backbone was assigned. The backbone assignment was achieved using the CBCA(CO)NH/CBCANH (Fig. 4.8) pair of 3D triple resonance experiments as well as the NH(CA)CO/NHCO and HBHA(CO)NH/HBHANH spectral pairs, as described in section 1.8. MARS [211], a software package for semi-automated backbone assignment, assigned about 50% of the backbone resonances in the HSQC spectra, using as input a list containing the C^{α}, C^{β}, CO, H^{α} and NH resonance frequencies as well as assignments of the glycine residues. Importing these assignments into CCPN Analysis and manually checking for mistakes led to only 2-5% disagreement with the automated assignment. The rest of the backbone was assigned manually; 97% of the ^{15}NH backbone resonances were identified.

During the assignment it was found that the protein contained the artificial cloning sequence EAAG at the N-terminus. Residues which did not give peaks in the ^{1}H, ^{15}N-HSQC spectra (Fig. 4.9) were N396, N399, C442 and D497, the first three of which belong to module 7. Severe overlap occurred between the backbone amide signals of C448 and C389, D413 and V445 as well as K474 and Y465, making temperature- and pH-titrations in small steps necessary to transfer the assignments for conditions used for oligosaccharide titrations (i.e. from 310 K, pH 5.0 to 298K, pH 7.4). Spectra were collected at pH 5.0 and 310 K, 307 K, 304 K, 301 K and 298 K as well as at 298 K and pH-values 5.0, 5.9, 6.4, 6.9 and 7.4. The backbone signals of K474 and Y465 remained unresolved under titration conditions, but the D413 and V445 pair separated and both peaks could be followed during the titrations.

The backbone ^{15}NH signals of K472, C477 and H417 only showed small overlap at assignment conditions but merged into one broad peak during the temperature titration. However, H417 could be followed under the ligand-titration conditions since it showed

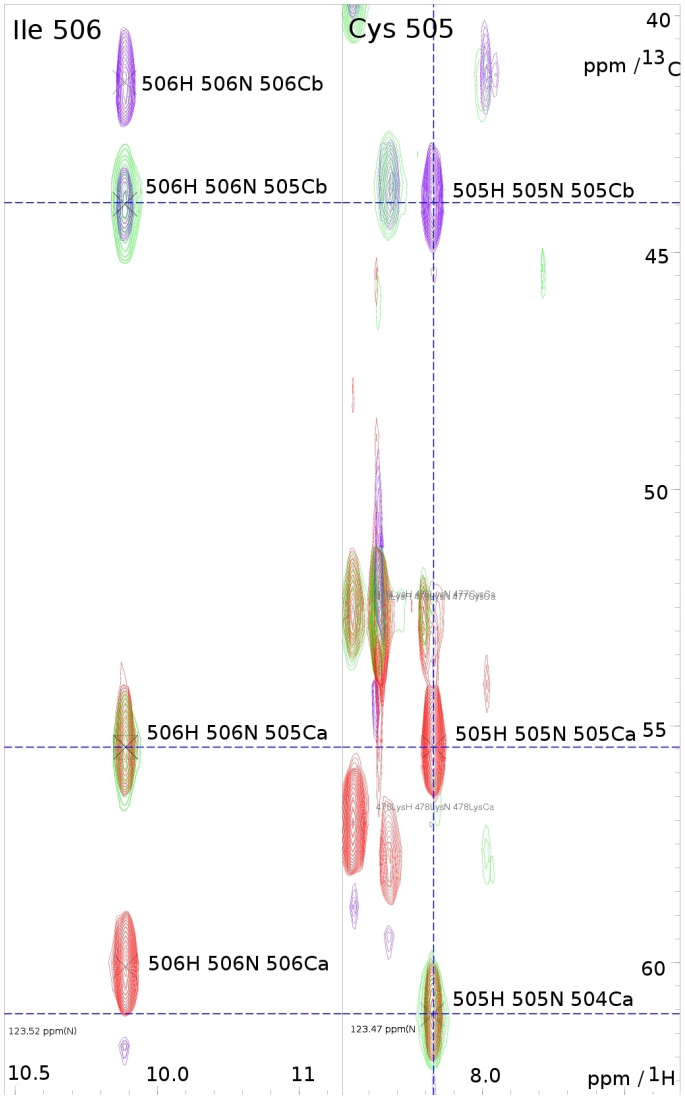

Figure 4.8: Example of two ^{1}H, ^{13}C-planes at two different ^{15}N frequencies in the overlaid CBCANH (purple/red) and CBCA(CO)NH (green) spectra, showing the stretches linked to the backbone ^{15}NH-resonance frequencies of Ile506 (left) and Cys505 (right). Cross peaks occurring in both spectra in the Ile506 stretch belong to residue Cys505, while cross peaks which are only present in the CBCANH spectrum belong to residue Ile506. Equally, cross peaks in both CBCANH and CBCA(CO)NH spectra in the Cys505 stretch belong to residue Thr504. The sequential order can be established because the CBCANH cross peaks in the Cys505 strip align with cross peaks in both CBCANH and CBCA(CO)NH in the Ile506 strip.

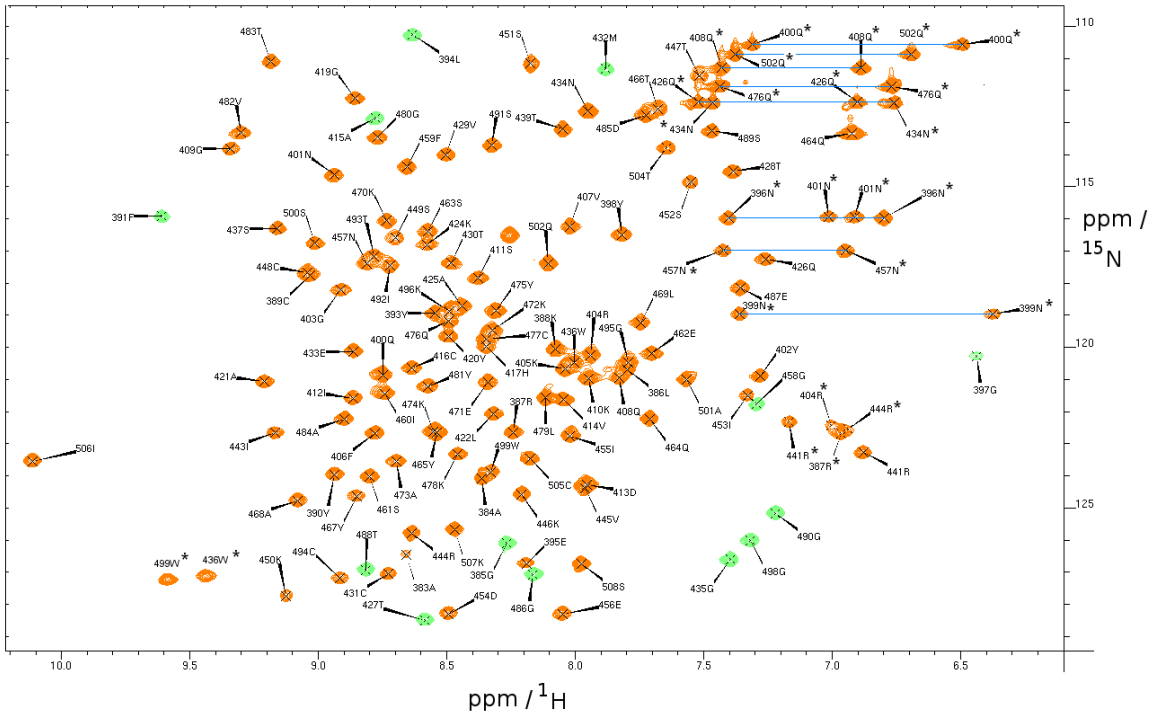

Figure 4.9: Assigned ^{1}H, ^{15}N-HSQC spectrum of ^{13}C,^{15}N-fH∼7-8Y at 310 K, 20 mM potassium phosphate, pH 5.0. Sidechain resonances are denoted by asterisks and connected by horizontal lines for Gln and Asn residues. Peaks with dashed crosses and colored green are folded and not displayed at their real ^{15}N frequency.

characteristic large chemical shift changes during the pH-titration that were also observed for this residue during pH-titrations of fH∼7Y/H where the H417 ^{15}NH cross peak did not overlap with other residues.

Other ^{15}NH cross peaks that could not be unambiguously assigned under the conditions used for oligosaccharide titrations (mostly due to excessive broadening at pH 7.4) belonged to residues L386, E395, G397, K424, S463 and K496. In total, information of chemical shift perturbation for the backbone NH cross peaks of about 10 non-Proline residues in fH∼7-8Y could not be given due to missing/unassigned peaks in the reference spectra. This represented about 8% of the total backbone signals. However, the peaks that remained unassigned because of spectral overlap (in particular the C448/C389 and

K472/C477) were still visible during the ligand titrations and did not experience significant shift changes during the titrations, implying that none of these residues were involved in binding.

A comparison of the assigned ^{1}H, ^{15}N-HSQC spectra of fH∼7Y and fH∼7-8Y showed that the biggest chemical shift movements are experienced by residues at the C-terminus of CCP 7 as would be expected from the end-to-end arrangement of the two modules. However, residues in the strand which contains a helical loop did also exhibit significant shift changes, suggesting small local structural changes (Fig. 4.10 and 4.11) upon addition of CCP 8.

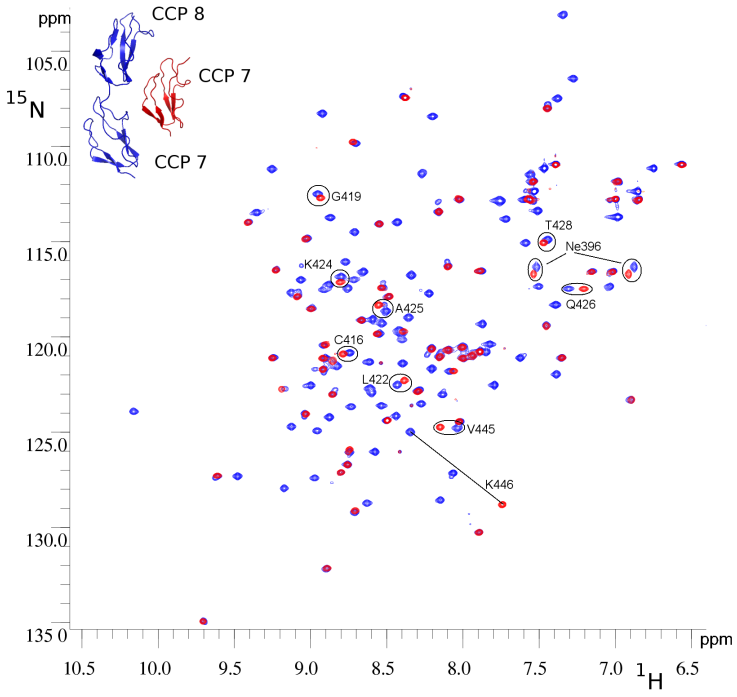

Figure 4.10: Overlay of ^{1}H, ^{15}N-HSQC spectra of fH∼7Y (red, PDB 2JGX) and fH∼7-8Y (blue, prepared from PDB 2UWN (fH∼6-8H)). K446 is the N-terminal residue in the fH∼7Y construct. The labels highlight residues in CCP 7 which experience combined chemical shift changes larger than the standard deviation of 0.03 ppm in the presence of CCP 8.

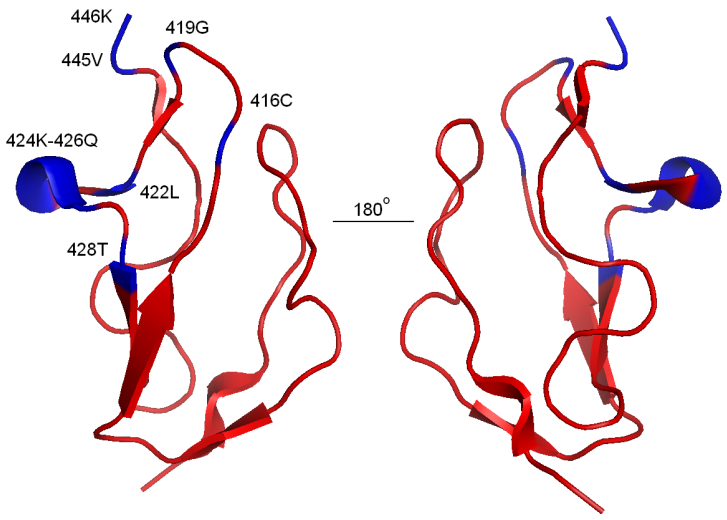

Figure 4.11: Visualization of chemical shift changes experienced by CCP 7 in the presence of CCP 8. Cross peaks of the residues highlighted in blue experience backbone combined chemical shift changes larger than the standard deviation of 0.03 ppm in the presence of CCP 8 (see Fig. 4.10).

4.3 Binding studies of fH∼7 and fH∼7-8

4.3.1 Titrations of fH∼7 and fH∼7-8 with heparin tetrasaccharide and sucrose octasulfate monitored by ^{1}H, ^{15}N-HSQC

The first oligosaccharide-protein titrations were done using ^{15}N-fH∼7Y and ^{15}N-fH∼7H, which were provided by Dr. Andrew Herbert. The conditions used were 20 mM potassium phosphate, pH 7.4 and 298K, the protein concentration was 50 µM and the initial ligand concentrations were between 13 and 25 µM, respectively. First, a reference spectrum without the ligand was recorded to account for small differences in pH and salt concentrations that could lead to small chemical shift changes compared to the spectra used for the assignment. This spectrum was also used to transfer the assignments obtained at more acidic conditions by Andrew Herbert for both variants of fH∼7. Titration points with increasing ligand concentrations were recorded, chemical shift differences were calculated and histograms showing these differences were produced (Fig. 4.12).

Residue 402, linked to AMD, appeared to be at the edge of the heparin binding site: while this residue's cross peak did not show significant chemical shift changes, residues 405 and 406 experienced large chemical shift changes.

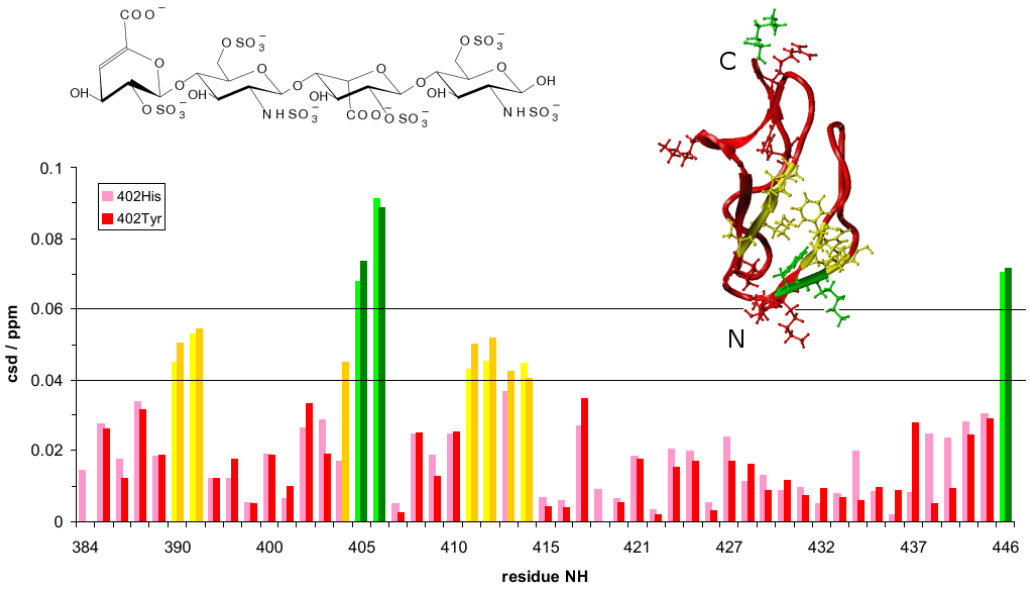

Figure 4.12: Combined amide chemical shift changes in ^{15}N-fH∼7Y (solid) and ^{15}N-fH∼7H (faint) upon titration with dp4C (see the inset) at 298 K, 20 mM potassium phosphate, pH 7.4. Residues with chemical shift changes >0.06 ppm (green) and chemical shift changes of 0.04-0.06 ppm (yellow) are mapped onto the fH∼7Y-structure and drawn in stick presentation (PDB 2JGX). Basic residues are also drawn in stick representation. Proline residues and other non-observable residues are omitted from the sequence.

Residues most affected by the binding are K405, F406 and K446 which experience chemical shift changes >0.06 ppm (larger than 3 times the average chemical shift change). Residues R404 (in the Y-form), S411, I412, D413, V414, Y390 and F391 also shift significantly (with chemical shift changes larger than 2 times the average chemical shift change). The region on CCP 7 which comprises all these residues except for K446 is denoted site 1 in the following text. When these residues were mapped onto the NMR structure of fH∼7Y [240] it became obvious that binding site 1 is in proximity to residue 402 and is orientated towards the N-terminus and module 6 rather than towards the

C-terminus and module 8. However, a large chemical shift change was also observed for the most C-terminal residue of the fH∼7-construct, K446, which is located in the linker region between fH∼7 and fH∼8. (The linker region is generally referred to as the amino acid sequence between the last Cys residue of one module and the first Cys residue in the following module, in this case 442C-IRVKT-448C). It was assumed that this second binding site (site 2) could be due to the artificial absence of module 8 in the fH∼7 constructs and an increased flexibility of the C-terminus. Consequently, fH∼7-8Y and fH∼7-8H were expressed and labelled with ^{15}N. Titrations under the same conditions as used for ^{15}N-fH∼7 showed that site 2 resonances exhibited larger chemical shift changes than those in site 1 (Fig. 4.13).

Judged by the chemical shift changes, the primary binding site in the double module therefore seemed to be located in the linker region and the hypervariable loop close to the C-terminus of module 7 (Fig. 4.13). The presence of CCP 8 completes the binding site in the linker between CCP 7 and CCP 8. Even more unexpected was the finding that the double module fH∼7-8 (neither the H- nor the Y-form) does bind dp2, while in the case of fH∼7 dp2 and dp4 showed very similar binding results. This behaviour, observed using ^{1}H, ^{15}N-HSQC titrations, was supported by GMSA data acquired by Jon Deakin and Dr. Malcolm Lyon (see below). More binding studies with sucrose octasulfate (SOS) were conducted and compared with the results obtained from the dp4C titrations as part of our collaboration with an x-ray group [241]. ^{15}N fH∼7-8 rather than fH∼7 was used in these studies since the titrations with dp4C suggested contributions from the linker region between CCP 7 and CCP 8. SOS is often employed as a heparin mimic in x-ray crystallography but a comparison with the more physiological heparin-derived tetrasaccharide dp4C was required, given the large chemical differences between both compounds (structures inserted in Fig. 4.13). The titration of ^{15}N-fH∼7-8Y was repeated using SOS instead of dp4C (Fig. 4.13, 4.14). No significant differences in the binding sites of these two ligands were found despite large differences in size and primary structure, suggesting that both molecules bind to the same regions of fH∼7-8Y and that charge density rather than subtle structural features is responsible for the

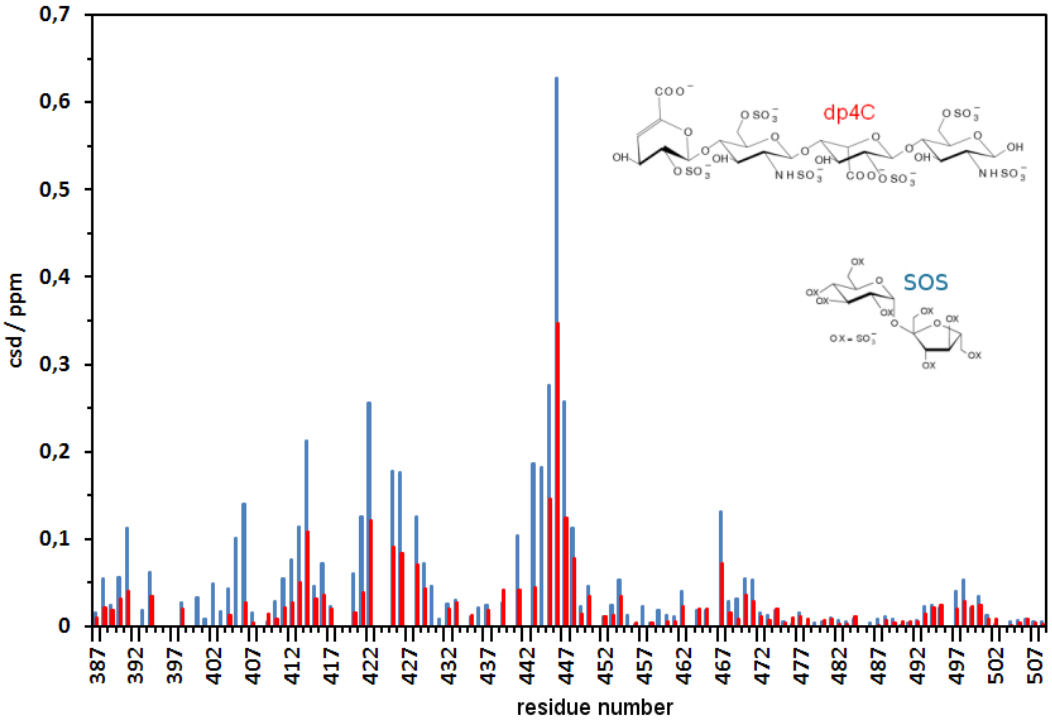

Figure 4.13: Combined amide chemical shift changes in ^{15}N-fH∼7-8Y upon titration with dp4C (red) and the heparin mimic sucrose octasulfate (SOS, blue) at 298 K, 20 mM potassium phosphate, pH 7.4. Site 1 observed in the single module fH∼7 (Fig. 4.12) shows smaller chemical shift changes than the double module. Concentrations used: 50 µM protein with 75 µM dp4C and 60 µM protein with 500 µM SOS. Missing residues are either proline residues or residues whose amide cross peaks were lost in the presence of the ligand due to line broadening.

binding. Chemical shift changes upon binding of the two ligands are quantified in the next section.

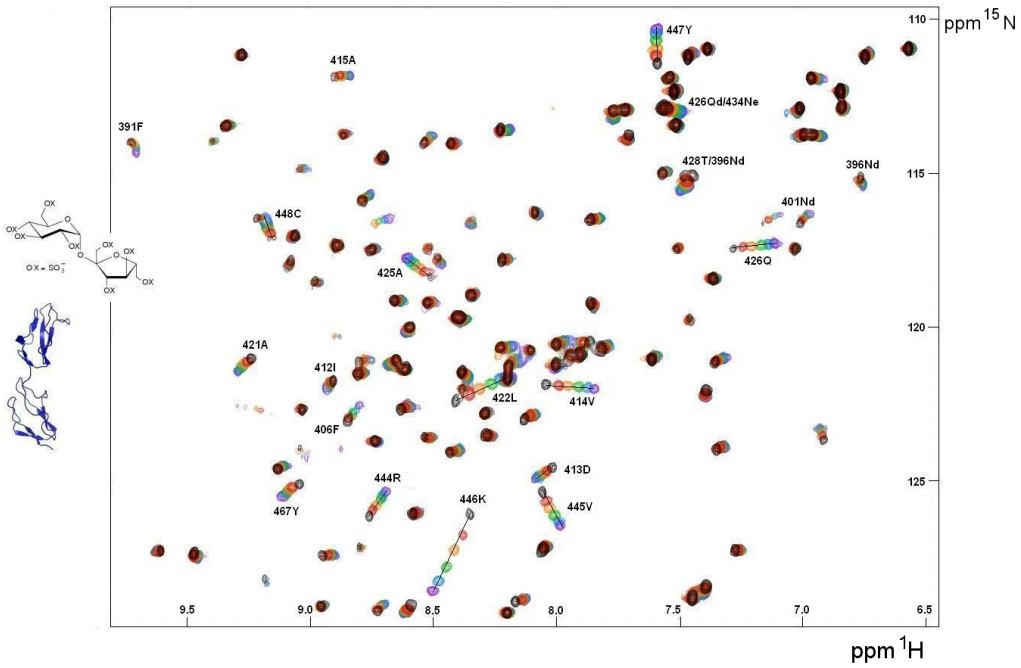

Figure 4.14: ^{1}H, ^{15}N-HSQC spectra of ^{15}N-fH∼7-8Y titration with sucrose octasulfate (SOS) at 298 K, 20 mM potassium phosphate, pH 7.4. Concentrations were 60 µM protein (ref) and 15.5, 31, 62, 124, 248 and 496 µM SOS.

4.3.2 Evaluation of the effects of Y402H SNP on the binding of dp4C and SOS to fH∼7-8

A direct comparison of the ^{1}H, ^{15}N-HSQC titrations of both fH∼7-8 variants with both dp4C and SOS shows that only small differences are observed between the protein variants and dp4C/SOS binding sites (Fig. 4.15). Subtle differences are found at the N-terminus of CCP 7. The chemical shift changes seem to be larger when using the H402 variant but it is possible that this is caused by slightly varying protein or GAG concentrations. During HSQC titrations samples are repeatedly taken from their NMR tubes, mixed with small volumes of ligand stock and, after pH adjustment, transferred back. This procedure leads to errors in concentration as the sample volume is permanently reduced. In complementary experiments like gel mobility shift assays and heparin affinity chromatography no significant differences in binding affinities were observed between the two fH∼7-8 variants. Both proteins eluted in one peak from a heparin affinity column after simultaneous injection. Thus, using the fH∼7-8 construct neither changes in the position of the binding site nor significant reduction in affinity were observed as a result of the Y402H SNP.

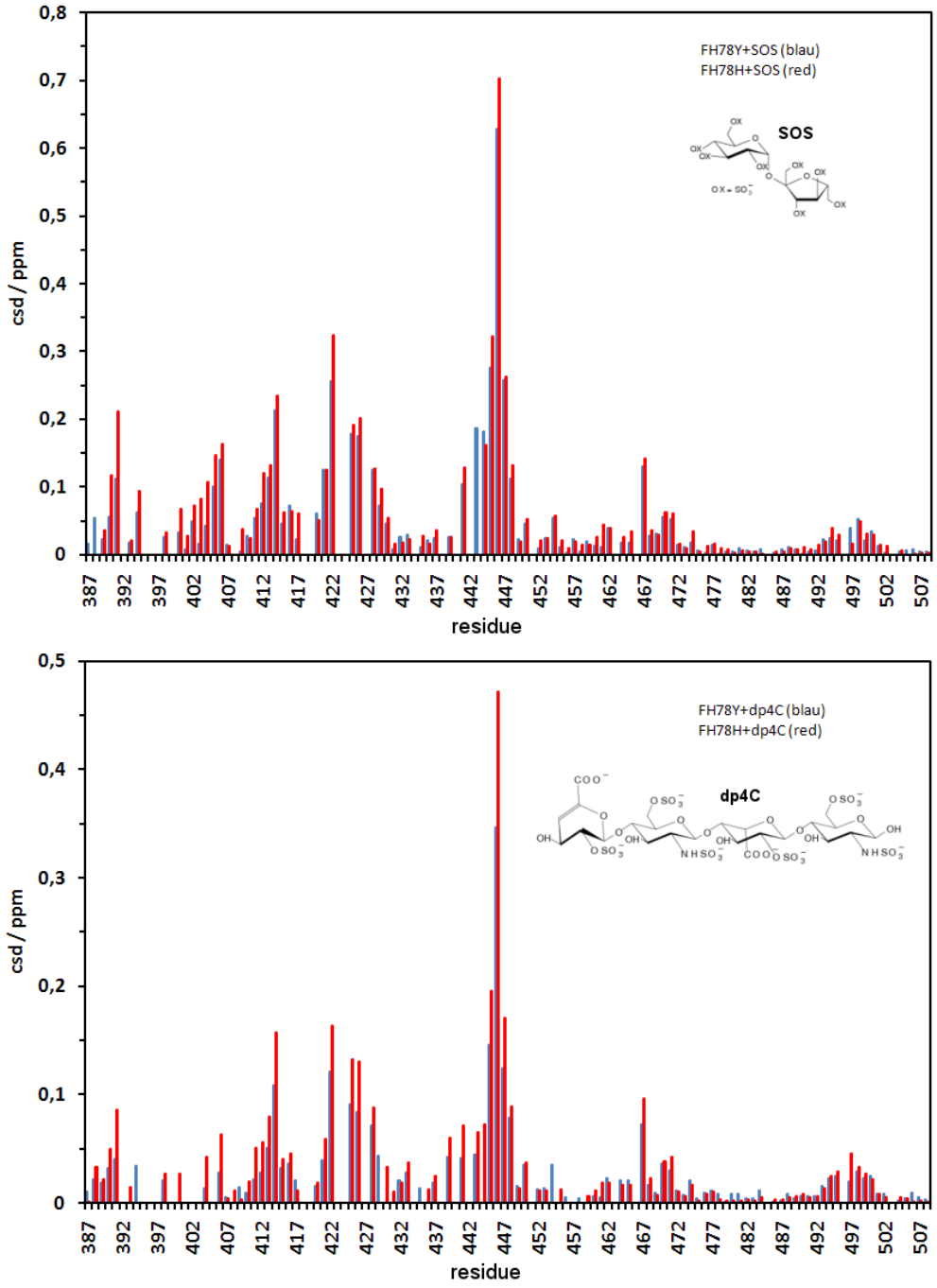

Figure 4.15: Comparison of the effect of SOS (top) and dp4C (bottom) fH∼7-8Y (blue) and fH∼7-8H (red).

4.3.3 Titration of fH∼7-8Y with dp6C

An attempt was made to titrate dp6C in fH∼7-8Y in order to answer two questions: whether the binding site for the hexasaccharide is larger and whether it binds more strongly than dp4C to the protein. While mainly the same residues seemed affected by the complex formation the binding affinity was higher as judged by intermediate broadening and disappearance of cross peaks (Fig. 4.16 and 4.17). At large excess of dp6C some cross peaks which had been broadened during early titration points sharpened, signalling saturation. Other peaks, however, did not re-appear during the last points of the titration (Fig. 4.17). The effect which temperature and field strength have on the exchange regime was not investigated further in this titration. During the titration of HBD2 with Fondaparinux, however, these two parameters were found to be important tools that allowed the chemical shift changes to be followed during intermediate exchange broadening (see below).

A comparison of the fH∼7-8Y titrations with dp4C and dp6C revealed similar binding sites for both oligosaccharides (Fig. 4.18). The largest chemical shift changes were still observed in the linker region between CCP 7 and CCP 8. The binding affinity of fH∼7-8Y for dp6C can be judged to be higher than the affinity for dp4C because the titration with dp6C showed signs of intermediate broadening. However, the N-terminal binding site in CCP 7 which was more pronounced in the dp4C titration of fH∼7 than in fH∼7-8 did become more prominent in this titration with dp6C. A possible interpretation of this observation is that with dp6C both binding sites can be bridged while with dp4C they compete for oligosaccharide binding. This observation matches the hypothesis that for dp6C the two binding sites termed site1 and site2 above simultaneously contribute to the complex, while dp4C is only able to contact one of them at any time. The length of a free tetrasaccharide is ca. 17 Å[165], which is not sufficient to span a whole CCP module of about 30-35 Å. dp6C is significantly longer (ca. 25 Å) than dp4 and almost as long as a CCP module, it might therefore be able to contact both site1 and site2 simultaneously. dp6C does also carry a larger negative charge (eleven anionic groups instead of eight in dp4C), although the charge density is the same. Physiological NS

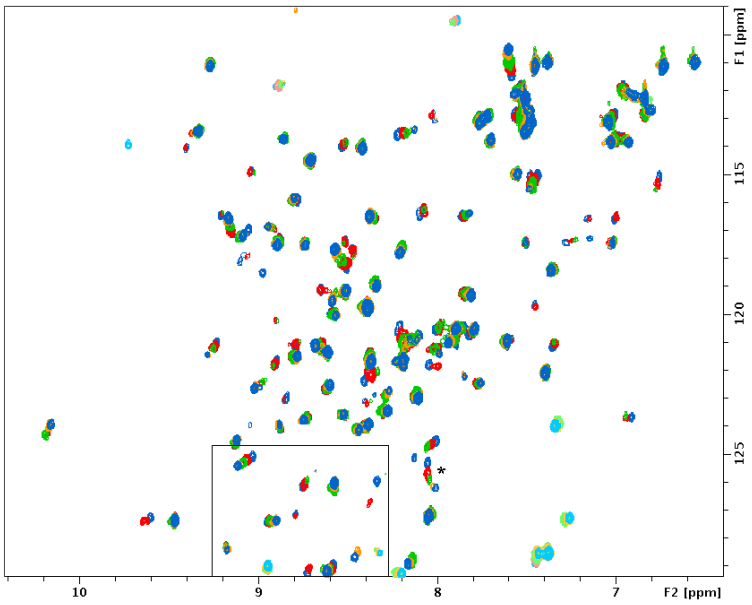

Figure 4.16: Titration of 25 µM fH∼7-8Y with dp6C at 298K, 20 mM potassium phosphate, pH 7.4, 800 MHz. Ligand concentrations were: 0 µM (blue), 12.5 µM (red), 37.5 µM (green), 50 µM (yellow), 100 µM (blue). The region of the spectrum within the black square is enlarged in Fig. 4.17. The asterisk denotes residue V445, whose shift changes follow a curved rather than a straight line, indicating that this residue experiences two different events during the titration. F2 refers to the ^{1}H, F1 to the ^{15}N dimension, respectively.

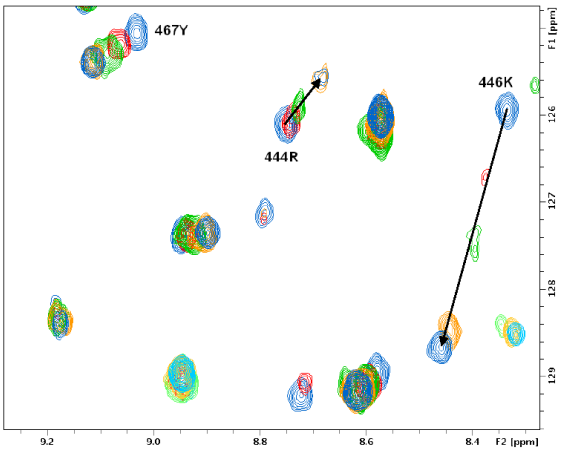

Figure 4.17: Enlargement of Fig. 4.16. Cross peaks undergo intermediate exchange broadening but sharpen at high ligand concentrations. FH∼7-8Y: 25 µM, dp6C: 0 µM (blue), 12.5 µM (red), 37.5 µM (green), 50 µM (yellow), 100 µM (blue). F2 refers to the ^{1}H, F1 to the ^{15}N dimension, respectively.

domains in HS comprise 8-9 disaccharide units, dp6 is therefore still short in comparison to physiological HS chains and the length of one CCP module could easily be spanned by physiological HS chains.

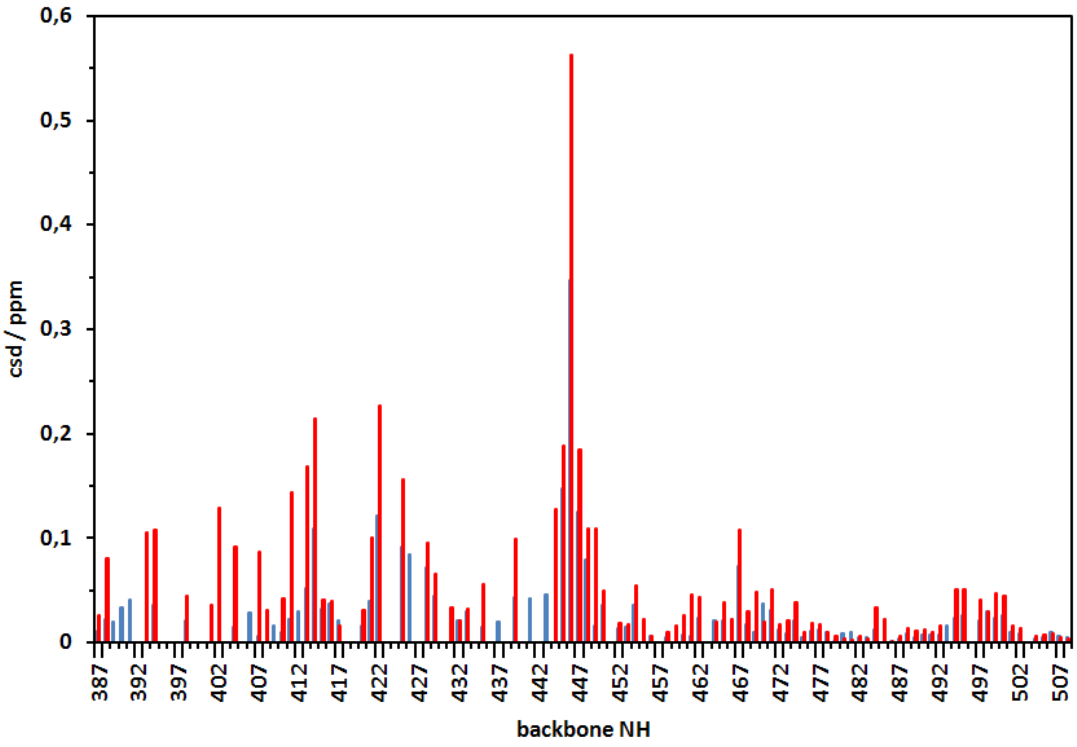

Figure 4.18: Comparison of the effect of dp4C (blue, protein:GAG ratio 1:3.5) and dp6C (red, protein:GAG ratio 1:4) on fH∼7-8Y.

4.3.4 Gel mobility shift assay and heparin affinity chromatography on fH∼7-8

The GMSA data, obtained by Jon Deakin, led to the same result as the titration experiments monitored by HSQC spectra (Fig. 4.19): the migration of dp2 in the electric field is not hindered by either of the fH∼7-8 mutants, implying that dp2 is not bound by these proteins and in the buffer usually employed in GMSA (PBS).

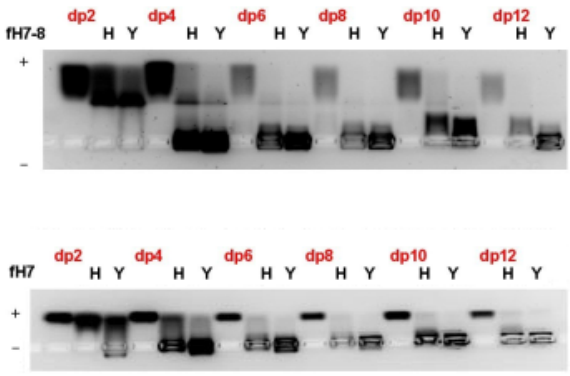

Figure 4.19: GMSA on fH∼7-8Y/H (top) and fH∼7Y/H (bottom, taken from [240]) with different length oligosaccharides. Free ligand migrates to the anode (+). Oligosaccharides which are are hold back by the proteins and do not migrate out of the wells.

This assay did not exhibit significant differences in the binding affinities for the same length oligosaccharides between the two isoforms fH∼7-8Y and fH∼7-8H. This result is in line with the titration results which showed that the main binding site in fH∼7-8 is located at the C-terminus of CCP 7, well away from residue 402.
Injection of a solution containing both fH∼7-8Y and fH∼7-8H in PBS onto a heparin affinity column (Poros HE20) followed by elution with a salt gradient resulted in a single peak, again implying that in the double module residue 402 is not part of the major heparin binding site (the same experiment yields separation of the two variants of the fH∼7 single module, [240]).

4.4 Oligomerisation of fH∼7 upon addition of dp4C

In many cases heparin binding proteins dimerise upon heparin binding [242, 243, 244]. This dimerisation slows down the molecular tumbling which is reflected in the change of ^{15}N relaxation times. The rotational correlation time, τ_c, can be determined from the ratio of the ^{15}N T_1 and T_2 relaxation times [245]. Chemical exchange can influence the results. However, in the cases presented here the majority of residues experience no or marginal chemical shift changes and can safely be described as being in the fast exchange regime in which no peak broadening is observed. In this case the resulting relaxation is given by the weighted average of the relaxation times of individual species [246]. The ^{15}N T_1 and T_2 relaxation times were measured based on 1D experiments sampling the envelope of NH resonances. A comparison with 2D ^{15}N relaxation data obtained for fH∼7Y showed good agreement with the 1D data. Therefore, the 1D method was used for this study. The measurement of ^{15}N T_1 and T_2 for different samples led to the following results:

sample	method	T_1 / ms	T_2 / ms	T_1/T_2	τ_c / ns
fH∼7Y 25 μM	1D, 800 MHz	712	145	4.9	4.7
fH∼7Y 25 μM	2D, 600 MHz	462	151	3.0	4.6
fH∼7Y 25 μM + dp4C 40 μM	1D, 800 MHz	981	74	13.3	8.4
fH∼7H 600 μM	1D, 800 MHz	662	121	5.5	5.0
fH∼7H 25 μM + dp4C 40 μM	1D, 800 MHz	1140	85	13.4	8.5

The drastic increase in the protein τ_c values in the presence of dp4C clearly imply that some form of oligomerisation (most likely dimerisation) is caused in both fH∼7Y and fH∼7H by the addition of the heparin-derived oligosaccharide dp4C. Large excess of dp2 caused some dimerisation of fH∼7Y (data not shown), however to a much smaller degree, which might be related to its size and the fact that it contains only three sulfate groups as opposed to six sulfate groups in dp4C. The measurements were all done using samples with pH values around 5, and it remains to be seen if this trend is also observed

at physiological pH. Practically identical rotational correlation times of the two proteins imply that their dimerisation in the presence of dp4C does not depend on the 402Y/H polymorphism. Sedimentation studies of fH∼6-8Y and H in the presence of heparin decasaccharide have recently been carried out and in these experiments a slightly larger self-association tendency was observed for the H402 form when compared to the Y402 form [247].

4.5 Cross linking of dp4C to fH∼7Y

4.5.1 The HISQC experiment

^{1}H, ^{15}N-HSQC titrations are commonly used to monitor chemical shift changes of backbone NH signals induced by increasing ligand concentrations. In the context of heparin binding proteins, however, lysine and arginine sidechains are of central importance. Due to the long sidechains of these two basic amino acids it does not seem ideal to observe backbone resonances only, as the backbone NH of a lysine residue whose ζ-NH_3^+-group is involved in an interaction does not necessarily undergo large chemical shift changes (and vice versa). The ^{1}H, ^{15}N-HSQC experiment alone does not confer the degree of atomic resolution that NMR is capable of. This work demonstrates how monitoring the backbone resonances can be complemented by sidechain experiments both in the non covalent protein heparin complex (H_2CN) and after chemical cross linking (HISQC).

One reason for the absence of arginine- and lysine sidechains in the HSQC experiment is the fast exchange of sidechain NH protons with bulk water. For arginine H^{ϵ} signals this exchange is slow enough at pH values below 6, thus arginine $N^{\epsilon}H^{\epsilon}$ cross peaks can be observed under slightly acidic conditions and are usually assigned during structure determination. For lysine H^{ζ} protons the exchange rate is even faster, and at pH 6 and room temperature $N^{\zeta}H^{\zeta}$ resonances are not observed unless they belong to buried sidechains or are protected from exchange by complex formation/oligomerisation. Ad-

ditionally, the ^{1}H, ^{15}N-HSQC experiment is usually performed with a nitrogen offset of around 110 ppm which is far away from the resonance frequency of N^{ζ} nitrogens at around 33 ppm. Consequently, ζ-NH_3^+ groups do not usually appear in ^{1}H, ^{15}N-HSQC spectra. These limitations are overcome by the HISQC pulse sequence (for Heteronuclear *In-phase* Single Quantum Coherence), which is optimised for lysine ζ-NH_3^+-groups rather than backbone NH groups [248]. The key of the HISQC experiment is the fact that during the nitrogen chemical shift labelling period the ^{15}N transverse coherence is kept in-phase with respect to the the attached protons via proton decoupling, hence the name of the sequence. The development of the HISQC experiment was motivated by a tight DNA-protein complex in which a lysine $N^{\zeta}H^{\zeta}$ cross peak was protected from exchange with water by the tightly bound ligand [248]. Here, it was investigated if a similar protection could be observed for the fH∼7Y-heparin interaction. Spectra were recorded at different temperatures and pH values to determine conditions that could be used for the assignment of all lysine $N^{\zeta}H^{\zeta}$ cross peaks in fH∼7Y. Only at 283 K and pH 3.0 was the water exchange sufficiently slow to observe and assign all $N^{\zeta}H^{\zeta}$ resonances, conditions which can hardly be used for physiological meaningful ligand titrations. This problem was overcome by covalent cross linking of the fH∼7Y-GAG complex and use of the H_2CN pulse sequence, as desribed in this section.

4.5.2 Preparation of a covalent GAG-protein complex

Standard ^{1}H, ^{15}N-HSQC titrations of proteins with GAGs follow the movement of backbone NH resonances. Out of nine basic residues (five lysines and four arginines) only three were among the residues that displayed significant movement of NH cross peaks during the titration of fH∼7Y with dp4C. The positions of the five lysine residues in the structure of fH∼7Y are indicated in Fig. 4.20. Three are in the N-terminal part of the protein (K388, K405 and K410) while two are in the C-terminal half (K424 and K446). In order to characterise the contributions from these lysine sidechains to binding of dp4C a zero-length cross linking strategy was applied. dp4C contains two carboxyl groups which can be activated with carbodiimides. Subsequently, a peptide bond can be

formed with protein lysine sidechains if the activated oligosaccharide forms a complex with the protein. A similar strategy was developed previously to characterise protein-protein complexes through cross linking of lysine and aspartate/glutamate sidechains, followed by tryptic digestion and mass spectrometry [249].

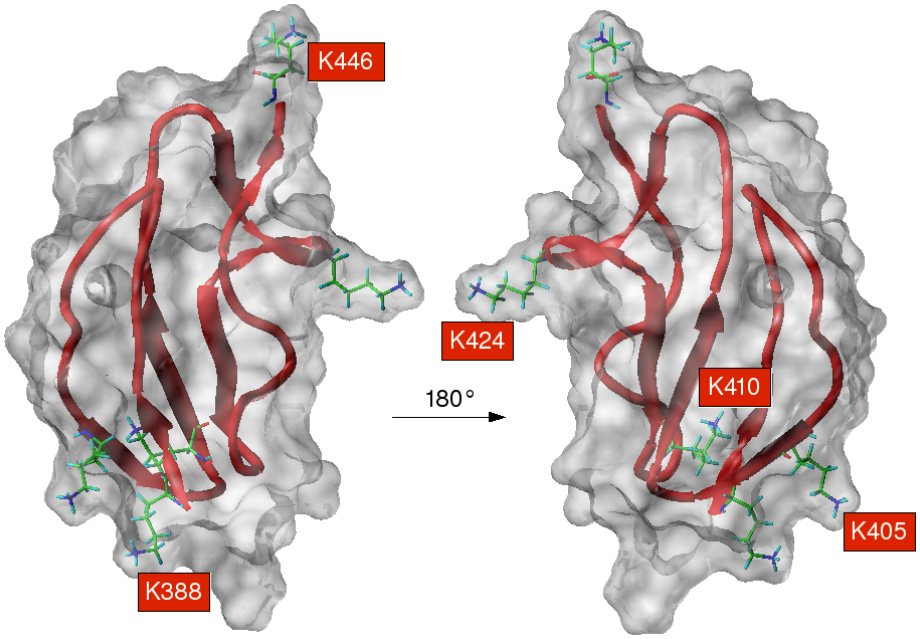

Figure 4.20: Position of the five lysine residues on fH7Y (PDB 2JGX).

Prior to cross linking, the purity of fH∼7Y was checked using cation exchange chromatography (Fig. 4.21). The activated dp4C was separated from excess cross linking reagent using a PD-10 desalting column. The products of the cross linking reaction were also loaded onto a MonoS column in order to separate free protein, whose binding properties should be unchanged, from the cross linked species. The cross linked protein, fH∼7Yxdp4C, was expected to bind with much lower affinity to the cation exchange column due to the negative charges of the dp4C moiety. Different fractions were subject to MALDI-TOF mass spectrometry and it was found that the cross linked protein did not bind to the MonoS column at all (Fig. 4.22). The masses of several minor species which bound better than fH∼7Yxdp4C but less well than the free protein were also

measured. It was concluded that they corresponded to species arising from chemical modifications of fH∼7Y with the cross linking reagents and dimerisation caused by cross linking of two protein molecules without a dp4C molecule. Thus, separation of activated dp4C from the cross linking reagents prior to addition of the protein appears to have been incomplete. Further purification of the MonoS column flow through using HiTrap desalting columns also revealed the presence of remaining cross linking reagents (Fig. 4.23). An accurate mass measurement after this desalting step and dialysis confirmed that the protein-GAG hybrid recovered from the MonoS flow through contained only mono-cross linked species (Fig. 4.24).

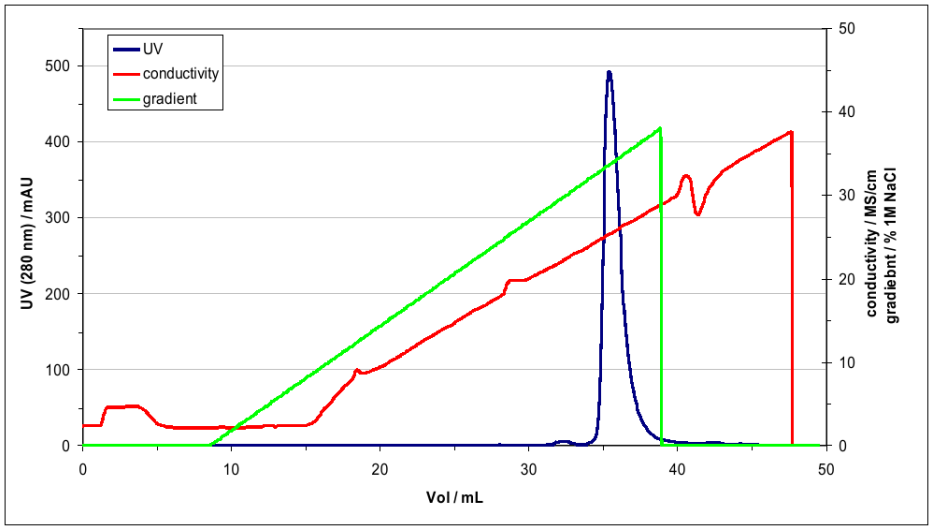

Figure 4.21: Purity check of fH∼7Y on a MonoS column prior to cross linking with dp4C, pH 6.0.

4.5.3 HISQC spectra of fH∼7Y and fH∼7Yxdp4C

The HISQC sequence is a variant of the HSQC sequence which is optimised for lysine ζ-NH_3^+ sidechains [248]. The presence of 5-10% deuterium for the lock signal as usually used in protein NMR can lead to complications in HISQC spectra because proton-deuterium exchange in ζ-NH_3^+-groups produces NH_2D and NHD_2 species which display isotope shifts in both dimensions and deuterium spliting in the ^{15}N dimension. These

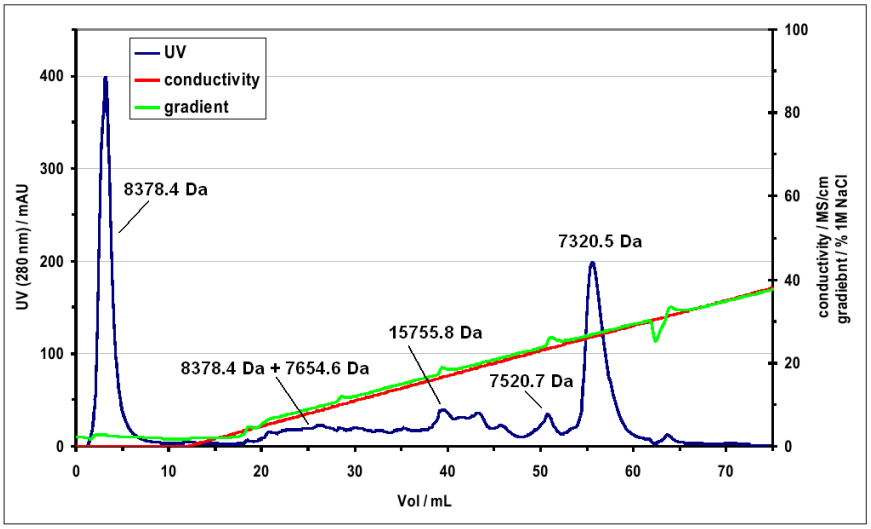

Figure 4.22: Purification of fH∼7Yxdp4C on a MonoS column at pH 6.0. The cross linked protein does not bind to the cation exchange column due to the presence of the negatively charged tetrasaccharide. Free protein elutes at the same salt concentration as previously (Fig. 4.21). A range of minor species was detected, corresponding to chemically modified fH∼7Y. The masses are approximate as they were acquired without previous calibration of the MALDI-TOF mass spectrometer.

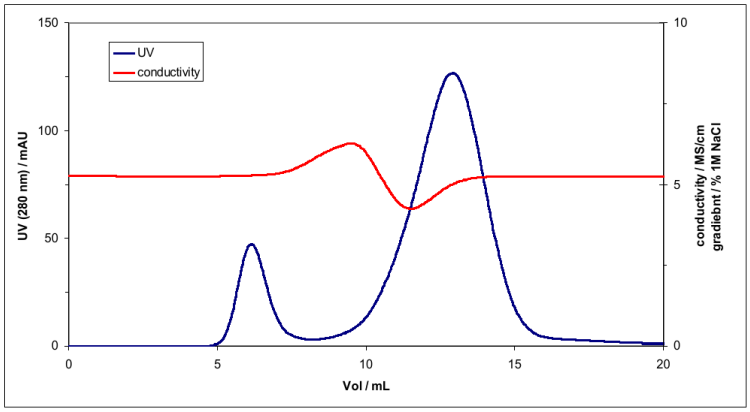

Figure 4.23: Desalting of fH∼7Yxdp4C on two consecutive HiTrap Desalting columns. The first peak corresponds to fH7Yxdp4C while the large second peak most likely contains remaining cross linking reagents.

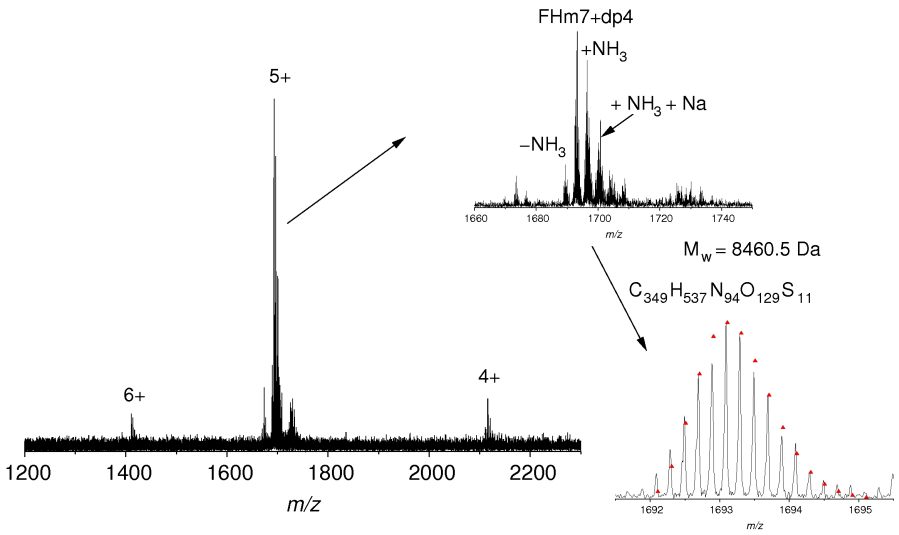

Figure 4.24: Accurate mass measurement of fH∼7Yxdp4C using tandem mass spectrometry. The simulation of a species with five postive charges (red triangles) corresponds well to the measured spectrum.

satellite peaks surround the main NH_3 cross peaks. While the NHD satellites that are usually observed for glutamine and asparagine sidechains in protein ^{1}H, ^{15}N-HSQC spectra aid in identifying sidechain signals, such satellites cause intensity loss and inconvenient overlap in HISQC spectra because the chemical shift dispersion of the $N^{\zeta}H^{\zeta}$ cross peaks is generally very small. Additionally, peak intensities can vary widely due to different proton exchange rates making the distinction between the satellites of strong peaks and the weak peaks problematic. To achieve satellite-free $N^{\zeta}H^{\zeta}$ cross peaks common 5 mm NMR tubes were replaced by capillary tubes (Norell). These consist of a 5 mm NMR outer tube and a smaller inner tube with an outer diameter of 2 mm and a volume of 60 µL. This capillary was filled with D_2O for the lock signal and no D_2O was added to the actual sample. This setup allows for the necessary number of deuterium spins for the lock signal while preventing proton-deuterium exchange. It should, however, be noted that if free hydrogen/deuterium exchange is allowed the relative intensities of the NH_3, NH_2D and NHD_2 peaks can be used to determine the flexibility of lysine sidechains, as proposed recently [250]. A practical solution to the problem can be achieved by using deuterium decoupling during the t_1 period. This removes the splitting

caused by deuterium but not the isotope shifts.

A ^{15}N-fH∼7Yxdp4C sample which had been prepared by Jon Deakin from ^{15}N-fH∼7Y and dp4C was subject to HISQC spectroscopy at pH 3.0 and 283 K. These conditions were found to be suited for the free fH∼7Y, i.e. the protein appeared to be folded at this pH as judged from the ^{1}H, ^{15}N-HSQC spectrum. At higher pH or temperature at least one of the five sidechain $N^{\zeta}H^{\zeta}$ cross peaks disappeared due to fast exchange with water. Overlay of the free and cross linked samples showed that only two of the five lysine residues were not affected by the cross linking (Fig. 4.25).

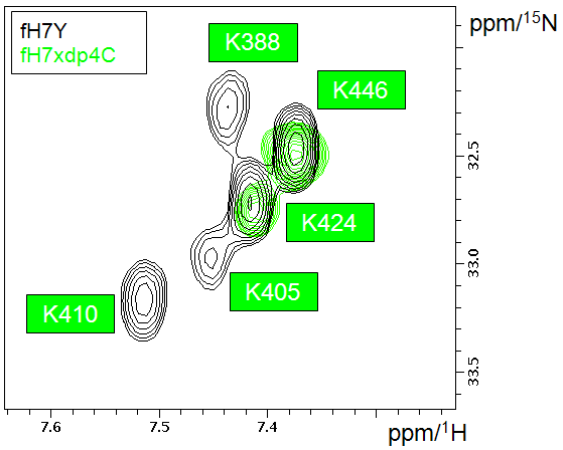

Figure 4.25: Overlay of the assigned HISQC spectra of free fH∼7Y (black) and cross linked fH∼7Yxdp4C (green). The two most C-terminal lysines, K424 and K446 seem unaffected by the cross linking.

Assignment of the HISQC spectrum was accomplished using a ^{13}C,^{15}N-sample of fH∼7Y and recording an (H)CCENH$_3$ spectrum [248]. This pulse sequence directs the TOCSY magnetisation transfer in the opposite direction compared to the flow of magnetisation in the C(CO)NH-TOCSY experiment, linking the C^{α} - C^{ϵ} resonances to the ζ-NH_3^+ group rather than to the backbone NH atoms. The spectrum was recorded using the free protein and the same pH and temperature conditions as used for the HISQC spectra. It was then compared to the conventional sidechain C(CO)NH-TOCSY spectrum of fH∼7Y, which had been assigned in our group prior to the start of this project (Fig. 4.26). The

C^{α} - C^{ϵ} chemical shifts for all lysine sidechains in the assigned C(CO)NH-TOCSY were compared to the C^{α} - C^{ϵ} chemical shifts for all residues in the unassigned (H)CCENH_3 spectra and lysine residues in the latter were identified based on the similarity of the five carbon chemical shifts in each sidechain. Comparison of the ^{13}C chemical shifts allowed transfer of the assignments despite the large difference in pH and temperature under which the original TOCSY had been recorded (pH 5.0, 298K). The similarity of the carbon chemical shifts does again point to the stability of the protein fold at pH 3.0. K446 was not present in the C(CO)NH-TOCSY spectrum because in this experiment lysine sidechain resonances are detected on the C-terminal neighbour, which is lacking for K446 in the fH∼7 construct. However, since the remaining four lysine sidechains could be unambiguously assigned using the described strategy the only remaining lysine sidechain in the (H)CCENH_3 could be positively identified as belonging to the remaining K446.

Based on the assigned HISQC spectrum of fH∼7Y the conclusion was drawn that neither K424 nor K446 was cross linked to dp4C in fH7Yxdp4C (Fig. 4.25). This result was surprising as K446 showed the largest movement of its backbone NH cross peak during the titrations with dp4C and SOS. An accurate mass measurement of ^{14}N-fH7Yxdp4C had shown that the mass of the cross linked protein corresponded to the linkage with one dp4C molecule only, thus the disappearance of the three N-terminal lysine sidechain ^{15}NH cross peaks in the HISQC spectrum could only be explained with exchange broadening. An ^{1}H, ^{15}N-HSQC spectrum of the cross linked sample did indeed exhibit large perturbations and bleaching of resonances in the whole N-terminal region of the protein (Fig. 4.27). It was therefore concluded that the tetrasaccharide was bound either to K388, K405 or K410 and that its proximity to the remaining two N-terminal lysine residues caused intermediate exchange broadening of their sidechain and backbone ^{15}NH cross peaks.

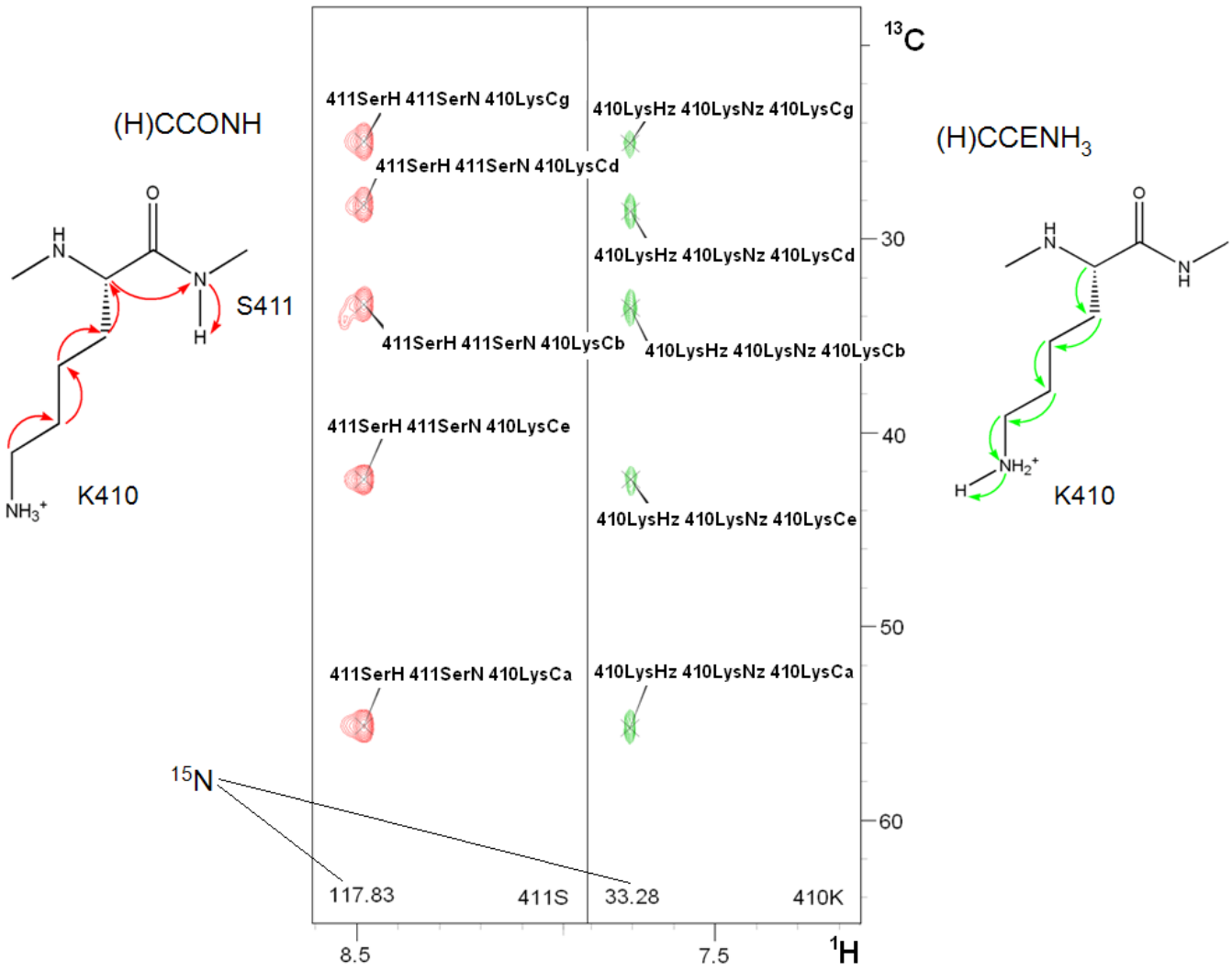

Figure 4.26: Example of the comparison of different fH∼7Y TOCSY spectra. The ^{13}C-resonances of the lysine sidechain are linked to the backbone ^{15}NH of the preceding residue in the CCONH-TOCSY (red) or the $^{15}N^{\zeta}H^{\zeta}$ of the lysine sidechain in the (H)CCENH$_3$ (green). The magnetisation transfer is shown schematically for both experiments in the respective color.

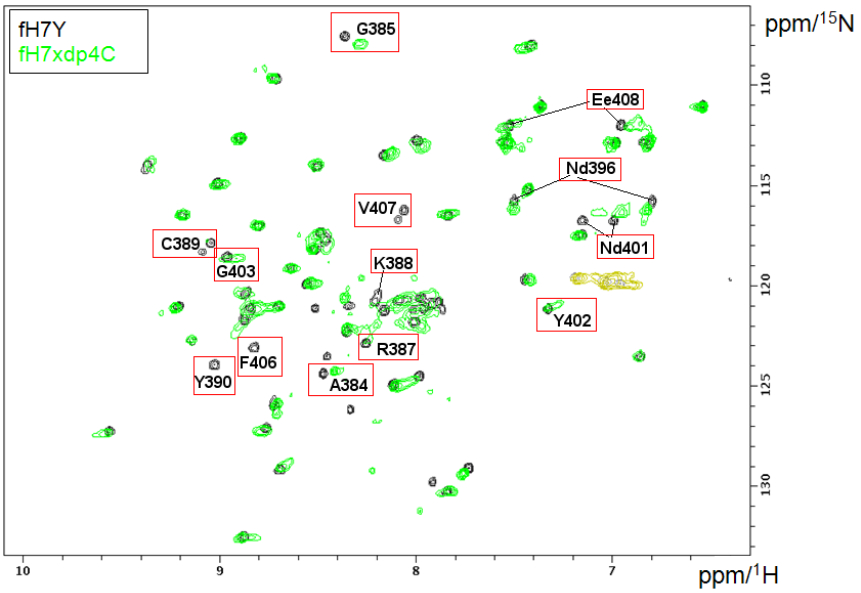

Figure 4.27: ^{1}H, ^{15}N-HSQC spectrum of free fH∼7Y (black) and cross linked fH7Yxdp4C (green). Many of the broadened and shifted NH cross peaks belong to residues in the N-terminal region of the protein (red boxes).

4.5.4 Trypsin digestion of fH∼7Y and fH∼7Yxdp4C

The analysis of the cross linked fH∼7Yxdp4C with the HISQC experiment had not unambiguously answered the question to which lysine sidechain dp4C was covalently linked, mass spectrometry and trypsin digestion were employed as complementary techniques. The protease trypsin is widely used as a biochemical tool for protein identification. It cleaves peptide bonds which lie C-terminal of arginine and lysine residues. Trypsin digestion of free fH∼7Y should thus yield two very short peptides (E_{382}AAGLR and CIRVK_{446}) for both the N- and C-terminus. These could not readily be identified by MALDI-TOF mass spectrometry due to increased noise levels below 1 kDa and small molecule impurities which made small peptide identification unreliable. Additionally, the middle stretch following residues K405 is also too short for reliable detection by MALDI-TOF (K_{405}FVQGK$_{410}$). These limitations also apply to fragments obtained through trypsin digestion of fH7Yxdp4C as long as they are not cross linked to dp4C, which would increase the molecular weight by more than 1 kDa. However, four large

peptides could clearly be identified in both samples after denaturation/reduction, alkylation and trypsin digestion (Fig. 4.28, 4.29):

(K_{388})CYFPYLENGYNQNYGR$_{404}$
S_{411}IDVACHPGYALPK$_{424}$
A_{425}QTTVTCMENGWSPTPR$_{441}$

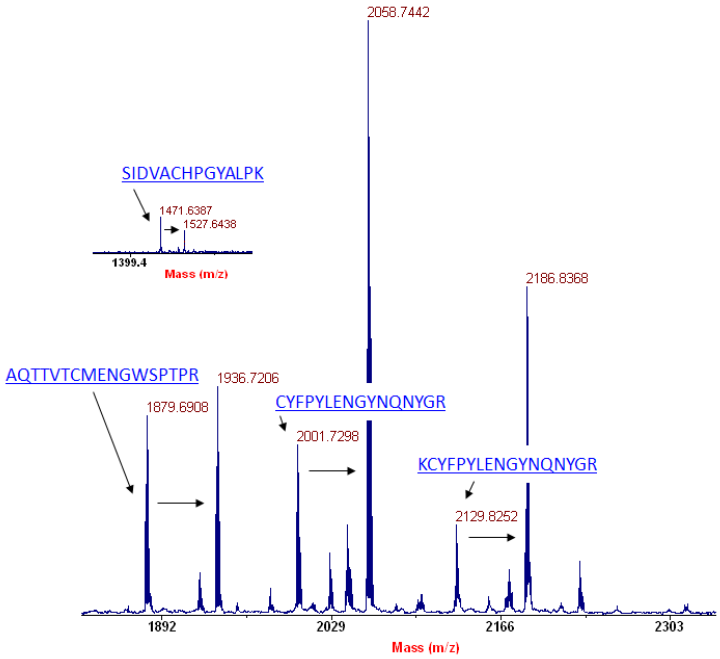

Figure 4.28: MALDI-TOF spectrum after fH∼7Y trypsin digestion. Horizontal arrows indicate alkylation products.

The presence of the first and the last cysteine residue in these peptides confirmed successful reduction. Since all three peptides were also found in the digestion of fH7Yxdp4C it was clear that neither the N-terminal lysine residue, K388, nor the hypervariable loop residue K424 was cross linked to dp4C. K410 was very likely also not cross linked because the presence of the second fragment starting with S411 proved that this residue was still recognized as lysine by the protease, which is unlikely for a lysine residue carrying a large covalent modification.

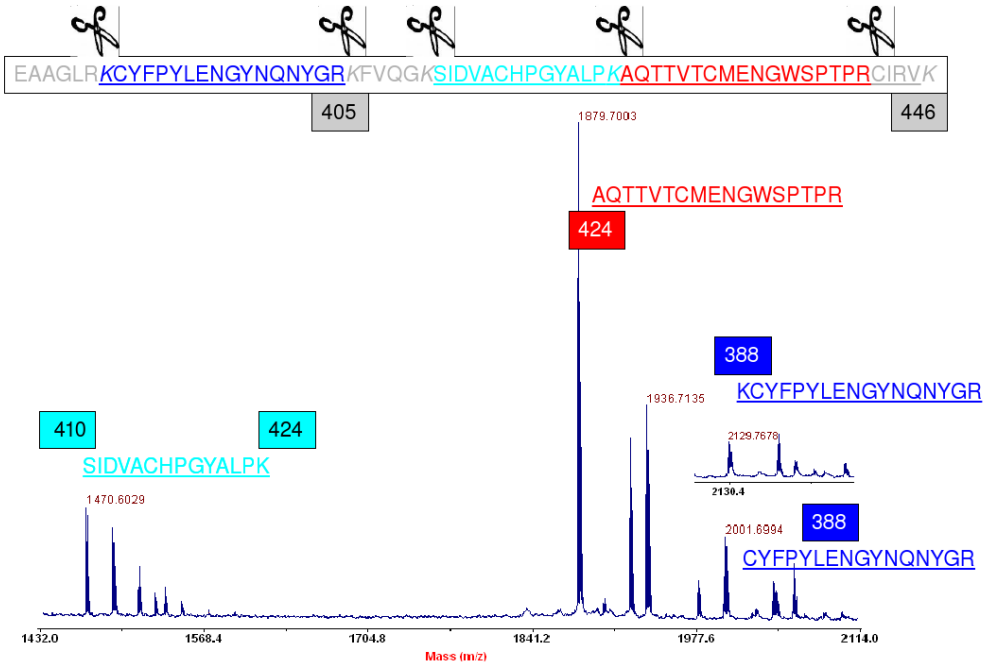

Figure 4.29: MALDI-TOF spectrum after fH7Yxdp4C trypsin digestion. Fragments which could be identified and which are not modified are highlighted in cyan, red, and blue. The framed numbers refer to the lysine residues which were identified as not cross linked on the basis of the presence of the respective fragment. The light stretches in the sequence (top) were not identified.

4.5.5 Microwave-assisted acid hydrolysis

As the last analytical step in the fH7Yxdp4C analysis microwave-assisted acid hydrolysis coupled with mass spectrometry (MAAH-MS) was used. This method provides an alternative route to protein sequencing where tandem mass spectrometry fails (which is often the case for membrane proteins). The sequence coverage was much better for the C-terminus of the cross linked protein than for the N-terminus (Fig. 4.30). The identified fragments included both the first and the last cysteine residue, again signalling that both disulfide bonds had been reduced under the drastic conditions of acid, DTT and microwave irradiation.

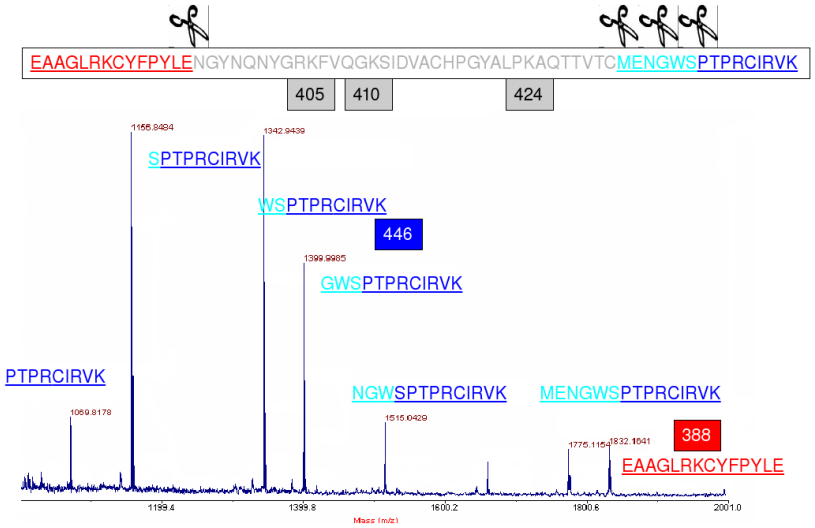

Figure 4.30: MALDI-TOF spectrum after microwave-assisted TFA hydrolysis of fH7Yxdp4C. Several C-terminal fragments (blue and cyan) and one N-terminal fragment (red) were identified. The C-terminal stretch that was contained in every single C-terminal fragment is highlighted in blue, the "growing" amino acid sequence in cyan. The framed numbers refer again to lysine sidechains present in the respective fragments and not modified.

Prolonging the time the cross linked protein was exposed to microwave irradiation did not improve the sequence coverage but increased the amount of small peptides below 1 kDa. The microwave-assisted TFA hydrolysis confirmed that residue K446 was not

linked to the tetrasaccharide, nor was residue K388. The first result confirmed the results of the NMR analysis but could not be obtained from trypsin digestion, while the latter result was in agreement with the trypsin digestion data but could not be obtained from the NMR analysis. Thus, the partial TFA-hydrolysis complemented the NMR and trypsin digestion data. Together the results from these three technique leave only one residue as the possible cross linked lysine: K405.

In the crystal structure of fH∼6-8H in complex with SOS [241] the N^{ζ} atom of K405 is in relative proximity (4.5 and 4.3 Å) to two sulfate groups of the only well-defined SOS molecule in the deposited PDB file (Fig. 4.31). Cross linking, on the other hand, requires a carboxyl group of dp4C to be in the vicinity of the lysine sidechain. SOS does not have any carboxyl groups, our results therefore hint to the existence of different binding modes for SOS and dp4C.

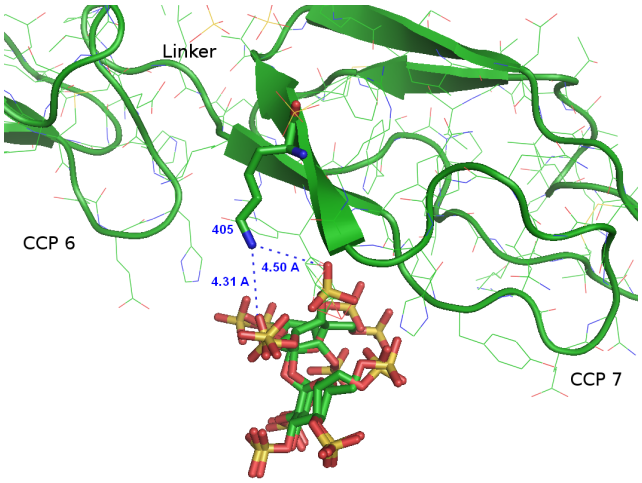

Figure 4.31: Residue 405 in the x-ray structure of fH∼6-8H in complex with SOS. Prepared from PDB 2UWN.

Trypsin cleavage before the modified residue K405 was possible because the protease identified the N-terminal neighbour, R404. The modified peptide, K_{405}(dp4C)FVQGK$_{410}$,

was not detected in the MALDI-TOF analysis of the digested fH7Yxdp4C and thus only indirect evidence leads to the conclusion that K405 was the cross linked residue. It is, however, not surprising that this fragment was not detected because of its large negative charge. After the cross linking seven negative charges remain in the dp4C moiety. This high number cannot be compensated by the positive charges which the short peptide K(dp4C)FVQGK adopts even in the presence of TFA (which is part of the MALDI-TOF matrix).

4.5.6 Titration of fH∼7Y with dp4C monitored by H_2CN experiments

During the titration of fH∼7Y with dp4C at pH 7.4 and 298 K none of the NH_3^+ cross peaks was detectable by the HISQC experiment. At lower pH addition of dp4C caused sample precipitation. A way around this problem was the usage of double labelled protein which allowed following the ^{15}N chemical shift changes during the titration using the H_2CN pulse sequence with the excitation and detection of H^ϵ protons [251].This pulse sequence links the N^ζ chemical shift to the C^ϵ and H^ϵ chemical shifts of lysine sidechains. This procedure allowed use of physiological conditions for the binding experiments. ^{13}C,^{15}N-fH∼7Y was titrated with dp4C and 2D planes of the H_2CN sequence were recorded. The assignment was transferred on the basis of the known N^ζ and C^ϵ chemical shifts. The results of this titration are displayed in Fig. 4.32. It can be seen that the $H^\epsilon C^\epsilon$ cross peaks of K446 and K424 did not change position during the titration. This is consistent with the results of the cross linking and the HISQC experiments. The chemical shifts of the remaining three residues were affected, starting with the smallest to the largest change in the order K410 < K388 < K405. A mutational study conducted on the N-terminus of fH found that the mutations K405A and R404A in fH∼6-8 each had a significant effect on binding to different heparin affinity columns, while K388A resulted in only a small decrease in this affinity for certain heparin preparations [252].

This study thus supports our observation that K405 is a central player in the N-terminal heparin binding site in CCP7 and validates the use of the H_2CN pulse sequence for the analysis of protein-GAG complexes.

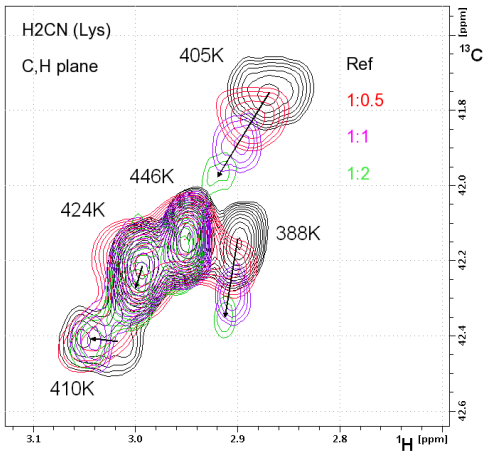

Figure 4.32: H_2CN titration of fH∼7Y with dp4C (CH-planes). The ratios given in the respective colours are protein:ligand. The least affected peaks are again K424 and K446.

The analysis of the cross linked complex, together with the presented ^{1}H ,^{15}N-HSQC and H_2CN titrations of fH∼7Y, lead to the conclusion that the backbone NH but not the sidechain NH_3^+ group of K446 contributes to heparin binding. It cannot be entirely excluded that K446 does not directly contribute to heparin binding at all and that the chemical shift changes observed in the ^{1}H ,^{15}N-HSQC titration are due to an allosteric effect. This is, however, an unlikely scenario because of the role of the nearby R444 sidechain in the crystal structure of fH∼6-8H in complex with SOS. Titrations have been conducted in our group using a TEMPO-radical labelled heparin disaccharide which caused paramagnetic relaxation enhancement of the K446 backbone NH cross peak, indicating the binding of the oligosaccharide at this position.

4.6 Discussion of heparin binding to the fH∼6-8 region

4.6.1 NMR methods for monitoring of lysine sidechains

It was found that the HISQC sequence was not useful for the monitoring of lysine sidechains in the non-covalent fH∼7Y:dp4C complex, possible due to unsufficient supression of exchange with water. Spectra were recorded at different temperatures and pH values to determine conditions that could be used for the assignment of all lysine $N^{\zeta}H^{\zeta}$ cross peaks in fH∼7Y. Only at 283 K and pH 3.0 was the water exchange sufficiently slow to observe and assign all $N^{\zeta}H^{\zeta}$ resonances. The fact that none of the lysine sidechains was observed in the HISQC experiment of the non-covalent fH∼7Y:dp4C complex at pH 7.4 and room temperature underlined the weakness of the interaction, which was found to have a K_d around 40-50 μM in the ^{1}H, ^{15}N-HSQC titration of fH∼7Y with dp4C. The conditions under which all lysine sidechain resonances were observed using the HISQC sequence were too acidic to be used for any physiological interpretation of the complex and where thus only used for the characterisation of the cross linked complex. In addition, protein precipitation was observed upon addition of the dp4C at pH 3.0.

In order to follow the sidechain signals during titrations at more physiological conditions another lysine sidechain experiment was employed, the so-called H_2CN experiment. This is an out-and-back type experiment which does not rely on the exchanging protons for excitation or detection but - unlike the HISQC experiment - requires ^{13}C labelling. In this experiment the aliphatic sidechain protons H^{ϵ} are excited and the magnetization is transferred to ^{13}C in an INEPT step, followed by a second INEPT transfer to ^{15}N. Nitrogen chemical shift labelling takes place while proton and carbon couplings do not evolve due to decoupling pulses on both the ^{1}H and ^{13}C channels. Following the t_1 period the magnetisation is back transferred to ^{13}C via another INEPT step, labelled with the ^{13}C chemical shift and eventually transferred back to the H^{ϵ} protons where the detection takes place. By leaving out the incrementation during the t_1 or t_2 period this experiment can be set up as a 2D ^{13}C,H- or ^{15}N,H-experiment, respectively

(dubbed $H_2C(N)$ and $H_2(C)N$). Comparison of the N^{ζ} chemical shifts in the 2D $H_2(C)N$ experiment with the previously assigned HISQC experiment allowed assignment of the N^{ζ} signals in the $H_2(C)N$ spectrum. Several $H_2(C)N$ spectra were recorded at different temperatures and pH values in order to transfer the assignment to physiological conditions, where a heparin titration was performed. The C^{ϵ} resonances in the $H_2C(N)$ spectrum were assigned based on the H^{ϵ} shifts in the $H_2(C)N$ spectrum. While they served as an additional assignment test via comparison of C^{ϵ} resonances in $(H)CCENH_3$ and CC(CO)NH spectra they cannot be relied on as a sole instrument for the assignment of the N^{ζ} resonances because of their generally very small chemical shift dispersion.

4.6.2 Heparin binding to the fH∼6-8 region

During this study it was found that heparin derived fully sulfated tetrasaccharide (dp4C) initiates similar chemical shift changes when titrated into fH∼7-8 (both variants) as the synthetic heparin mimic sucrose octasulfate. This finding served as justification for its use in a crystallographic study of fH∼6-8H in complex with SOS by our collaborators [241]. The backbone NH chemical shift changes during the titrations were almost indistinguishable for both ligands and both protein variants, fH∼7-8Y and fH∼7-8H. While on the first glance these findings seem surprising - given the large structural differences between dp4C and SOS - they are not incompatible with a model in which the heparin protein complex is dominated by charge-charge interactions: SOS has the same number of negative charges as dp4C, and because of its smaller size the charge density in SOS is even higher. It should not be forgotten that the possibility of binding non-physiological ligands to proteins is the basis of a large branch of pharmaceutical research and drug development, and that the binding affinity of SOS to fH∼7-8 is significantly lower than the affinity of dp4C, i.e. a larger amount of ligand is needed to achieve similar chemical shift changes.

A number of findings and conclusions from the x-ray study of fH∼6-8H with SOS were not supported by the previous work from our laboratory and or results obtained during

this project. One conclusion that was drawn from the crystal was a possible steric clash between the tyrosine sidechain of residue 402 and a SOS molecule [241]. In the crystal, the histidine sidechain of residue 402 in fH∼6-8H directly coordinates a SOS molecule. The hypothesis that a tyrosine sidechain could not maintain this interaction and would even sterically hinder the SOS binding to this particular binding site (there are several SOS binding sites in the x-ray structure) seemed supported by the failure of crystallization trials with fH∼6-8Y and SOS. Only the H402 variant was successfully crystallized and only in the presence of SOS, leading to the conclusion that a "stiffening" effect of SOS was needed for crystallization and that this effect could not be obtained in the tyrosine variant. We used the Y402 variant in the NMR and cross linking studies, yet the results of these experiments suggested binding of dp4C in the vicinity of 402, the critical tyrosine residue. If the isolated CCP 7 module alone (as used in the cross linking and H_2CN titration study) is able to hold a ligand molecule in place at the N-terminus of CCP 7 this binding site should also be occupied in the presence of CCP 6, which - according to the crystal structure - contributes to this binding site via a histidine residue in the C-terminal half of CCP 6. The assumption that binding to the binding site observed in the crystal structure of fH∼6-8H with SOS could not be accomplished by the tyrosine variant seems thus questionable. A binding study with different types of heparin/HS observed that each isoform of fH∼6-8 exhibits a higher affinity than the other one, depending on the type of heparin/HS that was used [252]. The solution structures of both fH∼7 variants [240] show that the overall structure of CCP 7 is not disturbed by the SNP Y402H. The lack of a single salt bridge from residue H402 to the ligand in the tyrosine variant might reduce the affinity of the tyrosine variant for SOS and heparin but most likely does not exclude formation of the complex, as proven by the experiments presented in this thesis.

More open questions surround the role of CCP 6 in the binding of heparin to fH. Unfortunately, fH∼6-7 proved to be too unstable for NMR spectroscopy. Thus, we were unable to conduct NMR studies on this construct in order to clarify the contributions

of CCP 6. In the crystal structure, this module seems to contribute more to the SOS binding than CCP 7 or CCP 8: one entire SOS binding site is contained in CCP 6, and the binding site involving the critical residue H402 is completed by a second histidine residue in CCP 6, H360. However, the binding studies which were presented along with the x-ray structure do, to some degree, contradict the conclusions drawn from the fH∼6-8H:SOS structure. The constructs fH∼5-6 and fH∼6-7 elute at lower salt concentrations from a heparin column than fH∼7-8, and the latter elutes at approximately the same salt concentration as fH∼6-8 (Fig. 3 in [241]). This behaviour points to larger binding contributions from CCP 8 than CCP 6. Additionally, a titration of partially deuterated ^{15}N-fH∼6-8Y with SOS in our laboratory did not find evidence for the large binding contributions of CCP 6 that were observed in the crystal. While CCP 6 was not assigned in this NMR study a comparison with the fH∼7-8Y titration revealed that most unassigned amide cross peaks (i.e. resonances from CCP 6) remained in place, confining the binding amides to CCP modules 7 and 8 (data not shown). Additional experiments using both heparin (in form of a heparin affinity column) or SOS (in the ^{1}H, ^{15}N-HSQC titration with fH∼6-8) do not therefore fully support the interpretation of the x-ray structure. A possible explanation for the potentially over rated involvement of CCP 6 in the binding is the particular orientation of protein molecules relative to each other seen in the protein crystal. While protein-ligand interactions in [241] are displayed and analysed using an isolated protein molecule in fact several protein molecules contribute to the binding of each SOS molecule in this crystal: the overall ratio of protein:ligand is 1:1, despite the observation of several SOS binding sites per protein molecule. This arrangement potentially allows minor contributions to the SOS binding from residues outside of the binding site(s) as observed in the liquid phase.

These arguments touch the fundamentals of x-ray crystallography - molecules need to interact in a network-like fashion to form crystals, and such contacts cannot simply be avoided in protein crystallography. In many cases, some of the interactions seen in crystals are representative of physiological interactions, e.g. in cases where proteins need to self-associate to be functional. Growth factors and virus capsid proteins, for example,

are known to self-associate (to dimers and multimers, respectively) in their physiological environment and crystal structures are representative of these interactions. Generally, by measuring the buried surfaces in protein crystals and with the help of complementary techniques it is possible to differentiate between physiological protein oligomerisation and additional contacts between multimers which only reflect crystal packing. This process is, however, often disputed in protein-ligand complexes and even more difficult when the ligand binds in a superficial manner, i.e. does not enter a binding pocket. The fH∼6-8H-molecules are "glued" together by SOS molecules in the reported crystal and the asymmetric unit contains at least three protein molecules. Which of the observed interactions can be maintained by a monomeric protein in solution need to be answered by complementary methods.

4.6.3 The protein-heparin cross linking strategy

In this study, a zero-length cross linking strategy previously employed to link protein-protein or protein-peptide complexes was tested on a protein-heparin complex. Heparin tetrasaccharide was activated with common chemical cross linking reagents and, after a basic purification step, mixed with fH∼7Y in a biochemical buffer at pH 6.0. The resulting hybrid was identified by chromatography (the covalent complex failed to bind to a cation exchange column) and mass spectrometry. While an accurate mass measurement proved a 1:1 stoichiometry, three lysine sidechain amine cross peaks disappeared from the HISQC spectrum upon cross linking. In theory, one single cross peak should disappear from the HISQC spectrum of the cross linked species and an additional peak should appear in the ^{1}H, ^{15}N-HSQC spectrum due to the creation of a new peptide bond. Instead, not only did three peaks disappear from the HISQC spectrum, but many more backbone amide cross peaks were broadened beyond detection in the ^{1}H, ^{15}N-HSQC spectrum, and additionally distored peak shapes were observed in this spectrum. Assignment of the HISQC spectrum (the ^{1}H, ^{15}N-HSQC spectrum had been assigned previously [240]) revealed that the observed changes in both spectra clustered at the N-terminus of fH∼7Y. Further investigation using mass spectrometry, trypsin digestion

and microwave-assisted acid hydrolysis yielded the following result: of the five lysine sidechains, four could unambiguously be identified as not affected in either the HISQC and/or yielding unmodified peptide fragments in the trypsin digestion/acidic hydrolysis. The only lysine sidechain which was not identified by either method was K405. These results are in good agreement with the fH∼6-8H:SOS crystal structure where this sidechain directly coordinates a SOS molecule. Thus, the techniques presented in this work have the potential to become useful tools in the characterisation of glycosaminoglycan-protein interactions. Together with ^{1}H, ^{15}N-HSQC- monitored ligand titrations these techniques are able to identify and distinguish between binding contributions from lysine sidechain amines and lysine backbone amides, thus complementing the common chemical shift mapping approach which relies on the analysis of ^{1}H, ^{15}N-HSQC spectra only.

It should be noted that, despite significant efforts using different mass spectrometry techniques, no fragmentation of the cross linked hybrid was accomplished by tandem mass spectrometry. The intact hybrid gave a broadened peak in the MALDI-TOF spectrum which was in fact made up of a number of peaks differing by 80 Da, arising from species with different numbers of SO_3 groups. After trypsin digestion and acid hydrolysis, no peptide could be identified which contained the heparin tetrasaccharide, and the cross linked fragment could only be identified in an indirect manner. All these difficulties are probably due to the large negative charge of the tetrasaccharide. This charge prevents the intact complex from accumulating enough positive charges to fragment in tandem mass spectrometry. For small peptides containing the tetrasaccharide the situation is even worse because the number of residues available for the protonation through the acidic matrix is smaller. Such fragments are probably not observed in a common positive mode MALDI-TOF because they cannot adopt a positive net charge. Efforts were made to use negative mode MALDI-TOF but the signal to noise ratio was worse and no peptides could be identified at all. Sulfated glycosaminoglycans are challenging targets for mass spectrometry because of their large negative charge and the lability of sulfate groups. It remains to be seen if the development of new methods in this field will aid

the identification of such compounds in the future. Such development would without doubt open a large range of possibilities for studies of heparin-protein complexes.

5 Studies on fH∼19-20

5.1 NMR binding studies of fH∼19-20

FH acts as a cofactor to factor I which proteolytically inactivates C3b covalently bound to host cell surfaces. Binding sites for C3b have been mapped to both the N- and the C-terminus of fH. While the N-terminus contains the regulatory domains of fH (CCP 1-4), which provide a platform for factor I binding [253], the C-terminal domains CCP 19 and CCP 20 are not directly involved in the inactivation of C3b. They do nevertheless play a critical role in the C3b deactivation process because the N-terminal modules of fH do not contribute to the cell recognition process. Thus, the C-terminus of fH acts as a sort of anchor for the protein, directing the functional domains CCP 1-4 into proximity of host cell-surface bound C3b which requires deactivation. This section summarizes a number of experiments which were conducted in order to elucidate the interplay of heparin oligosaccharides (which serve as a model for cell surfaces), C3b, and fH∼19-20.

The binding site on C3b for the C-terminal fH modules has previously been shown to reside in the so-called TED domain [67]. This domain, which contains the covalent attachment point of C3b, remains bound to cell surfaces after proteolytic C3b inactivation. It is then labelled C3d. Because the TED domain does not significantly alter its structure during the progression from C3b to C3d [254, 255] C3d was used instead of C3b in some of the following experiments. C3d is a convenient replacement for C3b because the latter has a large molecular weight of 175 kDa while C3d is much smaller (35 kDa). Binding of C3b to ^{15}N-fH∼19-20 caused uniform broadening and disappearance of many

peaks in the ^{1}H, ^{15}N-HSQC spectrum of fH∼19-20 because of the slow molecular tumbling of the resulting complex. This problem was circumvented by the use of C3d in lieu of C3b.

The backbone assignment of fH∼19-20 was obtained from Dr. Andrew Herbert [79] and transferred to different pH-values and salt concentrations using a set of pH and salt titrations. A number of assignments were lost during this procedure due to ambiguity caused by overlap or peak broadening. Residues whose cross peaks could not be unambiguously assigned are omitted in the following histograms.

5.1.1 Binding of C3b and C3d to fH∼19-20

To avoid working with a very concentrated stock of C3b, a 1:1 ratio of C3b and ^{15}N labelled fH∼19-20 was prepared and subsequently diluted with ^{15}N fH∼19-20. CRINEPT-HMQC-TROSY was employed during the titration because of the large molecular weight of the complex (around 200 kDa). Although the sensitivity of the spectra improved when compared to conventional ^{1}H, ^{15}N-HSQC spectra many peaks remained broadened and could not be detected. The relative small gain in sensitivity when using CRINEPT-HMQC-TROSY was likely due to the fact that the used proteins were not deuterated. In non-deuterated proteins efficient relaxation pathways remain in the form of many dipole-dipole interactions between NH protons and other proximate protons. CRINEPT and TROSY experiments are most effective when combined with deuteration [214]. In general, only cross peaks of the most flexible regions of fH∼19-20 were observed in early titration points, i.e. sidechains and terminal resonances. Upon dilution with increasing amounts of ^{15}N fH∼19-20 the peak broadening decreased in a dose-dependent manner (Fig. 5.1).

This observation is in accordance with a binding equilibrium in which the percentage of the free (smaller) protein increases, contributing more and more to the intensity of the

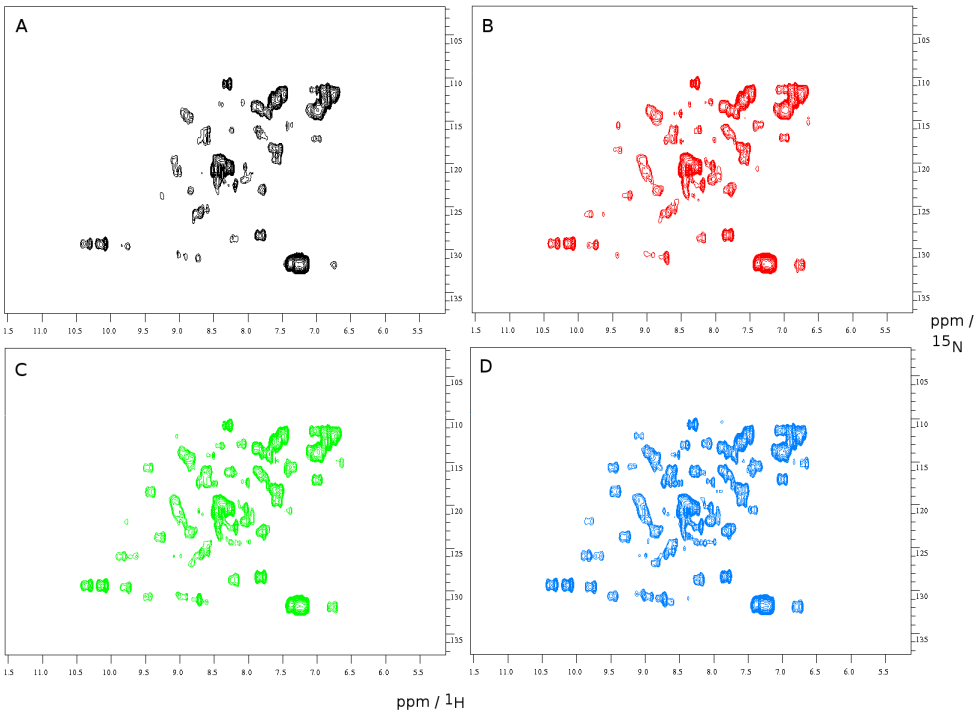

Figure 5.1: CRINEPT-HMQC-TROSY spectra of ^{15}N-fH∼19-20 and C3b. C3b:fH∼19-20 ratios were A) 1:1 (black), B) 1:1.3 (red), C) 1:1.8 (green) and D) 1:2 (blue.)

cross peaks which represent weighted averages of free fH∼19-20 and fH∼19-20:C3b complex. Because many peaks are only visible at excess of fH∼19-20 meaningful chemical shift changes, which are also weighted, could not be observed in this titration. Deuteration of ^{15}N-fH∼19-20 would have been needed to improve the spectral quality of the ∼200 kDa complex. A titration with the smaller C3d was conducted instead as C3d's molecular weight is not large enough to decrease the molecular tumbling of the protein-protein complex in the dramatic manner in which C3b does.

In a sample of C3d and ^{15}N fH∼19-20 (ratio 1.6:1) small chemical shift changes and significant line broadening of a defined set of backbone amide cross peaks were observed. Several residues could be identified which exhibited larger than average line broadening (see histogram below) or disappeared completely (highlighted residues in Fig. 5.2, blue residues in Fig. 5.4). This behaviour is a sign of intermediate chemical exchange which requires sufficient chemical shift difference between the bound and the free state and which has been reported before for other macromolecular interactions [256]. The broadened residues can be assumed to be at the interface of the complex formed between fH∼19-20 and C3d or to be indirectly affected by structural changes in fH∼19-20 upon binding to C3d. Residues whose backbone amide cross peak broadening was above the average broadening are listed in Table 5.1.

To quantify the individual intensity loss due to line broadening in the presence of C3d, a plot of peak volumes (and, for control, peak heights, which are not shown) for the individual fH∼19-20 amide backbone residues, compared to the respective values in a reference spectrum was created (Fig. 5.3). The average loss in peak volume was 85 %. This large general signal loss was partially due to the decreased concentration of fH∼19-20 in the sample with C3d. Possibly the larger molecular weight of the protein:protein complex did also contribute to a general loss in peak volume. Those residues whose signal loss was above the average value of 85 % are listed in Table 5.1 and highlighted on the NMR structure of fH∼19-20 (Fig. 5.4).

Generally, residues whose backbone amide cross peaks were significantly broadened (and

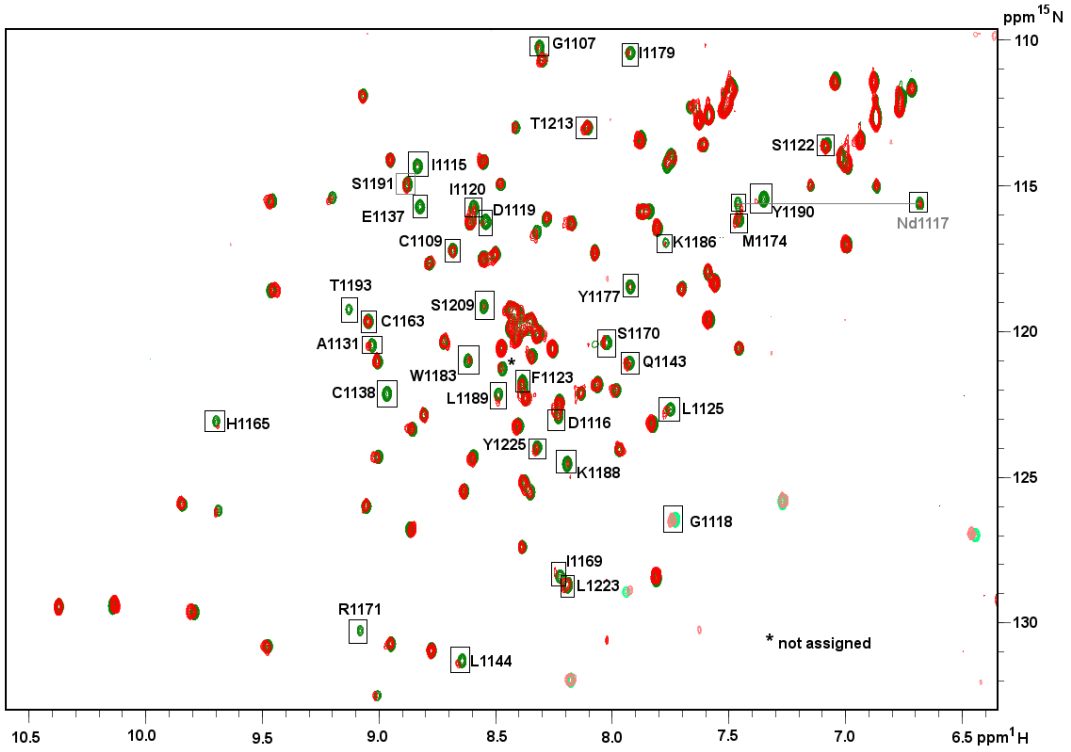

Figure 5.2: Comparison of ^{1}H, ^{15}N-HSQC spectra of fH∼19-20 in the absence of C3d (green) and with a 1.6fold excess of C3d (red). Cross peaks which exhibit larger than average line broadening in the complex with C3d are labelled. The grey line connects the resonances form a diastereotopic sidechain (residue N1117).

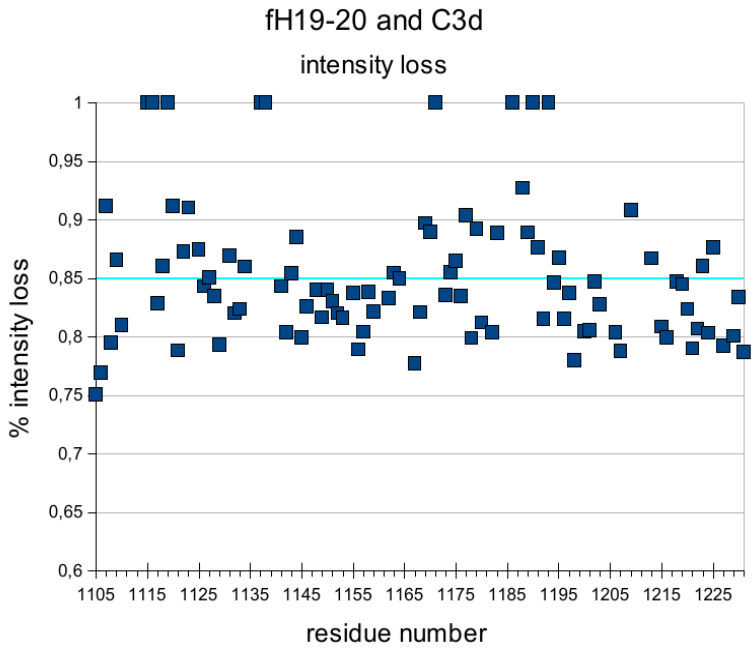

Figure 5.3: Histogram showing the impact of C3d on the signal intensity of amide backbone cross peaks in fH∼19-20. Cross peaks which are completely lost occur at y-axis value 1. The C3d:fH∼19-20 ratio was 1.6:1. The average peak volume loss in this experiment was 85 % between a fH∼19-20 reference spectrum and the complex due to higher protein concentration in the reference sample. Larger than average broadening was observed for residues above the cyan line.

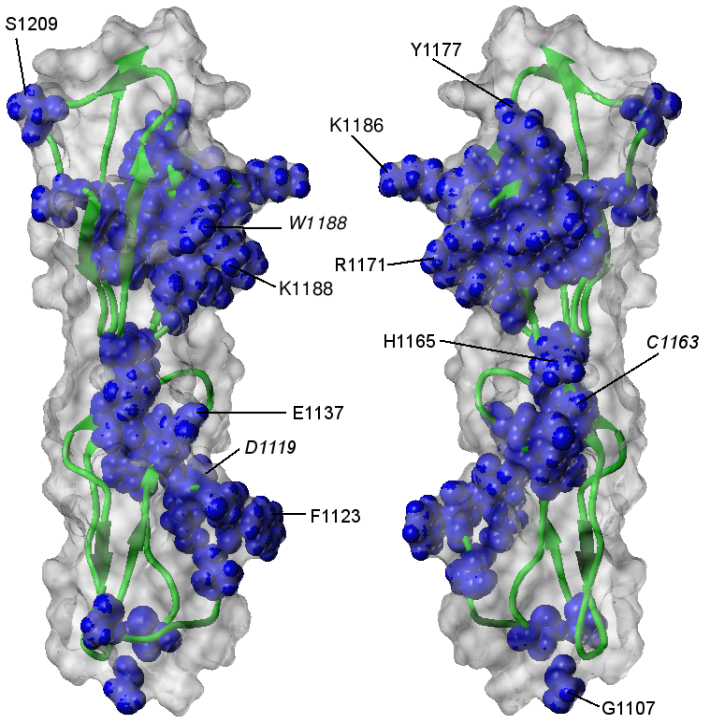

Figure 5.4: NMR structure of fH∼19-20 (PDB 2BMZ) (rotated by 180° between both views) with residues highlighted whose NH backbone cross peaks by broaden more than the average value in the presence of C3d (dark blue). These residues are also listed in Table 5.1. For clearness, only a subset of the residues is labelled to facilitate orientation. Italics type highlights residue which have been linked to aHUS. The C-terminus is at the top of the structures, the N-terminus at the bottom.

which, in some cases, experienced chemical shift changes) in the presence of C3d stretch along the full length of both CCP 19 and 20 but are mainly restricted to one face of the double module (Fig. 5.4). Interestingly, a number of perturbed residues are part of the linker between CCP 19 and CCP 20 or located in its close proximity. This could be a possible hint to a kink in the linker in the C3d-bound form of fH∼19-20 as seen for the linker between fH modules 3 and 4 in complex with C3b [253]. Unlike the C-terminal heparin binding site in fH, which is restricted to CCP 20, the C3d binding site covers both CCP 20 and CCP 19, lending an explanation to the observation of aHUS linked mutations in CCP 19.

A mutational study [257, 258] found decreased C3b/C3d affinities in the following variants of fH∼19-20: Q1139A, W1157L, R1182A, W1183L, T1184R (reduced binding to C3b only), K1188A, R1206A, and R1210A. Of these residues, two main chain NH cross peaks were not assigned/visible in ^{1}H, ^{15}N-HSQC spectra under conditions used here (T1184 and R1210). The main chain crosspeak of Q1139 could not be assigned under titration conditions but its sidechain NH_2 peaks were observed. Details of the fH∼19-20 ^{1}H, ^{15}N-HSQC spectrum in the presence of 1.6fold excess of C3d are shown for some of these residues in Fig. 5.5: for Q1139, a slight decrease in intensity was observed for one of the two diastereotropic sidechain amide protons (Fig. 5.5(c)). The main chain cross peak of W1157 appears not affected by the presence of C3d, its sidechain NH^{ϵ} cross peak, however, undergoes a small chemical shift change (Fig. 5.5(b)). R1182 is affected as well as the W1183 main chain peak, which belongs to the significantly broadened set of cross peaks (Fig. 5.5(a)). K1188 is also among the residues whose cross peaks are significantly attenuated in the presence of C3d, but K1188 continues to exhibit chemical shift changes when dp8 is added (Fig. 5.14, 5.5(g)). The R1206 main chain cross peak undergoes a chemical shift change comparable to the one experienced by the W1157 sidechain peak (Fig. 5.5(e)). Although R1210 could not be observed in the ^{1}H, ^{15}N-HSQC spectra the observation that the mutation R1210A leads to a decrease

Table 5.1: fH∼19-20 residues affected during titrations with C3d. Residues which have directly been linked to aHUS are highlighted in italics type.

CCP 19	CCP 20
G1107	I1169
C1109	S1170
I1115	R1171
D1116	M1174
G1118	Y1177
D1119	I1179
I1120	*W1183*
S1122	K1186
F1123	K1188
L1125	*L1189*
A1131	Y1190
E1137	S1191
C1138	T1193
Q1143	S1209
L1144	T1213
C1163	L1223
H1165 (linker)	Y1225

in C3d/C3b binding is important in the light of our results because the neighbouring residue, S1209, was found to exhibit line broadening as a result of binding of C3d. On the other hand, there is a mutation which was not found to have an impact on C3b/C3d binding in [257] but whose backbone NH cross peak is significantly broadened in the presence of C3d. This residue is D1119 (changed to G1119 in the mentioned study). The dramatic effect of C3d on D1119 can be seen in Fig. 5.5(f). R1215Q was described as a mutation which has no effect on C3b/C3d binding, which is in agreement with the fact that it is not broadened nor experiences chemical shift changes in the presence of C3d (Fig. 5.5(d)). The assumption that D1119 is a C3d-binding residues can lend an explanation to the observation that a mutation in CCP 19, D1119G, is a risk factor for aHUS [259]. Although no altered affinity had been found for C3d/C3b binding in [257], the same mutation had dramatic effects in a physiological complement assay [260] (see below).

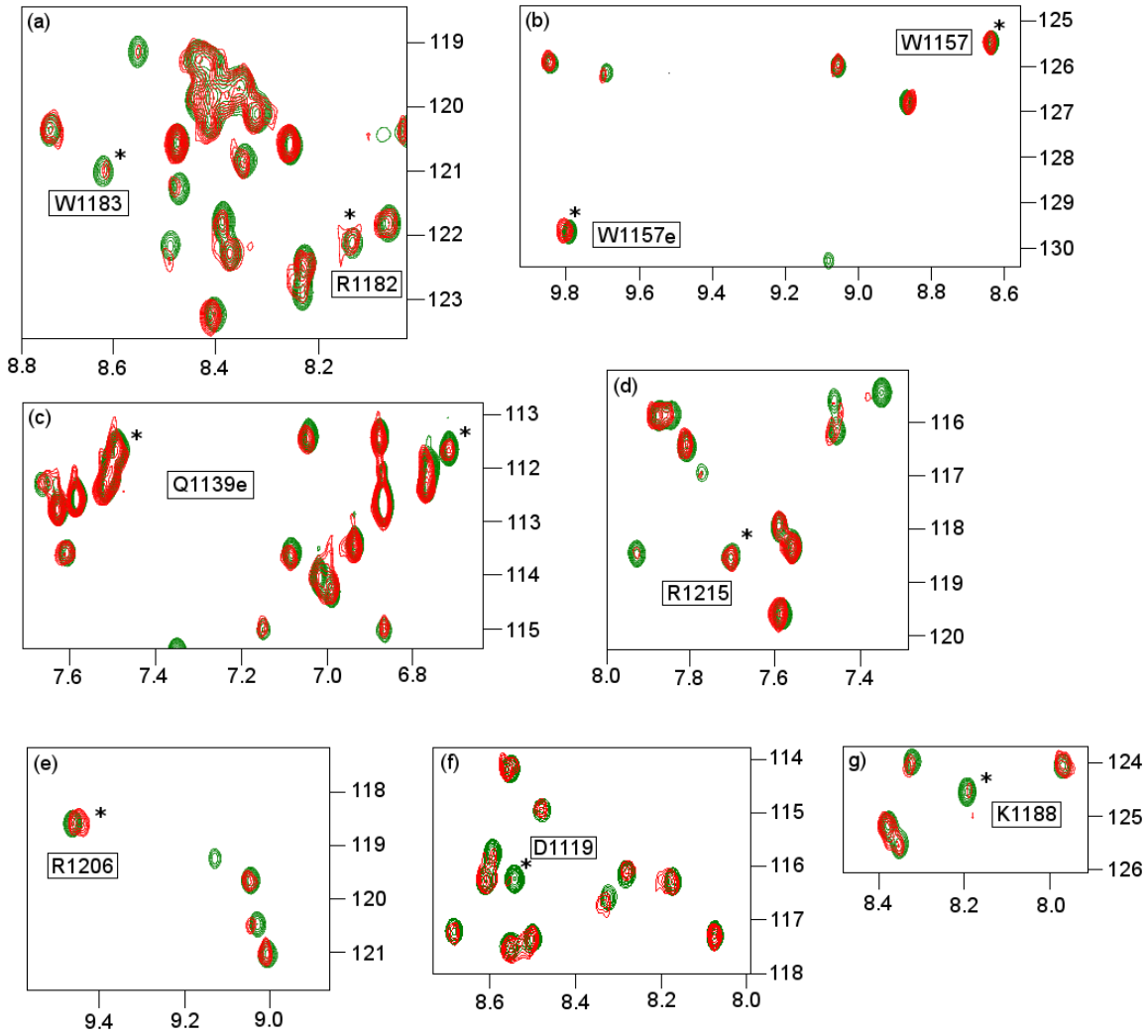

Figure 5.5: ^{1}H, ^{15}N-HSQC spectra of free ^{15}N fH∼19-20 (green) and in the complex with C3d (red) (Details of Fig. 5.2). The vertical axes correspond to ^{15}N, the horizontal axes to ^{1}H frequencies in ppm. (a) Backbone NH cross peaks for residues R1182 and W1183, (b) backbone NH and sidechain NH$^{\epsilon}$ (He) cross peaks for residue W1157, (c) sidechain NH_2 cross peaks for Q1139, (d)-(f) backbone NH cross peaks of R1215, R1206 and D1119, respectively. The mentioned cross peaks are highlighted with asterisks in the respective spectra.

5.1.2 ^{1}H, ^{15}N-HSQC monitored dp8-titration of fH∼19-20

In order to map the heparin binding site onto fH∼19-20 a titration with heparin derived octasaccharide, dp8, was carried out. dp8 was chosen in order to judge if the previously described heparin binding site (from a titration with dp4C, [79]) would turn out to be more extended when a longer oligosaccharide was used. The titration with dp8 revealed that this interaction, like the dp4C interaction with CCP 7 and CCP 7-8, falls into the fast exchange regime (Fig. 5.6, 5.7).

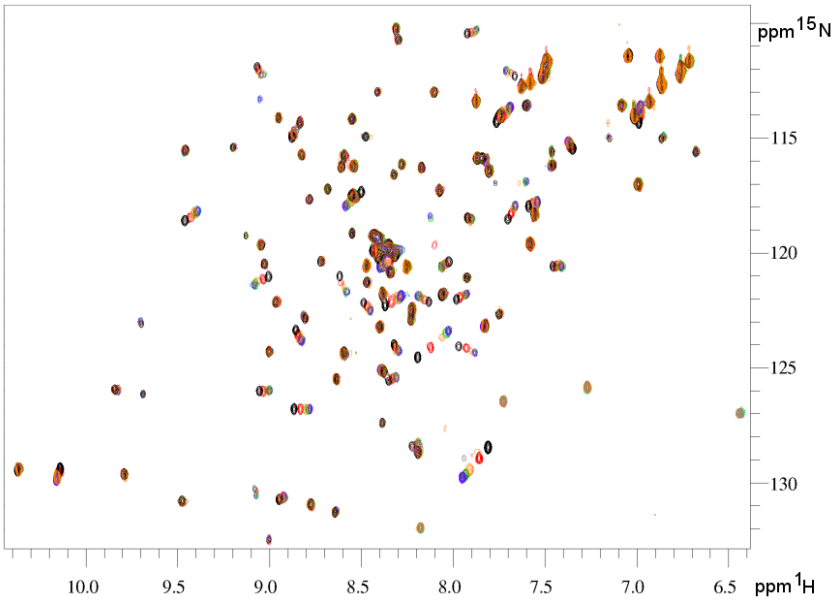

Figure 5.6: ^{1}H, ^{15}N-HSQC monitored titration of fH∼19-20 with dp8. Protein:ligand ratios are 1:0 (black), 1:0.5 (red), 1:1.1 (orange), 1:2.1 (green), 1:4.3 (blue) and 1:8.5 (purple).

When compared to the previously published chemical shift changes that were observed in an ^{1}H, ^{15}N-HSQC monitored titration of ^{15}N-fH∼19-20 and dp4C [79] a similar pattern was observed but the different absolute chemical shift scales are strikingly different (Fig. 5.8). Both oligosaccharides caused significant chemical shift changes in CCP 20 only, and numerous residues are significantly affected in both titrations, among them residues 1182-1188, 1198-1204, 1215, and 1227-1231. On the other hand, there are also significant differences in the ranking of affected residues: for dp4C, the C-terminal residue R1231 shows the largest shift changes (∼ 0.18 ppm), followed by T1184, while in the dp8

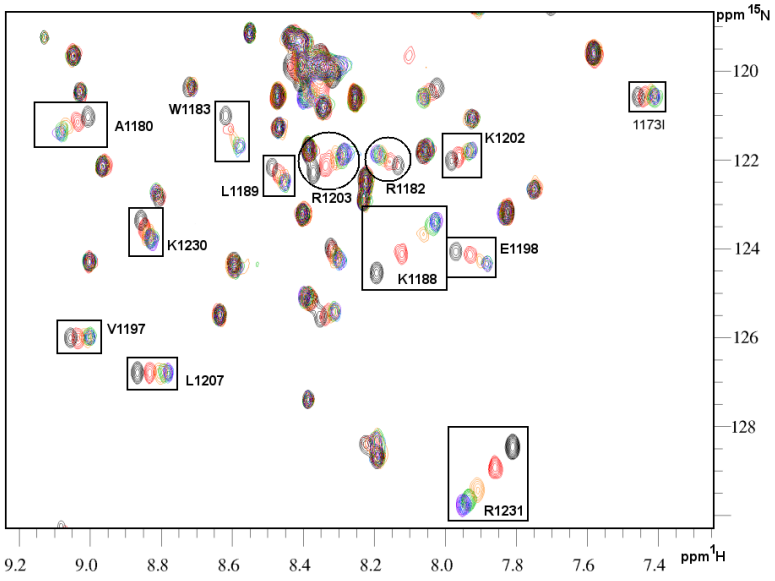

Figure 5.7: Detail of Fig. 5.6. Negative (folded) peaks were omitted for clarity.

titration residue Q1187 (∼ 0.42 ppm) is the most affected residue, followed by R1231 and K1188.

A detailed comparison of both oligosaccharide titrations is hindered by the incompleteness of both data sets. It can be seen that both oligosaccharides bind to the same face on CCP 20, although the same residues contribute differently to binding of the different oligosaccharides. In this context it should be emphasised that dp4C is a pure (fully sulfated) tetrasaccharide while for dp8 the heparin octasaccharide fraction was pooled without further purification. Additionally, different pH and salt conditions were used for these studies: the dp4C titration was done at pH 4.0 in 8 mM sodium acetate buffer containing 50 mM sodium chloride while for the dp8 titration pH 7.0 in 20 mM potassium phosphate was chosen; both titrations were conducted at 298 K. While the basic sidechains of arginine and lysine residues, which likely contribute most to the heparin binding, are quantitatively protonated at physiological pH subtle local electrostatic changes can be introduced on the protein by going from one buffer to another, affecting the chemical shift changes. Such effects have also to be taken into consideration when comparing the two oligosaccharide titrations. The residues which experience the largest

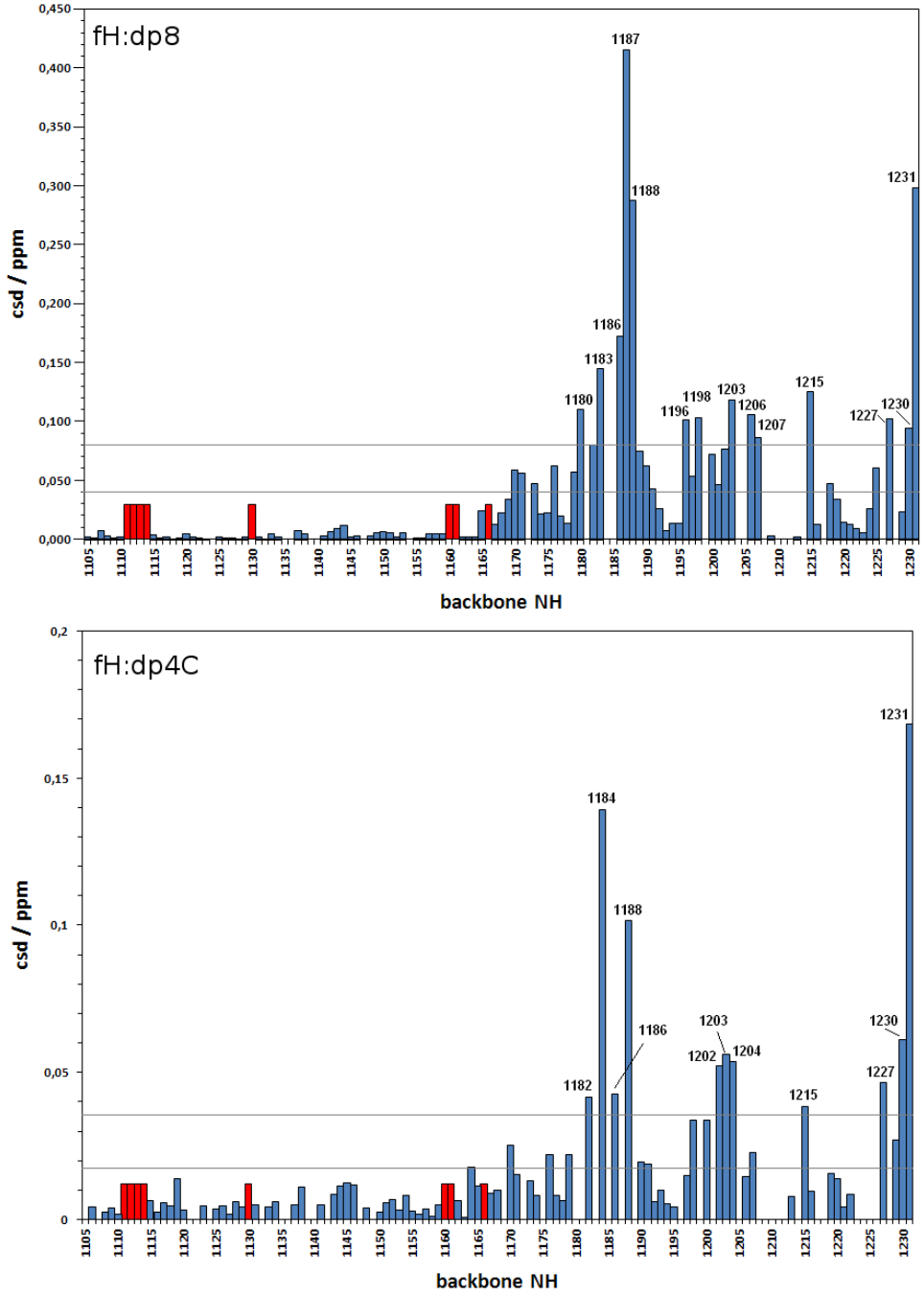

Figure 5.8: Comparison of the chemical shift changes induced on ^{15}N-fH∼19-20 by 8 equivalents of dp8 (top) and 12 equivalents of dp4C (bottom). Horizontal lines mark the average chemical shift changes and twice this value. Red bars indicate the position of proline residues for which no NH cross peaks are observed. The linker region extends from C1163 to C1167.

chemical shift changes in the titration with dp8 are mapped on the NMR structure of fH∼19-20 in Fig. 5.9 and listed in Table 5.2.

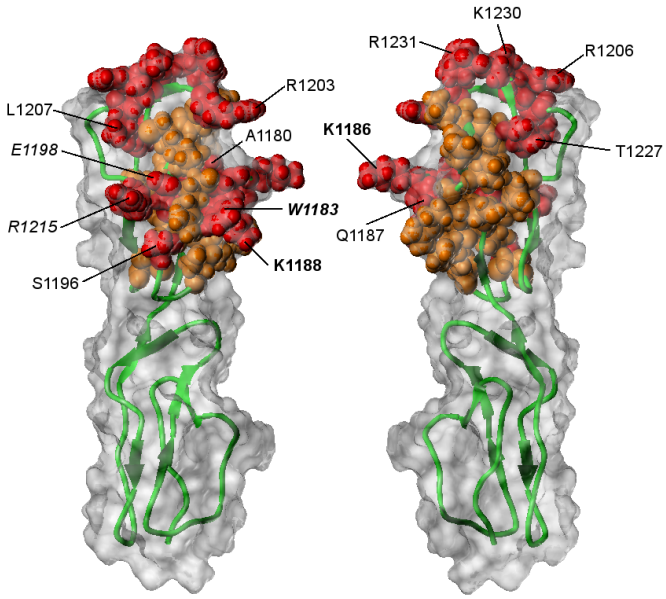

Figure 5.9: NMR structure of fH∼19-20 (PDB 2BMZ)(rotated by 180° between both views) with residues highlighted which experience chemical shift changes of 0.04 - 0.08 ppm (orange) and > 0.08 ppm (red) during the titration with dp8. Both groups are also listed in Table 5.2. Bold and italics type are used for residues whose cross peaks are affected by C3d and which have been linked to aHUS, respectively. The C-terminus is at the top of the structures, the N-terminus at the bottom.

Among the most affected cross peaks in the dp8 titration is R1231 (Fig. 5.7 bottom). A K_d value of 4.2 +/- 0.3 μM was calculated for the chemical shift changes observed for this residue. Similar values were measured for residues E1198 and K1188 (3.2 +/- 1.2 μM and 4.2 +/- 1.3 μM, respectively). The K_d of fH∼19-20 binding to dp4C, for comparison, has been determined as 9 +/- 2.5 μM in the above mentioned study [79]. These K_d values are in the same range, however, it should be noted that pH and salt concentration have a significant influence on the affinity of heparin oligosaccharides to fH∼19-20, making a direct comparison of the K_d values for dp4C and dp8 questionable.

It should also be taken into consideration that dp4C is a pure, fully sulfated species, while dp8 is only homogeneous in length and not a pure compound in the sense that its chemical composition reflects the complex primary structure of heparin. The calculated K_d is therefore an apparent K_d representing this particular mixture. It is very likely that the fully sulfated species binds more tightly.

The similarities and differences observed in both titrations are interesting when the absolute size of the oligosaccharides is compared to the size of the binding interface on fH∼19-20. dp4C is only $\sim$ 17 Å long and, if not significantly deviating from its structure in the free form, too short to cover the mapped binding surface. dp8, on the other hand ($\sim$ 34 Å in length) would be able to span the length of a full CCP module (30-35 Å, Fig. 5.10). It is likely that dp4C binds in multiple ways to the heparin binding site on fH∼19-20 since a surface area that is effectively larger than the ligand exhibits chemical shift perturbation upon dp4C binding. Another possibility is the binding of two dp4C molecules to fH∼19-20 at the same time.

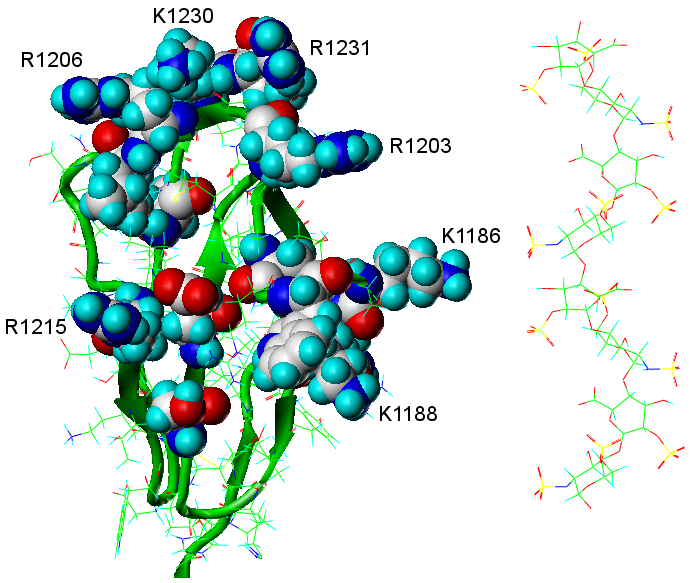

Figure 5.10: Detail of the NMR structure of fH∼19-20 (PDB 2BMZ) showing the proposed heparin binding site as delineated from the dp8 titration (same residues highlighted as red subset in Fig. 5.9). Heparin octasaccharide (PDB 1HPN) is drawn to scale for comparison of the relative sizes.

Table 5.2: fH∼19-20 residues affected during titrations with dp8. Residues which are also affected by C3d are highlighted in bold type, residues which have directly been linked to aHUS are highlighted in italics type. ppm values refer to combined chemical shift changes observed in the dp8 titration.

chemical shift changes > 0.08 ppm	chemical shift changes 0.04 -0.08 ppm
A1180	**S1170**
W1183	**R1171**
K1186	I1173
Q1187	N1176
K1188	**I1179**
S1196	*R1182*
E1198	***L1189***
R1203	**Y1190**
R1206	**S1191**
L1207	*V1197*
R1215	V1200
T1227	K1201
K1230	K1202
R1231	C1218
	Y1225

5.1.3 The ternary mixture of fH∼19-20, dp8 and C3d monitored by ^{1}H, ^{15}N-HSQC spectra

FH∼19-20 interacts *in vivo* with both cell-surface polyanions and C3b and both interactions are central for the host protection from complement. In an attempt to access the simultaneous interaction of fH∼19-20 with heparin and C3d a titration with the octasaccharide fraction of the heparin digestion (dp8) was conducted in the presence of C3d and the results were compared to titrations with dp8 alone and C3d alone. Addition of a 1.6fold excess of C3d to free ^{15}N-fH∼19-20 caused only very small chemical shift changes in the R1231 ^{15}NH-cross peak, R1231 is therefore most likely not directly involved in the binding of fH∼19-20 to C3d. This residue was therefore used as a probe for fH∼19-20 binding to dp8 in the presence of C3d (and compared to the binding in the absence of C3d). Interestingly, the chemical shift changes experienced by the cross peak of residue R1231 and other residues of the heparin binding site were smaller in the presence of C3d even at a large excess of dp8 (Fig. 5.11, 5.13). For the R1231 backbone NH cross peak, a plot of chemical shift changes as a function of ligand concentration highlights the different binding characteristics (Fig. 5.12). Fitting the chemical shift changes for the R1231 cross peak in absence and presence of C3d led to K_d-values of 4.2 μM +/- 0.3 μM and 44 μM +/- 0.3 μM, respectively. Interestingly, addition of 50 mM potassium chloride to the last titration point of the ternary mixture (containing fH∼19-20:C3d:dp8 = 1:1.6:15.0) reversed the chemical shift changes significantly (Fig. 5.11). Presumably, this observation is an indicator of the prevalently ionic contributions to this protein-heparin interaction.

The line broadening of those residues that where involved in the interaction with C3d was not affected by dp8, notably for the cross peaks of residues W1183, K1188 and L1189 in Fig. 5.14. This observation together with the conserved chemical shift trend of the heparin binding residues suggests that both interactions with fH∼19-20 were preserved in the ternary mixture. Addition of 50 mM potassium chloride did not have a visible effect on the peak broadening in fH∼19-20, suggesting that binding of C3d is not

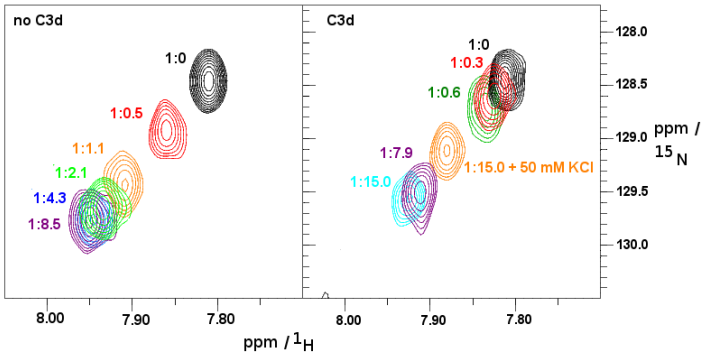

Figure 5.11: Comparison of chemical shift changes of residue R1231 of fH∼19-20 induced by dp8 in absence (left) and presence (right) of C3d. The fH∼19-20:dp8 ratios are listed for each titration point. Addition of 50 mM potassium chloride to the last titration point in the ternary mixture reversed the chemical shift changes to some degree (orange cross peak in right spectrum).

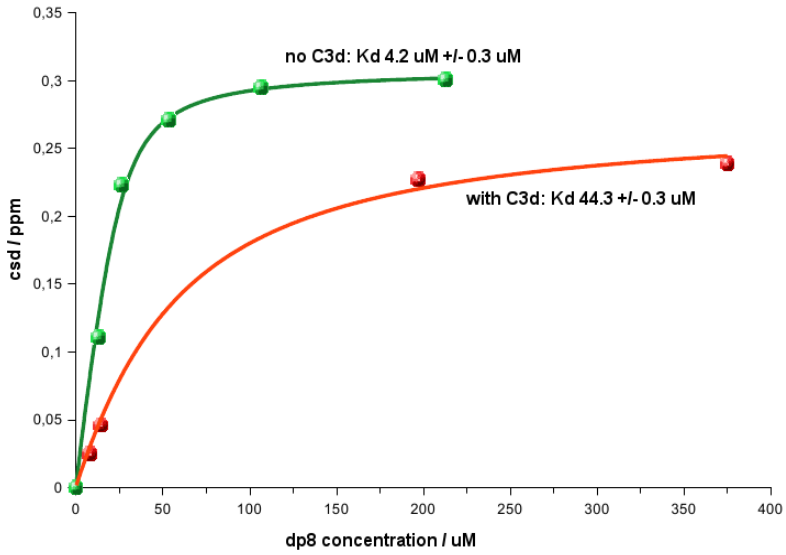

Figure 5.12: Combined chemical shift differences (csd) of the R1231 NH backbone cross peak induced by dp8 in the absence (green) and presence (red) of C3d. Experimental values: spheres, fitted curves: solid lines.

susceptible to small changes in ionic strength, while dp8 binding is.

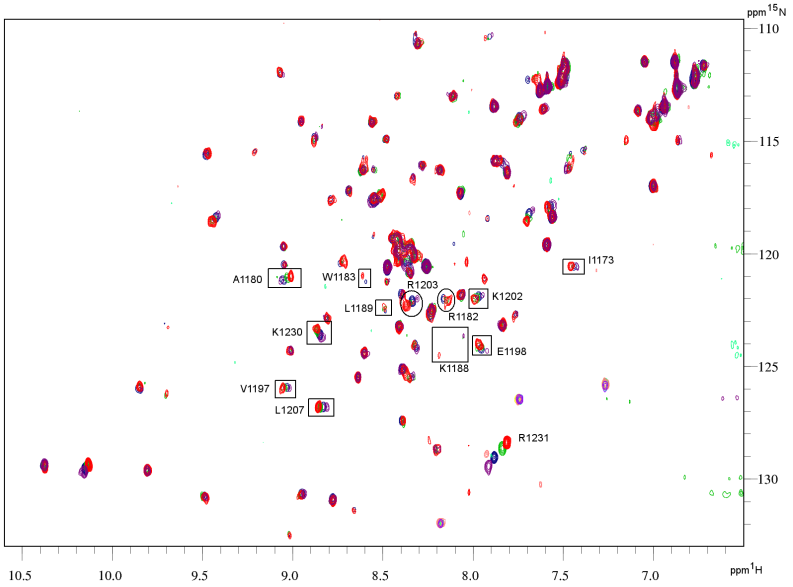

Figure 5.13: ^{1}H, ^{15}N-HSQC monitored titration of fH∼19-20 with dp8 in the presence of 1.6fold excess of C3d. fH∼19-20:dp8 ratios are 1:0 (black), 1:0.3 (red), 1:0.6 (green), 1:7.9 (purple) and 1:15 (cyan).

C3d was found to bind very weakly to a heparin affinity column at the pH used for the titration experiment (pH 7.0). The interpretation of the ternary interactions is complicated by this fact. The results do, however, exclude cooperativity between both proteins with respect to dp8 binding. Cooperative binding of C3d and dp8 to fH∼19-20 would, presumably, result in an apparent decrease in the K_d for the fH∼19-20:dp8 complex and saturation would consequently be reached at lower dp8 concentrations. Since the opposite is observed two explanations are possible: competition between fH∼19-20 and C3d with respect to heparin-dp8 (resulting in an apparent lower concentration of dp8 available for fH∼19-20 since some is bound to C3d) or reduced affinity of fH∼19-20 for dp8 in the presence of C3d. The latter scenario could be a consequence of partial overlap of the binding sites as previously suggested on the basis of mutational studies and modelling [258].

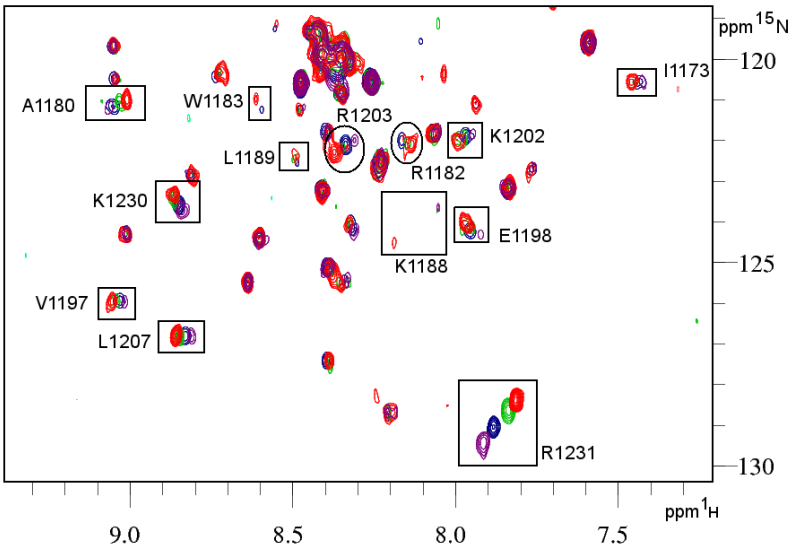

Figure 5.14: Detail of Fig. 5.13. The same detail is shown in the absence of C3d in Fig. 5.7.

5.2 Discussion of the NMR binding studies of fH∼19-20

In this study, C3d was titrated into ^{15}N-fH∼19-20 and ^{1}H, ^{15}N-HSQC spectra were recorded to delineate the binding site using intact wild type fH∼19-20. Decreased peak intensity and excessive line broadening was obsereved for several residues. Peaks which underwent excessive line broadening and disappeared or almost disappeared are listed in Table 5.1 (first column), together with residues whose cross peaks were less dramatically affected but still significantly broadened (second column). A number of residues in the linker region fall into this second category, and their broadening might arise from conformational change rather than direct binding of C3d. For example, a change in the relative orientation of CCP 19 and CCP 20 could cause such effects. It was observed previously that ligand binding can alter the relative orientation of neighbouring CCP modules [253]. Addition of heparin-derived octasaccharide (dp8) did not recover the intensity losses caused by C3d. The implication of this observation is that dp8 did not

outcompete C3d with respect to fH∼19-20 binding. It was, however, observed that the chemical shift changes which dp8 caused in free ^{15}N-fH∼19-20 were smaller in the presence of C3d (Fig. 5.12, 5.11). One residue, the C-terminal R1231, was chosen to monitor this effect (Fig. 5.11). This residue was chosen because it is one of the most prominent heparin-binders in CCP 20 and it was not affected by C3d binding. Mapping of the respective binding sites on the NMR structure of fH∼19-20 shows that this residue is not part of the C3d binding site (Fig. 5.4).

In separate titration experiments, nine residues in fH∼19-20 were affected by both C3d and dp8 (bold residues in Table 5.2). This observation and the altered affinity of fH∼19-20 for dp8 in the presence of C3d could be explained with a mechanism that has been proposed by [257] in which cell surface polyanions and C3b first form a transient encounter complex before the stronger C3b-fH interaction is accomplished. The role of the polyanions would be that of a mediator which brings factor H in proximity to cell bound C3b. Such a role is not uncommon for HS which acts as a scaffold for two proteins in a range of protein-protein interactions like the thrombin:antithrombin or the FGF:FGF-receptor complex.

A large and still growing number of polymorphisms and mutations in fH∼19-20 has been linked to the rare yet serious kidney disease aHUS (Fig. 5.15). In this study, several of the residues linked to aHUS were found to be directly affected by binding of C3d (D1119, W1183 and L1189) and/or dp8 (R1182, W1183, L1189, V1197, E1198, R1215). Many of the aHUS mutations of fH∼19-20 cluster around the heparin binding site in CCP 20, but many other aHUS mutations are found in parts of the protein that are relatively remote from this binding site, spanning almost the full length of both CCP modules (Fig. 5.15). A direct link between the aHUS mutations and heparin binding alone could not be established. Similarly, SPR experiments using the same panel of mutants and C3b failed to establish a correlation between the aHUS mutants and the affinity to C3b that would have included all aHUS-linked mutations. We have therefore concluded that

experiments which investigate all three components at the same time are required. As the first step towards this goal we have monitored the interaction between wild type fH∼19,20 and dp8 in the presence of C3d. These experiments yielded promising results and will be extended to the investigation of aHUS mutant proteins in the future.

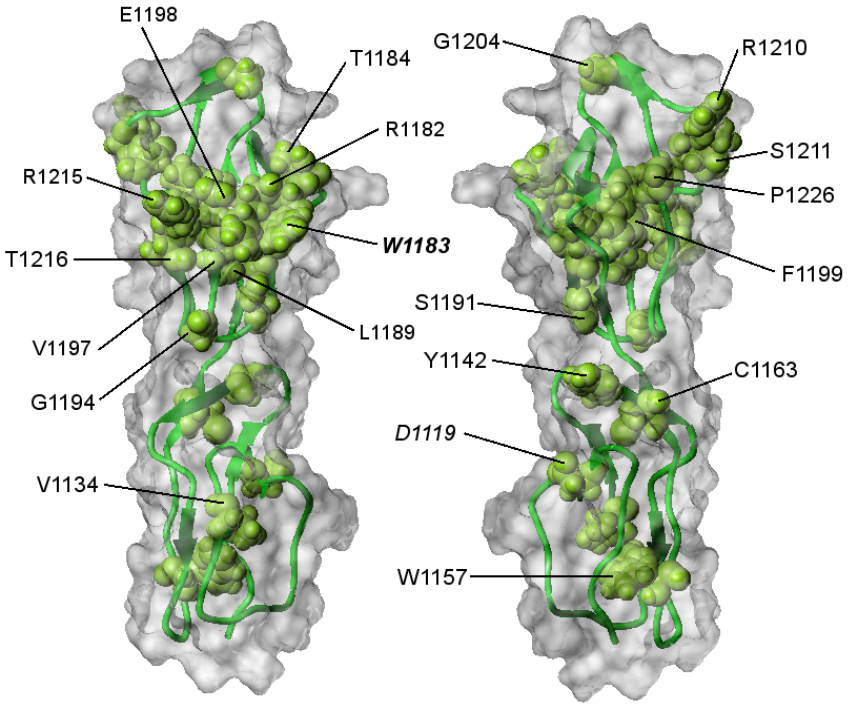

Figure 5.15: NMR structure of fH∼19-20 (PDB 2BMZ) with residues highlighted which have been linked to aHUS (source: aHUS mutational database, www.fh-hus.org). Bold and italics type are used to mark residues which are affected by dp8 and C3d binding or either of the two, respectively.

5.3 GMSA of fH∼19-20 aHUS mutants and dp4C/dp6

The interaction between fH∼19-20, C3b/C3d and cell-surface polyanions is of interest for kidney research because many mutations in CCP 19 and CCP 20 have been identified as risk-factors for atypical haemolytic uraemic syndrome (aHUS). Most biochemical research presented so far made use of mutations and subsequent binding experiments as well as models derived by molecular docking [258, 91, 79]. In particular, we [260] and others [257] used a large set of aHUS-related and designed mutants to map the binding sites of both heparin/cell surfaces and C3b/C3d onto fH. Both studies used heparin as a cell-surface polyanion mimic and [257] compared the results obtained from heparin binding assays with binding affinities to endothelial cells from mice. The overall ranking of the tested mutant proteins with respect to heparin on one hand and endothelial cells on the other hand were very similar, supporting the use of heparin as a cell-surface polyanion mimic.

A gel mobility shift assay (GMSA) was used to assess the relative binding affinities of fH∼19-20 WT and 13 aHUS mutants for heparin derived oligosaccharides. The mutant proteins were prepared by Dr. Andrew Herbert and Dr. David Kavanagh. In this assay, free heparin oligosaccharides migrates in an electric field through an agarose gel because of its large negative charge. The formation of a neutral/less charged complex in the presence of a heparin binding protein hinders this migration and retains the oligosaccharide in the well. Fluorescent labelling of the ligand allows localisation of it after short electrophoresis, leading to an estimate of how much ligand remained free (i.e. migrated in the electric field) and how much was retained in the well. For these experiments pure fully sulfated tetrasaccharide dp4C and dp6 fractions from heparinase I digestion of bovine intestinal heparin were used.

GMSA of fH∼19-20 with dp4C and dp6 yielded very similar results for the WT and mutant proteins. The GAG:protein ratio were chosen to give an excess of free GAG

(evident in the WT lanes as a faint band migrating to the anode). In this manner, both stronger binding (i.e. less free GAG migrating from the well) and weaker binding (i.e. more GAG migrating to the anode) could be assessed. In general, the addition of a positively charged residue (K or R) increased affinity to GAGs, while replacement of a positively charged residue by an uncharged residue weakened the binding (Fig. 5.16). This trend was observed for both designed non aHUS and aHUS mutants of fH∼19-20. From a purely biophysical point of view, this trend is consistent with the observation that heparin binding is mostly dominated by charge-charge interactions. However, from a medical point of view a coherent trend was not found for the aHUS mutants: R1210C (which was not misfolded despite the free cystein, as shown by ^{1}H, ^{15}N-HSQC spectra) and R1182S bound weaker than WT to GAGs, numerous mutants exhibited binding affinities similar to the WT, and L1189R, T1184R, and W1183R all bound stronger than WT. In other words, the aHUS linked mutations display either increased, normal or decreased heparin binding affinities and heparin binding alone cannot be used as an indicator for the disease risk associated with aHUS mutations. With the exception of R1210C, R1215G and R1215Q all mutants were tested for a homogeneously sulfated GAG, dp4C, as well as for a mixture with different sulfation levels, dp6. No striking difference between the relative affinities for the two GAGs was observed. In conclusion, the interaction of fH∼19-20 seems to be dominated by charge-charge interactions, and heparin binding alone does not correlate with the disease as both weaker and stronger binding was observed for the aHUS mutant proteins.

While some of the results described in the literature for aHUS-linked mutant proteins are in agreement, others contradict each other [260, 257]. A number of possible reasons can be found for these contradictions. Firstly, mutations which lead to an increase in binding affinity (as was observed for several mutations with respect to both C3b/C3d and heparin) can reflect an artificial enlargement of the natural binding site. In particular, this seems to be the case for heparin-binding, which is often mediated through relatively unspecific ionic interactions. Changing an aliphatic residue nearby the heparin binding site into a basic residue, for example, might lead to an increase in affinity even if

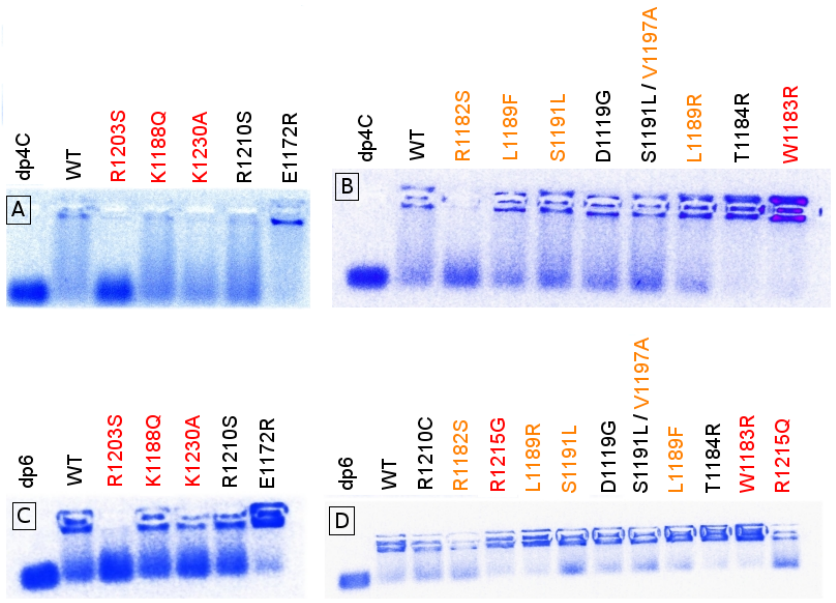

Figure 5.16: GMSA with dp4C (A and B) or dp6 (C and D) and fH∼19-20 WT and designed (A and C) and aHUS (B and D) mutants. Cathode and well: top of gels, anode:bottom of gels. Residues which are were found to be involved in dp8 binding in the ^{1}H, ^{15}N-HISQC titration are coloured red and orange for combined chemical shift changes $>$ 0.08 ppm and 0.04-0.08 ppm, respectively (same colour coding as in Fig. 5.9).

the mutated aliphatic residue was not previously part of the binding site. Secondly, the choice of the introduced amino acid is as important as the position at which it is introduced as different mutations of the same amino acid can lead to differently modulated affinities. Mutation of W1157 to a leucine residue, for example, was reported to impair C3b binding [257] while the choice of an arginine residue resulted in unaltered binding affinity [260]. Thirdly, a single amino acid change can lead to severe structural changes, affect the stability and the protein's tendency to self-associate, parameters which are often not checked carefully enough (in the case of [260], the structural integrity of all mutant proteins was checked by NMR). Additionally, binding contributions from main chain atoms are overlooked completely in cases were mutations do not lead to significant structural changes. The C3d-CR2 complex, for example, has been reported to be mediated by several main-chain carbonyl groups which were not picked up by earlier mutational studies [261]. Thus, mutant proteins only represent an indirect tool for a physiological map of the C3b/C3d and HS binding sites on fH and need validation by complementary techniques. NMR spectroscopy and x-ray crystallography are able to characterise the interactions of wild type proteins on an atomic level, helping to rationalize the results observed in functional assays.

5.4 aHUS mutants in a functional haemolytic assay

The ability of recombinantly expressed fH∼19-20 to impede the protective effect of full length fH in serum was tested using a complement assay for the alternative pathway [95, 69]. Because the human complement system does not recognize sheep erythrocytes (RBCs) as foreign such cells are protected from complement mediated lysis when exposed to human serum. Recombinant WT fH∼19-20 competes with full length fH for cell binding but is unable to stop the complement cascade because the truncated protein lacks the regulatory N-terminal domains. Thus, the addition of WT fH∼19-20 to a serum sample containing sheep RBCs causes dose-dependant cell lysis which can be

measured via the haemoglobin absorbance. Interestingly, aHUS associated mutants of fH∼19-20 cause different degrees of lysis. However, all mutant proteins tested - with the exception of R1210C - compete less effectively with full length fH than the WT fH∼19-20 (Fig. 5.17).

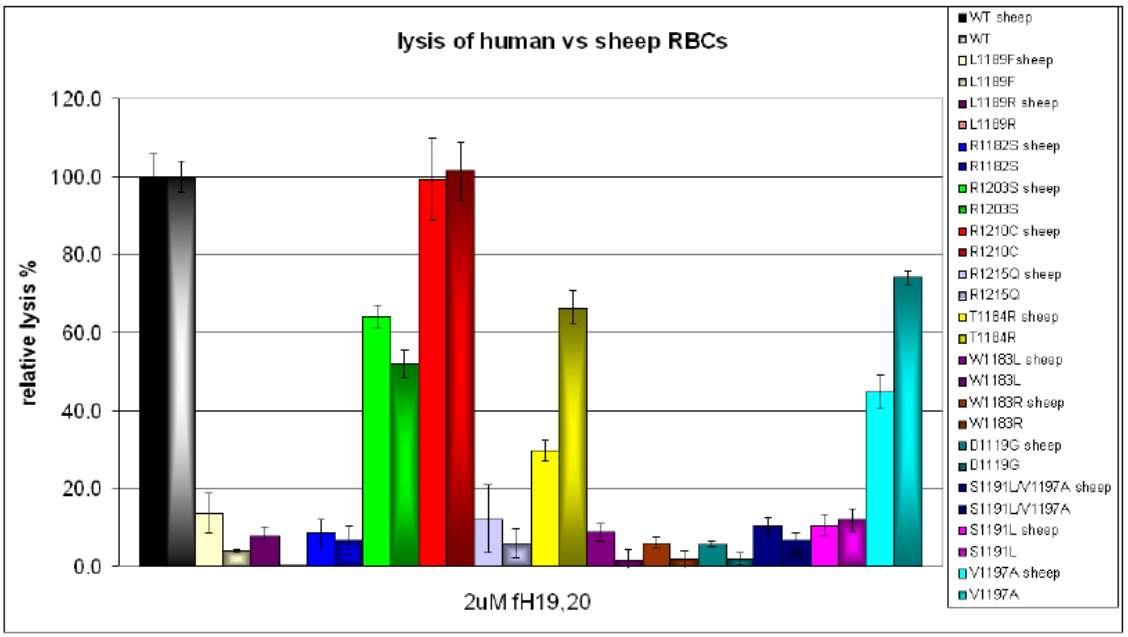

Figure 5.17: Haemolytic assay of different mutant proteins and WT fH∼19-20 with sheep and human erythrocytes. Solid bars: sheep RBCs, shaded bars: human RBCs.

An especially dramatic example is the D1119G mutant protein (Fig. 5.18). For comparison, the assay was simultaneously conduced with human RBCs (Fig. 5.17). Because human cells have additional, membrane-bound complement regulators of the alternative pathway the concentrations of recombinant protein that were needed to cause a certain amount of lysis are larger for human RBSc than for sheep cells. In order to compare the relative effects of the different mutant proteins the data were normalised. The amount of lysis caused by 2 µM WT fH∼19-20 was set to 100% and the lysis caused by the same concentration of mutant proteins were calculated with respect to this value.

The results of this functional assay show that all but one (R1210C) of the tested aHUS-linked variants of fH∼19-20 are less efficient in protecting cell surfaces from the al-

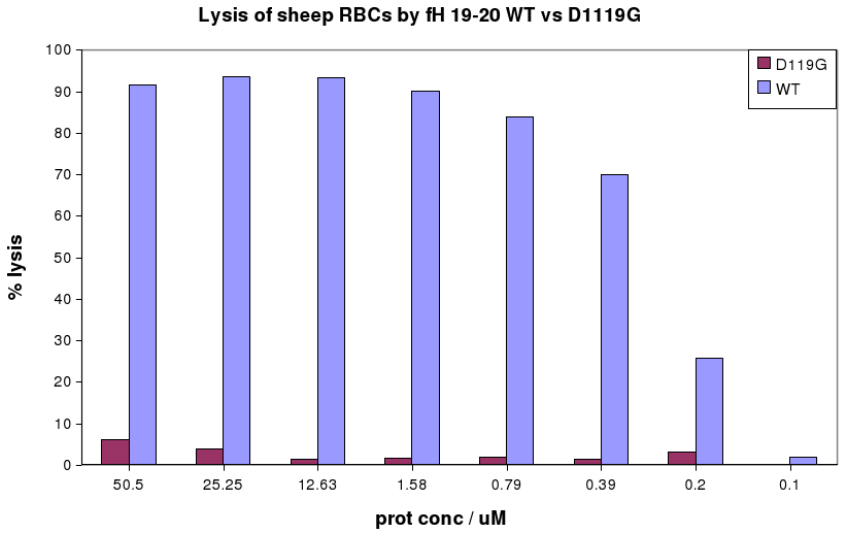

Figure 5.18: Haemolytic assay of D1119G and WT fH∼19-20 with sheep erythrocytes. Increasing concentrations of WT protein lead to complete lysis of the cells in human serum while the mutant protein barely causes any lysis of the sheep cells.

ternative pathway of the complement system (the Ca^{2+}-dependent classical and lectin pathways were silenced by addition of Mg-EGTA in this assay). Our collaborators have studied the interaction of the same panel of mutants with C3b and neither observed a clear trend: some aHUS variants bound to C3b with smaller, some with higher affinity [260]. Since no coherent trend was found in binding studies with aHUS variants and heparin or C3b alone it can be assumed that the reason for the risk associated with these variants is to be found in the ternary interaction rather than pairwise single interactions. One possible explanation is a model which has been proposed by [257] in which cell-bound HS first forms a transient encounter complex with fH∼19-20. In the proximity of cell-bound C3b this complex is "swapped" for an interaction with the C3b-TED domain on fH∼19-20 and this protein-protein complex is additionally stabilized by binding of fH∼1-4 to another face of C3b [253]. In such a scenario both weaker and stronger binding to HS as well as to C3b would be able to disrupt a fine equilibrium, giving rise to reduced protection of cell-surfaces.

6 GMSA study on the middle region of factor H

A third heparin binding site in fH has been reported previously to reside in CCP 9 [262]. In collaboration with Dr. Christoph Schmidt a set of recombinant proteins encompassing fH modules 8-15 were tested for heparin binding via GMSA. Fluorescently labelled size-fractionised heparin decasaccharide (dp10) was used in this study. None of the wide range of constructs (fH∼8-9, fH∼10-12, fH∼12-13, fH∼13-15, fH∼11-14, fH∼10-15 and fH∼8-15) bound heparin in this assay in which fH∼6-8 (H402) was used as a positive control (Fig. 6.1). Additionally, fH∼8-9 with a cloning artefact containing basic residues (EFTWPS*R*PS*R*IGT) was tested for heparin binding because the reported construct used in [262] contained such an aritfact. The artefact carrying fH∼8-9 construct was found to be retained by dp10, albeit not strongly. It was thus concluded that this cloning artefact was at least partially responsible for the previously observed heparin affinity in this construct. Similarly, no heparin binding was observed that could be attributed to CCP 13. This module has also been proposed to carry a heparin binding site, mainly based on indirect evidence [67]. A potential third heparin binding site is thus unlikely to exist in fH [68].

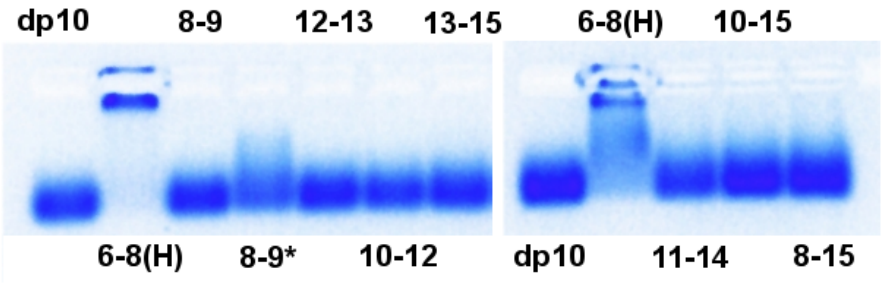

Figure 6.1: GMSA using several constructs form the middle region of fH. The numbering refers to fH modules, i.e. 6-8(H) stands for fH∼6-8 (H402). The construct 8-9* refers to fH∼8-9 carrying the N-terminal artefact EFTWPSRPSRIGT. 1 µ fluorescently labelled oligosaccharide and equimolar amounts of protein were used for each lane.

7 Heparin binding study on NK1

HGF/SF is a multifunctional growth factor which primarily controls epithelial/endothelial cell behaviour. Its functions are transduced through the single specific tyrosine kinase receptor Met. Heparin oligosaccharides act as cofactors in this ligand-receptor interaction. The primary heparin binding site of HGF/SF resides in the two most N-terminal domains, the N and the K1 domain. The construct of these two domains, which is also a naturally occurring splice variant of HGF/SF, is referred to as NK1. The physiological function of this splice variant is, to present, not fully understood. In addition to high affinity heparin binding it has been demonstrated that NK1 also binds to dermatan sulfate (DS), albeit with lower affinity [135]. The objective of this investigation was to determine whether NK1 has distinct or overlapping binding sites for the two GAGs. ^{15}N NK1 was titrated with heparin dp4C and DS hexasaccharide (DS-dp6) to assess to which degree both binding sites overlap and to compare the affinities for both oligosaccharides. These two oligosaccharides had previously been identified as the smallest binding ligands of either type of GAG [233].

NK1 was expressed and purified as described in section 2.1. The buffer used for NMR studies was 20 mM sodium acatate, 100 mM sodium chloride, pH 6.0 and spectra were recorded at 303 K on a 800 MHz Bruker Avance NMR spectrometer. The oligosaccharides were produced by Jon Deakin as described in section 2.2. For DS oligosaccharides, chondroitinase AC-1 and chondroitinase ABC were used instead of heparinase [172].

Since no assignment of NH signals was available for NK1, a histogram was produced in

which peaks were ordered according to the combined chemical shift perturbation that heparin dp4C binding caused. These "book keeping" assignments were transferred to the DS-dp6 titration in order to compare relative effects. In some regions of the spectra following the chemical shift changes of individual residues during the titrations was hindered due to overlap and peak broadening. A comparison of peaks which could be followed unambiguously during both titrations showed that both ligands affect a set of key residues to a similar degree (Fig. 7.1).

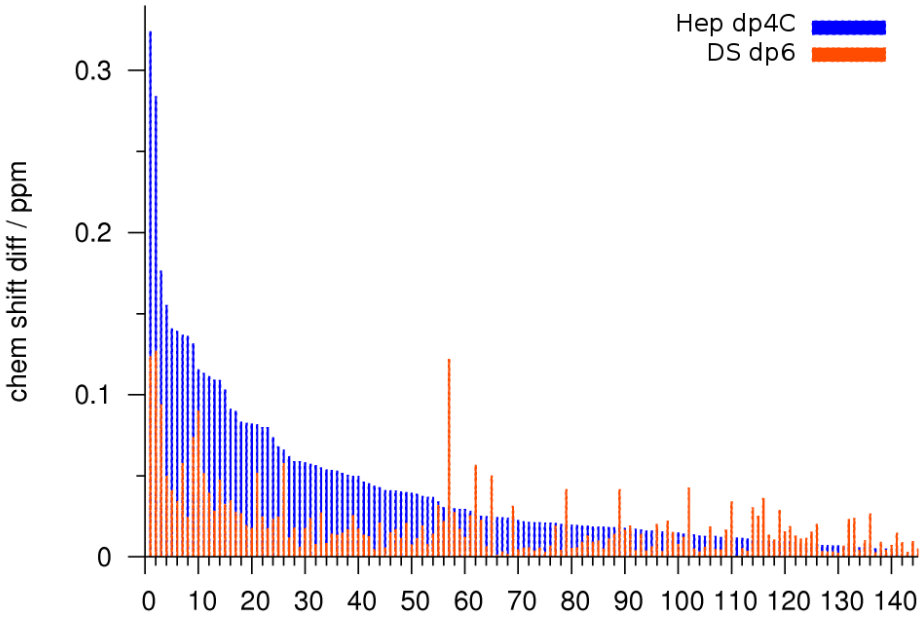

Figure 7.1: Histogram showing the relative effects of heparin dp4C (blue) and DS hexasaccharide (orange) on NK1 backbone ^{15}NH chemical shifts. The horizontal axis numbering does not refer to residue numbers because no assignment was available for NK1. Instead, the residues are ordered according to the combined chemical shift changes which heparin dp4C induced in the amide cross peaks (their numbering is thus only for "book keeping"). Ratios were 1:1 protein:oligosaccharide in both titrations.

The binding sites were found to be overlapping yet distinct. While some of the most affected residues in the dp4C titration were also significantly affected in the DS dp6

titration, some of the most affected residues in the DS dp6 titration were affected to a smaller extent by heparin dp4C (red exclamation marks in Fig. 7.2). Additionally, differences in line broadening were observed: some cross peaks disappeared in the DS-dp6 but not in the heparin dp4C titration (green asterisks in Fig. 7.2). Finally, differences could also be observed when comparing the directions in which cross peaks of residues in the overlapping binding site moved. Some cross peaks were affected in both titrations but exhibited different chemical shift trends (red asterisks in Fig. 7.2), pointing to different chemical environments. Addition of dp4C to NK1 above a ratio of 1:1 did not cause further chemical shift changes, implying that at this ratio saturation had been reached. In contrast, a 2.5fold excess of DS-dp6 had to be added to NK1 in order to reach saturation with this GAG, indicating lower affinity. Subsequently, a titration with DS tetrasaccharide (DS-dp4) was conducted, yielding very small chemical shift changes at a large (> 10-fold) excess of GAG (data not shown). Thus, DS oligosaccharides bind weaker to NK1 than heparin oligosaccharides of similar size, and the binding affinity for short DS oligosaccharides increases with their length, at least for such small oligosaccharides as tetra- and hexasaccharides.

Overall, the following observations were made: (i) for the majority of amide cross peaks the absolute changes were larger for heparin dp4 at equal concentrations, (ii) a similar trend was observed between the two titrations (i. e. the same set of key residues was affected), and (iii) a fraction of the movers observed in the DS-dp6 titration exceeded the chemical shift changes caused by heparin dp4C. This last point can be ascribed to the different nature and size of the two oligosaccharides. The data support the assumption that both ligands bind to overlapping sites of NK1, with the larger binding site for DS-dp6. Several NK1 residues of the heparin dp4C binding site are less important for DS-dp6 binding as judged by smaller movements of their NH cross peaks in HSQC spectra. This is likely caused by the different chemical nature and charge distribution of the two ligands. The relative binding affinities of NK1 for the three studied oligosaccharides are: heparin dp4C > DS-dp6 > > DS-dp4.

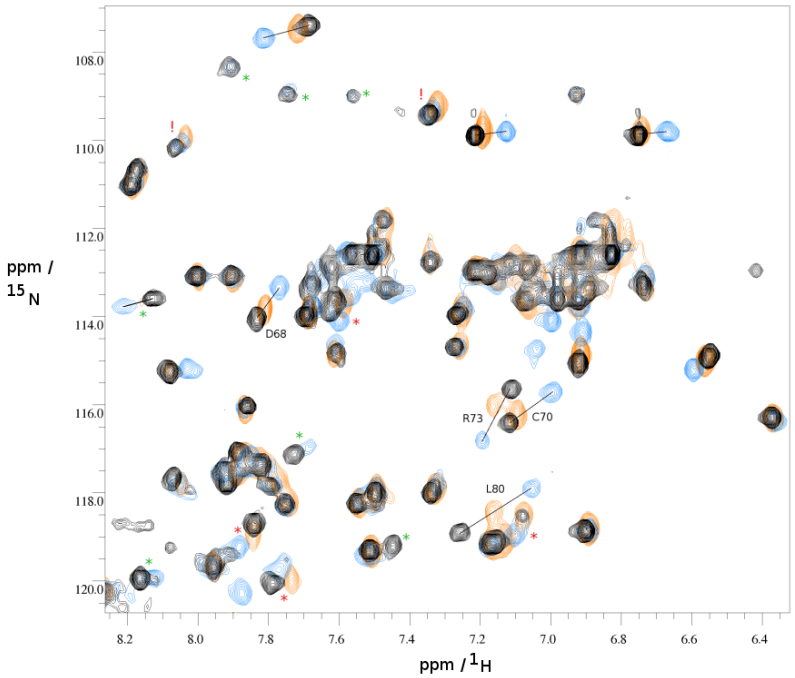

Figure 7.2: ^{1}H, ^{15}N-HSQC monitored titration of NK1 (free protein: black) with heparin dp4C (blue, 1:1) and DS hexasaccharide (orange, 1:1). Red exclamation marks highlight residues which are more affected by the weaker binding DS hexasaccharide, green asterisks highlight residues which experience intermediate broadening in the DS hexasaccharide titration and red asterisks highlight residues which are affected in both titrations yet cause different chemical shift perturbations.

A provisional assignment of some of the key residues of NK1 that are involved in DS-dp6 and heparin dp4C binding can be made based on a comparison with NMR data of NK1 published elsewhere [263]. Some of the most affected residues are C70, R73, L80 and, to a lesser extent, D68. R73 has been described as central in a x-ray structure of NK1 in complex with a heparin 14mer [126]. All of the above residues were affected during titrations of the N-domain with fractionated heparin (10-15mers) and SOS [263]. Only a tetrasaccharide moiety within the heparin 14mer establishes well-defined contact with NK1 residues in the crystal structure of the complex, suggesting that the size of the heparin binding site is in the range of the GAGs which were used in this work.

The observation that both heparin dp4C and DS-dp6 caused different chemical shift changes in ^{1}H, ^{15}N-HSQC spectra of NK1 is interesting when compared to the titrations of fH∼7-8 with heparin oligosaccharides and sucrose octasulfate (SOS). No significant differences between these two types of ligand were observed when titrated into fH∼7-8 despite the large chemical differences between SOS and heparin. SOS is very short, contains a furanose ring, lacks L-iduronic acid and D-glucuronic acid and has no nitrogen as well as no carboxy groups. In other words, it differs from the class of glycosaminoglycans in many ways. The fact that it nevertheless appears to bind in a very similar position and manner to fH∼7-8 when compared to dp4C while differences are observed between the two GAG-type ligands binding to NK1 points to a possible difference between the NK1 and fH∼7-8 GAG-binding sites. A comparison between complexes of SOS and fractionated heparin with the N-domain (which contributes most to GAG binding in NK1 and HGF/SF) is available in the literature [263]. While in NMR titrations both ligands affect overlapping sets of residues, there are notable differences in the chemical shift changes that the same residue experiences in titrations with different ligands. It can be concluded that heparin and DS oligosaccharides as well as SOS all bind to the same face of NK1 but can still be unique on an atomic level. During NMR-monitored titrations chemical shift changes are only used to identify and rank protein residues while

the shifts are usually not interpreted in terms of direction, i.e. high- or low-field shift in the proton or nitrogen dimension. If x-ray structures of NK1 in complex with DS and SOS were available this system seemed well suited to attempt a more detailed analysis of what chemical shift changes are able to report.

8 Relaxation time measurements of HBD2 and GAGs

The following investigation became part of a study on the human beta defensin 2 (HBD2) and two types of GAGs which was conducted by Dr. Emily Seo in our laboratory. The peptide was expressed and ^{15}N-labelled recombinantly by her [264]. Both DS-dp6 and Fondaparinux, the antithrombin binding heparin pentasaccharide (Fondaparinux is a commercial name), were found to bind to HBD2 but exhibited different affinities as judged by ^{1}H, ^{15}N-HSQC monitored titrations. Substantial differences were observed during these two titrations. The same set of residues was effected but chemical shift changes could be followed more easily in the complex with DS-dp6 which was characterised by the fast exchange regime. The titration with Fondaparinux, on the other hand, suffered from excessive line broadening. The conditions improved when a lower temperature (283 K instead of initially 298 K) and lower field strength (600 MHz instead of initially 800 MHz) were used. Nevertheless, some peaks still remained broadened in the latter titration. Additionally, some chemical shift changes followed a linear behaviour during the first titration points but seemed to report on a second process after a 0.5:1 ratio of Fondaparinux:HBD2 was reached. In order to clarify the question if protein oligomerisation could be caused by addition of the GAGs measurements of ^{15}N T_1 and T_2 relaxation times for free HBD2 and HBD2 in the presence of DS-dp6 and Fondaparinux were conducted at 283 K and 600 MHz, giving the following results:

sample	T_1 / ms	T_2 / ms	τ_c / ns
free HBD2	428 +/- 15	163 +/- 6	4.0 +/- 0.2
HBD2 + DS dp6 (1:4)	466 +/- 20	103 +/- 4	5.0 +/- 0.2
HBD2 + Fondaparinux (1:16)	560 +/- 21	131 +/- 5	6.8 +/- 0.2

Deviations in this table reflect uncertainties in the exponential fit. The rotational correlation times, τ_c, were calculated using the program R2R1 (Arthur G. Palmer III, Columbia University). Due to the relative small size of HBD2 (41 residues) 1D spectra were sufficient to estimate relaxation times from non-overlapping backbone amide peaks. Decays were fitted to a mono-exponential decay using gnuplot. During the titrations the overlap in the proton dimension worsened and some peaks which had been used for the relaxation times of the free HBD2 were no longer used. In total, the number of backbone amides used for each sample were 12 (free HBD2), 10 (DS-dp6) and 9 (Fondaparinux). Standard deviations were calculated to evaluate how big the differences in the relaxation times between the different peaks were:

sample	σ_{T_1} / ms	σ_{T2} / ms
free HBD2	10	18
HBD2 + DS dp6 (1:4)	27	8
HBD2 + Fondaparinux (1:16)	32	11

These values suggest that the spread of T_1 and T_2 is relatively small in all samples despite complex formation and chemical exchange. HBD2 was reported to have a roughly globular shape and to be monomeric in solution [117, 265, 266], which fits well with the narrow distribution of relaxation times - no large flexible features are present in HBD2. Presumably, such flexible stretches would cause more dramatic deviations in the individual relaxation times.

Both complexes used in this study were assumed to be close to saturation (despite the largely varying ratios) as determined by ^{1}H, ^{15}N-HSQC titrations. While the τ_c of HBD2 in the presence of DS-dp6 might be caused by the increase in mass through presence of the GAG in the complex, the more dramatic increase in τ_c for the sample containing

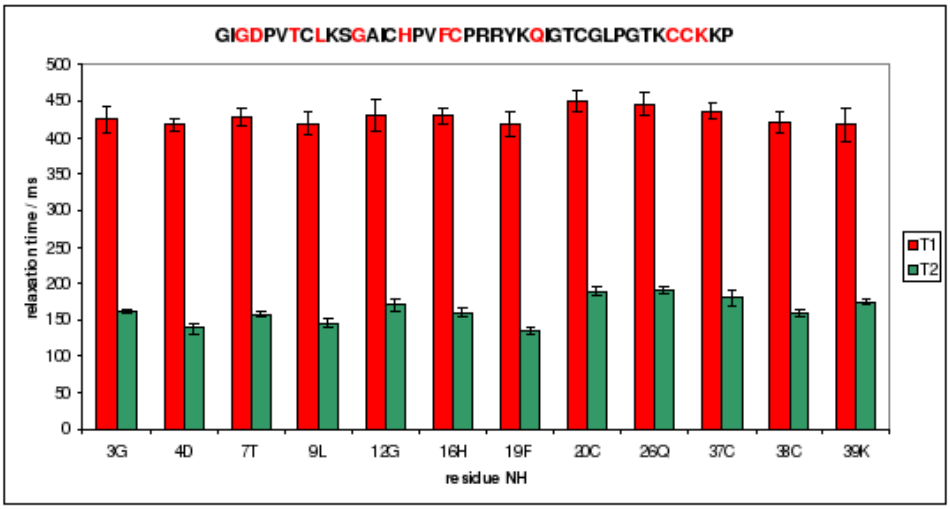

Figure 8.1: Individual relaxation times for backbone amide groups in unbound HBD2. Red residues in the primary sequence (top) were used, black residues were omitted due to overlap.

Fondaparinux very likely indicates an increased tendency to self-associate. The τ_c for free Fondaparinux has been determined elsewhere to be 0.55 ns [267]. It is likely that both GAGs have similar τ_c values due to their similar length and shapes [165, 268]. Thus, the difference in τ_c values of the complexes must point to different interaction modes. The assumption that Fondaparinux causes self-association of HBD2 is in agreement with DOSY measurements which revealed a slightly decreased diffusion coefficient in the presence of DS hexasaccharide and a larger decrease in the diffusion coefficient in the presence of Fondaparinux ($1.24*10^{-10}$ m^2/sec for free HBD2, $1.18*10^{-10}$ m^2/sec for the complex with DS-dp6, and $1.08*10^{-10}$ m^2/sec for the complex with Fondaparinux). Fondaparinux is a highly sulfated pentasaccharide with eight sulfate groups while DS-dp6 contains only three sulfate groups. The binding site on HBD2 that is identical for both GAGs is approximately as long as these oligosaccharides. Fondaparinux therefore presents sulfates on both sites of its heparin helix, opening the possibility of creating a trans dimer with two molecules of HBD2. Such an arrangement has been observed in the x-ray structure of FGF2 with dp4C [184].

9 Conclusions

Glycosaminoglycan (GAG)-protein interactions are involved in a large variety of biological processes ranging from the regulation of blood coagulation to cellular communication and pathogenic invasion of cells. The main subject of this work were GAG-protein complexes that are thought be central to the host cell recognition. The protein studied was factor H (fH), a crucial fluid phase regulator of the alternative pathway of the complement system, which is part of the innate immune defense. Recombinantly produced proteins and heparin derived oligosaccharides were used to structurally characterise the interaction of fH with heparin/heparan sulfate (HS) using NMR spectroscopy, mass spectrometry and other biophysical techniques. Complement assays served the purpose of putting these structural studies in the physiological context.

Two heparin/HS binding sites exist in factor H. One is located in the 20^{th}, C-terminal module of factor H, CCP 20, the other centers around CCP 7. Two more potential binding sites proposed previously in CCP 9 and CCP 13 were discarded during the course of this work [68]. Cloning artifacts were found to be responsible for the previously reported heparin binding in CCP 9 and this work contributed to this finding by means of gel mobility shift assay (GMSA). The heparin binding of CCP 20 was studied by GMSA and NMR. Dose-dependent chemical shift changes induced by the octasaccharide fraction of heparin, dp8, were observed by NMR using a ^{15}N labelled recombinant construct of the two C-terminal modules of fH, fH$\sim$19-20. The residues of module 20 identified previously as heparin binding site by using fully sulfated heparin tetrasaccharide, dp4C, were also found to be interacting with dp8. However, the chemical shift changes of backbone

NH protons exerted by both oligosaccharides varied dramatically. A comparison of the relative sizes of dp4C, dp8 and the heparin binding site in CCP 20 revealed that dp4C, unlike dp8, is not able to span the complete binding site. Multiple binding modes for shorter heparin oligosaccharides like dp4C might exist, a scenario that is not uncommon in protein-GAG interactions.

A number of disease related variants of CCP 19 and CCP 20 exist in the context of full-length fH, many of which have been linked to a severe kidney disease, atypical hemolytic uremic syndrome (aHUS). Here, some of the disease linked residues of fH are shown to be directly affected by dp8 binding. Altered heparin binding was confirmed for a range of aHUS variants of fH∼19-20 using GMSA and heparin oligosaccharides, dp4C and dp6. These data were in line with the results of heparin affinity chromatography [260]. Some aHUS mutants were binding more strongly, some more weakly, and some with affinity comparable to that of the wild type fH. The binding affinity was found to correlate almost exclusively with the charge density: a higher net positive charge of the protein led to increased affinity, a smaller net negative charge to decreased affinity. It was concluded that no casual link exists between the heparin binding strength and the aHUS linked variants of fH.

The aHUS related variants of fH∼19-20 exhibited non-uniform alterations not only in their heparin affinities but also with respect to a second binding partner of the C-terminus of factor H, C3b [260]. The heparin and C3b binding studies conducted in our laboratory and by our collaborators, respectively, showed that both heparin and C3b interactions of the aHUS related fH variants need to be investigated simultaneously in order to explain the molecular mechanism of aHUS. A possible route towards this goal is a NMR study of a ternary interaction between fH, C3b and heparin. As the first step, the complicated interplay of the three molecules was assessed in this work by using ^{15}N fH∼19-20 (wild type), C3d and dp8 [269]. The binding of C3b to fH∼19-20 was also monitored by NMR, but the size of C3b prevented its use in the study of the ternary

complex. C3d was used instead, as it is the part of C3b that interacts with fH∼19-20.

It was observed that an overlap of the heparin and C3d binding sites exists on CCP 20. The C3d binding site was found to span almost the full length of fH∼19-20. A number of NH cross peaks belonging to residues in the linker region between modules 19 and 20 were found to be broadened in the presence of C3d making it likely that a kink in the linker region exists in the fH∼19-20:C3d complex. The affinity of fH∼19-20 for dp8 is reduced in the presence of C3d, matching the observation that both binding sites overlap. Binding of dp8 but not C3d to fH∼19-20 could be reduced by a small increase in the buffer ionic strength, demonstrating that the heparin but not the C3d interaction of fH∼19-20 is dominated by charge-charge interactions [269].

While the heparin binding site in the C-terminus of factor H is contained within a single CCP module, CCP 20, the situation is more complicated for the second heparin binding site of fH centered in CCP 7. This module was thought to be surrounded by non-heparin binding modules prior to the onset of this work. Since then CCP 6 was implicated through the work of others, while we have shown that the linker region between modules 7 and 8 is the dominant heparin binding of the fH∼7-8 construct. Towards this end we have over expressed and ^{13}C, ^{15}N labelled fH∼7-8, assigned its back bone resonances and performed titrations with dp4C and the heparin mimic sucrose octasulfate (SOS). No contributions from CCP 8 to heparin binding were found beyond the linker region. A collaboration with an x-ray crystallography group during the course of this work confirmed that a residue in the linker region, R444, coordinates SOS in the crystal of fH∼6-8 (Y402 H) with SOS. Large contributions of CCP 6 to SOS binding were also observed in this crystal. These contributions were not matched by heparin binding affinities of protein constructs comprising different modules between CCP 6 and CCP 8. These results were corroborated by heparin titration of deuterated fH 6-8 triple module.

A substantial part of our efforts was dedicated to exploring the role of lysine side chains in protein GAG interactions [270]. Large chemical shift changes that we observed for the NH backbone cross peak of K446 in ^{1}H, ^{15}N-HSQC monitored titrations of fH∼7 and fH∼7-8 with dp4C, which would traditionally be interpreted as a strong indication of the interaction of this residue with heparin, were not confirmed by the x-ray data. The side chain of K446 does not interact with a SOS molecule. In order to better characterize the contributions of individual lysine residues to the binding, their side chain NH resonances were inspected during the titration experiments. The HISQC experiment, which monitors lysine $N^{\zeta}H^{\zeta}$ resonances, was found to have a limited use for the study of weak fH-heparin complexes as it requires low pH and low temperature, conditions that caused precipitation of our complex. Instead, a triple resonance H_2CN experiment that monitors H^{ϵ}, C^{ϵ}, and N^{ϵ} resonances was performed using double-labelled samples and physiological conditions. This experiment confirmed that the side chain of K446 is not involved in the interaction of fH∼7 with dp4C. This methodology is applicable to the study of GAG-protein complexes in general. It complements the routinely performed ^{1}H, ^{15}N-HSQC experiments that do not report on the behaviour of lysine side chains during the GAG titrations.

Using physiological conditions and zero cross linking methodology chemically activated dp4C was cross linked with a lysine side chain of unlabelled and ^{15}N labelled fH∼7. Accurate mass measurements confirmed that one tetrasaccharide was linked to one molecule of fH∼7. The low overall positive charge of this covalent complex was a likely reason for insufficient fragmentation by MS-MS experiments aimed at identifying the point of attachment. While this technique failed to identify the modified lysine, chemical and enzymatic degradation analysed by MALDI-TOF MS proved more informative. The combination of NMR (HISQC experiments) and MS data allowed the identification of K405 as the residue that was cross linked to the heparin tetrasaccharide.

Y402H is a single nucleotide polymorphism in CCP 7 which significantly increases the

risk of age related macula degeneration (AMD). For heparin binding studies with fH∼7 and fH∼7-8 two protein variants were produced and tested for each construct, Y402 and H402. No significant difference in the binding affinities of dp4C or SOS between the two variants was found during this study. We therefore suggest that heparin binding to CCP 7 is not affected by the 402 polymorphism, contrary to a hypothesis that was based on the x-ray structure of fH∼6-8 (H402) in complex with SOS. This conclusion is further supported by the result of our cross linking study in which the Y402 isoform was used. In this study a residue in close proximity to Y402, K405, was found to be cross linked to dp4C. While this result confirms the proximity of residue 402 to the heparin binding site it does not confirm the crystallographers' assumption that only a histidine residue in position 402 would allow heparin binding.

Further studies included in this thesis centered on the GAG binding characteristics of another protein of the innate immune system, human β-defensin 2 (HBD2), and the heparin binding domain of hepatocyte growth factor/scatter factor, NK1. For NK1, ^{1}H, ^{15}N-HSQC monitored titrations showed that the binding sites for heparin tetrasaccharide and dermatan sulfate hexasaccharide are overlapping, yet not identical [172]. The oligosaccharides used for the HBD2 study were the antithrombin binding heparin pentasaccharide Fondaparinux and dermatan sulfate hexasaccharide. T_1 and T_2 relaxation measurements revealed that Fondaparinux but not (or to a much smaller degree) dermatan sulfate likely initiates dimerisation of HBD2. This result was consistent with DOSY experiments conducted by Dr. Emily Seo [271].

The methods presented here, ranging from different NMR and MS techniques to chemical cross linking and physiological assays, can aid structural investigations in the growing field of GAG-protein interactions. This work contributes to the analysis of the interplay of factor H with a range of different binding partners, a complex equilibrium which allows this complement regulator to protect host cells from auto immune attack.

10 Appendix

10.0.1 Protein sequences

Sequences in brakets are cloning artefacts.

fH∼7Y (Uniprot database entry P08603):

390	400	410	420
(EAAG)LRKCY	FPYLENGYNQ	N*Y*GRKFVQGK	SIDVACHPGY
430	440	450	
ALPKAQTTVT	CMENGWSPTP	RCIRVK	

fH∼7-8Y (Uniprot database entry P08603):

390	400	410	420
(EAAG)LRKCY	FPYLENGYNQ	N*Y*GRKFVQGK	SIDVACHPGY
430	440	450	460
ALPKAQTTVT	CMENGWSPTP	RCIRVKTCSK	SSIDIENGFI
470	480	490	500
SESQYTYALK	EKAKYQCKLG	YVTADGETSG	SITCGKDGWS
510			
AQPTCIKS			

fH∼19-20 (Uniprot database entry P08603):

1110	1120	1130	1140
(EAEF)GKCG	PPPPIDNGDI	TSFPLSVYAP	ASSVEYQCQN
1150	1160	1170	1180
LYQLEGNKRI	TCRNGQWSEP	PKCLHPCVIS	REIMENYNIA
1190	1200	1210	1220
LRWTAKQKLY	SRTGESVEFV	CKRGYRLSSR	SHTLRTTCWD
1230	1240		
GKLEYPTCAK	R		

NK1 (sequence from [127], WT Uniprot database entry P14210):

30	40	50	60
YV(A in WT)E	GQRKRRNTIH	EFKKSAKTTL	IKIDPALKIK
70	80	90	100
TKKVNTADQC	ANRCTRNKGL	PFTCKAFVFD	KARKQCLWFP
110	120	130	140
FNSMSSGVKK	EFGHEFDLYE	NKDYIRNCII	GKGRSYKGTV
150	160	170	180
SITKSGIKCQ	PWSSMIPHEH	SFLPSSYRGK	DLQENYCRNP
190	200	210	
RGEEGGPWCF	TSNPEVRYEV	CDIPQCSEVE	

HBD2 (Uniprot database entry O15263):

30	40	50	60
GIGDPVT	CLKSGAICHP	VFCPRRYKQI	GTCGLPGTKC
70			
CKKP			

List of Figures

Bibliography

[1] J. Bordet, "Contribution à l'étude du sérum chez les animaux vaccinés," *Ann. Soc. Roy. Sci. Méd*, vol. 4, pp. 455–530, 1895.

[2] D. Mastellos, D. Morikis, S. N. Isaacs, M. C. Holland, C. W. Strey, and J. D. Lambris, "Complement: structure, functions, evolution, and viral molecular mimicry," *Immunol. Res.*, vol. 27, pp. 367–386, 2003.

[3] O. Barreiro, H. de la Fuente, M. Mittelbrunn, and F. Sànchez-Madrid, "Functional insights on the polarized redistribution of leukocyte integrins and their ligands during leukocyte migration and immune interactions," *Immunol. Rev.*, vol. 218, pp. 147–164, 2007.

[4] A. Aderem and D. M. Underhill, "Mechanisms of phagocytosis in macrophages," *Annu. Rev. Immunol.*, vol. 17, pp. 593–623, 1999.

[5] D. M. Underhill and A. Ozinsky, "Phagocytosis of microbes: complexity in action," *Annu. Rev. Immunol.*, vol. 20, pp. 825–852, 2002.

[6] J. Q. He, C. Wiesmann, and M. van Lookeren Campagne, "A role of macrophage complement receptor CRIg in immune clearance and inflammation," *Mol. Immunol.*, vol. 45, pp. 4041–4047, 2008.

[7] M. Cinco, R. Murgia, G. Presani, and S. Perticarari, "Integrin CR3 mediates the binding of nonspecifically opsonized Borrelia burgdorferi to human phagocytes and mammalian cells," *Infect. Immun.*, vol. 65, pp. 4784–4789, 1997.

[8] J. B. Cornacoff, L. A. Hebert, W. L. Smead, M. E. VanAman, D. J. Birmingham, and F. J. Waxman, "Primate erythrocyte-immune complex-clearing mechanism,"

J. Clin. Invest., vol. 71, pp. 236–247, 1983.

[9] Z. Chen, S. B. Koralov, and G. Kelsoe, "Complement C4 inhibits systemic autoimmunity through a mechanism independent of complement receptors CR1 and CR2," *J. Exp. Med.*, vol. 192, pp. 1339–1352, 2000.

[10] B. L. Myones, J. G. Dalzell, N. Hogg, and G. D. Ross, "Neutrophil and monocyte cell surface p150,95 has iC3b-receptor (CR4) activity resembling CR3," *J. Clin. Invest.*, vol. 82, pp. 640–651, 1988.

[11] D. Mevorach, J. O. Mascarenhas, D. Gershov, and K. B. Elkon, "Complement-dependent clearance of apoptotic cells by human macrophages," *J. Exp. Med.*, vol. 188, pp. 2313–2320, 1998.

[12] P. R. Taylor, A. Carugati, V. A. Fadok, H. T. Cook, M. Andrews, M. C. Carroll, J. S. Savill, P. M. Henson, M. Botto, and M. J. Walport, "A hierarchical role for classical pathway complement proteins in the clearance of apoptotic cells in vivo," *J. Exp. Med.*, vol. 192, pp. 359–366, 2000.

[13] L. A. Trouw, A. M. Blom, and P. Gasque, "Role of complement and complement regulators in the removal of apoptotic cells," *Mol. Immunol.*, vol. 45, pp. 1199–1207, 2008.

[14] M. C. Carroll, "The role of complement in B cell activation and tolerance," *Adv. Immunol.*, vol. 74, pp. 61–88, 2000.

[15] M. J. Walport, "Complement. Second of two parts," *N. Engl. J. Med.*, vol. 344, pp. 1140–1144, 2001.

[16] K. B. Reid, "Activation and control of the complement system," *Essays Biochem.*, vol. 22, pp. 27–68, 1986.

[17] A. R. Duncan and G. Winter, "The binding site for C1q on IgG," *Nature*, vol. 332, pp. 738–740, 1988.

[18] J. H. Lu, B. K. Teh, L. Wang, Y. N. Wang, Y. S. Tan, M. C. Lai, and K. B. Reid, "The classical and regulatory functions of C1q in immunity and autoimmunity," *Cell. Mol. Immunol.*, vol. 5, pp. 9–21, 2008.

[19] T. Fujita, M. Matsushita, and Y. Endo, "The lectin-complement pathway–its role in innate immunity and evolution," *Immunol. Rev.*, vol. 198, pp. 185–202, 2004.

[20] R. Roozendaal and M. C. Carroll, "Complement receptors CD21 and CD35 in humoral immunity," *Immunol. Rev.*, vol. 219, pp. 157–166, 2007.

[21] H. Müller-Eberhard, "Molecular organization and function of the complement system," *Annu. Rev. Biochem.*, vol. 57, pp. 321–347, 1988.

[22] J. Köhl, "Anaphylatoxins and infectious and non-infectious inflammatory diseases," *Mol. Immunol.*, vol. 38, pp. 175–187, 2001.

[23] Z. Fishelson, M. K. Pangburn, and H. J. Müller-Eberhard, "Characterization of the initial C3 convertase of the alternative pathway of human complement," *J. Immunol.*, vol. 132, pp. 1430–1434, 1984.

[24] S. Bhakdi and J. Tranum-Jensen, "Molecular nature of the complement lesion," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 75, pp. 5655–5659, 1978.

[25] R. G. DiScipio and C. Berlin, "The architectural transition of human complement component C9 to poly(C9)," *Mol. Immunol.*, vol. 36, pp. 575–585, 1999.

[26] H. J. Müller-Eberhard, "The membrane attack complex of complement," *Annu. Rev. Immunol.*, vol. 4, pp. 503–528, 1986.

[27] H. J. Müller-Eberhard, "Transmembrane channel-formation by five complement proteins," *Biochem. Soc. Symp.*, vol. 50, pp. 235–246, 1985.

[28] R. DiScipio and I. Schraufstatter, "The role of the complement anaphylatoxins in the recruitment of eosinophils," *Int. Immunopharmacol.*, vol. 7, pp. 1909–1923, 2007.

[29] M. L. Thomas, J. Janatova, W. R. Gray, and B. F. Tack, "Third component of human complement: localization of the internal thiolester bond," *Proc. Natl. Acad. Sci.*, vol. 79, pp. 1054–1058, 1982.

[30] M. Gadjeva, A. W. Dodds, A. Taniguchi-Sidle, A. C. Willis, D. E. Isenman, and S. K. Law, "The covalent binding reaction of complement component C3," *J. Immunol.*, vol. 161, pp. 985–990, 1998.

[31] B. J. Janssen, A. Christodoulidou, A. McCarthy, J. D. Lambris, and P. Gros, "Structure of C3b reveals conformational changes that underlie complement activity," *Nature*, vol. 444, pp. 213–216, 2006.

[32] B. J. Janssen, E. G. Huizinga, H. C. Raaijmakers, A. Roos, M. R. Daha, K. Nilsson-Ekdahl, B. Nilsson, and P. Gros, "Structures of complement component C3 provide insights into the function and evolution of immunity," *Nature*, vol. 437, pp. 505–511, 2005.

[33] R. B. Sim, T. M. Twose, D. S. Paterson, and E. Sim, "The covalent-binding reaction of complement component C3," *Biochem. J.*, vol. 193, pp. 115–127, 1981.

[34] A. Sahu, T. R. Kozel, and M. K. Pangburn, "Specificity of the thioester-containing reactive site of human C3 and its significance to complement activation," *Biochem. J.*, vol. 302 (Pt 2), pp. 429–436, 1994.

[35] M. K. Pangburn, R. D. Schreiber, and H. J. Müller-Eberhard, "Formation of the initial C3 convertase of the alternative complement pathway. Acquisition of C3b-like activities by spontaneous hydrolysis of the putative thioester in native C3," *J. Exp. Med.*, vol. 154, pp. 856–867, 1981.

[36] M. S. Winters, D. S. Spellman, and J. D. Lambris, "Solvent accessibility of native and hydrolyzed human complement protein 3 analyzed by hydrogen/deuterium exchange and mass spectrometry," *J. Immunol.*, vol. 174, pp. 3469–3474, 2005.

[37] P. Gros, F. J. Milder, and B. J. Janssen, "Complement driven by conformational changes," *Nat. Rev. Immunol.*, vol. 8, pp. 48–58, 2008.

[38] M. D. Kirkitadze and P. N. Barlow, "Structure and flexibility of the multiple domain proteins that regulate complement activation," *Immunol. Rev.*, vol. 180, pp. 146–161, 2001.

[39] L. Kask, B. O. Villoutreix, M. Steen, B. Ramesh, B. Dahlbäck, and A. M. Blom, "Structural stability and heat-induced conformational change of two complement inhibitors: C4b-binding protein and factor H," *Protein Sci.*, vol. 13, pp. 1356–1364, 2004.

[40] S. A. Smith, G. Krishnasamy, K. H. Murthy, A. Cooper, K. Bromek, P. N. Barlow, and G. J. Kotwal, "Vaccinia virus complement control protein is monomeric, and retains structural and functional integrity after exposure to adverse conditions," *Biochim. Biophys. Acta*, vol. 1598, pp. 55–64, 2002.

[41] H. G. Hocking, A. P. Herbert, D. Kavanagh, D. C. Soares, V. P. Ferreira, M. K. Pangburn, D. Uhrín, and P. N. Barlow, "Structure of the N-terminal region of complement factor H and conformational implications of disease-linked sequence variations," *J. Biol. Chem.*, vol. 283, pp. 9475–87, 2008.

[42] E. Hawrot, Y. Xiao, Q. L. Shi, D. Norman, M. Kirkitadze, and P. N. Barlow, "Demonstration of a tandem pair of complement protein modules in GABA(B) receptor 1a," *FEBS Lett.*, vol. 432, pp. 103–108, 1998.

[43] S. Blein, R. Ginham, D. Uhrín, B. O. Smith, D. C. Soares, S. Veltel, R. A. McIlhinney, J. H. White, and P. N. Barlow, "Structural analysis of the complement control protein (CCP) modules of GABA(B) receptor 1a: only one of the two CCP modules is compactly folded," *J. Biol. Chem.*, vol. 279, pp. 48292–48306, 2004.

[44] A. Ichinose, B. A. McMullen, K. Fujikawa, and E. W. Davie, "Amino acid sequence of the b subunit of human factor XIII, a protein composed of ten repetitive segments," *Biochemistry*, vol. 25, pp. 4633–4638, 1986.

[45] R. B. Sim, A. J. Day, B. E. Moffatt, and M. Fontaine, "Complement factor I and cofactors in control of complement system convertase enzymes," *Meth. Enzymol.*, vol. 223, pp. 13–35, 1993.

[46] R. G. DiScipio, "Ultrastructures and interactions of complement factors H and I," *J. Immunol.*, vol. 149, pp. 2592–2599, 1992.

[47] G. D. Ross, J. D. Lambris, J. A. Cain, and S. L. Newman, "Generation of three different fragments of bound C3 with purified factor I or serum. I. Requirements for factor H vs CR1 cofactor activity," *J. Immunol.*, vol. 129, pp. 2051–2060, 1982.

[48] S. Rodríguez de Córdoba, M. A. Díaz-Guillén, and D. Heine-Suñer, "An integrated

map of the human regulator of complement activation (RCA) gene cluster on 1q32," *Mol. Immunol.*, vol. 36, pp. 803–808, 1999.

[49] S. Rodríguez de Córdoba, D. M. Lublin, P. Rubinstein, and J. P. Atkinson, "Human genes for three complement components that regulate the activation of C3 are tightly linked," *J. Exp. Med.*, vol. 161, pp. 1189–1195, 1985.

[50] S. Rodríguez de Córdoba and P. Rubinstein, "New alleles of C4-binding protein and factor H and further linkage data in the regulator of complement activation (RCA) gene cluster in man," *Immunogenetics*, vol. 25, pp. 267–268, 1987.

[51] S. Meri, V. Koistinen, A. Miettinen, T. Törnroth, and I. J. Seppälä, "Activation of the alternative pathway of complement by monoclonal lambda light chains in membranoproliferative glomerulonephritis," *J. Exp. Med.*, vol. 175, pp. 939–950, 1992.

[52] M. C. Pickering, H. T. Cook, J. Warren, A. E. Bygrave, J. Moss, M. J. Walport, and M. Botto, "Uncontrolled C3 activation causes membranoproliferative glomerulonephritis in mice deficient in complement factor H," *Nat. Genet.*, vol. 31, pp. 424–428, 2002.

[53] V. P. Ferreira and M. K. Pangburn, "Factor H mediated cell surface protection from complement is critical for the survival of PNH erythrocytes," *Blood*, vol. 110, pp. 2190–2192, 2007.

[54] J. Esparza-Gordillo, J. M. Soria, A. Buil, L. Almasy, J. Blangero, J. Fontcuberta, and S. Rodríguez de Córdoba, "Genetic and environmental factors influencing the human factor H plasma levels," *Immunogenetics*, vol. 56, pp. 77–82, 2004.

[55] K. M. Morris, D. P. Aden, B. B. Knowles, and H. R. Colten, "Complement biosynthesis by the human hepatoma-derived cell line HepG2," *J. Clin. Invest.*, vol. 70, pp. 906–913, 1982.

[56] S. J. Perkins, A. S. Nealis, and R. B. Sim, "Oligomeric domain structure of human complement factor H by x-ray and neutron solution scattering," *Biochemistry*, vol. 30, pp. 2847–2857, 1991.

[57] R. Nan, J. Gor, I. Lengyel, and S. J. Perkins, "Uncontrolled zinc- and copper-induced oligomerisation of the human complement regulator factor H and its possible implications for function and disease," *J. Mol. Biol.*, vol. 384, pp. 1341–1352, 2008.

[58] M. K. Pangburn, N. Rawal, C. Cortes, M. N. Alam, V. P. Ferreira, and M. A. Atkinson, "Polyanion-induced self-association of complement factor H," *J. Immunol.*, vol. 182, pp. 1061–1068, 2009.

[59] F. Fenaille, M. Le Mignon, C. Groseil, C. Ramon, S. Riandé, L. Siret, and N. Bihoreau, "Site-specific N-glycan characterization of human complement factor H," *Glycobiology*, vol. 17, pp. 932–944, 2007.

[60] M. H. Jouvin, M. D. Kazatchkine, A. Cahour, and N. Bernard, "Lysine residues, but not carbohydrates, are required for the regulatory function of H on the amplification C3 convertase of complement," *J. Immunol.*, vol. 133, pp. 3250–3254, 1984.

[61] M. Aslam and S. J. Perkins, "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.*, vol. 309, pp. 1117–1138, 2001.

[62] C. Q. Schmidt, A. P. Herbert, H. D. T. Mertens, M. Guariento, D. C. Soares, D. Uhrín, A. J. Rowe, D. I. Svergun, and P. N. Barlow, "The central portion of factor H (modules 10-15) is compact and contains a structurally deviant CCP module," *J. Mol. Biol.*, doi:10.1016/j.jmb.2009.10.010.

[63] C. Estaller, W. Schwaeble, M. Dierich, and E. H. Weiss, "Human complement factor H: two factor H proteins are derived from alternatively spliced transcripts," *Eur. J. Immunol.*, vol. 21, pp. 799–802, 1991.

[64] W. Schwaeble, J. Zwirner, T. F. Schulz, R. P. Linke, M. P. Dierich, and E. H. Weiss, "Human complement factor H: expression of an additional truncated gene product of 43 kDa in human liver," *Eur. J. Immunol.*, vol. 17, pp. 1485–1489, 1987.

[65] R. Misasi, H. P. Huemer, W. Schwaeble, E. Sölder, C. Larcher, and M. P. Dierich, "Human complement factor H: an additional gene product of 43 kDa isolated from human plasma shows cofactor activity for the cleavage of the third component of complement," *Eur. J. Immunol.*, vol. 19, pp. 1765–1768, 1989.

[66] A. K. Sharma and M. K. Pangburn, "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.*, vol. 93, pp. 10996–11001, 1996.

[67] T. S. Jokiranta, J. Hellwage, V. Koistinen, P. F. Zipfel, and S. Meri, "Each of the three binding sites on complement factor H interacts with a distinct site on C3b," *J. Biol. Chem.*, vol. 275, pp. 27657–27662, 2000.

[68] C. Q. Schmidt, A. P. Herbert, D. Kavanagh, C. Gandy, C. J. Fenton, B. S. Blaum, M. Lyon, D. Uhrín, and P. N. Barlow, "A new map of glycosaminoglycan and C3b binding sites on factor H," *J. Immunol.*, vol. 181, pp. 2610–2619, 2008.

[69] V. P. Ferreira, A. P. Herbert, H. G. Hocking, P. N. Barlow, and M. K. Pangburn, "Critical role of the C-terminal domains of factor H in regulating complement activation at cell surfaces," *J. Immunol.*, vol. 177, pp. 6308–6316, 2006.

[70] M. K. Pangburn and H. J. Müller-Eberhard, "Complement C3 convertase: cell surface restriction of beta1H control and generation of restriction on neuraminidase-treated cells," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 75, pp. 2416–2420, 1978.

[71] D. T. Fearon, "Regulation by membrane sialic acid of beta1H-dependent decay-dissociation of amplification C3 convertase of the alternative complement pathway," *Proc. Natl. Acad. Sci.*, vol. 75, pp. 1971–1975, 1978.

[72] N. Okada, T. Yasuda, and H. Okada, "Restriction of alternative complement pathway activation by sialosylglycolipids," *Nature*, vol. 299, pp. 261–263, 1982.

[73] S. Ram, A. K. Sharma, S. D. Simpson, S. Gulati, D. P. McQuillen, M. K. Pangburn, and P. A. Rice, "A novel sialic acid binding site on factor H mediates serum resistance of sialylated *Neisseria gonorrhoeae*," *J. Exp. Med.*, vol. 187, pp. 743–752, 1998.

[74] M. D. Kazatchkine, D. T. Fearon, J. E. Silbert, and K. F. Austen, "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. Exp. Med.*, vol. 150, pp. 1202–1215, 1979.

[75] S. Meri and M. K. Pangburn, "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.*, vol. 87, pp. 3982–3986, 1990.

[76] M. K. Pangburn, K. L. Pangburn, V. Koistinen, S. Meri, and A. K. Sharma, "Molecular mechanisms of target recognition in an innate immune system: interactions among factor H, C3b, and target in the alternative pathway of human complement," *J. Immunol.*, vol. 164, pp. 4742–51, 2000.

[77] T. S. Jokiranta, Z. Z. Cheng, H. Seeberger, M. Józsi, S. Heinen, M. Noris, G. Remuzzi, R. Ormsby, D. L. Gordon, S. Meri, J. Hellwage, and P. F. Zipfel, "Binding of complement factor H to endothelial cells is mediated by the carboxy-terminal glycosaminoglycan binding site," *Am. J. Pathol.*, vol. 167, pp. 1173–1181, 2005.

[78] A. Richards, D. Kavanagh, and J. P. Atkinson, "Inherited complement regulatory protein deficiency predisposes to human disease in acute injury and chronic inflammatory states the examples of vascular damage in atypical hemolytic uremic syndrome and debris accumulation in age-related macular degeneration," *Adv. Immunol.*, vol. 96, pp. 141–177, 2007.

[79] A. P. Herbert, D. Uhrín, M. Lyon, M. K. Pangburn, and P. N. Barlow, "Disease-associated sequence variations congregate in a polyanion recognition patch on human factor H revealed in three-dimensional structure," *J. Biol. Chem.*, vol. 281, pp. 16512–16520, 2006.

[80] R. J. Klein, C. Zeiss, E. Y. Chew, J. Y. Tsai, R. S. Sackler, C. Haynes, A. K. Henning, J. P. SanGiovanni, S. M. Mane, S. T. Mayne, M. B. Bracken, F. L. Ferris, J. Ott, C. Barnstable, and J. Hoh, "Complement factor H polymorphism in age-related macular degeneration," *Science*, vol. 308, pp. 385–389, 2005.

[81] J. L. Haines, M. A. Hauser, S. Schmidt, W. K. Scott, L. M. Olson, P. Gallins,

K. L. Spencer, S. Y. Kwan, M. Noureddine, J. R. Gilbert, N. Schnetz-Boutaud, A. Agarwal, E. A. Postel, and M. A. Pericak-Vance, "Complement factor H variant increases the risk of age-related macular degeneration," *Science*, vol. 308, pp. 419–421, 2005.

[82] A. O. Edwards, R. Ritter, K. J. Abel, A. Manning, C. Panhuysen, and L. A. Farrer, "Complement factor H polymorphism and age-related macular degeneration," *Science*, vol. 308, pp. 421–424, 2005.

[83] W. K. Scott, S. Schmidt, M. A. Hauser, P. Gallins, N. Schnetz-Boutaud, K. L. S. J. R. Gilbert, A. Agarwal, E. A. Postel, J. L. Haines, and M. A. Pericak-Vance, "Independent effects of complement factor H Y402H polymorphism and cigarette smoking on risk of age-related macular degeneration," *Ophthalmology*, 2007.

[84] S. Rodríguez de Córdoba, J. Esparza-Gordillo, E. Goicoechea de Jorge, M. Lopez-Trascasa, and P. Sánchez-Corral, "The human complement factor H: functional roles, genetic variations and disease associations," *Mol. Immunol.*, vol. 41, pp. 355–367, 2004.

[85] H. Jarva, T. S. Jokiranta, J. Hellwaage, P. F. Zipfel, and S. Meri, "Regulation of complement activation by C-reactive protein: targeting the complement inhibitory activity of factor H by an interaction with short consensus repeat domains 7 and 8-11," *J. Immunol.*, vol. 163, p. 3957, 1999.

[86] C. Mold, H. Gewurz, and T. W. Du Clos, "Regulation of complement activation by C-reactive protein," *Immunopharmacology*, vol. 42, pp. 23–30, 1999.

[87] S. Hakobyan, C. L. Harris, C. W. van den Berg, M. C. Fernandez-Alonso, E. G. de Jorge, S. R. de Cordoba, G. Rivas, P. Mangione, M. B. Pepys, and B. P. Morgan, "Complement factor H binds to denatured rather than to native pentameric C-reactive protein," *J. Biol. Chem.*, vol. 283, pp. 30451–30460, 2008.

[88] M. J. Walport, "Complement. First of two parts," *N. Engl. J. Med.*, vol. 344, pp. 1058–1066, 2001.

[89] C. J. Fang, A. Richards, M. K. Liszewski, D. Kavanagh, and J. P. Atkinson, "Ad-

vances in understanding of pathogenesis of aHUS and HELLP," *Br. J. Haematol.*, vol. 143, pp. 336–348, 2008.

[90] S. Rodríguez de Córdoba and E. Goichoechea de Jorge, "Translational Mini-Review Series on Complement Factor H: Genetics and disease associations of human complement factor H," *Clin. Exp. Immunol.*, vol. 151, pp. 1–13, 2008.

[91] T. S. Jokiranta, V. P. Jaakola, M. J. Lehtinen, M. Pärepalo, S. Meri, and A. Goldman, "Structure of complement factor H carboxyl-terminus reveals molecular basis of atypical haemolytic uremic syndrome," *EMBO J.*, vol. 25, pp. 1784–1794, 2006.

[92] M. Józsi, S. Heinen, A. Hartmann, C. W. Ostrowicz, S. Hälbich, H. Richter, A. Kunert, C. Licht, R. E. Saunders, S. J. Perkins, P. F. Zipfel, and C. Skerka, "Factor H and atypical hemolytic uremic syndrome: mutations in the C-terminus cause structural changes and defective recognition functions," *J. Am. Soc. Nephrol.*, vol. 17, pp. 170–177, 2006.

[93] S. Heinen, P. Sánchez-Corral, M. S. Jackson, L. Strain, J. A. Goodship, E. J. Kemp, C. Skerka, T. S. Jokiranta, K. Meyers, E. Wagner, P. Robitaille, J. Esparza-Gordillo, S. Ródriguez de Córdoba, P. F. Zipfel, and T. H. Goodship, "De novo gene conversion in the RCA gene cluster (1q32) causes mutations in complement factor H associated with atypical hemolytic uremic syndrome," *Hum. Mutat.*, vol. 27, pp. 292–293, 2006.

[94] P. Sánchez-Corral, D. Pérez-Caballero, O. Huarte, A. M. Simckes, E. Goicoechea, M. López-Trascasa, and S. R. de Córdoba, "Structural and functional characterization of factor H mutations associated with atypical hemolytic uremic syndrome," *Am. J. Hum. Genet.*, vol. 71, pp. 1285–1295, 2002.

[95] P. Sánchez-Corral, C. González-Rubio, S. Rodríguez de Córdoba, and M. López-Trascasa, "Functional analysis in serum from atypical Hemolytic Uremic Syndrome patients reveals impaired protection of host cells associated with mutations in factor H," *Mol. Immunol.*, vol. 41, pp. 81–84, 2004.

[96] V. P. Ferreira, A. P. Herbert, C. Cortes, K. A. Mckeel, Blaum, B. S., D. Uhrín, P. N. Barlow, M. K. Pangburn, and D. Kavanagh, "The binding of Factor H to

a complex of physiological polyanions and C3b on cells is impaired in atypical hemolytic uremic syndrome," *J. Immunol.*, vol. 182(11), pp. 7009–18, 2009.

[97] E. Goicoechea de Jorge, C. L. Harris, J. Esparza-Gordillo, L. Carreras, E. A. Arranz, C. A. Garrido, M. López-Trascasa, P. Sánchez-Corral, B. P. Morgan, and S. Rodríguez de Córdoba, "Gain-of-function mutations in complement factor B are associated with atypical hemolytic uremic syndrome," *Proc. Natl. Acad. Sci.*, vol. 104, pp. 240–245, 2007.

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

[99] K. P. Magnusson, S. Duan, H. Sigurdsson, H. Petursson, Z. Yang, Y. Zhao, P. S. Bernstein, J. Ge, F. Jonasson, E. Stefansson, G. Helgadottir, N. A. Zabriskie, T. Jonsson, A. Bjornsson, T. Thorlacius, P. V. Jonsson, G. Thorleifsson, A. Kong, H. Stefansson, K. Zhang, K. Stefansson, and J. R. Gulcher, "CFH Y402H confers similar risk of soft drusen and both forms of advanced AMD," *PLoS Med.*, vol. 3, pp. 109–114, 2006.

[100] R. F. Mullins, S. R. Russell, D. H. Anderson, and G. S. Hageman, "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.*, vol. 14, pp. 835–846, 2000.

[101] L. V. Johnson, S. Ozaki, M. K. Staples, P. A. Erickson, and D. H. Anderson, "A potential role for immune complex pathogenesis in drusen formation," *Exp. Eye Res.*, vol. 70, pp. 441–449, 2000.

[102] P. T. Johnson, K. E. Betts, M. J. Radeke, G. S. Hageman, D. H. Anderson, and

L. V. Johnson, "Individuals homozygous for the age-related macular degeneration risk-conferring variant of complement factor H have elevated levels of CRP in the choroid," *Proc. Natl. Acad. Sci.*, vol. 103, pp. 17456–17461, 2006.

[103] R. J. Ormsby, S. Ranganathan, J. C. Tong, K. M. Griggs, D. P. Dimasi, A. W. Hewitt, K. P. Burdon, J. E. Craig, J. Hoh, and D. L. Gordon, "Functional and structural implications of the complement factor H Y402H polymorphism associated with age-related macular degeneration," *Invest. Ophthalmol. Vis. Sci.*, vol. 49, pp. 1763–1770, 2008.

[104] A. Day, "Y402H variants of complement factor H have different GAG-binding sites within human macula," Oral Communication, 67th Harden Conference, Cambridge 2009.

[105] K. A. Howes, Y. Liu, J. L. Dunaief, A. Milam, J. M. Frederick, A. Marks, and W. Baehr, "Receptor for advanced glycation end products and age-related macular degeneration," *Invest. Ophthalmol. Vis. Sci.*, vol. 45, pp. 3713–3720, 2004.

[106] A. King, E. Gottlieb, D. G. Brooks, M. P. Murphy, and J. L. Dunaief, "Mitochondria-derived reactive oxygen species mediate blue light-induced death of retinal pigment epithelial cells," *Photochem. Photobiol.*, vol. 79, pp. 470–475, 2004.

[107] K. G. Schmidt, H. Bergert, and R. H. Funk, "Neurodegenerative diseases of the retina and potential for protection and recovery," *Curr. Neuropharmacol.*, vol. 6, pp. 164–178, 2008.

[108] P. Kraiczy and R. Würzner, "Complement escape of human pathogenic bacteria by aquisition of complement regulators," *Mol. Imm.*, vol. 43, p. 31, 2006.

[109] J. Shaughnessy, L. A. Lewis, H. Jarva, and S. Ram, "Factor H SCR 6 binds to Factor H binding protein from neisseria meningitidis: functional comparison with binding of factor H SCRs 18-20 to neisseria gonorrhoeae porin," *Infect. Immun.*, 2009.

[110] H. G. Boman, "Peptide antibiotics and their role in innate immunity," *Annu. Rev.*

Immunol., vol. 13, pp. 61–92, 1995.

[111] J. J. Schneider, A. Unholzer, M. Schaller, M. Schäfer-Korting, and H. C. Korting, "Human defensins," *J. Mol. Med.*, vol. 83, pp. 587–595, 2005.

[112] D. M. Hoover, K. R. Rajashankar, R. Blumenthal, A. Puri, J. J. Oppenheim, O. Chertov, and J. Lubkowski, "The structure of human beta-defensin-2 shows evidence of higher order oligomerization," *J. Biol. Chem.*, vol. 275, pp. 32911–32918, 2000.

[113] J. M. Schröder and J. Harder, "Human beta-defensin-2," *Int. J. Biochem. Cell Biol.*, vol. 31, pp. 645–651, 1999.

[114] M. V. Sawai, H. P. Jia, L. Liu, V. Aseyev, J. M. Wiencek, P. B. McCray, T. Ganz, W. R. Kearney, and B. F. Tack, "The NMR structure of human beta-defensin-2 reveals a novel alpha-helical segment," *Biochemistry*, vol. 40, pp. 3810–3816, 2001.

[115] F. Bauer, K. Schweimer, E. Klüver, J. R. Conejo-Garcia, W. G. Forssmann, P. Rösch, K. Adermann, and H. Sticht, "Structure determination of human and murine beta-defensins reveals structural conservation in the absence of significant sequence similarity," *Protein Sci.*, vol. 10, pp. 2470–2479, 2001.

[116] B. L. Kagan, M. E. Selsted, T. Ganz, and R. I. Lehrer, "Antimicrobial defensin peptides form voltage-dependent ion-permeable channels in planar lipid bilayer membranes," *Proc. Natl. Acad. Sci.*, vol. 87, pp. 210–214, 1990.

[117] D. M. Hoover, K. R. Rajashankar, R. Blumenthal, A. Puri, J. J. Oppenheim, O. Chertov, and J. Lubkowski, "The structure of human beta-defensin-2 shows evidence of higher order oligomerization," *J. Biol. Chem.*, vol. 275(42), pp. 32911–8, 2000.

[118] L. Naldini, E. Vigna, R. P. Narsimhan, G. Gaudino, R. Zarnegar, G. K. Michalopoulos, and P. M. Comoglio, "Hepatocyte growth factor (HGF) stimulates the tyrosine kinase activity of the receptor encoded by the proto-oncogene c-MET," *Oncogene*, vol. 6, pp. 501–504, 1991.

[119] L. Naldini, E. Vigna, R. Ferracini, P. Longati, L. Gandino, M. Prat, and P. M.

Comoglio, "The tyrosine kinase encoded by the MET proto-oncogene is activated by autophosphorylation," *Mol. Cell. Biol.*, vol. 11, pp. 1793–1803, 1991.

[120] C. T. To and M. S. Tsao, "The roles of hepatocyte growth factor/scatter factor and met receptor in human cancers," *Oncol. Rep.*, vol. 5, pp. 1013–1024, 1998.

[121] C. Migliore and S. Giordano, "Molecular cancer therapy: can our expectation be MET?," *Eur. J. Cancer*, vol. 44, pp. 641–651, 2008.

[122] E. Gak, W. G. Taylor, A. M. Chan, and J. S. Rubin, "Processing of hepatocyte growth factor to the heterodimeric form is required for biological activity," *FEBS Lett.*, vol. 311, pp. 17–21, 1992.

[123] J. S. Rubin, R. M. Day, D. Breckenridge, N. Atabey, W. G. Taylor, S. J. Stahl, P. T. Wingfield, J. D. Kaufman, R. Schwall, and D. P. Bottaro, "Dissociation of heparan sulfate and receptor binding domains of hepatocyte growth factor reveals that heparan sulfate-c-met interaction facilitates signaling," *J. Biol. Chem.*, vol. 276, pp. 32977–32983, 2001.

[124] H. Zhou, M. J. Mazzulla, J. D. Kaufman, S. J. Stahl, P. T. Wingfield, J. S. Rubin, D. P. Bottaro, and R. A. Byrd, "The solution structure of the N-terminal domain of hepatocyte growth factor reveals a potential heparin-binding site," *Structure*, vol. 6, pp. 109–116, 1998.

[125] H. Zhou, J. R. Casas-Finet, R. Heath Coats, J. D. Kaufman, S. J. Stahl, P. T. Wingfield, J. S. Rubin, D. P. Bottaro, and R. A. Byrd, "Identification and dynamics of a heparin-binding site in hepatocyte growth factor," *Biochemistry*, vol. 38, pp. 14793–14802, 1999.

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

[127] D. Y. Chirgadze, J. P. Hepple, H. Zhou, R. A. Byrd, T. L. Blundell, and E. Gherardi, "NK1 fragment of human hepatocyte growth factor / scatter factor (HGF/SF)

at 2.5 Ångstrom resolution," *Nat. Struct. Biol*, vol. 6, p. 72, 1999.

[128] M. Kinosaki, K. Yamaguchi, A. Murakami, M. Ueda, T. Morinaga, and K. Higashio, "Identification of heparin-binding stretches of a naturally occurring deleted variant of hepatocyte growth factor (dHGF)," *Biochim. Biophys. Acta*, vol. 1384, pp. 93–102, 1998.

[129] T. Nakamura, H. Teramoto, and A. Ichihara, "Purification and characterization of a growth factor from rat platelets for mature parenchymal hepatocytes in primary cultures," *Proc. Natl. Acad. Sci. USA*, vol. 83, p. 6489, 1986.

[130] M. Lyon, J. A. Deakin, K. Mizuno, T. Nakamura, and J. T. Gallagher, "Interaction of hepatocyte growth factor with heparan sulfate. Elucidation of the major heparan sulfate determinants," *J. Biol. Chem.*, vol. 269, p. 11216, 1994.

[131] H. Rahmoune, P. S. Rudland, J. T. Gallagher, and D. G. Fernig, "Hepatocyte growth factor/scatter factor has distinct classes of binding site in heparan sulfate from mammary cells," *Biochemistry*, vol. 37, p. 6003, 1998.

[132] M. Lyon, J. A. Deakin, H. Rahmoune, D. G. Fernig, T. Nakamura, and J. T. Gallagher, "Hepatocyte growth factor/scatter factor binds with high affinity to dermatan sulfate," *J. Biol. Chem.*, vol. 273, pp. 271–278, 1998.

[133] D. Béchard, T. Gentina, M. Delehedde, A. Scherpereel, M. Lyon, M. Aumercier, R. Vazeux, C. Richet, P. Degand, B. Jude, A. Janin, D. G. Ferning, A.-B. Tonnel, and P. Lassalle, "Endocan is a novel chondroitin/dermatan sulfate proteoglycan that promotes hepatocyte growth factor/scatter factor nitrogenic activity," *J. Biol. Chem.*, vol. 276, p. 48341, 2001.

[134] R. H. Schwall, L. Y. Chang, P. J. Godowski, D. W. Kahn, K. J. Hillan, K. D. Bauer, and T. F. Zioncheck, "Heparin induces dimerization and confers proliferative activity onto the hepatocyte growth factor antagonists NK1 and NK2," *J. Cell Biol.*, vol. 133, pp. 709–718, 1996.

[135] M. Lyon, J. A. Deakin, and J. T. Gallagher, "The mode of action of heparan and dermatan sulfates in the regulation of hepatocyte growth factor/scatter factor," *J.*

Biol. Chem., vol. 277, pp. 1040–1046, 2002.

[136] J. A. Deakin and M. Lyon, "Differential regulation of hepatocyte growth factor/scatter factor by cell surface proteoglycans and free glycosaminoglycan chains," *J. Cell. Sci.*, vol. 112 (Pt 12), pp. 1999–2009, 1999.

[137] G. Hartmann, T. Prospero, V. Brinkmann, C. Ozcelik, G. Winter, J. Hepple, S. Batley, F. Bladt, M. Sachs, C. Birchmeier, W. Birchmeier, E. Gherardi, and O. Ozcelik, "Engineered mutants of HGF/SF with reduced binding to heparan sulphate proteoglycans, decreased clearance and enhanced activity in vivo," *Curr. Biol.*, vol. 8, pp. 125–134, 1998.

[138] H. Sakata, S. J. Stahl, W. G. Taylor, J. M. Rosenberg, K. Sakaguchi, P. T. Wingfield, and J. S. Rubin, "Heparin binding and oligomerization of hepatocyte growth factor/scatter factor isoforms. Heparan sulfate glycosaminoglycan requirement for Met binding and signaling," *J. Biol. Chem.*, vol. 272, pp. 9457–9463, 1997.

[139] K. Mizuno, H. Inoue, M. Hagiya, S. Shimizu, T. Nose, Y. Shimohigashi, and T. Nakamura, "Hairpin loop and second kringle domain are essential sites for heparin binding and biological activity of hepatocyte growth factor," *J. Biol. Chem.*, vol. 269, pp. 1131–1136, 1994.

[140] V. Cioce, K. G. Csaky, A. M.-L. Chan, D. P. Bottaro, W. G. Taylor, R. Jensen, S. A. Aaronson, and J. S. Rubin, "Hepatocyte growth factor (HGF)/NK1 is a naturally occuring HGF/scatter factor variant with partial agonist/antagonist activity," *J. Biol. Chem.*, vol. 271, p. 13110, 1996.

[141] N. A. Lokker and P. J. Godowski, "Generation and characterization of a competitive antagonist of human hepatocyte growth factor, HGF/NK1," *J. Biol. Chem.*, vol. 268, pp. 17145–17150, 1993.

[142] R. Montesano, J. V. Soriano, K. M. Malinda, M. L. Ponce, A. Bafico, H. K. Kleinman, D. P. Bottaro, and S. A. Aaronson, "Differential effects of hepatocyte growth factor isoforms on epithelial and endothelial tubulogenesis," *Cell Growth Differ.*, vol. 9, pp. 355–365, 1998.

[143] J. L. Jakubczak, W. J. LaRochelle, and G. Merlino, "NK1, a natural splice variant of hepatocyte growth factor/scatter factor, is a partial agonist in vivo," *Mol. Cell. Biol.*, vol. 18, pp. 1275–1283, 1998.

[144] A. Imberty, H. Lortat-Jacob, and S. Perez, "Structural view of glycosaminoglycan-protein interactions," *Carbohydr. Res*, vol. 342, p. 430, 2006.

[145] D. L. Rabenstein, "Heparin and heparan sulfate: structure and function," *Nat. Prod. Rep.*, vol. 19, pp. 312–331, 2002.

[146] C. Prussin and D. D. Metcalfe, "IgE, mast cells, basophils, and eosinophils," *J. Allergy Clin. Immunol.*, vol. 111(4), pp. S486–494, 2003.

[147] G. Nilsson, M. Johnell, C. H. Hammer, H. L. Tiffany, K. Nilsson, D. D. Metcalfe, A. Siegbahn, and P. M. Murphy, "C3a and C5a are chemotaxins for human mast cells and act through distinct receptors via a pertussis toxin-sensitive signal transduction pathway," *J. Immunol.*, vol. 157, pp. 1693–1698, 1996.

[148] K. J. Bame, I. Venkatesan, H. D. Stelling, and S. Tumova, "The spacing of S-domains on HS glycosaminoglycans determines whether the chain is a substrate for intercellular heparanases," *Glycobiology*, vol. 10, p. 715, 2000.

[149] M. Lyon and J. T. Gallagher, "Biospecific sequences and domains in heparan sulfate and the regulation of cell growth and adhesion," *Matrix Biology*, vol. 17, p. 485, 1998.

[150] K. J. Murphy, C. L. R. Merry, M. Lyon, J. E. Thompson, I. S. Roberts, and J. T. Gallagher, "A new model for the domain structure of heparan sulfate based on the novel specificity of K5 lyase," *J. Biol. Chem.*, vol. 279, pp. 27239–45, 2004.

[151] R. D. Rosenberg, "Biochemistry of heparin antithrombin interactions, and the physiologic role of this natural anticoagulant mechanism," *Am. J. Med.*, vol. 87, pp. 2S–9S, 1989.

[152] J. D. Esko and S. B. Selleck, "Order out of chaos: assembly of ligand binding sites in heparan sulfate," *Annu. Rev. Biochem.*, vol. 71, pp. 435–471, 2002.

[153] M. Kato, H. Wang, M. Bernfield, J. T. Gallagher, and J. E. Turnbull, "Cell surface

syndecan-1 on distinct cell types differs in fine structure and ligand binding of its heparan sulfate chains," *J. Biol. Chem.*, vol. 269, pp. 18881–18890, 1994.

[154] F. H. Blackhall, C. L. Merry, E. J. Davies, and G. C. Jayson, "Heparan sulfate proteoglycans and cancer," *Br. J. Cancer*, vol. 85, pp. 1094–1098, 2001.

[155] K. Fjeldstad and S. O. Kolset, "Decreasing the metastatic potential in cancers–targeting the heparan sulfate proteoglycans," *Curr. Drug Targets*, vol. 6, pp. 665–682, 2005.

[156] A. G. Marguiles, V. S. Klimberg, S. Bhattacharrya, D. Gaddy, and L. J. Suva, "Genomics and proteomics of bone cancer," *Clin. Cancer Res.*, vol. 12, pp. 6217s–6221s, 2006.

[157] E. Feyzi, T. Saldeen, E. Larsson, U. Lindahl, and M. Salmivirta, "Age-dependent modulation of heparan sulfate structure and function," *J. Biol. Chem.*, vol. 273, pp. 13395–13398, 1998.

[158] K. B. Komosińska-Vassev, K. Winsz-Szczotka, K. Kuznik-Trocha, P. Olczyk, and K. Olczyk, "Age-related changes of plasma glycosaminoglycans," *Clin. Chem. Lab. Med.*, vol. 46, pp. 219–224, 2008.

[159] P. N. Sanderson, T. N. Huckerby, and I. A. Nieduszynski, "Conformational equilibria of alpha-L-iduronate residues in disaccharides derived from heparin," *Biochem. J.*, vol. 243, pp. 175–181, 1987.

[160] D. R. Ferro, A. Provasoli, M. Ragazzi, B. Casu, G. Torri, V. Bossennec, B. Perly, P. Sinaÿ, M. Petitou, and J. Choay, "Conformer populations of L-iduronic acid residues in glycosaminoglycan sequences," *Carbohydr. Res.*, vol. 195, pp. 157–167, 1990.

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

[162] M. Hricovíni, M. Guerrini, A. Bisio, G. Torri, M. Petitou, and B. Casu, "Conformation of heparin pentasaccharide bound to antithrombin III," *Biochem. J.*, vol. 359, pp. 265–272, 2001.

[163] W. L. Chuang, M. D. Christ, J. Peng, and D. L. Rabenstein, "An NMR and molecular modeling study of the site-specific binding of histamine by heparin, chemically modified heparin, and heparin-derived oligosaccharides," *Biochemistry*, vol. 39, pp. 3542–3555, 2000.

[164] A. Canales, J. Angulo, R. Ojeda, M. Bruix, R. Fayos, R. Lozano, G. Giménez-Gallego, M. Martín-Lomas, P. M. Nieto, and J. Jiménez-Barbero, "Conformational flexibility of a synthetic glycosylaminoglycan bound to a fibroblast growth factor. FGF-1 recognizes both the $^{1}C_{4}$ and $^{2}S_{0}$ conformations of a bioactive heparin-like hexasaccharide," *J. Am. Chem. Soc.*, vol. 127, pp. 5778–5779, 2005.

[165] B. Mulloy, M. J. Forster, and C. Jones, "N.m.r. and molecular-modeling studies of the solution conformation of heparin," *Biochem. J.*, vol. 293, pp. 849–858, 1993.

[166] Z. Zhang, S. A. McCallum, J. Xie, L. Nieto, F. Corzana, J. Jiménez-Barbero, M. Chen, J. Liu, and R. J. Linhardt, "Solution structures of chemoenzymatically synthesized heparin and its precursors," *J. Am. Chem. Soc.*, vol. 130, pp. 12998–13007, 2008.

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

[168] L. Jin, M. Hricovíni, J. A. Deakin, M. Lyon, and D. Uhrín, "Does the flexibility of the iduronic acid affect the molecular shape of heparin?,"

[169] R. Raman, V. Sasisekharan, and R. Sasisekharan, "Structural insights into biological roles of protein-glycosaminoglycan interactions," *Chem. Biol.*, vol. 12, pp. 267–277, 2005.

[170] J. Kreuger, S. D., J. P. Li, and U. Lindahl, "Interactions between heparan sulfate

and proteins: the concept of specificity," *J. Cell Biol.*, vol. 174(3), pp. 323–7, 2006.

[171] U. Lindahl, "Heparan sulfate-protein interactions - A concept for drug design?," *Thromb. Haemost.*, vol. 98(1), pp. 109–15, 2007.

[172] J. A. Deakin, Blaum, B. S., J. T. Gallagher, D. Uhrín, and M. Lyon, "The binding properties of minimal oligosaccharides reveal a common heparan sulfate/dermatan sulfate-binding site in hepatocyte growth factor/scatter factor that can accommodate a wide variety of sulfation patterns," *J. Biol. Chem.*, vol. 284, pp. 6311–6321, 2009.

[173] S. O. Kolset, K. Prydz, and G. Pejler, "Intracellular proteoglycans," *Biochem. J.*, vol. 379, pp. 217–27, 2004.

[174] W. C. Lamanna, I. Kalus, M. Padva, R. J. Baldwin, C. L. Merry, and T. Dierks, "The heparanome–the enigma of encoding and decoding heparan sulfate sulfation," *J. Biotechnol.*, vol. 129, pp. 290–307, 2007.

[175] S. F. Penc, B. Pomahac, T. Winkler, R. A. Dorschner, E. Eriksson, M. Herndon, and R. L. Gallo, "Dermatan sulfate released after injury is a potent promoter of fibroblast growth factor-2 function," *J. Biol. Chem.*, vol. 273, pp. 28116–28121, 1998.

[176] K. R. Taylor, J. A. Rudisill, and R. L. Gallo, "Structural and sequence motifs in dermatan sulfate for promoting fibroblast growth factor-2 (FGF-2) and FGF-7 activity," *J. Biol. Chem.*, vol. 280, pp. 5300–5306, 2005.

[177] U. Priglinger, M. Geiger, E. Bielek, E. Vanyek, and B. R. Binder, "Binding of urinary protein C inhibitor to cultured human epithelial kidney tumor cells (TCL-598). The role of glycosaminoglycans present on the luminal cell surface," *J. Biol. Chem.*, vol. 269, pp. 14705–14710, 1994.

[178] G. Mascellani, L. Liverani, P. Bianchini, B. Parma, G. Torri, A. Bisio, M. Guerrini, and B. Casu, "Structure and contribution to the heparin cofactor II-mediated inhibition of thrombin of naturally oversulphated sequences of dermatan sulphate," *Biochem. J.*, vol. 296 (Pt 3), pp. 639–648, 1993.

[179] M. M. Maimone and D. M. Tollefsen, "Structure of a dermatan sulfate hexasaccharide that binds to heparin cofactor II with high affinity," *J. Biol. Chem.*, vol. 265, pp. 18263–18271, 1990.

[180] M. C. Bourin and U. Lindahl, "Glycosaminoglycans and the regulation of blood coagulation," *Biochem. J.*, vol. 289(2), pp. 313–330, 1993.

[181] J. C. Whisstock, R. N. Pike, L. Jin, R. Skinner, X. Y. Pei, R. W. Carrell, and A. M. Lesk, "Conformational changes in serpins: II. The mechanism of activation of antithrombin by heparin," *J. Mol. Biol.*, vol. 301, pp. 1287–1305, 2000.

[182] L. Jin, J. P. Abrahams, R. Skinner, M. Petitou, R. N. Pike, and R. W. Carrell, "The anticoagulant activation of antithrombin by heparin," *Proc. Natl. Acad. Sci.*, vol. 94, pp. 14683–14688, 1997.

[183] A. D. DiGabriele, I. Lax, D. I. Chen, C. M. Svahn, M. Jaye, J. Schlessinger, and W. A. Hendrickson, "Structure of heparin-linked biologically active dimer of fibroblast growth factor," *Nature*, vol. 393, pp. 812–817, 1998.

[184] S. Fahamn, R. E. Hileman, J. R. Fromm, R. J. Linhardt, and D. C. Rees, "Heparin structure and interactions with basic fibroblast growth factor," *Science*, vol. 271, pp. 1116–11620, 1996.

[185] L. Pellegrini, D. F. Burke, F. Delft, B. Mulloy, and T. L. Blundell, "Crystal structure of fibroblast growth factor receptor ectodomain bound to ligand and heparin," *Nature*, vol. 407, pp. 1029–1034, 2000.

[186] S. J., A. N. Plotnikov, O. A. Ibrahimi, A. V. Eliseenkova, B. K. Yeh, A. Yayon, R. J. Linhardt, and M. M., "Crystal structure of a ternary FGF-FGFR-heparin complex reveals a dual role for heparin in FGFR binding and dimerization," *Mol. Cell*, vol. 6, pp. 743–750, 2000.

[187] B. Mulloy and R. J. Linhardt, "Order out of complexity - protein structures that interact with heparin," *Curr. Opin. Strcut. Biol.*, vol. 11, pp. 623–628, 2001.

[188] N. J. Harmer, D. Chirgadze, K. H. Kim, K. L. Pellegrini, and T. L. Blundell, "The structural biology of growth factor receptor activation," *Biophys. Chem.*, vol. 100,

pp. 545–553, 2003.

[189] K. R. Catlow, J. A. Deakin, Z. Wei, M. Delehedde, D. G. Fernig, E. Gherardi, J. T. Gallagher, M. S. Pavão, and M. Lyon, "Interactions of hepatocyte growth factor/scatter factor with various glycosaminoglycans reveal an important interplay between the presence of iduronate and sulfate density," *J. Biol. Chem.*, vol. 283, pp. 5235–5248, 2008.

[190] A. E. I. Proudfoot, T. M. Handel, Z. Johnson, E. K. Lau, P. LiWang, I. Clark-Lewis, F. Borlat, T. N. Wells, and M. H. Kosco-Vilbois, "Glycosamnioglycan binding and oligomerization are essential for the in vivo activity of certain chemokines," *Proc. Natl. Acad. Sci. USA*, vol. 100, pp. 1885–90, 2003.

[191] T. M. Handel, Z. Johnson, S. E. Crown, E. K. Lau, M. Sweeney, and A. E. Proudfoot, "Regulation of protein function by glycosaminoglycans as exemplified by chemokines," *Annu. Rev. Biochem.*, vol. 74, pp. 385–410, 2005.

[192] E. K. Lau, C. D. Paavola, Z. Johnson, J. P. Gaudry, E. Geretti, F. Borlat, A. J. Kungl, A. E. Proudfoot, and T. M. Handel, "Identification of the glycosaminoglycan binding site of the cc chemokine, mcp-1: implications for structure and function in vivo," *J. Biol. Chem.*, vol. 279, pp. 22294–305, 2004.

[193] R. J. Linhardt and T. Toida, "Role of glycosaminoglycans in cellular communication," *Acc. Chem. Res.*, vol. 37, pp. 431–438, 2004.

[194] A. Coppi, R. Tewari, J. R. Bishop, B. L. Bennett, R. Lawrence, J. D. Esko, O. Billker, and P. Sinnis, "Heparan sulfate proteoglycans provide a signal to plasmodium sporozoites to stop migrating and productively invade host cells," *Cell Host Microbe*, vol. 2, pp. 316–327, 2007.

[195] A. Schulze, P. Gripon, and S. Urban, "Hepatitis B virus infection initiates with a large surface protein-dependent binding to heparan sulfate proteoglycans," *Hepatology*, vol. 46, pp. 1759–1768, 2007.

[196] M. Tyagi, M. Rusnati, M. Presta, and M. Giacca, "Internalization of HIV-1 tat requires cell surface heparan sulfate proteoglycans," *J. Biol. Chem.*, vol. 276,

pp. 3254–3261, 2001.

[197] M. R. Schenauer, Y. Yu, M. D. Sweeney, and J. A. Leary, "CCR2 chemokines bind selectively to acetylated heparan sulfate octasaccharides," *J. Biol. Chem.*, vol. 282, pp. 25182–25188, 2007.

[198] R. J. Linhardt, K. G. Rice, Y. S. Kim, D. L. Lohse, H. M. Wang, and D. Loganathan, "Mapping and quantification of the major oligosaccharide components of heparin," *Biochem. J.*, vol. 254, no. 3, pp. 781–787, 1988.

[199] A. Linker and P. Hovingh, "Structural studies on heparin. Tetrasaccharides obtained by heparinase degradation," *Carbohydr. Res.*, vol. 127, no. 1, pp. 75–94, 1984.

[200] E. Kupče and R. Freeman, "Distant echoes of the accordion: Reduced dimensionality, GFT-NMR, and projection-reconstruction of multidimensional spectra," *Concepts Magn. Res*, vol. 23A, pp. 63–75, 2004.

[201] S. Kim and T. Szyperski, "GFT NMR, a new approach to rapidly obtain precise high-dimensional NMR spectral information," *J. Am. Chem. Soc.*, vol. 125, pp. 1385–1393, 2003.

[202] D. Malmodin and M. Billeter, "Multiway decomposition of NMR spectra with coupled evolution periods," *J. Am. Chem. Soc.*, vol. 127, pp. 13486–13487, 2005.

[203] Y. Shen, R. Vernon, D. Baker, and A. Bax, "De novo protein structure generation from incomplete chemical shift assignments," *J. Biomol. NMR*, vol. 43, pp. 63–78, 2009.

[204] P. Robustelli, A. Cavalli, and M. Vendruscolo, "Determination of protein structures in the solid state from NMR chemical shifts," *Structure*, vol. 16, pp. 1764–1769, 2008.

[205] M. H. Levitt, *Spin Dynamics*. Wiley, 2001.

[206] G. A. Morris and R. Freeman, "Enhancement of nuclear magnetic resonance signals by polarization transfer," *J. Am. Chem. Soc.*, vol. 101, p. 760, 1979.

[207] S. Grzesiek and A. Bax, "Correlating backbone amide and sidechain resonances

in larger proteins by multiple relayed triple resonance NMR," *J. Am. Chem. Soc.*, vol. 114, p. 6291, 1992.

[208] M. Wittekind and L. Mueller, "A high sensivity 3D NMR experiment to correlate amide proton and nitrogen resonances with the alpha- and beta-carbon resonances in proteins," *J. Magn. Reson. B*, vol. 101, p. 201, 1993.

[209] S. Grzesiek and A. Bax, "Amino acid type determination in the sequential assignment procedure of uniformely $^{13}C/^{15}N$-Enriched Proteins," *J. Biomol. NMR*, vol. 3, p. 185, 1993.

[210] S. Grzesiek and A. Bax *J. Magn. Res*, vol. 96, p. 432, 1992.

[211] Y.-S. Jung and M. Zweckstetter, "Mars - robust automatic backbone assignment of proteins," *J. Biomol. NMR*, vol. 30, p. 11, 2004.

[212] L. E. Kay, L. K. Nicholson, F. Delagio, A. Bax, and D. Torchia, "Pulse sequences for removal of the effects of cross correlation between dipolar and chemical-shift anisotropy relaxation mechanisms on the measurement of heteronuclear T1 and T2 values in proteins," *J. Magn. Res*, vol. 97, pp. 359–375, 1992.

[213] N. A. Farrow, R. Muhandiram, A. U. Singer, S. M. Pascal, C. M. Kay, G. Gish, S. E. Shoelson, T. Pawson, J. D. Forman-Kay, and L. E. Kay, "Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by ^{15}N NMR relaxation," *Biochemistry*, vol. 33, pp. 5984–6003, 1994.

[214] R. Riek, K. Pervushin, and K. Wüthrich, "TROSY and CRINEPT: NMR with large molecular and supramolecular structures in solution," *Trends Biochem. Sci.*, vol. 25, pp. 462–468, 2000.

[215] K. Pervushin, R. Riek, G. Wider, and K. Wüthrich, "Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution," *Proc. Natl. Acad. Sci.*, vol. 94, pp. 12366–12371, 1997.

[216] R. Riek, G. Wider, K. Pervushin, and K. Wüthrich, "Polarization transfer by cross-correlated relaxation in solution NMR with very large molecules," *Proc. Natl. Acad.*

Sci., vol. 96, pp. 4918–4923, 1999.

[217] S. Grzesiek and A. Bax, "Amino acid type determination in the sequential assignment procedure of uniformly $^{13}C/^{15}N$-enriched proteins," *J. Biomol. NMR*, vol. 3, pp. 185–204, 1993.

[218] D. R. Muhandiram and L. E. Kay, "Gradient-enhanced triple-resonance three-dimensional NMR experiments with improved sensitivity," *J. Magn. Reson. B*, vol. 103, pp. 203–216, 1994.

[219] S. Grzesiek and A. Bax, "An efficient experiment for sequential backbone assignment of medium-sized isotopically enriched proteins," *J. Magn. Reson.*, vol. 99, pp. 201–207, 1992.

[220] S. Grzesiek and A. Bax, "Improved 3D triple-resonance NMR techniques applied to a 31 kDa protein," *J. Magn. Reson.*, vol. 96, pp. 432–440, 1992.

[221] L. E. Kay, G. Y. Xu, and T. Yamazaki, "Enhanced-sensitivity triple-resonance spectroscopy with minimal H_2O saturation," *J. Magn. Reson. A*, vol. 109, pp. 129–133, 1994.

[222] J. Schleucher, M. Sattler, and C. Griesinger, "Coherence Selection by Gradients without Signal Attenuation: Application to the Three-Dimensional HNCO Experiment," *Angew. Chem. Int. Ed.*, vol. 32, pp. 1489–1491, 1993.

[223] R. T. Clubb, V. Thanabal, and G. Wagner, "A constant-time three-dimensional triple-resonance pulse scheme to correlate intraresidue ^{1}HN, ^{15}N, and $^{13}C'$ chemical shifts in ^{15}N—^{13}C-labelled proteins," *J. Magn. Reson.*, vol. 97, pp. 213–217, 1992.

[224] W. F. Vranken, W. Boucher, T. J. Stevens, R. H. Fogh, A. Pajon, M. Llinas, E. L. Ulrich, J. L. Markley, J. Ionides, and E. D. Laue, "The CCPN data model for NMR spectroscopy: development of a software pipeline," *Proteins*, vol. 59, p. 687, 2005.

[225] L. J. McGuffin, K. Bryson, and D. T. Jones, "The PSIRED protein prediction server," *Bioinformatics*, vol. 16, p. 404, 2000.

[226] T. L. Hwang and A. J. Shaka, "Water supression that works. Excitation sculpting

using arbitrary wave-forms and pulsed-field fradients," *J. Magnet. Reson. Series A*, vol. 112, pp. 275–279, 1995.

[227] J.-R. Brisson, S.-C. Sue, W.-G. Wu, G. McManus, P. T. Nghia, and D. Uhrín, "NMR of carbohydrates: 1D homonuclear selective methods," *NMR spectroscopy of glycoconjugates*, pp. 59–93, 2003.

[228] H. Kessler, C. Oschkinat, C. Griesinger, and W. Bermel, "Transformation of homonuclear two-dimensional NMR techniques into one-dimensional techniques using Gaussian pulses," *J. Magnet. Reson.*, vol. 70, pp. 106–133, 1986.

[229] J. Stonehouse, P. Adell, J. Keeler, and A. J. Shaka, "Ultrahigh-quality NOE spectra," *J. Am. Chem. Soc.*, vol. 116, pp. 6037–6038, 1994.

[230] K. Stott, J. Stonehouse, J. Keeler, T.-L. Hwang, and A. J. Shaka, "Excitation sculpting in high-resolution nuclear magnetic resonance spectroscopy: application to selective NOE experiments," *J. Am. Chem. Soc.*, vol. 117, pp. 4199–4200, 1995.

[231] A. J. Shaka, C. J. Lee, and P. A., "Iterative schemes for bilinear operators: applications to spin decoupling," *J. Magnet. Reson.*, vol. 77, pp. 274–293, 1988.

[232] P. Jackson, "The analysis of fluorophore-labeled carbohydrates by polyacrylamide gel electrophoresis," *Mol. Biotechnol.*, vol. 5, pp. 101–123, 1996.

[233] M. Lyon, J. A. Deakin, D. Lietha, E. Gherardi, and J. T. Gallagher, "The interaction of hepatocythe growth factor/scatter factor and its NK1 and NK2 variants with glycosaminoglycans using a modified gel mobility shift assay," *J. Biol. Chem.*, vol. 279, p. 43560, 2004.

[234] D. Kavanagh, R. Burgess, D. Spitzer, A. Richards, M. L. Diaz-Torres, J. A. Goodship, D. E. Hourcade, J. P. Atkinson, and T. H. Goodship, "The decay accelerating factor mutation I197V found in hemolytic uraemic syndrome does not impair complement regulation," *Mol. Immunol.*, vol. 44, pp. 3162–3167, 2007.

[235] Z. Grabarek and J. Gergely, "Zero-length crosslinking procedure with the use of active esters," *Anal. Biochem.*, vol. 185, pp. 131–135, 1990.

[236] H. Zhong, S. L. Marcus, and L. Li, "Microwave-assisted acid hydrolysis of proteins

combined with liquid chromatography MALDI MS/MS for protein identification," *J. Am. Soc. Mass Spectrom.*, vol. 16, pp. 471–481, 2005.

[237] J. Jiménez-Barbero and T. Peters, eds., *NMR spectroscopy of glycoconjugates*. Wiley-VCH, Weinheim, 2003.

[238] L. Jin, "Conformation of glycosaminoglycans by ion mobility mass spectrometry." 1st Year Report, University of Edinburgh, 2004.

[239] W. Chai, E. F. Hounsell, C. J. Bauer, and A. M. Lawson, "Characterisation by LSI-MS and ^{1}H NMR spectroscopy of tetra-, hexa- and octasaccharides of porcine intestinal heparin," *Carbohydr. Res.*, vol. 269, pp. 139–56, 1995.

[240] A. P. Herbert, J. A. Deakin, C. Q. Schmidt, B. S. Blaum, C. Egan, V. P. Ferreira, M. K. Pangburn, M. Lyon, D. Uhrín, and P. N. Barlow, "Structure shows glycosaminoglycan- and protein-recognition site in factor H is perturbated by age-related macular degeneration-linked SNP," *J. Biol. Chem.*, vol. 287, pp. 18960–68, 2007.

[241] B. E. Prosser, S. Johnson, P. Roversi, A. P. Herbert, Blaum, B. S., J. Tyrrell, T. A. Jowitt, S. J. Clark, E. Tarelli, D. Uhrín, P. N. Barlow, R. B. Sim, A. J. Day, and S. M. Lea, "Structural basis for complement factor H-linked age-related macular degeneration," *J. Exp. Med.*, vol. 204(10), pp. 2277–83, 2007.

[242] M. Gralle, C. L. Oliviera, L. H. Guerreiro, W. J. McKinstry, D. Galatis, C. L. Masters, R. Cappai, M. W. Parker, C. H. Ramos, I. Torriani, and S. T. Ferreira, "Solution conformation and heparin-induced dimerization of the full-length extracellular domain of the human amyloid precursor protein," *J. Mol. Biol.*, vol. 357(2), pp. 493–508, 2006.

[243] C. J. Robinson, N. J. Harmer, S. J. Goodger, T. L. Blundell, and J. T. Gallagher, "Cooperative dimerization of fibroblast growth factor 1 (FGF1) upon a single heparin saccharide may drive the formation of 2:2:1 FGF1*FGF2c*heparin ternary complexes," *J. Biol. Chem.*, vol. 280, pp. 42274–82, 2005.

[244] K. Tan, M. Duquette, J. Liu, A. J. Shanmugasundaram, A. Joachimiak, J. T.

Gallagher, A. C. Rigby, J. Wang, and J. Lawler, "Heparin-induced *cis*- and *trans*-dimerization mode of the thrombospondin-1 N-terminal domain," *J. Biol. Chem.*, vol. 283, pp. 3932–41, 2008.

[245] Y. Chi, S. K. Krishnaswamy, I. Chiu, and C. Yu, "^{15}N relaxation studies of free and ligand bound human acidic fibroblast growth factor," *J. Biol. Chem.*, vol. 275(50), pp. 39444–50, 2000.

[246] G. G. K. Editor Roberts, "NMR of Macromolecules, A practical Approach," *Oxford university press*, p. 164 ff., 1993.

[247] A. N. Fernando, P. B. Furtado, S. J. Clark, H. E. Gilbert, A. J. Day, R. B. Sim, and S. J. Perkins, "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.*, vol. 368, 2007.

[248] J. Iwahara, Y. S. Jung, and G. M. Clore, "Heteronuclear NMR spectroscopy for lysine NH_3 groups in proteins: unique effect of water exchange on ^{15}N transverse relaxation," *J. Am. Chem. Soc.*, vol. 129(10), pp. 2971–80, 2007.

[249] J. Rappsilber, S. Siniossoglou, E. C. Hurt, and M. Mann, "A generic strategy to analyze the spatial organization of multi-protein complexes by cross-linking and mass spectrometry," *Anal. Chem.*, vol. 72(2), pp. 267–75, 2000.

[250] J. H. Tomlinson, S. Ullah, P. E. Hansen, and M. P. Williamson, "Characterization of salt bridges to lysines in the protein G B1 domain," *J. Am. Chem. Soc.*, vol. 131(13), pp. 4674–84, 2009.

[251] I. André, S. Linse, and F. A. Mulder, "Residue-specific pKa determination of lysine and arginine side chains by indirect ^{15}N and ^{13}C NMR spectroscopy: application to apo calmodulin," *J. Am. Chem. Soc.*, vol. 129(51), pp. 15805–13, 2007.

[252] S. J. Clark, V. A. Higman, B. Mulloy, S. J. Perkins, S. M. Lea, R. B. Sim, and A. J. Day, "His-384 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.*, vol. 281(34), pp. 24713–24720, 2006.

[253] J. Wu, Y. Q. Wu, D. Ricklin, B. J. Janssen, J. D. Lambris, and P. Gros, "Structure of complement fragment C3b-factor H and implications for host protection by complement regulators," *Nat. Immunol.*, vol. 10(7), pp. 728–33, 2009.

[254] B. J. Janssen, A. Christodoulidou, A. McCarthy, D. J. Lambris, and P. Gros, "Structure of C3b reveals conformational changes that underlie complement activity," *Nature*, vol. 444(7166), pp. 213–6, 2006.

[255] B. Nagar, R. G. Jones, R. J. Diefenbach, D. E. Isenman, and J. M. Rini, "X-ray structure of C3d: a C3 fragment and ligand for complement receptor 2," *Science*, vol. 280(5367), pp. 1277–81, 1998.

[256] H. Matsuo, K. J. Walters, K. Teruya, T. Tanaka, G. T. Gassner, S. J. Lippard, Y. Kyogoku, and G. Wagner, "Identification by NMR spectrescopy of residues at contact surfaces in large, slowly exchanging macromolecular complexes," *J. Am. Chem. Soc.*, vol. 121, pp. 9903–04, 1999.

[257] M. J. Lehtinen, A. L. Rops, D. E. Isenman, J. van der Vlag, and T. S. Jokiranta, "Mutations of factor H impair regulation of surface-bound C3b by three mechanisms in atypical hemolytic uremic syndrome," *J. Biol. Chem.*, vol. 284(23), pp. 10–8, 2009.

[258] J. Hellwaage, T. S. Jokiranta, M. A. Friese, T. U. Wolk, E. Kampen, P. F. Zipfel, and S. Meri, "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.*, vol. 169, pp. 6935–6944, 2002.

[259] A. Richards, M. R. Buddles, R. L. Donne, B. S. Kaplan, E. Kirk, M. C. Veninng, C. L. Tielemans, J. A. Goodship, and T. H. Goodship, "Factor H mutations in hemolytic uremic syndrome cluster in exons 18-20, a domain important for hist recognition," *Am. J. Hum. Genet.*, vol. 68(2), pp. 485–90, 2001.

[260] V. P. Ferreira, A. P. Herbert, C. Cortes, K. A. McKee, Blaum, B. S., S. T. Esswein, D. Uhrín, P. N. Barlow, M. K. Pangburn, and D. Kavanagh, "The binding of factor H to a complex of physiological polyanions and C3b on cells is impaired in atypical hemolytic uremic syndrome," *J. Immunol.*, vol. 182(11), pp. 7009–18, 2009.

[261] G. Szakonyi, J. M. Guthridge, D. Li, K. Young, V. M. Holers, and X. S. Chen, "Structure of complement receptor 2 in complex with its C3d ligand," *Sience*, vol. 292(5522), pp. 1725–8, 2001.

[262] R. J. Ormsby, T. S. Jokiranta, T. G. Duthy, K. M. Griggs, T. A. Sadlon, E. Giannakis, and D. L. Gordon, "Localization of the third heparin-binding site in the human complement regulator factor h," *Mol. Immunol.*, vol. 43, pp. 1624–1632, 2006.

[263] H. Zhou, J. R. Casas-Finet, R. H. Coats, J. D. Kaufman, S. J. Stahl, P. T. Wingfield, J. S. Rubin, D. P. Bottaro, and A. Byrd, "Identification and dynamics of a heparin-binding site in hepatocyte growth factor," *Biochemistry*, pp. 14793–14802, 1999.

[264] E. S. Seo, T. Vargues, D. J. Clarke, D. Uhrín, and D. J. Campopiano, "Preparation of isotopically labelled recombinant beta-defensin for nmr studies," *Protein Expr Purif*, vol. 65, pp. 179–84, 2009.

[265] M. V. Sawai, H. P. Jia, L. Liu, V. Aseyev, J. M. Wiencek, P. B. McCray, T. Ganz, W. R. Kearney, and B. F. Tack, "The NMR structure of human beta-defensin-2 reveals a novel alpha-helical segment.," *Biochemistry*, vol. 40(13), pp. 3810–6, 2001.

[266] F. Bauer, K. Schweimer, E. Kluver, J. R. Conejo-Garcia, W. G. Forssmann, P. Rosch, K. Adermann, and H. Sticht, "Structure determination of human and murine beta-defensins reveals structural conservation in the absence of significant sequence similarity.," *Prot. Sci.*, vol. 10(12), pp. 2470–9, 2001.

[267] M. Hricovini and G. Torri, "Dynamics in aqueous solutions of the pentasaccharide corresponding to the binding site of heparin for antithrombin III studied by NMR relaxation measurements.," *Carbohydr. Res.*, vol. 268(2), pp. 159–75, 1995.

[268] A. Silipo, Z. Zhang, F. J. Canada, R. J. Linhardt, and J. Jiménez-Barbero, "Conformational analysis of a dermatan sulfate-derived tetrasaccharide by nmr, molecular modeling, and residual dipolar couplings," *Chembiochem*, vol. 9(2), pp. 240–52, 2008.

[269] Blaum, B.S., A. Herbert, C. Johansson, P. Barlow, and D. Uhrín, "Atomic resolution study in solution confirms overlapping binding sites for heparin and C3d on the C-terminal module of complement factor H," in preparation.

[270] Blaum, B.S., J. Deakin, C. Johansson, A. Herbert, P. Barlow, M. Lyon, and D. Uhrín, "Lysine and arginine side-chains in glycosaminoglycan-protein complexes investigated by nmr, cross-linking and mass spectrometry. a case study of the factor H : heparin interaction," submitted.

[271] E. S. Seo, Blaum, B. S., T. Vargues, J. A. Deakin, M. Lyon, D. J. Campopiano, and D. Uhrín, "Interaction studies of human β-defensin 2 (HBD2) with glycosaminoglycans," in preparation.