UNIVERSITY OF LIVERPOOL

Institute of Integrative Biology

Selective labelling of arginine residues in protein sulfated glycosaminoglycan binding sites

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

BUI PHUONG THAO

JULY 2019

To my parents with love

Author's declaration

I declare that the work in this dissertation was carried out in accordance with the regulations of the University of Liverpool. The work described is original and has not been submitted for any other degree. All aspects of the experimental design and planning for the study were conducted by me in conjunction with my supervisors, Dr. Quentin M. Nunes, Dr. Mark C. Wilkinson and my primary supervisor Professor David G. Fernig. The experimental work in this dissertation has been undertaken by me, with specific contributions that I have indicated below and in the text.

Any views in this thesis are those of the author and in no way represent those of the University of Liverpool. This thesis has not been presented to any other university for examination in the United Kingdom or overseas.

Acknowledgements

I would like to express my sincerest gratitude and appreciation to those who have been generous to guide, advice, and support me for my PhD study through the valuable four years.

This work would not have been possible without the input of many people but first and foremost, I would like to express my thankfulness to my supervisor Prof **Dave Fernig** for his wealth of knowledge and experience that helped me tremendously with my work. He has been supporting me in terms of encouragement, guidance and advice, not only just in academic work, but also in a challenging moment in my personal life. Next, my gratitude delivers to my second supervisors **Dr Mark C. Wilkinson** and **Dr Quentin M. Nunes** for their kind support and guide. Moreover, I express thankful gratitude to the Vietnamese Ministry of Education, 911 scholarship and Newton Fund of British Council for their generous sponsorship during the four years of study.

I would also like to thank the colleagues I have worked with in these four years, especially my friends from Lab D. I have had lots of advice from them, which has been very useful in my work. Thank Dr Yong Li for his kind guidance at the beginning of my PhD. Thanks to Pawin Ngamlert, Dr Baradji Aïseta, Dr Dunhao Sun and Dr Krithika Ramakrishnan for their daily support and friendships. My special thanks to Dr Tram Nguyen for being my friend and for supporting me through many challenges over the past 4 years.

Last but not least, special thanks to my parents, Mr **Bui Quang Thoi** and Mrs **Luu Thi Lam**, as well as my family for their tireless support during this amazing journey, for always being by my side.

Abstract

In animals, the interaction of numerous proteins with glycosaminoglycans (GAGs) regulates many aspects of their functions, and so organism development and homeostasis, as well many related pathologies. The interactions between the GAG heparan sulfate (HS) and the fibroblast growth factors (FGFs) have become a major paradigm for how such interactions regulate biological outcomes. Thus, binding HS controls FGF stability, its diffusion between cells and the generation of intracellular signals *via* the cognate FGF receptors (FGFRs). The basic arginine and lysine amino acids on the surface of FGFs have an important role in binding the polysaccharide via its numerous sulfate and carboxyl groups. Indeed electrostatic bonds between FGF and GAG dominate the interaction, at least kinetically. However, the contribution of arginine residues on the protein surface to GAG binding has generally not been established. This has been addressed directly in this thesis. A means to selectively label and so identify arginine side chains involved in binding heparin as an approximation for cellular GAGs was developed and evaluated, using the two most studied FGFs, FGF1 and FGF2. Two chemicals which selectively and specifically react with the guanidino group of arginine, phenylglyoxal (PGO) and hydroxyphenylglyoxal (HPG) were chosen for the purpose of this work. By combining mass spectrometry and an automatic programming language, the multiple products of PGO's reaction with arginine were readily deconvoluted.

The method was then applied to the other thirteen paracrine FGFs, which were produced as recombinant proteins. In addition, lysine selective labelling data were acquired for those FGFs where this was lacking. Thus, a complete description of the surface electrostatic map on fifteen paracrine FGFs guiding heparin engagement was obtained. The addition of arginine residues into the map of binding makes adjustments to the definition of heparin binding sites (HBSs) in some FGFs, for instance, FGF4. The refinement of the assignment of residues to secondary

HBSs is consistent with data obtained by others on the ability of particular FGFs (FGF4, FGF6, FGF17, FGF18 and FGF10) to cross-link HS chains. IN addition, the data raise some interesting properties of GAG binding, for example, the dual specificity of the secondary heparin binding site, HBS-3, in some FGFs, which is also involved in binding the FGFR. While the results generally support the idea that the conservation of structures for GAG binding in FGFs is related to their evolutionary divergence, it also suggests that changes in the structures in GAG binding may in some instances have diverged more rapidly since an amino acid substitution between members of a subfamily enables a significant alteration in heparin engagement. This suggests that in some instances, changes to the heparin binding properties of an FGF may contribute to a diversification of function within a subfamily.

A further method was developed, whereby the entire selective labelling was done in solution, without the need for a heparin affinity column. The method was applied to the interactions of members of the FGF7 subfamily (FGF3, FGF7, and FGF10) with heparin, followed by chondroitin sulfate (CS), dermatan sulfate (DS). This allowed the analysis of lysine residues involved in binding physiologically relevant GAGs rather than heparin. Finally, the wider applicability of the method was acquired by determining the lysine and arginine residues that bind heparin in a 'phage display antibody', HS4C3. Due to the selectivity of HS4C3 for anticoagulant structures in heparin through its CDR3 loop, it was possible to use the data to propose a means to model protein-GAG binding. This was demonstrated using the knowledge of the binding specificity of HS4C3 for 3-*O*-sulfate and the positions of residues identified by the selective labelling. Taken together, the work in this thesis provides new insight into the involvement of arginine and lysine residues in the engagement of proteins to GAGs. This may enable future 'omics' approaches to identify GAG binding sites in proteins in cells and tissues, and to predict from principles where such sites are on a protein surface.

Contents

List of Figures:

List of Tables:

List of Abbreviations

ACN: Acetonitrile

Arg: Arginine

BSA: Bovine serum, albumin

CS: Chondroitin sulfate

DMSO: Dimethyl sulphoxide

DMEM: Dulbecco's modified Eagle medium

dp: Degree of polymerization

DS: Dermatan sulfate

DSF: Differential scanning fluorimetry

DTT: Dithiothreitol

ECM: Extracellular matrix

EDTA: Ethylenediamine tetra-acetic acid

FA: Fibronectin adhesion-promoting peptide

FGF: Fibroblast growth factor

FGFR: Fibroblast growth factor receptor

HA: Hyaluronic acid

HBS: Heparin binding site

HPG: P-hydroxyl phenylglyoxal

HS: Heparan sulfate

HSPGs: Heparan sulfate proteoglycans

GAG: Glycosaminoglycans

GlcNAc: N-acetylglucosamine

GlcNS: N-sulfoglucosamine

IdoA: L-iduronic acid

IPTG: Isopropyl β-D-1-thiogalactopyranoside

LB: Lysogeny broth

Lys: Lysine

MW: Molecular weight

NHS: N-hydroxysuccinimide

PAGE: Polyacrylamide gel electrophoresis

PBS: Phosphate-buffered saline

PCR: Polymerase chain reaction

PDB: Protein data bank

PGO: Phenylglyoxal

PI: Phosphatidylinositol

RE: Reducing end

SDS: Sodium dodecyl sulphate

TB: Terrific Broth

TCA: Trichloroacetic acid

TEMED: N,N,N',N',Tetramethylethylenediamine

TFA: Trifluoroacetic acid

TM: Melting temperature

Tris: Tris (hydroxymethyl) methylamine

Tween 20: Polyoxyethylenesorbitan monolaurate

CHAPTER 1 GENERAL INTRODUCTION

1.1 OVERVIEW

Communication through the assembly of ligand and receptor complex present at an extracellular interface is common in multicellular organism. The specificity in binding between FGF, its unique subset of FGFRs and the cofactor HS tightly controls the response to FGF simulation [1]. The engagement of FGF with HS also regulates the transport of different FGFs in extracellular matrix (ECM) [2]. HS is ubiquitous in the ECM and are important actors to guide complex biological phenomena, such as embryonic development and organism homeostasis. Besides HS, other two proteoglycans CS and DS also interact with FGFs, however, the structural properties of these interactions have been obscured. What are the structures of FGFs, HS, CS and DS?? What do we know about these interactions in terms of structures and functions? How have we studied them? Those three points will be briefly introduced later.

1.2 FIBROBLAST GROWTH FACTORS

Fibroblast growth factors (FGFs) are a family of 22 members in mammals (FGF1 to FGF23, with humans possessing FGF19, but not FGF15 and mice FGF15, but not FGF19). These proteins have 13 % to 71 % sequence identity and a highly conserved central β trefoil structure. First identified were basic FGF (FGF2) and acidic FGF (FGF1) which were isolated from pituitary and brain [3][4][5]. Thereafter, twenty-two *fgf* genes have been added to the list of FGF family, *fgf1-23*. Among them, fifteen of the FGFs are paracrine effectors, which bind the glycosaminoglycan HS and these are the focus of this study.

1.2.1 FGF sub-families

The original classification of FGFs was made according to their amino acid sequence conservation [6]. Following the Human Genome Project, which defined the human *fgf* genes, the same approach was used to establish phylogeny of the entire family. Accordingly, the 22 FGFs in mammals are classified into 7 sub-families, FGF1/2; FGF4/5/6; FGF7/3/10/22; FGF8/17/18; FGF9/16/20; FGF11/12/13/14 and FGF19/21/23 [1]. However, the classification of FGF3 and FGF5 is debated and more recently, other classifications have been proposed. First, the work based on synteny in invertebrates [7] and vertebrates [8] suggested that FGF3 would be in the same subfamily as FGF4 and FGF6, and further grouped with the FGF19/21/23 subfamily to form a single subfamily FGF3/4/6/19/21/13. The same analysis indicates that FGF5 would belong to the FGF1/2 subfamily. A deeper approach was taken by Oulion and colleagues [9], who used many characteristics of FGFs including gene content, phylogenetic distribution, and conservation of synteny to classify FGF3 into a new single family. This study also used a deeper synteny analysis of human FGF5 locus to claim that this FGF belongs to FGF4 family rather than FGF2 family.

Based on their location, their binding partners and so their functions, FGFs are classified as intracellular, endocrine and paracrine. Paracrine FGFs are the largest group comprising the FGF1, FGF4, FGF7, FGF8, and FGF9 subfamilies. These all bind HS and possess a dual receptor system of HS co-receptor and FGF receptor (FGFR) tyrosine kinase. The FGF19 subfamily (in humans FGF19, FGF21 and FGF23) do not bind HS and signal through a dual receptor system consisting of members of Klotho family, α-Klotho or β-Klotho and an FGFR [10]. The absence of binding to HS enables their systemic release. Bearing high sequence and structural identity to FGFs are the FGF homologous factors (FHFs) or so called intracellular FGFs, also sometimes called (incorrectly) intracrine FGFs. These are the FGF11 subfamily with FHF1 = FGF12, FHF2 = FGF13, FHF3 = FGF11, FHF4 = FGF14 [11][12]. Unlike other

FGFs that function through FGFR-dependent manner, FHFs bind to intracellular proteins and modulate their function. For example, FHF1, FHF2 and FHF4 bind to the cytosolic carboxyl terminal tail of voltage-gated sodium channels (VGSCs) [13] and FHF2 to a neuronal MAP kinase scaffold protein, islet-brain-2 (IB2) [14].

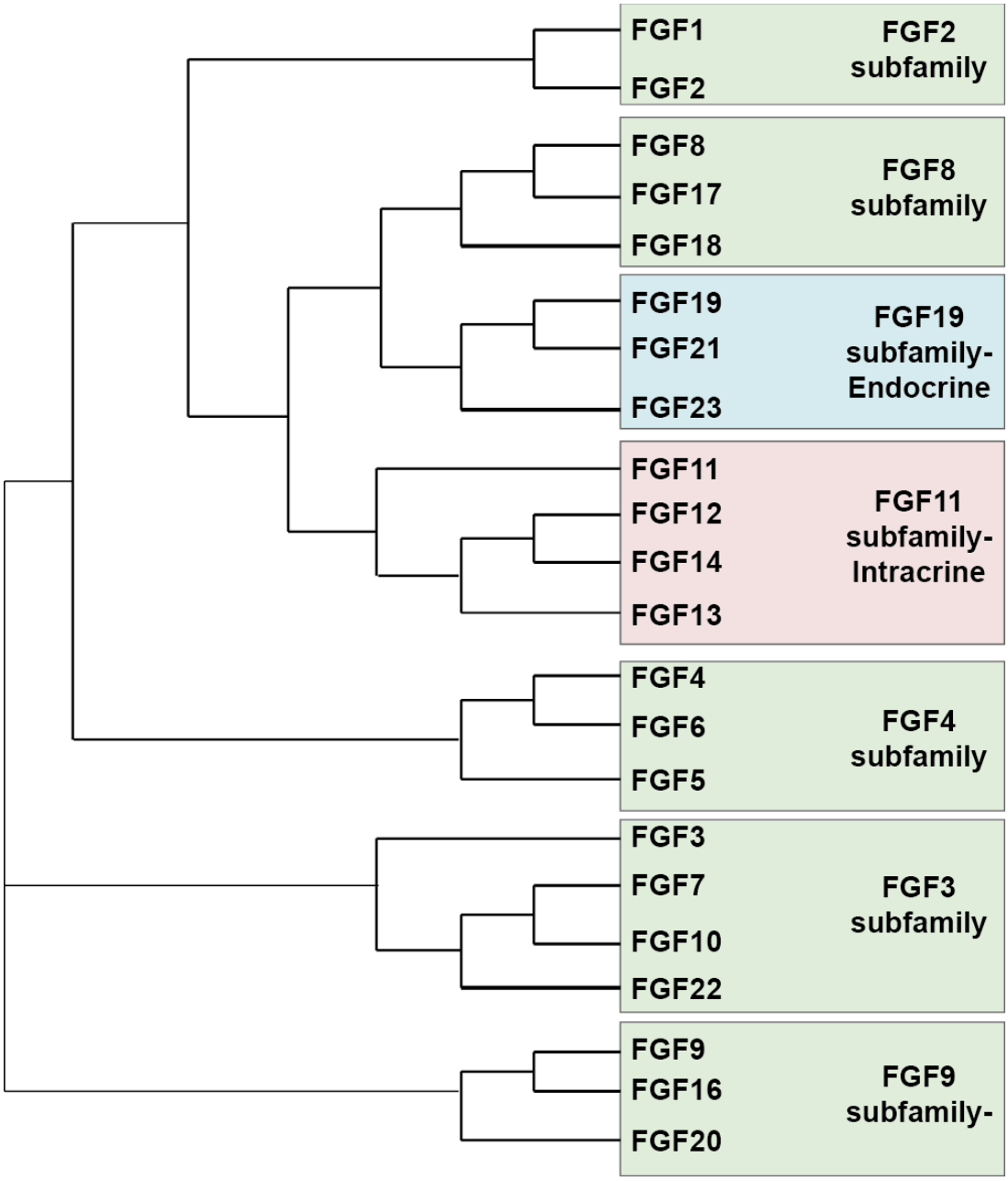


Figure 1.1: Phylogenetic tree of the paracrine FGFs. They can be arranged into five subfamily: FGF1, FGF4, FGF7, FGF8, FGF9. Branch lengths are proportinal to the evolutionary distance between each genes.

1.2.2 Genes encoding FGFs

In general, *fgf* genes comprise three to five coding exons, named as number 1, 2 to 5 but the introns between them are extremely varied from 4.9 kb (in *fgf3* and *fgf4*) to over 100 kb (in *fgf12*) [15]. Some *fgf* genes have unique properties. *fgf2* and *fgf3* have additional 5' transcribed region. In *fgf8*, exon 1 is subdivided into two, three or four alternatively spliced sub-exons, denoted as 1A, B, C, D [16]. When the coding region starts with exon 1, it initiates with methionine (coded as AUG), however, when it has 5' transcribed region, the start is upstream CUG codons [17]. For members of the *fgf8* family, AUG is always the initiation codon [16] . This gene organisation is conserved in mammals.

1.2.3 Evolution of FGFs

The fact that most of the *fgfs* in humans locate in paralogons (sets of paralogous chromosal regions) indicates that they follow the general scheme of evolution of the human genome, which included two large-scale duplications [1]. The consequence of the first series of duplication from a single or few *acheo-FGFs* or alternatively FGF ancestor genes generated eight proto-FGFs, but losses occurred later bringing the number of proto-FGFs to two in *protostomians*. This was followed by the second major expansion in early vertebrates. The number of *fgf* genes increased dramatically before gene losses happened again and resulted in the 22 present day *fgfs*.

The arguments for synteny are somewhat convoluted. Thus, it is argued that gene linkage will exist between members of each sub-family, whereas any such linkage between members of different FGF subfamilies arises by chance [1] [8]. For example, *fgf3, fgf4,* and *fgf19* are close in the q13.3 region of chromosome 11, and *fgf6* and *fgf23* are linked on chromosome 6. On the other hand, *fgf1* locates in chromosome 5, but *fgf2* is on chromosome 4. *fgf4, fgf5* and *fgf6* of

the FGF4 family are located in three different chromosomes, chromosome 11, 4 and 12, respectively.

Given the short length of the FGF domain and the divergence of the sequences between the different lineages, the phylogenetic distribution of *FGF* genes into eight subfamilies was employed to solve the orthology relationship between the different members of this family [9]. As a result, there are two major evolutionary scenarios. The first one suggests that the ancestral eumetazoan contained an *fgf* gene set of at least two genes, which are orthologs of *fgf1/2* and *fgf8/17/18/24*. Indeed, the present diversity of the *fgf* genes into eight subfamilies was generated from the important chordate-specific duplications [9]. The second hypothesis starts from the maximum (eight genes in the eumetazoan ancestor). It implies the presence of two *fgf* genes, orthologs of the *fgf9/16/20* and *fgf8/17/18/24* families in the ancestor of nematodes. Moreover, it highlighted a higher degree of gene loss during metazoan evolution including six gene losses in cnidarians, five in protostomes and five in ambulacrarians. However, one important aspect missing from these analyses is 'b' and 'c' FGFR isoform selectivity. In an extant nematode, *C. elegans*, there are just two *fgf*s and one *fgfr*; alternative splicing of the *fgfr* leads to selectivity for the two *fgf*s [9]. This is a recurrent theme in metazoan development: 'b' isoform *fgfr* is generally produced by epithelial tissues, and 'b' isoform selective *fgf* ligand by the neighbouring mesenchyme, whereas the opposite is true for 'c' isoform *fgfr* and 'c' isoform selective FGF ligand. This is an essential feature of mesenchyme-epithelial communication during development (Section 1.2.1). So what can be concluded from the above is the occurrence of events of gene duplication and gene loss at different times and in different evolutionary lineages during evolutionary history of FGF gene family.

1.2.4 Roles of paracrine FGFs in development

A number of functions of paracrine FGFs in development have been established by many studies on *Fgf* knockout mice. Most essential during embryonic developments are *Fgf4* and *Fgf8* genes when the mouse with these two genes knocked out all died [18] [19]. The germline mutation of *Fgf2* gene is related to a reduced number of cortical neurons in the adult. Indeed, during the middle to late stage of neurogenesis, the downregulation of *Fgf2* and *Fgfr1* has been observed [20]. *Fgf5* gene knockout results in abnormal long hair in mice [21] whereas *Fgf7* gene is involved in kidney development [22]. Related to the development of mesenchyme are *Fgf9, Fgf10* and *Fgf18,* with gene knock out leading to the death of mice shortly after birth. *Fgf-10*-deficient mice fail to initiate limb buds [23]. These mice also showed perinatal lethality due to complete lack of lungs. The disruption of *Fgf9* resulted in a reduced lung mesenchyme and decreased branching of airways in mice lungs, suggesting that *Fgf9* has impact on lung size through stimulating mesenchymal proliferation [24]. *Fgf18*-deficient mice exhibited several abnormal phenomena, including the delay of the progress of cranial suture closure, the decrease of the proliferation of calvarial osteogenic mesenchymal cells, and the postponement of terminal differentiation of these cells to calvarial osteoblasts. These observations suggested functions of FGF18 in the regulation of cell proliferation and differentiation in osteogenesis and in chondrogenesis [25]. The functions of *Fgf3* in the development of the ear and tail in mice have been addressed by gene targeting experiments [26]. In these, mice carrying a targeted insertion of a *neo*^{r} gene in the *int-2* (*Fgf-3*) did not often survive to adulthood. Mice with knockout of *Fgf16* were viable, but had a decrease in heart weight and cardiomyocyte cell numbers at 6 months of age [27]. Moreover, *Fgf1* knock out no phenotype in development, but did have in the adult if it had been stress with a high-fat diet, suggesting its potential role in nutrient homeostasis [28].

1.2.5 Structural features

Structural studies of FGFs identified from the outset are that they possess a β trefoil structure of 12 antiparallel β strands (120-130 amino acids) arranged as four-stranded β sheets arranged in a triangular array in the conserved core region (Figure 2) [29]. This folding pattern is similar to that of the interleukins IL-1β and IL-1α [30]. Some structure have β strand XI as an α-helix but others have very poor definition here, for example FGF 10 [31].

Even though 3-D structures of FGFs represented a high level of similarity across families, there are some undeniable differences. For example, FGF7 and FGF10 in the FGF7 subfamily [32] [31] have a longer β strand I (11 amino acids) than that of FGF1 and FGF2 (3 amino acids) [29] [33]. In addition, the length of loop between β strands I-II and β strands IX-X of FGF1 and FGF2 is much shorter than that of FGF9 [34], FGF7 and FGF10 [31].

The HS canonical binding sites, so called HBS-1, locates in the region between β10 and β12,which is rich in arginine and lysine [35][36] [37] [38].

Variations in loops and N-, C-termini

The function of the N-terminus, including N-terminal spliced regions of FGF8 and N-terminal alternatively translated extensions of FGF2 and FGF3 are poorly understood. There is very little structural information, as the N-terminus is often disordered. In FGF2, an NMR study suggests that the N-terminus of the protein translated form the AUG adopts two structures that are in equilibrium [39]. Higher-molecular-weight forms (22.5, 23.1, and 24.2 kDa) of FGF2 contain the colinear N-terminal extensions of up to 53 amino acids which then direct, post-translationally, the extended forms of FGF2 to the nucleus [40] [41]. Members of FGF1 subfamily, FGF1 and FGF2 do not have the signal sequence but because they are present in the cell surface and within ECM, hence they are secreted by a typical mechanism called alternative exocytotic that bypasses the endoplasmic reticulumn Golgi pathway [17]. In the FGF9 family,

the N-terminus is found to adopt an α helical structure in crystals and is thought to contribute to the dimerization of these FGFs and in this way regulates binding to the FGFR [42] [34]. The N-terminal spliced variants of FGF8 have been found to contribute to the interaction with the FGFR.

The structure and function of the C-terminus in the paracrine FGFs is also not understood. It has been demonstrated that in FGF1, the deletion of C-terminus caused a decrease in protein synthesis simulated by FGF1 on the cells [43]. Among the paracrine FGFs, FGF3 and FGF5 possess the longest C-tails with a high density of positive charge residues. In particular, dominant in FGF5's C-tail is lysine and for FGF3, is arginine. This indicates the potential for binding to a negatively charged partner such as HS. Members of the FGF8 sub-family (FGF8, FGF17 and FGF18) also have extended C-termini, which are relatively conserved and include 4 to 6 basic amino acids.

For the FGFs secreted conventionally via the endoplasmic reticulum/Golgi, the signal peptide is located at the N-terminus [6]. In most of these FGFs the secretory signal peptide is believed to be cleaved, as seen in FGFs 3-8 [6], FGF10, FGF17-19, 21 and 23 [44]. By contrast, some FGFs do not contain a cleavable N-terminal signal peptide. Thus, the signal peptide remains associated with the native protein, but the protein is secreted and is not membrane-associated. This is the case for the FGF9 subfamily (Fig. 1). For FGF22, its putative N-terminal signal peptide is cleaved, but then apparently remains attached to the cell surface [44].

Another difference between the structures of these FGFs is the conformation of the in N- and C-termini. The two termini in the FGF-1 sub-family are disordered in crystal structures to the extent that these proteins are expressed as truncated variants in order to obtain crystals. In contrast, in FGF-9, the N-terminal contains an α helix and the C-terminal of FGF-9 (17 amino

acids), which is much longer than that of FGF-1 (4 amino acids) and FGF-2 (3 amino acids) also contains a short helical structure

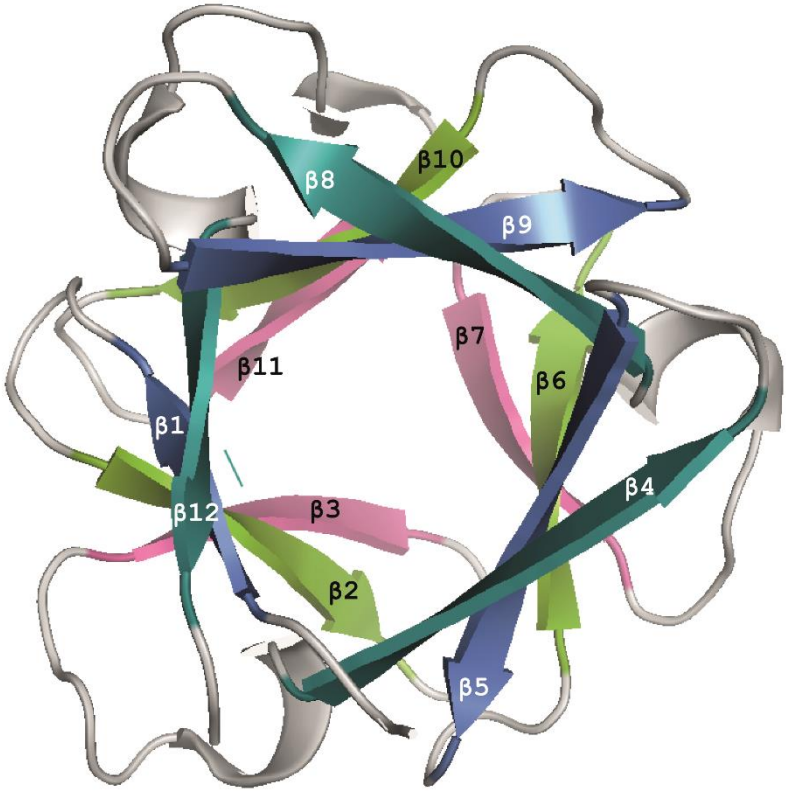


Figure 1.2: The structure of FGF2 illustrating the core beta trefoil. The typical beta trefoil (FGF2, PDB 1BFG [29]) has 12 antiparallel β strands (120-130 amino acids) arranged as four-stranded β sheets arranged in a triangular array.

1.3 PROTEOGLYCANS

Proteoglycans (PGs) are *O*-glycosylated proteins which are secreted into the ECM, inserted into the plasma membrane, or stored in secretory granules. Perlecan, glypicans, decorin, and syndecans are classic PGs. Proteoglycans comprises two elements, first, a core protein, second, a glycosaminoglycan (GAG) chain linked to the side chain of a serine side chain and in some instances to that of an asparagine. Glycosaminoglycans are linear polysaccharides mainly consisting of disaccharide building blocks of an amino sugar (*N*-acetyl-α-D-glucosamine or *N*-acetyl-β-D-galactosamine) and a uronic acid (β-D-glucuronic acid or α-L-iduronic acid) or

an amino sugar and galactose (Figure 1.3). The family of GAGs includes 4 members based on their disaccharide repeat units and biosynthetic pathway: chondroitin sulfate (CS) and dermatan sulfate (DS), hyaluronic acid (HA), keratan sulfate (KS), heparan sulfate (HS) and heparin [45].

1.3.1 Structural properties of GAGs

Hyaluronic acid (HA) consists of β-D-glucuronic acid linked to 2-acetamido-2-deoxy-β-D-glucopyranose (*N*-acetyl glucosamine) though alternating 1→3 and 1→4 linkages. It is the only GAG that is synthesized on the plasma membrane, not built on a core protein in the Golgi, and that does not contain sulfate groups (Figure 1.3). In the case of KS, the chains are linked to either Asn or Ser/Thr residues in core proteins. The main component of the chain is sulfated poly-*N*-acetyl-galactosamine (Figure 1.3).

The GAG chains of CS/DS and HS/heparin are attached to the serine residues in core proteins. The building block of CS/DS is a disaccharide of sulfate-substituted *N*-acetyl galactosamine (GalNAc)-GlcA polymerized into long chains. Owing to the common disaccharide unit, DS was known as chondroitin sulfate B, however, it is distinguished from chondroitin sulfates-A (4-*O*-sulfated) and -C (6-*O*-sulfated) by the presence of iduronic acid (IdoA), which is formed post-polymerisation by epimerisation of the GlcA (Figure 1.3).

HS is composed of disaccharide repeating units comprising 1,4 linked β-D-glucuronate, (GlcA) and N-acetyl-α-D-glucosamine (GlcNAc) [46] with diverse modifications including sulfation at 2-*O*-, 3-*O*-, 6-*O*- and *N* positions and epimerisation of GlcA to L-iduronic acid (IdoA) (Figure 1.3) [47].

Heparin shares with HS the disaccharide building block 1,4 linked β-D-glucuronate, (GlcA) and α-D-N-acetyl glucosamine (GlcN)NAc, however, it has much more extensive levels of modifications.

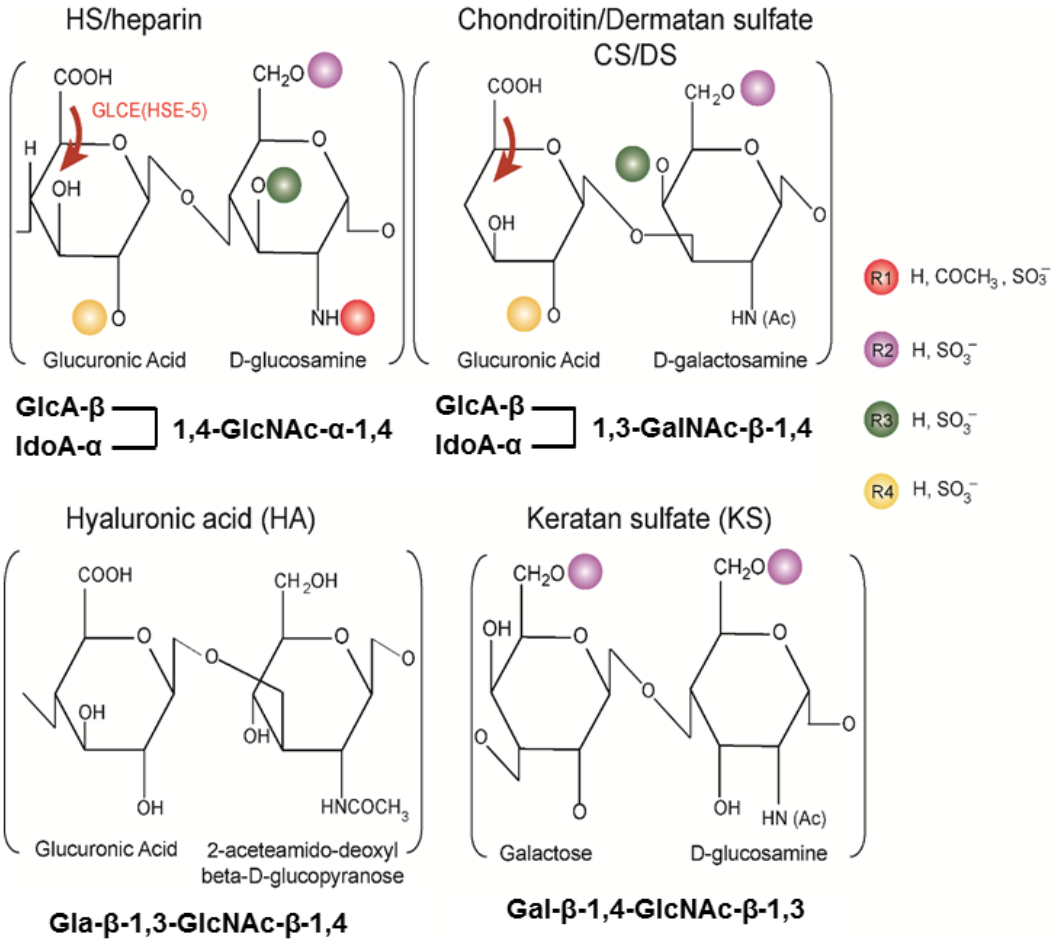


Figure 1.3: Disaccharide repeating unit of GAG members. For HS and heparin, the repeating disaccharide unit contains: a D-glucuronic acid or L-iduronic acid and a D-glucosamine. For CS and DS, the repeating disaccharide unit contains: a D-glucuronic acid or L-iduronic acid and D-galactosamine. HA is composed of the repeating units of a D-glucuronic acid and 2-aceteamido-deoxyl- β-D-glucopyranose. For KS, the repeating disaccharide unit contains: a galactose and a D-glucosamine.

1.3.2 Hierarchical biosynthesis of GAGs

The biosynthesis of two classes of GAGs (CS/DS and HS/heparin) have a common chain initiation biosynthesis pathway. The discrimination between them occurs at the last step of chain initiation, which then leads to two distinct pathways of chain elongation and modification.

1.3.2.1 Chain initiation in HS and CS/DS synthesis

CS/DS/HS/heparin are synthesized as covalent complexes, in which, the first step is the assembly of a linker tetrasaccharide attached to core protein of proteoglycans (PGs) at a specific Ser-Gly motif flanked by at least two acidic amino acid residues. The linker tetrasaccharide β Xyl-β GlcA-1,3-β Gal-1,3-β Gal (Figure 1.4) is formed by the attachment of a xylose (Xyl) by Xylosyltransferase to the serine side chain, followed by the addition of two galactose (Gal) residues by galactosyltransferases I and II (GalTI and GalTII), and then finally by the addition of glucuronic acid (GlcA) by glucuronosyltransferase I (GlcATI) linkage [45][48]. The subcellular localization of each step in chain initiation is not identical. While the attachment of xylose to the core protein is believed to occur in the endoplasmic reticulum (ER), the further assembly of the linkage region and the other parts of the chain has been shown to be in the Golgi apparatus.

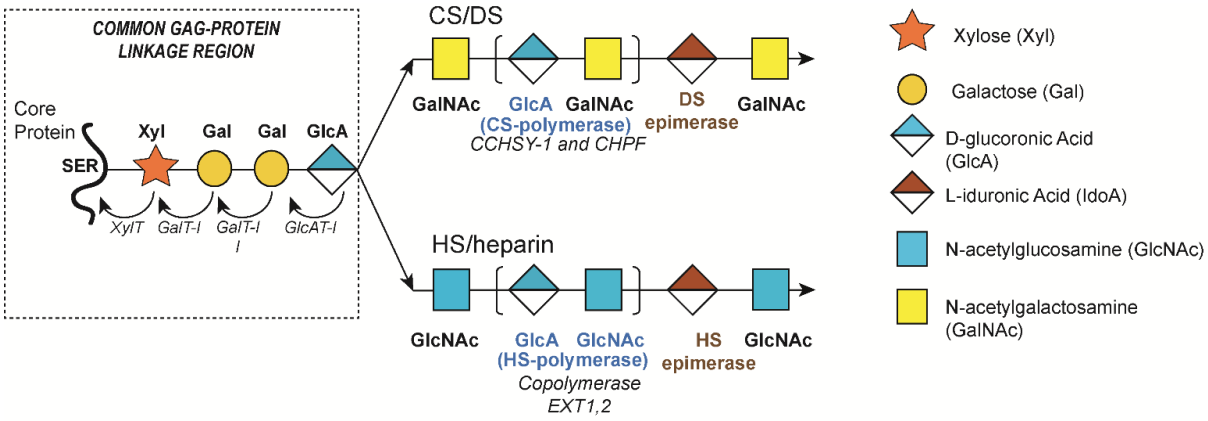


Figure 1.4: Chain initiation of HS/CS/DS in synthesis. The synthesis of HS, CS and DS shares the initiation step of synthesis the GAG-core protein linkage region. They start to branch out from the step of chain elongation.

1.3.2.2 Next stages in the biosynthesis of HS

1.3.2.2.1 Chain elongation (chain polymerization)

The elongation step is where the biosynthesis of heparan sulfate/heparin and chondroitin sulfate/dermatan sulfate diverge (Figure 1.4), which occurs by the attachment of a *N*-acetylglucosamine (GlcNAc) residue by O-linked N-acetylglucosamine (O-GlcNAc) transferase to direct the synthesis of a chain of heparan sulfate/heparin or a *N*-acetylgalactosamine (GalNAc) residue by *N*-acetylgalactosaminyl transferase (chondroitin sulfate/dermatan sulfate), which results in the synthesis of a chain of CS/DS (Figure 1.4). There follows the stepwise addition of GlcA and GlcNAc residues in the case of heparan sulfate/heparin and for CS/DS the addition of GlcA and GalNAc residues.

1.3.2.2.2 Chain modification

The diversity in arrangement of post-polymerization modifications along the HS and hepairn chain leads to substantial structural variability (Figure 1.5) and is mediated by four classes of sulfotransferases and an epimerase.

The next stage, is catalysed by the GlcA C5 epimerase (heparosan-*N*-sulfate-glucuronate 5-epimerase), which epimerises β-D-GlcA to α-L-IdoA. For substrate recognition, the GlcNS residue has to be linked to the non-reducing side of a potential GlcA [49]. This epimerase (Glce) required in HS and heparin synthesis is different from the one involved in epimerisation of dermatan sulfate, which is Dse1-2. Some IdoA units are then sulfated at C2 by HS2ST1. The occasional addition of 2-*O*-sulfate groups to GlcA blocks the epimerization reaction.

O-sulfation can occur at positions 3 and 6 of the glucosamine [50]. There are three glucosaminyl 6-*O*-transferases (6OSTs) that act on GlcNAc and GlcNS. The 6-*O* sulfation of GlcNS blocks 2-*O* sulfation of a neighbouring IdoA. A number of 3-*O*-sulfated disaccharides are produced by the seven glucosaminyl 3-*O*-sulfotransferases (3OSTs) [51]–[53], though the

products of these enzymes have not been dissected in depth: GlcA-GlcNS(3S±6S) is produced by 3OST1 and 3OST5, IdoA(2S)-GlcNH$_2$(3S±6S) is produced by 3OST3a, 3OST3b, 3OST5 and 3OST6, and 3OST2 and 3OST4 produce GlcA/IdoA(2S)-GlcNS(3S). Despite being the largest family of HS modification enzymes, 3-*O*-sulfation is observed less often in HS.

The N-deacetylation/*N*-sulfation of GlcNAc residues to form GlcNS is commonly the first reaction, as there is a strong dependence of subsequent modifications on GlcNS. It is catalyzed by members of a family of four GlcNAc N-deacetylase/N-sulfotransferase enzymes (NDSTs). A study on mice where the NDST2 gene was knocked out showed that the mast cells are abnormal and lack heparin, although HS from other tissues appeared unaffected, indicating that this isoform is responsible for the synthesis of heparin [54]. On the other hand, the N-sulfation of HS is mainly performed by NDST-1 [55]. These observations lead to the conclusion that the regulation of enzyme activities at the level of transcription, translational and activity through posttranslational modification is likely to contribute to the formation of different patterns of N-acetylation.

1.3.2.2.3 Major and minor pathways for HS

The pathway of HS or heparin biosynthesis discussed above is widely accepted, but the sequence of reactions, starting with an NDST is unable to explain all the disaccharides found in HS/heparin, for example IdoAGlcNAc,6S [56]. Therefore, it has been proposed [56] that there are two branches of HS biosynthesis. One, described above has been called the major branch, since it accounts for the majority of disaccharides found in the polysaccharide. The other, called the minor branch is responsible for the reminder of the modification patterns observed. This bifurcation of the synthesis has been ascribed to the preference of the HS epimerase in converting GlcA-GlcNS to IdoA-GlcNS efficiently, but also GlcA-GlcNAc to IdoA-GlcNAc inefficiently (Figure 1.5).

An important feature of HS biosynthesis is that the hierarchical dependence of modifications of the major pathway on *N*-sulfation and directionality imposed by, for example, 6-*O* sulfation of GlcNS

preventing subsequent 2-*O* sulfation of IdoA, coupled to the partial processivity (apparent if not mechanistic) of NDSTs leads to a domain structure. Thus, the NDSTs tend to modify shorter or longer (as in the case of NDST1) tracts of disaccharide units. Before and after these segments they 'slip', modifying only around one GlcNAc in two and in between they appear to 'pause' not modifying any substrate. The result is that HS chains have NA domains, where every disaccharide has a GlcNAc residue, transition domains, where around one in two disaccharides contain GlcNS and sulfated domains where every disaccharide contains GlcNS in minor pathway [57] (Figure 1.5) [56].

1.3.2.2.4 Degradation of surface-bound and extracellular HS

The 6-*O*-sulfate groups in HS in the extracellular space can be removed by the extracellular sulfatases family (Sulfs), SULF1 and SULF2 which are HS-specific 6-*O*-endosulfatases [58]. Across the family, there are evolutionarily conserved amino acid regions of hydrophilic domains (HD) which could anchor Sulf to the cell surface, bind to HS substrates, and mediate HS 6-*O*-endosulfatase enzymatic activity. Degradation of 6-*O*-sulfation of cell surface HS chains promotes Wnt signaling and inhibits growth factor signaling in embryonic tissue patterning [59], so this is an important aspect of functional regulation.

In mammals, extracellular HS may be processed by the secreted enzyme heparanase, an endo-glucuronidase, which cleave the HS chains into smaller and potentially bioactive HS oligosaccharides [60]. This includes the formation of heparin (10–20 kDa) from the newly synthesized heparin chains (60–100 kDa) on the serglycin core protein [60].

HS chains may also be released from the cell surface by shedding. This occurs through the action of matrix metalloproteinases cleaving, for example, the core protein of Syndecans [61] or by the action of the notum protein and phospholipases in the case of the lipid-anchored glypicans [62] [63].

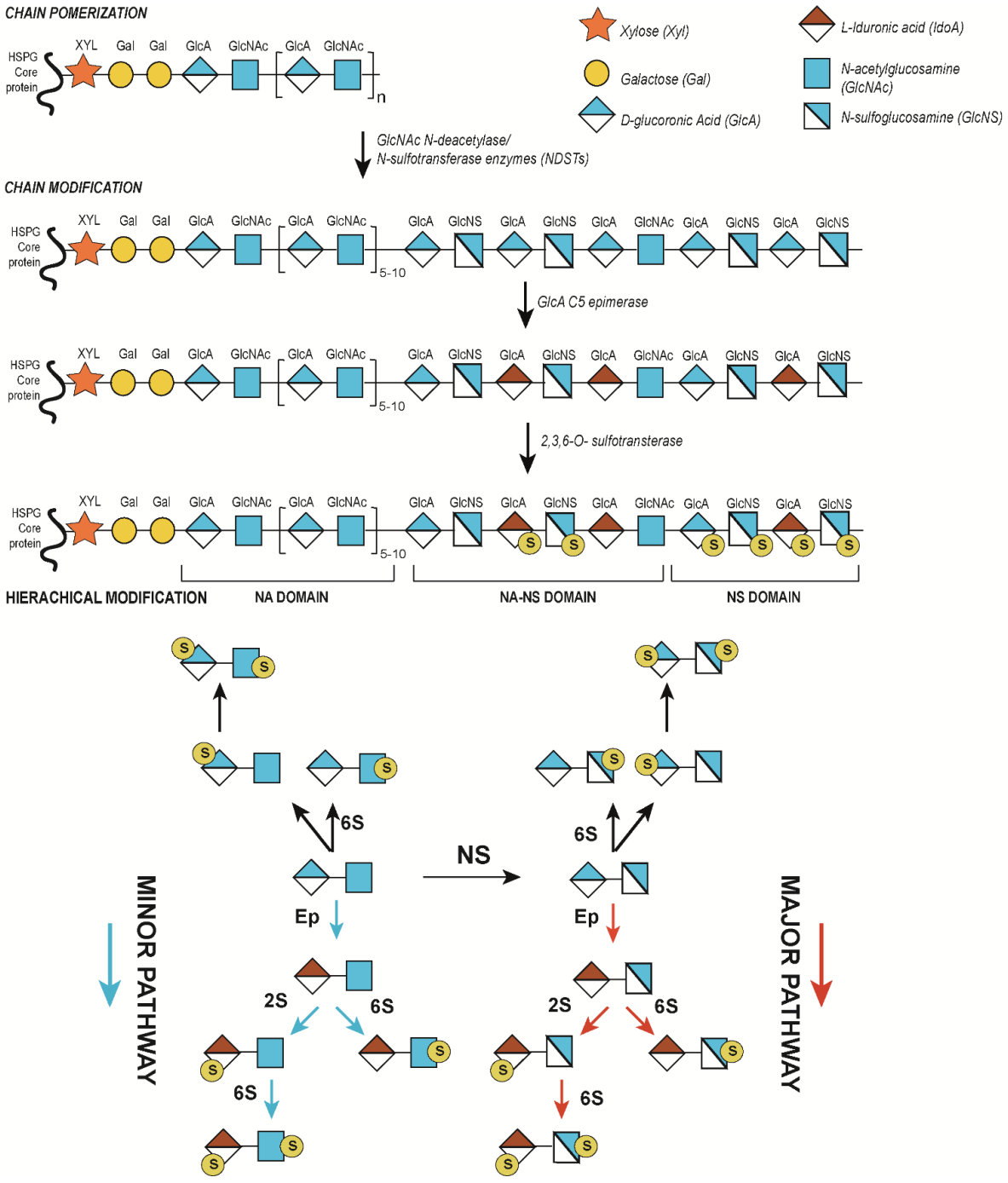


Figure 1.5: Synthesis of HS. Major and minor pathways of HS biosynthesis (after [56]). The upper panel presented the common synthesis of HS. Whereas, the lower panel showed the conceptual modifications happening during the synthesis using major pathway (right) or minor pathway (left).

1.3.2.3 Next stages in the bioynthesis of CS/DS

1.3.2.3.1 CS- Chain elongation (chain polymerization)

The polymerization of CS and DS is catalyzed by one or more bifunctional enzymes (chondroitin synthases- ChSy) that have β3- glucuronyltransferase-II (GlcATII) and β4-*N*-acetylgalactosaminyltransferase-II (GalNAcTII) activities to add individual sugars stepwise to the nonreducing end of the growing chain. This results in a chain of up to 70 kDa. Chondroitin polymerization also requires the concomitant expression of chondroitin polymerizing factor (ChPF), a protein that lacks catalytic activity, but acts as a specific activating factor for ChSy to promote the formation of polymers over oligomers [64].

1.3.2.3.2 CS- Chain modification

Sulfation of CS/DS in vertebrates is variable and along with epimerisation of GlcA, defines the name given to the polymer; the sulfation may occur on GalNAc at position 4- or 6-, or both, in the same glycosaminoglycan chains of CS with some GlcA 2- or 3-*O*-sulfated. DS often has GalNAc 4-*O*-sulfate and IdoA 2-*O*-sulfate along with some nonsulfated GalNAc and some rare GalNAc 6-*O*-sulfate [48]. Unlike HS, CS/DS chain shows no domain structure, but long tracts of fully modified disaccharides.

The modification patterns of a disaccharide unit composing CS has been characterised as letter, A, B (DS), C, D, E, H and K (Figure 1.6 and detail was provided in Table 1). The ordinary mature CS, designed CS-A, CS-C and hybrid CS chains are composed of monosulfated disaccharide units A and C (Table 1) in various proportions. The peculiar disulfated disaccharide units named D, E, H and K units depicted in Table 1 are components of oversulfated CS chains [65]. The H unit has also been named as iE, in which the abbreviation ‘i’ in ‘iE’ stands for IdoUA [66]. Interestingly, the K unit possesses 3-*O*-sulfation on GlcA (Table 1).

While many IdoA residues of DS possess 2-*O*-sulfation, this sulfation is relatively rare in GlcA of CS. The activity of 2-sulfotransferase on IdoA is likely dependent on the sulfation of the adjacent GalNAc. Thus, IdoA adjacent to nonsulfated GalNAc is generally not sulfated, the sulfation is slightly increased with 6-*O*-sulfated GalNAc, and it is strongest when IdoA is adjacent to 4-*O*-sulated GalNAc [67] [48] .

In mammals, there is no report of 3-*O*-sulfation in CS or DS, however, this has been found in king crab cartilage [68]. It is also clear that no *N*-sulfate was observed in these two GAGs.

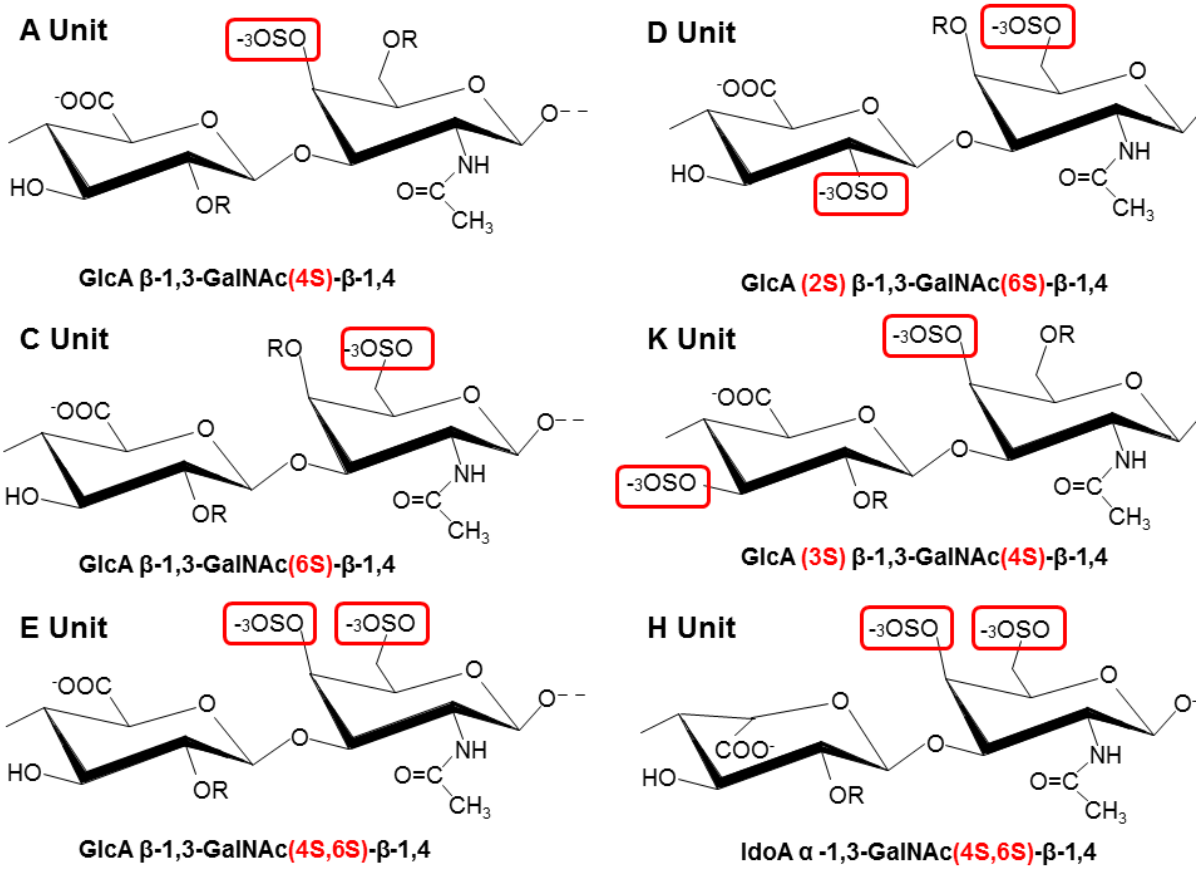


Figure 1.6: Structures of 6 defined CS disaccharide units (After [65]).

Table 1.1: Different CS disaccharide units. The sulfation of disaccharide units and epimerisation of GlcA is used to classify CS. CS-B is an old name for DS. The occurrence of a single more complex disaccharide, in the context of a simpler disaccharide, is sufficient to define the chain. Thus a single IdoA in a chain results in it being called DS.

Chain type	**Site of sulfation**
CSA	Carbon 4 of the N-acetylgalactosamine (GalNAc) sugar
CSB (DS)	Carbon 2 of the iduronic acid and carbon 4 of the GalNAc
CSC	Carbon 6 of GalNAc
CSD	Carbon 2 of the glucuronic acid and 6 of the GalNAc
CSE	Carbon 4 and 6 of the GalNAc
CS K	Carbon 3 of the glucuronic acid and 6 of the GalNAc
CS H	The iduronic acid and Carbon 4 and 6 of the GalNAc

1.3.2.3.3 Epimerization of DS

DS is derived from CS by the epimerization of GlcA to IdoA. This event can only occur at a low level following polymerisation, since the non-sulfated GAG is a poor substrate for the 2-*O* sulfotransferase, however, following 4-*O*-sulfation, the epimerization is significantly accelerated [69]. For example, a study on human skin fibroblasts showed that CS/DS had all IdoA adjacent to GalNAc-4S and only GlcA adjacent to nonsulfated GalNAc residues [70].

Chain termination

It seems that there is no specific saccharide terminating the synthesis of CS/DS. Some studies proposed that GalNAc 4-*O*-sulfate at the nonreducing end may be a signal for termination [71], however, this hypothesis is challenged by the study on cartilage microsomal system, which found poor 4-sulfation and no 6-sulfation of terminal GalNAc [72].

In the case of HS, the nonreducing end is often, but not always, occupied by NS domains. There is an observation that when a short segment of *N*-sulfated disaccharides dominates the nonreducing end, the further synthesis of the chain can be blocked [73], though as for CS there is no clear rule for chain termination.

1.3.2.3.4 The Interplay between HS and CS Biosynthesis

A relationship between HS and CS biosynthesis has been indicated by a number of studies. For example, non-functional *uxs1* (UDP-glucuronic acid decarboxylase 1) and *b3gat3* (Galactosylgalactosylxylosylprotein 3-beta-glucuronosyltransferase 3) mutants which were $uxs1^{hi954}$ and $b3gat3^{hi307}$, respectively in zebrafish impaired biosynthesis of both HS and CS, but on different levels. The larvae with these mutants failed to synthesize CS, but still produced 50% of the wild-type level of HS [74]. In sprouting angiogenesis using mouse embryonic stem cells lacking EXT1, the embryoid bodies showed a complete loss of HS production and an increase in CS biosynthesis [75]. In these and other studies there is, however, no clear answer as to how and why the synthesis of CS and HS are related.

1.3.3 Structural comparison between HS, heparin, and CS and DS

1.3.3.1 Consequence of biosynthetic pathways differences between heparin and HS

Heparin and HS share structural similarities in terms of the composition of an amino hexose and uronic acid, either GlcA or IdoA, joined by (1→ 4) glycosidic linkages.

1.3.3.1.1 Biosynthetic pathways differences between heparin and HS

During synthesis, due to the extensive tracts of GlcNAc acted on by NDST1, the uronic acid epimerization and sulfation on heparin are more extensive than that on HS. Moreover, heparin is excised from the full heparin chain by heparanase, a β-D glucuronidase (Section 1.3.2.2). Since GlcA is more common adjacent to GlcNAc, this will increase the sulfation of the ~17 kDa heparin product. Hence, in terms of composition, heparin contains mainly α-L-iduronate (over 70%), whereas in HS, β-D-glucuronate is predominant. While most of the D-glucosamine residues in HS are *N*-acetylated, the majority are *N*-sulfated in heparin. After epimerization, most of the newly formed iduronate residues and *N*-sulfated glucosamine residues in heparin are sulfated at C-2 and C-6. However, the level of sulfation in HS in regions where GlcNS is present is lower, and since NDST2-4 sulfates shorter tracts that NDST1, there are frequent segments of contiguous GlcNAc containing disaccharides [76]. These differences result in the degree of sulfation of 2.3–2.8 sulfates/disaccharide in heparin, but only 0.6–1.5 sulfates per disaccharide in HS [77]. These fundamental differences in composition lead to distinct conformations such as glycosidic linkage geometry (torsion angles, ϕ and ψ) and pyranose ring conformational equilibria between these closely related GAGs. In general, HS chains, as well as being longer, are more flexible than heparin.

1.3.3.1.2 Domain structure of HS and evidence for variation in S & NA & NS domains

Owing to NDSTs acting on tracts of GlcNAc containing disaccharides and the dependence on GlcNS of subsequent modifications in the major pathway of HS biosynthesis (Fig. 1.5, Section 1.3.2.2), HS exhibits a domain structure. Thus, there are stretches of low/non sulfated

disaccharide repeats, rich in GlcA-GlcNAc (termed NA-domains); shorter, highly sulfated regions where every glucosamine is *N*-sulfated glucosamine and L-iduronate is common (termed S-domains); and intervening mixed regions with alternating GlcNAc and GlcNS disaccharides (termed NA/NS domains) (Figure 1.5) [78]. The definition of these domains has been obtained experimentally through the actions of heparinases I, II, III and K5 lyase and the latter enzyme requires an octasaccharide of GlcA-GlcNAc [78]. NA/NS domains can contain up to three GlcA-GlcNAc disaccharides between GlcA-GlcNS disaccharides. In contrast, heparin is characterized as a more or less continuous NS-domain due to the lack of alternating NA- and NS-domains, though it does contain short stretches of NA/NS domains.

1.3.3.1.3 Critique of using heparin as a proxy for an S-domain.

Owing to the commercial availability of heparin and heparin affinity chromatography matrices, heparin is widely used as a proxy for HS S-domains, however, there are drawbacks to this approach, because HS and heparin are different in their level of sulfation and chain domain structure (Section 1.3.3.1.2). First, the use of heparin to determine the sugar binding preferences for interacting proteins may bias the data towards ionic interactions, because of the high sulfation of heparin. Indeed, there are examples of where relatively low sulfated HS-derived oligosaccharides have similar interacting properties as heparin [79]. Besides ionic bonds, there are a variety of different interactions. Thus, hydrogen bonds can also make significant contributions to the engagement of HS by proteins. For instance, ionic bonds only contribute 6 % to the binding free energy between brain natriuretic peptide (BNP) and heparin, whereas 94 % of the free energy comes from the hydrogen bonding between polar amino acids of BNP and heparin [80]. There are also data indicating contributions from other bonds including van der Waal's bonding, e.g., between FGF2 and heparin [81], but it is not clear whether these are from protein-sugar interactions or from protein conformational change. There is, as yet, no example of a protein binding to NA domains, though this may be due to the

use of heparin affinity columns. However, GlcNAc-containing disaccharides and small NA domains have one function, which is to act as flexible spacer, which allow flanking NA/S or S domains to interact with bipartite binding sites on proteins that are at 90° to 180° to each other on the protein surface. This is the case for interferon γ [82] and vascular endothelial growth factor [83].

Another issue is that the variation of sulfation is likely the key factor for the specificity of HS, which cannot be reflected by heparin [84] [37] [85]. For example, the studies on the modifications required for FGF2 binding with chemically modified heparin or native HS preparations demonstrated that *N*-sulfate on GlcNAS and 2-*O*-sulfate of IdoA 2 are essential for binding, but not the 6-*O*-sulfate of GlcNS [86].

Compared to heparin, HS is more ubiquitous, whereas heparin is produced by connective-tissue type mast cells [87] and almost all cells expresses HS. It is obvious that HS rather than heparin is the physiological binding partner for proteins. However, there are polysaccharides isolated from non-mammalian organisms that seem to be hybrids of HS/heparin [88]. For example, a heparin/HS-like polysaccharide was isolated from the heads of *Litopenaeus vannamei* shrimp. It has the high degree of *N*- and 6-*O*-sulfation and low levels of *N*-acetylation of glucosamine, like heparin but is similar to HS in its glucuronic acid content [89]. Often detected in this context is IdoA-2S.

There are further structural differences between HS and heparin. Heparin chains (~17.5 kDa) are much shorter than HS, chains (25 kDa- 100 kDa), because they are excised from the heparin chains on serglycin [90]. Moreover, the presence of GlcNAc and NA domains in HS, as noted above, makes HS chains more flexible around the glycosidic linkages (torsion angles ϕ and ψ, Fig 1.6) than heparin, but heparin is more conformationally plastic, due to its proportionally higher IdoA content and this residue is uniquely in equilibrium between two conformers, $^{1}C_{4}$ and $^{2}S_{0}$ (Fig 1.6) [91]. ϕ is the dihedral angle defined by H1–C1–O–C4 and ψ is the dihedral

angle defined by C1–O–C4–H4. To study the permissible ϕ and ψ angles of various disaccharides, molecular-modelling has been used to examine the glycosidic bonds of GlcNS-IdoA, GlcNAc- IdoA, IdoA-GlcNS and IdoA-GlcNAc where the GlcN(S) is in the $^4C_1$ conformation, whereas the IdoA is in both the $^1C_4$ and the $^2S_0$ conformations [92]. The results show that IdoA-(1,4)-GlcN glycosidic bonds have more freedom that GlcN-(1,4)-IdoA glycosidic bonds, while the other disaccharides have ϕ and ψ angles that are quite limited due to steric interactions. Indeed, the $^1C_4$ and the $^2S_0$ conformations adopt different ϕ and ψ angles. Moreover, protein binding can have a major affect on the sugar conformation. For example as measured by ^{1}H-NMR spectroscopy when bound to FGF2, the ϕ and ψ angles of ΔUA(2S)-GlcNS(6S) and GlcNS(6S)-IdoA(2S) –are significantly different compared to the free form of the polysaccharide, whereas the ϕ and ψ angles of GlcNS(6S)-IdoA(2S)-GlcNS(6S) did not change compared to the free sugar [93]. A key point made by such work is that polysaccharide conformation can play an important role in protein interactions and that the sugar's pendant sulfate groups, while important, are insufficient on their own to explain binding.

1.3.3.2 Difference in spatial position of sulfate and carboxylate in CS/DS

Owing to the dimer repeat, the CS glycosidic backbone has two distinct (ϕ1, ψ1) for 1-3 glycosidic linkage and (ϕ2, ψ3) for 1-4 glycosidic linkage (Figure 1.7). The charged group combination is a key factor for the conformational preferences. A study comparing the disaccharides β-D-Gal-(1 → 4)-β-GlcNAc, 6S and β-D-Gal-(1 → 4)-β-GlcNAc, 4S, 6S showed a slight repulsive effect between the 6-*O*-sulfate groups and the carboxylate group and a small change of the preferred population of the glycosidic linkage. In contrast, the sulfate group at C-4 of β-D-GlcA-(1→3)-β-D-Gal disaccharides and the carboxylate group had an attractive interaction that led to a change of the conformation of the glycosidic linkage by about 30°. When both 4-*O* and 6-*O* were sulfated, the rotamer population of the C-6 groups can be observed [94].

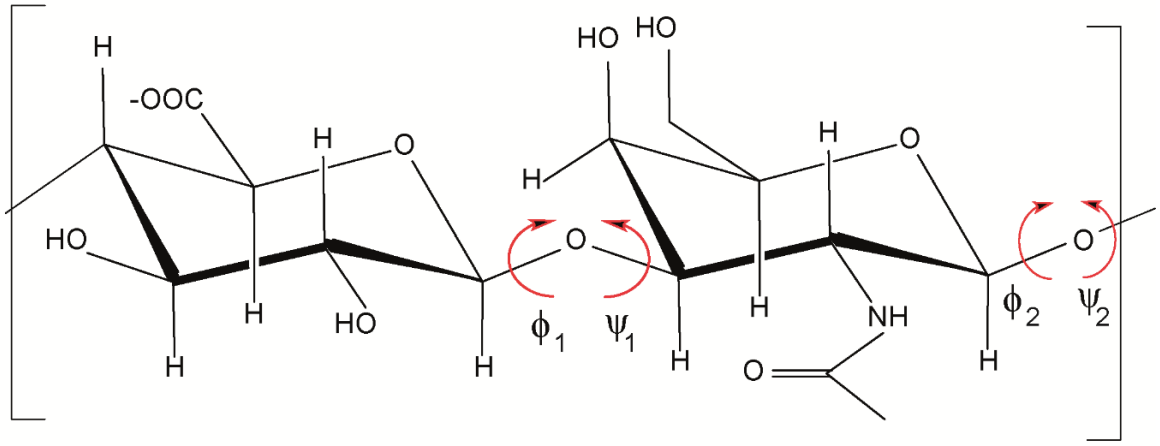


Figure 1.7: The two distinct degrees of disaccharide [(-4GlcAβ1-3GalNAcβ1-)] of CS backbone conformational freedom (ϕ_1, ψ_1) and (ϕ_2, ψ_2) [95]. For O1 which is the 1 → 3 linkage, ϕ1 = *O*-ring-1-*O*-glycosidic-3 and ψ1 = 1- *O*-glycosidic-3-4 ("(ϕ1, ψ1)"), and for O2 which is the 1 → 4 linkage, ϕ2 = *O*-ring-1-*O*-glycosidic-4 and ψ2 = 1- *O*-glycosidic-4-5 ("(ϕ2, ψ2)").

In disaccharides with the β-(1→4)-linkage and the 4-*O* and 6-*O* sulfations, the sulfate group(s) and the carboxylate group were on opposite sides, so that the molecules could adopt a conformation similar to those described for uncharged β-(1→4)-linked disaccharides [94]. In contrast, the sulfate groups and the carboxylate groups were on the same side the β-(1→3)-linked disaccharides [94].

1.3.3.3 Chain flexibility

It has been argued that heparin has unusual mobility (in the iduronate ring) and considerable rigidity (in the glycosidic conformations), in part because the iduronate pyranose rings can adopt at least two conformations depending on pattern of sulfations without altering the overall shape of the heparin chain [96]. However, co-crystal structures of protein-oligosaccharide complexes suggest that the $^{2}S_{0}$ of IdoA does alter chain direction, but up to 90° over a trisaccharide, e.g. 1AXM [97]. NMR analysis of a single sugar within heparin observed

unsubsituted IdoA as either $^{1}C_{4}$ or $^{4}C_{1}$ chair form. For IdoA-2-sulfate (IdoA-2S), $^{1}C_{4}$ and $^{2}S_{0}$ co-existed (Figure 1.8). In a combination of IdoA-2S and preceding a 3-*O*-sulfated glucosamine, $^{2}S_{0}$ skew boats dominated, whereas when the IdoA 2S was at the non-reducing direction at the end of the chain, this equilibrium changed towards $^{1}C_{4}$ form. In the $^{4}C_{1}$ form, there are 1,3 diaxial non-bonded interactions, where four of the substituents are axially oriented and only the carboxylate group is equatorial. Hence, this conformer is unfavourable for the stability of the chains [98] [93]. One chain may show multiple conformation of the iduronate pyranose rings, for example, a mixture of $^{1}C_{4}$ and $^{2}S_{0}$ form of the central iduronate residues were defined when polydisperse heparin interacts with foot and mouth virus (PDB code 1QQP) [99] [98], while in a co-crystal structure of FGF2 and a heparin-derived hexasaccharide, there is an IdoA residue in the $^{2}S_{0}$ confromation, alongside two in $^{1}C_{4}$ [100].

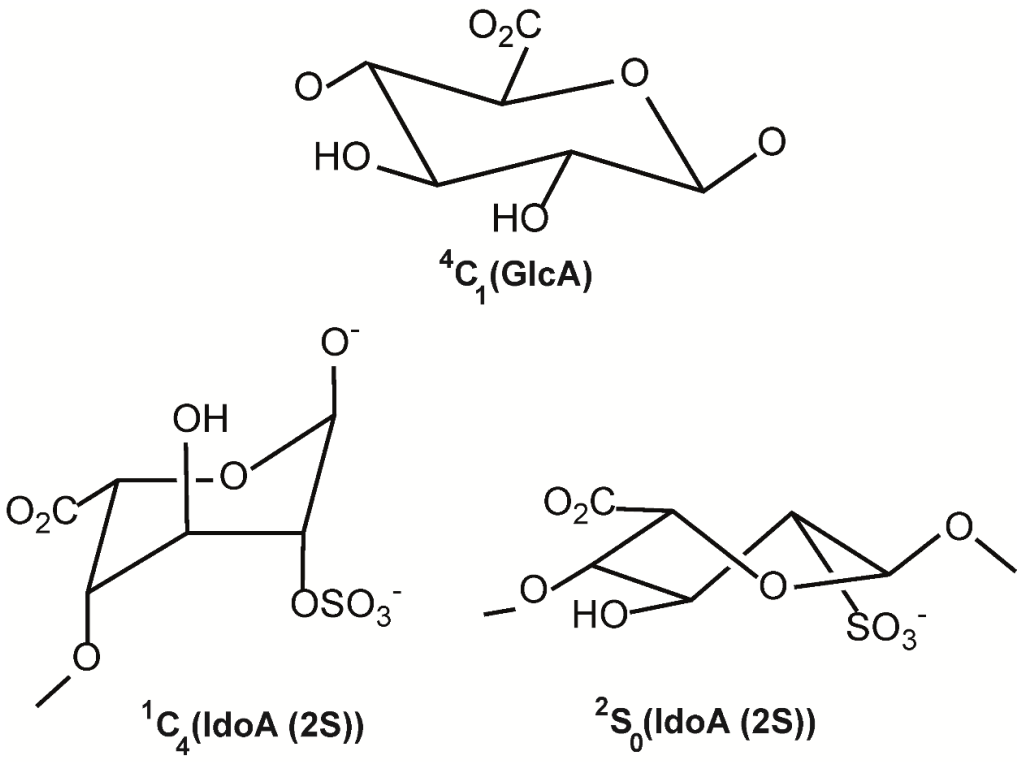


Figure 1.8: Conformation of β-D glucuronate acid and α-L idoronate acid. Pyranose ring of each glucuronic acid could adopt three different conformation, which are $^{1}C_{4,}$ $^{4}C_{1}$ and $^{2}S_{0}$.

In C4S and C6S the glucuronic acid and galactosamine both adopt the $^{4}C_{1}$ chair conformation. The conformation of IdoA in DS has been unresolved for a long time due to two conflicting pieces of evidence provided from the studies of IdoA pyranose ring conformations in the solid state and in solution. While the $^{4}C_{1}$ and $^{1}C_{4}$ conformations were incompatible with the rise per residues and the number of residues per turn, the $^{2}S_{0}$ conformation gave a left-handed helix similar to those of the other GAG solid-state structures. More recent NMR data has indicated a distorted $^{1}C_{4}$ chair, evidenced by the preference of oxidized DS toward vicinal equatorial hydroxyls, as found in a $^{4}C_{1}$ chair. In equilibrium, the mixture of $^{1}C_{4}$ and $^{2}S_{0}$ with small proportions of $^{4}C_{1}$ was proposed by the examination of NMR proton-proton coupling constants [101].

1.4 HEPARIN/HEPARAN SULFATE AND PROTEIN INTERACTIONS

1.4.1 General overview of heparin binding proteins

A large number of proteins, including cytokines and chemokines, enzymes and enzyme inhibitors, extracellular matrix proteins, and membrane receptors are classified as heparin-binding proteins (HBPs), since they have been shown experimentally to bind this polysaccharide. The interactions between HS and HBPs regulate many of the key functions associated with multicellularity, such as cell migration, cell differentiation, morphogenesis, organogenesis, and contribute to many disease states such as inflammatory conditions.

1.4.2 How do heparin/HS recognize a protein?

NS or NA/NS domain are the protein binding structures of HS chains for the reason that the interaction between HS and HBP depends on the sulfate groups and on the conformational versatility of the iduronate residues. For instance, the HS binding FGF-2 requires IdoA2S and GlcNS [102]. The IdoA2S residues in a co-crystal of FGF2 (PDB code 1BFC [100]) with a heparin-derived hexasaccharide are in two conformations, while the third iduronate ring adopts a $^{1}C_{4}$ chair conformation, the fifth ring adopts a $^{2}S_{0}$ skew boat conformation [100]. In a

different arrangement, the crystal structure of a heparin tetrasaccharide with annexin V, shows that the IdoA2S which interacts with the protein is in a $^{2}S_{0}$ conformation, whereas IdoA2S in the 1C4 conformation does not interact with annexin V [103]. Thus, the ability of IdoA to exchange between two conformations enables the polysaccharide to fit different protein surfaces. This conformation flexibility in the modified area may explain the reason many proteins, e.g., FGF3, bind to DS, but not to CS, despite the similar charge density of these polysaccharides [37].

The conformation and dynamics of a heparin pentasaccharide, representing a binding site of heparin for antithrombin III, have been investigated both in the crystal [104] and solution states [105]. Both show that the iduronate residues adopt the $^{2}S_{0}$ skew-boat conformation in the protein-bound status. In contrast, NMR studies of the complex of a heparin tetrasaccharide with AT claimed that the binding to protein causes a distinct change in conformation of the glycosidic linkage and the iduronate residue then adopt the $^{1}C_{4}$ chair conformation [106].

A considerable body of evidence shows that different HBPs recognise distinct HS through often overlapping patterns of sulfation in HS. For example, while FGF-1 and FGF-2 both bind to sequences containing IdoA2S-GlcNS, FGF-1 also binds well to oligosaccharides containing IdoA-GlcNS, 6S [107] [108]. In addition, platelet factor 4 (PF-4) and interleukin 8 (IL-8) form 1:1 complexes with long sugar sequences of 12-20 saccharides, whereas vascular endothelial growth factor (VEGF), interferon-gamma and chemokines require short NS domains separated by spacer region containing N-acetylated residues (NA/NS domains) [83] [45]. The underlying reason is the ability of the N-acetyl–rich 'spacer' or NA-domain (six to seven saccharides) in HS to allow much more conformational freedom for the simultaneous.

The level of structural specificity required for HS-protein interactions remains an open question. There are several factors that have hindered progress. One on the protein side is the absence of a systematic analysis of the heparin binding sites of a large number of unrelated

proteins, due to a lack of such data. On the GAG side, the complexity of polysaccharide chain structure in solution has confounded analysis, which often simply considers the sequence of saccharides.

1.4.3 How do proteins engage GAGs?

The binding of heparin and HS to proteins is driven by electrostatic bonds [47] between the sulfate and carboxyl groups on the polysaccharide and the basic amino acids arginine and lysine present in the HBS of most proteins [109]. While these interactions may not dominate energetically (Section 1.4.2), they do dominate kinetically. In an attempt to determining a consensus sequence of basic amino acids in heparin-binding sites, Cardin and Weintraub compared the HBS of 21 proteins [110] and proposed two consensus sequences of amino acids: XBBXBX and XBBBXXBX, where B is a basic residue and X is a hydropathic residue. Molecular modelling studies assumed the sequence XBBXBX as a β-strand conformation with the basic amino acids on one face of the β-strand. Similarly, the sequence XBBBXXBX folds into an α-helix with the basic amino acids displayed on one side and hydrophobic amino acids pointing back into protein core. Another consensus proposed was: XBBBXXBBBXXBBX, where 'B' represents a cationic residue coming from a molecular modelling of von the Willebrand factor HBS [111]. However, such sequences of amino acids derived from the linear sequence of HBSs may not be representative, since these patterns of residues are not found in many HBPs. Margalit *et al* suggested that a distance of around 20 Å between basic residues is required for the interaction [112]. Moreover, many HBSs are formed by amino acids distant in sequence, but adjacent in structure, so the sequence of a protein does not necessarily predict the interaction with the polysaccharide [109].

One challenge in studying GAG binding sites is the high variation of the interactions, as the structural similarity and evolutionary relatedness of proteins is not always reflected by common GAG binding sites. Taking platelet factor 4 (PF4) and IL-8 as an example, these two proteins

are members of the α-chemokine family sharing common monomeric three dimensional structures, with anti-parallel β-strands and an α-helix in the C-terminus. However, while PF4 engages heparin through the sequence 'KKIIKK', the equivalent heparin/HS binding domain of IL-8 is 'KENWVQRVVEKFLKR' [113] [98].

One important consequence of the above is that, there is at present no bioinformatic tool that is able to predict heparin binding sites in proteins.

1.4.4 Electrostatic interactions (arginine and lysine)

Both the guanidino group of arginine and prime amine of lysine have pKa over 10, therefore, they are positively charged at physiological pH. These properties enables lysines and arginines to be involved in hydrogen-bonding, ion pair, as short range interactions, or electrostatic interactions as long range ones [114] [115] [116]. In terms of their basicity, the lysine side chain primary amino group is symmetric and the charge focussed in a small volume. This results in lysine being able to form a single bond (Figure 1.9). In contrast the guanidinino group of arginine is planar and asymmetric (N^{ε}, $N^{\eta 1}$, $N^{\eta 2}$), with the conjugation occurring between the double bond and the nitrogen pairs. This allows the delocalization of the positive charge and, therefore, multiple ion pair and H-bonding interactions with sulfate/carboxylate groups (Figure 1.8). The electrostatic interactions formed by the guanidino group with sulfate are relatively stronger than the ones formed by the primary amine, based on Pearson's concept of soft acid soft base interactions (Figure 1.9) [117] [114]. In addition, there are up to 5-6 H-bonds that can be formed by the guanidino group whereas H-bonds formed by lysine are not only fewer but also much more dynamic [118]. In terms of energy, contact ion pairs formed by arginine can be more energetically favourable compared to those of lysine [119]. The interaction between carboxylates and arginine can formed from the side-on and end-on interactions (bidentate configuration) or the backside interaction (monodentate configuration) involving the Nη1 hydrogen closest to Nε [114].

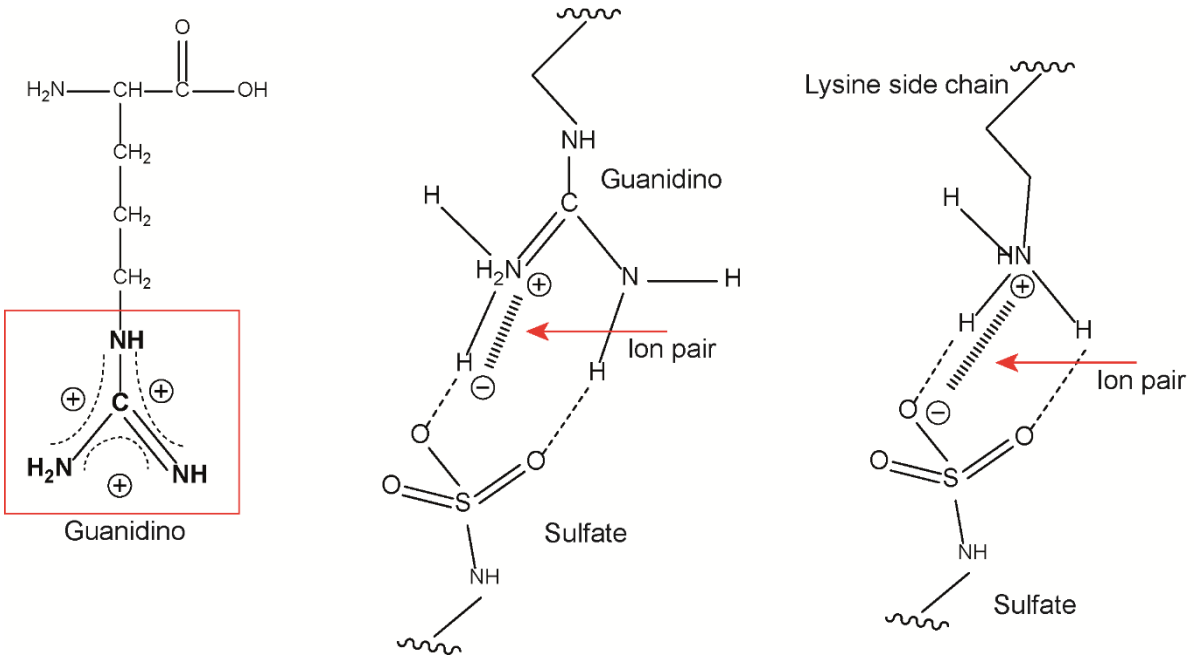


Figure 1.9: Interactions of arginine and lysine side chains with sulfate groups. (A) Charge delocalization in the guanidino group of arginine. (B) Lys and Arg ion pair interactions with sulfates. The chemical structure was drawn by ChemDraw.

1.4.5 The canonical HBSs and the secondary HBSs

The theory of the secondary HBS, including HBS-2, HBS-3 and HBS-4 emerged from the study on FGF1 subfamily [120] [36], in which HBS-2 locates at the area of β-strand IX and X, HBS-3 tends to position at the N-terminus and C-terminus of the protein, finally, the work on FGF7 [36] introduced HBS-4 which form a large T-shape patch relatively to HBS-1, hence stopping a single polysaccharide chain from crossing over. Compared to the canonical HBS-1, the affinity of those binding sites to heparin is modest but they contribute significantly to the functions. One example is the deletion of HBS-3 resulted in the dramatic reduction in the mitogenic activity of FGF-2 [121].

1.4.6 Cellular functions of heparin/heparan sulfate and protein interactions

The interaction of effector proteins and HS can be considered to have two types of function. One relates to the regulation of the localization and diffusion of the HBPs. The other relates to the assembly of functional complexes of HBPs and polysaccharide.

1.4.6.1 Localisation and diffusion

As a regulator for the movement of effectors between cells, HS was shown to control the transport of FGF2 across Descemet's membrane (DM), the basement membrane of the corneal endothelium with fast, reversible binding [122]. Morphogen-HS interactions were also demonstrated to contribute to gradient formation and so regulating the development of the embryo in contexts ranging from vertebrates to *Drosophila* [2][123][124]. As an example, the lower binding affinity of FGF7 compared to FGF10 for HS results in the formation of two distinguishable gradients that lead to distinct activities during branching morphogenesis [2]. The mechanism of the regulation remains unclear, however, some recent work has provided some insights. The study of nanoparticle-labelled FGF2 demonstrated that the spatial distribution of HS binding sites on proteoglycans of the pericellular matrix is not homogeneous, but heterogeneous and clustered over length scales ranging from 22 nm and above [125]. Moreover, in this environment, FGF2 undergoes five distinct types of movement, including immobile/highly confined, confined diffusion, simple diffusive, slow directed diffusion and long and fast directed diffusion. Recently, biophysical experiments have shown that some effectors, such as FGF2 and CXCL12γ which possess more than one binding sites for HS can cross-link the polyanion chains [126]. Moreover, the level of binding, the spatial distribution of binding sites and the diffusion in the pericellular matrix of fibroblasts are remarkably different for five paracrine FGFs [127]. The interpretation of these data has given rise to the idea that HS chains in extracellular matrix possess long-range (supramolecular) structure due

to the chains forming a dynamic network of interlinked molecules resulted from the high multiplicity of interactions among endogenous HBPs.

1.4.6.2 HS as a co-receptor in signalling complexes

For many effectors, HS functions as a co-receptor. The classic system in this respect is the FGF signalling system [128] [129], where binding to HS can determine the strength and duration of the signalling [85] [130] [131]. There are two models derived from co-crystal structures of FGF ligand, model heparin oligosaccharides and FGFR extracellular domain, the symmetric 2:2:2 and asymmetric 1:1:1 stoichiometry complexes [132][130] [133], though only the former is currently considered to reflect what may occur at the cell surface. What is agreed is that FGF ligands can bind FGFR in the absence of HS and generate intracellular signals [134] [135], but that a full mitogenic response requires the engagement of the HS co-receptor. Experimentally, heparin is found to bind both the FGF ligand and the FGFR, suggesting that engagement of all three species in a complex is likely to increase its stability with respect to dissociation [136].

1.4.7 The binding interfaces of the ternary complex, FGF-HS-FGFR

The knowledge of the binding interface of the ternary complex comes mainly crystal structure studies, hence, quite limited. In general, there are four areas of binding reported. For example, the asymmetric FGF-1-FGFR2-heparin DP10 crystal structure (PDB: 1E0O) presented the engagement at the areas of first, β1 and loop βI-βII (Y30, Y35 and Y37), second, the loop βIII-IV (R50 and R52) and loop IV-V (V66), third, βVIII to loop βVIII-IX (E102, 104-LEE-106 and 108-HYN-109), last, on β XII (L148 and L150).

In the crystal of FGF2-FGFR1-heparin [130], the pivotal function of 6-*O*-sulfation of heparin was highlighted when this sulfation formed the hydrogen bonds with both FGF and FGFR. Three areas on FGF2 involved in binding, the β I-II loop (Asn-27), the β IX-X loop (Arg-120, Thr-121), the β XI-XII (Lys-125, Lys-129, Gln-134, Lys-135, Ala-136). In addition to

interacting with the FGF-FGFR preformed complex, heparin was shown to form 5 hydrogen bonds with the adjoining receptor across the 2-fold dimer.

Besides a small number of studies which investitage the terrary complex as whole, there are several attempts of investigating the interaction between FGFs and their cognate receptors. The crystal structures of FGF10 and FGFR2b (PDB: 1nun [31]) and FGF8b and FGFR2C (PDB: 2fdb [137]) individually reported four areas of engagement on protein surface which are aligned in terms of sequence to those in FGF1 and FGF2. Those areas on FGF10 are: first, the N-terminus and β strand I (71-HLLQGDVR-78, R80, F83 and F85), second, loop of β strands III-IV to β-strand VI (K102, E104, 113-ITSVEIG-119, V121 and Y131), third, the loop of β strands VII-VIII to β strand IX (F146, 154-ERI-156 and 159-NGY-161), last, β strand XII (L202 and M204). Four areas on FGF8b are: first, the N-terminus and β strand I (containing 11 residues: F50, H53, V54, Q57, D62, L64, R66, L68, R70, Y75 and R77), second, the loop of β strands IV-V (V106, T108, F111 and S113), third, the loop of β strands VIII-IX (E159, V161, L162 and 165-NYT-167), last, β strand XII (193-MKR-195). The consistence in binding pattern of FGF to FGFR demonstrated the conservation evolution impact on this family.

1.5 APPROACHES FOR IDENTIFICATIONS OF GAG BINDING SITES IN PROTEIN

Proteins have in the past been identified as binding HS serendipitously. The result was a gradual growth of the number of HBPs, which in 2008 were catalogued as numbering 214 [45]. An affinity proteomics approach increased this number to 435 [138] and subsequently to 831 [139]. An understanding of the functional significance of the protein-polysaccharide interaction lags behind. For example, the interaction between FGF-1 and FGF-2 with heparin was critical in their original purification and the cloning of their cDNAs [15] [140]. However, it was only in 1991 that the requirement for HS in the assembly of the FGF2-FGF receptor signalling complex was discovered. Regarding to the specificity and selectivity of protein and

HS interactions, improvements have only made through the development of new experimental approaches. These are discussed below.

1.5.1 Low throughput methods

The identification of the binding site of the polysaccharide on the protein has in the past depended on structural and molecular biology approaches that are low throughput e.g. NMR spectroscopy, site-directed mutagenesis [141] [142] and X-ray crystallography [100], [143], [144]. Of these, X-ray crystallography has been the largest contributor, as seen by the number of models of protein-heparin oligosaccharide co-crystals uploaded onto the PDB (https://www.rcsb.org/). In addition to being low throughput and requiring the formation of crystals, a further weakness of this approach is that X-ray crystallography will select for a rigid protein conformation, should several conformations exist in equilibrium. This is highlighted by the identical structures of FGF2 and FGF2 in complex with a heparin hexasaccharide [145] [100], despite the long-established fact that FGF2 bound to heparin is more rigid, possessing more secondary structure than the unbound protein [146][147], which has been used more recently to measure in high throughput interactions with different model GAG structures [102]. With NMR and the actual conformation of the protein in the bound state could be obtained with the information of the rearrangements of protein and sugar structure [148]. However, NMR spectroscopy is limited by the size of protein that can be analysed, as well as being low throughput.

1.5.2 High throughput methods

Screening by surface noncovalent affinity mass spectrometry (SNA-MS)

The development of mass spectrometric (MS) approaches has abled this to go further than the analysis of proteins to forward the characterization of many biological components, including heparin/heparan sulfate-like glycosaminoglycan (HLGAG), resulting the emerge of the

approach named surface noncovalent affinity mass spectrometry (SNA-MS) [149]. The overall strategy for SNA-MS requires the immobilization of protein on a thin hydrophobic film, followed by the absorbance of target proteins onto MS targets. Then, oligosaccharides were eluted and their structures were identified by MALDI-TOF using chelation with $(RG)_{19}R$ [149]. This method allowed the rapid parallel screening of an oligosaccharide library against multiple proteins, however, it is limited only to proteins actively immobilized on the target.

Identification of protein–GAG interactions by surface plasmon resonance

SPR binding assays use the change in refractive index at a surface to measure the interaction between a partner bound to the surface and one in solution [150]. These assays have been used to identify new protein–GAG interactions and in competition format can identify preferences for GAG structures [151]. The identification of binding sites in proteins requires either the use of competing synthetic peptides [121] or the use of mutant proteins [82].

Pull-down and 2D gel electrophoresis

Recent studies of the interactions between proteins and polyanions such as heparin, actin, tubulin and DNA used simple pull-down experiments performed on COS-7 cell extracts in combination with the two-dimensional electrophoretic methods and antibody arrays [152]. The number of heparin binding protein identified here was 944 out of a total of 1,751 which was double the number of actin and tubulin interactors [152]. However, because using the chromatographic method, this work could not distinguish the proteins which directly engages to heparin and the proteins which binds to heparin binding proteins.

1.5.3. Bioinformatics tools

Molecular docking

Computational docking has been used to study protein interactions such as those with GAGs. The process basically involves two steps: 1) molecular docking (for example, AUTODOCK software) and 2) molecular dynamic simulation (for instance, MD simulation refinement) [150]. Step 1 is able to identify the position of a rigid fragment on the protein surface, then step 2 is used to refine the docked model. There are many successful examples of using this process on addressing the binding sites of heparin on proteins such as endostatin [153] and the engagement of HS to FGFs [154]. A study docking different oligosaccharides to the potential surface of chemokines which are positively charged demonstrated that the different selected oligomers contact different clusters of basic amino acids, giving some important cues for the specificity of GAG-chemokine interactions [155].

A heparin docking server is available [156]. This method overcomes the general restriction of other docking strategies that search for a binding pocket, which limits the search around the defined heparin binding sites. In the future with more availability of the structural data, it will be possible to generate more precise models of interaction between GAG and proteins, more potentially for a group of related proteins, that fit with available data. If successful, this work will be extremely useful for the design of new therapeutics. However, validation of docking has still relied on low throughput methods.

1.5.4 The lysine selective labelling approach

A high throughput method has been developed for identifying lysine residues in heparin binding sites (HBSs) of HBPs was called "protect and label" and using mass spectrometry identification of labelled lysine residues [120]. In the protection step, the protein is bound to heparin in a chromatography column and exposed lysine residues are reacted with NHS-

acetate. Thus, the lysine side chains not involved in binding heparin are protected from further reaction. The protein is then eluted from the heparin affinity column and any lysine side chains protected by the interaction with heparin are then labelled by reaction with NHS-biotin. Biotinylated peptides may be purified quantitatively and are then identified by mass spectrometry. An interesting feature of this method is that it identifies canonical, high affinity heparin binding sites in proteins and lower affinity secondary binding sites [120][36]. Since, there are far more identified HBPs than HBPs with identified polysaccharide binding sites and in the absence of a bioinformatics approach to HBSs other than by alignment with a well characterised member of the same protein family, the lysine protect and label method fills a major gap.

However, the lysine protect and label method has its limitations. The "protect and label" approach only detects lysine side chains involved in binding of the polysaccharide, but arginine residues are equally important (Section 1.4.4). In addition, the heparin binding sites in some proteins may contain arginine residues, e.g., the canonical heparin binding site of FGF-22 predicted by sequence alignment with other FGFs [36]. As the method depends on the use of a heparin affinity chromatography matrix, it is also unable to probe protein interactions with more physiologically relevant HS or other GAGs.

1.6 RATIONALE AND AIMS OF THE THESIS

Ionic interactions play an important role in the interactions of proteins and GAGs evidenced by the work on the lysine selective labelling on many FGFs. However, some proteins, for example FGF22, do not contain any lysine but arginine amino acids in their binding sites. Indeed, the contribution of arginine to the electrostatic interactions is essential to comprehensively understand the engagement of protein to sugar.

Hence, the first aim of this thesis is to develop an approach which can quickly and efficiently identify the arginine residues on protein responsible for the electrostatic engagement to the sulfated GAGs. The second aim is to produce a complete panel of all fifteen paracrine FGFs and determine the structure in these proteins involved in heparin binding using the method of selective labelling arginine and lysine. This piece of information was then used to justify whether FGFs in the same subfamily may share some common features in this kind of interaction. Moreover, the positions of labelled arginine and lysine residues on protein draw a map of electrostatic distribution on the protein surface which, together with the information of the selectivity/specificity of that protein toward the modifications on the sugar, is valuable to predict the potential binding pattern of sugar on a specific protein.

For long, heparin has been used as the proxy to the S-domain of HS. Indeed, interest in interactions of the protein with heparin and HS overwhelmed other GAGs, such as CS, DS due to some technical limitations. Here, the third purpose is to overcome the heavy dependence of the selective labelling approach on heparin beads by sensible alteration of the original version to the selective labelling "in-solution" using commercially available heparin, CS and DS. This method is first tested on the FGF7 sub-family.

As the fourth aim, one of the known HS-binding proteins is selected to apply both selective labelling of arginine and lysine. This shows that selective labelling is a handy approach which could be used to study the electrostatic interactions of many proteins with GAGs.

Longstanding questions are how specific and selective protein-glycosaminoglycan interactions are, to what extent can HBSs be predicted, and, if the residues in an HBS are known, what might be the sugar specificity of that protein. A model of the engagement between protein and heparin is desired to develop using the knowledge gained here of the basic residues on the protein surface and the selectivity toward sulfations of polysaccharide chain.

CHAPTER 2 MATERIALS AND METHODS

2.1 ELECTROPHORESIS

2.1.1 Agarose electrophoresis

Agarose gels (1.2 %, w/v) were made by melting agarose (0.4 g) (Bioline, London, UK) in 48 mL TAE buffer (40 mM Tris-Cl, 20 mM acetic acid, and 1 mM EDTA, pH 8). When the agarose had cooled slightly, 4 μL 10,000x SYBR (New England Biolab, UK) was added. The molten agarose was poured into the gel making tray and allowed to cool. The samples and DNA ladder (1 kb New England Biolab, Herts, UK) were loaded on the gel after the gel tray was placed in the gel tank and covered with TAE buffer. Electrophoresis was carried out at 100 V, 30 min for each gel.

2.1.2 SDS PAGE

To prepare samples for SDS-PAGE, sample (30 μL) was mixed with 10 μL 4x SDS-PAGE loading buffer (50 % (v/v) glycerol, 10 % (w/v) SDS, 25 % (v/v) 2-mercaptoethanol in 0.3 mM Tris-Cl, pH 6.8 and coloured with bromophenol blue), and then heated at 95°C for 10 min. Ten μL of the samples along with 10 μL SDS-PAGE markers (SDS7-1VL, Sigma) were loaded onto a 12 % (w/v) SDS-polyacrylamide gel (Table 2.1) and the gels were run at 30 mA (per gel), 200 V for 50 min with running buffer (50 mM Tris-Cl, 192 mM glycine and 0.1 % (w/v) SDS).

2.1.3 SILVER STAIN

After electrophoresis, the gels were soaked in fixative (40 % (v/v) ethanol, 10 % (v/v) acetic acid) for 1 hour following by soaking twice in 10 % (v/v) ethanol (5 min/each) and washing in RO water three times, each for 5 min.

Following incubation in 0.2 % (v/v) silver nitrate for 30 min, the gels were washed with water for 15 s and then dipped in freshly made developer solution (2.5 % (w/v) Na_2CO_3, 0.03 % (v/v) formaldehyde) until the solution went brown. New developer was then used to further stain the gel until bands were stained to the required intensity. Stop solution (1 % v/v acetic acid) was used to stop the reaction between silver and formaldehyde. The gels were then washed with water six times for 5 min each. Freshly made reducer (0.6 % w/v sodium thiosulphate, 0.3 % w/v potassium ferricyanide, 0.1 % w/v sodium carbonate) was used to remove the excess silver and clear the background. The gels were then quickly washed with a large volume of water and followed by five washes with water (5 min/each). When required, gels were re-stained to increase the sensitivity of detection, starting by the addition of 0.2 % (v/v) silver nitrate for 30 min.

Table 2.1: SDS-PAGE Gel preparation

SDS-PAGE Resolving Gel (ingredients for 12 % gel – 2 gels)	**Total 10 mL**
Acrylamide/ bis-acrylamide stock (30 %, w/v)	4.0 mL
Tris-HCl (3 M), pH 8.85	2.5 mL
DD- Water	3.5 mL
10 % Sodium dodecyl sulfate (SDS) w/v	100 µL
TEMED(N, N, N', N', Tetramethylethylene diamine)	10 µL
Ammonium persulphate 50 mg/mL (freshly made)	100 µL

SDS-PAGE Stacking Gel (ingredients for 12 % gel – 2 gels)	**Total 5 mL**
Acrylamide/ bis-acrylamide stock (30 %, w/v)	650 µL
Tris-Cl (1.25 M), pH 6.8	500 µL
DD- Water	3.5 mL
10 % Sodium dodecyl sulfate (SDS) w/v	50 µL
TEMED (N, N, N', N', Tetramethylethylene diamine)	10 µL
Ammonium persulphate 50 mg/mL (freshly made)	50 µL

The gels were incubated in Coomassie stain (50 % (v/v) methanol v/v, 10 % (v/v) acetic acid, 0.25 % (w/v) CBB R-250) for 60 min, then soaked in destaining buffer (40 % (v/v) methanol, 10 % (v/v) acidic acid) until the bands became clear.

2.2 BACTERIAL CULTURE STRAINS

2.2.1 Vectors and cDNAs

Existing expression vectors were used: pET-14b containing cDNAs encoding FGF1 and FGF2 [36] and pET-M11 vector containing FGF7 cDNA encoding FGF3, FGF10, FGF16, FGF17, FGF20, FGF6, FGF5, FGF8 and FGF22 as well as HaloTag [102]. All of the protein sequences corresponding to the above cDNAs are listed in Table 3.1. Bacterial cells: DH5α, BL21 (DE3) pLysS and SoluBL21 were a gift from Olga Mayans, University of Liverpool.

2.2.2 Medium for bacterial culture

Lysogeny Broth (LB) and LB Agar culture medium was made following the instructions of the manufacturer (Merk, East Yorkshire, UK).

2.2.3 Competent cell list and culture

DH5α cells, CL41 cells and BL21 (DE3) were stored in -80 °C. Competent cells were thawed on ice. Ten µL was plated out onto LB agar plates. Plates were incubated overnight at 37 °C.

2.2.4 Competent cell preparation

A single colony (Section 2.2.3) was picked from the plate and cultured in 8 mL LB overnight at 37°C. The next day, 1 ml of bacterial culture was transferred into 100 ml LB and this culture was incubated at 37 °C, until the absorbance at 600 nm reached 0.3-0.5. The culture was immersed in ice for 5 minutes and divided into 4 pre-chilled sterile centrifuge tubes and then centrifuged at 6000 rpm for 5 minutes, 4 °C, to collect the cells. The pellets were re-suspended in 20 mL Transformation buffer I (Tbf I- 30 mM potassium acetate (KOAc), 100 mM RbCl, 10 mM $CaCl_2$, 50 mM $MnCl_2$, 15% (v/v) glycerol, pH 5.8) and then were put on ice for 5 min, followed by centrifugation at 4°C, 6000 rpm, for 5 min to collect the cells. Finally, cells were re-suspended in 2 mL Transformation buffer II (Tbf II- 10 mM MOPS, 70 mM $CaCl_2$, 10 mM RbCl, 15% (v/v) glycerol, pH 6.5). The competent cells were put on ice for 15 min before they were placed in aliquots in pre-chilled tubes and stored at -80 °C.

2.2.5 Transformation

A stock of competent cells and plasmids were thawed on ice. A mixture of 70 µL competent cells and 70 ng (~1 µL) plasmid was placed on ice for 30 minutes, followed by heating at 42 °C for 45 seconds and then immediately placed again on ice for 2 minutes. In a new tube, 1 ml Lysogeny Broth (LB) medium (Merck, Watford, UK) was added to the mixture and incubated for 60 minutes at 37 °C on a 250 rpm shaker. The tube was centrifuged to collect the pellet, which was suspended in 100 µL LB. On a LB Agar-antibiotic plate (antibiotic dependent on the plasmid, either ampicillin or kanamycin from Sigma-Aldrich Ltd. Dorset, UK, 100 mg/mL stock), 20 µL of this mixture was spread and the dish was incubated at 37 °C for 16 hours. After that, the plate could be stored in the cold room.

2.2.6 Miniprep

A single colony was taken (Section 2.2.5) and cultured in 8 mL of LB broth overnight at 37°C (shaken at 240 rpm). Plasmids were purified the next day using the Qiagen Miniprep Kit 250 (Qiagen, Manchester, UK) following the manufacturer's instructions and then stored at -20 °C.

2.3 PROTEIN EXPRESSION

2.3.1 IPTG induction

A single colony (Section 2.2.5) was cultured in a tube of 8 mL of LB Broth and 8 µL of the appropriate antibiotic from a 1000x stock at 37 °C overnight (shaken at 240 rpm). Generally, six colonies were expanded in this way. In the next step, the culture tubes were pooled into a 2 L flask containing 800 mL of LB broth with 800 µL of the appropriate antibiotic and then cultured at 37°C with shaking (240 rpm) until the absorbance at 600 nm reached 0.4-0.5. For all FGFs except for FGF2, the temperature was then reduced to 16 °C. When the absorbance at 600 nm was 0.6, 800 µL IPTG (1 mM final concentration) was added to each flask and the bacteria were grown at the same temperature for 3 h (37 °C) or overnight (16 °C) with shaking (240 rpm).

2.3.2 Cell harvesting

The bacteria were harvested by centrifugation at 6000 rpm for 15 min (4°C) in Sorvall RC6 centrifuge (Thermo Scientific, UK) using a Fiberlite™ F10-6 x 500y Fixed-Angle Rotor. The cell pellets were re-suspended in 30 mL lysis buffer (0.1 to 0.3 M NaCl, 50 mM Tris-Cl (pH 7.2)) for each centrifuge tube.

2.4 CHROMATOGRAPHY

2.4.1 Chromatography columns

The bacterial pellet suspended in lysis buffer (Section 2.3.2) was kept on ice prior to sonication with a Soniprep 150 Plus (MSE, UK). Cells were disrupted by six cycles of 30 s sonication at 12 amplitude and 60 s break between each cycle). To remove cell debris, the disrupted cells were centrifuged at 15000 x g for 40 min in a Sorvall RC-5B refrigerated super-speed centrifuge. The supernatant was then filtered through a 0.45 μM pore filter prior to chromatography.

2.4.2 Heparin affinity chromatography

An affinity column was made with 3 mL heparin agarose (Affi-Gel Heparin, BioRad, UK). It was first equilibrated with washing buffer (50 mM Tris-Cl, 0.3 M NaCl, pH 7.2) at a flow rate of 1.0 mL/min. The supernatant after filtration (Section 2.4.1) was applied to the column and the flow-through fraction was collected. After that, the column was washed with 5 volume of each of two or three washing buffers (0.4M NaCl, 0.5M NaCl, and 0.6M NaCl, all in 50 mM Tris-Cl, pH 7.2), depending on the FGF purified. Higher ionic strength elution buffers 1.0 M NaCl, 1.5 M NaCl or 2.0 M NaCl (depending on the FGF) in 50 mM Tris-Cl, pH 7.2 were applied to the column to elute the bound proteins. During purification, the column effluent was monitored at 280 nM with an Econo UV monitor (Bio-rad, Hertfordshire, UK).

Owing to the requirement of subsequent experiments, phosphate-buffered saline PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 2.0 mM KH_2PO_4, pH 7.4) was used in place of 50 mM

Tris-Cl buffer in some instances, but the heparin affinity chromatography procedure was otherwise unchanged.

2.4.3 Ion exchange chromatography

A cation-exchange column (1 mL HiTrap SP-HP, GE Healthcare Life Sciences, UK) or DEAE Sepharose (1 mL) (GE Healthcare Life Sciences, UK) column for anion-exchange chromatography was equilibrated with 0.2 M NaCl, 50 mM Tris-Cl, pH 7.2 for a minimum of 5 min at a flow rate of 1.0 mL/min. The fractions from heparin affinity chromatography (Section 2.4.2) were diluted with 50 mM Tris-Cl, pH 7.2, so that the final concentration of NaCl in the loading solution was less than 0.1 M, and applied to the ion-exchange column. The column was then washed with steps of increasing NaCl (0.3 M, 0.4 M and 0.5 M) in 50 mM Tris-Cl, pH 7.2. Depending on the FGF, proteins were eluted with 0.6 M, 0.8 M or 1.0 M NaCl in 50 mM Tris-Cl, pH 7.2. For the requirement of the subsequent experiments, in some preparations solutions were buffered with PB buffer (17.9 mM Na_2HPO_4, 2.1 mM NaH_2PO_4, pH 7.8) rather than Tris-Cl.

2.4.4 Nickel (Ni^{2+}) chelation chromatography for His-tag purification

Ni^{2+} Sepharose (from GE Healthcare Life Sciences, UK) was used for the purification of some His-tag proteins. The column was equilibrated with 5 to 10 column volumes 20 mM sodium phosphate, 0.3 M NaCl, 50 mM imidazole, pH 7.4. The fractions from heparin affinity chromatography were diluted with 20 mM sodium phosphate, 50 mM imidazole, pH 7.4, so that the final concentration of NaCl in the loading solution was around than 0.1M, and applied to the Ni^{2+} column. The column was then washed with steps of increasing imidazole (0.1 M, 0.2 M and 0.3 M) in 20 mM sodium phosphate, 0.3 M NaCl, pH 7.4. Depending on the FGF, proteins were eluted with 0.4 M or 0.5 M imidazole in 20 mM sodium phosphate, 0.3 M NaCl, pH 7.4.

2.5 SELECTIVE LABELLING ON LYSINE RESIDUES

2.5.1 Protection of lysine residues

The method of Ori [120] was followed. To make a mini-column, a plastic air filter was placed at the end of a P10 tip (P50) and packed with 30 µL AF-heparin beads (Tosoh Biosciences GmbH, Stuttgart, Germany; binding capacity of 4 mg Antithrombin III/mL, generally, the loading capacity of FGFs to resin was estimated at 15 mg/mL). PB150 buffer (150 mM NaCl, 17.9 mM Na_2HPO_4, 2.1 mM NaH_2PO_4, pH 7.8) (4 x 50 µL) was used to equilibrate the heparin column. A minimum of 10 µg FGF protein was loaded onto the column. The loading was repeated 3 times to ensure the binding between protein and heparin beads. After binding, 200 µL (4 x 50 µL) PB150 buffer was used to wash the column and remove the unbound components.

Lysine remaining exposed to solvent were acetylated by a quick rinse of the column with 20 µL PB 150 containing 50 mM sulfo-NHS-acetate and followed by and incubation for 5 min with 20 µL of the above solution at room temperature. The minicolumn was then washed with 200 µL PB150 buffer (4 x 50 µL). Bound proteins were collected in 2 x 20 µL elution buffer (2 M NaCl, 45 mM Na_2HPO_4, 5 mM NaH_2PO_4, pH 7.8).

2.5.2 HBS Lysine Biotinylation

Owing to the high concentration of NaCl in elution buffer, the acetylated protein was diluted with 200 µL PB buffer (17.9 mM Na_2HPO_4, 2.1 mM NaH_2PO_4, pH 7.8) and concentrated on a 3.5 kDa-MWCO centrifugal filter (Sartorius Ltd., Epsom, UK), which was centrifuged for 25 min at 13200 rpm until the volume was reduced to 37.2 µL. The biotinylation of lysine residues was accomplished by adding to the sample 2.8 µL 145 mM NHS-biotin in DMSO and incubating for 30 min at room temperature. The reaction was quenched with 4 µL 1 M Tris-Cl, pH 7.5.

The excess reagents were removed by 3 cycles of dilution on 3.5 kDa-MWCO centrifugal filters with 400 µL 10-times diluted PB150 buffer and 3 cycles of dilution with 400 µL HPLC water by centrifugation at 13200 rpm for 10 min. Products (40 µL) were then applied to a desalting centrifugal column (7K MWCO, Thermo Scientific, The UK) and covered by 70 µL HPLC grade water. The column was then centrifuged for 2 min at 1,200 x g. After freezing at -80 °C for 20 min, the sample was lyophilized for an hour.

2.5.3 Protein Digestion

Dried sample was dissolved in 8 M urea, 400 mM NH_4HCO_3, pH 7.8 (25 µL) and 45 mM DTT (2.5 µL) and incubated for 15 min at 56°C. Proteins were diluted with 70 µL HPLC grade water and digested overnight with 0.1µg of the appropriate enzyme, chymotrypsin or trypsin (both mass spectrometry grade, Promega, Southampton, UK).

2.5.4 Biotinylated Peptide Purification

A minicolumn of 40 µL Strep-Tactin Sepharose beads (IBA GmbH) was packed as above and equilibrated with 4 x 50 µL of 500 mM urea, 25 mM NH_4HCO_3. Digested protein was diluted to a final volume of 200 µL with HPLC grade water and applied to the column three times to ensure binding. The mini-column was washed with 3 x 50 µL 0.5 M urea, 25 mM NH_4HCO_3, followed by 3 x 50 µL HPLC grade water to remove unbound peptides. Biotinylated peptides were eluted with 2 x 20 µL 80% (v/v) acetonitrile 20% (v/v) triflouroacetic Acid (TFA), 5 mM biotin (Thermo Fisher Scientific).

2.6 SELECTIVE LABELLING ON ARGININE RESIDUES

The detailed protocol is uploaded onto **Protocol.IO** an open access methods repository.

https://www.protocols.io/view/selective-protection-and-labelling-of-arginine-lys-qqmdvu6

2.6.1 Protection of arginine side chains

Step 1: Binding

Protein was loaded onto an AF-heparin mini affinity column as for lysine protect and label (Section 2.7.1).

Step 2: Protection of arginine side chains

PGO (Merck Ltd., UK, 97%) was used in the dark, as it is light sensitive. PGO was freshly prepared in 50% (v/v) DMSO, 50% (v/v) HPLC grade water at 1 M, which was then diluted to 0.5 M and then 0.2 M with 0.2 M $NaHCO_3$, pH 9.5. The pH was adjusted with 0.1 M NaOH to between 9.1 and 9.5, to ensure optimal reaction. The heparin mini column was rinsed with 30 µL 0.2 M PGO solution to exchange buffers. A further 30 µL PGO solution was added to the column and the bound protein was allowed to react for 60 min at room temperature in the dark. The reaction was quenched with 5 µL 0.1% (v/v) Trifluoroacetic acid (TFA) in water, so that the final concentration of TFA was 0.01% (v/v). The mini-column was then washed with 200 µL Na-1 buffer (4 x 50 µL). Bound proteins were eluted with 2 x 20 µL Na-2 buffer (2 M NaCl, 0.2M $NaHCO_3$, pH 9.5) containing 0.1% (w/v) RapiGest SF Surfactant (Waters, UK). The addition of surfactant was important to ensure protein recovery in this and subsequent steps, due to the increased hydrophobicity of proteins following PGO conjugation to arginine side chains.

2.6.2 Labelling of Arginine side chain

Step 3: Labelling of Arginine side chain by HPG

The preparation of HPG was performed in the dark room, as it is even more light-sensitive than PGO, following a procedure identical to that used for PGO. The eluted protein was diluted with 400 µL 0.2M $NaHCO_3$, pH 9.5 and concentrated on a 3.5 kDa MWCO centrifugal filter (Merk Millipore, UK) to a final volume of 70~80 µL. The reaction with HPG was performed by incubating 80 µL diluted protein with 20 µL 0.5M HPG so that the final concentration of HPG in the reaction was 0.1 M. The pH was maintained at over 9.0. The reaction was performed for 60 min at room temperature in the dark and then was quenched with 5 µL 0.1% (v/v) TFA in water.

2.6.3 Sample preparation for mass spectrometry

Step 4: Sample preparation for mass spectrometry

Protein was buffer-exchanged by four cycles of dilution on 3.5 kDa MWCO centrifugal filters with 400 µL 10-fold diluted 0.2 M $NaHCO_3$, pH 9.5 containing 0.1 % (w/v) RapiGest and 3 cycles of dilution with 400 µL HPLC water containing 0.1% (w/v) RapiGest by centrifugation at 13200 g for 10 min. After freezing at -80 °C for 30 min, the sample was lyophilized for an hour.

Step 5: Incubation with proteasses

Chymotrypsin/trypsin: The freeze-dried protein was dissolved in a mixture of 80 µL 25 mM NH_4HCO_3 and 10 µL 1 % (w/v) RapiGest (~ 0.1 % w/v in final solution) and heated at 80 °C for 10 min. The mixture was quickly centrifuged at 3200 g for 30 seconds before 10 µL 50 mM DTT was added (5 mM final concentration) and incubated for 15 min at 56 °C. After cooling the sample to room temperature, proteins were carbamidomethylated with 5 µL 0.1 M iodoacetamide (freshly made) for 30 min in the dark. Proteins were then digested overnight

with mass spectrometry grade chymotrypsin or trypsin (Promega, Southampton, UK) at a ratio of 1:100 (w/w).

Arg-C: The dried sample was dissolved in 400 mM NH_4HCO_3, pH 7.8 (25 µL) and 45 mM DTT (2.5 µL), and incubated for 15 min at 56 °C. After cooling to room temperature, 75 µL incubation buffer (50 mM Tris-Cl, 5 mM $CaCl_2$, 2 mM EDTA, pH 7.8) was added to dilute the urea to 2.0 M. Arg-C protease (Promega, Southampton, UK) was freshly prepared in incubation buffer and then added to the protein solution at a ratio of 1:100 (w/w). Activation buffer 10X (50 mM Tris-Cl, 50 mM DTT, 2 mM EDTA, pH 7.8) was added to give a final concentration of 1X. The mixture was mixed gently and centrifuged briefly before allowing digestion to proceed overnight at 37 °C.

2.7 PROTECT AND LABEL WITH POLYSACCHARIDE IN SOLUTION

2.7.1 Buffer exchange proteins

Step 1: Buffer exchange proteins

The protein (~10 µg) was buffer-exchanged with 400 µL (1) 0.2M $NaHCO_3$, pH 9.5 (Na-1 buffer) for arginine and for lysine (2) PB150 buffer (See Section 2.7.1) when the protein had a slow dissociation form the GAG relative to labelling time or (3) PB buffer (no NaCl) if the protein apparently dissociated faster in a saline than the reaction time [157]. Thus, the concentration of NaCl could be adjusted depending on the dynamic of the individual protein and GAG interaction. The mixture was then concentrated on a 3.5 kDa MWCO centrifugal filter (Merk Millipore, UK) to a final volume of 70~80 µL. This step was repeated 4 to 5 times to ensure buffer exchange.

2.7.2 Protection of lysine/arginine exposed to solvent

Step 2: Interaction between protein and GAG

GAGs, including heparin, CS, DS were stored in water at 1mg/ml. GAG was added to protein at the ratio of 100:1 to 20:1 (GAG: protein Molar concentration), gently mixed and left at room temperature for 5 min.

Step 3: Protection of lysine/arginine residues

Sulfo-NHS-acetate (Section 2.7.1) or NHS-acetate (Thermo Scientific, UK) was prepared at 250 mM in PBS. This reagent was added to the mixture of GAG-protein for final concentration of 50 mM, followed by incubation for 5 min at room temperature. NHS-acetate was preferred to sulfo-NHS for protein GAG interactions that dissociated over the timescale of the labelling step, since it is less charged and does not compete for binding.

PGO (Section 2.8.1) was at 0.5 M in 0.2 M $NaHCO_3$, pH 9.5. The pH was adjusted with 0.1 M NaOH to between 9.1 and 9.5 to ensure optimal reaction. PGO was added to the mixture of GAG-protein at a final concentration of 0.2 M and was allowed to react for 60 min at room temperature in the dark. The reaction was quenched with 0.1% (v/v) trifluoroacetic acid (TFA) in water so that the final concentration of TFA was 0.01% (v/v).

The excess NHS-acetate or PGO was removed by 3-4 cycles of dilution on 3.5 kDa-MWCO centrifugal filters with 400 μL PBS buffer (for lysine) and 0.2 M $NaHCO_3$ (for arginine) using centrifugation at 13200 rpm for 10 min/each cycles. The final volume is recommended to around 70-100 μL.

2.7.3 Selective labelling of lysine/arginine of binding sites

Step 4: Dissociation of the protein-GAG interaction

The GAG:protein complexes were dissociated by the addition of one volume of 6 M NaCl, 45 mM Na_2HPO_4, 5 mM NaH_2PO_4, pH 7.8 to a final concentration 3 M of NaCl and left for 5 min.

Step 5: Labelling of lysine/arginine of binding sites

For arginine, HPG was prepared at 0.5 M (Section 2.8.4) then added to the mixture, so that the final concentration of HPG in the reaction was 0.1 M, and then left for 60 min at room temperature in the dark. The reaction was quenched by adding 0.1% (v/v) TFA in water to a final concentration of 0.01%.

For lysine, 150 mM NHS-biotin in DMSO was used. The final concentration of NHS-biotin was 10 mM and the reaction time was 30 min at room temperature. The reaction was quenched with 1 M Tris-Cl, pH 7.5 at 1:10 (v/v).

2.7.4 Sample preparation for MALDI-TOF

Step 6: Removal of excess reagent

This step is similar to Section 2.7.2 (for lysine) and Section 2.8.3 (for arginine). After centrifugation, samples were kept at -80°C overnight followed by lyophilization for an hour.

Then, the preparation of samples was identical to that described in Section 2.8.3.

2.8 MALDI-TOF FOR IDENTIFICATIONS OF LABELLED PEPTIDES

Peptides were concentrated by rotary evaporation to a final volume of 10 µL and desalted using C18 Zip-Tips (Millipore). C18 Zip Tips were first pre-wetted with 2 x 10 µL 100 % (v/v) acetonitrile and then pre-equilibrated with 2 x 10 µL 0.1% (w/v) TFA in water. The peptides were loaded on the Zip Tip, the loading was repeated 7 to 8 times to ensure binding. The Zip Tip was washed with 10 µL 0.1% (w/v) TFA. Finally, the peptides were eluted with 2 µL of

5mg/mL α-cyano-4-hydroxycinnamic acid (CHCA, > 99% purity, Sigma) in 50:50 acetonitrile/water + 1% (w/v) TFA, straight onto a 96 spot MALDI (matrix-assisted laser desorption/ionisation) target plate.

Analyses were performed on a Synapt G2-Si instrument (Waters, Manchester, UK) with MALDI source equipped with a frequency tripled Nd:YAG UV laser (λ = 355 nm), operating at 1 kHz. The spectrum acquisition time was 120 seconds, with 1 second scan rates, laser energy of 150 Au. The MS spectra were extracted by MASSLYNX v.4.1 (Waters, Manchester, UK) with the spectrum range from 500 Da to 4000 Da. The spectra were then processed using automatic peak detection including background subtraction.

2.9 PYTHON SCRIPTS FOR THE IDENTIFICATIONS OF LABELLED RESIDUES

The scripts for the identification of labelled residues have been uploaded onto **GitHub**.

https://github.com/bpthao/PGO-HPG-mass-predictor_2

and https://github.com/bpthao/Matchmaker_2

Protein Prospector (v 5.19.1 developed at the USCF Mass spectrometry Facility) and Peptide Mass (ExPASy) were used to predict the possible peptides after incubation with an enzyme with the following parameters: enzyme, chymotrypsin or trypsin; maximum missed cleavages, 5; mass range, 500 to 4000 Da; monoisotopic; instrument, MALDI-Q-TOF (Figure 1-B1). The list of peptides after enzyme cleavage was filtered to remove the peptides without arginine residues.

Arginine:

Because products from the reaction between PGO/HPG and arginine residues bring different additional masses to the peptides, the prediction of the peptide masses after the modification was achieved using a script, written in Python (version 3.5.3 released on January 17th, 2017,

available at http://www.python.org), named "*PGO-HPG mass predictor*" (available at Github: https://github.com/bpthao/PGO-HPG-mass-predictor_2). Based on the number of arginine residues in each peptide, *PGO-HPG mass predictor* script" considers all possible reaction products of PGO and HPG with arginine, and generates a list of predicted mass of the modified peptides (Figure 1-B2).

PGO-HPG mass predictor has two input files. First, the list of predicted peptides from the native proteins cleaved by enzyme from either Protein Prospector or Peptide Mass (ExPASy). The file has two columns, the native sequence of the peptide and the corresponding mass. Second, the file of modifications also has two columns, the mass shift of each modification and the description of products. Using the loop, *PGO-HPG mass predictor* automatically adds up all potential mass shifts to each peptide having one or more arginine residues in the sequence. All combinations of mass changes were covered. The output file has four columns, 1-the native sequence of the peptides, 2- the original mass, 3-the final mass after modifications, 4-modifications, which make it easier for further processing.

The observed list is the original mass of FGF peptides after modifications (Figure 1-B3), which provided the mass and intensity of each peak. The match between predicted and the observed list was carried out with a second Python script, "*Matchmaker*", (available at Github in following link: https://github.com/bpthao/Matchmaker_2) with a mass difference tolerance set to 0.1~0.5 Da (Figure 1-B3), as recommend by Mascot.

The output is a list of matching peptides with the native sequence, the predicted mass with modifications, the specific modifications associated with arginine residues and the actual observed mass.

Lysine:

Each reaction between NHS-acetate with the amine side chain of lysine leads to a mass shift of 42.01 Da for the peptide. The mass change in case of biotinylation on lysine is 226.08. Both modifications on lysine containing peptides are considered by the script written in Python (available at Github in following link: **https://github.com/bpthao/** *Lysine mass predictor***).** This script automatically adds up all potential mass shifts to each peptides which have lysine residues in the sequence. The output file has four columns, 1-the native sequence of the peptides, 2- the original mass, 3-the final mass after modifications, 4-modifications. Because the lysine products have only a limited number of combinations, then the final mass after modifications was compared to what was observed in the spectra manually.

CHAPTER 3 PURIFICATION OF PARACINE FGFs

3.1 INTRODUCTION

The initial step in this work required the production of recombinant FGFs in *E.coli*. The protocol for purification of many paracrine FGFs has been described before [37] [36], including the use of Halo-tag as a solubilisation tag with some FGFs that were otherwise poorly expressed. Thirteen of the 15 paracrine FGFs had been successfully purified previously [120][36][37], but not FGF5 and FGF8. Moreover, only 11 paracrine FGFs have had their HBSs analysed by selective labelling of lysine residues, and data are lacking for FGF16 and FGF22 in addition to the two FGFs that had not been expressed in-house.

In this thesis FGF5 and FGF8 were successfully purified as Halo-tag fusion proteins. All other FGFs were produced following existing protocols, except for FGF7, FGF10 and Halo-FGF20 which were produced by Pawin Ngamlert, University of Liverpool.

3.2 PURIFICATIONS OF PROTEINS

The list of the recombinant FGFs used in this study is in Table 3.1

Table 3.1: Recombinant human FGF proteins and their N-terminal tag. FGF names, ID number and amino acid numbering are according to the Uniprot entry. The molecular weight of each FGF with the tag is provided. "Halo" is Halo-tag. FGFs with an N-terminal Halo-tag are called "Halo-FGF" here.

Name	UniProt accession number	Molecular mass (kDa)	N-terminal Tag	Chromatography used
FGF1	P05230	19.1	His	Heparin & Cation-exchange
FGF2	P09038-2	17.3	None	Heparin & Cation-exchange
FGF3	P11487-1	26.89	His	Heparin & Ni^{2+}
FGF4	P08620	19.2	His	Heparin & Cation-exchange
FGF5	P12034-1	58.95	Halo	Heparin & DEAE
FGF6	P10767	51.1	Halo	Heparin & Cation-exchange
FGF7	P21781	22.2	His	
FGF8b	P55075-3	57.5	Halo	Heparin & DEAE
FGF9	P31371	23.4	His	Heparin & Cation-exchange
FGF10	O15520	22.7	His	
FGF16	O43320	21.4	His	Heparin & Ni^{2+}
FGF17	O60258-1	23.3	His	Heparin & Cation-exchange
FGF18	O76093	24.03	His	Heparin & Ni^{2+}
FGF20	Q9NP95	58.8	Halo	
FGF22	Q9HCT0	52.3	Halo	Heparin

3.2.1. Purification of FGF-2

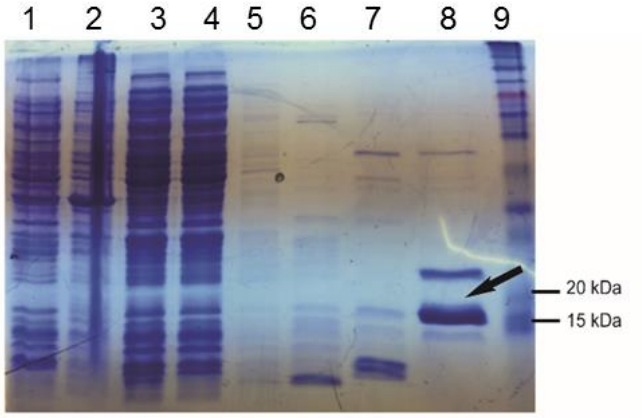
FGF-2 was produced from a culture of 800 µL x 4 flasks (Section 2.3.1). The protein eluting from the Affi-Heparin agarose column (Section 2.4.2) was then diluted before applying to a cation-exchange column (Section 2.4.3). The fractions collected from the columns along with samples corresponding to the starting material and unbound protein were analyzed by SDS-PAGE. The results from heparin chromatography showed that a band, of ~17 kDa was present in fraction of 2.0 M NaCl (Fig 3.1A). This 2 mL was then diluted before applying on the cation-

exchange column. This chromatography showed a single band of ~17 kDa in fractions of 0.6 M NaCl, 1.0 M NaCl, 2.0 M NaCl but not 0.8 M NaCl (Fig 3.1A). The concentration of protein in the fraction of 2.0 M NaCl (Fig 3.1B) was 0.7 mg/mL, as measured by the absorbance at 280 nm. The fractions with FGF2 were pooled and stored as 100 µL aliquots in elution buffer at -80°C.

For FGF2, the purification was repeated using phosphate buffer (Section 2.3.1) (Fig 3.1C). Upon heparin affinity chromatography, the bands at the molecular weight of FGF2 (~17 kDa) were detected in fractions eluting with 0.6 M NaCl and 1.0 M NaCl but not 2.0 M NaCl (Fig 3.1C). The concentration of protein in fraction of 0.6 M NaCl (Fig 3.1C) was 0.5 mg/mL.

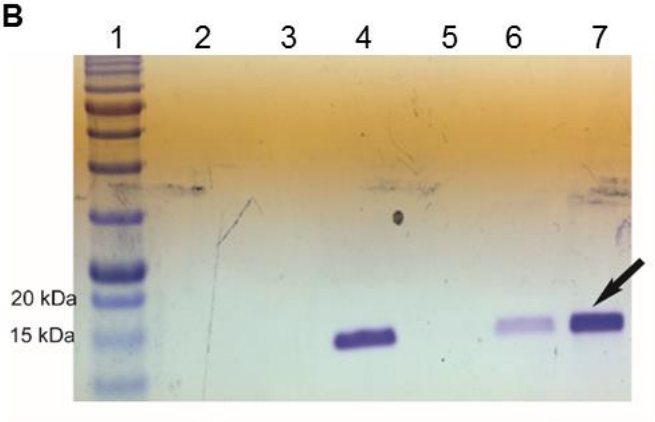
FGF2

A



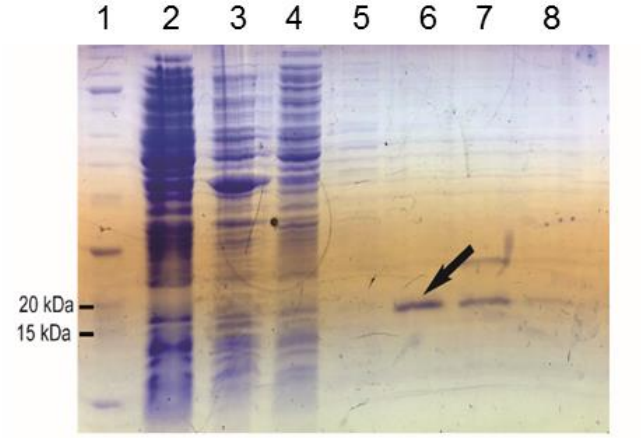
Heparin affinity
1: before IPTG
2: Pellet
3: Loading
4: F/T
5: Wash 0.3 M NaCl
6: Wash 0.6 M NaCl
7: Elute 1.0 M NaCl
8: Elute 2.0 M NaCl
9: Maker

B



Cation-exchange
1: Maker
2: Wash 0.3 M
3: Wash 0.4 M
4: Elute 0.6 M
5: Elute 0.8 M
6: Elute 1.0 M
7: Elute 2.0 M

C



Heparin affinity (phosphate buffer)
1: Maker
2: Pellet
3: Loading
4: F/T
5: Wash 0.3 M
6: Elute 0.6 M
7: Elute 1.0 M
8: Elute 2.0 M

Figure 3.1: Expression and purification of FGF2. (FGF2 molecular weight: 16.7 kDa)

A) Purification of FGF2 by heparin-affinity chromatography using Tris buffer (Section 2.4.2). Lane 1, bacteria before induction by IPTG. The following lanes all after induction by IPTG, all solutions buffered with 50 mM Tris-Cl, pH 7.2: lane 2, Pellet; lane 3, sonicated whole cell lysate; lane 4, flow through (F/T); wash with 0.3 M NaCl (lane 5); 0.6 M NaCl (lane 6); lane 7, elution with 1.0 M NaCl; lane 8, elution with 2.0 M NaCl; lane 9: makers. Black arrow, band corresponding to the molecular weight of FGF2.

B) Further purification of FGF2 by cation-exchange chromatography (Section 2.4.3). After heparin-affinity chromatography, the sample was further purified on a 1 mL HiTrap SP-HP column with all solutions buffered with 50 mM Tris-Cl, pH 7.2. Lane 1, markers; lane 2, 0.3 M NaCl; lane 3, 0.4 M NaCl; lane 4, elution (with 0.6 M NaCl; lane 5, elution with 0.8 M NaCl; lane 6: elution with 1.0 M NaCl; lane 7: elution with 2.0 M NaCl. Black arrow, band corresponding to the molecular weight of FGF2.

C) Similar protocol as panel A) however, instead phosphate buffer (Section 2.4.2) pH 6.8 was used in place of 50 mM Tris-Cl.

3.2.2. Purification of His-tag FGF-1

His-FGF-1 was produced using the plasmid described above (Section 2.2.1) from a culture of 800 µL x 4 flasks (Section 2.3.1). The protein eluting from the Affi-Heparin agarose column (Section 2.4.2) in 2.0 M NaCl, 50 mM Tris-Cl, pH 6.8 was then diluted with the same buffer before applying it to a 1 mL HiTrap SP-HP cation-exchange column and eluting using phosphate buffered solutions (Section 2.4.3). The fractions collected from the second column were analyzed by SDS-PAGE. The results from cation-exchange chromatography showed that a band of ~19.1 kDa was present when the concentration of NaCl in the buffer was above 0.6 M (Fig 3.2). A single band of ~19 kDa was in the 1.5 M NaCl elutate, whereas in other fractions

additional bands were present. The concentration of protein in fraction the 1.5 M NaCl eluate was 1.2 mg/mL as measured by the absorbance at 280 nm. These fractions were stored as 100 µL aliquots in elution buffer at -80 °C.

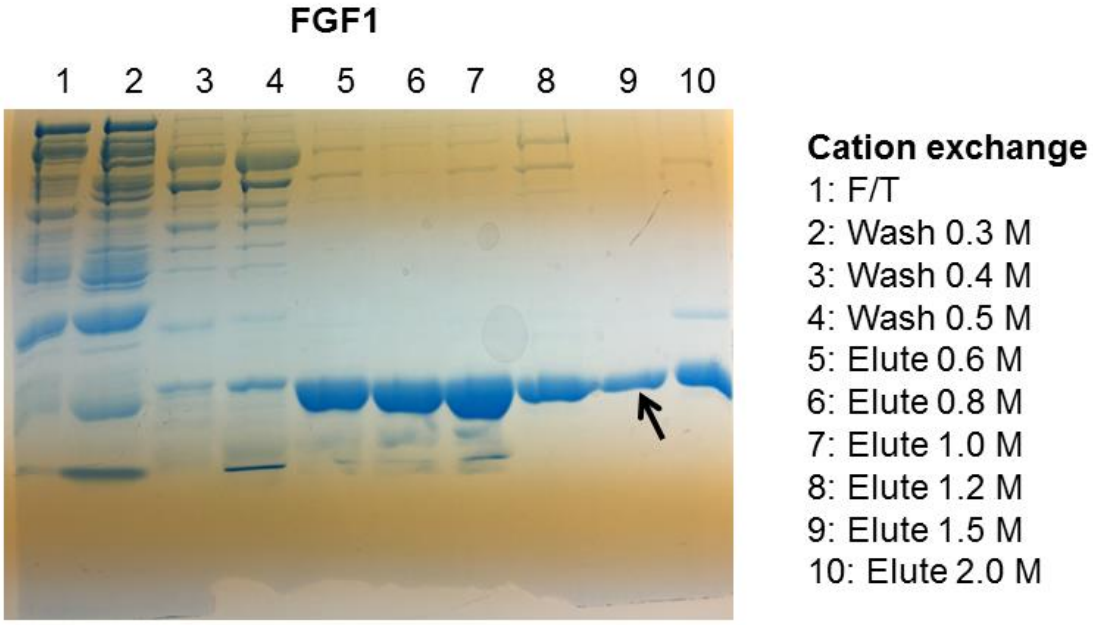


Figure 3.2: Expression and purification of His-FGF1. (His-FGF1 molecular weight: 19.1 kDa). His-FGF1 was first subjected to heparin-affinity chromatography and the eluate was diluted with 50 mM Tris-Cl, pH 6.8 and applied to a 1 mL HiTrap SP-HP. The column was then washed and eluted with phosphate buffered solutions (Section 2.4.1). Lane 1: flow through; lane 2: (0.3 M NaCl; lane 3: 0.4 M NaCl; lane 4: 0.5 M NaCl; lane 5: elution (0.6 M NaCl); lane 6: elution (0.8 M NaCl); lane 7: elution (1.0 M NaCl); lane 8: elution (1.2 M NaCl); lane 9: elution (1.5 M NaCl); lane 10: elution (2.0 M NaCl). Black arrow, band corresponding to the molecular weight of His-FGF1.

3.2.3. Purification of His-tag FGF-4

His-FGF-4 was produced using the plasmid described above (Section 2.2.1) from a culture of 800 µL x 4 flasks (Section 2.3.1). The protein from cell lysis was loaded onto a Affi-Heparin agarose column (Section 2.4.2) in Tris-buffer, and the eluate from this subjected to cation-exchange chromatography on a 1 mL HiTrap SP-HP column, which was developed using phosphate buffer (Section 2.4.3). The fractions collected from both columns along with samples corresponding to the starting material and unbound protein were analyzed by SDS-PAGE. FGF4 was eluted from heparin affinity column at the concentration NaCl of 1.5 M and 2.0 M (Fig 3.3A). Those two fractions were diluted in phosphate buffer and then applied to the cation-exchange column (Section 2.4.3). The results from cation-exchange chromatography showed that a band of ~19 kDa was present when the concentration of NaCl in the buffer was above 0.6 M (Fig 3.3B). The fractions with presence of pure protein were pooled and stored as 100 µL aliquots in elution buffer at -80 °C.

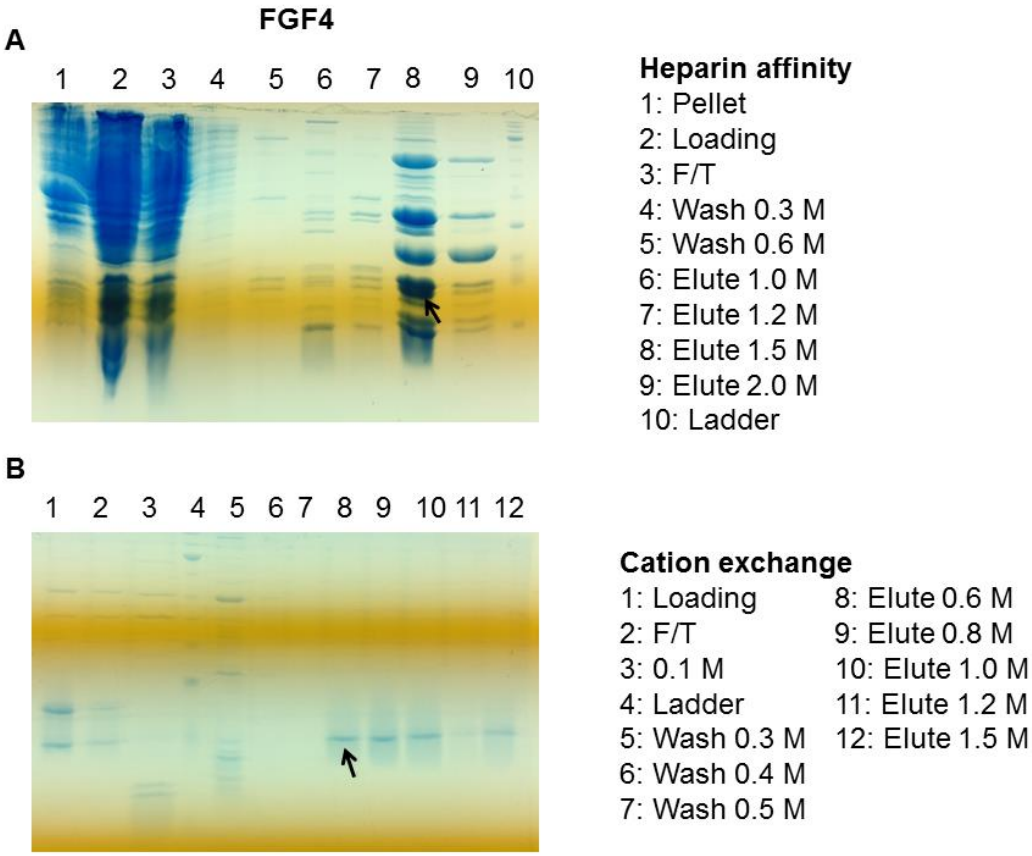


Figure 3.3: Expression and purification of His-FGF4. (His- FGF4 molecular weight: 19.2 kDa). **A)** Purification of FGF4 by heparin-affinity chromatography using Tris-Cl buffer (Section 2.3.2). Lane 1, Pellet; lane 2, sonicated whole cell lysate; lane 3, flow through after loading cell lysate onto the column; lane 4, wash fraction (0.3 M NaCl); lane 5, wash fraction (0.6 M NaCl); Lane 6, elution (1.0 M NaCl); Lane 7, elution (1.2 M NaCl); Lane 8, elution (1.5 M NaCl); Lane 9, elution (2.0 M NaCl). Lane 10, Ladder. **B)** Further purification of FGF4 by cation-exchange chromatography (Section 2.3.3). The eluate from heparin-affinity chromatography was diluted 20-fold in PB buffer, pH 6.8 and applied to a 1 mL HiTrap SP-HP column. Lane 1, load; lane 2, flow through; lane 3, wash fraction (0.1 M NaCl); lane 4, molecular weight markers; lane 5, wash fraction (0.3 M NaCl); lane 6, wash fraction (0.4 M NaCl; lane 7, wash fraction (0.5 M NaCl); lane 8, elution (0.6 M NaCl); lane 9, elution (0.8 M NaCl)); lane 10, elution (1.0 M NaCl)); lane 11, elution (1.2 M NaCl).); lane 12, elution (1.5 M NaCl). Black arrow, His-FGF4.

3.2.4. Purification of Halo-FGF6, His-FGF9 and His-FGF17

Halo-FGF6, His-FGF9 and His-FGF17 were all purified similarly to FGF4. Following heparin affinity chromatography, a major band was observed in the fractions eluting at the higher NaCl concentrations (Fig 3.4A, 3.5A and 3.6A). These fractions were diluted in 50 mM Tris-Cl, pH 6.8 to 0.1 M NaCl and applied to a 1 mL HiTrap SP-HP column. Following washing, elution fractions were collected and analysed by SDS-PAGE. Each of these FGFs was successfully purified in this way (Figs 3.4B, 3.5C and 3.6C).

3.2.4.1 Purification of FGF6

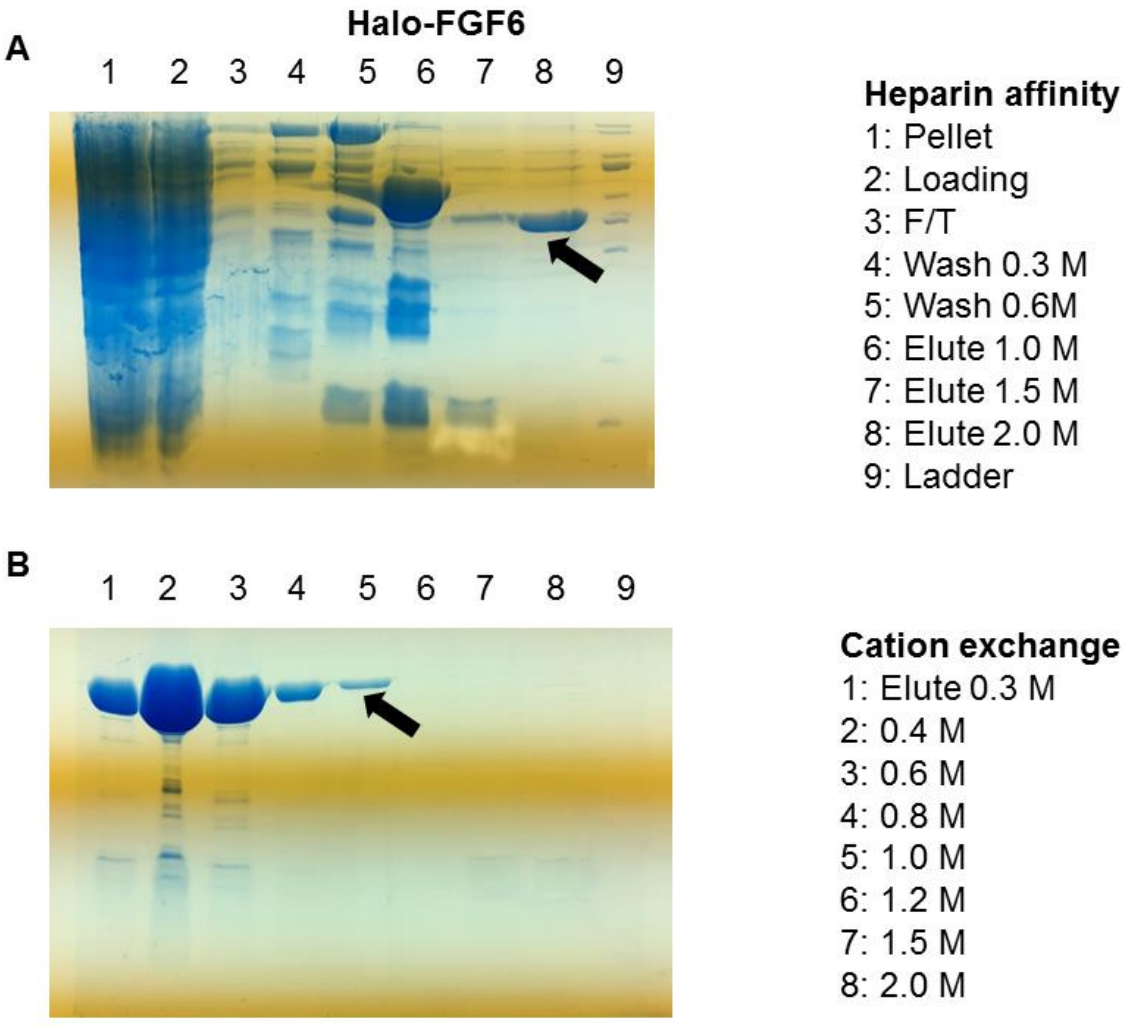


Figure 3.4: Expression and purification of Halo-FGF6. (Halo-FGF6 molecular weight: 51.1 kDa). A) Purification of Halo-FGF6 by heparin-affinity chromatography using 50 mM Tris-Cl pH 6.8 buffer (Section 2.4.2). Lane 1, pellet; Lane 2, sonicated whole cell lysate; Lane 3, flow through after loading cell lysate onto the column; Lane 4, wash fraction (0.3 M NaCl); Lane 5, wash fraction (0.6 M NaCl); Lane 6, elution (1.0 M NaCl); Lane 7, elution (1.2 M NaCl); Lane 8, elution (1.5 M NaCl); Lane 9: Ladder. **B)** Further purification of Halo-FGF6 by cation-exchange chromatography using Tris-Cl buffer (Section 2.4.3). Lane 1, wash fraction (0.3 M NaCl); Lane 2, wash fraction (0.4 M NaCl); Lane 3, wash fraction (0.6 M NaCl); Lane 4, elution fraction (0.8 M NaCl); Lane 5, elution fraction (1.0 M NaCl); Lane 6, elution fraction (1.2 M NaCl); Lane 7, elution fraction (1.2 M NaCl); Lane 8, elution fraction (1.5 M NaCl); Lane 9, elution fraction (2.0 M NaCl). Black arrow, Halo-FGF6.

3.2.4.1 Purification of His-FGF9

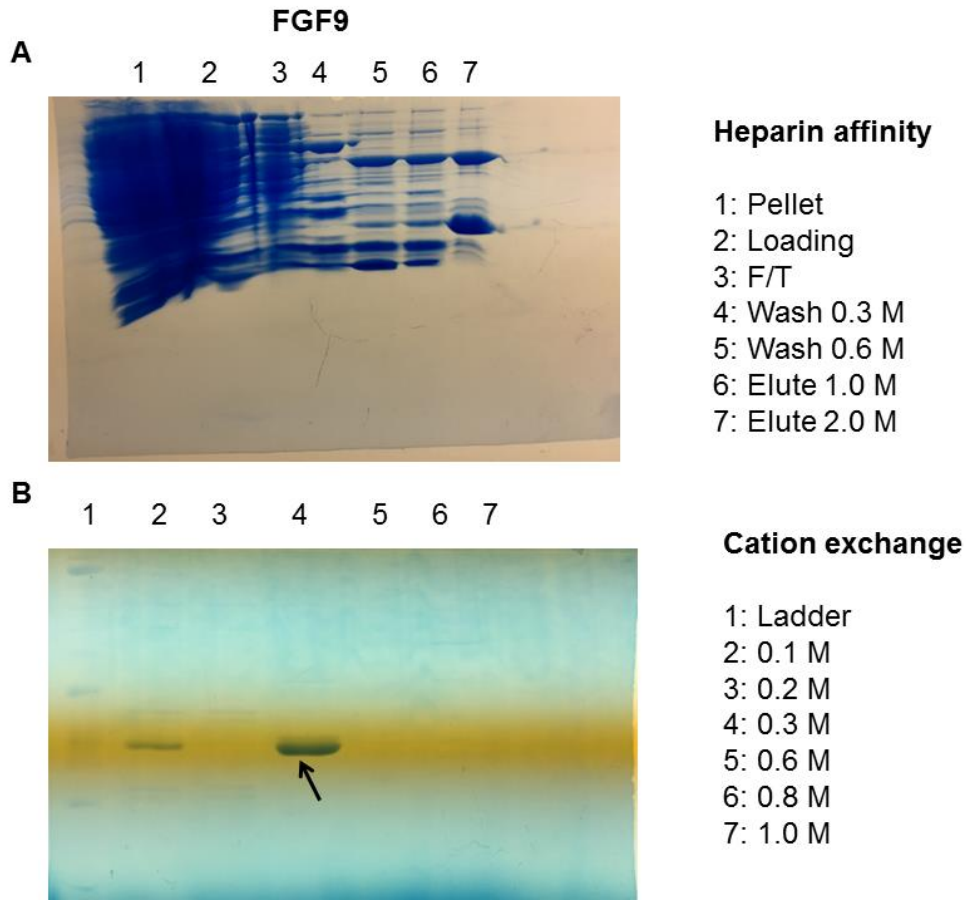


Figure 3.5: Expression and purification of His-FGF9. (His-FGF9 molecular weight: 23.4 kDa) A) Purification of His-FGF9 by heparin-affinity chromatography using 50 mM Tris-Cl pH 6.8 (Section 2.4.2). Lane 1, pellet; Lane 2, sonicated whole cell lysate; Lane 3, flow through after loading cell lysate onto the column; Lane 4, wash fraction (0.3 M NaCl); Lane 5, wash fraction (0.6 M NaCl); Lane 6, elution (1.0 M NaCl); Lane 7, elution (2.0 M NaCl). **B)** Further purification of His-FGF9 by cation-exchange chromatography (Section 2.4.3). After heparin-affinity chromatography, the sample was further purified on HiTrap SP HP column. The 2.0 M NaCl fraction was diluted 20 times then applied to a 1 mL Hi Trap SP-HP column. Lane 1, molecular weight markers; Lane 2, wash fraction (0.1 M NaCl,); Lane 3, wash fraction (0.2 M NaCl); Lane 4, washing fraction (0.3 M NaCl); Lanes 5 and 6, elution (0.6 M NaCl); Lane 7, elution (1.0 M NaCl). Black arrow, His-FGF9.

3.2.4.1 Purification of His-FGF17

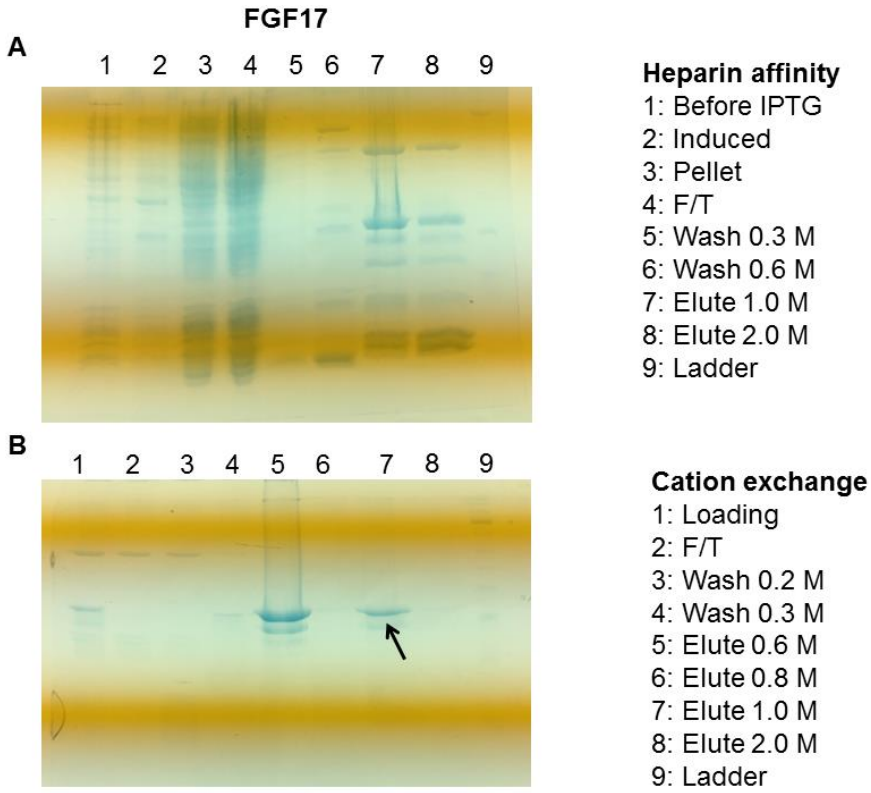


Figure 3.6: Expression and purification of His-FGF17. (His-FGF17 molecular weight: 23.3 kDa) A) Purification of His-FGF17 by heparin-affinity chromatography using 50 mM Tris-Cl, pH 6.8 buffer. Lane 1, cells before induction by IPTG; Lane 2, sonicated whole cell lysate; Lane 3, pellet; Lane 4, Flow through; Lane 5, wash fraction (0.3 M NaCl); Lane 6, wash fraction (0.6 M NaCl); Lane 7, elution (1.0 M NaCl); Lane 8, elution (2.0 M NaCl); Lane 9, molecular weight markers. **B)** Further purification of FGF17 by cation-exchange chromatography (Section 2.4.3). After heparin-affinity chromatography, the sample was further purified on a 1 mL HiTrap SP-HP column. FGF17 in the 1.0 M NaCl fraction following heparin affinity chromatography was diluted 10 times and then applied to the cation-exchange column. Lane 1, load after dilution; Lane 2, Flow through; Lane 3, wash fraction (0.2 M NaCl); Lane 4, wash fraction (0.3 M NaCl); Lane 5, elution (0.6 M NaCl); Lane 6, elution (0.8 M NaCl); Lane 7, elution (1.0 M NaCl); Lane 8, elution (2.0 M NaCl). Black arrow, His-FGF17.

3.2.5. Purification of His-FGF3, His-FGF16 and His-FGF18

His-FGF3, His-FGF16 and His-FGF18 were first purified from bacterial cell lysates by heparin affinity chromatography (Section 2.4.1). The fractions containing a major band (Fig 3.7A, 3.8A and 3.9A) were diluted in 50 mM Tris-Cl, pH 6.8 to 0,1 M NaCl and applied to a 1 mL Ni^{2+} affinity column (Section 2.4.3). Following washing, elution fractions were collected and analysed by SDS-PAGE, which demonstrates the successful purification of these FGFs (Figs 3.7B, 3.8C and 3.9C).

3.2.5.1 Purification of His-FGF3

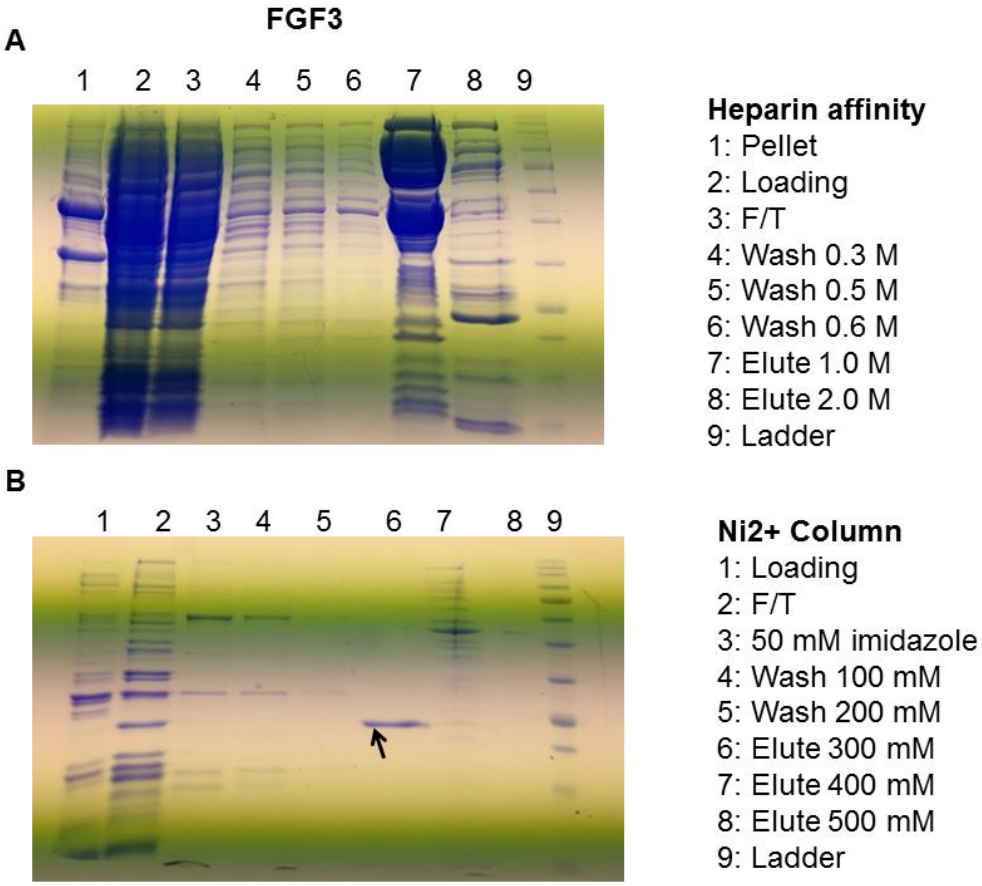


Figure 3.7: Expression and purification of His-FGF3. (His-FGF3 molecular weight: 26.89 kDa) A) Purification of His-FGF3 by heparin-affinity chromatography using 50 mM Tris-Cl buffer, pH 6.8. Lane 1: Pellet; lane 2: sonicated whole cell lysate; lane 3: flow through application of cell lysate onto the column; lane 4, wash fraction (0.3 M NaCl); lane 5, wash fraction (0.5 M NaCl); lane 6, wash fraction (0.6 M NaCl); lane 7, elution (1.0 M NaCl); lane 8, elution (2.0 M NaCl). Lane 9, molecular weight markers. **B)** Further purification of His-FGF3 by Ni^{2+} chromatography (Section 2.4.4). Lane 1: Load: Lane 2, Flow through; Lane 3, wash fraction (50 mM imidazole); Lane 4, wash fraction (0.1 M imidazole); lane 5, wash fraction (0.2 M imidazole); lane 6, elution (0.3 M imidazole); lane 7, elution (0.4 M imidazole); lane 8, elution (0.5 M imidazole); lane 9: molecular weight markers. Black arrow, His-FGF3.

3.2.5.1 Purification of His-FGF16

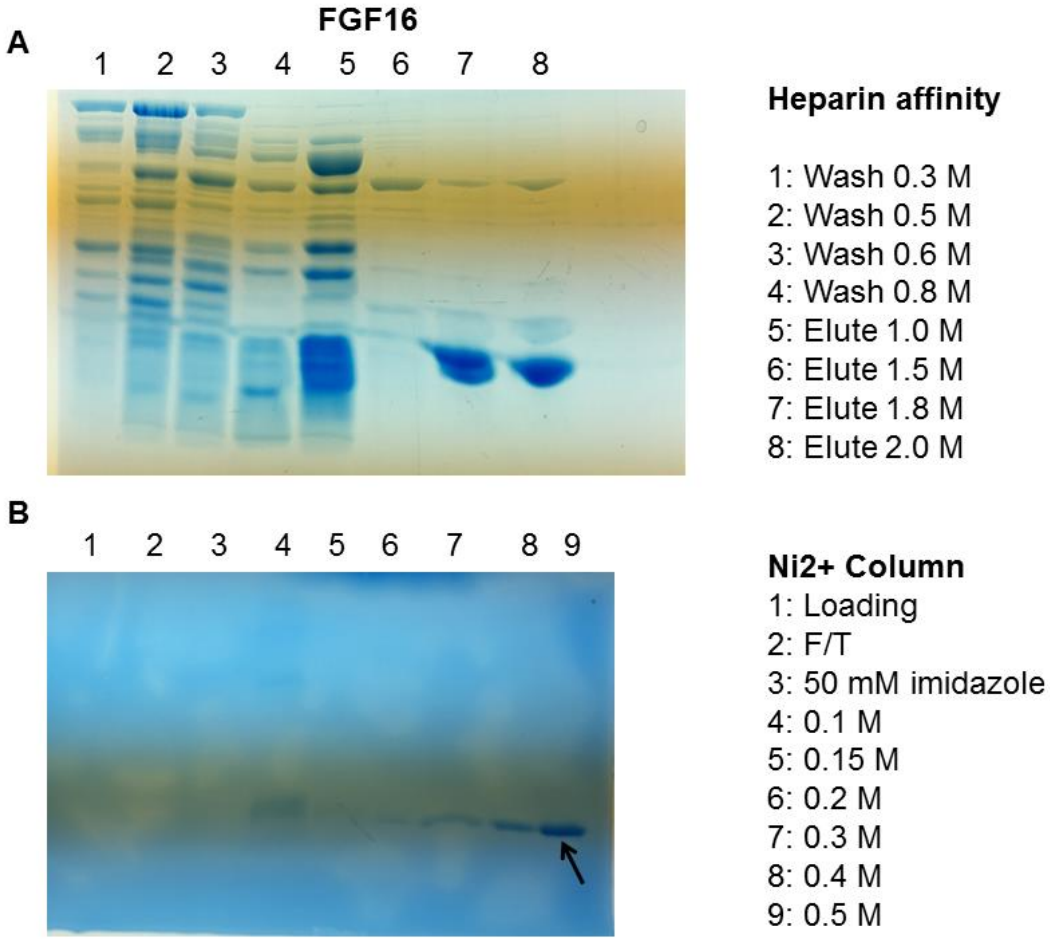


Figure 3.8: Expression and purification of His-FGF16. (His-FGF16 molecular weight: 21.4 kDa) A) Purification of FGF16 by heparin-affinity chromatography using 50 mM Tris-Cl pH 6.8 buffer. Lane 1, wash fraction (0.3 M NaCl); Lane 2, wash fraction (0.5 M NaCl); Lane 3, wash fraction (0.6 M NaCl); Lane 4, wash fraction (0.8 M NaCl); Lane 5, elution (1.0 M NaCl); Lane 6, elution (1.5 M NaCl); Lane 7, elution (1.8 M NaCl); Lane 8 elution (2.0 M NaCl). **B)** Further purification of His-FGF16 by Ni^{2+} chromatography (Section 2.4.4). Protein from the 1.8 M and 2.0 M NaCl fractions were combined, diluted 20 times and then applied to the Ni^{2+} column. Lane 1, Load: Lane 2 Flow through; Lane 3, wash fraction (50 mM imidazole); Lane 4, elution (0.1 M imidazole); Lane 5, elution fraction (0.15 M imidazole); Lane 6, elution fraction (0.2 M imidazole); Lane 7, elution fraction (0.3 M imidazole); Lane 8, elution fraction (0.4 M imidazole); Lane 9, elution fraction (0.5 M imidazole). Black arrow, His-FGF16.

3.2.5.1 Purification of His-FGF18

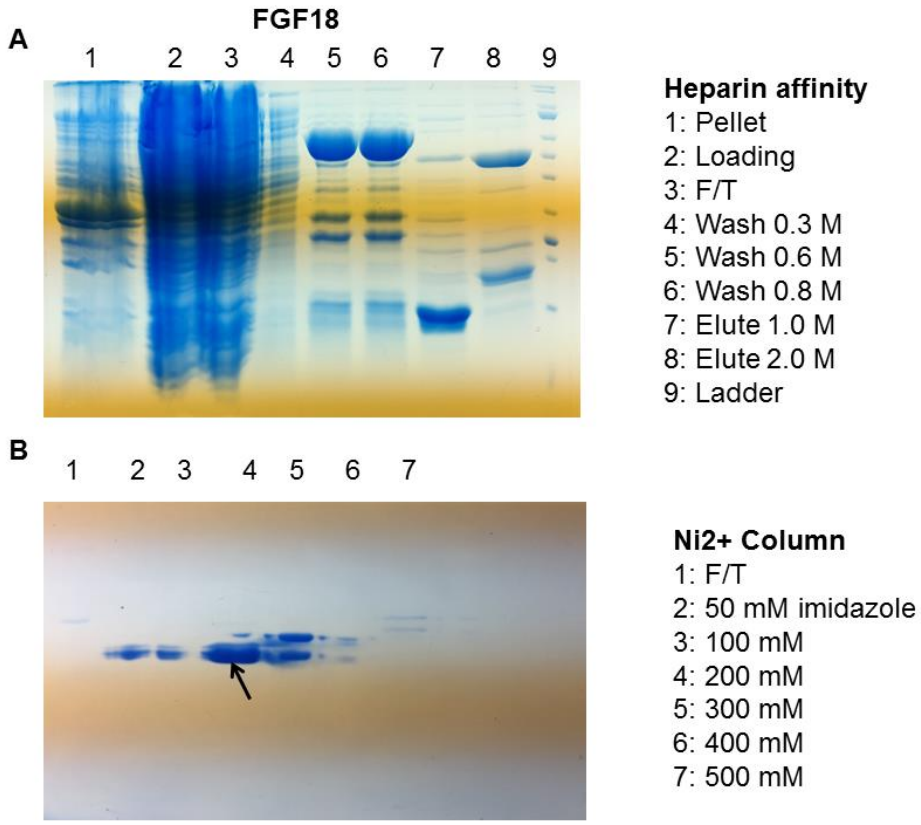


Figure 3.9: Expression and purification of His-FGF18. (His-FGF18 molecular weight: 24.03 kDa) A) Purification of His-FGF18 by heparin-affinity chromatography using 50 mM Tris-Cl buffer, pH 6.8. Lane 1, Pellet; Lane 2, sonicated whole cell lysate; Lane 3, Flow through; Lane 4, wash fraction (0.3 M NaCl); Lane 5, wash fraction (0.6 M NaCl); Lane 6, wash fraction (0.8 M NaCl); Lane 7, elution (1.0 M NaCl); Lane 8, elution (2.0 M NaCl); Lane 9, molecular weight markers. **B)** Further purification of FGF18 by Ni^{2+} chromatography. After heparin-affinity chromatography, FGF18 from fractions eluting between 0.6 M NaCl and 2.0 M NaCl was diluted 10 times before applying to the Ni^{2+} column. Lane 1, Flow through; Lane 2, wash fraction (50 mM imidazole); Lane 3, elution fraction (0.1 M imidazole); Lane 4, elution fraction (0.2 M imidazole); Lane 5, elution fraction (0.3 M imidazole); Lane 6, elution fraction (0.4 M imidazole); Lane 7, elution fraction (0.5 M imidazole). Black arrow, His-FGF18.

3.2.6. Purification of Halo-FGF5

Halo-FGF5 was produced using the plasmid described in Section 2.2.1. To optimize the amount of protein, the expression was conducted at 16 ^{o}C for 4 hours. After sonication (Section 2.4.1), the insoluble material was removed by centrifugation. There was no band around the expected molecular weight of Halo-FGF5 (~59 kDa) in the pellet (Figure 3.10A, lane 1), but a band was apparent in the supernatant (Figure 3.10A, lane 2). Upon application to a heparin affinity column (Section 2.4.2), this protein remained bound, since no equivalent band was detectable in the flow through fraction (Figure 3.10A, lane 3). However, the band corresponding to FGF5 was present in the elution fractions when the concentration of NaCl was increased to 1.0 M, although some contaminants were observed as well (Figure 3.10B, lanes 1 to 9). It could be seen that the elution range of the Halo-FGF5 was broad. There were two hypothesises for this phenomena. First, the amount of Halo-FGF5 was over the binding capability of the heparin column. Second, Halo-FGF5 molecules have different binding affinity to heparin due to their

folding in different conformations or to recognising with different affinities the different structures present in heparin, which is polydisperse (Section 1.3).

To remove the contaminants, Halo-FGF5 in the fractions of 1.2 M and 1.5 M NaCl were diluted to 0.05 M NaCl with 50 mM Tris-Cl, pH 6.8 and then applied to a 1 mL DEAE-Sepharose column (Section 2.4.3) for further separation. DEAE chromatography takes advantage of the acidic isoelectric point of the Halo-tag moiety of the fusion protein. No protein was observed in the load and flow through fractions, likely due to the dilution of the sample (Fig 3.10C lane 2, 3). A small amount of Halo-FGF5 could be seen in the 0.2 M NaCl fraction (Figure 3.10C, lanes 5 and 6). However, the majority of the Halo-FGF5 was eluted at 0.3 M NaCl (Figure 3.10C, lane 7) as a single band. There were no detectable polypeptides in the higher ionic strength eluates (Figure 3.10C, lanes 8 and 9). The concentration of protein in lane 7 was 1.2 mg/mL.

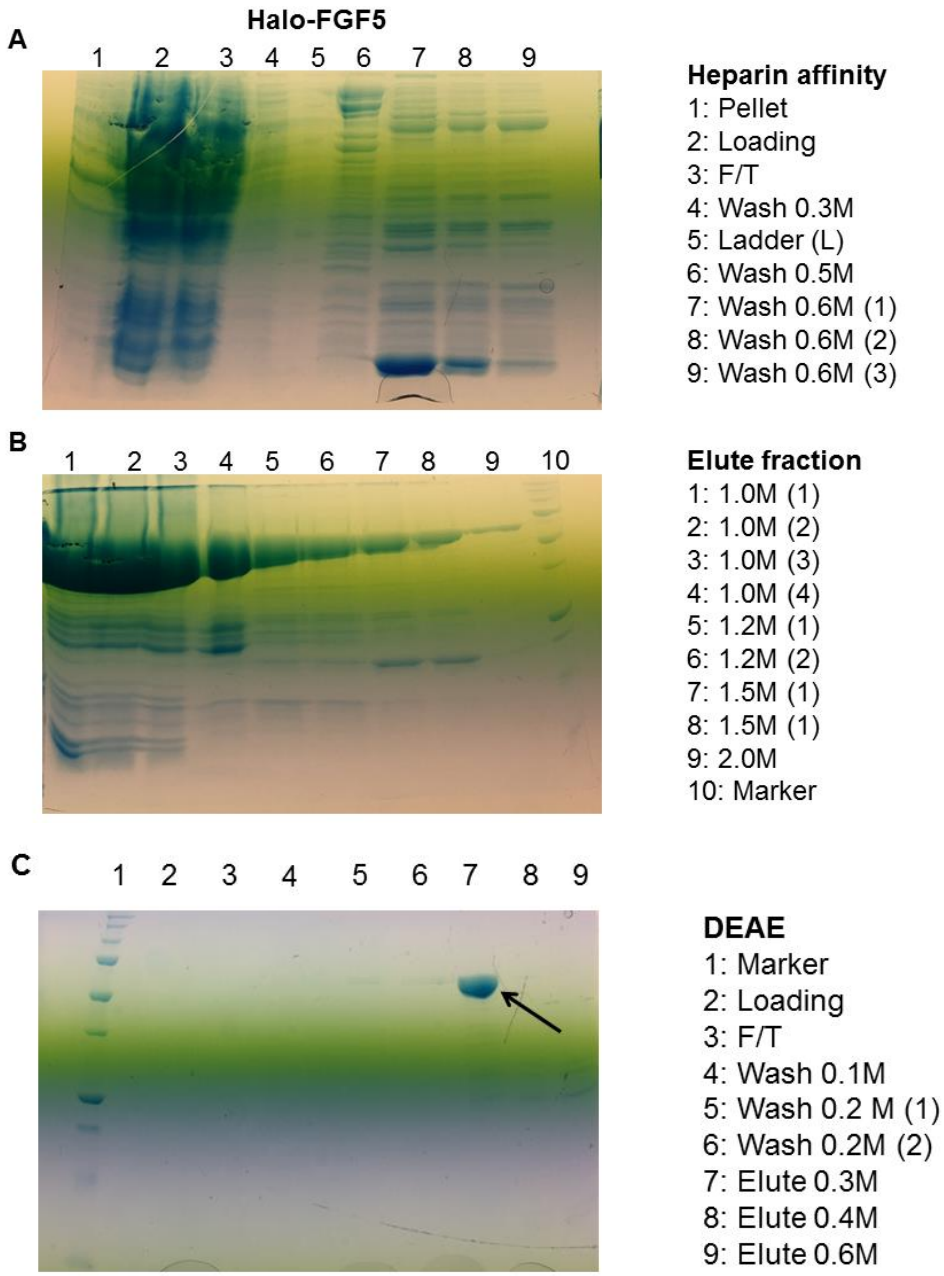


Figure 3.10: Purification of Halo-FGF5 (Molecular weight 58.95 kDa). A, B. Heparin affinity chromatography. **A**) Lane 1, Pellet; Lane 2, sonicated whole cell lysate; Lane 3, Flow through; Lane 4, wash (0.3 M NaCl); Lane 5, molecular weight markers; Lane 6, wash fraction (0.5 M NaCl); Lanes 7 to 9: wash fraction (0.6 M NaCl). **B**) Eluates. Lane 1 to 4, serial 1 mL elutions (1.0 M NaCl); Lanes 5 and 6, elution (1.2 M NaCl); Lanes 7 and 8, elution (1.5 M NaCl); Lane 9, elution (2.0 M NaCl); Lane 10 molecular weight markers. **C.** DEAE chromatography. Lane 1: molecular weight markers; Lane 2, load following dilution; Lane 3,

Flow through; Lane 4, wash (0.1 M NaCl); Lanes 5 and 6, wash (0.2 M NaCl); Lane 7, elution (0.3 M NaCl); Lane 8, elution (0.4 M NaCl); Lane 9, elution (0.6 M NaCl). Arrow: Halo-FGF5.

3.2.7. Purification of Halo-FGF8

The 500 mL cultures of transformed with the plasmid encoding Halo-tag FGF8 were cultured for 16 hours at 16 ^{o}C (Sections 2.3.1). After sonication and centrifugation (Section 2.4.1), there was no band around the expected molecular weight of Halo-FGF8 (~57.5 kDa) in the pellet (Figure 3.11A, lane 1). The applied sample and the flow through from the column contained a mixture of polypeptides, hence it was difficult to distinguish a band corresponding to Halo-FGF8 (Figure 3.11A, lane 2, 3). The wash using Tris buffer containing 0.3 M and 0.4 M NaCl (Section 2.3.2) remove some contaminants (Figure 3.11A, lane 4, 5 and 6). Two large bands were apparent in the 0.5 M NaCl eluate (Figure 3.11A, lane 7). However, those bands were not detected in the eluate of 0.6 M NaCl (Figure 3.11A, lane 8). When the concentration of NaCl was increased to 1.0 M, a single major band equivalent to the expected size of Halo-FGF8 was detected (Figure B, lanes 2, 3 and 4). However, contaminants were also present, so elution at higher concentrations of NaCl was performed. The fractions eluted at 1.5 M NaCl (Figure 3.11A, lane 7 and 8) and 2.0 M (Figure 3.11A, lane 9) had far less protein, but appeared to be sufficiently for use in experiments.

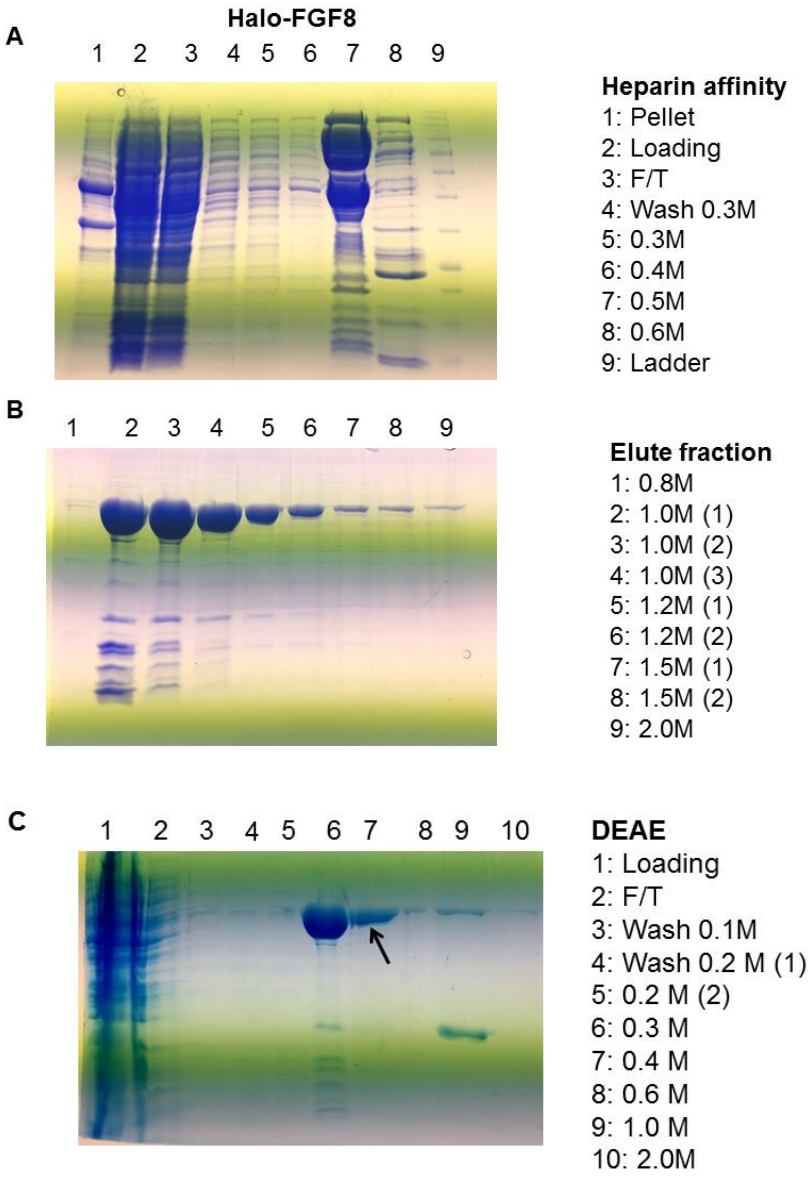


Figure 3.11: Purification of Halo-FGF8 (Molecular weight 57.5 kDa). A, B. Heparin affinity chromatography. A) Lane 1, Pellet; Lane 2, sonicated whole cell lysate; Lane 3, Flow through; Lane 4, wash fraction (0.3 M NaCl); Lane 5, washing fraction (0.3 M NaCl); Lane 6, wash fraction (0.4 M NaCl); Lane 7, wash fraction (0.5 M NaCl); Lane 8, wash fraction (0.6 M NaCl); Lane 9, molecular weight markers **B**) Eluates Lane 1, 0.8 M NaCl; Lane 2 to 4, 1.0 M NaCl; Lanes 5 and 6, 1.2 M NaCl; Lanes 7 and 8, 1.5 M NaCl; Lane 9, 2.0 M NaCl. **C)** DEAE chromatography (Section 2.4.4). Lane 1, load after dilution; Lane 2, Flow through; Lane 3, wash (0.1 M NaCl); Lanes 4 and 5, wash (0.2 M NaCl); Lane 6, elution (0.3 M NaCl); Lane

7, elution (0.4 M NaCl); Lane 8, elution (0.6 M NaCl); Lane 9, elution (1.0 M NaCl); Lane 10, elution (2.0 M NaCl). Arrow, Halo-FGF8.

There was a large amount of protein in the 1.0 M NaCl eluate from the heparin affinity column, which also contained substantial levels of contaminants, so it was decided to further purify the Halo-FGF8 using DEAE Sepharose, taking advantage once again of the acidic isoelectric point of the Halotag protein. The 1.0 M NaCl eluates from the heparin column were combined and diluted to 0.1 M NaCl with 50 mM Tris, pH 7.4 and applied to a DEAE Sepharose column (1 mL) (Section 2.4.4). After washing with 0.1 M NaCl (Fig 3.11C, lane 3), a step gradient, starting at 0.2 M NaCl (Fig 3.11C, lane 4 to 10) was used to elute proteins. A polypeptide corresponding to the expected size of Hal-FGF8 started to be eluted at 0.3 M NaCl (Fig 3.11C, lane 6) although other bands were detected in this fraction as well. The purest Halo-FGF8 was eluted with 0.4 M NaCl (Fig 3.11C, lane 7). There was a small amount of protein eluting at higher concentrations of NaCl when 0.6 M, 1.0 M and 2.0 M was loaded onto the column (lanes 8, 9 and 10).). The concentration of protein in lane 7 was 1.05 mg/mL.

3.2.8. Purification of Halo-FGF22

Halo-FGF22 was produced using the plasmid described in Section 2.2.1. Without the addition of IPTG, there was no band corresponding to the molecular size of Halo-FGF22 (~52.3 kDa) (Fig 3.12A, lane 1). When the culture condition was set at 37 ^{o}C for 1 h or 4 h, there was no protein corresponding to Halo-FGF22 detectable (Fig 3.12A lane 2 and 3). Hence, this protein was expressed at 16 ^{o}C for 4 hours as done for Halo-FGF5 (Section 3.2.11). After sonication and centrifugation (Section 2.4.1), the supernatant was applied to a heparin affinity column (Section 2.4.2). The pellet contained a mixture of proteins (Fig 3.12A, lane 4). The flow through fraction (Fig 3.12A, lane 5) and the two subsequent washes (0.3 M NaCl) (Fig 3.12A, lane 6), and 0.5 M NaCl (Fig 3.12A, lane 7) did not contain a detectable band corresponding

to Halo-FGF22. The 0.6 M NaCl and 0.8 M NaCl fractions contained a band corresponding to Halo-FGF22, as well as a number of other polypeptides (Fig 3.12A, lane 8 and 9).

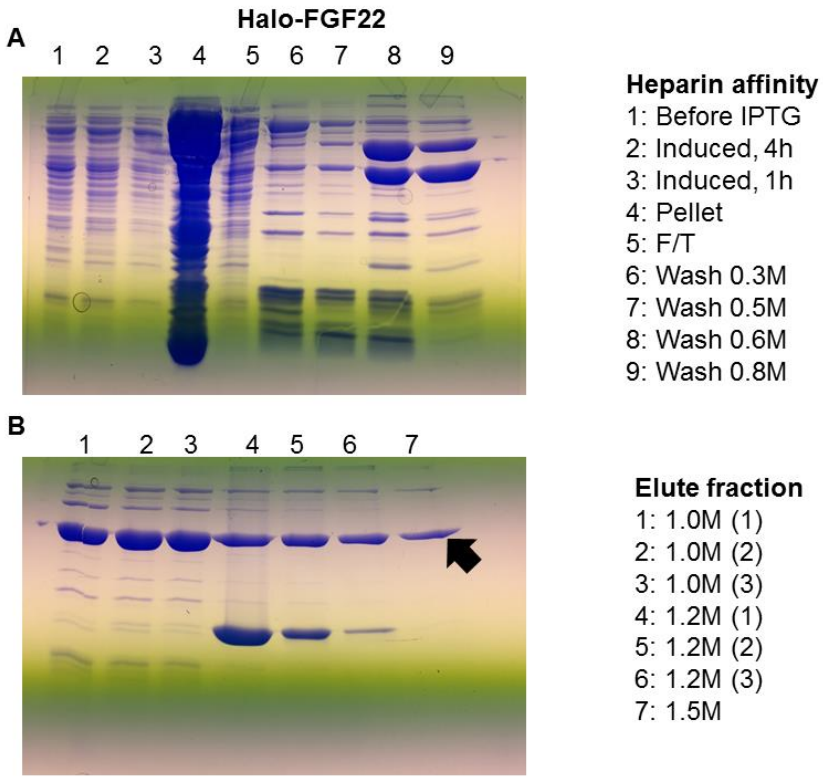


Figure 3.12: Purification of Halo-FGF22 (Molecular weight 52.3 kDa). A, B. Heparin affinity chromatography. A) Lane 1, Cells before the induction with IPTG; Lane 2, Cell lysate after 4 hours induction; Lane 3, Cell lysate after 1 hour induction; Lane 4, Pellet; Lane 5, Flow through; Lane 6, wash (0.3 M NaCl); Lane 7, wash (0.5 M NaCl); Lane 8, wash (0.6 M NaCl); Lane 9: wash (0.8 M NaCl) **B)** Elutions: Lanes 1 to 3, 1.0 M NaCl; Lanes 4 to 6, 1.2 M NaCl; Lane 7, 1.5 M NaCl. Arrow: Halo-FGF22.

At 1.0 M NaCl (Fig 3.12B, lane 1, 2 and 3), the lower molecular weight contaminants observed in the 0.6 M fractions (Panel A, lane 8, 9) were greatly reduced. At 1.2 M NaCl there were two major bands detected, including the band of Halo-FGF22 and one band of lower molecular weight (Fig 3.12B, lane 4, 5 and 6). At 1.5 M NaCl, (Lane 7), a single band at the molecular weight of the Halo-FGF22 was observed and it was judged to be sufficiently pure for use in experiments.

CHAPTER 4 SELECTIVE LABELLING OF ARGININE RESIDUES ENGAGED IN BINDING SULFATED GLYCOSAMINOGLYCAN

4.1 INTRODUCTION

Arginine is an abundant residue in proteins, some of them participate in electrostatic interactions with the negative charged patterns, some contribute to π-cation interactions, other may engage in the intramolecular network of hydrogen bonds maintaining the local structure of proteins. Moreover, the heparin binding sites of some proteins have arginines, but no lysine residues, e.g., fibroblast growth factor FGF22 (Uniprot ID: Q9HCT0-1). To establish their functions clearly, a method for selective labelling of arginine has been developed, taking advantage of mass spectrometry. Selective labelling was first carried-out for lysine due to the specificity of on the reaction of N-hydroxysuccinimide (NHS) with the amino side chain of lysine residues. However, NHS cannot react with the guanidino side chains of arginine residues, leading us to other potential chemicals for the modification of arginine amino acid. Eventually, PGO and hydroxyl-phenylglyoxal (HPG) were chosen to develop a method for selective labelling of arginines with the assistance of modern mass spectrometry and automated analysis of spectra to overcome the weakness of multiple reaction products of PGO and guanidino groups of arginine.

This approach is initially applied on two well-known FGFs, FGF1 and FGF2, to distinguish the arginine residues which form electrostatic interactions with sulfate groups of sugar chain from the one not involved in binding. The condition and analysis were optimized. The results were additionally verified using the previous works on interaction of FGF1 and FGF2 with heparin.

4.2 THE CHEMICAL MODIFICATION OF ARGININE

4.2.1 Arginine modification, reagents, and conditions

***Reaction mechanisms**:* There are limited works on the chemical modification of arginine residues in proteins and the majority of them based on the reaction of vicinal dicarbonyl compounds including phenylglyoxal, p-hydroxyphenylglyoxal, 2,3-butanedione, 1,2-cyclohexanedione, and methylglyoxal with guanidine group to form cyclic adducts [158].

The arginine-specific reactions are commonly carried out in 25°C over a pH range of 7–10 with the wide range of buffer and reaction times ranging from 15 min to 24 hours. The rate of the reaction is relatively slow under alkaline conditions and the products are reversible in the presence of diones [159]. Universally, dicarbonyl compounds react with arginine in the dark to avoid possible photoactivation of these molecule, which could enhance nonspecific reactions with groups other than arginine [160] [161] [162]. Considered as a whole, the accessible reactions conditions are almost ideal to study proteins under native conditions.

Choice of reagents: Several dicarbonyl compounds were considered for their suitability for the approach. As priority, the compounds must be specific for the guanidine side chain of arginine and avoid the side reactions with the side chain of lysine or histidine. Products have to be stable especially under mild acid condition (pH 2.0 of 0.1% TFA used in mass spectrometry) and enable unambiguous identification of peptides by mass spectrometry. Two reactants that are identical in terms of their reaction with arginine are required, which can be distinguished by mass spectrometry; ideally one will allow separation, e.g., using biotin-streptactin as with lysine directed protect and label.

There are four promising candidates to arginine specific ‘protect and label’: Phenylglyoxal (PGO); 2,3 butanedione (BD); 1,2 cyclohexanedione (CHD) and p-hydroxyphenylglyoxal (HPG)

4.2.1.1 Phenylglyoxal

The ability of phenylglyoxal (PGO) to selectively modify guanidine group of arginine was first discovered by Takahashi [163] and has been utilized in numerous investigations, especially in enzymatic activity studies [164][165][166].

According to the model suggested by Takahashi, the ratio of PGO to arginine in the reaction determines the product. The reaction starts with the formation of an adduct of PGO with the guanidino group and is followed by the addition of a second PGO molecule (if sufficient reactant) to form the final product. Depending on the conformation and folding state of proteins, the degree of surface-accessibility of arginine changes and then affects the final products. A problem emerges in case of proteins containing a large number arginine residues such as creatine kinase and yeast hexokinase in which only one mol of [^{14}C]-PGO for each subunit of the protein was observed. A reason may be that those residues are difficult to access by the reagent.

A study on the effect of buffer and pH on the model reaction between PGO and L-arginine showed that the rate of the reaction increases with increasing pH from 7.5 to 11.5 and much faster in bicarbonate buffer than in borate, phosphate, or Tris buffer [167]. In addition, it has been confirmed that PGO does not react with α- NH_2 group of L-Arg, indicating that this reagent selectively modifies the guanidine side chain only [168].

Borate changes the final products of the reaction. The borate buffer is an acid-base buffer used in biochemistry to maintain the pH within a relatively narrow range from 8 to 10 [169]. On the other hand, this buffer may affect the degree of hydration of PGO. The remove of borate from the modified protein solution by the Sephadex column could lead to the regeneration of some modified arginine residues [170]. It indicates that borate buffer system is not suitable for the analysis by MS where the buffer should be removed before enzyme digestion.

The product of labelling arginine with PGO has been identified by MS in case of rat creatine kinase (CK) [171], ribonuclease A [168] implying the stability of the product. On the model of peptides, it has been shown that PGO treatment did not change the ESI charge distribution of the peptides

4.2.1.2 2,3 butanedione

Among the reagents for modification of arginine, 2, 3-butanedione (BD) is becoming increasingly popular. The first study introduced BD as an arginine-modifying reagent 40 years ago [172]. BD has been utilized to modify arginine residues in a wide range of proteins including penicillin acylase P06875 (PAC_ECOLX), carboxylpeptidase A P15085 (CBPA1_HUMAN). The study of Riordan [173] showed that a 150-fold molar excess of BD with the presence of borate buffer were required to totally inactive the peptidase activity of carboxypeptidase A in 30 min [173]. The optimal concentration of borate was 0.05 M, and increasing or decreasing the concentration of borate resulted in incomplete reaction of arginine residues.

However, the modification of arginine by BD has a number of serious disadvantages. Modification with monomeric BD is reversible. In addition, the constant presence of borate is required for stabilizing the reaction products which creates several final products causing additional inconveniences. It has been reported that in the absence of light, BD could modify certain Lysine [174]. The modification with the BD trimer may be accompanied by side reactions at free --SH and amino groups.

4.2.1.3 1,2 cyclohexanedione

The first work introducing 1,2 cyclohexanedione (CHD) as arginine modifying reagent was conducted in 1965 [162]. The final product was found to be stable in borate buffer (pH 8-9) but it slowly decomposed in neutral or alkaline solutions. At pH 9.0, the arginine residues are converted quantitatively into N^7, N^8- (l,2-dihydroxycyclohex-l,2-ylene) arginine. Stabilization

of the modification product in borate buffer and the high rate of the reaction were explained as the formation of a complex of borate ions with the vicinal cis-hydroxy groups of the product [175]. Completely modified lysozyme totally lost enzymatic activity in the presence of borate ions. The observation was interpreted as the consequence of the neutralization of a positive charge in the anion-binding site by borate ions. A similar phenomena was seen in case of BD.

However, the product of 2 molecules of CHD and 1 molecule of L-arginine in borate buffer has been observed [168].

4.2.1.4 p-hydroxyphenylglyoxal (HPG)

Junkova et al suggested a method using PHG to increase the extension of modifying reaction between arginines and the reagent [176]. In this study, the technique was applied on matrix protein Mason-Pfizer monkey virus (M-PMV) (M-PMV) containing five arginine residues. The MS result demonstrated that all five arginines were 100% modified. In addition, the peptide spectra were used to confirm the result, and all of the individual modified arginines (R10, R22, R55, R57, and R58) were found.

Another study used LC/MS with a LCQ-ESI ion trap instrument to estimate enzyme activity of amadoriase II toward fructosylglycine. It showed that chemical modification of amadoriase II with p-hydroxyphenylglyoxal, resulted in an inhibition of enzyme activity [177].

In general, HPG modifies arginine in a similar manner to PGO. However, the rate of the reaction is higher than the reaction of PGO and it increases considerably with a rise in the pH. Moreover, at pH 9-10 p-hydroxyphenylglyoxal forms a single product with the guanidino group of arginine, which is stable at this pH.

4.2.2 Choice of reagents and the challenges

Final choice of reagents: Phenylglyoxal and its close analog p-hydroxyphenylglyoxal were chosen for their high rate reactions with arginine under mild conditions as well as their high specificity and well-established stoichiometry making the identification of products by MS more convenient. Another advantage is the stability of the products in the absence of borate buffer and under mild acid condition. Noted that the MS is carried on at pH 2.0 to 3.0 with the presence of TFA.

Reaction conditions:

Arginine residues of proteins are reacted with PGO and HPG in 0.2M bicarbonate buffer (pH 9.5). The reaction is carried on the dark, at room temperature for 60 min. The reagents are used at 100 fold excess of proteins.

4.3 PAPER: SELECTIVE LABELLING OF ARGININE RESIDUES ENGAGED IN BINDING SULFATEDGLYCOSAMINOGLYCAN

Thao P. Bui, Yong Li, Quentin M. Nunes, Mark C. Wilkinson, David G. Fernig (2019). **Selective labelling of arginine residues engaged in binding sulfatedglycosaminoglycan**

Preprint: BIORXIV/2019/574947

Contributions:

Thao P. Bui: Production of FGF proteins, Protect and Label, Co-wrote the paper.

Yong Li: Develop the model of method on peptides (NA, FA). Edited paper.

Quentin M. Nunes: Edited paper.

Mark C. Wilkinson: Edited paper.

David G. Fernig: Conceived study and co-wrote the paper.

Selective labelling of arginine residues engaged in binding sulfatedglycosaminoglycans

Thao P. Bui[1], Yong Li[1], Quentin M. Nunes[2], Mark C. Wilkinson[1], David G. Fernig[1]

[1]Department of Biochemistry, Institute of Integrated Biology, University of Liverpool, Crown Street, Liverpool L69 7ZB, United Kingdom.

[2]Liverpool Pancreatitis Research Group, Department of Molecular &Clin. Cancer Medicine, Institute of Translational Medicine, University of Liverpool, 2nd Floor Sherrington Building, Ashton Street, Liverpool L69 3GE, United Kingdom.

Affiliations

Keywords: arginine, phenylglyoxal, p-hydroxyl phenylglyoxal, fibroblast growth factor, heparansulfate, heparin binding site.

Address for correspondence: David G.Fernig. Institute of Integrative Biology, Biosciences Building, Crown Street, University of Liverpool, Liverpool, L69 7ZB, UK.

Email: dgfernig@liv.ac.uk

http://www.liv.ac.uk/~dgfernig/

I- Abstract

The activities of hundreds of proteins in the extracellular space are regulated by binding to the glycosaminoglycan heparan sulfate (HS). These interactions are driven by ionic bonds between sulfate and carboxylate groups on the polysaccharide and the side chains of basic residues in the protein. Here we develop a method to selectively label the guanidino side chains of arginine residues in proteins that engage the anionic groups in the sugar. The protein is bound to heparin (a common experimental proxy for HS) on an affinity column. Arginine side chains that are exposed to solvent, and thus involved in binding, are protected by reaction with the dicarbonyl phenylgyoxal (PGO). Elution of the bound proteins then exposes arginine side chains that had directly engaged with anionic groups on the polysaccharide. These are reacted with hydroxyl-phenylglyoxal (HPG). PGO was found to generate three products: a 1:1 product, the 1:1 water condensed product and a 2:1 PGO:arginine product. These three reaction products and that of HPG had distinct masses. Scripts were written to analyse the mass spectra and so identify HPG labelled arginine residues. Proof of principle was acquired on model peptides. The method was then applied to the identification of heparin binding arginine residues in fibroblast growth factors (FGF) 1 and 2. The data demonstrate that four out of eleven arginine residues on FGF2 and five out of six arginine residues of FGF1 engage heparin. Our approach provides a rapid and reliable means to identify arginines involved in functional complexes such as those of proteins with heparin.

II- Introduction

In the extracellular space, interactions between proteins and the glycosaminoglycan heparan sulfate (HS) regulate activities of hundreds of the proteins [77]. These proteins include growth factors, cytokines, chemokines, morphogens, enzymes, enzyme inhibitors, receptors, and extracellular matrix proteins, forming the network of HS-binding proteins, termed the heparan sulfate interactome [178]. Binding to the polysaccharide may, for example, regulate ligand diffusion [125], [127], formation of signalling ligand-receptor complexes [130], [179] and enzyme activity [180].

The engagement of HS to protein often occurs on the surface or in shallow grooves of proteins with a major contribution from ionic interactions, due to the highly anionic nature of HS. These ionic interactions occur between the negatively charged sulfate and carboxyl groups on the polysaccharide chain with the positively charged residues, lysine, and arginine in proteins [114], [181]. Proteins may have more than one HS binding site, and the amino acids contributing to binding though adjacent on protein surface are usually not continuous in the amino acid sequences [36], [37], [126], [142], [184] [120].

The structural basis of the interaction of a protein with HS is often probed using the related polysaccharide, heparin, as an experimental proxy. Many approaches such as NMR spectroscopy [184], site-directed mutagenesis [185]–[187] and X-ray crystallography [100], [130], are low throughput. In the absence of a robust bioinformatics method to predict heparin binding sites in proteins, higher throughput experimental methods have been developed. These latter include a method to selectively label lysine residues involved in heparin binding, called “protect and label”, which was coupled with mass spectrometric identification of labelled peptides [120]. This was based on the reaction of N-hydroxysuccinimide (NHS) with the amino side chain of lysine residues and it has been applied successfully to a number of protein-heparin

interactions [36], [37], [188] and also to an electrostatically driven protein-protein interaction [189]. An interesting feature of this method is that it identifies canonical, high-affinity heparin binding sites in proteins, as well as lower affinity secondary binding sites [36], [37], [188].

NHS cannot react with the guanidino side chains of arginine residues, yet these are at least as important as lysine side chains in mediating the interactions of proteins with anionic partners such as sulphated GAGs [114]. Moreover, the heparin binding sites of some proteins have arginines, but no lysine residues, e.g., fibroblast growth factor FGF22 (Uniprot ID: Q9HCT0-1). Since the heparin binding sites (HBS) on proteins consist of residues not necessarily adjacent in sequence, the identification of arginines, as well as lysines engaged with heparin would provide a comprehensive picture of the ionic interactions.

The modification of arginine residues in proteins is a challenge, because the guanidino functional group of arginine has a pKa between 11.5 ~ 12.5, which makes it the most basic side chain in a protein and a poor nucleophile. There are limited studies on the chemical modification of arginine residues in proteins and the majority of them rely on the reaction of vicinal dicarbonyl compounds with the guanidino group to form cyclic adducts [158]. The ability of phenylglyoxal (PGO) to selectively modify the guanidino group of arginine was first discovered by Takahashi [190], and has been utilized since, especially in studies of enzyme activity [166], [191], [192]. In addition, it has been confirmed that PGO does not react with the α- NH_2group of lysine, indicating that this reagent selectively modifies the guanidino side chain rather than primary amines [168]. The reaction of PGO with the guanidino side chain of arginine is a quantitativereaction but produces several products which may have contributed to the lack of popularity of this approach. However, modern mass spectrometry combined with automated analysis of spectra should allow the deconvolution of multiple reaction products.

Therefore, we have used PGO and hydroxyl-phenylglyoxal (HPG) to develop a method for selective labelling of arginines that are directly involved in binding heparin. PGO was reacted with the arginine residues that are not involved in binding to heparin and so protect them, then HPG was used to selectively label arginine residues in binding sites. The PGO-arginine reaction has three products, and these were readily distinguishable by mass spectrometry. The method was initially tested on the model of peptides and then on two FGFs that have extremely well characterised heparin binding sites (HBSs), FGF1 and FGF2. The data demonstrated the ability of PGO/HPG to quantitatively and selectively modify the guanidino group of arginine residues on proteins. The arginine residues in the primary HBS of FGF1 and FGF2 were all selectively labelled by HPG. The selectively labelling by HPG of arginines in the secondary binding sites of FGF1 and 2 provides a full structural definition of ionic bonding of these sites to the polysaccharide. Interestingly, the data demonstrated the potential for competition by arginine residues on the HBS3 of FGF1 for the binding to HS and the FGF receptor (FGFR) tyrosine kinase [179].

III- Materials and Methods

Heparin-binding proteins

Recombinant human FGF1 (UniProt accession P05230 – residues 1-155) and FGF2 (UniProt accession P09038 – residues 134-288) were expressed in C41 Escherichia coli cells using the pET-14b system (Novagen, Merck, Nottingham, UK) and purified, as described previously [145].

SDS-PAGE and Silver stain

Samples were separated on 15% (w/v) polyacrylamide-SDS gels. Silver staining was done according to Heukeshoven and Dernick[193].

Selective protection and labelling of arginine side chains in HBSs of proteins using PGO and HPG (Fig 1A)

A step by step guide is available at *protocol.io* at the following link:

https://www.protocols.io/view/selective-protection-and-labelling-of-arginine-lys-qqmdvu6

Step 1: Binding

AF-heparin beads (Tosoh Biosciences GmbH, Stuttgart, Germany; binding capacity of 4 mg antithrombin III/mL) previously used in the lysine selective labelling protocol [120] were employed. A mini affinity column was made by placing a plastic air filter as a frit at the end of a P10 pipette tip (Star Lab Ltd., Milton Keynes, UK) and then packed with 20 μL AF-heparin beads. The mini-column was equilibrated with 4 × 50 μL 200 mMNaCl, 0.2 M $NaHCO_3$, pH 9.5 (Na-1 buffer). The buffer was dispensed slowly into the column using a 2 mL sterile syringe. A minimum of 10 μg FGF protein was loaded onto the column (generally, the loading capacity of FGFs to resin was estimated at 15 mg/mL). The loading was repeated 3 times to

ensure the binding between protein and heparin beads. After binding, the column was washed with 200 µL (4 x 50 µL) Na-1 buffer to remove any unbound protein.

Step 2: Protection of arginine side chains

PGO (Merck Ltd., UK, 97%) was used in the dark, as it is light sensitive. PGO was freshly prepared in 50 % (v/v) DMSO, 50 % (v/v) HPLC grade water at 1 M, which was then diluted to 0.5 M and then 0.2 M with 0.2 M $NaHCO_3$, pH 9.5. The pH was adjusted with 0.1 M NaOH to between 9.1 and 9.5 to ensure optimal reaction. The FGF loaded heparin mini column was rinsed with 30 µL 0.2 M PGO solution to exchange buffers. A further 30 µL PGO solution was added to the column and the bound protein was allowed to react for 60 min at room temperature in the dark. The reaction was quenched with 5 µL 0.1 % (v/v) trifluoroacetic acid (TFA) in water. The mini-column was then washed with 200 µL Na-1 buffer (4 x 50 µL). Bound proteins were eluted with 2 x 20 µL Na-2 buffer (2 M NaCl, 0.2 M $NaHCO_3$, pH 9.5) containing 0.1 % (w/v) RapiGest SF Surfactant (Waters, UK). The addition of surfactant was important to ensure protein recovery in this and subsequent steps, due to the increased hydrophobicity of proteins following PGO conjugation to arginine side chains.

Step 3: Labelling of Arginine side chain by HPG

The preparation of HPG was performed in the dark room, as it is even more light-sensitive than PGO, following a procedure identical to that used for PGO. The eluted protein was diluted with 400 µL 0.2 M $NaHCO_3$, pH 9.5 and concentrated on a 3.5 kDa MWCO centrifugal filter (Merk Millipore, UK) to a final volume of 70~80 µL. The reaction with HPG was performed by incubating 80 µL diluted protein with 20 µL 0.5 M HPG so that the final concentration of HPG in the reaction was 0.1 M. The pH was maintained at over 9.0. The reaction was performed for 60 min at room temperature in the dark and then was quenched with 5 µL 0.1 % (v/v) TFA in water.

Step 4: Sample preparation for Mass Spectrometry

Protein was buffer-exchanged by four cycles of dilution on 3.5 kDa MWCO centrifugal filters with 400 µL 10-fold diluted 0.2 M $NaHCO_3$, pH 9.5 containing 0.1 % (w/v) RapiGest and 3 cycles of dilution with 400 µL HPLC water containing 0.1% (w/v) RapiGest by centrifugation at 13200 rpm for 10 min. After freezing at -80 °C for 30 min, the sample was lyophilized for an hour.

Step 5: Incubation with proteases

Chymotrypsin/trypsin: The freeze-dried protein was dissolved in a mixture of 80 µL 25 mM NH_4HCO_3 and 10 µL 1 % (w/v) RapiGest (~ 0.1 % w/v in final solution) and heated at 80 °C for 10 min. The mixture was quickly centrifuged at 3200 rpm for 30 seconds before 5 µL 50 mM DTT was added (5 mM final concentration) and incubated for 15 min at 56 °C. After cooling the sample to room temperature, proteins were carbamidomethylated with 5 µL 0.1 M iodoacetamide (freshly made) for 30 min in the dark. Proteins were then digested overnight with chymotrypsin (Promega Ltd., UK) or trypsin (Sigma, UK) at a ratio of 1:100 (w/w).

Incubation with Arg-C

The dried sample was dissolved in 400 mM NH_4HCO_3, pH 7.8 (25 µL) and 45 mM DTT (2.5 µL), and incubated for 15 min at 56 °C. After cooling to room temperature, 75 µL incubation buffer (50 mM Tris-Cl, 5 mM $CaCl_2$, 2 mM EDTA, pH 7.8) was added to dilute the urea to 2.0 M. Arg-C protease (Promega, Southampton, UK) was freshly prepared in incubation buffer and then added to the protein solution at a ratio of 1:100 (w/w). Activation buffer 10X (50 mM Tris-Cl, 50 mM DTT, 2 mM EDTA, pH 7.8) was added to give a final concentration of 1X. The mixture was mixed gently and centrifuged briefly before allowing digestion to proceed overnight at 37 °C.

Step 6: Mass spectrometry for the identification of peptides

Peptides were concentrated by rotary evaporation to a final volume of 10 µL and desalted using C18 Zip-Tips (Millipore). C18 Zip Tips were first pre-wetted with 2 x 10 µL 100 % (v/v) acetonitrile and then pre-equilibrated with 2 x 10 µL 0.1% (w/v) TFA in water. The peptides were loaded on the Zip Tip, the loading was repeated 7 to 8 times to ensure binding. The Zip Tip was washed with 10 µL 0.1% (w/v) TFA. Finally, the peptides were eluted with 2 µL of 5 mg/mL α-cyano-4-hydroxycinnamic acid (CHCA, > 99 % purity, Sigma) in 50:50 acetonitrile/water with 1 % (v/v) TFA, directly on to a 96 spot MALDI (matrix-assisted laser desorption/ionisation) target plate.

Analyses were performed on a Synapt G2-Si (Waters, Manchester, UK) with MALDI source equipped with a frequency tripled Nd:YAG UV laser (λ = 355 nm), operating at 1 kHz. The spectrum acquisition time was 120 seconds, with 1 second scan rates, laser energy of 150 Au. The MS spectra were extracted by MASSLYNX v.4.1 (Waters, Manchester, UK) with the spectrum range from 500 Da to 4000 Da. The spectra were then processed using automatic peak detection, including background subtraction.

Identification of modifications of peptides

Protein Prospector (v 5.19.1 developed at the USCF Mass spectrometry Facility) and Peptide Mass (ExPASy) were used to predict the possible peptides after incubation of protein with an enzyme with the following parameters: enzyme, chymotrypsin or trypsin; maximum missed cleavages, 5; mass range, 500 to 4000 Da; monoisotopic; instrument, MALDI-Q-TOF (Fig 1-B1). The list of peptides after enzyme cleavage was filtered to remove the peptides without arginine residues. Because products from the reaction between PGO/HPG and arginine residue bring different additional masses to the peptides, the prediction of the peptide masses after the modification was achieved using a script, written in Python (version 3.5.3 released on January

17th, 2017, available at http://www.python.org), named "*PGO-HPG mass predictor*" (available at Github in following link: https://github.com/bpthao/PGO-HPG-mass-predictor_2). Based on the number of arginine residues in each peptide, "*PGO-HPG mass predictor*" script considers all possible reaction products of PGO and HPG with arginine, and generates a list of predicted masses of the modified peptides (Fig 1-B2).

PGO-HPG mass predictor used two inputs. First, the list of predicted peptides from the native proteins cleaved by enzyme from either Protein Prospector or Peptide Mass (ExPASy). The first file is composed of two columns, the native sequence of the peptide and the corresponding mass. Second, the file of modifications also comprises of two columns, the mass shift of each product and its descriptions. Using a loop, *PGO-HPG mass predictor* automatically adds up all potential mass shifts to each peptide with has arginine residues. All combinations of mass changes were covered. The output file has four columns: 1-the native sequence of the peptides; 2- the original mass; 3-the final mass after modifications; 4-the description of the reaction product(s), to facilitate.

The observed list is the original mass of FGF peptides after modifications (Fig 1-B3), which provided the mass and intensity of each peak. The match between predicted and the observed list was carried out with a second Python script, "Matchmaker", (available at Github in following link: https://github.com/bpthao/Matchmaker_2) with a mass difference tolerance set to 0.1 Da (Fig 1-B3), as recommend by Mascot.

The output is a list of matching peptides with the native sequence, the predicted mass with modifications, the specific modifications associated with arginine residues and the actual observed mass.

IV- Results and discussion

PGO reactions with arginine in model peptides

The reaction of PGO with arginine, though highly selective, has been demonstrated to result in several products, influenced by the ratio of reactants and the adjacent amino acids [167], [171], [191], [194], [195]. Thus, according to Takahashi [190], the ratio of PGO to arginine in the reaction determines the product (Fig 2A). The reaction starts with the formation of an adduct of PGO with the guanidino group (Fig 2A, product {1}), which can then reversibly release a hydroxyl group from glyoxal, resulting in the alternative product (Fig 2A, product {2}). Following this is the reactionof a second PGO molecule (if sufficient reactant) onto either a nitrogen of the guanidino group of arginine (Fig 2A, products {3}) or the carbonyl group of the first PGO, which may reversibly release a hydroxyl group from glyoxal to form the further products (Fig 2A, products {4}). All of the reactions are highly dynamic, in the presence of even or excess of PGO, so a mixture of all products is formed, though in solution, product {2} predominates over product {1}[190].

In contrast, the reaction of HPG with arginine only has one product [176]. Since two reagents are required for selective labelling, one to protect arginines not involved in binding and a second to label arginines engaged in non covalent bonds with the polysaccharide, it was important to not just optimise these reactions, but also to establish which products are formed following reaction with PGO. To determine these parameters, model peptides were used in the first instance. These peptides were specifically chosen to represent arginine residues in particular contexts. Peptide FA (WQPPRARI) has two arginine residues, separated by a single alanine, which will provide insight into any differences in a reaction that may occur due to proximity of arginines on the surface of a protein (Table 1). Peptide NA provides the means to generate an N-terminal arginine (p-Glu-LYENKPRRPYIL) by pre- digestion with trypsin and

so identify any influence of the adjacent free amine of the N-terminus and the impact of lysine on PGO reaction with arginine (Table 1B).

The reaction of PGO with arginine residues in peptide FA

First, three concentrations of PGO were tested with peptide FA, which contains two arginines, to identify the conditions required for the reaction to go to completion (Table 1). Two reaction times (5 min and 30 min) were used, each with three different concentrations of PGO. Peptide FA was unmodified after reaction with 100 mM PGO for 5 min or 30 min(Table 1). However, with 200 mM PGO the peak of the original FA peptide was not detected demonstrating that arginines in the peptide were fully reacted with PGO (Fig 2B). No peaks were identified when the peptide was reacted with 1 M PGO, which could be due to the aggregation of the peptide induced by the high concentration of PGO (Table 1). Thus, 200 mM was considered as the optimal concentration of PGO for the modification of arginine in this model peptide.

Interestingly, six reaction products of peptide FA with PGO were detected (Table 1A). The mass of unreacted FA is 1023.58 Da, and after reaction three major peaks of 1139.60 Da, 1255.61 Da and 1275.58 Da were detected, equivalent to a mass shift of 116.02 Da of Product {2} (Fig. 2A) (Peak 1, Fig. 2B, Table 1A); a mass shift of 232.04 Da (Product {4}, Fig. 2A; Peak 2, Fig 2B, Table 1A); and 252.0 Da (Product {3}, Fig. 2A; Peak 3, Fig 2B, Table 1A), respectively.

Peptide FA has two arginine residues and their reaction with PGO also produced three combinations of products. The combinations of product {2} and product {3} resulted in the mass of peak 4of 1388.62 Da (Fig 2B, Table 1A); of product {4} and product {3} resulted in peak 5 of 1410.60 (Fig 2B, Table 1A); finally, when both arginines formed product {3} with PGO, a mass shift corresponding to peak 6 of 1523.67 was observed (Fig 2B) (Table 1A).

The reaction of PGO with argininesin peptide NA

It is clear from the above that the position of arginine resides can affect the reaction product. Since the side chain of an arginine at the terminus of a peptide has more steric freedom, it is more likely that the dicarbonyl group of PGO could form a reversible bond at a 2:1 ratio with the guanidino group of arginine, as described in Takahashi's study [190]. To establish if this was the case, two peptides: peptide NA-I (RPYIL, molecular mass 661.38) and peptide NA-II (LYENKPR, molecular mass 1030.53) were produced by treating peptide NA with trypsin (Table 1B). This yielded one peptide with arginine at the N terminus (peptide NA-I) and another one with arginine at the C terminus (peptide NA-II). Both peptides were reacted with 200 mM PGO in the dark for 30 min (Figs 2 C,D).

For peptide NA-I, the product of this peptide with one PGO had amass shift of 116.02 Da (Product {2}, Fig. 1A; Peak 1, Table 1B) resulting in the product of 777.42 Da (Fig 2C), while, the peak at 795.40 Da (Fig 2C) indicated the presence of product {1} from the reaction between arginine and PGO. The addition of a second PGO resulted in the product of 896.41 Da, giving a mass shift of 232.03 Da (Product {4}, Fig 2A, Table 1B) and the product of 911.1 Da due to a mass shift of 250.04 Da (Product {3}, Fig 2A, Table 1B).

In a similar manner, the products of peptide NA-II with PGO showed the mass shifts of all four products between arginine and PGO (Fig 2D, Table 1B). The peak of 1146.55 Da (Table 1B) corresponds to the mass shift of 116.02 Da (Product {2}, Fig 2A); the mass shift due to product {1} (Fig 2A) is detected as the peak at 1164.52 Da (Fig 2D, Table 1B); the product of two PGO reacting with one arginine are observed as the peak at 1262.58 Da (Fig 2D, Table 1B) of product {4} and the peak at 1280.56 (Fig 2D, Table 1B) was due to product {3}. In addition, for the mass shifts observed, only the arginine of peptide II was modified by PGO and there

was no reactivity of its lysine or N-terminus towards PGO, in agreement with previous work [168].

The use of peptides with arginine residues in defined contexts demonstrated that 200 mM PGO was likely to fully protect all of the arginine residues in a protein regardless of the adjacent sequence. Moreover, the position of arginine in the peptides affected the formation of the Schiff's base and types of products produced. However, these are resolvable by mass spectrometry. The next step was to determine whether HPG could similarly fully react with arginine residues.

HPG reactions with arginine in model peptides

For the labelling step, HPG, which has the same dicarbonyl moiety as PGO, was chosen. In general, HPG modifies arginine in a similar manner to PGO, but the rate of the reaction is faster than that of PGO and it increases with pH, where pH 9-10 is optimal [176]. HPG forms a single product with the guanidino group of arginine, which is stable at this pH (Fig 3A). The two peptides FA and NA were used again to validate the reactivity of HPG toward to guanidino group of arginine. HPG was dissolved in DMSO at 1 M and diluted stepwise to 500 mM and then 100 mM with 0.2M $NaHCO_3$ pH 9.5 and then reacted with the peptides FA and NA for 10 and 30 min. Peptide FA was not entirely modified after reaction with 100 mM HPG for 10 min (Fig 3B), as evidenced by the presence of two major peaks of unmodified and HPG-modified peptide FA, indicating that this time was too short for the reaction to go to completion. However, 100 mM HPG for 30 min modified all of the arginines in the FA peptide (Fig 3B), since the peak of the unmodified peptide FA was no longer detected. In a similar manner, HPG at 100 mM after 10 min reaction did not modify fully the arginines of peptide NA, evidenced by the observation of two major peaks of unmodified and HPG-modified peptide NA, whereas after 30 min, they were fully modified, as only one major peak was seen (Fig 3C). These

reactions with the model peptides demonstrated that HPG modifies arginine to form a single product. The difference in the reaction between HPG and arginine compared to that of PGO may be due to the presence of the hydroxyl group in HPG.

Because the model peptides used here contained a small number of arginine residues in a reasonably simple environment, it was not known whether these reaction conditions would be applicable to a protein. Thus, the next issue to address was the ability of PGO to modify all arginine side chains in a protein. Arg-C was used to digest the protein after modification and so identify the unmodified arginine residues. These experiments used FGF2, which contains eleven arginine residues, as a model protein (Fig 4).

The reaction of FGF2 with PGO

The reaction between FGF2 and PGO was first conducted in the absence of heparin beads. After buffer exchange, half of the protein was reacted with 200 mM PGO in the dark, at 25^0C for 60 min (Material and Methods, step 2) and the other half was used as a control. Native FGF2 migrates as a band at round 18 kDa on SDS-PAGE (Fig 4A lane 1), but after digestion by chymotrypsin or by Arg-C no bands were apparent, demonstrating the expected cleavage of the protein by these enzymes (Fig 4A lanes 3 and 5). Reaction with PGO did not change appreciably the migration of the FGF2 (Fig 4A lane 2, PGO treated FGF2, no enzyme), and the modified protein was cleaved by chymotrypsin (Fig 4A lane 4). In contrast, PGO-modified FGF2 was not cleaved by Arg-C, since a band corresponding to FGF2 was clearly visible (Fig 4A lane 6). This suggests that all arginine residues of FGF2 were modified.

When FGF2 (Fig 4B- L) was loaded onto the heparin mini-column, no protein was observed in the flow through (Fig 4B- lane FT) and wash fractions, indicating that FGF2 bound as expected (Fig 4B- lanes FT and W). After the on-column reaction with 200 mM PGO, the excess PGO was removed (Fig 4B- lane PGO PROTECT FT) and the column was washed by

Na-1 buffer (Fig 4B- lane PGO PROTECT W) before elution with Na-2 buffer (Fig 4B- lane PGO PROTECT E) (Materials and Methods, step 2). No protein was detected in the PGO PROTECT FT and PGO PROTECT W fractions, indicating that the FGF2 remained bound to heparin during the reaction. A band was observed in the elution fraction (Fig 4B-lane PGO PROTECT E), which was similar in size and amount as the loading control (Fig 4B lane L). This result indicated that protein was efficiently eluted from the column. A quarter of the protein in the eluted fraction was incubated overnight with Arg-C or chymotrypsin and the products of digestion were analysed by SDS-PAGE. In both cases, no band was apparent, indicating that the enzymes had cleaved the PGO reacted FGF2 (Fig 4B PGO PROTECT lanes 1 and 2). Thus, when FGF2 is reacted with PGO in solution, there is a complete modification of arginine residues, but when the reaction is performed on FGF2 bound to heparin, the modification is incomplete. This suggests that only arginine residues exposed to solvent in heparin bound FGF2 were able to react with PGO.

The remaining half of the eluted protein (~20 µL) was diluted with Na-1 buffer until the final concentration of NaCl was less than 0.2 M and then reacted in solution with PGO at 100 mM final concentration. Half of this product was applied onto a mini heparin affinity column. The flow-through fraction contained a band of almost identical intensity to the reaction product (Fig 4C, lane FT) and there was no protein detected in the wash (Fig 4C, lane W) and eluted fractions (Fig 4C, lane E) indicating that the FGF2 no longer bound to heparin. Thus, after the second reaction (equivalent to the labelling step in the original lysine selective protect and label [120]) the arginine residues in HBSs were blocked by PGO. The FGF2 reacted with PGO a second time and recovered from the flow through fraction was also probed with proteases. While chymotrypsin digested the FGF2 (Fig 4C lane 1), Arg-C was unable to do so, since there was a band (Fig 4C lane 2) at the same size as the initial reaction product (Fig 4D -L) and labelled

FGF2 (Fig 4C - FT). Hence, the arginine residues protected after the initial on-column reaction with PGO were successfully blocked in the second reaction of this FGF2 with PGO in solution.

These data suggest that under the reaction conditions used PGO successfully modified the side chains of all 11 arginine residues of FGF2, and that following PGO modification of those arginine residues that remain exposed when FGF2 binds to heparin, the remaining arginine residues on the eluted protein may also be successfully modified.

The reaction of FGF1 with PGO

The same series of experiments were repeated with FGF1. In the absence of heparin beads, after buffer exchange, half of the protein was reacted with 200 mM PGO in the dark, at room temperature for 60 min (Material and Methods, step 2) and the other half was used as a control. The expected cleavage of the native FGF1 (Fig 4D lane 1) by chymotrypsin or by Arg-C was observed, evidenced as no detectable bands (Fig 4D lanes 3 and 5). Whereas the PGO-modified FGF1 was cleaved by chymotrypsin (Fig 4D lane 4), it was not cleaved by Arg-C, since a band corresponding to FGF1 was clearly visible in this case (Fig 4D lane 6). This suggests that all arginine residues of FGF1 were modified.

When FGF1 (Fig 4E- L) was loaded onto the heparin mini-column, no protein was detected in the flow through (Fig 4E lane - FT) and wash fractions,nor when PGO was applied (Fig 4E - PGO PROTECT FT and PGO PROTECT W lanes), indicating that the FGF1 remained bound to heparin during the reaction. A band was observed in the elution fraction (Fig 4E-lane PGO PROTECT E) lane, which was similar in size and amount as the loading control (Fig 4E L lane). This result indicated that protein was efficiently eluted from the column. When a quarter of the PGO reacted FGF1 was incubated overnight with Arg-C or chymotrypsin, no band was apparent (Fig 4E PGO PROTECT lanes 1 and 2), demonstrating that there remained unreacted arginine residues. After a second reaction of eluted protein with PGO at 200 mM final

concentration, chymotrypsin digested the FGF1 (Fig 4F lane 1), but Arg-C was unable to do so, since there was a band (Fig 4F lane 2) at the same size as the initial reaction product (Fig 4F, lane L) and FGF1 reacted with PGO in solution (Fig 4F, FT lane). Moreover, this eluted FGF1 subjected to a second reaction in solution with PGO failed to bind to heparin column (Fig 4F, lane FT). Hence, the arginine residues engaged with heparin in the initial on-column reaction with PGO were successfully reacted in the second reaction of this FGF1 with PGO in solution.

Protect and Label strategy for the identification of arginine residues in FGF1 and FGF2 that bind heparin

There is a large body of published structural, biophysical and biological data relating to the interactions of FGF1 and FGF2 with heparan sulfate and its experiment proximal heparin. FGF1 and FGF2 share a high level of similarity in structure and sequence, though they possess very different isoelectric points, 6.52 and 11.18, respectively. FGF1 and FGF2 were loaded on to AF-heparin mini columns and reacted *in situ* with 200 mM PGO (Fig 1) (Material and Methods, step 1). The eluted proteins were then reacted for 60 min with 100 mM HPG in the dark (Material and Methods, step 3) and processed for mass spectrometry (Material and Methods, step 4). In parallel, FGF1 and FGF2 were reacted with 200 mM PGO in the absence of heparin beads (Material and Methods, step 2). The native and modified proteins were then cleaved by chymotrypsin (Material and Methods, step 5).

Peptides produced from digestions were predicted using two protein identification and analysis tools, Protein Prospector (v 5.19.1 developed by the USCF Mass spectrometry Facility) and Peptide Mass (ExPASy). In both cases, these provided parameters included the mass range from 500 Da to 4000 Da, maximum missed cleavages 5, monoisotopic only and the enzyme

used. For Protein Prospector, the oxidation of methionine was considered as the variable modification, whereas it was not included by Peptide Mass (ExPASy).

To identify the reaction products of the arginine side chain and PGO, their structures and masses were evaluated (Table 2). Depending on the stoichiometry of the reactions and the loss of water, there were four possible products: 1:1 (1PGO: 1arginine), (Product 1- Fig 2A); 1:1 water-condensed, (Product 2- Fig 2A) and 2:1 product water-condensed, (Product 3, 4- Fig 2A). Product 3 and 4 had different structures, but they led to the same mass shift for the reacted arginine residue. The corresponding mass shifts resulting from these products are provided in Table 2. The 250 kDa product (Supplementary Fig 2C) was not considered, as neither of the two FGFs possesses a terminal arginine. To identify arginine residues protected by heparin binding, a second reagent was required, that would react similarly with arginine side chains, but yield a product with a different mass. HPG was chosen for this purpose, and this yields just a single, 1:1 product (Table 2, Fig 3).

Peptides from the native and modified proteins produced by enzyme cleavage were analysed by MALDI-Q-TOF mass spectrometry. Following was the analysis of the mass spectra using scripts *PGO-HPG mass predictor* and *Matchmaker*.

Identification of arginine residues in FGF2 involved in binding heparin

FGF2 has eleven arginine residues in the sequence, R^{31}, 42, 48, 53, 69, 81, 106, 116, 118 and 129. The peptides generated from native FGF2 by cleavage with chymotrypsin were predicted by Peptide Mass and Prospector, then filtered with the script PGO-HPG mass predictor to remove peptides without arginine residues. Prospector predicted 158 peptides and there were 126 peptides predicted by Peptide Mass with 100% sequence coverage, demonstrating that all arginine residues could be analysed.

Protection of arginine residues on FGF2 by PGO in solution

To understand the accessibility of the reagent to arginine, FGF2 was reacted with 200 mM PGO for 60 min in 0.2 M $NaHCO_3$, pH 9.5. Peptides of the modified FGF2 were generated by cleavage with chymotrypsin and analysed by mass spectrometry. The modification on each arginine residue was identified by *PGO-HPG mass predictor* and *Matchmaker*. The resulting peptides with information about modification, sequence, and final m/z are presented in Supplementary Tables 1 and 2 and the spectra are in Supplementary Fig 3.

Among peptides predicted by Prospector, nine of them contained the reacted arginine residues (Supplementary Table 1, Supplementary Fig 3). PGO product 1 (Fig 2A) was observed in peptide 2 with R^{129}, peptides 4, 5 and 8 with R^{116} and R^{118}, peptide 7 with R^{69} and R^{81}. R^{90} of peptide 3 reacted with PGO to generate product 2 (Fig 2A). Peptide 1 with R^{31}, peptide 6 with R^{106} showed a mass shift corresponding to products 3, 4 (Fig 2A). A mixture of products was found in peptide 9 with R^{42}, 48 and 53, as a combination of products 1 and 3. It was noticed that only two out of three arginines on peptide 9 reacted with PGO, indicating that one arginine was apparently not accessible to PGO in this context. Hence, the products of ten arginine residues with PGO were identified, though a lack of reactivity of one arginine to the reagents was also observed.

Using the predicted peptides from ExPASy, the number of peptide with modified arginine residues was six, which completely overlapped with the list generated by Prospector (Supplementary table 2).

Selective labelling of arginine residues on FGF2 by PGO and HPG on heparin mini column

PGO-HPG mass predictor and Matchmaker scripts were used to identify the peptides containing the arginine residues that are labelled by HPG and therefore engaged with heparin.

The resulting peptides with information about modification, sequence, and final m/z are presented in Tables 3 and 4 and the spectra are in Supplementary Figs S4 and S5).

After the protection step on the heparin affinity column with PGO and labelling with HPG ten peptides with modified ariginines were identified (Table 3) (Supplementary Fig 4). Of these, four peptides (peptides 4, 8, 9 and 10) contained PGO protected arginine residues only; peptides 1, 2, 3, 6 and 7 had a mass shift corresponding to the product of HPG (132.02 Da) whereas, in peptide 5, a mixture of PGO and HPG products was observed (Table 3).

The first arginine on the sequence is R^{31}, situated in the disordered N-terminal region adjacent to beta strand I. Peptide 7 containing R^{31} had a mass shift corresponding to HPG (Table 3), indicating that this arginine engages heparin. From previously defined HBSs of FGF2, R^{31} is part of HBS-3 which locates to the N-terminus of the beta trefoil FGF (Fig 7).

Following on in the sequence are R^{42}, R^{48}, and R^{53}, which were detected on peptide 10. The mass shift of this peptide corresponded to two 1:1 condensed products between arginine and PGO. Hence, only two of the three arginines were modified by PGO, whereas one remained unreactive to PGO/HPG (Table 3). This peptide was also observed with just two modifications by PGO when the reaction was performed with FGF2 in solution in the absence of heparin (Supplementary table 1, peptide 9). The presence of an unlabelled arginine is surprising, since Arg-C could not cleave FGF2 after reaction with PGO and HPG (Fig 4A-C). The question of which arginine was not accessible to PGO, HPG or trypsin is addressed later.

Next, Arg^{69} and Arg^{81} were protected by PGO, as observed in two sister peptides 8 and 9. They cover the sequence from the loop between β-strand IV-strand V to a loop between strand V-strand VI, in which K^{75} has been defined as part of HBS-2. Peptide 9 is two amino acids longer than its sister, but their reaction products are distinct. While peptide 8 showed a single mass shift corresponding to the 2:1 condensed product, the 1:1 product was observed in peptide 9,

but only one of it's two arginine reacted with PGO and the other failed to react with PGO or HPG. The MALDI-TOF data are not conclusive as to which arginine was modified in each case and to whether PGO reacted with one arginine in both cases or it reacted with different arginine residues. Interestingly, when FGF2 reacted with PGO in-solution, Arg^{69} and Arg^{81} were both modified by PGO generating two 1:1 products (Supplementary table 1, peptide 7). These differences between reaction products may be due to the effect of either higher concentration of electrolytes in the reaction of the eluted protein with HPG or to a long-lasting effect on protein conformation of heparin binding.

Two sister peptides, 2 and 3, containing R^{90} from strand VI/strand VII loop (Fig 7) were labelled by HPG (Table 3), indicating that this residue was involved in the binding of FGF2 to heparin. R^{90} has not been shown in previous publications to be part of any HBSs (Fig 7).

Peptide 4 covers the sequence from F^{105} to Y^{115}, containing Arg^{108} in strand VIII (Fig 7) which formed the 1:1 condensed product with PGO, resulting in the mass shift of 116 Da.

Peptides 1 and 5 contain two arginine residues, R^{116} and R^{118} in strand IX, assigned previously though sequence alignment to HBS-2 (Fig 7) [36]. Peptide 1 showed a mass shift corresponding to one HPG, indicating that one of these two adjacent arginines engaged with heparin (Table 3). In peptide 5 there were two modifications, one from reaction with HPG, one from reaction with PGO (Table 3). In solution, both R^{116} and R^{118} were modified by PGO generating two 1:1 water condensed products (peptide 4 and peptide 5, Supplementary table 1). To answer the question of which of the arginine residues was bound to heparin, a protection reaction with PGO on a heparin mini-column was followed by sequential digestion for 5 hours with chymotrypsin then overnight with Arg-C. The peptides were analysed by MALDI_TOF MS. As result, R^{118} of peptide 5 (Table 3) was cut by Arg-C, resulting in peptide $^{113}NTYRSR^{118}$ with a mass before modification of 795.4 and a mass after modification of R^{116} by PGO of

1027.13 (Supplementary Fig 6). This observation demonstrated that R^{118} binds to heparin whereas R116 does not. The different behaviour of peptides 1 and 5 may be due to the N-terminal location of the arginine residue causing the loss of a PGO in some instances.

The last arginine on the sequence, R^{129} in strand X/strand XI loop (Fig 7) showed a mass shift of 132.02 corresponding to the modification by HPG (peptide 6, Table 3). This agrees with its prior assignment to HBS-1[120][36].

Using Peptide Mass (ExPASY) to predict peptides generated from native FGF2 by cleavage with chymotrypsin, Matchmaker found 12 modified peptides, summarized in Table 4 (Supplementary Fig 4). Ten of them overlap with the list generated by Prospector prediction. The extra two peptides were sisters that showed a mass shift corresponding to HPG on R^{129} of the canonical HBS-1 (Table 4).

Locations of modified arginine residues in the FGF2 structure

The canonical HBS of FGF2, HBS-1, has been characterized by many different methods, including X-ray crystallography, NMR spectroscopy, and site-directed mutagenesis, while evidence for its two secondary binding sites, HBS-2 and HBS-3, has also been acquired by independent approaches [100], [143], [186], [187], [197], [120]. The primary binding site (HBS1) consists of Lys^{35} and Asn^{36} and the group of 17 amino acids from 128-144 (Fig 7) [185]. In addition, HBS-3 was identified towards the N-terminus, which consists of 5 amino acids in the region 25-30, and HBS-2 is formed by Lys^{75} and Gly^{76}, $^{82}LMAK^{86}$ and Lys^{119} [197][120].

HBS-1, Arg129 and Arg90

One arginine in HBS-1, $R^{129,}$ was labelled by HPG (peptide 6_Table 3; peptides 4, 5 and 8_Table 4). HBS-1 is an extremely basic surface on FGF2 (Fig 5A) formed by Lys35, Asn36 on strand I/strand II loop and residues in strand X/strand XI loop, strand XI and strand XI/strand

XII loop (Fig 7). The double mutation R129A/K143A dramatically reduced the binding of the protein to a heparin affinity column [185]. In addition, R^{129} is highly conserved inthe FGF family and it aligns to K^{128} of FGF1, which engages heparin, as shown previously[36].

The location of R^{90} on the surface of FGF2 has suggested that this arginine is a part of HBS-1 (Fig 5A). Peptides 2 and 3 (Table 3 and Table 4, respectively) demonstrated that R^{90} reacted with HPG and hence is bound to heparin. The involvement of R^{90} in heparin binding has been suggested by a docking model to be an indirect interaction through an intervening water molecule to the $GlcNSO3^-$ group of glucosamine 5 (GlcNS, -5) at the reducing end of the sugar ligand [185][186].

HBS-2, Arg118

HBS-2 comprises amino acids scattered in sequence, Lys^{75}, K^{86} and $^{116}RSRK^{119}$. Sequence alignment, following the selective labelling of lysine residues, suggested that R^{118} and R^{116} were part of HBS2 [30]. However, the selective labelling of arginine residues demonstrates R^{118} but not R^{116} is bound to heparin. $^{118}RK^{119}$ is separated from HBS1 by an acidic boundary (Fig 5B), and hence likely to constitute a distinct binding site.

Re-assignment of K^{86} to HBS1

The selective labelling of lysine has identified K^{86} as part of the secondary HBS2 [36]. However, on the surface, K^{86} is adjacent to R^{90} of HBS1 and situated on a positively charged region (Fig 5B). These observations prompted us to propose the re-assignment of K^{86} to HBS1 rather than HBS2.

HBS-3, Arg31

HBS-3 is located N terminal to strand beta I (Fig 7). The lysine selective labelling of FGF2 found K^{30} to be biotinylated and K^{27} to be acetylated, implying the involvement of K30 in

heparin engagement [36]. R^{31} reacted with HPG (Peptide 7 – Table 3, Peptide 9 – Table 4) so can be considered to be part of HBS3.

Re-assignment of K^{75} to HBS3

Interestingly, K^{75}, which was considered as part of HBS2 by the selective labelling of lysine [36], is distant from $^{118}RK^{119}$ of HBS2. On the other hand, it is close to K^{30} and R^{31} of HBS3 (Fig 5C) in a continuous positively charged area of the protein's surface. Hence, we propose that K^{75} of FGF2 is part of HBS3, not HBS2.

HBS3 has two mutually exclusive partners: HS and FGFR

The structure of a co-crystal of FGF2 and FGFR1c (1CVS) [133] indicated that N-terminal segment of FGF2 interacts with the third immunoglobin loop of FGFR1. In particular, K^{31} of HBS3 forms a hydrogen bond with the side chain of Gln-284 and Asp-282 of FGFR1 [133]. In addition, sequence alignment data demonstrated that K^{30} and R^{31} were aligned to R^{84} and R^{85} of FGF4, respectively, which have been shown to interact with FGFR1 [197]. These observations imply that HS and FGFR are mutually exclusive binding partners of HBS3, and these amino acids may switch partners during the formation of the receptor signalling complex.

Arg42 of FGF2 is trapped in an intramolecular network of hydrogen bonds

R^{42} of FGF2 (peptide 3 – Table 4) was not modified by PGO in the in solution context or on the heparin affinity mini column and was not a site for Arg-C cleavage. This implies that this residue may have intramolecular interactions, which are sufficient to prevent significant interactions with solvent (Supplementary Fig 7). The stick structure of FGF2 illustrates that the side chain of R^{42} is 0.29 nm and 0.38 nm from the side chain of D^{50}; 0.29 nm and 0.38 nm to that of D^{57} and 0.34 nm and 0.38 nm to that of V^{52} (Supplementary Fig 7). These measurements suggested that the guanidino group of R^{42} is engaged in an intramolecular hydrogen bond network with the side chains of D^{50} and D^{57}, the invariant residues in the FGF family, with the

side chain of V^{49} forming a hydrophobic environment for the aliphatic portion of the arginine side chain. These interactions hold the side chain of R^{42} and so prevent the access of Arg-C and PGO/HPG. This intramolecular network of R^{42}, D^{50}, V^{52}, and D^{57} may serve to restrict the conformational freedom of the four antiparallel β strands I, II, III and IV.

Identification of arginine residues on FGF1 engaging heparin by Protect and Label strategy

FGF1, originally called acidic fibroblast growth factor due to its isoelectric point (pI) 6.52, has six arginine residues of which, only R^{134} and R^{137} located in HBS1 have been previously identified as interacting with heparin. The other arginine residues are R^{39}, 50, 52 and 103. Solvent-exposed arginine residues in heparin-bound FGF1 were protected with PGO and following elution of the FGF1, any arginine side chains engaged with the polysaccharide were labelled with HPG. The resulting peptides with information about modification, sequence, and final m/z are presented in Tables 5 and 6 and the spectra are in Supplementary Figs 8 and 9).

Prospector

Prospector predicted 148 peptides after cleavage of FGF1 by chymotrypsin with 100% sequence coverage, indicating that all six arginine residues would be included in the analysis. However, analysis of the data with Prospector did not identify a peptide containing R^{39} (Table 5) indicating that coverage was incomplete. Subsequently, the filter and analysis identified ten peptides with HPG-labelled arginines (Table 5) (Supplementary Fig 8).

R^{50} and R^{52} from beta strand III/beta strand IV loop (Fig 7) were found in peptide 8 (Table 5) and this showed a mass shift of two HPG products, indicating that both arginines were bound to heparin. R^{103} from beta strand VIII (Fig 7) reacted with HPG as observed in peptides 1, 2, 9 (Table 5), indicating the binding of this arginine to heparin.

The engagement of R^{134} and R^{137} in the HBS1 of FGF1 to heparin was demonstrated by HPG modification and a mass shift of two HPG products were found in peptides 4, 5, 6 and 10 (Table 5). However, peptide 3, which also contains R^{134} and R^{137}, only showed the mass shift of one HPG (Table 5, Supplementary Fig 13), meaning that one arginine had not reacted. Peptide 3 is identical to peptide 4 and overlaps with peptides 6 and 10. The mass shift of peptide 7 (Table 5, Supplementary Fig 8) is a combination of one HPG product and one 2:1 condensed product of reaction with PGO. Together these data suggest that while R^{134} and R^{137} are engaged to heparin, one of them may dissociate in the time of the protection step and so react with PGO, and moreover, may engage in intramolecular bonds rendering it resistant to reaction.

Peptide Mass (ExPASY)

In the case of FGF2, Prospector and ExPASY demonstrated a high level of overlap in terms of peptides containing modified arginine residues but, their predications are more diverse for FGF1. Nine peptides with modified arginines were identified when Peptide Mass (ExPASY) was used as the starting point for the analyses (Table 6, Supplementary Fig 9), but only two of them appeared in the list generated by Prospector, peptides 2 and 8.

R^{39}, R^{50}, and R^{52} were in peptides 7 and 8 (Table 6). Three products are observed in peptide 7, in which, only two products of arginine and HPG were detected, the other is a 2:1 product of PGO and arginine. In the case of peptide 8, the mass shift was attributed to two products of arginine with HPG, implying a lack of long-lasting reaction between one arginine with PGO or HPG. With the additional evidence from peptide 8 (Table 3) generated by Prospector, it could be concluded that HPG reacted with R^{50} and R^{52}.

R^{103} from beta strand VIII was labelled with HPG, resulting in a mass shift of HPG in five peptides, 1, 2, 4, 5 and 9 (Table 6), demonstrating its involvement in heparin binding.

R^{134} and R^{137} of HBS-1 were labelled by HPG as observed in peptides 3 and 6 (Table 6).

Locations of modified arginine residues in the FGF1 structure

Previous work demonstrated that FGF1 has three regions on its surface that engage heparin: the canonical HBS-1 and the secondary binding sites, HBS-2 and HBS-3 [36], [198], [199]. The core of the canonical heparin binding site of FGF1 is almost aligned to that in FGF2 (Fig 7), and extends from beta strand IX/beta strand X loop to beta strand XI/beta strand XII loop (Fig 7) as evidenced by X-ray crystallography and NMR spectroscopy[33], [36], [200], [201]. Lysine directed protect and label identify HBS-2 (K^{116}, K^{117}) in beta strand IX of FGF1, but do not include β strand VI as in FGF2 [36] and HBS-3, which locates towards the N-terminus, and consists of K^{24}, K^{25} and K^{27}.

HBS-1, Arg134 and Arg137

The canonical HBS-1 is highly conserved within the FGF sub-family in terms of both sequence alignment and structure. In terms of sequence, HBS-1 in FGF1 has a high density of basic residues. The selective labelling of lysines identified Lys-127, 128, 133 and 143 as interacting with heparin. The present data demonstrate that R^{134} and R^{137} (peptides 3, 4, 5, 6, 7 and 10, Table 5; peptides 3 and 6, Table 6; Fig 6A) also bound to heparin. The involvement of R^{134} in heparin binding of FGF1was indirectly evidenced in an NMR structure of FGF1 with inositol hexasulfate as a substitute for heparin [200]. Subsequently a direct involvement was shown using a synthetic heparin hexasaccharide which caused a chemical shift perturbation of R^{137} [33], [201].

FGF1 may have a single continuous HBS1.

Although FGF1 has an acid pI, its charged residues are segregated on its surface so that it possesses a largely basic face and a largely acidic face (Fig 6). With the additional data provided by arginine selective labelling we were able to re-examine the assignment of the basic residues to the different HBSs.

R^{103} in beta strand VIII (Fig 7) was labelled by HPG (Tables 5, 6).On the surface of FGF1, this arginine is adjacent to HBS1 and connected to HBS-2, $^{116}KK^{117}$ (Fig 6 B – lower panel). This suggests that R^{103} is part of HBS1. Both R^{50} and R^{52} in FGF1 were labelled by HPG, indicating that they interact with heparin. R^{50} and R^{52} locate on the loop between beta strand III and beta strand IV (Fig 7). Although separated by an acidic Asp51 residue (Fig 7), inspection of the surface electrostatic potential shows that there is no acidic boundary between these arginines (Fig 6A, C), presumably because the Asp side chain is involved in a local hydrogen bonding network. Notably, R^{50} and R^{52} are on the same basic face of the protein, as is the previously defined HBS3 (Fig 6 C). These arginines connect HBS1 and $^{24}KK^{25}$ and K^{27} in HBS3 of FGF1 (Fig 6 C), implying that R^{50} and R^{52} as well as $^{24}KK^{25}$ and K^{27} are part of HBS1.

In conclusion, we propose here that FGF1 has a single long, continuous HBS1, which comprises $^{24}KK^{25}$ and K^{27}at the N-terminus of beta strand I, R^{50} and R^{52} between β strand III/β strand IV, going through K^{127}, K^{128}, R^{134}, R^{137}, and K^{143} in β strand X/β strand XI, R^{103} in beta strand VIII and $^{116}KK^{117}$ in beta strand IX. Thus, although the position of most of these residues in the FGF1 structure is similar to that in the structure of FGF2, differences in the surface distribution of the acidic side chains in the two proteins lead to these sites likely coalescing in FGF1.

These data demonstrate that although FGF1 and FGF2 are in the same sub-family and, possess a high level of sequence conservation, they may differ in their interactions with heparin/HS. FGF2 has the primary and the secondary binding sites separated by acidic boundaries implying that they may engage different polysaccharide chain. Indeed, this is supported by the demonstration that FGF2 can crosslink HS chains [202]. However, FGF1 possess a continuous HBS suggesting that it can bind a single HS. Analysis of the interactions of FGF1 with HS demonstrates that unlike FGF2, it does not crosslink HS chain [203]. This is consistent with FGF1 possessing a single, large HS binding site.

Conclusion

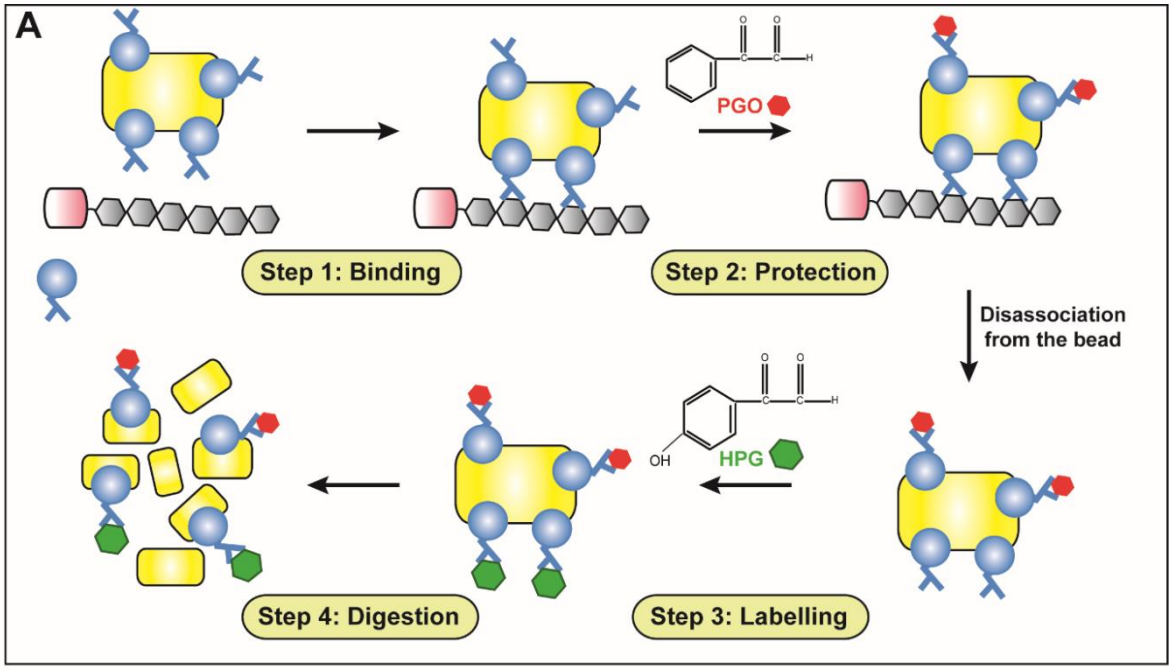
Our strategy for selective labelling of arginine side chains involved in binding to heparin, in combination with our published data on lysines, allows a greater dissection of the electrostatic bonds that drive the interactions between proteins and GAGs. The results with the model peptides illustrated the highly selectivity and specificity of PGO/HPG modification of the guanidino group of arginine. One of the main limitations is the multiple products from reaction with PGO, but our strategy of integrating mass spectrometry and automated analysis successfully tackled this issue. The data on proteins show that: 1) heparin binding effectively protects the arginine residues engaged with the polysaccharide, allowing the protection of uninvolved arginines with PGO;2) the recovery of protein after the elution from heparin affinity mini-column is reasonably quantitative; 3) the automated analysis is sensitive enough to identify the modifications of each arginine residue. The method is rapid and so should be applicable for any protein-heparin interactions. Moreover, like lysine selective labelling, our arginine labelling method identifies the secondary, low affinity binding sites. The likely importance of these in contributing to the structure of extracellular matrix and the regulation of ligand diffusion is suggested by a recent biophysical analysis [202]. As well as protein-sulfated GAG interactions, the method should be adaptable to any interactions involving arginine residues such as protein-nucleic acid and protein-phospholipid.

The identification of arginines engaging heparin in FGF1 and FGF2 alongside the lysines involved in binding provides a number of new insights. For example, we are able to propose the reassignment of HBSs in both FGF2 and FGF1 and intriguingly, FGF1 would appear to have just a single, large HBS1, similar to FGF9 [36]. In addition, the data demonstrate that the arginine just N-terminal to β-strand I, which in at least some instances are involved in binding the cognate FGFR, can alternatively bind heparin.

Figure legends

Figure 1: Workflow for selective labelling of arginine residues in proteins

A. The arginine protect and label experimental workflow. There are four main steps in the workflow. In the first step, the engagement between protein and heparin affinity beads results in the exposure to solvent of only the non-binding-involved arginine residues. The second step is protection by PGO of these exposed arginine side chains. The protein is then disassociated from heparin with 2 M NaCl, so that the arginine residues in heparin binding sites are available for labelling with HPG in the third step. Finally, proteins are digested by enzymes and peptides are analysed by MALDI-Q-TOF mass spectrometry. B Analytical workflow for the identification of modified peptides: B1-1, 2. The corresponding mass and sequence of peptides after enzyme digestion were predicted by Prospector and Peptide Mass. B1-3. The peptides were then processed by the *PGO-HPG mass predictor* Python script, which added the possible mass shifts after modification by PGO or HPG on each peptide based on the number of arginine residues in the native sequence. B1-4. This step provides the theoretical list of modified peptides. B2. The second step starts with the list observed list of peptides from MALDI-TOF MS data, which contains the information of the m/z and intensity of each peak. B3. Components in the predicted and observed lists are then matched by the *Matchmaker* Python script with a tolerance of difference of less than 0.1 Da. B4. The modified arginine residues were mapped onto the 3D structure of proteins back to identify their locations.


A
PGO
Step 1: Binding
Step 2: Protection
Disassociation from the bead
HPG
OH
Step 4: Digestion
Step 3: Labelling

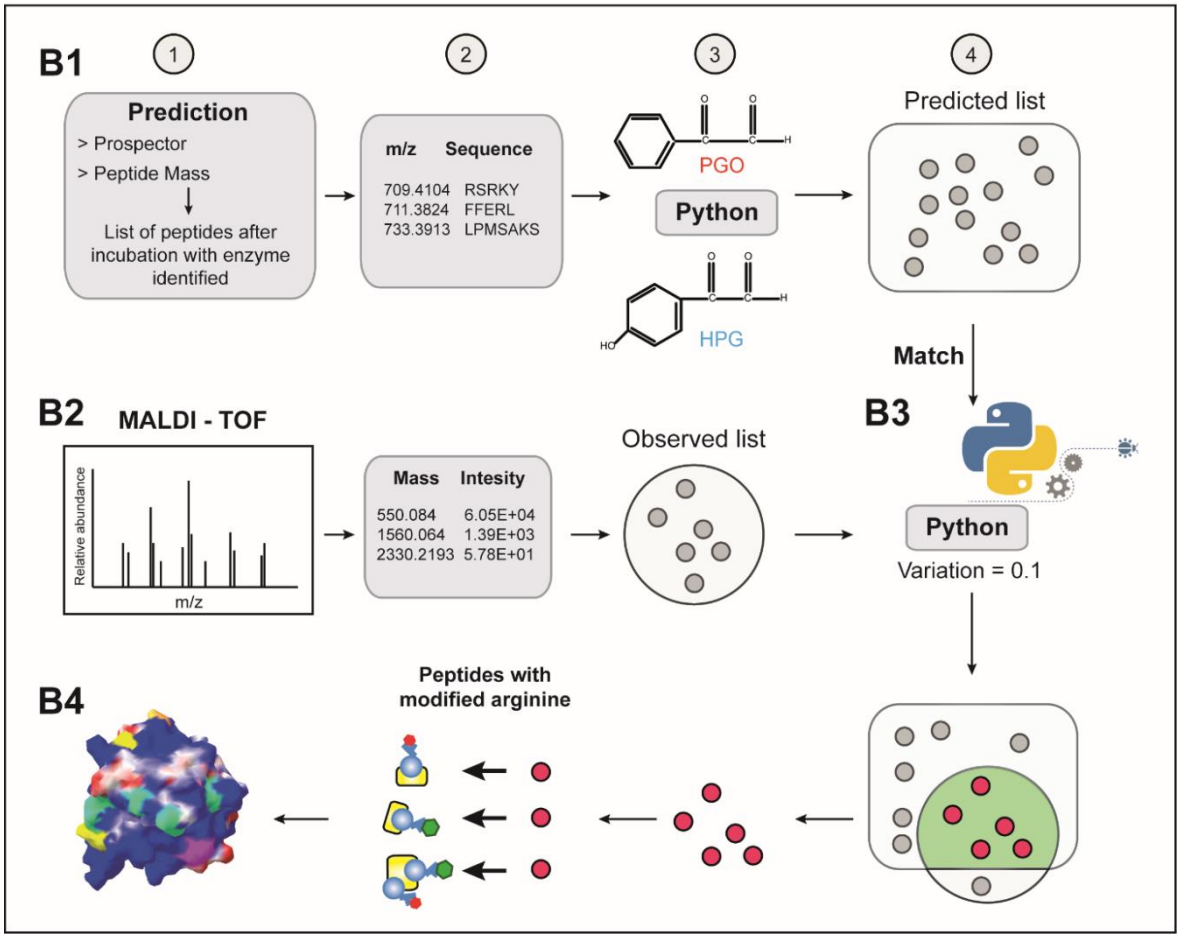

B1
1
2
3
4
Prediction
> Prospector
> Peptide Mass
List of peptides after incubation with enzyme identified
m/z Sequence
709.4104 RSRKY
711.3824 FFERL
733.3913 LPMSAKS
PGO
Python
HPG
Predicted list
Match
B2
MALDI - TOF
Relative abundance
m/z
Mass Intesity
550.084 6.05E+04
1560.064 1.39E+03
2330.2193 5.78E+01
Observed list
B3
Python
Variation = 0.1
B4
Peptides with modified arginine

Figure 2: Optimization of the reaction between PGO and arginine residues on peptides. A. The description of the stoichiometry of PGO with arginine. Initially, PGO reacts with the guanidine group of arginine to form a first product (Product 1), which may then condens with the loss of a water molecule through a reversible process to yield product 2. The 2:1 water condensed product of PGO and arginine may appear in the presence of excess PGO (product 3 or product 4), which have the same mass. B. Mass spectra of peptide FA reacted with PGO. The reaction between FA and 200 mM PGO was conducted in the dark, for 30 min at 25°C. There are six products observed. The products between FA and one PGO were demonstrated as three major peaks of 1139.60 Da, 1255.61 Da and 1275.58 Da. The reaction of FA with two PGO results in three peaks, numbered 4, 5 and 6 with the mass of 1388.62; 1410.60 and 1523.67. C. Reactions of arginine residues located at N- or C-terminus of a peptide. Peptide NA was cleaved by trypsin to generate two peptides: i) NA-I: RPYIL, mass 661.38 Da with arginine at the N terminus and ii) NA-II: LYENKPR, mass 1030.53 Da with arginine at C terminus. Both peptides were reacted with 200mM PGO in the dark for 30 min. After the reaction, each peptide showed 4 products, indicated by the arrows and numbered 1, 2, 3 and 4.

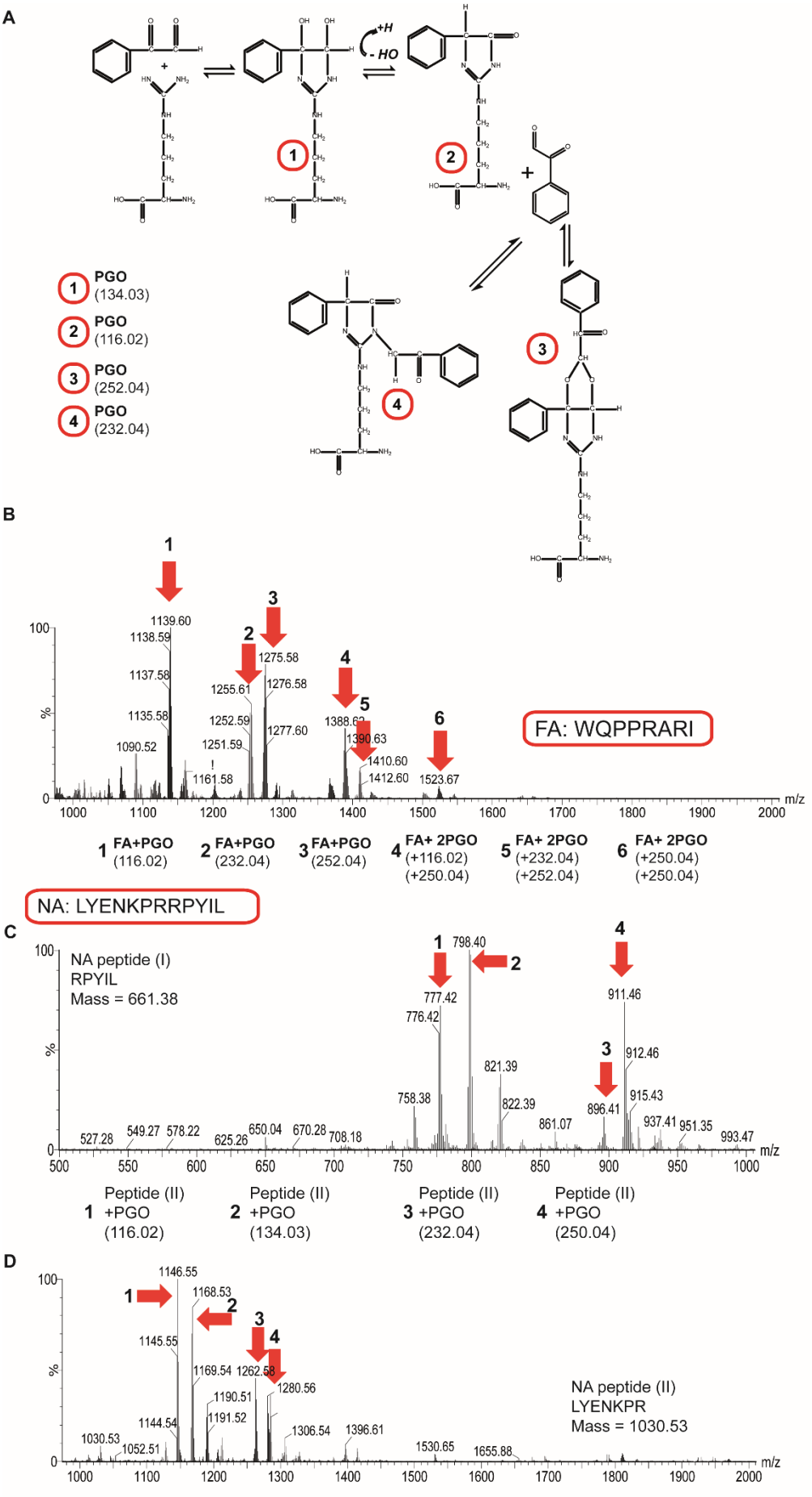

A
1 PGO (134.03)
2 PGO (116.02)
3 PGO (252.04)
4 PGO (232.04)
B
FA: WQPPRARI
1 FA+PGO (116.02)
2 FA+PGO (232.04)
3 FA+PGO (252.04)
4 FA+ 2PGO (+116.02) (+250.04)
5 FA+ 2PGO (+232.04) (+252.04)
6 FA+ 2PGO (+250.04) (+250.04)
NA: LYENKPRRPYIL
C
NA peptide (I)
RPYIL
Mass = 661.38
1 Peptide (II) +PGO (116.02)
2 Peptide (II) +PGO (134.03)
3 Peptide (II) +PGO (232.04)
4 Peptide (II) +PGO (250.04)
D
NA peptide (II)
LYENKPR
Mass = 1030.53
m/z
%

Figure 3: Optimization of the reaction between HPG and arginine residues on peptides. A. The reaction of HPG and arginine. There is a single product generated after reaction between HPG and guanidino side chain of arginine residues. B. Mass spectra of FA reacted with HPG. The reaction between FA and 100 mM HPG was conducted in the dark, for 10 min (Upper panel) and 30 min (Lower panel), at 25°C. At 10 min, the original FA with the mass of 1023.5 is observed along with and (FA+HPG) with a mass shift of 1156.6. At 30 min, only the product (FA+HPG) was observed. C. Mass spectra of NA reacted with HPG. The reaction between FA and 100 mM HPG was conducted in the dark, for 10 min (Upper panel) and 30 min (Lower panel), at 25°C. At 10 min, the original NA and along with the reaction product (NA+HPG) were detected. At 30 min, only the product (NA+HPG) was observed.

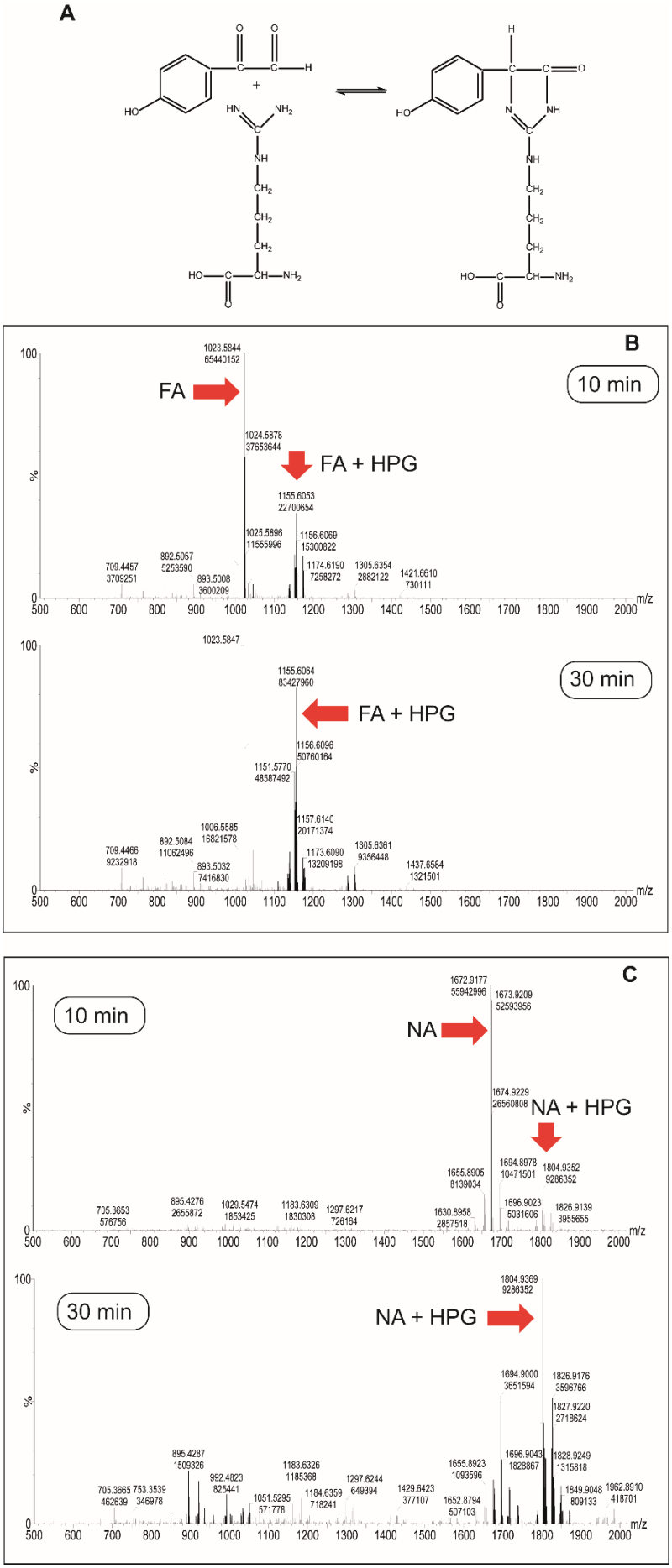

A
B
10 min
FA
1023.5844
65440152
1024.5878
37653644
FA + HPG
1155.6053
22700654
1025.5896
11555996
1156.6069
15300822
892.5057
5253590
709.4457
3709251
893.5008
3600209
1174.6190
7258272
1305.6354
2882122
1421.6610
730111
30 min
1023.5847
1155.6064
83427960
FA + HPG
1156.6096
50760164
1151.5770
48587492
1157.6140
20171374
1006.5585
16821578
892.5084
11062496
709.4466
9232918
893.5032
7416830
1173.6090
13209198
1305.6361
9356448
1437.6584
1321501
C
10 min
1672.9177
55942996
1673.9209
52593956
NA
1674.9229
26560808
NA + HPG
1655.8905
8139034
1694.8978
10471501
1804.9352
9286352
705.3653
576756
895.4276
2655872
1029.5474
1853425
1183.6309
1830308
1297.6217
726164
1630.8958
2857518
1696.9023
5031606
1826.9139
3955655
30 min
1804.9369
9286352
NA + HPG
1694.9000
3651594
1826.9176
3596766
1827.9220
2718624
895.4287
1509326
1183.6326
1185368
1655.8923
1093596
1696.9043
1828867
1828.9249
1315818
992.4823
825441
1297.6244
649394
705.3665
462639
753.3539
346978
1051.5295
571778
1184.6359
718241
1429.6423
377107
1652.8794
507103
1849.9048
809133
1962.8910
418701
%
m/z
100
0
500
600
700
800
900
1000
1100
1200
1300
1400
1500
1600
1700
1800
1900
2000

Figure 4: Selective labelling of arginine residues in the heparin binding sites of FGF2 and FGF1. The reaction between PGO and FGF2 or FGF1 was conducted in solution and or when bound to heparin beads in the dark, at 25°C for 60 min. The protein eluted from the column was further reacted with PGO in the same conditions. The products collected after reactions were incubated with chymotrypsin or Arg-C, at 37°C overnight. All samples were analysed by SDS-PAGE. A. Reactions of FGF2 with PGO. Lane 1: 1/10 of the native FGF2 used in the in solution experiment; lane 2: 1/10 of the protein after the reaction with 200 mM PGO; lane 3: products of the digestion of native FGF2 by chymotrypsin; lane 4: products of the digestion by chymotrypsin of PGO-modified FGF2; lane 5: products of the digestion by Arg-C of native FGF2; lane 6: products of the digestion by Arg-C of PGO-modified FGF2. B. The reaction of FGF2 bound to heparin affinity column with PGO. L, 1/10 of the protein loaded onto the column; FT, 1/10 of the flow through fraction; W, 1/10 of the fraction washed with Na-1 buffer; PGO PROTECT FT, flow through fraction after the reaction between FGF2 and PGO; PGO PROTECT W, wash fraction after the reaction with PGO; PGO PROTECT E, Eluate from the column with Na-2 buffer; PGO PROTECT lane 1: the protein in the elution was digested with chymotrypsin; PGO PROTECT lane 2: the protein in elution was digested with Arg-C. C. Labelling of arginine residues in binding sites of FGF2. FGF2 from the eluate (panel B, fraction PGO PROTECT E) was reacted with PGO at a final concentration of 100 mM for 60 min. The product from this the second reaction was loaded onto a heparin affinity minicolumn. L, 1/10 of the FGF2 labelled by PGO; FT, the flow through the minicolumn; W, 1/10 of the wash with Na-1 buffer; E, Elution with buffer Na-2; lane 1, the labelled FGF2 in FT fraction was digested by chymotrypsin; lane 2, the labelled FGF2 in FT fraction was digested by Arg-C.

D. Reaction of FGF1 with PGO. Lane 1: 1/10 of the native FGF1 used in the in solution experiment; lane 2: 1/10 of the protein after the reaction with 200 mM PGO; lane 3: products of the digestion of native FGF1 by chymotrypsin; lane 4: products of the digestion by

chymotrypsin of PGO-modified FGF1; lane 5: products of the digestion by Arg-C of the native FGF1; lane 6: products of the digestion by Arg-C of PGO-modified FGF1. E. The reaction of FGF1 bound to heparin affinity column with PGO. L, 1/10 of the protein loaded onto the column; FT, 1/10 of the flow through fraction; W, 1/10 of the fraction washed with Na-1 buffer; PGO PROTECT FT, flow through fraction after the reaction between FGF1 and PGO on the mini-column; PGO PROTECT W, wash fraction after the reaction with PGO; PGO PROTECT E, Eluate from the column with Na-2 buffer; PGO PROTECT lane 1: the protein in the elution was digested with chymotrypsin; PGO PROTECT lane 2: the protein in elution was digested with Arg-C. F. Labelling of arginine residues in binding sites of FGF1. FGF1 from the eluate (panel D, fraction PGO PROTECT E) was reacted with PGO at a final concentration of 200 mM for 60 min, the product of this second reaction was loaded onto a heparin affinity mini-column. L, 1/10 of the FGF1 labelled by PGO; FT, the flow through the mini-column; W, 1/10 of the wash with buffer Na-1; E, Elution with buffer Na-2; lane 1, the labelled FGF1 in FT fraction was digested by chymotrypsin; lane 2, the labelled FGF1 in FT fraction was digested by Arg-C.

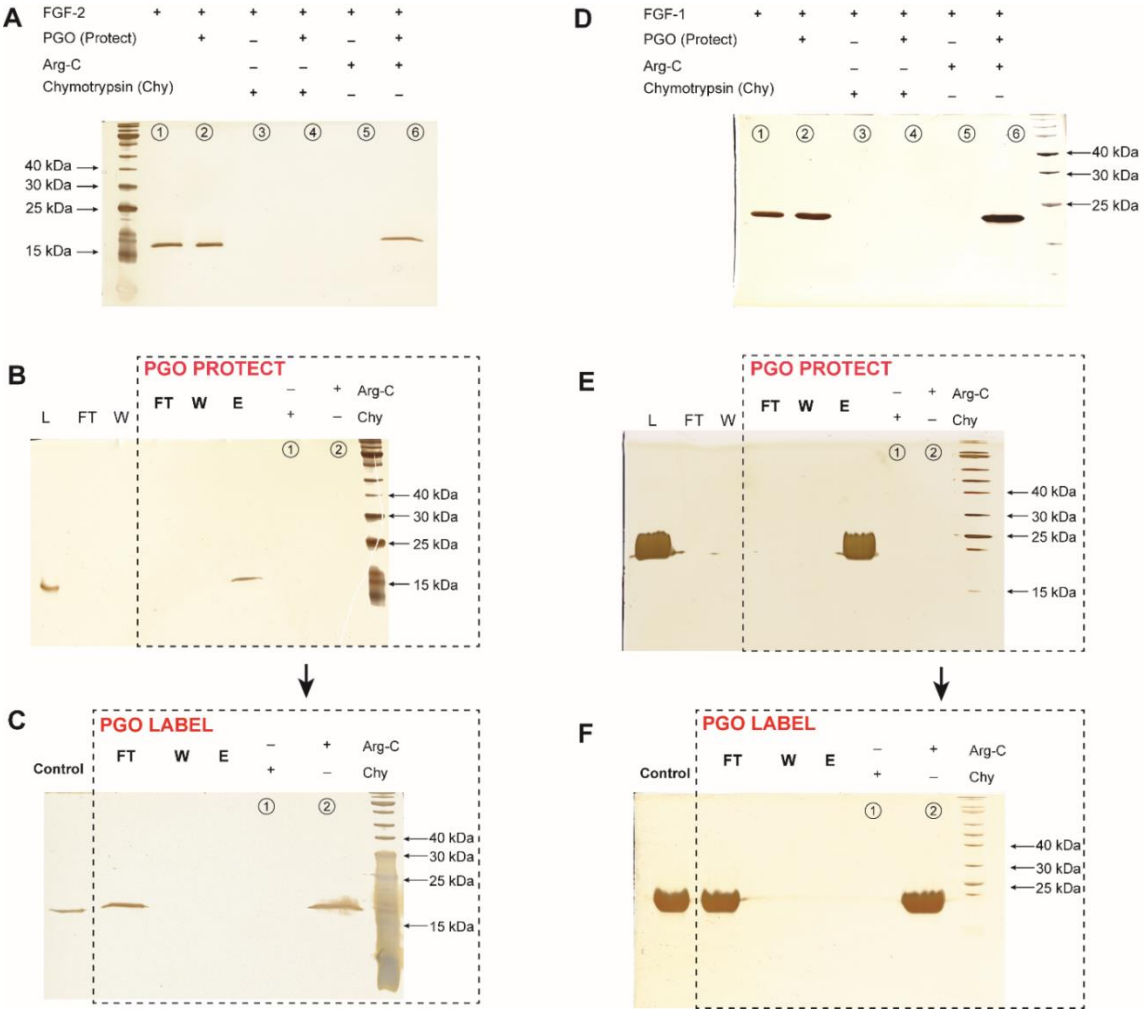

A
FGF-2
PGO (Protect)
Arg-C
Chymotrypsin (Chy)
40 kDa
30 kDa
25 kDa
15 kDa
D
FGF-1
B
PGO PROTECT
L FT W
FT W E
Arg-C
Chy
E
C
PGO LABEL
Control
F

Figure 5: Labelled arginine and lysine residues in heparin-binding sites of FGF2. The FGF2 structure (PBD code 1bfg [44]) is shown as a surface. The electrostatic potential of FGF2 and FGF1 was computed using Poisson-Boltzmann algorithm by Swiss-PDBView with the positively charged areas coloured *blue* and the negatively charged areas coloured *red*. Labelled lysine residues [36] [120] are presented in *yellow*. Labelled arginine residues (Tables 3, 4) are coloured *green*. This colour scheme for electrostatic potential is used throughout the paper including Figs 5, 6. A. Basic residues in HBS1. R^{129} of the canonical HBS-1. Other basic residues in HBS1 include K^{128}, K^{134}, K^{138}, K^{144} [36] [120]. R^{90} and K^{86} which are now assigned to be part of HBS1 by this work are shown. B. Basic residues in HBS2. Location of R^{118} and K^{119} are shown on the surface in relative to the rest of HBS-2. All the basic residues in HBS1 can also be seen from this viewpoint as well. C. Basic residues in HBS3. Location of K^{30}, R^{31} and K^{75}.

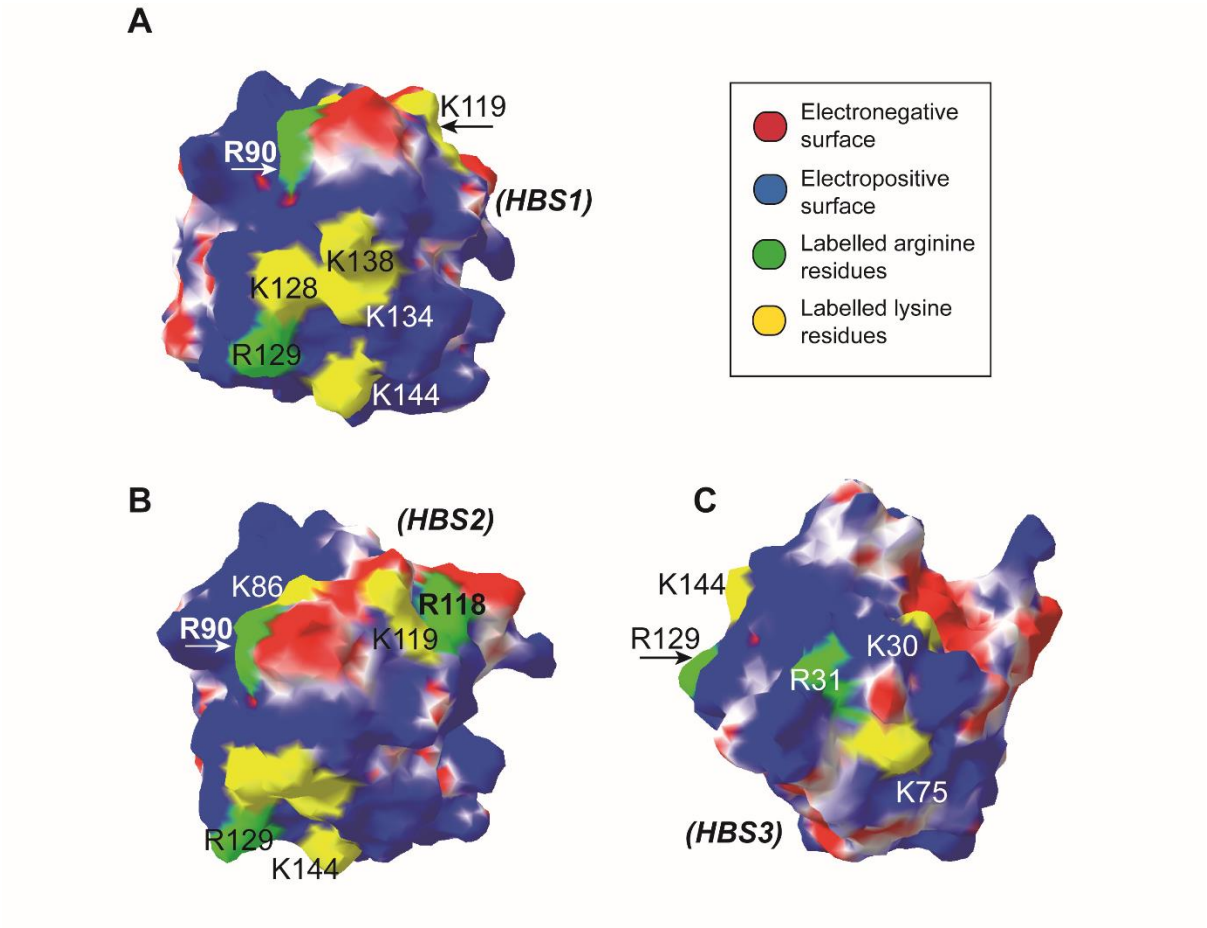


Figure 6: Labelled arginine and lysine residues in the heparin binding sites of FGF1. The FGF1 structure (PBD code 2erm [33]) is shown as a surface. A. Location of labelled R^{50}, R^{52}, R^{134} and R^{137} on the surface as part of the canonical HBS-1. The other basic residues are K^{127}, K^{128}, K^{143}, K^{116} and K^{117} [36]. K^{116} and K^{117}, which previously were part of HBS-2, now are assigned to be in HBS1 by this work. B. (Upper) The location of is mapped to the surface, on the electronegative surface. (Lower) The connection of R^{103} to HBS1. C. Locations of K^{24}, K^{25} and K^{27} [36] situated at the N-terminal are shown along with locations of R^{50} and R^{52}. K^{24}, K^{25} and K^{27} have been reassigned from HBS-3 [36] to HBS-1 in this study.

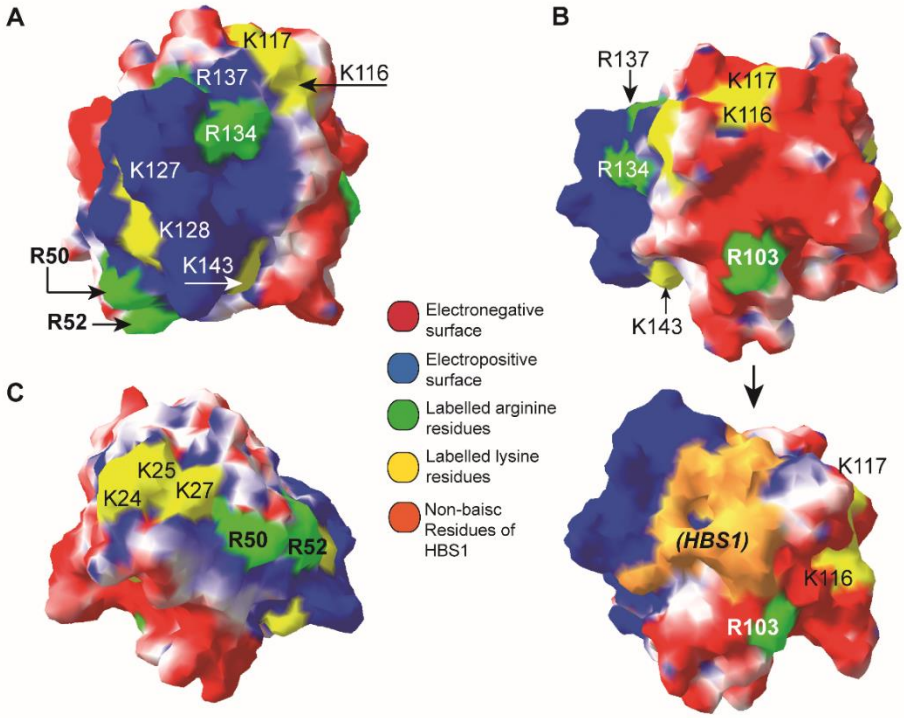


Figure 7: Sequence alignment of FGF2, FGF1 and location of labelled arginine and lysine residues in the heparin binding sites. Uniprot ID of FGF2 sequence is, P09038-2, Uniprot ID of FGF1 sequence is P05230. Labelled arginine residues are coloured in *red*, labelled lysine residues are coloured in *blue*, the HBSs are highlighted in red boxes. The β strands are presented as the arrows.

Tables

Table 1: Reactions of arginine residues in peptide FA with PGO

A. Screen for the conditions for PGO reaction on model peptide FA. Peptide FA at three different concentrations (100 mM, 200 mM, 1M) reacted with PGO at two different times (5 min and 30 min). (X) corresponds to "does not react" and (V) indicates that the reaction between peptide and PGO was detected.

	Concentration	100 mM	200 mM	1M
	Time			
FA *WQPPRARI* MW: *1023.58 Da*	5 min	X	**V**	X
	30 min	X	**V**	X
	PGO 200 mM for 30 min			
	Products	**Mass (Da)**		**Shift**
	1	1139.60		116.02
	2	1255.61		232.03
	3	1275.58		252.0
	4	1388.62		116.02+250.04
	5	1410.60		232.03+250.04
	6	1523.67		250.04+250.04

B. Screen for the products for PGO reaction on model peptide NA. Peptide NA was pre-treated by trypsin to generate two peptides: NA-I and NA-II. The resulting peptides reacted with PGO at concentration of 200 mM for 30 min.

Peptide NA (p-Glu-LYENKPRRPYIL) (**PGO200 mM for 30 min)**					
peptide **NA-I**: RPYIL MW: *661.38 Da*			peptide **NA-II**: YENKPR MW: *1030.53* Da		
Products	**Mass (Da)**	**Shift**	**Products**	**Mass (Da)**	**Shift**
1	777.42	116.02	**1**	1146.55	116.02
2	795.40	134.03	**2**	1164.52	134.03
3	896.41	232.03	**3**	1262.58	232.03
4	911.46	250.04	**4**	1280.56	250.04

Table 2: Products of PGO/HPG and arginine side chains with corresponding mass shift

The summary of the products between PGO/HPG and arginine side chains.The first PGO reacts with the guanidino group to produce1:1 product (Fig 2A, product [1]), which could then reversibly release a hydroxyl group from glyoxal, resulting in the 1:1 product water-condensed product (Fig 2A, product [2]). Following is the reaction of a second PGO molecule (if sufficient reactant) onto either a nitrogen of the guanidino group of arginine (Fig 2A, products [3]) or the carbonyl group of the first PGO (Fig 2A, products [4]) to form the final 2:1 product water-condensed product.

# of product	**Product**	**Mass shift**
Figure 2A 1	1:1 product	134.13
Figure 2A 2	1:1 product water-condensed	116.12
Figure 2A 3	2:1 product	252.23
Figure 2A 4	2:1 product water-condensed	232.26
Figure 3A	HPG	132.12

Table 3: FGF-2 peptide analysis based on prediction by Prospector. The reference number of the peptide is followed by the predicted m/z and with the sequence of the peptide following cleavage of native FGF-2 by chymotrypsin. The two columns under "Native FGF-2" present the predicted m/z and sequences of peptides after chymotrypsin digestion of FGF-2.The first two columns under "FGF-2 protect and label (P&L)" present the observed and predicted m/z for FGF-2 after modifications of arginine residues. The third column is the assignment of the peptide to one of the three HBSs of FGF2[36], [120]. The final columns indicates the modification occurring on the arginine residues of the peptides.

	Native FGF-2		**FGF-2 protect and label (P&L)**[1]			
	m/z theoretical	**Sequence**	**m/z observed**	**m/z theoretical**	**Published HBS**[2]**.**	**Modifications**
1	996.54	^{116}**RSRKY**TSW^{123}	1144.62	1144.56	HBS2 (R^{116} and R^{118})	1 R + 132 **(HPG)** and oxidation
2	1032.55	^{83}**AMK**EDGRLL^{92}	1180.59	1180.59	HBS2 (R^{90})	1 R + 132 **(HPG)** and oxidation
3	1032.55	^{82}**LAMK**EDGRL^{91}	1180.59	1180.59	HBS2 (R^{90})	1 R + 132 **(HPG)** and oxidation
4	1318.61	^{105}FFERLESNNY^{115}	1434.63	1434.61		1 R +116**(PGO)**
5	1374.69	^{113}NTY**RSRKY**TW^{123}	1738.82	1738.74	HBS2 (R^{118})	2 R + 132 **(HPG)** + 232
6	1439.84	^{124}YVAL**KRTGQYKL**^{135}	1587.85	1587.86	HBS1 (R^{129})	1 R + 132 **(HPG)** and oxidation
7	1525.79	^{27}**KDPKR**LYCKNGGF^{40}	1673.79	1673.82	HBS3 (R^{31})	1 R + 132 **(HPG)** and oxidation
8	1979.01	^{65}QAEERGVVS**IKG**VCANR Y^{82}	2211.15	2211.08	HBS2	1 R +232**(PGO)**
9	2220.16	^{63}QLQAEERGVVS**IKG**VCA NRY^{82}	2353.17	2353.15	HBS2 (R^{69} and R^{81})	R + 133 **(PGO)**
10	2537.41	^{41}LRIHPDGRVDGVREKSD PHIKL^{62}	2769.45	2769.45		2 R + 116 x 2**(PGO x2)**

[1] P&L: FGF-2 reacted in solution with PGO when bound to a heparin affinity column (protection step) and then, following elution, reaction of any protected arginine residues with HPG (labelling step).

[2] The bolded amino acids are a part of the known HBSs which were noted in column 3rd of FGF-2 on column.

Table 4: FGF-2 MS analysis based on prediction by Peptide Mass (ExPASy). The reference number of the peptide is followed by the predicted m/z and with the sequence of the peptide following cleavage of native FGF-2 by chymotrypsin. The two columns under "Native FGF-2" present the predicted m/z and sequences of peptides after chymotrypsin digestion of FGF-2.The first two columns under "FGF-2 protect and label (P&L)" present the observed and predicted m/z for FGF-2 after modifications of arginine residues. The third column is the assignment of the peptide to one of the three HBSs of FGF2 [36], [120]. The final columns indicates the modification occurring on the arginine residues of the peptides.

	Native FGF2		**FGF-2 (P&L)**[3]			
	m/z theoretical	**Sequence**	**m/z observed**	**m/z theoretical**	**Published HBS**[4]	**Modifications**
1	996.54	^{116}**RSRKY**TSW^{123}	1144.63	1144. 56	HBS2 (R^{116} and R^{118})	1 R + 132 **(HPG)** and oxidation
2	1032.55	^{82}**LAMK**EDGRL^{91}	1180.59	1180.58	HBS2 (R^{90})	1 R + 132 **(HPG)** and oxidation
3	1032.55	^{83}**AMK**EDGRLL^{92}	1180.59	1180.58	HBS2 (R^{90})	1 R + 132 **(HPG)** and oxidation
4	1198.66	^{124}YVAL**KRTGQY**^{134}	1362.61	1362.67	HBS1 (R^{129})	1 R + 132 **(HPG)** and oxidation
5	1276.77	^{125}VAL**KRTGQYKL**^{135}	1408.71	1408.77	HBS1 (R^{129})	1 R + 132 **(HPG)**
6	1318.61	^{105}FFERLESNNY^{115}	1434.63	1434.61		1 R + 116 **(PGO)**
7	1374.69	^{113}NTY**RSRKY**TW^{123}	1738.82	1738.74	HBS2	2 R + 132 **(HPG)** + 232
8	1439.84	^{124}YVAL**KRTGQYKL**^{135}	1587.85	1587.86	HBS1 (R^{129})	1 R + 132 **(HPG)**
9	1525.79	^{27}**KDPKR**LYCKNGGF^{40}	1673.79	1673.82	HBS3 (R^{31})	1 R + 132 **(HPG)**
10	1979.01	^{65}QAEERGVVS**IKG**VC ANRY^{82}	2211.15	2211.08	HBS2	1 R + 232**(PGO)**
11	2220.15	^{63}QLQAEERGVVS**IKG** VCANRY^{82}	2369.17	2369.15	HBS2 (R^{69} and R^{81})	2 R + 13**3(PGO)**
12	2537.41	^{41}LRIHPDGRVDGVREK SDPHIKL^{62}	2769.27	2769.26		2 R + 116 x 2**(PGOx2)**

[3] P&L: FGF-2 reacted in solution with PGO when bound to a heparin affinity column (protection step) and then, following elution, reaction of any protected arginine residues with HPG (labelling step).

[4] The bolded amino acids are a part of the known HBSs which were noted in column 3rd of FGF-2 on column.

Table 5: FGF-1 MS analysis based on prediction by Prospector. The reference number of the peptide is followed by the predicted m/z and with the sequence of the peptide following cleavage of native FGF-1 by chymotrypsin. The two columns under "Native FGF-1" present the predicted m/z and sequences of peptides after chymotrypsin digestion of FGF-1.The first two columns under "FGF-1 protect and label (P&L)" present the observed and predicted m/z for FGF-1 after modifications of arginine residues. The third column is the location of arginine residues of HBS-1 of FGF-1. The final columns indicates the modification occurring on the arginine residues of the peptides.

	Native FGF-1		**FGF-1 (P&L)**[5]			
	m/z theoretical	**Sequence**	**m/z observed**	**m/z theoretical**	**Arginine of HBS1**	**Modifications**
1	1349.65	^{100}FLE**R**LEENHY^{109}	1481.67	1481.63	R^{103}	1 R + 132 **(HPG)**
2	1467.65	^{102}E**R**LEENHYNTY^{109}	1599.67	1599.69	R^{103}	1 R + 132 **(HPG)**
3	1901.03	^{124}VGLKKNGSCK**R**GP**R**THY^{140}	2033.05	2033.07	R^{134} and R^{137}	1 R + 132 **(HPG)**
4	1901.03	^{124}VGLKKNGSCK**R**GP**R**THY^{140}	2165.07	2165.07	R^{134} and R^{137}	2 R + 132 **(HPG)** x 2
5	2048.10	^{123}FVGLKKNGSCK**R**GP**R**THY^{140}	2312.14	2312.14	R^{134} and R^{137}	2 R + 132 **(HPG)** x 2
6	2389.30	^{127}KKNGSCK**R**GP**R**THYGQKAIL F^{147}	2653.34	2653.30	R^{134} and R^{137}	2 R + 132 **(HPG)** x 2
7	2389.30	^{127}KKNGSCK**R**GP**R**THYGQKAIL F^{147}	2753.39	2753.30	R^{134} and R^{137}	2R + 132 **(HPG)** + 232
8	2633.37	^{40}ILPDGTVDGT**R**D**R**SDQHIQLQ L^{59}	2897.42	2897.39	R^{50} and R^{52}	2 R + 132 **(HPG)** x 2
9	3105.55	^{102}E**R**LEENHYNTYISKKHAEKN WFVGL^{126}	3253.57	3253.49	R^{103} And HBS-2	1 R + 132 **(HPG)**
10	3197.74	^{127}KKNGSCK**R**GP**R**THYGQKAIL FLPLPVSSD^{155}	3461.88	3461.78	R^{134} and R^{137}	2 R + 132 **(HPG)** x 2

[5] P&L: FGF-1 reacted in solution with PGO when bound to a heparin affinity column (protection step) and then, following elution, reaction of any protected arginine residues with HPG (labelling step).

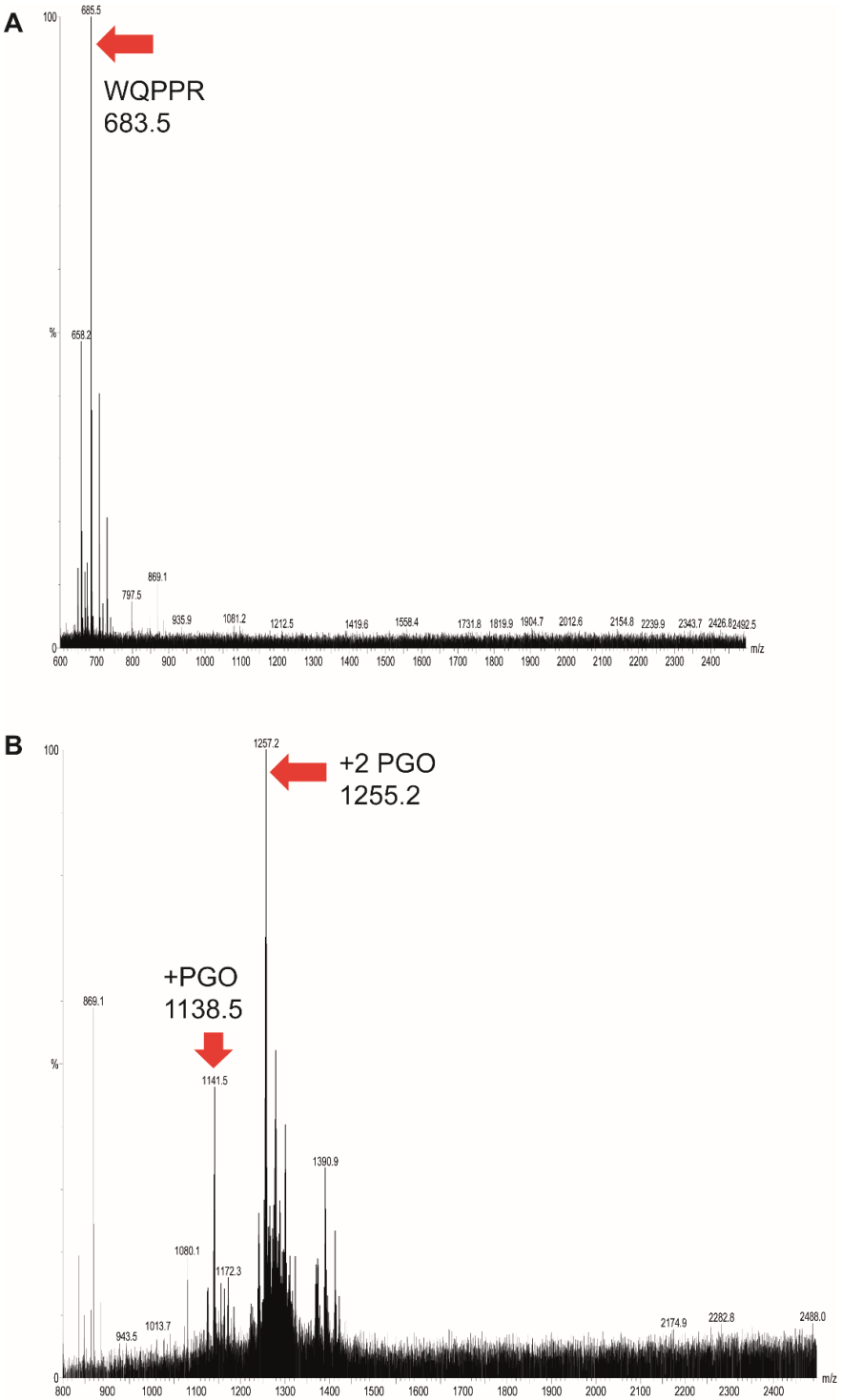
Table 6: FGF-1 MS analysis based on prediction by Peptide Mass. The reference number of the peptide is followed by the predicted m/z and with the sequence of the peptide following cleavage of native FGF-1 by chymotrypsin. The two columns under "Native FGF-1" present the predicted m/z and sequences of peptides after chymotrypsin digestion of FGF-1.The first two columns under "FGF-1 protect and label (P&L)" present the observed and predicted m/z for FGF-1 after modifications of arginine residues. The third column is the location of arginine residues of HBS-1 of FGF-1. The final columns indicates the modification occurring on the arginine residues of the peptides.

	Native FGF-1		**FGF-1 (P&L)**[6]			
	m/z theoretical	**Sequence**	**m/z observed**	**m/z theoretical**	**Arginine of HBS1**	**Modifications**
1	1089.50	^{102}E**R**LEENHY^{109}	1221.62	1221.65	R^{103}	1 R + 132 **(HPG)**
2	1349.65	^{100}FLE**R**LEENHY^{109}	1481.67	1481.64	R^{103}	1 R + 132 **(HPG)**
3	1631.85	^{127}KKNGSCK**R**GP**R**TH Y^{141}	1895.90	1895.82	R^{134} and R^{137}	2 R + 132 **(HPG)** x 2
4	1727.80	^{100}FLE**R**LEENHYNTY^{112}	1859.82	1859.73	R^{103}	1 R + 132 **(HPG)**
5	1898.90	^{89}YGSQTPNEECLFLE **R**L^{104}	2030.92	2030.83	R^{103}	1 R + 132 **(HPG)**
6	2242.23	^{127}KKNGSCK**R**GP**R**THYGQ KAIL^{146}	2506.28	2506.32	R^{134} and R^{137}	2 R + 132 **(HPG)** x 2
7	2505.32	^{38}L**R**ILPDGTVDGT**R**D**R**SDQ HIQL^{59}	3019.42	3019.39	R^{39,} R^{50} and R^{52}	2 R + 132 **(HPG)** x 2 + 250 **(PGO)**
8	2633.37	**^{39}R**ILPDGTVDGT**R**D**R**SDQ HIQLQL^{59}	2897.43	2897.40	R^{39,} R^{50} and R^{52}	2 R + 132 **(HPG)** x 2
9	3105.55	^{102}E**R**LEENHYNTYISKKHA EKNWFVGL^{126}	3267.58	3267.61	R^{103}	1 R + 132 **(HPG)** And 2 oxidation

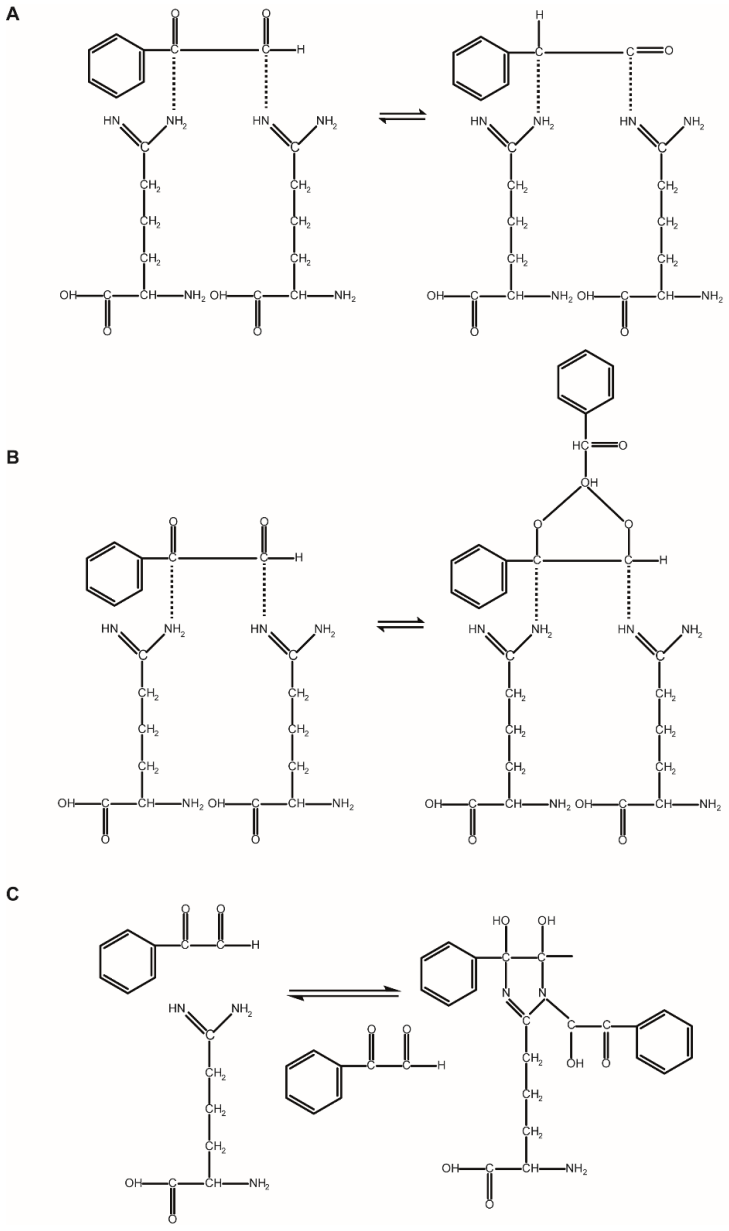
[6] P&L: FGF-1 reacted in solution with PGO when bound to a heparin affinity column (protection step) and then, following elution, reaction of any protected arginine residues with HPG (labelling step).

Supplementary data

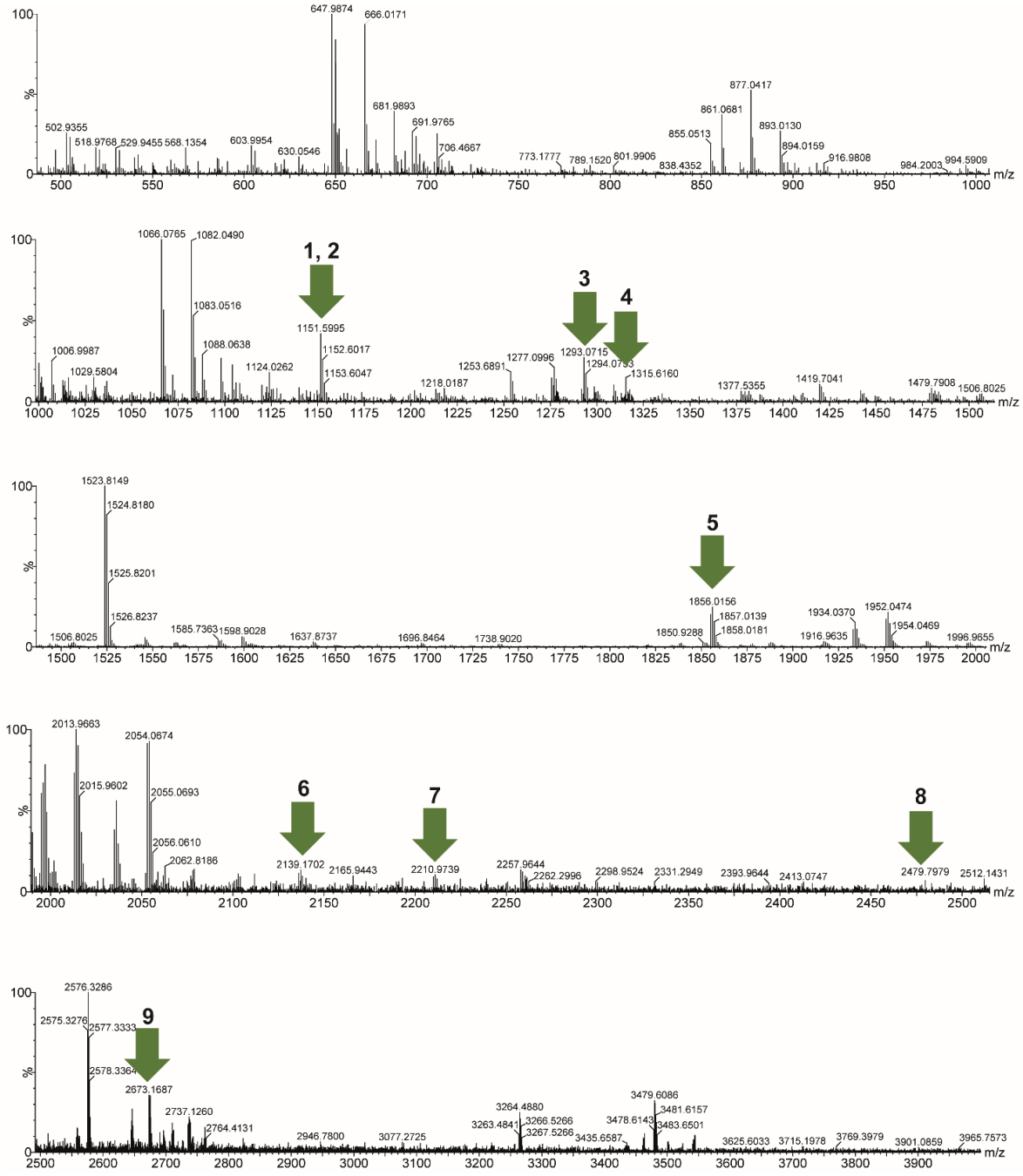
Supplementary figure 1: Mass spectra of the trypsin digestion products of peptide FA. A. FA was digested by trypsin. The larger part of the sequence was identified as “WQPPR”. B. Peptide FA after the reaction with PGO was digested with trypsin.



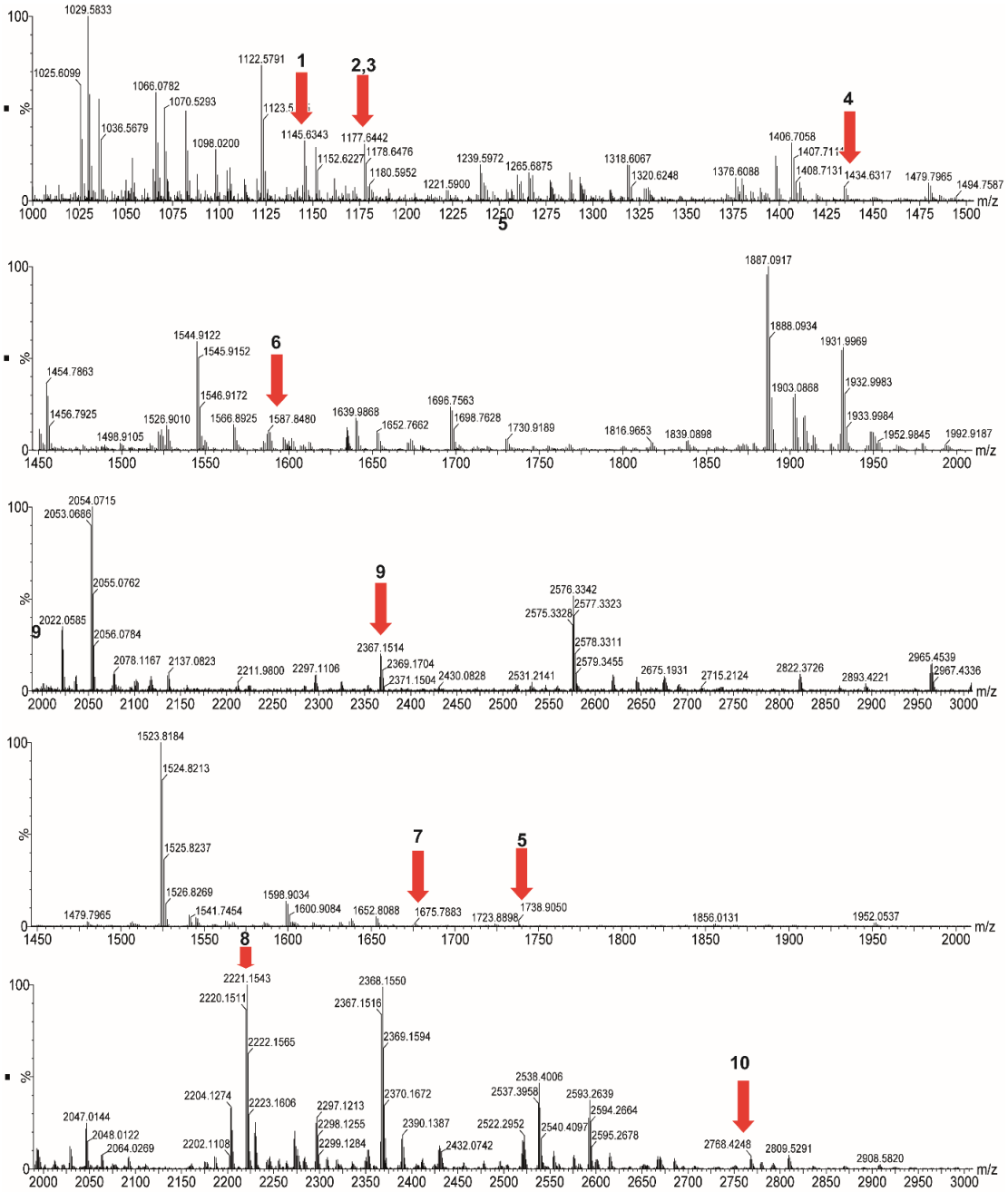
Supplementary figure 2: The stoichiometry of reaction of PGO with arginine residues in specific sequence contexts. A. B. PGO reacts to two adjacent arginine residues which are separated by a small sized amino acid, alanine. (A) When PGO reactsindividually with each arginine; (B) When the dicarbonyl group of PGO reacts with NH_2 groups on different arginine residues; (C) When PGO reacts to arginine at the N- or C-terminus forming 2:1 product of PGO:Arginine.



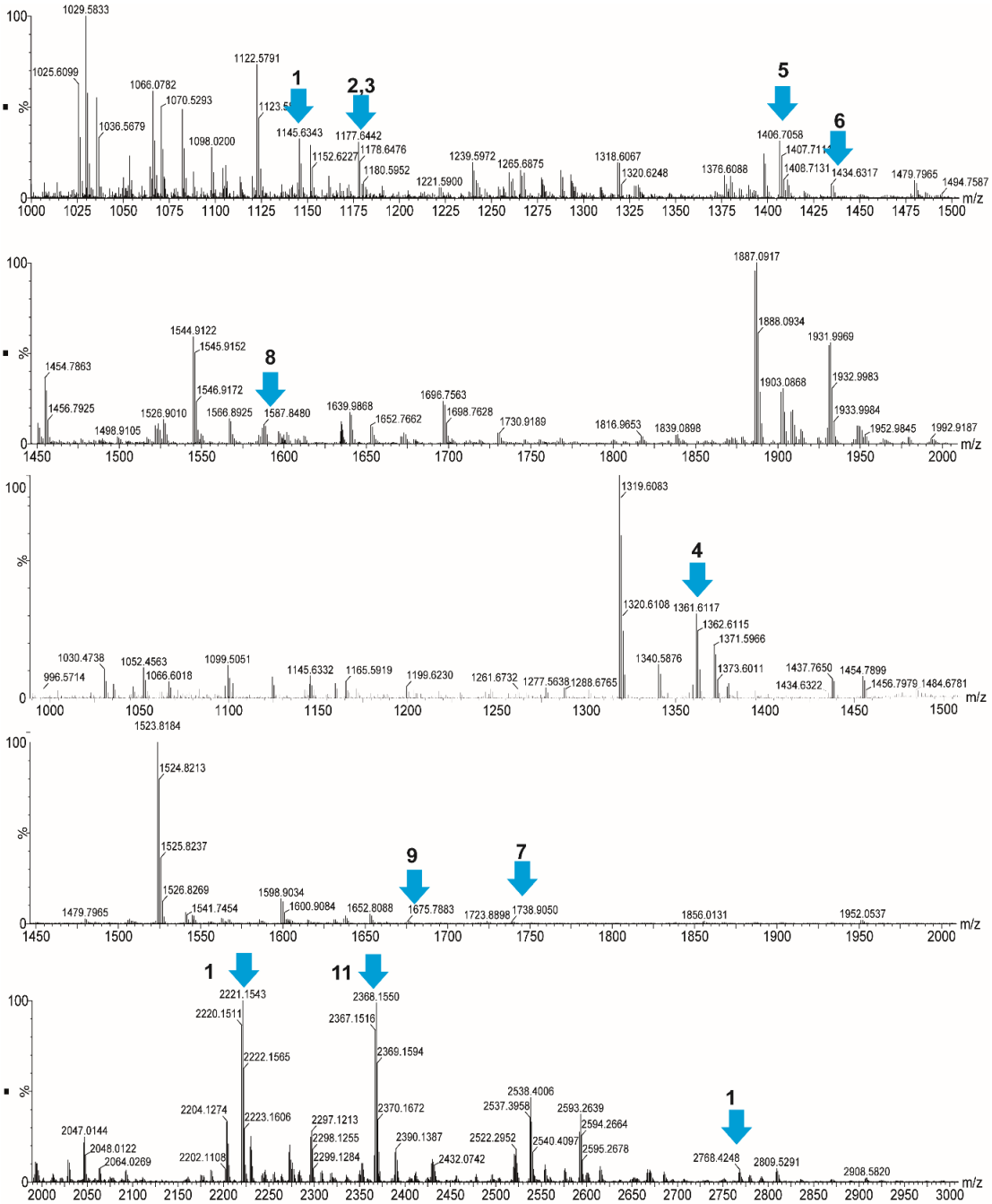
Supplementary figure 3: FGF2 MS spectra of peptides after in solution modification with PGO based on prediction by Prospector. The reference number of the peptide is followed by the observed m/z of the peptides produced by cleavage of FGF2 reacted in solution with PGO.



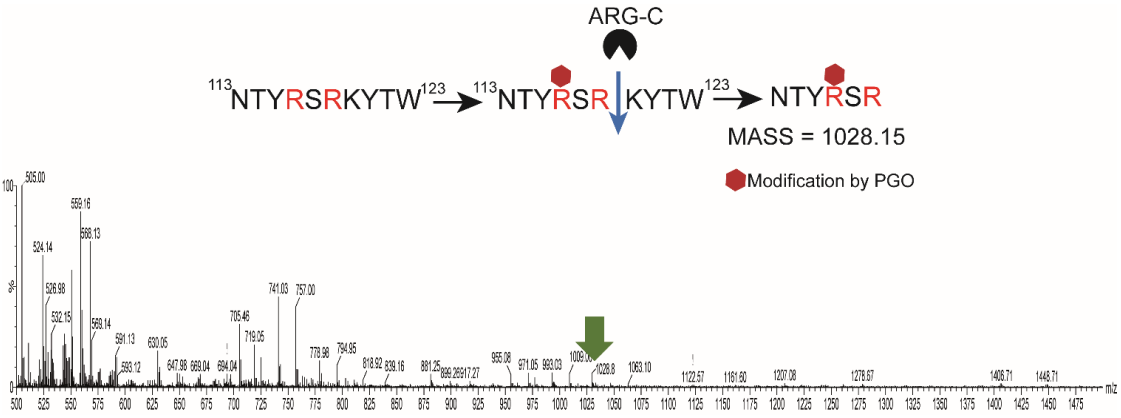
Supplementary figure 4:FGF2 MS spectra of peptides after modification based on prediction by Prospector. The reference number of the peptide is followed by the observed m/z of the peptides produced by the cleavage of FGF2which was before reacted with PGO when bound to a heparin affinity column (protection step) and then, following elution, reaction with HPG (labelling step).



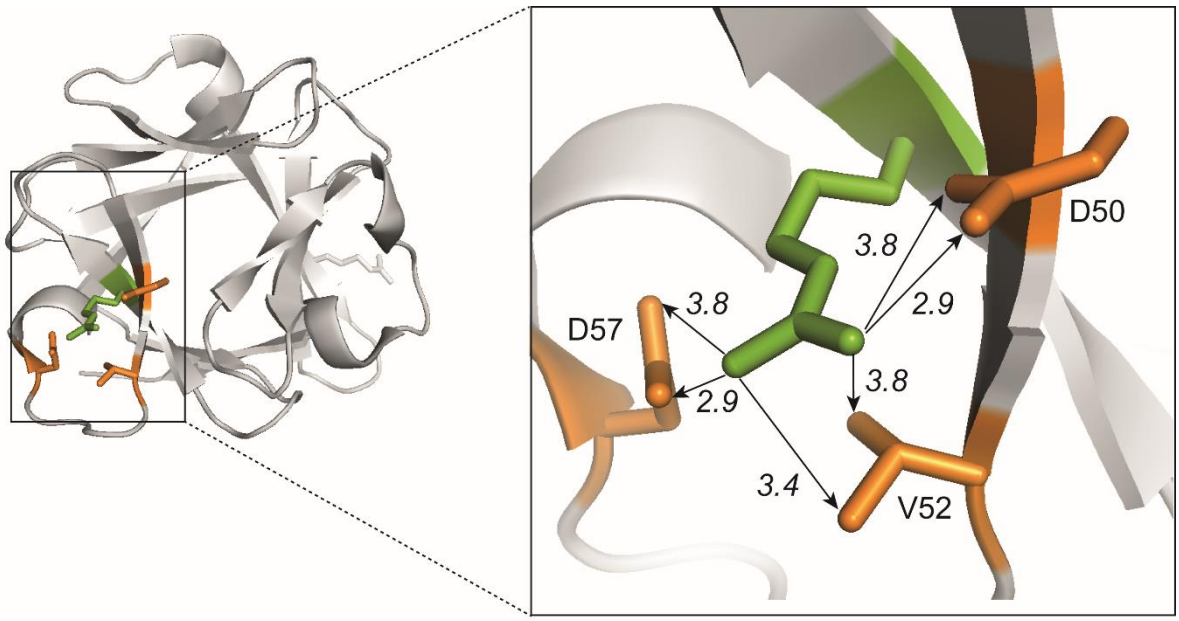
Supplementary figure 5: FGF2 MS spectra of peptides after modification based on prediction by Peptide Mass (ExPASy). The reference number of the peptide is followed by the observed m/z of the peptides produced by the cleavage of FGF2 which was before it reacted with PGO when bound to a heparin affinity column (protection step) and then, following elution, reaction with HPG (labelling step).



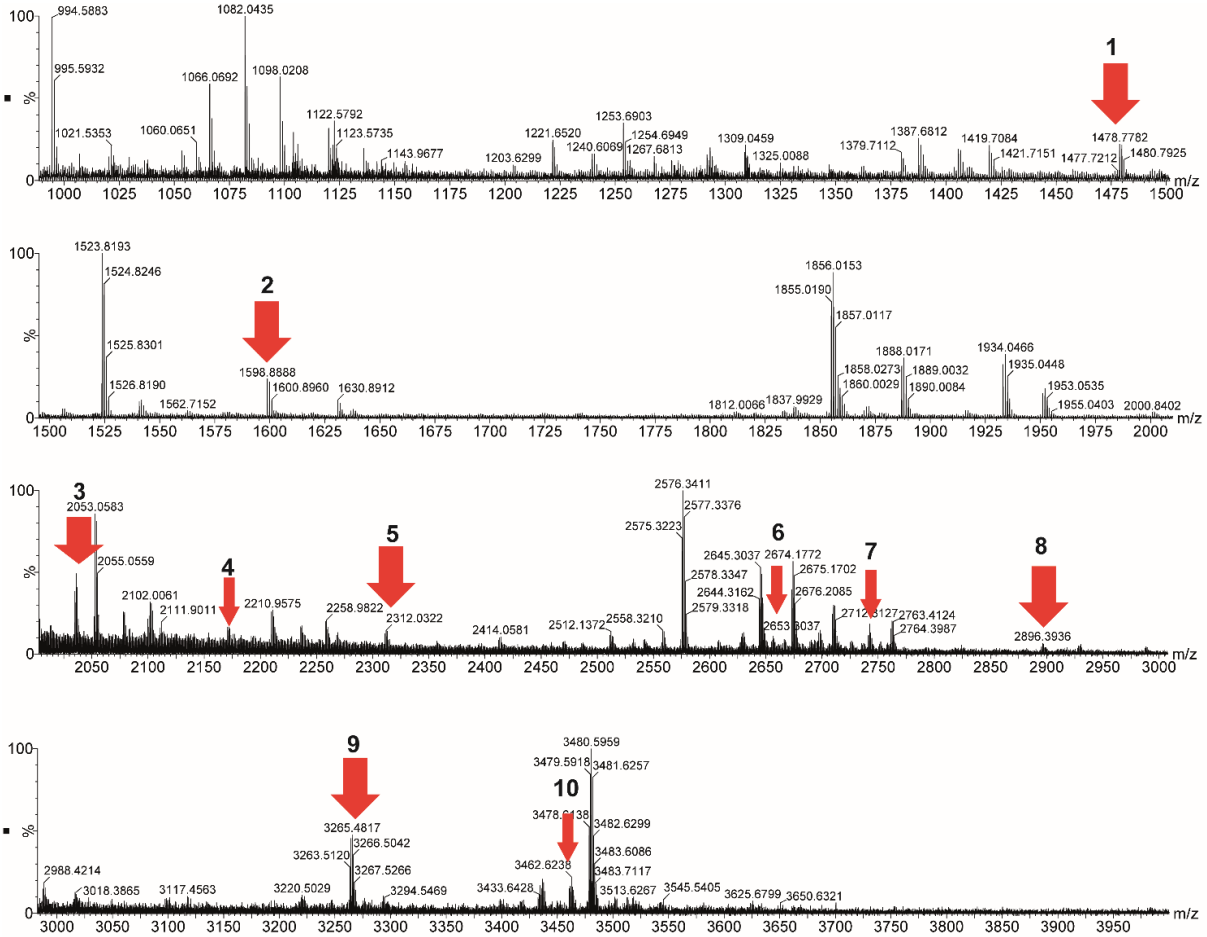
Supplementary figure 6: In FGF2, R118 engages to heparin whereas R116 does not. Double digestion of FGF2 by chymotrypsin and Arg-C. After the protection of exposed arginine residues on the mini-column by 200 mM PGO, protein was eluted from the column and then cleaved for 5 hours by chymotrypsin and overnight by Arg-C. The red diamond presents for the modification by PGO on R^{116}. R^{118} was cleaved by Arg-C resulting in peptide NTYRSR with the mass 1028.15, observed in the spectrum.



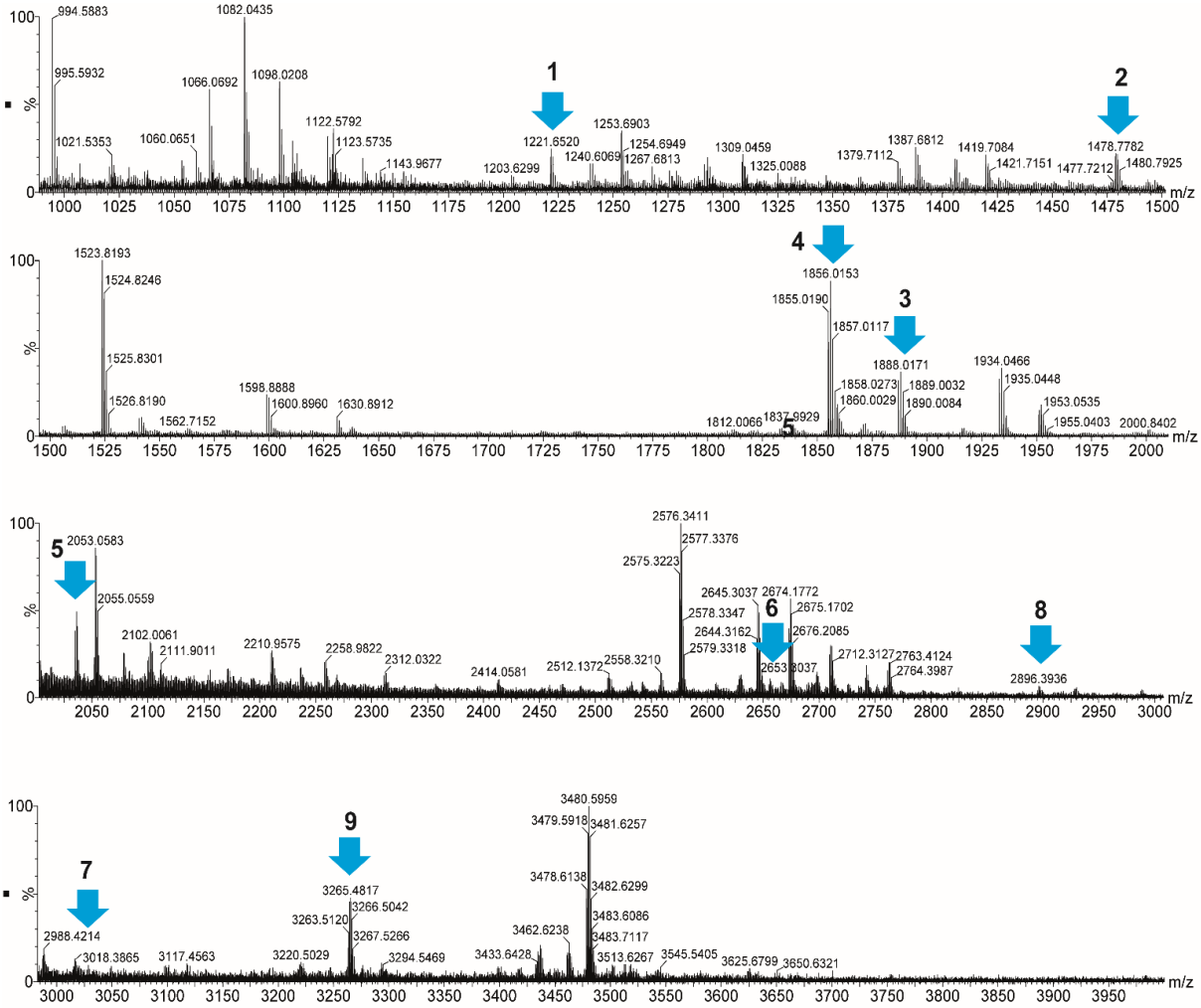
Supplementary figure 7: The intramolecular network of hydrogen bonds of R^{42} with D^{50}, V^{52}, and D^{57}. The *green* presents R^{42}; the *orange* shows the D^{50}, V^{52}, and D^{57}. Their potential hydrogen bonds are presented as backlines with arrow.



Supplementary figure 8: FGF1 MS spectra of peptides after modification based on prediction by Prospector. The reference number of the peptide is followed by the observed m/z of the peptides produced by the cleavage of FGF1which was before reacted with PGO when bound to a heparin affinity column (protection step) and then, following elution, reaction with HPG (labelling step).



Supplementary figure 9: FGF1 MS spectra of peptides after modification based on prediction by Peptide Mass (ExPASy). The reference number of the peptide is followed by the observed m/z of the peptides produced by the cleavage of FGF1which was before reacted with PGO when bound to a heparin affinity column (protection step) and then, following elution, reaction with HPG (labelling step).



Supplementary table 1: FGF-2 MS analysis based on prediction by Prospector

The reference number of the peptide is followed by the predicted m/z and with the sequence of the peptides following cleavage of native FGF-2 chymotrypsin. The two columns under "Native FGF-2" present the predicted m/z and sequences of peptides after chymotrypsin digestion of FGF-2.The first two columns under "FGF-2 in-solution" present the observed and predicted m/z for FGF-2 after modification by PGO[7]. The final column indicates the modification occurring on the arginine residues of the peptides.

	Native FGF2		**FGF-2 in solution**		
	m/z theoretical	**Sequence**	**m/z observed**	**m/z theoretical**	**Modifications (PGO)**
1	919.54	^{27}**KDPKR**LY^{33}	1151.60	1151.69	R + 232
2	1035.59	^{125}VAL**KRTGQY**^{134}	1151.62	1151.69	R + 116
3	1161.63	^{83}**LAMK**EDGRLL^{95}	1294.62	1294.62	1 R + 133 and oxidation
4	1083.70	^{116}**RSRKY**TSW^{123}	1315.62	1315.65	2 R + 116 x 2
5	1624.79	^{113}NTY**RSRKY**TW^{123}	1856.84	1857.01	2 R + 116 x 2
6	1907.90	^{92}LASKCVTDECFFFERL^{107}	2139.97	2139.80	R + 232
7	1979.01	^{65}QAEERGVVS**IKG**VCANRY^{82}	2211.00	2211.00	2 R + 116 x 2
8	2248.00	^{109}ESNNYNTY**RSRKY**TSW^{123}	2479.80	2479.90	2 R + 116 x 2
9	2424.32	^{41}LRIHPDGRVDGVREKSDPHIKL^{62}	2673.30	2673.17	2R + 116 + 133

Supplementary table 2: FGF-2 MS analysis based on prediction by Peptide Mass (ExPASY).

[7]In solution: FGF-2 reacted in solution with PGO for 60 min in the dark.

The reference number of the peptide is followed by the predicted m/z and with the sequence of the peptides following cleavage of native FGF-2 chymotrypsin. The two columns under "Native FGF-2" present the predicted m/z and sequences of peptides after chymotrypsin digestion of FGF-2.The first two columns under "FGF-2 in-solution" present the observed and predicted m/z for FGF-2 after modification by PGO[8]. The final column indicates the modification occurring on the arginine residues of the peptides.

	Native FGF-2		**FGF-2 in solution**		
	m/z theoretical	**Sequence**	**m/z observed**	**m/z theoretical**	**Modifications (PGO)**
1	1161.63	^{83}**LAMK**EDGRLL^{95}	1294.62	1294.62	1 R + 133 and oxidation
2	1624.79	^{113}NTY**RSRKY**TW^{123}	1856.84	1857.01	2 R + 116 x 2
3	1907.90	^{92}LASKCVTDECFFFERL^{107}	2139.97	2139.79	R + 232
4	1979.01	^{65}QAEERGVVS**IKG**VCANRY^{82}	2211.06	2211.00	2 R + 116 x 2
5	2248.00	^{109}ESNNYNTY**RSRKY**TSW^{123}	2480.06	2479.80	2 R + 116 x 2
6	2684.48	^{40}FLRIHPDGRVDGVREKSDPH IKL^{62}	3264.64	3264.49	3 R + 116 + 232 x 2

[8]In solution: FGF-2 reacted in solution with PGO for 60 min in the dark.

CHAPTER 5 THE IDENTIFICATION OF LYSINE AND ARGININE RESIDUES INVOLVED IN HEPARIN BINDING IN THE FIFTEEN PARACRINE FIBROBLAST GROWTH FACTORS

5.1. INTRODUCTION

In humans there are 22 fibroblast growth factors (FGFs), which regulate many aspects of embryonic development and cellular homeostasis. The 15 paracrine FGFs elicit signals by forming a ternary complex with fibroblast growth factor receptors (FGFRs) and their heparan sulfate (HS) co-receptors. In addition to its role as a co-receptor, HS also regulates the transport of FGFs between source and target cells and so their bioavailability and gradients [204] [2][36].

Thus, HS is a key regulatory partner of the paracrine FGFs. HS is composed of disaccharide repeating units comprising 1,4 linked –β-D glucuronate (GlcA) and α-D-*N*-acetyl glucosamine (GlcNAc) with diverse modifications: de-acetylation/*N* sulfation and *O* sulfation at C3 and C6 of glucosamine, epimerisation of GlcA to L-iduronic acid (IdoA) and 2-*O* sulfation of this uronic acid [47]. The clustering of these modifications results in HS chains possessing a domain structure: S-domains, where all glucosamine residues are *N*-sulfated, followed by transition domains with alternating GlcNS and GlcNAc containing disaccharides, and unmodified NA domains [57]. Heparin is widely used *in vitro* as a proxy for S-domains due to its high degree of sulfation. On each saccharide, the sulfate and carboxylate groups are completely ionized at physiological pH, and this high negative charge density results in electrostatic attraction driving the interaction between HS/heparin and protein [47].

The so-called heparin binding sites (HBSs) on proteins are enriched in arginine and lysine residues and consequently basic [114] [36][109]. However, it is well established that the interaction of proteins with sulfated GAGs is based on affinity and not simply ion-exchange

[148]. Some attempts have been made to determine a consensus sequence of basic amino acids representative of HBSs [110] [112]. However, such linear sequences of amino acids have not been generally useful. This is because HBSs are often formed by residues that are physically adjacent, but distant in the sequence [109]. For example, HBSs of FGFs locate to various β-strands and loops [120] [36] [37]. It is also noteworthy that proteins often possess multiple HBSs. In at least some instance, this allows the protein to engage multiple HS chains, reducing the protein mobility [36]. Moreover, there are far more identified HBPs than HBPs with defined HBSs, so there is currently limited scope for developing a means to predict HBSs.

The identification of the binding sites of the polysaccharide on proteins includes low throughput methods such as NMR spectroscopy [184], site-directed mutagenesis [185] [36][186] and X-ray crystallography [100] [143] [144]. A higher throughput method to identify lysine residues, called "protect and label" uses selective labelling and mass spectrometry [120][36][37], and has increased the number of characterised HBSs considerably. An interesting feature of this method is that it identifies both canonical, high affinity HBSs in a protein and lower affinity secondary binding sites [36] [37]. However, this approach is limited due to its selectivity for lysine and so can provide no information on arginine residues.

Recently, an analogous selective labelling approach has been developed to identify arginine residues engaged with heparin (Section 4.2). The application of both lysine and arginine selective labelling thus enables the complete characterisation of the key basic residues in an HBS. In previous work [36][120] [37], we determined the lysine residues contributing to heparin binding in 12 of the 15 paracrine FGFs and arginine residues in FGF1 and FGF2. This provided insight into the architecture of the HBSs and their evolutionary relationship. Here, we complete the analysis of the lysine and arginine residues involved in heparin binding in all 15 paracrine FGFs.

The results underline the importance of the area between β-strands X and XII in engaging heparin and the critical contributions of other basic residues scattered across the sequence. The data demonstrate that while some FGFs have a single, continuous canonical HBS-1, many possess multiple HBSs, separated by an acidic border and changes in the positions of acidic residues can make important contributions to the diversification of HBSs in FGFs. In addition, HBS-3, which locates N-terminal to the core β-trefoil of FGFs is found in many of the FGFs and appears likely to overlap an FGFR binding site.

5.2. RESULTS

The selective labelling of lysine and arginine residues involved a protection step, in which the solvent exposed residues of protein bound to a heparin affinity columns are reacted. Following elution, any residues that were engaged with heparin were now exposed to solvent and could be labelled using the same reaction, but a different reagent. For lysine residues, the analyses are straight forward, since the reaction of N-hydroxy-succinimide functional groups with primary amines yields a single acidic stable product. Moreover, the use of NHS-biotin simplifies analysis, since biotinylated peptides are readily purified prior to mass spectrometry [120].

In contrast, for arginine residues the analysis is more complex. Phenylglyoxal (PGO) is used to protect arginine side chains exposed to solvent in a heparin-bound protein. Hydroxyl-phenylglyoxal (HPG) is then used to label exposed arginine side chains in the eluted protein. The complexity arises from the fact that the reaction of PGO with arginine side chains produces multiple products, which are acid labile [192]. The multiple products can be deconvoluted using newly developed software tools (Section 4.1), whereas the acid lability imposes a restriction on post-reaction processing of peptides.

We have used the lysine and arginine protect and label methods to complete the analysis of HBSs in the 15 paracine FGFs. The identified heparin-binding residues were then mapped onto existing modelled structures of the FGFs. This allowed a definitive assignment of the labelled residues between primary, canonical HBS and secondary HBSs based on the presence and absence of isolating acidic borders. The mass spectrometry data are in supplementary.

5.2 RESULTS

5.2.1 FGF4 subfamily (FGF4/5/6)

5.2.1.1 FGF4

Evaluation of the conditions used in the reaction between FGF4 and PGO

The efficiency of the reaction between PGO and the arginines in FGF4 was examined first, using Arg-C, a protease which specifically cleaves at the arginine amino acids and the resulting products were analyased by SDS-PAGE and silver staining (Fig 5.1). The process is described for FGF4 only, but this evaluation step was applied to all analysed FGFs (Supplementary Figs 5.1-5).

The reaction between FGF4 and PGO was initially tested in solution at room temperature and for 60 mins. The reaction products were subsequently subjected to cleavage by chymotrypsin as a control, and by Arg-C to be able to confirm the completion of the reaction (Fig 5.1A). There was no band detectable when FGF4 was incubated with chymotrypsin or Arg-C, demonstrating that the enzymes were active. When FGF4 was reacted with PGO, chymotrypsin was still able to cleave the protein, but Arg-C failed to do so (Fig 5.1A). These results demonstrated that all arginine residues exposed on the surface of FGF4 had reacted with PGO. Subsequently, the reaction between FGF4 and PGO was performed on the mini heparin affinity column (Fig 5.1B, C). The FGF4 eluted from the heparin affinity column remained sensitive to digestion by Arg-C (Fig 5.1B), showing that the arginine residues not engaged were now

exposed to solvent. The eluted FGF4 was then the subjected to a second reaction with PGO. The product of this reaction was then applied to a heparin affinity column, and it failed to bind (Fig 5.1C). The flow through fraction was then treated with chymotrypsin and Arg-C. A band at equivalent molecular weight and concentration as the protein in the flow through fraction was observed in the land of Arg-C cleave product (Fig 5.1C), demonstrating that all arginine residues of FGF4 had reacted with PGO (Fig 5.1C). Thus, while binding to heparin protects some arginine residues, once the protein was eluted it did not affect their reaction with this dialdehyde.

Selective labelling of arginine residues on FGF4

Solvent-exposed arginine residues in heparin-bound FGF4 were protected with PGO and, following elution of the FGF4, arginine side chains engaged with the polysaccharide were labelled with HPG. After processing the protein, chymotryptic peptides were generated and analysed by MALDI-TOF mass spectrometry. The resulting peptides with information about modification, sequence, and final m/z are presented in Tables 5.2A and 5.2B and the spectra are in Supplementary Figs 5.1A and 5.1B.

Prospector

Using the list of peptides generated by Prospector, ten modified-arginine-containing peptides were identified (Table 5.2A). N-terminal to β-strand I, R^{45} and R^{46} were found to be labelled by HPG (peptide 1 and peptide 5, Table 5.2A) and so engaged to heparin, whereas R^{57} reacted with PGO (peptide 8, Table 5.2A), so was not involved in binding to heparin. R^{82}, R^{84} and R^{85} are close on the sequence, but the corresponding peptide possessed one HPG and two PGO products (peptide 2, Table 5.2A), indicating that only one of them was engaged to heparin. Neither R^{103} nor R^{112} were bound to heparin, since peptide 7 had two PGO products (peptide 7, Table 5. 2A). There were three peptides containing R^{123} and R^{134} in their sequence, peptides

4, 6 and 9 (Table 5. 2A). They all showed one modification by PGO and one by HPG, indicating that one arginine was bound to heparin and one was solvent exposed. However, R^{134} is uniquely in peptide 3, which possessed just the product of the PGO reaction. Together, this indicated that R^{123} bound to heparin, but R^{134} did not. Located in the highly conserved area of the canonical HBS-1 is R^{192}, which, as expected, reacted with HPG (peptide 10, Table 5. 2A), and thus was engaged to heparin.

Peptide Mass (ExPASY)

In contrast to Prospector, Peptide-Mass predicted eight peptides (Table 5.2B) (Supplementary figure 1B) and the binding status of arginine residues was consistent with that predicted by Prospector (Table 5.2A), although peptides containing R^{192} and R^{57} were missing. Thus, peptide 1 (Table 5.2B), which contained R^{45} and R^{46}, was found to carry two HPG molecules, indicating that both arginine residues were engaged to heparin. One of three arginine residues, R^{82}, R^{84} and R^{85} of the β-strand I reacted with HPG (peptides 2 and 6– Table 5.2B), whereas the other two reacted with PGO, due to the observed mass shifts of two PGO reaction products. R^{103} and R^{112} reacted with PGO, so were exposed to the solvent and not involved in binding to heparin (peptide 4, Table 5.2B). R^{123} in the loop between β-strands IV and V (peptides 3, 5, 7 and 8, Table 5. 2B) reacted with HPG, hence were bound to heparin.

Organisation of HBSs

The original analyses of FGF1 and FGF2, using peptide mapping and mutagenesis indicated that the core of the canonical HBS-1 was located in the loop between β-strands X and XI, through β-strand XI until the loop of β-strands XI–XII [100], [143], [144]. Subsequent work, including X-ray crystallography and NMR confirmed this [29] [148]. Sequence alignment showed that this region was very basic in all the paracrine FGFs. The analysis of the lysine residues using the protect and label method led to the canonical heparin binding site, which

encompasses this region being called HBS1, with secondary binding sites in FGFs labelled HBS2-4 [36] [102]. In the case of FGF4, it was considered to have, in addition to the canonical HBS1, a secondary HBS, HBS3, situated towards the N-terminus of the beta-trefoil structure [102].

Thus, previous work demonstrated that HBS1 comprised K^{183}, K^{186}, K^{188} and K^{189} in the core along with K^{142} and K^{144} on the loop of β-strands VI – VII and K^{173} [37] (Fig 5.2F). Even though distant in the sequence, the labelled R^{123} in the loop between β-strands IV – V loop is adjacent to K^{158} on the surface (Fig 5.2B, C), suggesting that this arginine also belongs to HBS-1. Thus, the present work adds $R^{123}$and R^{192} to HBS1 (Figs 5.2A, F).

FGF4's secondary HBS, HBS3, was considered to comprise K^{81} in β-strand I (Fig 5.2B, C) and K^{158} in β-strand VIII (Fig 5.2B). One of three arginine residues, R^{82}, R^{84} or R^{85} of β-strand I reacted with HPG (peptide 2– Table 5. 2A and peptide 2 and 6– Table 5. 2B). Of these, R^{82} is adjacent to both K^{81} and R^{123}, which indicates that R^{82} is most likely the one to have reacted with HPG and it would then belong to the same HBS (Fig 5.2B, C). Inspection of the surface charge distribution shows some acidic areas between HBS1 and what was defined as HBS3 (Fig 5.2 A, B, C), which is what led to the original assignment of K^{81} and K^{158} to HBS3 [37]. However, these acidic areas may not be able to truly separate these heparin binding regions on the surface of FGF4 (Fig 5.2E). Thus, although K^{158} of β-strand VIII (Fig 5.2B) is separated from to K^{173}, assigned to HBS1 (Fig 5.2B, C) by an electronegative surface (Fig 5.2B, C), this is due to the peptide backbone and Y^{172} (Fig 5.2E), rather than a negatively charged amino acid side chain. It is unlikely to repel the negative charges on the sugar and may be able to interact through H-bonding with it. This implies that the sugar chain could bridge between K^{173} and K^{158} and so consequently it is likely that K^{158} and the adjacent K^{81} are both part of HBS-1. Moreover, many of the lysine and arginine side chains in FGF4 found to interact with heparin

protrude from the protein surface, which would reduce the effects of the backbone on sugar binding (Fig 5.2D). The alternative arrangement, whereby the polysaccharide chain binds on the other face of FGF4, bridging from the conserved area of HBS1 (K^{140}, K^{188}, K^{189} and R^{192}) to K^{81} and R^{82} (Fig 5.2D) is not supported by the present data. Though there are three arginine residues R^{103}, R^{112} and R^{134} in between these areas, they were all found to have reacted with PGO (R^{103}, R^{112} (peptide 4, Table 5.2B) and R^{134} (peptide 3, Table 5.2A)), and so were not involved in binding heparin.

Therefore, there is only one way the polysaccharide can bind FGF4, and the most parsimonious explanation is that the residues previously assigned to HBS3 are in fact an extension of HBS1. This would require the sugar chain to bend around the protein, which is seen in a crystal structure of a heparin decasaccharide with FGF1 [201] and in the interaction of HS with VEGF [205] [83] and IFN-gamma [206].

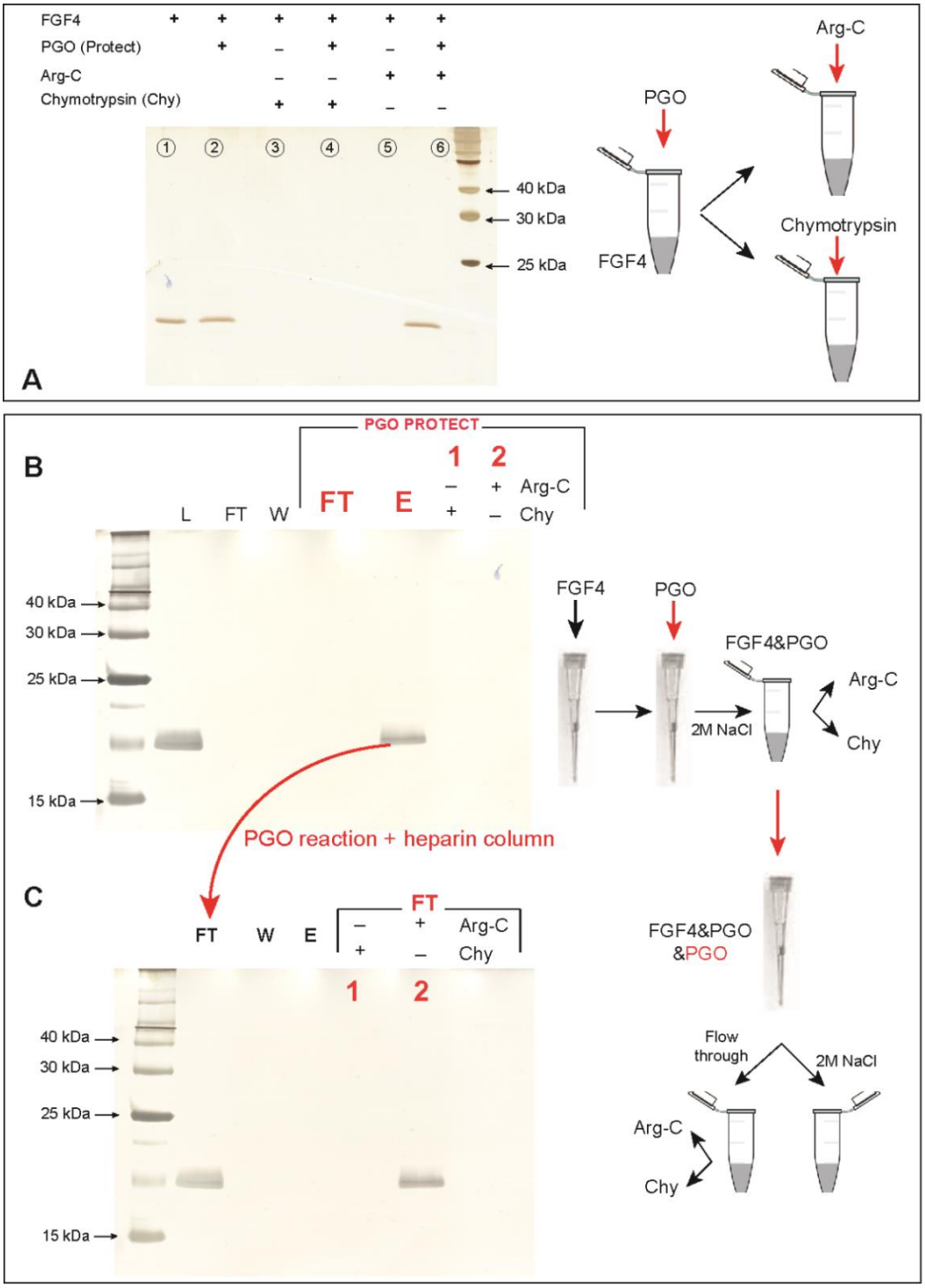


Figure 5.1: Selective labelling of arginine residues in the heparin binding sites of FGF4.

The reaction of PGO with FGF4 in the dark, at room temperature (~25 ^{o}C) for 60 min. The efficiency of the reaction was evaluated by cleavage of chymotrypsin and Arg-C (incubation at 37 $^{\circ}$C overnight). Equal amount of protein were then analysed by SDS-PAGE. **A) The**

analysis of the products of reaction of FGF4 in solution with PGO by Arg-C digestion. FGF4 (~10 µg) was reacted in solution with 200 mM PGO (Right panel). Lane 1: The native FGF4, 1/10 equivalent of that used in the reaction; lane 2: 1/10 of the FGF4 after the reaction with 200 mM PGO; lane 3: products of the digestion of 1/10 equivalent of native FGF4 by chymotrypsin (Chy); lane 4: products of the digestion by chymotrypsin of 1/10 equivalent of PGO-reacted FGF4; lane 5: products of the digestion by Arg-C of 1/10 equivalent of native FGF4; lane 6: products of the digestion by Arg-C of 1/10 equivalent of PGO-reacted FGF4. **B) Analysis of the products of reaction of FGF4 bound to a heparin affinity column with PGO**. FGF4 (~10 µg) was applied to a heparin affinity mini column. The flow through (FT) was collected then PGO (200 mM) was applied. The eluate was then subjected to digestion with Arg-C and Chy (Right panel). Lane L, 1/10 equivalent of the protein loaded onto the column; lane FT, 1/10 equivalent of the flow through fraction; lane W, 1/10 equivalent of the fraction washed with Na-1 buffer; lane PGO PROTECT FT, flow through fraction after the reaction between FGF4 and PGO; lane PGO PROTECT E, Eluate from the column with Na-2 buffer; lane PGO PROTECT 1: the protein in the elution after incubation with chymotrypsin; lane PGO PROTECT 2: the protein in elution after incubation with Arg-C. **C) Analysis of PGO reacted with arginine residues in binding sites of FGF4**. FGF4 bound to heparin (panel B, fraction PGO PROTECT E) was reacted with 200 mM PGO for 60 min, then the product from this second reaction was loaded onto a new heparin affinity mini-column. Lane FT, 1/10 equivalent of the flow through from the mini-column; lane W, 1/10 equivalent of the wash with Na-1 buffer; lane E, 1/10 equivalent of elution with buffer Na-2; lane FT 1, 1/10 equivalent of the FGF4 from the FT fraction after incubation with chymotrypsin; lane FT 2, 1/10 equivalent of the FGF4 from the FT fraction after incubation with by Arg-C.

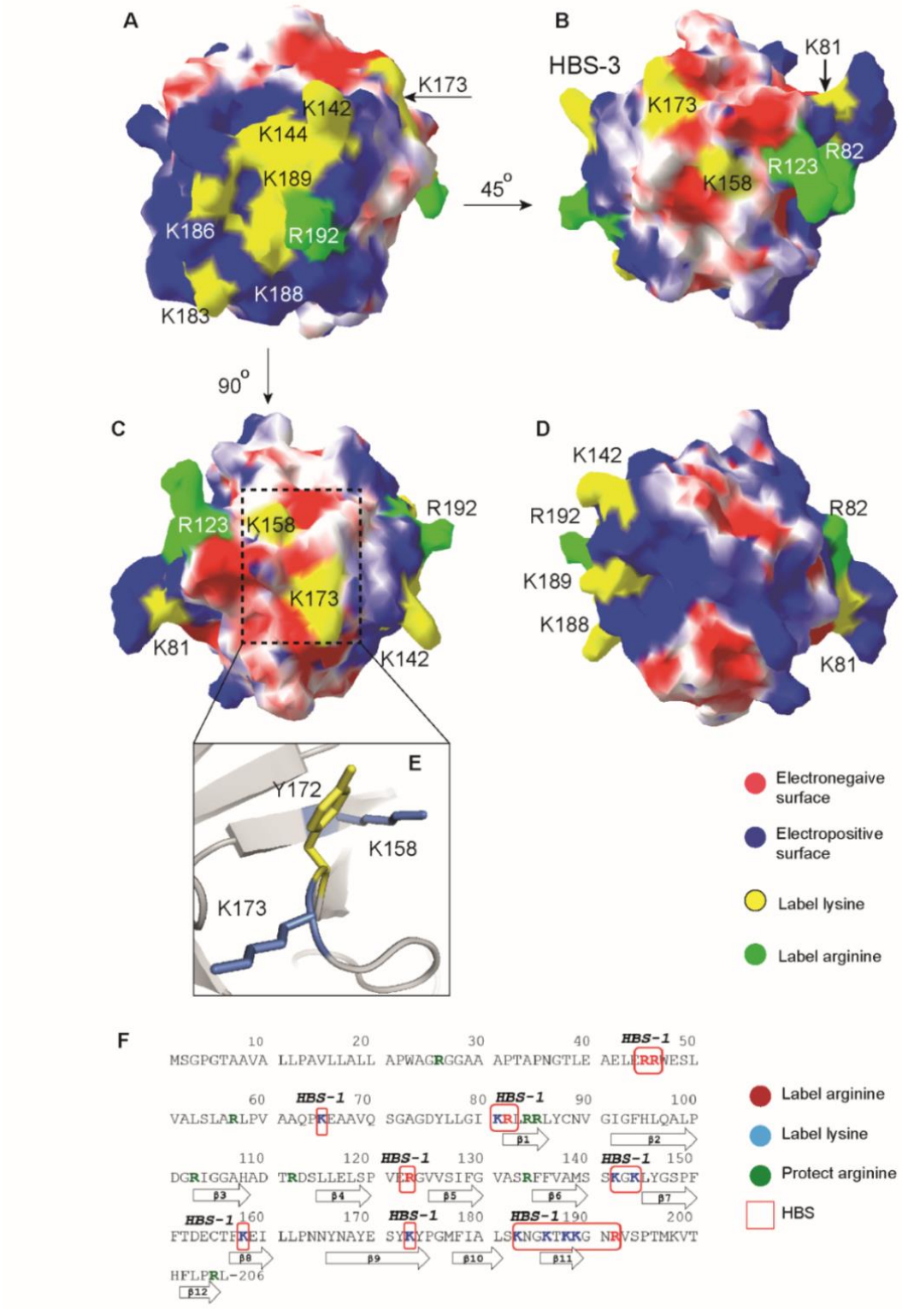

A
K173
K142
K144
K189
K186
R192
K188
K183
45°
B
HBS-3
K81
K173
R82
R123
K158
90°
C
R192
R123
K158
K173
K81
K142
D
K142
R192
R82
K189
K188
K81
E
Y172
K158
K173
Electronegaive surface
Electropositive surface
Label lysine
Label arginine
F
HBS-1
HBS-1
HBS-1
HBS-1
HBS-1
HBS-1
HBS-1
MSGPGTAAVA LLPAVLLALL APWAGRGGAA APTAPNGTLE AELERRWESL
VALSLARLPV AAQPKEAAVQ SGAGDYLLGI KRLRRLYCNV GIGFHLQALP
DGRIGGAHAD TRDSLLELSP VERGVVSIFG VASRFFVAMS SKGKLYGSPF
FTDECTFKEI LLPNNYNAYE SYKYPGMFIA LSKNGKTKKG NRVSPTMKVT
HFLPRL-206
β1
β2
β3
β4
β5
β6
β7
β8
β9
β10
β11
β12
Label arginine
Label lysine
Protect arginine
HBS

Figure 5.2: Heparin binding arginine and lysine residues in FGF4.

The FGF4 structure (PBD code **1ijt** [207]) is shown as a surface. The electrostatic potential of FGF4 was computed using the Poisson-Boltzmann algorithm in Swiss-PDBView with the positively charged areas coloured *blue* and the negatively charged areas coloured *red*. Labelled lysine residues from [102] are presented as *yellow*. Labelled arginine residues (Tables 5.2A, B) are coloured *green*. This colour scheme for electrostatic potential is used throughout. **A)** The position of R^{192} (peptide 10, Table 5. 2A) relative to the core canonical HBS-1 K^{189}, K^{188}, K^{186}, K^{183}, K^{144} and K^{142} [102]. **B)** Location of R^{123} (peptide 4, 6 and 9, Table 5. 2A) (peptides 3, 5, 7 and 8, Table 5. 2B) and R^{82} (peptide 8, Table 5. 2A) (peptides 2 and 6– Table 5. 2B) along with K^{158} and K^{81} previously assigned to HBS-3 are shown on the surface. **C)** The surface between K^{173} and the core HBS-1 is positively charged, whereas the surface between K^{173} and K^{158} is negatively charged. **D)** K^{81} and R^{82} are separated from K^{189}, K^{188}, K^{142} and R^{192} of HBS-1 by a positively charged area, and these side chains protrude from protein surface. **E)** The electronegative path between K^{158} and K^{173} has no acidic residues but a tyrosine side chain, Y^{172}. **F) Sequence, secondary structure and location of labelled arginine and lysine residues.** Uniprot ID of FGF4 sequence is **P08620-1**. Labelled arginine and lysine [102] residues are coloured in *red* and *blue*, respectively and the protected arginine residues are *green*. The HBSs assigned in reference [102] are highlighted in red boxes. The β strands are presented as arrows. This colour scheme for heparin engaging basic residues is used throughout the thesis.

5.2.1.2 FGF6

N-terminal to β-strand I are R^{60} and R^{62}, which were both modified by HPG (peptides 1 and 6, Table 5. 3A and peptide 1, Table 5. 3B). Adjacent are R^{84}, R^{86} and R^{87} in peptide 3 (Tables 5.3A, B), which possessed two products from reaction with PGO and one product from reaction

with HPG, indicating that only one of these arginines was engaged to heparin. The same was observed in FGF4, as these four residues "KR(L/Q)RR" are conserved between these two FGFs. Hence, it is likely that as in FGF4, R^{84} of FGF6 engages with heparin. One of R^{125} or R^{134} was found to have reacted with HPG and the other with PGO (peptide 7, Table 5.3A and peptide 8, Table 5.3B). On the loop from β-strand X to β-strand XII, two arginine residues, R^{188} and R^{191}, reacted with HPG (peptide 4 – Table 5.3A and peptides 4, 7- Table 5.3B), hence interacted with heparin. R^{205} of the disordered C-terminus (peptide 2, tables 3A and 3B) was also bound to heparin, since it was modified by HPG.

Organisation of HBSs

As for FGF4, previous lysine selective labelling and sequence alignment were interpreted to indicate that FGF6 had a canonical HBS1 and a secondary HBS3, N terminal to beta strand 1 [37]. The current data add five arginine residues to the HBSs of FGF6, and the independence of the residues previously assigned to HBS3 from HBS1 seems less likely. The HBS1 of FGF6 according to previous work (Figure 3A) comprises K^{185}, K^{190}, K^{194} of β-strand X – β-strand XI [37]. Located in this region and so part of this canonical binding site are two arginine residues R^{188} and R^{191} (Fig 5.3A, F). K^{144} of β-strand VII is next to R^{191} (Fig 5.3A, E) and was also previously assigned to HBS-1 [37].

Between K^{185} of HBS-1 and R^{205} is a positively charged surface (Fig 5.3C) and the distance between R^{205} and K^{185} is 2.4 nm, which is just over a tetrasaccharide. These observations suggest that R^{205} is likely part of HBS-1. There is also a basic path running from R^{188} of HBS-1 to K^{83} (Fig 5.3C), but arginine residues (R^{134}, R^{146} (peptide 5, Table 5.3A, B)), which lie along this path were protected rather than labelled. Thus, the polysaccharide is unlikely to engage FGF6 in this orientation. Neighbouring R^{205} on the surface are K^{83} and R^{84} (Fig 5.3D). K^{83} was previously assigned to HBS-3 [37], however, with the addition of R^{205} and R^{84} to

HBS1, this lysine should be reallocated to HBS-1. On the same basic surface with K^{83} and R^{84} is R^{125} of β-strand V (Fig 5.3D), indicating that R^{125} belongs to HBS-1 as well. R^{125} and K^{158} are close on the surface (Fig 5.3D), hence heparin can interact with both R^{125} and K^{158}. K^{158} of β-strand VIII is separated from K^{144} and K^{194} of HBS-1 by an acidic border (Fig 3B), which was one reason for its assignment to HBS3 [37]. Here, there is a major difference between FGF4 and FGF6 in this area, since in FGF4 is the basic K^{173} of β-strand IX, but in FGF6 an acidic D^{174} is present. Thus, FGF6 has a true acidic border on this part of its surface. However, K^{158} could be connected to HBS-1 from the opposite direction through R^{125} as mentioned above, making this lysine part of HBS-1. Moreover, the side chains of R^{205}, R^{84}, R^{125} and K^{158} protrude from protein surface (Fig 5.3B, C, D). Consequently, the simplest assignment of labelled residues to HBSs in FGF6 is that those residues previously assigned to HBS3 are in fact an extension of HBS1. A single long HBS1 is consistent with FGF6 and FGF4 requiring a DP12 for binding (length ~ 6 nm) [37]. Moreover, this conclusion supports the measurement of crosslinking of HS brushes [126] performed on FGF6, which was found not to cross-link HS chains, a property which is thought to require at least two independent HBSs [203].

Unassigned are R^{60} and R^{62} in the N-terminus of FGF6, which cannot be included in a structural model, as the corresponding template, FGF4 [207], does not include this part of the protein.

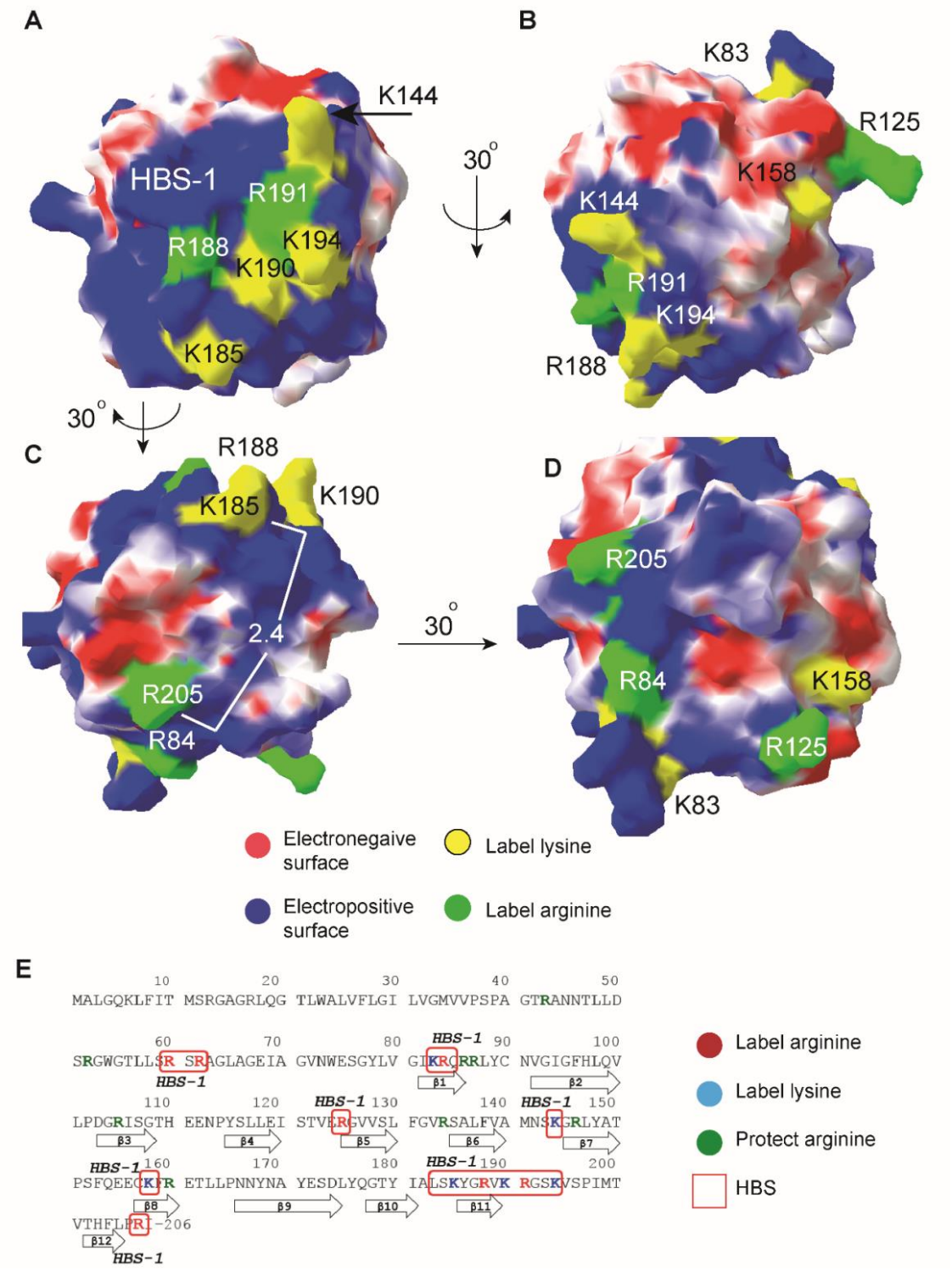

A
K144
HBS-1
R191
R188
K194
K190
K185
30°
B
K83
R125
K158
K144
R191
K194
R188
30°
C
R188
K185
K190
2.4
R205
R84
30°
D
R205
R84
K158
R125
K83
Electronegaive surface
Label lysine
Electropositive surface
Label arginine
E
MALGQKLFIT MSRGAGRLQG TLWALVFLGI LVGMVVPSPA GTRANNTLLD
SRGWGTLLSR SRAGLAGEIA GVNWESGYLV GIKRQRRLYC NVGIGFHLQV
LPDGRISGTH EENPYSLLEI STVERGVVSL FGVRSALFVA MNSKGRLYAT
PSFQEECKYR ETLLPNNYNA YESDLYQGTY IALSKYGRVK RGSKVSPIMT
VTHFLPR-206
HBS-1
β1
β2
β3
β4
β5
β6
β7
β8
β9
β10
β11
β12
Label arginine
Label lysine
Protect arginine
HBS

Figure 5.3: Heparin binding arginine and lysine residues in FGF6. The surface of FGF6 is generated on the FGF4 crystal structure (PDB code: **1ijt** [207]). The electrostatic potential of FGF6 was computed using the Poisson-Boltzmann algorithm in Swiss-PDBView with the positively charged areas coloured *blue* and the negatively charged areas coloured *red*. Labelled lysine residues from [102] are presented as *yellow*. Labelled arginine residues (Tables 3A, B) are coloured *green*. This colour scheme for electrostatic potential is used throughout. **A)** The position of R^{188} and R^{191} (peptide 4 – tables 3A and peptide 4, 7- Table 5.3B) relative to the core canonical HBS-1 K^{144}, K^{185}, K^{190} and K^{194} [102]. **B)** Location of R^{125} (peptide 7, Table 5.3A and peptide 8, Table 5.3B) and R^{84} (peptide 3 (Table 5.3A, B)) along with K^{83} previously assigned to HBS-3 are shown on the surface. **C)** Location of R^{205} of the disorder C-terminus (peptide 2, tables 3A and 3B) is shown on the surface. The surface between R^{205} and the core of HBS-1 is positively charged. The distance between R^{205} and K^{185} of HBS-1 is presented. **D)** K^{158} and R^{125} are separated from K^{83}, R^{84} and R^{205} of the extended area of HBS-1 by a positively charged area. **E) Sequence, secondary structure and location of labelled arginine and lysine residues.** Uniprot ID of FGF6 sequence is: P10767-1. Labelled arginine and lysine [102] residues are coloured in *red* and *blue*, respectively and the protected arginine residues are *green*. The HBSs assigned in reference [102] are highlighted in red boxes. The β strands are presented as arrows.

5.2.1.3 FGF5

FGF5 has not previously been analysed. Consequently, both lysine and arginine targeted protect and label were performed on the protein.

Lysine

Located N-terminal to β-strand I are K^{24} (peptide 3, Table 5.4C) and K^{29} (peptide 8, Table 5.4C), which were biotinylated, indicating their engagement to heparin. Peptide 1 (Table 5.4C)

possessed three lysine residues, K^{146}, K^{147} and K^{149} on the loop of beta-strands VI-VII and all were biotinylated, indicating that these bound to heparin. Similarly, $K^{194,}$ K^{197} and K^{199} in the area covering the loop of beta-strands X-XI through to β-strand XI and the loop of β-strands XI–XII all reacted with NHS biotin (peptide 9, Table 5.4C). K^{183} (peptide 6, Table 5.4C) was found to react with NHS-biotin, indicating its engagement to heparin. K^{220} located at the end of β-strand XII was also engaged to heparin as observed in peptide 4 (Table 5.4C). Finally, on the disordered C-terminus, K^{254} and K^{260} (peptide 5, Table 5.4C) were both biotinylated, indicating their binding to heparin.

Arginine

Adjacent to K^{24} described above is R^{25}, which reacted with HPG, so was also engaged to heparin (peptide 4, Table 5.4B). On β-strand I, there were two labelled residues R^{86} and R^{87} (peptides 2, 7 and 8, Table 5.4A) (peptide 5, Table 5.4B). Next in the sequence are R^{163} and R^{165} in β-strand VIII, which both bound to heparin (peptides 3 and 4, Table 5.4A) (peptide 2, Table 5.4B). On the loop of β-strands IX-X, R^{180} and R^{186} (peptide 5, Table 5.4A) (peptide 7, Table 5.4B) were found to react with HPG, indicating interaction with heparin. Two arginine residues R^{195} and R^{200}, both reacted with HPG (peptides 9 and 10, Table 5.4A) (peptide 8, Table 5.4B), following by R^{218} (peptide 1, Table 5.4B). Lastly, on the disordered C-terminus, R^{253}, R^{262} and R^{266} (peptide 1 and 6, Table 5.4A) all interacted with heparin, since peptides 1 and 6 possessed the mass shift of 3 HPG modifications.

Organisation of HBSs

There is no prior study on the binding sites of FGF5 to sulfated GAGs, and the only information is the prediction on the basis of sequence alignment with FGF4 and FGF6 [37]. Hence, with the data on lysine and arginine residues that were bound to heparin presented here, the identification of heparin binding sites on FGF5 was addressed for the first time. There is also

no crystal structure of this FGF, consequently, a template had to be selected for the modelling of FGF5 surface. It is actually challenging, because unlike other FGFs (except FGF3), there is a controversy in the classification of FGF5 in the FGF4 or FGF2 subfamily [9]. Consequently, two model structures of FGF5 were built, using the crystal structure of FGF2 (4oee) (Fig 5.4A) and FGF4 (1ijt) (Fig 5.4B) as templates.

The core of HBS1 is well conserved across the FGF family and locates in the loop of β-strand XI and β-strand XII. Hence, K^{194}, R^{195}, K^{197}, K^{199} and R^{200} together delineate the core of FGF5's HBS-1 (Fig 5.4A, B). Vicinal to this area on the surface are K^{146}, K^{147} and K^{149} on the loop of beta-strands VI-VII. This indicates that these lysine residues are part of HBS-1 (Fig 5.4A, B).

R^{180}, K^{183} and R^{186} on the loop between β-strand IX and β-strand X together with R^{163} and R^{165} of β-strand VIII form a cluster of basic amino acids (Fig 5.4A, B). However, in this area, the two model structures differ substantially in the loop between β-strands IX and X, which includes R^{180}, R^{186} and K^{183} (Fig 5.4A, B). The orientations of those three residues differ, and of these, R^{180} is the most different (Fig 5.4A, B). This loop in FGF4 and FGF6 is relatively short (only 2 amino acids) (Fig 5.2F, Fig 3E) compared to that of FGF5 (10 residues) (Fig 5.4D) and FGF2 (5 amino acids) or FGF1 (7 amino acids). In the model built on the FGF4 template the side chains of R^{180}, R^{186} and K^{183} are closer to each other than in the model built on the FGF2-template (Fig 5.4C). Indeed, HBS-2 of FGF2 covers areas of the loop of β-strand IX-X with R^{118} (Section 4) with K^{119} being labelled [120], which is consistent with the labelling of R^{180}, R^{186} and K^{183} in FGF5. Thus, for this part of the FGF5 model the FGF2 template is likely to be a better approximation. Despite the differences mentioned above, this cluster of basic amino acids is likely part of HBS-1. In the model built from the FGF4 template, there is no acidic border between this cluster and the area of HBS-1 mentioned above (Fig 5.4B), so they form a single, contiguous binding site. Engaging these residues would require the sugar

chain to bend around the protein, which would be helped by the fact that, as in FGF4, the side chains of many labelled lysine and arginine residues protrude from protein surface (Fig 5.4B). In the model built from the FGF2 template, there is an acidic pocket in between the area of HBS-1 with K^{146}, K^{147}, and R^{180}, R^{186} and K^{183} of FGF5. Since this pocket is formed by E^{182} and E^{187}, it likely blocks the path of the sugar chain. However, between R^{163}, R^{165} and R^{195}, R^{200} is an electropositive surface (Fig 5.4A), indicating no barrier for sugar chain. Indeed, in this model, R^{180} protrudes dramatically on the surface (Fig 5.4A), supporting the notion that a sugar chain could be able to reach the side chain of this arginine. Thus both models are consistent with these residues being an extension of HBS1.

R^{218} neighbours to R^{86}, R^{87} in β-strand I, and together, they form another cluster of basic residues (Fig 5.4A, B). Moreover, the surface between this cluster and the cluster of R^{163}, R^{165}, R^{180}, K^{183} and R^{186} is positively charged, so presenting no barrier to a heparin chain. The two clusters are parallel to each other, separated by 2.25 nm (for FGF4-template model) and 2.8 nm (for FGF2-template model) and one polysaccharide chain would be unlikely to engage both of them at the same time, due to the limitations of torsion angles around the glycosidic bond [208] [209]. The space between R^{218} and the edge of HBS-1(R^{195}) is electropositive (Fig 5.4A, B), however, located there are two protected residues (R^{205} and K^{207}), indicating that the polysaccharide does not engage to this area. Thus, it could be proposed that R^{218}, R^{86}, R^{87} are independent of HBS-1, and their position in sequence would mean they form an HBS-3 [120]. The neighbour of R^{218} on the surface, K^{220} would then also be part of HBS-3.

K^{24}, R^{25}, and K^{29} are N-terminal to β-strand I and cannot be modelled on the template structure of FGF4 [207], which lacks the corresponding residues. Given the position of these residues in sequence, they may be part of HBS-3. In the β trefoil structure of FGFs, the C- and N-termini are vicinal. Consequently, R^{253}, R^{262}, K^{254}, K^{260} and R^{266} in the C-terminus are assigned to

HBS-3. Whether the HBS3 of FGF5 is indeed large, as suggested by the above, or the folding of the N- and C-termini enable a single chain of the polysaccharide to bind to the residues assigned above to HBS1 and/or HBS2 remains to be determined. This could be done through measuring the interaction of FGF5 with an HS brush to determine if it is able to cross link HS chains [126] [203].

FGF5 has a long N-terminal that extends well beyond β-strand I of the β-trefoil structure and possesses the longest C-terminal tail of the FGF family. The function of C-terminus is not understood, but the present data suggest that it interacts with heparin and it is at least involved in binding HS.

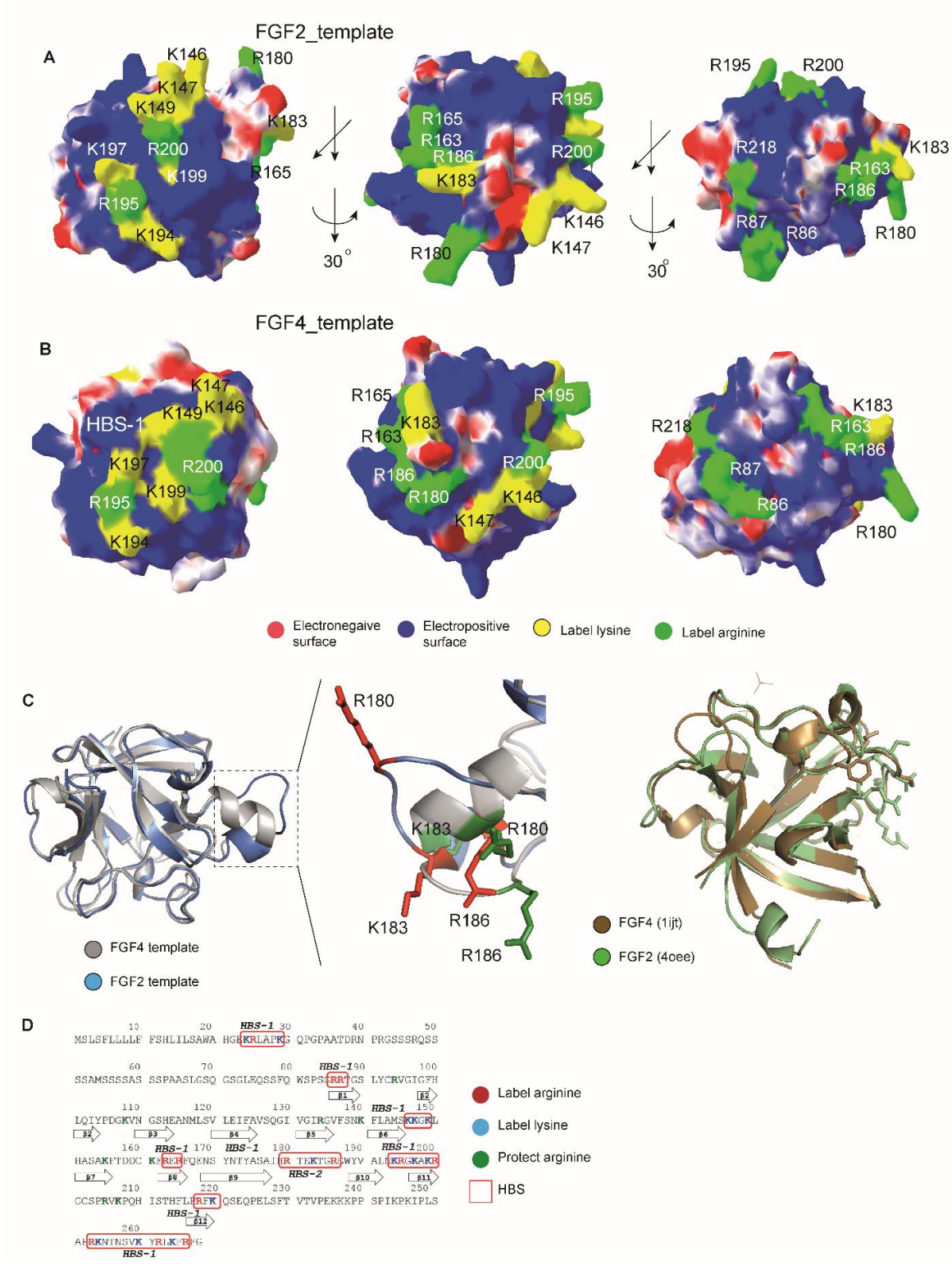

FGF2_template
A
K146
R180
K147
K149
K183
K197
R200
K199
R165
R195
K194
R165
R163
R186
R195
R200
K183
K146
K147
R180
30°
R195
R200
R218
K183
R163
R186
R87
R86
R180
30°
FGF4_template
B
HBS-1
K147
K149
K146
K197
R200
K199
R195
K194
R165
K183
R163
R186
R180
R195
R200
K146
K147
K183
R218
R163
R186
R87
R86
R180
Electronegaive surface
Electropositive surface
Label lysine
Label arginine
C
R180
K183
R180
K183
R186
R186
FGF4 template
FGF2 template
FGF4 (1ijt)
FGF2 (4oee)
D
HBS-1
HBS-2
Label arginine
Label lysine
Protect arginine
HBS

Figure 5.4: Heparin binding arginine and lysine residues in FGF5.

The surface of FGF5 is generated on the templates of FGF2 (PDB code: **4oee** [210]) and FGF4 crystal structure (PDB code: **1ijt** [207]) The electrostatic potential of FGF5 was computed using the Poisson-Boltzmann algorithm in Swiss-PDBView with the positively charged areas coloured *blue* and the negatively charged areas coloured *red*. Labelled lysine residues (Table 5.4C) are presented as *yellow*. Labelled arginine residues (Tables 4A, B) are coloured *green*. **A)** The locations of heparin binding arginine and lysine residues in FGF5 on the surface generated from FGF2 template. The positions of R^{195} and R^{200} (peptides 9 and 10, Table 5.4A) (peptide 8, Table 5.4B) of HBS-1 are presented on the surface**.** The lysine residues of HBS-1 are K^{146}, K^{147}, K^{149} (peptide 1, Table 5.4C), K^{183} (peptide 6, Table 5.4C), K^{194}, K^{197} and K^{199} (peptide 9, Table 5.4C). On the surface, residues of HBS-2 R^{163} and R^{165} (peptides 3 and 4, Table 5.4A) (peptide 2, Table 5.4B) and R^{180}, R^{183}, R^{186} (peptide 5, Table 5.4A) (peptide 7, Table 5.4B) are shown. The labelled residues on the N-terminus R^{86}, R^{87} (peptides 2, 7 and 8, Table 5.4A) (peptide 5, Table 5.4B) and on C-terminus R^{218} (peptide 1, Table 5.4B) and K^{220} (peptide 4, Table 5.4C) are presented. **B)** The locations of heparin binding arginine and lysine residues in FGF5 on the surface generated from FGF4 template. The orientations of the surface were maintained the same as for FGF2 template. **C)** The superimposition of two FGF5 models generated from FGF2 and FGF4 templates. Residues of HBS-2 including R^{180}, R^{183}, R^{186} are highlighted for comparison. **D) Sequence, secondary structure and location of labelled arginine and lysine residues.** Uniprot ID of FGF5 sequence is P12034-1.

5.2.2 FGF9 subfamily (FGF9/16/20)

The lysine residues contributing to heparin binding in FGF9 and FGF20 have previously been determined [37] and suggest that this family may possess a single extended HBS1, with FGF20 potentially having a single, isolated lysine forming a secondary HBS.

5.2.2.1 FGF9

Reacting with HPG and thus bound to heparin in FGF9 were: R^{173}, R^{177} and R^{180} (peptide 11, Table 5.5A) (peptide 7, Table 5.5B) on the loop of β-strands X to XII; R^{160} and R^{161} on the loop between β-strands X-XI (peptide 4 and 8, Table 5.5A) (peptide 5, Table 5.5B); R^{137} on β-strand VIII (peptide 3, 7, Table 5.5A) (peptide 4, Table 5.5B); R^{108} on β-strand V (peptide 3, Table 5.5B); R^{91} on the loop of beta-strands III-IV and R^{86} on β-strand III (peptide 10, Table 5.5A) (peptide 6, Table 5.5B); R^{62}, R^{63} and R^{64} on β-strand I (peptide 2 and 5, Table 5.5A) (peptide 2, Table 5.5B).

Organisation of HBS

The area covering the loop between β-strands X and XI, β-strand XI and the loop from β-strands XI to XII is the core of the canonical HBS-1. Together with the previous data on labelled lysine residues and sequence alignment of all FGFs, it could be concluded that R^{173}, $R^{177,}$ and R^{180}, along with K^{154}, K^{168} and K^{179} belong to HBS1 [36]. On the surface, R^{160}, $R^{161,}$ and R^{137} are adjacent to lysines previously assigned to HBS-1 (K^{154} and K^{179}) [36] and R^{180} mentioned above (Figure 5A), indicating that they are in the canonical HBS1 as well. These results are supported by the co-crystal structure of FGF9 with the mesenchymal "c" splice isoform of FGFR1 [211]. In this structure, R^{173} and R^{177} were bound to a sulfate group, and R^{180}, R^{160}, R^{161}, $R^{137,}$ and K^{154} interact with other two sulfate groups (Fig 5.5D). On the other side of surface are K^{87} of HBS-1 which was biotinylated in the previous work [36] and its

neighbours R^{86} and R^{91} which were labelled by HPG, and all of them are also close to the canonical binding site. Thus, these residues are most likely an extension of HBS-1 (Fig 5.5B).

On β-strand I are located three arginine residues, R^{62}, R^{63} and R^{64}, which all reacted with HPG (peptide 2 and 5, Table 5.5A) (peptide 2, Table 5.5B). These arginines occupy a position analogous to HBS3 in other FGFs, for example, FGF2 [120]. However, they are not separated from HBS-1 by an acidic border (Fig 5.5C), but an intervening electropositive surface. Indeed these residues, with those assigned to HBS-1, are part of a continuous basic surface on FGF9. On this basic surface, there is also the labelled R^{108}, indicating that this arginine is part of HBS-1. This surface requires the polysaccharide to bend considerably since it is not flat. Heparin, due to its many *N*-sulfated glucosamines and so S-domains of HS, has been considered to be relatively rigid [92]. However, a co-crystal of FGF1 and a heparin decasaccharide (1AXM) [97] indicated that this is not always the case. Moreover, a single N-acetylated glucosamine will enable the sugar chain to bend considerably. Thus, a single very extended HBS-1 on FGF9 could indeed bind a single chain of HS. This is supported by the demonstration that in a HS brush, FGF9 cannot crosslink HS chains, unlike FGF2, which has three independent HBSs [202].

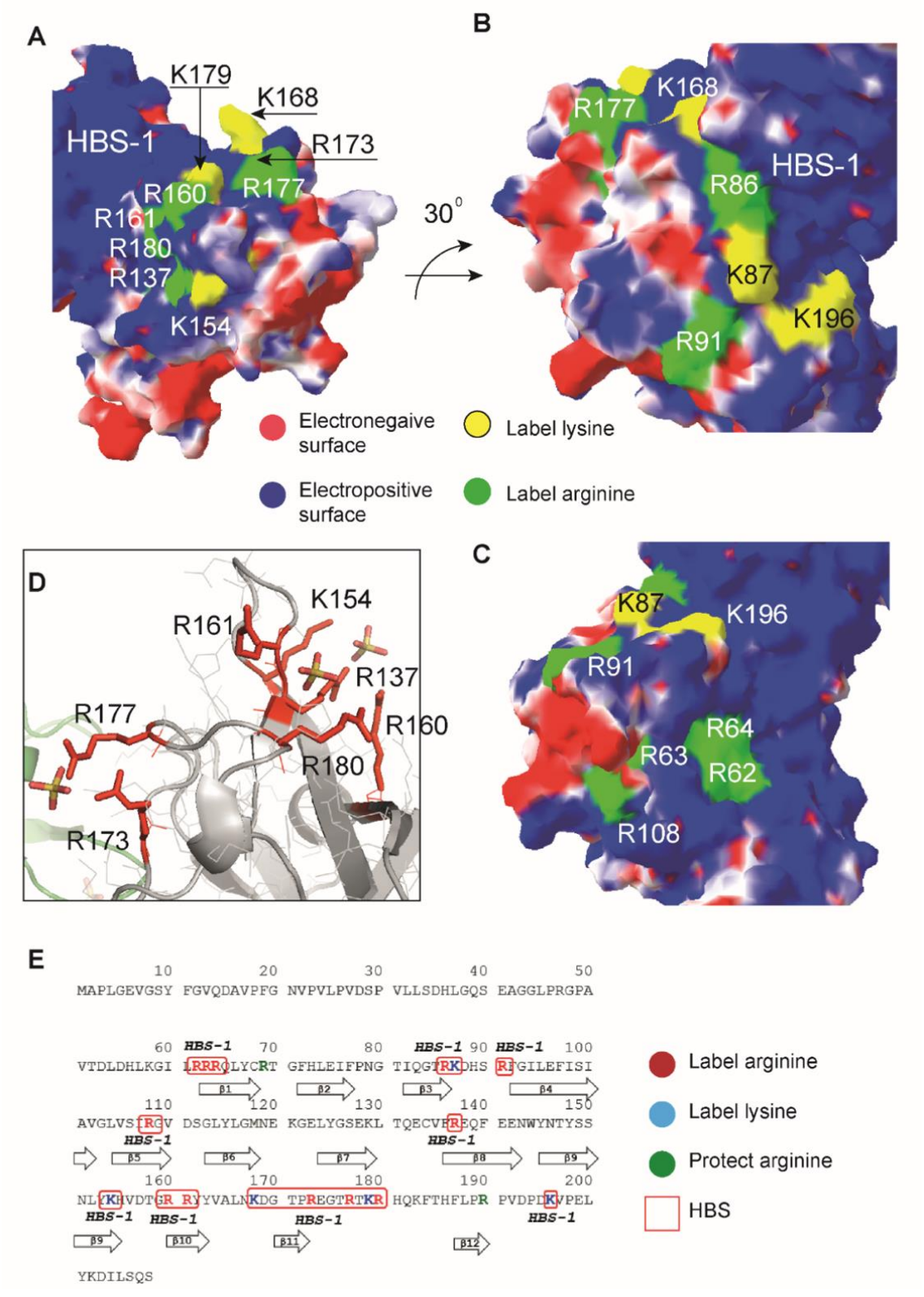

A
K179
K168
HBS-1
R173
R160
R177
R161
R180
R137
K154
30°
B
R177
K168
HBS-1
R86
K87
K196
R91
Electronegaive surface
Label lysine
Electropositive surface
Label arginine
D
K154
R161
R137
R177
R160
R180
R173
C
K87
K196
R91
R64
R63
R62
R108
E
10 20 30 40 50
MAPLGEVGSY FGVQDAVPFG NVPVLPVDSP VLLSDHLGQS EAGGLPRGPA
60 HBS-1 70 80 HBS-1 90 HBS-1 100
VTDLDHLKGI LRRRQLYCRT GFHLEIFPNG TIQGTRKDHS RTGILEFISI
β1 β2 β3 β4
110 120 130 140 150
AVGLVSIRGV DSGLYLGMNE KGELYGSEKL TQECVFREQF EENWYNTYSS
HBS-1 HBS-1
β5 β6 β7 β8 β9
160 170 180 190 200
NLYKHVDTGR RYYVALNKDG TPREGTRTKR HQKFTHFLPR PVDPDKVPEL
HBS-1 HBS-1 HBS-1 HBS-1
β9 β10 β11 β12
YKDILSQS
Label arginine
Label lysine
Protect arginine
HBS

Figure 5.5: Heparin binding arginine and lysine residues in FGF9.

The FGF9 structure (PBD code **5w59** [212]) is shown as a surface. The electrostatic potential of FGF9 was computed using the Poisson-Boltzmann algorithm in Swiss-PDBView with the positively charged areas coloured *blue* and the negatively charged areas coloured *red*. Labelled lysine residues from [36] are presented as *yellow*. Labelled arginine residues (Tables 5A, B) are coloured *green*. **A)** The positions of R^{137} (peptide 3, 7, Table 5.5A) (peptide 4, Table 5.5B), R^{160}, R^{161} (peptide 4 and 8, Table 5.5A) (peptide 5, Table 5.5B), R^{173}, R^{177} and R^{180} (peptide 11, Table 5.5A) (peptide 7, Table 5.5B) on the surface relative to K^{154}, K^{168} and K^{179} [36] **B)** The positions of R^{86} and R^{91} (peptide 10, Table 5.5A) (peptide 6, Table 5.5B) relative to K^{87} [36] and K^{196} [36] **C)** The positions of R^{62}, R^{63} and R^{64} (peptide 2 and 5, Table 5.5A) (peptide 2, Table 5.5B) as well as R^{108} (peptide 3, Table 5.5B) on the surface. **D)** The interactions of R^{161}, R^{177} and R^{180} with the sulfate groups in the crystal structure **E) Sequence, secondary structure, location of labelled arginine and lysine residues in heparin-binding sites.** Uniprot ID of FGF9 sequence is: P31371.

5.2.2.2 FGF20

In FGF20, HPG labelled arginine residues were found as follows: R^{176}, R^{180} and R^{183} on the core of the canonical HBS-1 (peptide 8 and 9, Table 5.6A) (peptide 9 and 10, Table 5.6B); R^{163} and R^{164} on β-strand X (peptide 4 and 7, Table 5.6A) (peptide 4, 7 and 8, Table 5.6B); R^{140} on β-strand VIII (peptide 3, Table 5.6A); R^{111} on β-strand V (peptide 6, Table 5.6A) (peptide 3, Table 5.6B); R^{89} in β-strand III (peptide 11, Table 5.6B); R^{72} on the loop between β-strands I-II (peptide 2, Table 5.6A) (peptide 2 and 11, Table 5.6B); in the disordered C-terminus, R^{199} (peptide 12, Table 5.6B); and $^{65}RRR^{67}$ in β-strand I (peptide 1, Table 5.6A and B).

Organisation of HBS

In the core of the canonical HBS-1, which covers the loop from β-strand X to XI, through β-strand XI to the loop of β-strands XI–XII, all basic residues engage to heparin, K^{171}, K^{182} [37] and R^{176}, R^{180}, and R^{183}. R^{163} and R^{164} of β-strand X are adjacent to K^{182} and R^{183}, hence they are part of HBS-1 (Fig5.6A). Next to R^{163} is R^{140} on β-strand VIII, followed by K^{157} of the β-strand IX [37], so both will contribute to the canonical binding site (Fig 5.6A). In the other direction on the surface, close to K^{171} are R^{72} in the loop between of β-strands I-II and R^{89} in β-strand III, implying that they are also likely an extension of HBS-1 (Fig 5.6B). At the C-terminus, because of R^{199} is close to R^{89}, it too is in HBS-1 (Fig 5.6A).

One feature of FGF20 highlighted previously is K^{124}, which was previously identified as a possible independent heparin binding site [37], because it was surrounded by an acidic surface. However, this seems unlikely. K^{124} is adjacent to R^{180} (Fig 5.6F), which is identified here as heparin binding, and is bound to a sulfate ion in an FGF20 crystal structure (PBD code 3f1r [42]), supporting its interaction with the polysaccharide. K^{124} and R^{180} both protrude from the protein surface, whereas the intervening acidic patch is recessed (Fig 5.6D). Thus, it seems most likely that K^{124} is an extension of the canonical HBS1 of FGF20.

The biotinylated K^{132} on the loop of β-strands VII–VIII is quite close to HPG labelled R^{111} on the β-strand V and also $^{65}RRR^{67}$ in β-strand I (Figure 5.6C), hence they may form an HBS-3. Between these residues and the HBS-1 are three acidic amino acids E^{141}, E^{144} and E^{145} forming an acidic border (Fig 5.6C). The sequence alignment showed that all but one labelled arginine of FGF20 are in the equivalent positions to those identified in FGF9 (Figure 5E and Figure 6D), it could be expected that the assignment of these residues into HBSs is identical to that of FGF9. However, in FGF9 the heparin-binding R^{91} on the loop of beta-strands III-IV provides a means for the sugar chain to bridge the acidic residues, whereas in FGF20 this arginine is

replaced by a leucine. Consequently, this single change in the sequence may result in FGF20, unlike FGF9, possessing a truly independent HBS3.

One issue relating to the involvement of R^{65}, R^{66} and R^{67} and heparin binding is that in the crystal structure of FGF20 (PBD code 3f1r [42]), the proposed dimer interface includes R^{65}, R^{66} and R^{67} (Fig 5.6E). It is known that FGF20 exits as a homodimer in solution [37], however, the side chains of these arginines cannot be buried in a dimer interface, since here they were found to bind react with HPG (peptide 1, Tables 5.6A and B). This observation suggests that FGF20 forms a monomer upon heparin binding or that the interface of the dimer in solution [37] differs from that in the crystal structure. The alternative, that the dimer dissociates at the high concentrations of electrolytes used to release the FGF20 from the heparin column seems unlikely, since these arginine residue in the crystal structure are involved in H-bonds with the backbone (Val 195) and π- stacking interactions with aromatic residues Trp147 and Tyr148.

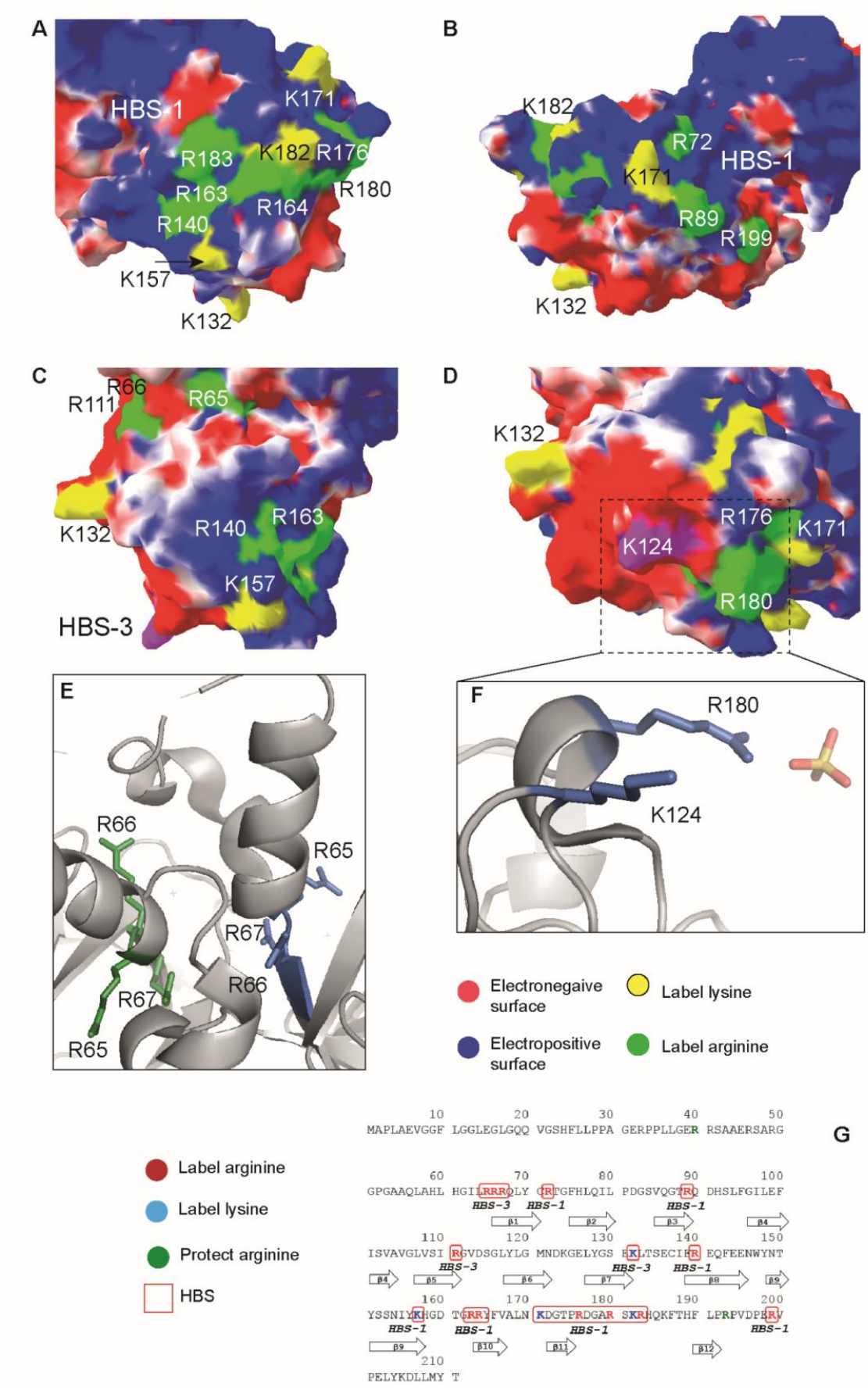

A
HBS-1
K171
R183
K182
R176
R180
R163
R164
R140
K157
K132
B
K182
R72
HBS-1
K171
R89
R199
K132
C
R66
R111
R65
K132
R140
R163
K157
HBS-3
D
K132
R176
K171
K124
R180
E
R66
R65
R67
R66
R67
R65
F
R180
K124
Electronegaive surface
Label lysine
Electropositive surface
Label arginine
Label arginine
Label lysine
Protect arginine
HBS
G
10 20 30 40 50
MAPLAEVGGF LGGLEGLGQQ VGSHFLLPPA GERPPLLGER RSAAERSARG
60 70 80 90 100
GPGAAQLAHL HGILRRRQLY QRFGFHLQIL PDGSVQGTRQ DHSLFGILEF
HBS-3 HBS-1 HBS-1
β1 β2 β3 β4
110 120 130 140 150
ISVAVGLVSI RGVDSGLYLG MNDKGELYGS EKLTSECIFR EQFEENWYNT
HBS-3 HBS-3 HBS-1
β4 β5 β6 β7 β8 β9
160 170 180 190 200
YSSNIYKHGD TGRRYFVALN KDGTPRDGAR SKRHQKFTHF LPRPVDPERV
HBS-1 HBS-1 HBS-1 HBS-1
β9 β10 β11 β12
210
PELYKDLLMY T

Figure 5.6: Heparin binding arginine and lysine residues in FGF20.

The FGF20 structure (PBD code 3f1r [42]) is shown as a surface. The electrostatic potential of FGF20 was computed using the Poisson-Boltzmann algorithm in Swiss-PDBView with the positively charged areas coloured *blue* and the negatively charged areas coloured *red*. Labelled lysine residues from [102] are presented as *yellow*. Labelled arginine residues (Tables 6A, B) are coloured *green*. **A)** The positions of basic residues R^{163}, R^{164} (peptide 4 and 7, Table 5.6A) (peptide 4, 7 and 8, Table 5.6B), R^{176}, R^{180} and R^{183} (peptide 8 and 9, Table 5.6A) (peptide 9 and 10, Table 5.6B) relative to K^{171}, K^{182} defined in [102] on the surface. **B)** The positions of basic residues R^{89} (peptide 11, Table 5.6B), R^{72} (peptide 2, Table 5.6A) (peptide 2 and 11, Table 5.6B) and R^{199} (peptide 12, Table 5.6B) relative to K^{171} [102] on the surface. **C)** The positions of basic residues R^{140} (peptide 3, Table 5.6A) relative to K^{157} [102]. Moreover, the positions of R^{111} (peptide 6, Table 5.6A) (peptide 3, Table 5.6B), $^{65}RRR^{67}$ (peptide 1, Table 5.6A and B) and K^{132} [102] on the surface **D)** The position of K^{124} [102] and its neighbour residues R^{176}, R^{180} as well as the surface between them. **E)** The dimer interface of FGF20 with the positions of $^{65}RRR^{67}$. **F)** The distance between K^{124} [102] and R^{180}. **G) Sequence, secondary structure, location of labelled arginine and lysine residues in heparin-binding sites.** Uniprot ID of FGF20 sequence is: Q9NP95.

5.2.2.3 FGF16

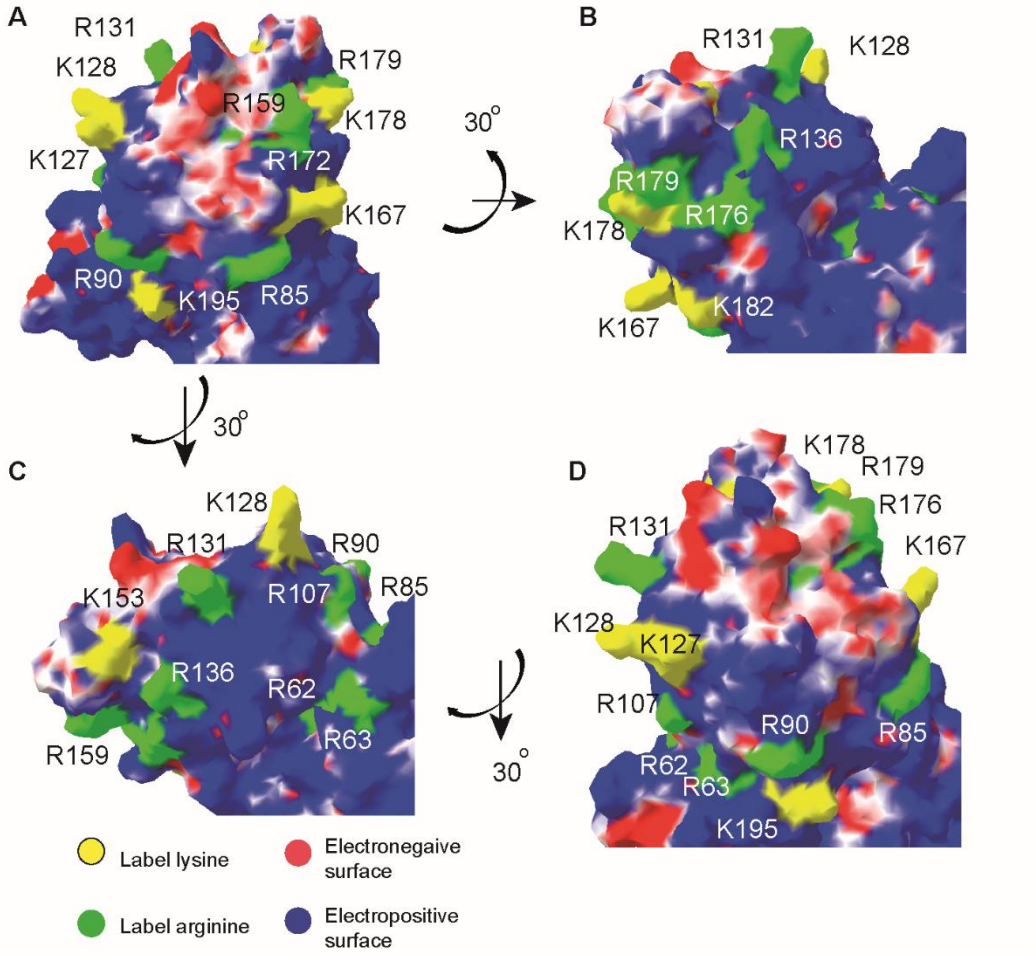
Selective labelling of lysine had not been previously established for FGF16. In this study, five lysine residues were found to be biotinylated, indicating their interaction with heparin, and all of them are assigned to the canonical HBS-1. They are K^{153} on β-strand IX (peptide 2, Table 5.7C), K^{167} on the loop between β-strands X-XI (peptide 1, Table 5.7C), K^{178} and K^{182} on the loop between β-strands XI-XII (peptide 4, Table 5.7C) and K^{195} at the C-terminus (peptides 4, Table 5.7C).

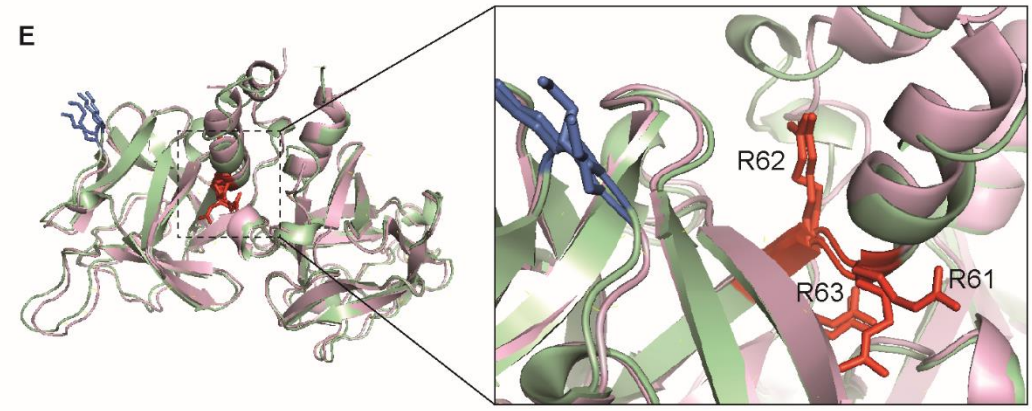
Twelve arginine residues were labelled, comprising R^{107} in β-strand V (peptide 8, Table 5.7A) (peptides 11, 13, Table 5.7B); R^{136} in β-strand VIII (peptide 3, Table 5.7A) (peptide 4, Table 5.7B); R^{159} of β-strand X (peptide 8, Table 5.7B); R^{172}, R^{176} and R^{179} on the loop of β-strands XI-XII (peptides 2, 6 and 7, Table 5.7A) (peptides 3, 6, 9 and 14, Table 5.7B); $^{61}RRR^{63}$ in β-strand VII; R^{85} in β-strand III (peptide 1, Table 5.7B); R^{90} on the loop between β-strands III-IV (peptide 2, Table 5.7B); R^{131} on the loop of β-strands VII-VIII (peptides 3 and 10, Table 5.7A) (peptides 4 and 7, Table 5.7B), (Figure 7B).

Organisation of HBS

Located in the core of the canonical HBS-1 are three lysine residues K^{167}, K^{178} and K^{182} , and three arginines R^{172}, R^{176} and R^{179}, which all were labelled (Fig 5.7A, B). Adjacent to R^{172} are R^{159} on β-strand X (Fig 5.7A) and R^{136}, which is close to R^{176} of HBS-1, indicating that both are part of HBS-1 (Fig 5.7A). R^{131} neighbours to R^{136}, hence this residue is part of HBS-1. The engagement of HBS-1 continues with R^{107}, R^{62}, R^{63}, R^{90} and K^{195} (Fig 5.7C, D), as they are all adjacent to the preceding HBS-1 residues and close to each other. Lastly, R^{85} of β-strand III is separated from R^{90} by a positively charged surface, and on the other side it is near K^{167} (Fig 5.7D), indicating that this arginine is an extension of HBS-1 also. Thus, it appears that similar to FGF9, FGF16 possesses a single, large HBS-1, which would bind a single sugar chain.

Owing to the lack of a crystal structure, the model of FGF16 was built on the templates of both FGF9 and FGF20. Comparison of two models agree on the locations of R^{61}, R^{62}, R^{63} at the dimer interface (Fig 5.7E), but the side chains of R^{61} point at different directions. Even though, R^{61} is not visible from the surface in the dimer structures, the present result demonstrates that this arginine side chain was labelled by HPG, so it cannot be buried. Hence, if FGF16 is a dimer similar to that of FGF20, FGF16 may form a monomer upon heparin binding or that the dimer interface in solution differs from that in the crystal structure.


A
R131
K128
R179
R159
K178
K127
R172
K167
R90
K195
R85
30°
B
R131
K128
R136
R179
R176
K178
K182
K167
30°
C
K128
R131
R90
K153
R107
R85
R136
R62
R63
R159
30°
D
K178
R179
R176
R131
K167
K128
K127
R107
R90
R85
R62
R63
K195
Label lysine
Electronegaive surface
Label arginine
Electropositive surface


E
R62
R63
R61

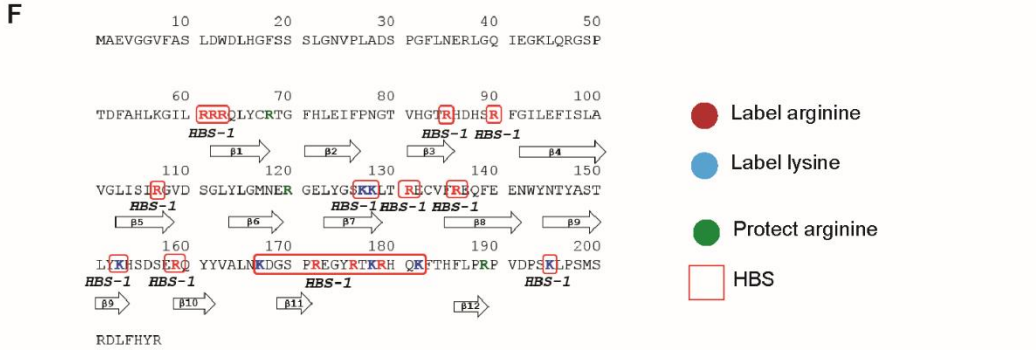

F
MAEVGGVFAS LDWDLHGFSS SLGNVPLADS PGFLNERLGQ IEGKLQRGSP
TDFAHLKGIL RRRQLYCRTG FHLEIFPNGT VHGTRHDHSR FGILEFISLA
VGLISIRGVD SGLYLGMNER GELYGSKKLT RECVFREQFE ENWYNTYAST
LYKHSDSERQ YYVALNKDGS PREGYRTKRH QKFTHFLPRP VDPSKLPSMS
RDLFHYR
HBS-1
Label arginine
Label lysine
Protect arginine
HBS

Figure 5.7: Heparin binding arginine and lysine residues in FGF16.

The FGF16 structure is modelled on the template of FGF9 crystal structure (PBD code **5w59** [212]). The electrostatic potential of FGF16 was computed using the Poisson-Boltzmann algorithm in Swiss-PDBView with the positively charged areas coloured *blue* and the negatively charged areas coloured *red*. Labelled lysine residues (Tables 7C) are presented as *yellow*. Labelled arginine residues (Tables 7A, B) are coloured *green*. **A)** The positions of the basic residues in HBS1 R^{85} (peptide 1, Table 5.7B), R^{90} (peptide 1, Table 5.7B), R^{131} (peptides 3 and 10, Table 5.7A) (peptides 4 and 7, Table 5.7B), R^{172} and R^{179} (peptides 2, 6 and 7, Table 5.7A) (peptides 3, 6, 9 and 14, Table 5.7B) relative to K^{153} (peptide 1, Table 5.7C), K^{167}, K^{178} and K^{182} (peptide 2, Table 5.7C) on the surface **B)** The positions of the basic residues R^{136} (peptide 3, Table 5.7A) (peptide 4, Table 5.7B), R^{176} and R^{179} (peptides 2, 6 and 7, Table 5.7A) (peptides 3, 6, 9 and 14, Table 5.7B) relative to K^{167}, K^{178} on the surface. **C)** The positions of the basic residues $^{61}RRR^{63}$ (peptide 1, Table 5.7B), R^{85} (peptide 1, Table 5.7B), R^{90} (peptide 2, Table 5.7B), R^{131} and R^{136} on the surface. **D)** The surface between $^{61}RRR^{63}$, R^{85}, R^{90}, R^{131} and K^{128} to the core of HBS-1**. E)** The dimer interface of FGF16 with the positions of $^{61}RRR^{63}$. **F) Sequence, secondary structure, location of labelled arginine and lysine residues in heparin-binding sites.** Uniprot ID of FGF16 sequence is: O43320.

5.2.3 FGF8 subfamily (FGF8/17/18)

5.2.3.1 FGF8

FGF8 has not had its interactions with heparin analysed previously, and its HBSs were predicted based on alignment with the two members of this subfamily where heparin-binding lysines had been identified, FGF17 and FGF18 [36][37].

Reacting with NHS-biotin in FGF8 were: K^{71} on β-strand I; six lysine residues K^{112}, K^{113}, K^{115}, K^{119}, K^{123} and K^{125} (peptides 1 and 2, Table 5.8C) on the loop between β-strands VI-VII through

β-strand VII to the loop of β-strands VII–VIII; K^{156}, K^{161}, K^{164} (peptide 6, Table 5.8C) and K^{176} (peptide 5, Table 5.8C) on the loop between β-strands X to β-strand XII. Hence, those residues contribute to the engagement of the protein to heparin.

At the N-terminus R^{37} (peptide 2, Table 5.8A), (peptide 5, Table 5.8B), and R^{48}, R^{49} and R^{52} on the N-terminus and on β-strand I reacted with HPG (peptide 1 and peptide 8, Table 5.8A; peptide 5, Table 5.8B), hence interacted with heparin. On β-strand I, there are three HPG-reacting arginine residues, R^{96}, R^{98}, and R^{100} (peptides 3 and 4, Table 5.8A) (peptide 1, 2 and 7, Table 5.8B). R^{59} on the loop between β-strands I–II (peptide 6, Table 5.8B), R^{72} in β-strand III (peptides 5 and 11, Table 5.8A) (peptides 6 and 8, Table 5.8B) reacted to HPG. Finally, in the core of the HBS-1 are four heparin binding arginine residues R^{155}, R^{158}, R^{160} and R^{170} (peptide 10, Table 5.8A).

Organisation of HBS

The sequence alignment of FGF8 with FGF17 and FGF18 predicted that the members of this subfamily possess the canonical HBS-1 and a secondary HBS-2. In FGF8, the core of the canonical HBS-1 comprises four arginine residues R^{155}, R^{158}, R^{160} and R^{170}, and three lysine residues K^{156}, K^{161} and K^{164} (Figure 8 A, B, E). K^{112}, K^{113} and K^{115} on the loop between β-strands VI and VII are adjacent to K^{161} and K^{164} (Figure 8A, B, E), indicating that they are part of HBS-1. This agrees with the sequence alignment with on FGF17 [36] and FGF18 [102]. Adjacent on the sequence, K^{119}, K^{123} and K^{125} are distant from K^{112}, K^{113} and K^{115} in the structure. However, there is no acidic barrier between these two groups of lysine residues, suggesting that they are an extension of the canonical HBS-1 (Figure 8A, B). Besides, scattered in the sequence, but likely part of HBS-1, are R^{59} on the loop between β-strands I–II, which is adjacent to R^{155} and R^{170} (Figure 8A), and R^{72} of β-strand III, which is close to K^{156} and R^{158} (Figure 8A). Collectively these residues form a single HBS-1. Whereas the participation of R^{59}

in heparin binding was expected from sequence alignment, that of R^{72} was not [36][37]. The participation of residues in β-strand VII and the loop between β-strands VII–VIII in binding heparin is only seen in this subfamily but not in others.

The assignment based on the sequence alignment identified R^{49} in β-strand I, R^{100} on β-strand V, K^{176} in β-strand XII and R^{179} in C-terminus as forming a potential secondary HBS-2 in FGF8. However, R^{179} was not labelled, indicating that it is not involved in heparin binding. K^{176} is a neighbour of R^{170}, located at the edge of HBS-1 (Fig 5.8C). The surface between these two residues is positively charged (Fig 5.8C), indicating the potential of this lysine as an extension of HBS-1. R^{48}, R^{49} and R^{52} of β-strand I were all labelled and on the surface they are adjacent to K^{176} with no acidic border between them (Fig 5.8C). This observation implies they are part of HBS-1 rather than forming an independent HBS-2. On the other side of the surface, on β-strands IV and V are the HPG-reacting R^{96}, R^{98} and R^{100}. On the surface, R^{96} neighbours K^{123} and K^{125} of HBS-1 and R^{98} is adjacent to R^{48} and R^{49} (Fig 5.8D). R^{100} is isolated, but is located on the same electropositive surface as R^{96}, R^{98}, hence is likely part of HBS-1 (Fig 5.8D). Thus, inspection of the surface charge shows no acidic barrier among all labelled residues of FGF8, which would appear to indicate that FGF8 has a single, large HBS1 rather than a secondary HBS-2.

Taking the relative positions of all residues of HBS-1 under consideration indicated that the binding path of the polysaccharide on FGF8 results in an HBS-1 with a dogleg (Fig 5.8F). To clarify, a dogleg is composed of two paths, in which the longer part would consist of R^{155}, R^{158}, R^{160}, R^{170}, K^{156}, K^{161} and K^{164} (Fig 5.8C, D, F) and the shorter part of R^{96}, R^{48}, R^{49}, R^{52}, R^{96}, R^{98} and R^{100} (Fig 5.8C, D, F). Bear in mind that the dogleg is continuous and covers all residues above. The corner of the dogleg is formed by K^{125} and R^{98} (Fig 5.8F). The distance between these two residues is 0.5 nm. The polysaccharide may make a sharp bend by means of an

intervening GlcNAc residue or it may skirt around the bend, with a sulfate or carboxylic acid on the outer edge of the polysaccharide contacting by K^{125} and on the following saccharide an anionic group on the inside edge contacting R^{98}.

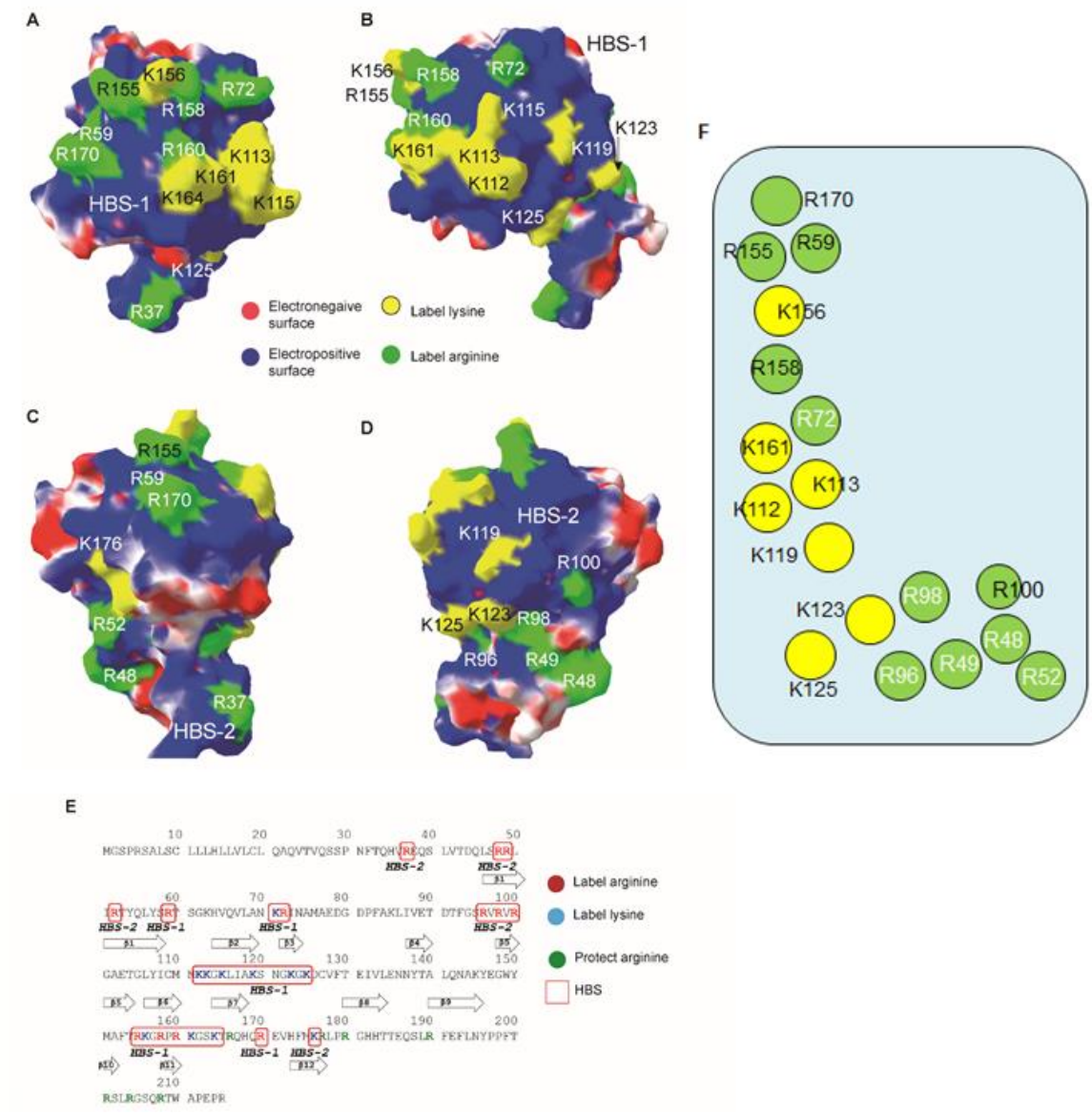


Figure 5.8: Heparin binding arginine and lysine residues in FGF8.

The FGF8 structure (PBD code **2fdb** [213]) is shown as a surface. The electrostatic potential of FGF8 was computed using the Poisson-Boltzmann algorithm in Swiss-PDBView with the positively charged areas coloured *blue* and the negatively charged areas coloured *red*. Labelled

lysine residues (Tables 8C) are presented as *yellow*. Labelled arginine residues (Tables 8A, B) are coloured *green*. **A)** The positions of the basic residues R^{155}, R^{158}, R^{160} and R^{170} (peptide 10, Table 5.8A) along with the labelled K^{156}, K^{161} and K^{164} (peptide 6, Table 5.8C) as well as K^{113}, K^{115} (peptides 1 and 2, Table 5.8C) on the surface. **B)** The positions of K^{112}, K^{113}, K^{115}, K^{119}, K^{123} and K^{125} (peptides 1 and 2, Table 5.8C) together with R^{72} (peptides 5 and 11, Table 5.8A) (peptides 6 and 8, Table 5.8B) on the surface **C)** The positions of R^{37} (peptide 2, Table 5.8A), (peptide 5, Table 5.8B) along with R^{48}, R^{49} and R^{52} (peptide 1 and peptide 8, Table 5.8A; peptide 5, Table 5.8B) on the surface. **D)** The locations of residues assigned to HBS-2 by the sequence alignment R^{100} (peptides 3 and 4, Table 5.8A) (peptide 1, 2 and 7, Table 5.8B), K^{176} (peptide 5, Table 5.8C) in FGF8 surface. **E. Sequence, secondary structure, location of labelled arginine and lysine residues in heparin-binding sites.** Uniprot ID of FGF8 sequence is: P55075-3. **F)** The dogleg binding site formed by all the labelled basic residues of FGF8.

5.2.3.2 FGF17

Arginine residues of FGF17 that reacted with HPG were: R^{49} and R^{50} (peptide 5, Table 5.9A) (peptide 1, Table 5.9B) on the N-terminus and R^{60} (peptide 6, Table 5.9B) on the loop of β-strands I-II; R^{96}, R^{98} in β-strand V (peptide 5, Table 5.9B); R^{113} (peptide 1, Table 5.9A) in the loop between β-strands VI and VII; R^{145} in β-strand IX (peptide 4, Table 5.9A) (peptides 2 and 4, Table 5.9B); R^{155}, R^{158}, R^{160}, R^{164} and R^{166} in the core of the canonical HBS-1 (peptide 6, Table 5.9A) (peptide 8, Table 5.9B); R^{177} on β-strand XII (peptide 7, Table 5.9B). Therefore, these arginines all engage heparin.

Organisation of HBS

The prior study defined two HBSs for FGF17, the canonical HBS-1 and a secondary HBS-2 [37]. First, with the canonical HBS-1, there are no lysine residues in the core of the canonical HBS-1, but only arginine residues, hence, the definition of this area was based on sequence

alignment data. This work confirmed that the core of the canonical HBS1 is in the area from the loop of β-strands X–XI to the loop of β-strands XI–XII, and consists of five arginine residues, R^{155}, R^{158}, R^{160}, R^{164} and R^{166} (peptide 6, Table 5.9A) (Figure 9B). The extended region of HBS-1 defined by previous work comprised five lysine residues, K^{112}, K^{115}, K^{119}, K^{123}, K^{125} [37] and the present work adds one arginine residue, R^{113} (Fig 5.9A, D) to this cluster. R^{145} of the β-strand IX is adjacent to K^{125} (Fig 5.9A) and R^{98} is next to K^{123} (Fig 5.9A), hence these two arginine residues are also part of HBS-1. Adjacent to the core of the canonical HBS-1 is R^{60} (Fig 5.9B) which is a neighbour of R^{155}. This arginine was not included in the binding site from the sequence alignment based assignment, but it is aligned to R^{59} in FGF8, where it was labelled as well.

The HBS-2 of FGF17 was defined as comprising K^{100} on β-strand V and K^{176} on β-strand XII [37]. However, the identification of the arginine residues involved in binding heparin suggests that FGF17 may not have an independent secondary HBS. R^{177}, which reacted with HPG neighbours K^{176}. Moreover, on the surface these two residues are adjacent to R^{155}, R^{158}, R^{60} of HBS-1 (Figure 9C). Thus, K^{176}, R^{177} are likely an extension of HBS-1 rather than forming a distinct HBS. Towards the N-terminus, K^{176} and R^{177} are separated from R^{49} and R^{50} by a basic surface formed by Q^{51} and Y^{55}, hence a single polysaccharide chain can bind both sets of residues (Fig 5.9C). Adjacent to R^{49} and R^{50} are R^{98}, K^{100} of β-strand V (Fig 5.9A). K^{100} is distant from R^{98} and R^{50}, however, there is no acidic border between them (Fig 5.9A), hence it too is likely to contribute to the same binding site, HBS-1.

These observations showed that all labelled residues of FGF17 are on the same positively charged surface with no acidic barrier. Hence, it can be interpreted that FGF17 possesses a single, long, continuous HBS-1. However, similar to FGF8, the relative positions of the heparin engaging residues of FGF17 form a dogleg path for the polysaccharide (Fig 5.9E). The longer

part of the dogleg consists of R^{155}, R^{158}, R^{160}, R^{164}, R^{166}, R^{113}, K^{112}, K^{115} and K^{119} (Fig 5.9E) and the shorter part is formed by K^{100}, K^{123}, R^{49}, R^{50}, R^{98} and R^{96} (Figs 5.9A, E) as for FGF8. However, FGF17 has been found to partially crosslink HS chains [203]. This suggests that the above interpretation is true part of the time, but there is a dynamic equilibrium with a binding mode of FGF17 involving two polysaccharide chains. In this later binding mode, it would seem likely that each polysaccharide chain engages independently each arm of the dogleg.

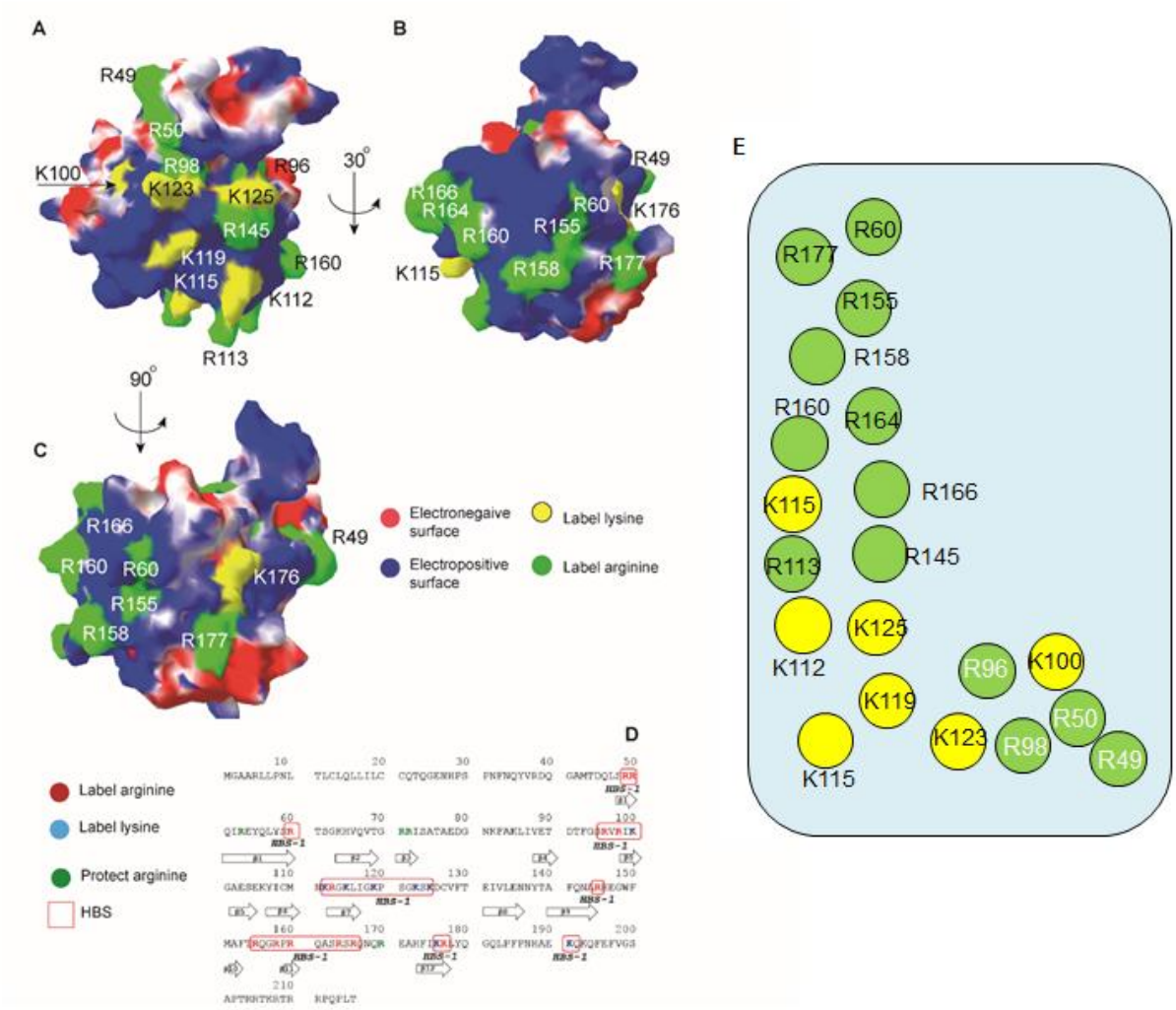


Figure 5.9: Heparin binding arginine and lysine residues in FGF17.

The FGF17 structure is modelled on the template of FGF8 crystal structure (PBD code **2fdb** [213]). The electrostatic potential of FGF17 was computed using the Poisson-Boltzmann algorithm in Swiss-PDBView with the positively charged areas coloured *blue* and the negatively charged areas coloured *red*. Labelled lysine residues [102] are presented as *yellow*. Labelled arginine residues (Tables 9A, B) are coloured *green*. **A)** The position of basic residues R^{49}, R^{50} (peptide 5, Table 5.9A) (peptide 1, Table 5.9B)**,** R^{96} (peptide 5, Table 5.9B), R^{113} (peptide 1, Table 5.9A) and R^{145} (peptide 4, Table 5.9A) (peptides 2 and 4, Table 5.9B) relative to K^{112}, K^{115}, K^{119}, K^{123}, K^{125} [102] on the surface. **B)** The position of basic residues R^{155}, R^{158}, R^{160}, R^{164}, R^{166} (peptide 6, Table 5.9A) (peptide 8, Table 5.9B), R^{177} on β-strand XII (peptide 7, Table 5.9B) and R^{60} (peptide 6, Table 5.9B) together with K^{176} [102] on the surface **C)** The surface between K^{176} and R^{177} and the core of the HBS-1. **D) Sequence, secondary structure, location of labelled arginine and lysine residues in heparin-binding sites.** Uniprot ID of FGF17 sequence is: O60258-1**. E)** The dogleg binding sites formed by all the labelled basic residues of FGF17.

5.2.3.3 FGF18

Arginine selective labelling identified seven residues bound to heparin. They are: R^{60} in β-strand I (peptide 3, Table 5.10A); R^{98} in β-strand V (peptide 6, Table 5.10B); on the loop of β-strands VI to VII, R^{112} (peptide 1 and 2, Table 5.10B); in the core of the canonical HBS-1, R^{158}, R^{160} and R^{166} (peptide 7, Table 5.10A) (peptide 9, Table 5.10B); C-terminal to β-strand XII R^{177} (peptide 6, Table 5.10A), (peptide 3, Table 5.10B). Supporting some of these data is the observation in a crystal structure of FGF18 of R^{158} and R^{160} engaging sulfate anions [214].

Organisation of HBS

The core HBS1 has K^{155}, K^{156}, K^{161} and K^{164} identified as heparin-binding by previous work [36] and here R^{158}, R^{160} and R^{166} (Figs 5.10A, B, D) were also found to bind heparin. Characteristic of the FGF8 sub-family is the involvement of basic residues in the loop of β-strands VI-VII, through β-strand VII to the loop of β-strands VII–VIII in heparin binding. In the case of FGF18, one arginine, R^{112} (Fig 5.10A, B), is added to the four lysine residues K^{113}, K^{115}, K^{119} and K^{125} [36] in this region. They are adjacent to the core of HBS-1 and on the same electropositive surface (Fig 5.10A). Distant in the sequence, but adjacent on the surface to this enlarged HBS1, is R^{98} in β-strand V (Fig 5.10A, B), indicating that this arginine is part of HBS-1 as well. In the other direction, R^{60} in β-strand I neighbours K^{155}, K^{156}, followed by R^{177} and K^{176}, which are C-terminal to β-strand XII (Fig 5.10B). Hence, HBS-1 includes R^{60} and R^{177}.

The previous assignment of heparin binding sites in this subfamily classified K^{82} and K^{102} as an independent HBS, HBS-2 [36]. They are isolated from the basic residues of HBS-1 by acidic amino acids. In detail, between them and the closest residue of HBS1, R^{98}, are E^{103} and E^{105} (Fig 5.10E) and between them and K^{176} and R^{177} of HBS1 is E^{78} (Fig 5.10E). Hence, these two lysines are still likely separated from HBS-1 (Fig 5.10C, D). Consequently, FGF18 likely possesses a fully independent secondary HBS-2. It has been argued previously that the switch of Asp^{121} in FGF18 to Ser^{121} in FGF17 [36] altered the negative barrier on the protein surface, resulting in the subtle difference in binding sites between FGF17 and FGF18. Since in FGF8 the corresponding residue is Asn^{139} in FGF8, it would appear that Asp^{121} is an important distinguishing feature in this FGF subfamily with respect to heparin binding.

A further distinguishing feature of FGF18 is that the residues of HBS1 do not form a dogleg shaped path on the protein surface, unlike HBS1 residues in FGF8 and FGF17. Thus although FGF18, like FGF17 was found to partially cross-link HS [203], the immobilisation parameters

were shown to be different between FGF17 and FGF18 [203]. Briefly, FGF17 and FGF18 both bound slowly HS brushes, but after rinsing while FGF18 was not released from brush, FGF17 exhibited a slow rate of dissociation into the bulk medium. This difference may relate to FGF17 dynamically exchanging between binding modes involving a single HBS and two HBSs, whereas in FGF18 the HBSs are independent.

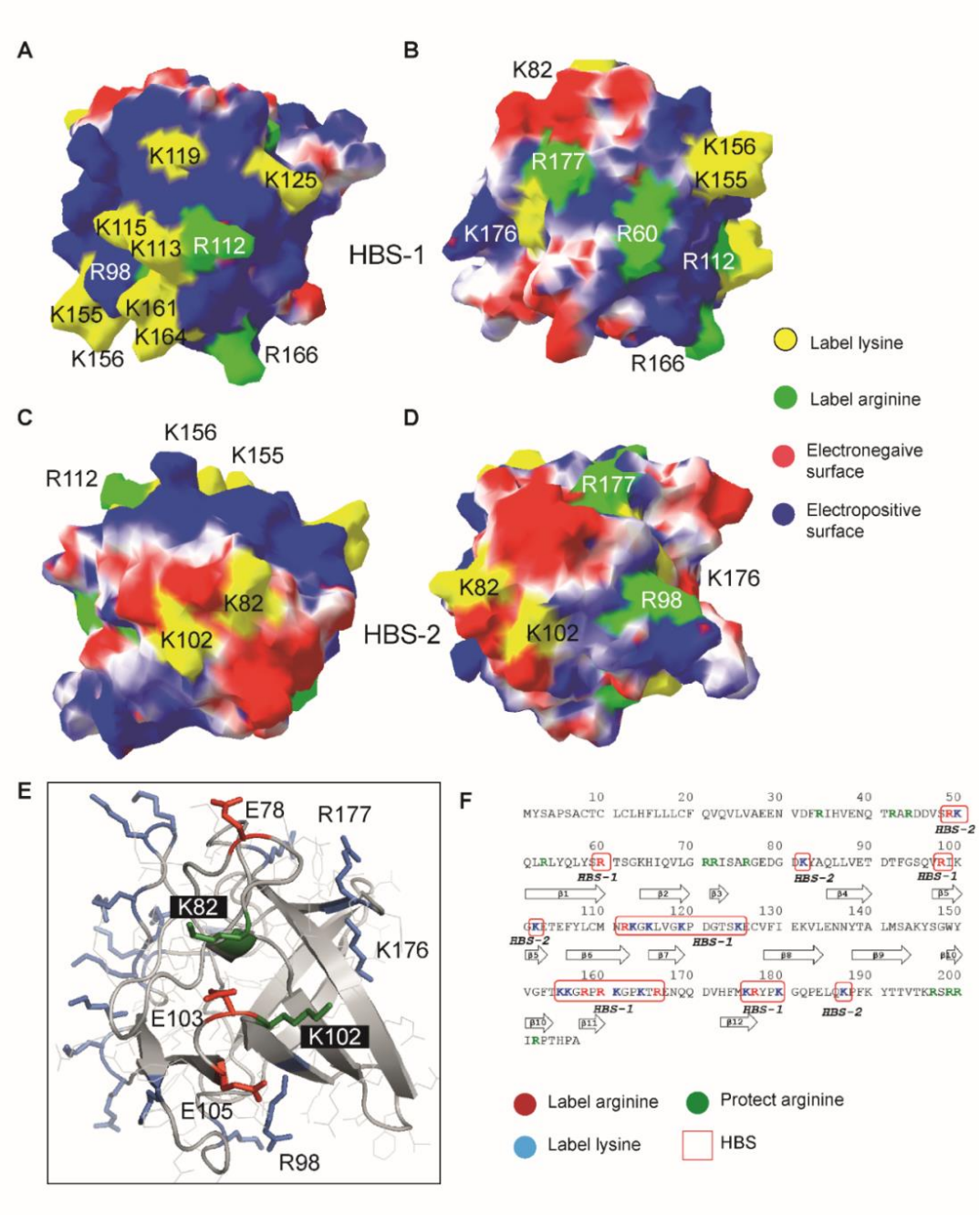


Figure 5.10: Heparin binding arginine and lysine residues in FGF18.

The FGF18 structure (PBD code **4cjm** [214]) is shown as a surface. The electrostatic potential of FGF18 was computed using the Poisson-Boltzmann algorithm in Swiss-PDBView with the positively charged areas coloured *blue* and the negatively charged areas coloured *red*. Labelled lysine residues [36] are presented as *yellow*. Labelled arginine residues (Tables 10A, B) are coloured *green*. **A)** The position of basic residue R^{112} (peptide 1 and 2, Table 5.10B) and R^{166} (peptide 7, Table 5.10A) (peptide 9, Table 5.10B) along with labelled K^{155}, K^{156}, K^{161} and K^{164} [37] on the surface. **B)** The position of basic residues R^{60} (peptide 3, Table 5.10A)**,** R^{158}, R^{160} (peptide 7, Table 5.10A) (peptide 9, Table 5.10B) and R^{177} (peptide 6, Table 5.10A), (peptide 3, Table 5.10B)) along with labelled K^{155}, K^{156} and K^{176} [36] are shown on the surface. **C)** The surface between K^{82} and K^{102} [37] to the residues of HBS-1, including R^{112}, K^{155}, and K^{156}. **D)** The position of basic residues R^{98} (peptide 6, Table 5.10B) and K^{82}, K^{102} and K^{176} **E)** Between K^{82} and K^{102} which were classified as an independent HBS, HBS-2 and the residue R^{98} of HBS-1 are E^{103} and E^{105}. Lie between K^{82}, K^{102} and the residues K^{176}, R^{177} of HBS-1 is E^{78}. **F) Sequence, secondary structure, location of labelled arginine and lysine residues in heparin-binding sites.** Uniprot ID of FGF18 sequence is: O76093.

5.2.4 FGF7 subfamily (FGF3/7/10/22)

5.2.4.1 FGF7

In FGF7 six arginines were labelled with HPG: R^{65}, R^{67} and R^{68} (peptides 5 and 7, Table 5.11A) (peptide 3, Table 5.11B), which are towards the N terminus; R^{72} on the loop of β-strand I-II (peptides 1, 2 and 3, Table 5.11A) (peptides 1 and 2, Table 5.11B); R^{78} and R^{82} on the loop of β-strand II-III (peptide 4, Table 5.11A); R^{101} on β-strand IV (peptide 6, Table 5.11A); and finally, R^{175} on the loop between β-strand XI-XII (peptide 8, Table 5.11A), (peptide 6, Table 5.11B).

Organisation of HBS

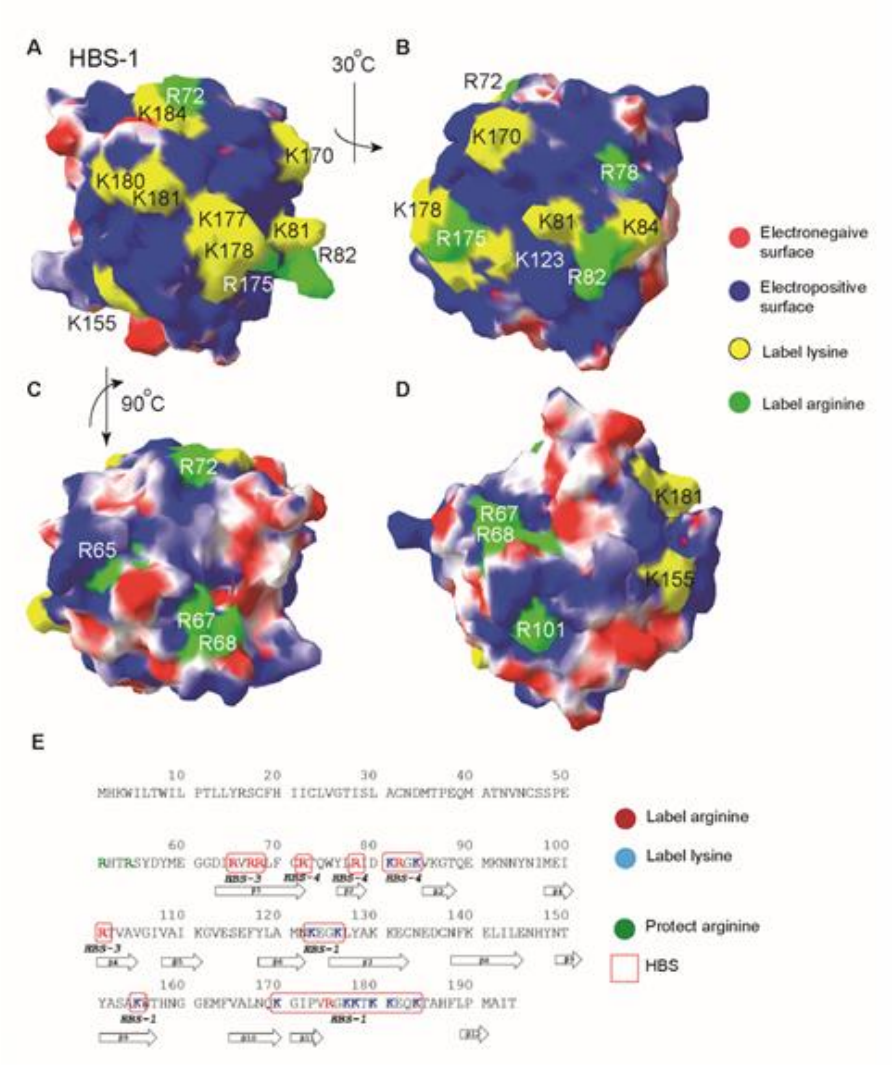
The selective labelling of lysines defined two HBSs in FGF7, the canonical HBS-1 and a secondary HBS-4, which occupies the loop of β-strand II-III and β-strand III. In addition, based on sequence alignment, a HBS3 was also proposed to exist.

First, in the core of HBS1, there are six lysine residues K^{170}, K^{177}, K^{178}, K^{180}, K^{181} and K^{184} [36] and one arginine R^{175}, which was labelled (peptide 8, Table 5.11A and peptide 6, Table 5.11B, Fig 5.11A). Previous work [36] also assigned K^{155} on β-strand IX and K^{123} on the loop of β-strands VI-VII (Figure 11A, B) to HBS-1. K^{123} is adjacent to R^{175} of HBS-1 whereas K^{155} is quite distant from the rest of HBS-1 (Figure 11A). Also some distance from the core of HBS-1 is R^{101}, which also reacted with HPG (Fig 5.11D). However, because there is no acidic barrier between this residue and the closest residues of HBS1, K^{155}, K^{180} and K^{181} (Fig 5.11D), it is likely this in HBS-1.

In previous work [36], it was argued that K^{81} and K^{84} were part of HBS-4. They formed a linear basic structure that was connected to residues in HBS-1, K^{123}, K^{177} and K^{178}. Visually, unlike the dogleg shape HBS-1 observed in FGF8 and FGF17, the HBS-1 and HBS-4 of FGF7 form a "T" shape (Fig 5.11 F), with HBS-4 consisting of K^{81}, R^{82}, K^{84} and R^{78} that meet HBS1 at K^{123}, K^{175}, K^{177} and K^{178} ([36] and Fig 5.11B, F). Consequently, R^{82} is assigned to HBS-4 and R^{78} is an extension of this secondary binding site (Fig 5.11F). The distance from the join of a "T" to K^{155} is 2.8 nm, making the binding pattern of FGF7 different from the dogleg shaped basic patch observed in FGF8 and FGF17 when one sugar chain could not engage K^{155} and turn back to engage to the other edge of the "T". A 'T' shape of FGF7 requires two GAG chains to be able to bind all labelled lysine and arginine residues. There are two possible models for the overall engagement. In the first model (blue line, Fig 5.11F), one sugar chain will engage to residues of the central of HBS-1 and K^{155} and another chain engage to the linear charged

HBS-4 of K^{81}, R^{82}, K^{84} and R^{78}. In the second model (purple line, Fig 5.11F), one sugar chain could bind to residues of HBS-4 and K^{155} and another chain engage remaining HBS-1 residues HBS-1. In at least the first model it seems likely that a requirement is an interaction with a sulfated unit at the non-reducing end of the polysaccharide chain.

Previous work suggested that FGF7 might have an HBS3 towards the N-terminus [36][37], consisting of R^{65}, R^{67} and R^{68}. Here, they were indeed found to be labelled by HPG. On the surface, these residues are isolated from HBS-1 and HBS-4 by acidic residues. Thus, R^{65}, R^{67} and R^{68} form an HBS3, and so FGF7 likely has three HBSs.


A
HBS-1
30°C
R72
K184
K170
K180
K181
K177
K81
K178
R82
R175
K155
B
R72
K170
R78
K178
K81
K84
R175
K123
R82
Electronegaive surface
Electropositive surface
Label lysine
Label arginine
C
90°C
R72
R65
R67
R68
D
K181
R67
R68
K155
R101
E
10 20 30 40 50
MHKWILTWIL PTLLYRSCFH IICLVGTISL ACNDMTPEQM ATNVNCSSPE
60 70 80 90 100
RHTRSYDYME GGDIKVRLFF CKTCWYLRID KNGKVKGTQE MKNNYNIMEI
110 120 130 140 150
RTVAVGIVAI KGVESEFYLA MNKEGKLYAK KECNEDCNFK ELILENHYNT
160 170 180 190
YASAKHTHNG GEMFVALNQK GIPVRGKKTK KEQKTAHFLP MAIT
Label arginine
Label lysine
Protect arginine
HBS

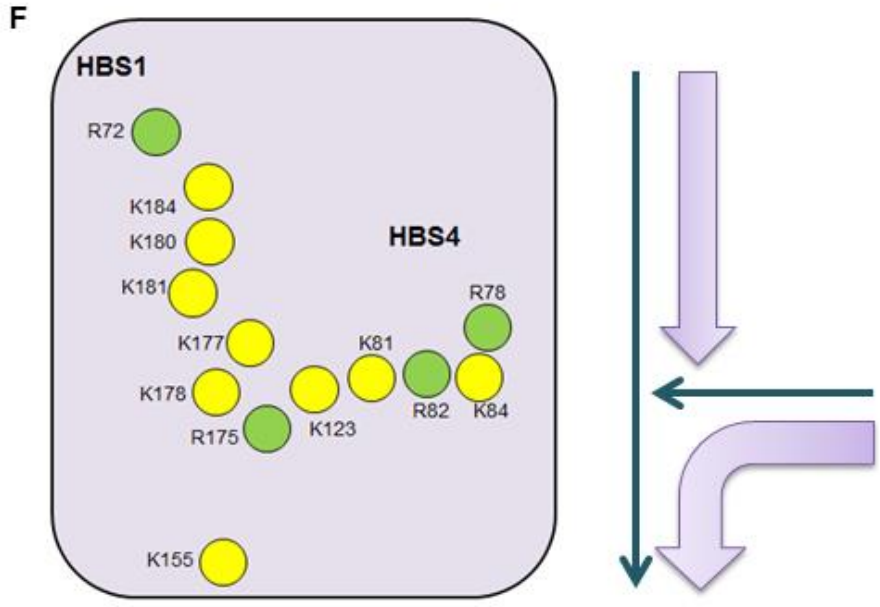

F
HBS1
R72
K184
K180
K181
K177
K178
R175
K123
HBS4
R78
K81
R82
K84
K155

Figure 5.11: Heparin binding arginine and lysine residues in FGF7.

The FGF7 structure (PBD code **1qqk** [31]) is shown as a surface. The electrostatic potential of FGF7 was computed using the Poisson-Boltzmann algorithm in Swiss-PDBView with the positively charged areas coloured *blue* and the negatively charged areas coloured *red*. Labelled lysine residues [36] are presented as *yellow*. Labelled arginine residues (Tables 11A, B) are coloured *green*. **A)** The positions of the basic residues R^{82} (peptide 4, Table 5.11A) and R^{72} (peptides 1, 2 and 3, Table 5.11A) (peptides 1 and 2, Table 5.11B) along with previously labelled K^{81}, K^{170}, R^{175}, K^{177}, K^{178}, K^{180}, K^{181} and K^{184} [36] on the surface. **B)** The positions of the basic residues R^{175} (peptide 8, Table 5.11A), (peptide 6, Table 5.11B) and R^{78} (peptide 4, Table 5.11A) together with K^{84} [36] on the surface. **C)** The positions of the basic residues R^{65}, R^{67} and R^{68} (peptides 5 and 7, Table 5.11A) (peptide 3, Table 5.11B) along with R^{72} on the surface. **D)** The location of R^{101} (peptide 6, Table 5.11A) on the surface. The surface between R^{101} and R^{67}, R^{68} is negatively charged. The surface between R^{101} and K^{155} (of HBS-1) is positively charged. **E) Sequence, secondary structure and location of labelled arginine and lysine residues in heparin-binding sites.** Uniprot ID of FGF7 sequence is: P21781. **F) Schematic showing the location of the residues forming the T shaped HBS1/HBS4 on FGF7.** Possible paths of two polysaccharide chains shown in blue and purple.

5.2.4.2 FGF10

There are six arginine residues in FGF10 that reacted with HPG. They are R^{155} in β-strand VIII (peptide 7, Table 5.12B), R^{174} in β-strand IX (peptides 2, 3 and 6, Table 5.12A, and peptides 3 and 5, Table 5.12B), R^{187}, R^{188}, R^{193} and R^{194} (peptide 7, Table 5.12A and peptide 6, Table 5.12B) situated in the core of HBS-1.

Organisation of HBSs

The selective labelling of lysine residues [37] and sequence alignment indicated that FGF10 had three HBSs, HBS-1, HBS-3 which locates on β-strand I, and HBS-4, as found in FGF7.

The core of the HBS-1 comprises three lysine residues identified previously K^{183}, K^{191} and K^{195} [37] and four arginine residues identified here: R^{187}, R^{188}, R^{193} and R^{194}. Lying between K^{191} and R^{194} is the labelled R^{174} in β-strand IX (Figure 12A), so this residue is part of HBS-1. Next to R^{194} and K^{195} is R^{155} of the β-strand VIII, indicating that this HPG-modified residue (peptide 7, Table 5.12B) is an extension of HBS-1 (Figure 11A). K^{136}, K^{137} and K^{139} in the loop between β-strands VI-VII (Figure 11A, B) and K^{87} on loop of β-strand I-II (Figure 11B) were assigned as HBS-1 residues [37].

Similar to FGF7, FGF10 possesses a HBS-4 which together with HBS-1 formed a "T" shape on the surface. The "T" shaped basic patch is composed of two groups of basic residues. HBS4 is composed of K^{136}, K^{137}, K^{139}, K^{183}, K^{87}, K^{94} and K^{97} [36] (Fig 5.12D) and HBS-1 (Fig 5.12B) R^{194}, K^{195}, R^{193}, R^{174}, K^{183}, K^{191} and R^{78}, with R^{187}, R^{188} potentially belonging to either HBS. As for FGF7, the geometric relationship of these two groups of basic residues makes it difficult to envisage their engaging a single heparin chain. Thus, even though these residues that engage heparin in FGF10 are connected on the continuous basic surface, they are most likely to form two individual HBSs. This result supports the observation that FGF-10 can partially cross-link HS chains in an HS brush [203].

Whereas the expectation from sequence alignment was that FGF10, like FGF7, possessed a HBS-3, located on β-strand I, R^{78} and R^{80} were found to react with PGO (peptide 5, Table 5.12A), hence were not involved in heparin binding. Thus, there is no HBS-3 in FGF10.

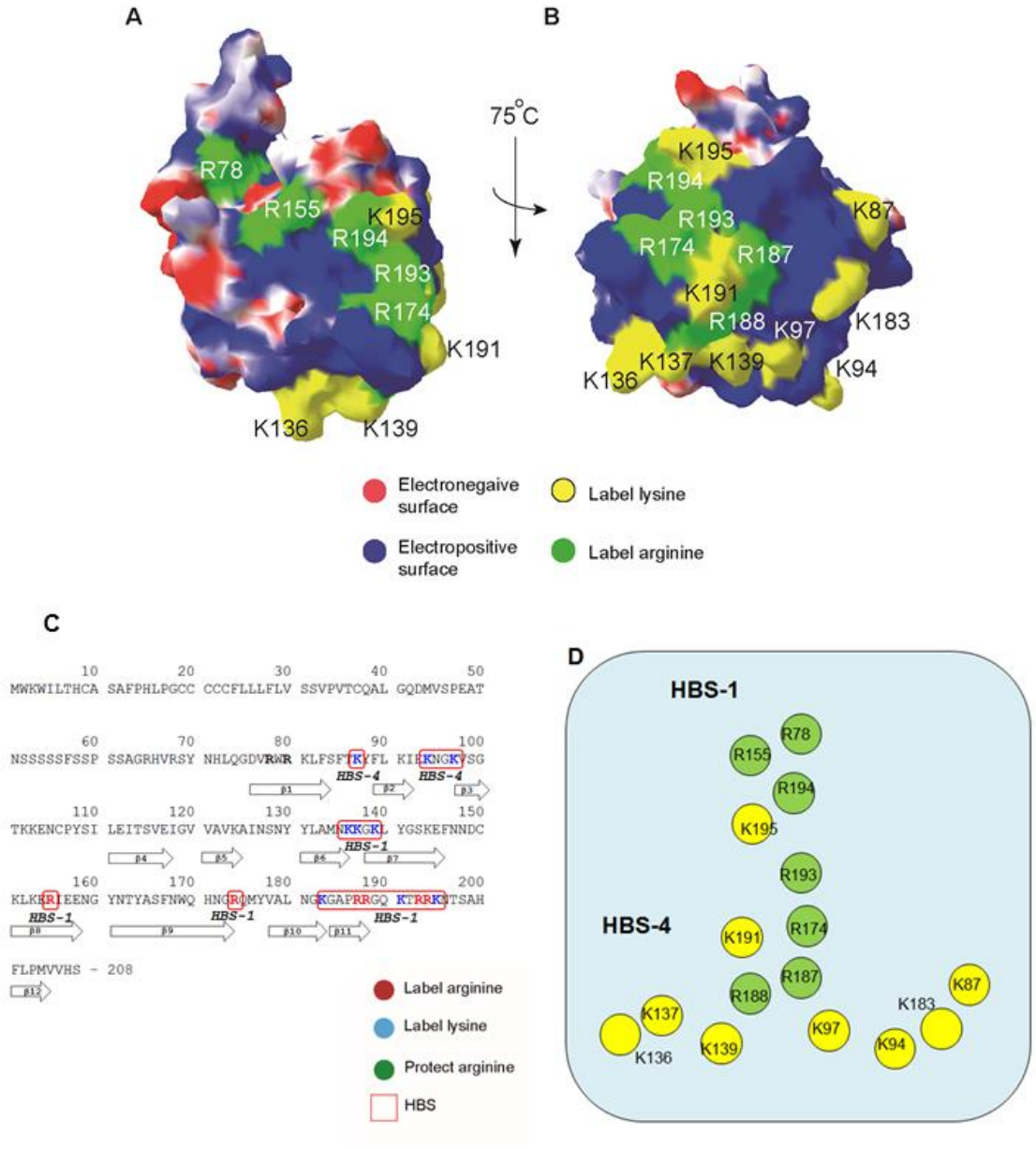


Figure 5.12: Heparin binding arginine and lysine residues in FGF10.

The FGF10 structure (PBD code **1nun** [31]) is shown as a surface. The electrostatic potential of FGF10 was computed using the Poisson-Boltzmann algorithm in Swiss-PDBView with the positively charged areas coloured *blue* and the negatively charged areas coloured *red*. Labelled lysine residues [102] are presented as *yellow*. Labelled arginine residues (Tables 5.12A, B) are coloured *green*. **A)** The positions of basic residues R155 (peptide 7, Table 5.12B), R^{174} (peptides 2, 3 and 6, Table 5.12A, and peptides 3 and 5, Table 5.12B), R^{193}, R^{194} (peptide 7,

Table 5.12A and peptide 6, Table 5.12B) along with K^{94} and K^{97}, K^{136}, K^{139}, $K^{191}$and K^{195} defined previously [102] on the surface. **B)** The positions of basic residues R^{187}, R^{188}, R^{193}, R^{194} (peptide 7, Table 5.12A and peptide 6, Table 5.12B) together with K^{94} and K^{97} [102] on the surface. **C) Sequence, secondary structure, location of labelled arginine and lysine residues in heparin-binding sites.** Uniprot ID of FGF10 sequence is: O15520. **D)** Schematic showing the relative location of the lysine and arginine residues forming the "T" shaped HBS1/HBS4 of FGF10.

5.2.4.3 FGF3

In FGF3, engaged to heparin were: R^{44}, R^{45} and R^{46} (peptide 1, Table 5.13A) (peptide 9, Table 5.13B); R^{63} on the loop between β-strands II-III (peptides 4 and 5, Table 5.13A) (peptide 6, Table 5.13B), R^{89} in β-strand III (peptide 12, Table 5.13B), R^{102} and R^{104} on the loop of β-strands VI-VII (peptides 2, 8 and 9, Table 5.13A) (peptide 1, Table 5.13B); R^{120} in β-strand VIII (peptide 10, Table 5.13B); R^{132} in β-strand IX (peptide 3, Table 5.13A) (peptide 2, Table 5.13B); R^{162}, R^{164}, R^{165}, R^{170} and R^{171} in the core area of HBS-1 (peptides 6 and 7, Table 5.13A) (peptides 3 and 11, Table 5.13B); R^{186} and R^{192}, which are C-terminal to β-strand XIII also reacted with HPG (peptide 4, Table 5.13B).

Organisation of HBS

The results of previous lysine labelling and sequence alignment suggested that FGF3, like FGF7 and FGF10, possessed HBS-1 and HBS-3 which locates on β-strand I. However, the aligned area of HBS-4, between β-strands II-III, in FGF3 contained arginine residues, hence it was not clear whether FGF3 also possessed a HBS4.

The core HBS-1 of FGF3 consists of five arginine residues R^{162}, R^{164}, R^{165}, R^{170} and R^{171} (Figure 13A, B), as well as K^{160} and K^{168} [37]. The addition of those arginine residues to the previous defined HBS-1 with only two lysine residues expands this binding site (Figure 13A,

B). Adjacent to R^{170} and R^{171} on the surface is R^{132} in β-strand IX (Figure 13A), suggesting this residue is part of HBS-1. R^{120} in β-strand VIII neighbours R^{171}, hence this residue is an extension HBS-1. Similar to the aligned residues on FGF7 and FGF10, R^{102} and R^{104} on the loop between β-strand VI-VII in FGF3 (Figure 13B) were labelled, hence, together with K^{101} defined before [37] are an extension of HBS-1. On the surface, those amino acids are next to R^{164} and R^{165} (Figure 13A, B).

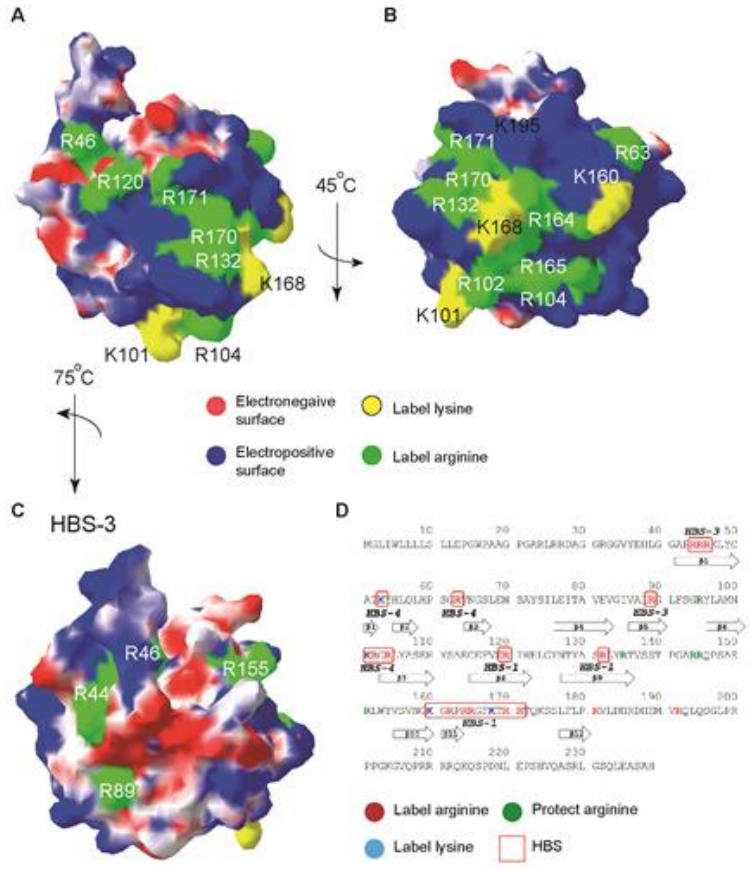
Besides the core of HBS-1, R^{63}, R^{102} and R^{104} were also labelled. From sequence alignment it was expected that R^{63} on the loop between β-strands II-III would be part of HBS-4 [37]. The position of R^{63}, R^{102} and R^{104} on the surface relative to that of the core of HBS-1 indicated that FGF3 was similar other two members, FGF7 and FGF10 in possessing a HBS-4 which joins HBS1 to form a "T" shape (Fig 5.13E). In this "T" arrangement, the basic patch is composed of two binding sites. HBS-1 would comprise R^{120}, R^{132}, R^{164}, R^{165}, K^{168}, R^{170} and R^{171} (Fig 5.12B), and HBS4 would be formed by K^{101}, R^{102}, R^{104}, R^{63} and K^{160}. The residues at the junction, R^{164} and R^{165} (Fig 5.13E) could be assigned to either HBS-1 and HBS-4.

As for FGF7, the geometric relationship of these two groups of basic residues makes it difficult to envisage their engaging a single heparin chain. Thus, even though these residues that engage heparin in FGF3 are connected on the surface, they actually form two individual HBSs.

The prediction of an HBS-3 in FGF3 is borne out by the present results. Thus, N-terminal to β-strand I are three labelled arginine residues, R^{44}, R^{45} and R^{46}. They are isolated from R^{155}, the nearest residue of HBS-1, by an acidic border (Fig 5.13C), hence likely form an independent HBS, HBS-3. In addition, R^{89} on β-strand V is also isolated from HBS1 and HBS4 by acidic residues (Figure 13C), so is likely to contribute to HBS-3.

FGF3 possesses one of the longest C-terminal tails of the FGFs, and it contains nine arginine residues and two lysine residues. However, only R^{186} and R^{192} reacted with HPG (peptide 4,

Table 5.13B), indicating that they bind to heparin. Since this part of the sequence is not covered by a structure, it is not possible to assign these residues to a particular HBS.



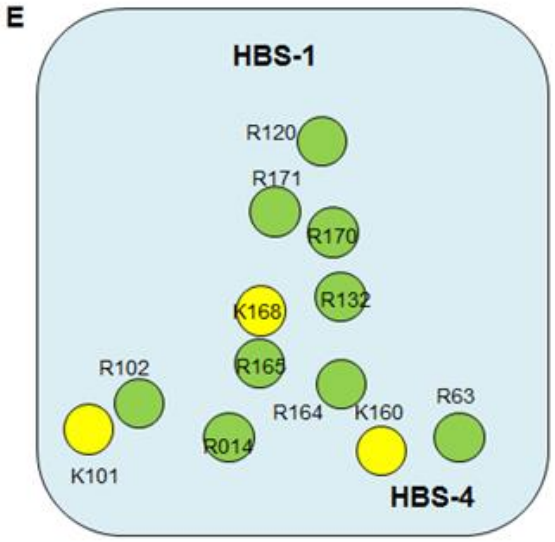


Figure 5.13: Heparin binding arginine and lysine residues in FGF3.

The model of FGF3 structure is built on the template of FGF10 crystal structure (PBD code **1nun** [31]). The electrostatic potential of FGF3 was computed using the Poisson-Boltzmann algorithm in Swiss-PDBView with the positively charged areas coloured *blue* and the negatively charged areas coloured *red*. Labelled lysine residues [102] are presented as *yellow*. Labelled arginine residues (Tables 5.13A, B) are coloured *green*. **A)** The positions of basic residues R^{104}**,** (peptides 2, 8 and 9, Table 5.13A) (peptide 1, Table 5.13B), R^{120} (peptide 10, Table 5.13B), R^{132} (peptide 3, Table 5.13A) (peptide 2, Table 5.13B), R^{170} and R^{171} (peptides 6 and 7, Table 5.13A) (peptides 3 and 11, Table 5.13B) along with K^{101} and K^{168} defined previously [102] on the surface. **B)** The positions of basic residues R^{63} (peptides 4 and 5, Table 5.13A) (peptide 6, Table 5.13B), R^{102}, R^{104} (peptides 2, 8 and 9, Table 5.13A) (peptide 1, Table 5.13B), R^{162}, R^{164}, R^{165}, R^{170} and R^{171} (peptides 6 and 7, Table 5.13A) (peptides 3 and 11, Table 5.13B) along with K^{160} defined previously [102] on the surface. **C)** The positions of basic residues R^{44}, R^{45}, R^{46} (peptide 1, Table 5.13A) (peptide 9, Table 5.13B) and R^{89} (peptide 12, Table 5.13B). On surface between these arginine residues and other residues is negatively charged. **D) Sequence, secondary structure, location of labelled arginine and lysine residues in heparin-binding sites.** Uniprot ID of FGF3 sequence is: P11487-1. **E)** The "T" shape formed by the labelled residues of FGF3.

5.2.4.4 FGF22

FGF22 possesses a single lysine residue in the sequence, which was found to contribute to heparin binding, since it was biotinylated (Table 5.14C), thus there was little information other than sequence alignment to determine its HBSs. The labelled arginine residues were: R^{41}, ^{43}R on β-strand I (peptides 6 and 7, Table 5.14B); R^{54} on β-strand II (peptide 2, Table 5.14A) (peptide 3, Table 5.14B); R^{60} on the loop between β-strands II–III (peptide 2, Table 5.14A)

(peptide 3, Table 5.14B); R^{77} on β-strand V (peptide 8, Table 5.14B); $^{99}RR^{100}$ and R^{102} on the loop between β-strands VI–VII (peptide 3, Table 5.14A), (peptide 2, Table 5.14B); R^{114} (peptide 8, Table 5.14A), R^{116}, R^{118} (peptide 6, Table 5.14A) on β-strand VIII; R^{131}, $^{133}RRR^{135}$ (peptide 6, Table 5.14A) (peptide 5, Table 5.14B) on the loop between β-strands XI–X; R^{145}, R^{146}, R^{150}, R^{154}, R^{156} and R^{157} (peptide 4 and 5, Table 5.14A) (peptide 4, Table 5.14B) in the core of HBS-1.

Organisation of HBS

The core of the canonical HBS-1 comprises six arginine residues R^{145}, R^{146}, R^{150}, R^{154}, R^{156} and R^{157} that reacted with HPG and so were engaged to heparin. However, despite of their closeness in the sequence, R^{145}, R^{146} are distant from R^{150}, R^{154}, R^{156} and R^{157} on the surface (Figure 14B), but there are no acidic amino acids between these two groups, indicting they are part of HBS-1. Alignment predicted that the residues on the loop between β-strands IX–X to be part of HBS-1 [36][37]. In FGF22, this region contains four arginines R^{131}, $^{133}RRR^{135}$ and all of them reacted with HPG. R^{114}, R^{116}, R^{118} on β-strand VIII are adjacent to R^{131}, $^{133}RRR^{135}$, indicating that they are an extension of HBS-1, though there are no aligned basic residues in this area in other members of FGF7 subfamily that engage heparin [36][37].

Like the other three members of the sub-family, FGF22 likely possesses a HBS-1 and a HBS-4. However, the HBS4 in FGF22 is more extensive and four of its residues are parallel to HBS-1 (Fig. 5.14). The HBS-1 in this model of binding is composed of R^{150}, R^{154}, R^{156} and R^{157} as well as R^{145} and R^{146} on the loop between β-strands X–XI (Fig 5.14E). HBS-4 in FGF22 is not linear, but forms a dogleg with the corner at R^{131}. This arrangement may lend itself to a number of different binding modes. In one, HBS-1 would run from R^{157}/R^{145} on the central of HBS-1 to R^{134} on the loop between β-strands IX and X, and HBS-4 from R^{118} on β-strand VIII, through R^{100}, R^{102} to finish at R^{54} on the loop between β-strands II and III. An alternative would be for

HBS-1 to include R^{102}, R^{99} on the loop between β-strands VI and VII, R^{60} and R^{54} and for HBS-4 to terminate at R^{133}/R^{100}. It seems unlikely that a single polysaccharide chain could engage these residues because the distance from R^{116} to R^{146} is 3.2 nm, whereas the distance between charged groups on opposite sides of the sugar chain is ~0.8 nm.

The difference observed here between FGF22 and the other three members of the FGF7 subfamily is largely due to six arginine residues in FGF22, R^{114}, R^{116}, R^{118} on β-strand VIII and R^{130}, R^{132}, R^{133} on the loop between β-strands IX and X, that bind heparin. In which, R^{118} is conserved among members and corresponding to R^{130} in FGF22 is K^{140} in FGF7, leaving four unique arginines in FGF22. On the surface, these six residues are adjacent, and together form the part of HBS-4 that is parallel to HBS1 (Fig 5.14E). The other members of the subfamily only have one residue in each of these areas that engages heparin. Hence, the four arginine residues unique to FGF22 make its HBS4 different from that of the other members of the FGF7 sub-family. Consequently, in terms of heparin binding, FGF22 is the most distinct member of the subfamily.

In addition to HBS-4, FGF22 possesses a HBS-3 on β-strand I. In this area, even though R^{41}, R^{43} and R^{44} are close in the sequence, modifications on the peptides (peptides 6 and 7, Table 5.14B) indicates that only two of them reacted with HPG and the other reacted with PGO. Inspection of the surface shows that R^{41} and R^{43} are adjacent, but R^{44} is distant from any labelled residues. This observation suggested that the former two are heparin-binding residues. R^{41} and R^{43} are separated from HBS-1 by an acidic barrier (Figure 14C), so likely forming an independent HBS-3. K^{87} (data) and R^{77} (peptide 8, Table 5.14B) were found to engage heparin. On the surface they are close to R^{41} and R^{43}, so are part of HBS-3.

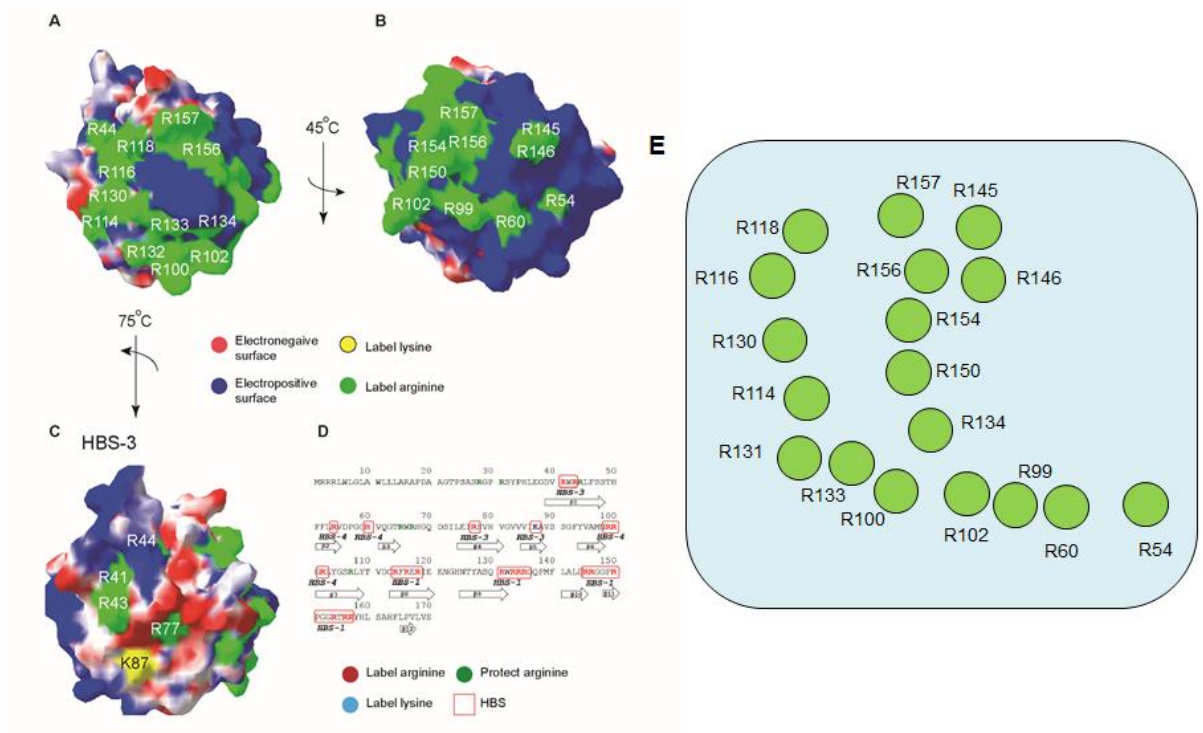


Figure 5.14: Heparin binding arginine and lysine residues in FGF22.

The model of FGF22 structure is built on the template of FGF10 crystal structure (PBD code **1nun** [31]). The electrostatic potential of FGF22 was computed using the Poisson-Boltzmann algorithm in Swiss-PDBView with the positively charged areas coloured *blue* and the negatively charged areas coloured *red*. Labelled lysine residues (Table 5.14C) are presented as *yellow*. Labelled arginine residues (Tables 5.14A, B) are coloured *green*. **A)** The positions of basic residues R^{100}, R^{102} (peptide 3, Table 5.14A), (peptide 2, Table 5.14B), R^{114} (peptide 8, Table 5.14A), R^{116}, R^{118} (peptide 6, Table 5.14A), R^{132}, R^{133}, R^{134} (peptide 6, Table 5.14A) (peptide 5, Table 5.14B), R^{156} and R^{157} (peptide 4 and 5, Table 5.14A) (peptide 4, Table 5.14B) on the surface. **B)** The positions of R^{54} (peptide 2, Table 5.14A) (peptide 3, Table 5.14B), R^{60} (peptide 2, Table 5.14A) (peptide 3, Table 5.14B), R^{99} (peptide 3, Table 5.14A), (peptide 2, Table 5.14B), R^{145} and R^{146} (peptide 4 and 5, Table 5.14A) (peptide 4, Table 5.14B) on the surface. **C)** The positions of basic residues K^{87} (Table 5.14C), R^{77} (peptide 8, Table 5.14B), R^{41}, R^{43} (peptides 6 and 7, Table 5.14B) on the surface of FGF22. On surface, between these

arginine residues and other residues is negatively charged. **D) Sequence, secondary structure, location of labelled arginine and lysine residues in heparin-binding sites.** Uniprot ID of FGF22 sequence is: Q9HCT0. **E)** Schematic of the relative positions of the labelled residues of FGF22.

5.3 DISCUSSION

5.3.1 Architecture of HBSs in the paracrine FGFs

The alignment of the amino acid sequences of the 15 paracrine FGFs with ClustalX was done to gain an insight into the diversification of their HBSs. The core beta trefoil structure is fairly constant across the 15 members of the 5 sub-families, but, as expected, the amino acid sequences of these FGFs vary to a greater extent. GAG binding regulates important aspects of FGF function: stability, diffusion and formation of signalling complexes (Sections 1.4.6). Thus, it is now possible to examine how changes in amino acid sequence that alter GAG binding relate to the evolution of the FGFs.

HBSs are more conserved within a subfamily than between subfamilies in terms of length, size, and shape. These factors will affect the selectivity and potentially the binding kinetics of the FGF for structures in HS. The type of basic residue (arginine or lysine) likely contributes mainly to the binding kinetics. For example, the HBS1 of FGF10 has more arginine residue than that of FGF7 and this results in a reduced dissociation rate of FGF10 from heparin [215]. The HBS-1 of members of the FGF8 sub-family is the most distinct. FGF8, FGF17 and FGF18 have heparin binding residues located in the area from β-strand VI till the loop between β-strands VII-VIII (Fig 5.15), in addition to the residues in the core of HBS-1, which aligns across the FGF family. Those two parts make the HBS-1 in members of FGF8 subfamily relatively wide and result in a dogleg shape. The HBS1 of the FGF7 sub-family also has a unique architecture, since it abuts the secondary HBS-4. On the surface, the labelled basic residues

form a “T” shaped basic patch of binding and there are potentially different binding modes for two polysaccharide chains to this structure.

5.3.2 Differences among members of a subfamily due to amino acid substitutions

The core of HBS-1 varies in sequence, but is present in all the paracrine FGFs and in this way is relatively invariant. This may reflect in part the need for this core to engage the polysaccharide in the ternary signalling complex with the FGFR [144] [216], which may constrain the amount of change possible. Thus, the greatest changes to heparin binding occur elsewhere on the surface and here, small differences in one or a few amino acids can have a major effect on the structure of HBS-1 and the presence or absence of secondary binding sites.

In the FGF1 subfamily, there is a major difference in terms of binding sites, since FGF1 possesses a single HBS-1, whereas there are three HBSs assigned in FGF2 isolated by acidic borders. The border between HBS-1 and HBS-2 in FGF2 is absent in FGF1, due to the substitution of E^{86} in FGF2 by T^{83}, whereas the substitution of E^{68} in FGF2 by S^{65} in FGF1 abolishes the barrier between HBS-1 and HBS-3 (Section 4.2).

For the FGF4 subfamily, FGF4 and FGF6 both engage heparin through a single HBS-1. However, a closer examination reveals some district features of these two FGFs. In detail, the aligned residue of labelled K^{173} on FGF4 is L^{176} in FGF6. This change disconnects the aligned K^{158} of FGF6 (R^{170}) to the core of HBS-1, whereas the core of HBS-1 and K^{158} of FGF4 are linked. In fact, R^{170} reacted with PGO (Section 5.2.1.2). This alters the direction of sugar engagement on FGF6, compared to FGF4, so that FGF6 engages to heparin in the direction connecting K^{185} and R^{205} (Section 5.2.1.2).

For the FGF9 subfamily, only FGF20 possesses a secondary HBS, HBS-3, consisting of R^{65}, R^{66} and R^{67} (Section 5.2.2.2). This difference stems from the substitution of R^{91} on the loop between the β-strands III-IV in FGF9 by L^{97} in FGF20, resulting in the polysaccharide likely

adopting a different orientation on the surface of FGF20 and so a separation of the HBS-3 region from that of HBS-1 in FGF20. Only FGF20 in this subfamily had a labelled K^{124}, whereas the corresponding K^{121} of FGF9 and R^{120} of FGF16 were protected (Section 5.2.2.2). Although it was suggested previously that this lysine in FGF20 may act as an individual HBSs [37], the addition of the arginine residues that bind heparin into HBS-1 indicates that this lysine is actually an extension of HBS-1.

In the FGF8 subfamily, FGF8 and FGF17 likely engage to heparin through a dogleg shaped basic patch. However, FGF18 is different from other members due to the substitution of D^{121} in FGF18 for S^{121} in FGF17 and N^{139} in FGF8, which creates a negatively charged barrier on the protein surface in FGF18, resulting in the independence of HBS-2 from the primary HBS-1 (Section 5.2.3.3). On β-strand I, R^{52} of FGF8 was labelled, whereas the corresponding arginines in FGF17 and FGF18 were not. This arginine on FGF8 contributes to the extension of HBS-1 (Fig 5.8).

In the FGF7 subfamily, FGF3, FGF7 and FGF10 have a conserved arrangement of HBS-1 and HBS-4, which form a "T" shape on the protein surface. However, in FGF22, this is altered. Thus, while there are still two independent binding sites, the structure is more 'Y' shaped, with two arms of HBS-1 and HBS-4 being parallel. This is due to the engagement of heparin by the unique R^{116}, R^{118} on β-strand VIII and R^{132}, R^{133} on the loop between β-strands IX and X of FGF22. Finally, FGF10 differs from the rest of the subfamily in that it does not possess and HBS-3 (Section 5.2.4.2).

5.3.3. Heparin binding sites and the classification of FGFs into subfamilies

It should be recalled that there is a debate regarding the subfamily to which FGF3 and FGF5 belong. From the alignment of amino acid sequence, FGF3 was assigned to the FGF7 subfamily and FGF5 to the FGF4 subfamily, whereas consideration of synteny assigned FGF3 to the FGF4 subfamily and FGF5 to the FGF1 subfamily (Section 1.2.1).

FGF5 possesses a longer loop between β-strands IX and X than that FGF4 and FGF6 (3 residues), hence, the labelling of arginine residues in this area of FGF5 results in a difference in binding to heparin. FGF2 possesses a secondary HBS-2, which locates on β-strand IX and the loop between β-strands IX –X, however, this loop is also short in FGF2. In contrast, this loop is relatively longer in FGF1 with 6 residues, but none of them engage heparin. Thus, the data on HBSs in FGF5 do not add evidence to its assignment to a particular subfamily. It should be noted that this loop is only long in FGF5 and FGF3 and the classification of both of them is still not conclusive.

FGF3 possesses a HBS-3 and a HBS-4, which are similar in architecture to the members of the FGF7 subfamily. In contrast, FGF4 and FGF6 possess a single, long HBS-1, with no evidence for a ‘T’ shaped HBS-1/HBS-4 structure or a HBS3 on their surface. Therefore, the present data support the assignment of FGF3 by sequence alignment to the FGF7 subfamily and consequently the observed synteny would have arisen by chance.

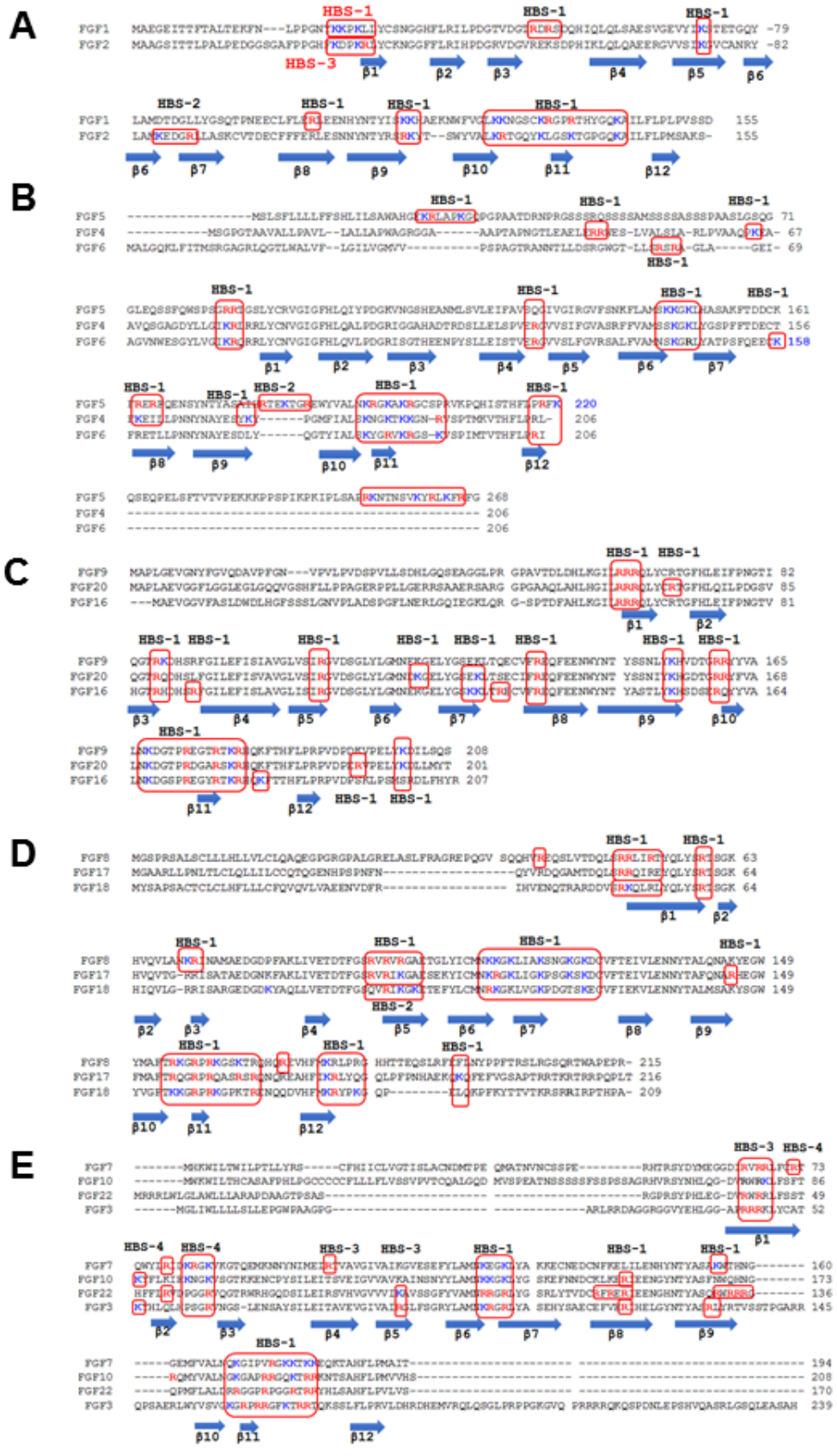

A
B
C
D
E

Figure 5.15: Positions of labelled arginine and lysine residues in the picture of the subfamily sequence alignment. The sequence alignment was performed with ClustalX and the input sequences were taken from Uniprot. Labelled lysine residues are presented as *blue*. Labelled arginine residues are coloured *red*. The secondary structures of the FGF1 subfamily is assigned from a FGF1 structure (PBD code 2erm [33]), FGF4 subfamily from a FGF4 structure [207] **(**PBD code **1ijt)**, FGF9 subfamily from a FGF9 structure [212] PBD code **5w59**), FGF8 subfamily from a FGF8b structure [213] (PBD code **2fdb**) and FGF7 from a FGF7 crystal structure (PBD code **1qqk**) [31]. The assigned HBSs are highlighted in *red box*es.

5.3.4 Assignment of HBSs in the FGFs is supported by measurement of cross-linking of HS chains

The present data together with those acquired previously on lysine residues [120] [36][37] provide insight into the basic residues in the fifteen paracrine FGFs that bind heparin. The existence of secondary HBSs in FGFs emerged from the analysis of the interactions of nested synthetic peptides covering the amino acid sequence of FGF2 and was supported by later work, including lysine protect and label experiments [36]. Analysis of further FGFs identified that there were likely three different secondary HBSs in the paracrine FGFs, HBS-2 located at the area of β-strands IX and X, HBS-3 at the N terminus of β-strand I and sometimes including residues from the C-terminus and, unique to the FGF7 subfamily, HBS4, which with HBS1 formed a T-shape patch, which could only be bound by two heparin chains [36]. The identification of the secondary binding sites led to the idea of the engagement of a single FGF to multiple polysaccharide chains at the same time and this was vindicated by a biophysical analysis, which showed that FGF2 (HBS-1, -2 and -3), but not FGF9 (only HBS-1) was able to cross-link HS chains *in vitro* [126]. More recent work suggested that this interpretation was either too simplistic or that what had been considered to be independent secondary HBSs [37] were in fact extensions of HBS1 [203].

The data acquired here provide a resolution of this issue. Thus, there would appear to be a single HBS-1 on FGF4 and FGF6 and no secondary HBS. This re-assignment of residues is due to what was considered previously to be an electronegative region between the HBSs arising from the peptide backbone and/or a hydrophilic side chain, rather than an acidic side chain. The updated assignment agrees with data that have been acquired on cross-linking HS chains in a bush: FGF4 and Halo-FGF6 were found not to cross-link HS chains [126] [203]. In some cases only partial cross-linking was observed, e.g., FGF17 [203], but there was no structure-based explanation for this observation. The identification of an dogleg shaped basic binding site (Sections 5.2.4) suggests that this FGF may have two binding modes, one to a single chain with a substantial bend in the polysaccharide, the other to two chains. FGF10 was demonstrated to partially cross-link HS chains as well [203]. Since FGF10 does not possess HBS-3, the partial cross-link observed in this FGF10 is entirely due to HBS-1 and HBS-4. Therefore, their simultaneous engagement to two polysaccharide chains would appear to be only one binding mode, in other(s) the heparin-binding residues of either one of HBS1 or HBS4 in FGF10 would not engage the polysaccharide. The structure of available binding sites in HS would then likely determine the partitioning and equilibrium between these binding modes and so the degree of cross-linking observed.

The presence of secondary HBSs in particular paracrine FGFs is likely to affect the functions of the growth factors. Differences in the diffusion of FGFs in pericellular matrix [217] [125] are likely to arise from the interplay between the kinetics of binding of HBS1, the presence of secondary HBSs and the availability of unoccupied binding sites in matrix HS. However, the relationship between the ability of an FGF to cross-link HS chains and the diffusion of that FGF in extracellular matrix remains to be established.

5.3.5 The definition of secondary HBS

The identification of both arginine and lysine residues that bind heparin has allowed refinement of the assignment of residues to the primary HBS-1 and secondary HBSs in the FGFs. There are at least two factors that determine whether heparin binding residues are in a secondary HBS or part of the primary HBS-1. One is the presence of an acidic border made by acidic side chains, rather than the sum of partial charges of the backbone amides and polar side chains, such as alcohols. The second is the geometrical relationship of the basic residues on the surface of the FGF that are not separated by an acidic border. While heparin, and certainly HS has considerable flexibility [84] [37] [85] [91], there are limits to this. Thus, whereas a sugar chain may accommodate a 90° bend in a binding site, such dogleg binding sites, as found on FGF8 subfamily members, need to be examined carefully to establish whether a NA domain of HS could be able to bend over the angle. In contrast, 'T' shaped binding structures, as found in the FGF7 subfamily would appear to require two polysaccharide chains.

5.3.6 HS and FGFR may compete for HBS3 residues

The observation that heparin binding sites on some FGFs include residues that have been found in co-crystals to bind to receptors has been reported in some other proteins, including chemokines [155], FGF2 [218] [130], FGF1 [36][143]. In FGF8 [213], R^{48} has been shown to interact with the alternative spliced region of D3 whereas, R^{52} engages to D3 of FGFR2c. On the loop of β-strand I and II, R^{59} bound to D2 region of FGFR2c [213]. In FGF9, while R^{62} and R^{64} are part of HBS-1, they are also involved in binding to FGFR1c [219]. In FGF7, R^{65} of HBS-3 is defined as a residue interacting with FGFR2IIIb [220]. Interestingly, the latter study also found another function of R^{65} besides interacting with D3, whereby it forms an intramolecular hydrogen bond, which is potentially required for maintaining the local conformation of FGF7 [220]. The fact that this arginine was labelled indicates that any such

intramolecular bonding is dynamic and that the side chain of this arginine is exposed to solvent at least part of the time. When an arginine is involved in stable intramolecular bonding it does not react with PGO/HPG and is not a substrate for Arg-C, as found for R^{42} in FGF2 (Section 4.2). Thus, the heparin-binding residue of HBS-3 and some others may have two competing partners, HS and FGFR. This suggests that upon the formation of the ternary signalling complex, FGF:HS:FGFR, that there may be a rearrangement of HS:FGF interactions to allow full interaction with the FGFR. The interactions with polysaccharide may also, in some cases, prevent the formation of a signalling complex with the FGFR.

5.3.7 Basic residues at the C-terminus of FGFs that bind heparin

FGF5 possesses both a long N-terminus extending well beyond β-strand I of the β-trefoil structure and the longest C-terminal tail of the FGF family. The functions of these parts of FGF5 are unknown, but the data here suggest that they may be involved in binding HS. FGF3 also has a long C-terminal tail, which is very basic, with nine arginine residues and two lysine residues, though only two arginines R^{186} and R^{192} reacted with HPG (peptide 4, Table 5.13B), indicating that they bind to heparin. The functions of the other lysines and arginines remain unresolved. The residues of C-terminus contributing to the engagement with heparin are also observed in FGF17 (Section 5.2.3.2), FGF9 and FGF20 (Section 5.2.2).

Table 5.1: List of paracrine FGFs. FGF names along with their amino acid numbering of the sequences are according to the UniProt entry. FGF1 is an N-terminal truncated protein [221]. FGF2 does not possess a secretory signal sequence. The full length protein sequence of FGF9, FGF16 and FGF20 was expressed because they contain no signal peptide recognised in Uniprot. FGF5 was expressed as the full length sequence as well. For all other FGFs, the protein expressed was without the Uniprot determined secretory signal sequence. The N-terminal tag of each FGF was indicated.

Name	UniProt accession number	Residues	Molecular mass (kDa)	Tag
FGF1	P05230	16–155	19.1	No
FGF2	P09038-2	1-155	17.3	No
FGF3	P11487-1	18-239	26.887	6xHis
FGF4	P08620	31–206	19.29	6x His
FGF5	P12034-1	1-268	29.551	Halo
FGF6	P10767	38-208	51.1	Halo
FGF7	P21781	32–194	22.2	6x His
FGF8b	P55075-3	23-215	57.5	Halo
FGF9	P31371	1–208	23.4	6x His
FGF10	O15520	38-208	22.7	6x His
FGF16	O43320	1-207	21.4	6x His
FGF17	O60258-1	23-216	23.3	6x His
FGF18	O76093	28–207	20.9	6x His
FGF20	Q9NP95	1-211	58.8	Halo
FGF22	Q9HCT0	23-170	52.3	Halo

Table 5.2A: FGF4 MS analysis based on prediction by Prospector. The reference number (Ref.no.) of the peptide is column 1. The two columns under "Native FGF4" present the predicted m/z and sequence of each peptide after digestion of FGF4 by chymotrypsin. The bold arginines in the 3rd column ("Sequence") are those identified as labelled with HPG. The first two columns under "FGF4 P&L" present the predicted and observed m/z for FGF4 after selective labelling. FGF4 bound to a heparin affinity column (protection step) was reacted with PGO, following elution, then arginine residues of binding sites exposed to solvent were reacted with HPG (labelling step). The 3rd column identifies the peptide as being considered from sequence alignment in the literature [102] to be in an HBS of FGF4. The final column indicates the modification occurring on the arginine residues of peptides.

	Native FGF4		FGF4 P&L			
Ref.no.	**Predicted native peptide m/z**	**Sequence**	**Predicted modified peptide m/z**	**m/z observed**	**Match HBS**	**Modifications**
1	976.50	^{44}E**RR**WESL^{50}	1240.55	1240.62	HBS3 R^{45} and R^{46}	2R+ 132 **(HPG)** + 132 (**HPG)**
2	1237.86	^{77}LLGIK**R**L**RR**L^{86}	1601.93	1601.94	HBS3 R^{82}	2R + 132 **(HPG)** + 116 + 116
3	1700.92	^{130}GVAS**R**FFVAMSSKGKL^{145}	1834.94	1834.94	R^{134}	R + 134
4	1806.99	^{119}SPVE**R**GVVSIFGVAS**R**F^{135}	2055.03	2055.11	HBS3 R^{123}	2R + 132 **(HPG)** + 116
5	1902.01	^{39}EAELE**RR**WESLVALSL^{53}	2166.06	2165.99	HBS3 R^{45} and R^{46}	2R + 132 **(HPG)** + 132 (HPG)
6	2049.11	^{117}ELSPVE**R**GVVSIFGVAS**R**F^{135}	2413.20	2413.11	HBS3 R^{123}	2R + 132 (**HPG)** + 232
7	2313.21	^{95}HLQALPDG**R**IGGAHADT**R**DSLL^{116}	2577.25	2577.38	R^{103} and R^{112}	2R + 132 (HPG) + 132 (HPG)
8	2412.29	^{54}SLA**R**LPVAAQPKEAAVQSGAGDYL^{77}	2644.36	2644.36	R^{57}	R + 232
9	2309.27	^{116}LELSPVE**R**GVVSIFGVAS**R**FF^{136}	2673.35	2673.30	HBS3 R^{123}	2R + 132 **(HPG)** + 232
10	2658.49	^{179}IALSKNGKTKKGN**R**VSPTMKVTHF^{202}	2791.51	2791.47	HBS3 R^{192}	2R + 132 (**HPG**)

Table 5.2B: FGF4 MS analysis based on prediction by PeptideMass. The reference number (Ref.no.) of the peptide is column 1. The two columns under "Native FGF4" present the predicted m/z and sequence of each peptide after digestion of FGF4 by chymotrypsin. The bold arginines in the 3rd column ("Sequence") are those identified as labelled with HPG. The first two columns under "FGF4 P&L" present the predicted and observed m/z for FGF4 after selective labelling. FGF4 bound to a heparin affinity column (protection step) was reacted with PGO, following elution, then arginine residues of binding sites exposed to solvent were reacted with HPG (labelling step). The 3rd column identifies the peptide as being considered from sequence alignment in the literature [102] to be in an HBS of FGF4. The final column indicates the modification occurring on the arginine residues of peptides.

	Native FGF4		FGF4 P&L			
Ref.no.	**Predicted native peptide m/z**	**Sequence**	**Predicted modified peptide m/z**	**m/z observed**	**Match HBS**	**Modifications**
1	1257.69	^{44}E**RR**WESLVAL^{53}	1523.74	1523.74	HBS3 R^{45} and R^{46}	2 R + 132 **(HPG)** + 132 **(HPG)**
2	1400.92	^{78}LGIK**R**L**RR**L^{86}	1997.08	1997.00	HBS3 R^{82}	3 R + 132 **(HPG)** + 232 + 232
3	1806.99	^{119}SPVE**R**GVVSIFGVAS**R**F^{135}	2055.03	2055.10	HBS3 R^{123}	2 R + 132 **(HPG)** + 116
4	2062.06	^{97}QALPDG**R**IGGAHADT**R**DSLL^{116}	2312.12	2312.07	R^{103} and R^{112}	2 R + 134 + 116
5	2049.11	^{117}ELSPVE**R**GVVSIFGVAS**R**F^{135}	2413.20	2413.11	HBS3 R^{123}	2 R + 132 **(HPG)** + 232
6	2115.21	^{79}GIK**R**L**RR**LYCNVGIGFHL^{96}	2513.28	2513.19	HBS3 R^{82}	3 R + 132 **(HPG)** + 134 + 134
7	2496.33	^{117}ELSPVE**R**GVVSIFGVAS**R**FF^{136}	2762.38	2762.46	HBS3 R^{123}	2 R + 132 **(HPG)** + 134
8	2855.56	^{119}SPVE**R**GVVSIFGVAS**R**FFVA^{139}	3219.65	3219.65	HBS3 R^{123}	2 R + 132 **(HPG)** + 232

Table 5.3A: FGF6 MS analysis based on prediction by Prospector. The reference number (Ref.no.) of the peptide is column 1. The two columns under "Native FGF6" present the predicted m/z and sequence of each peptide after digestion of FGF6 by chymotrypsin. The bold arginines in the 3rd column ("Sequence") are those identified as labelled with HPG. The first two columns under "FGF6 P&L" present the predicted and observed m/z for FGF6 after selective labelling. FGF6 bound to a heparin affinity column (protection step) was reacted with PGO, following elution, then arginine residues of binding sites exposed to solvent were reacted with HPG (labelling step). The 3rd column identifies the peptide as being considered from sequence alignment in the literature [102] to be in an HBS of FGF6. The final column indicates the modification occurring on the arginine residues of peptides.

	Native FGF6		**FGF6 P&L**			
Ref.no.	**Predicted native peptide m/z**	**Sequence**	**Predicted modified peptide m/z**	**m/z observed**	**Match HBS**	**Modifications**
1	859.51	^{58}LS**R**S**R**AGL^{65}	1224.61	1224.60	HBS3 R^{60} and R^{62}	2R + 133 + 232
2	1083.63	^{199}TVTHFLP**RI**^{206}	1215.65	1215.67	HBS3 R^{205}	1R + 132 **(HPG)**
3	1288.80	^{80}VGIK**R**Q**RR**LY^{89}	1652.87	1652.96	HBS3 R^{84}	3R + 132 **(HPG)** + 116 + 116
4	1414.83	^{187}G**R**VK**R**GSKVSPIM^{199}	1678.87	1678.94	HBS1 R^{188} and R^{191}	2R + 132 **(HPG)** + 132 **(HPG)**
5	1853.02	^{131}FGV**R**SALFVAMNSKG**R**L^{147}	2201.12	2201.11	R^{134} and R^{146}	2R + 116 + 232
6	2028.10	^{55}GTLLS**R**S**R**AGLAGEIAGVNW^{74}	2292.14	2292.14	HBS3 R^{60} and R^{62}	2R + 132 **(HPG)** + 132 **(HPG)**
7	2580.422	^{118}LEISTVE**R**GVVSLFGV**R**SAL^{137}	2944.56	2944.51	HBS3 R^{125}	2R + 132 **(HPG)** + 232

Table 5.3B: FGF6 MS analysis based on prediction by PeptideMass. The reference number (Ref.no.) of the peptide is column 1. The two columns under "Native FGF6" present the predicted m/z and sequence of each peptide after digestion of FGF6 by chymotrypsin. The bold arginines in the 3rd column ("Sequence") are those identified as labelled with HPG. The first two columns under "FGF6 P&L" present the predicted and observed m/z for FGF6 after selective labelling. FGF6 bound to a heparin affinity column (protection step) was reacted with PGO, following elution, then arginine residues of binding sites exposed to solvent were reacted with HPG (labelling step). The 3rd column identifies the peptide as being considered from sequence alignment in the literature [102] to be in an HBS of FGF6. The final column indicates the modification occurring on the arginine residues of peptides.

	Native FGF6		**FGF6 P&L**			
Ref. no.	**Predicted native peptide m/z**	**Sequence**	**Predicted modified peptide m/z**	**m/z observed**	**Match HBS**	**Modifications**
1	859.51	^{58}LS**R**S**R**AGL^{65}	1224.61	1224.60	HBS3 R^{60} and R^{62}	2R + 133 + 232
2	1083.63	^{199}TVTHFLP**RI**^{206}	1215.65	1215.67	HBS3 R^{205}	1R + 132 **(HPG)**
3	1288.80	^{80}VGIK**R**Q**RR**LY^{89}	1652.87	1652.96	HBS3 R^{84}	3R + 132 **(HPG)** + 116 + 116
4	1414.83	^{187}G**R**VK**R**GSKVSPIM^{199}	1678.87	1678.94	HBS1 R^{188} and R^{191}	2R + 132 **(HPG)** + 132 **(HPG)**
5	1853.02	^{131}FGV**R**SALFVAMNSKG**R**L^{147}	2201.12	2201.11	R^{134} and R^{146}	2R + 116 + 232
6	2028.10	^{55}GTLLS**R**S**R**AGLAGEIAGVNW^{74}	2292.14	2292.14	HBS3 R^{60} and R^{62}	2R + 116 + 232
7	2479.44	^{187}G**R**VK**R**GSKVSPIMTVTHFL^{203}	2744.49	2744.49	HBS1 R^{188} and R^{191}	2R + 132 **(HPG)** + 132 **(HPG)**
8	2580.422	^{118}LEISTVE**R**GVVSLFGV**R**SAL^{137}	2944.56	2944.51	HBS3 R^{125}	2R + 132 **(HPG)** + 232

Table 5.4A: FGF5 MS analysis based on prediction by Prospector. The reference number (Ref.no.) of the peptide is column 1. The two columns under "Native FGF5" present the predicted m/z and sequence of each peptide after digestion of FGF5 by chymotrypsin. The bold arginines in the 3rd column ("Sequence") are those identified as labelled with HPG. The first two columns under "FGF5 P&L" present the predicted and observed m/z for FGF5 after selective labelling. FGF5 bound to a heparin affinity column (protection step) was reacted with PGO, following elution, then arginine residues of binding sites exposed to solvent were reacted with HPG (labelling step). The 3rd column identifies the peptide as being considered from sequence alignment in the literature [102] to be in an HBS of FGF5. The final column indicates the modification occurring on the arginine residues of peptides.

	Native FGF5		**FGF5 P&L**			
Ref. no.	**Predicted native peptide m/z**	**Sequence**	**Predicted modified peptide m/z**	**m/z observed**	**Match HBS**	**Modifications**
1	866.54	^{262}**R**LKF**R**F^{267}	1132.59	1132.46	HBS3 R^{262} and R^{266}	2R + 132 **(HPG)** + 132 **(HPG)**
2	1017.54	^{82}SPSG**RR**TGSL^{91}	1281.59	1281.63	HBS3 R^{86} and R^{87}	2R + 132 **(HPG)** + 132 **(HPG)**
3	1228.57	^{163}**R**E**R**FQENSY^{171}	1492.61	1492.66	HBS2 R^{163} and R^{165}	2R + 132 **(HPG)** + 132 **(HPG)**
4	1316.61	^{157}TDDCKF**R**E**R**F^{166}	1680.69	1680.89	HBS2 R^{163} and R^{165}	2R + 132 **(HPG)** + 232
5	1641.85	^{175}ASAIH**R**TEKTG**R**EW^{188}	1905.89	1905.86	HBS2 R^{180} and R^{186}	2R + 132 **(HPG)** + 132 **(HPG)**
6	1909.08	^{250}SAP**R**KNTNSVKY**R**LKF**R**^{266}	2305.14	2305.13	HBS3 R^{253}, R^{262} and R^{266}	3R + 132 **(HPG)** x 3
7	1912.98	^{82}SPSG**RR**TGSLYC**R**VGIGF^{99}	2509.24	2509.32	HBS3 R^{86} and R^{87}	3R + 132 **(HPG)** + 232 + 232
8	2227.12	^{80}QWSPSG**RR**TGSLYC**R**VGIGF^{99}	2723.25	2723.35	HBS3 R^{86} and R^{87}	3R + 132 **(HPG)** + 132 **(HPG)** + 232
9	2632.45	^{193}NK**R**GKAK**R**GCSP**R**VKPQHISTHF^{215}	3130.57	3130.57	HBS1 R^{195} and R^{200}	3R + 132 **(HPG)** + 132 **(HPG)** + 232
10	2915.64	^{190}VALNK**R**GKAK**R**GCSP**R**VKPQHISTHF^{215}	3546.80	3546.77	HBS1 R^{195} and R^{200}	3R + 132 **(HPG)** + 250 +250

Table 5.4B: FGF5 MS analysis based on prediction by PeptideMass. The reference number (Ref.no.) of the peptide is column 1. The two columns under "Native FGF5" present the predicted m/z and sequence of each peptide after digestion of FGF5 by chymotrypsin. The bold arginines in the 3rd column ("Sequence") are those identified as labelled with HPG. The first two columns under "FGF5 P&L" present the predicted and observed m/z for FGF5 after selective labelling. FGF5 bound to a heparin affinity column (protection step) was reacted with PGO, following elution, then arginine residues of binding sites exposed to solvent were reacted with HPG (labelling step). The 3rd column identifies the peptide as being considered from sequence alignment in the literature [102] to be in an HBS of FGF5. The final column indicates the modification occurring on the arginine residues of peptides.

	Native FGF5		FGF5 P&L			
Ref .no.	**Predicted native peptide m/z**	**Sequence**	**Predicted modified peptide m/z**	**m/z observed**	**Match HBS**	**Modifications**
1	532.32	^{216}LP**R**F^{219}	764.39	764.22	HBS3 R^{218}	1R + 232
2	607.33	^{163}**R**E**R**F^{166}	871.37	871.23	HBS2 R^{163} and R^{165}	2R + 132 **(HPG)** + 132 **(HPG)**
3	751.39	^{93}C**R**VGIGF^{99}	867.42	867.40	R^{94}	1R + 116
4	810.46	^{20}AHGEK**R**L^{26}	942.48	942.53	HBS3 R^{25}	1R + 132 **(HPG)**
5	1017.54	^{82}SPSG**RR**TGSL^{91}	1281.59	1281.63	HBS3 R^{86} and R^{87}	2R + 132 **(HPG)** + 132 **(HPG)**
6	1364.73	^{250}SAP**R**KNTNSVKY^{261}	1496.75	1496.67	HBS3 R^{253}	1R + 132 **(HPG)**
7	1641.85	^{175}ASAIH**R**TEKTG**R**EW^{188}	1905.89	1905.87	HBS2 R^{180} and R^{186}	2R + 132 **(HPG)** + 132 **(HPG)**
8	2632.45	^{193}NK**R**GKAK**R**GCSP**R**VKPQHISTHF^{215}	3130.57	3130.57	HBS1 R^{195} and R^{200}	3R + 132 **(HPG)** + 132 **(HPG)** + 232

Table 5.4C: Selective labelling of lysine residues. FGF5 peptide analysis based on prediction by PeptideMass. The reference number of the peptide is followed by the predicted m/z and with the sequence of the peptide following cleavage of native FGF5 by chymotrypsin. The two columns under "Native FGF5" present the predicted m/z and sequences of peptides after chymotrypsin digestion of FGF5. The first two columns under "FGF5 protect and label (P&L)" present the observed and predicted m/z for FGF5 after modifications of lysine residues. The third column is the assignment of the peptide to one of the three HBSs of FGF5 [36], [183]. The final columns indicates the modification occurring on the lysine residues of the peptides.

	Native FGF5		FGF5 protect and label (P&L)			
Ref .no.	**Predicted native peptide m/z**	**Sequence**	**Predicted modified peptide m/z**	**m/z theoretical**	**Published HBS**	**Modifications**
1	660.44	^{145}S**KK**G**K**L^{150}	1356.55	1356.47	HBS1 K^{146}, K^{147} and K^{149}	3K + 232 **(biotin) x 3**
2	1316.61	^{157}TDDC**K**FRERF^{166}	1358.54	1358.62	K^{161}	1K + 42 (acetyl)
3	810.46	^{20}AHGE**K**RL^{26}	1042.46	1042.47	HBS3 K^{24}	1K + 232 **(biotin)**
4	1471.79	^{216}LPRF**K**QSEQPEL^{227}	1703.80	1703.88	HBS3 K^{220}	1K + 232 **(biotin)**
5	1633.91	^{250}SAPR**K**NTNSV**K**YRL^{263}	2097.93	2098.12	HBS3 K^{254} and K^{260}	2K + 232 **(biotin) x 2**
6	1641.85	^{175}ASAIHRTE**K**TGREW^{188}	1874.02	1873.86	HBS1 K^{183}	1K + 232 **(biotin)**
7	2009.95	^{100}HLQIYPDG**K**VNGSHEANM^{117}	2051.96	2052.07	K^{108}	1K + 42 (acetyl)
8	2785.34	^{27}AP**K**GQPGPAATDRNPRGSSSRQSSSSAM^{54}	3017.47	3017.35	HBS3 K^{29}	1K + 232 **(biotin)**
9	2632.45	^{193}N**K**RG**K**A**K**RGCSPRV**K**PQHISTHF^{215}	3370.69	3370.59	HBS1 K^{194}, K^{197} and K^{199}	4K + 232 **(biotin) x 3** + 42 (acetyl)

Table 5.5A: FGF9 peptide analysis based on prediction by Prospector. The reference number (Ref.no.) of the peptide is column 1. The two columns under "Native FGF9" present the predicted m/z and sequence of each peptide after digestion of FGF9 by chymotrypsin. The bold arginines in the 3rd column ("Sequence") are those identified as labelled with HPG. The first two columns under "FGF9 P&L" present the predicted and observed m/z for FGF9 after selective labelling. FGF9 bound to a heparin affinity column (protection step) was reacted with PGO, following elution, then arginine residues of binding sites exposed to solvent were reacted with HPG (labelling step). The 3rd column identifies the peptide as being considered from sequence alignment in the literature [102] to be in an HBS of FGF9. The final column indicates the modification occurring on the arginine residues of peptides.

	Native FGF9		FGF9 protect and label (P&L)			
Ref .no.	**Predicted native peptide m/z**	**Sequence**	**Predicted modified peptide m/z**	**m/z theoretical**	**Published HBS**1	**Modifications**
1	996.472	^{67}YC**R**TGFHL^{74}	1132.02	1131.50	R^{69}	1R + 134
2	891.53	^{62}**RRR**QLY^{67}	1288.12	1287.59	HBS3 R^{62}, R^{63} and R^{64}	3R + 132 **(HPG) x 3**
3	1287.58	^{130}TQECVF**R**EQF^{140}	1419.79	1419.61	HBS1 R^{137}	1R + 132 **(HPG)**
4	1131.60	^{154}KHVDTG**RR**Y^{162}	1495.17	1495.69	HBS1 R^{160} and R^{161}	2R + 132 **(HPG)** + 232
5	1504.92	^{55}DHLKGIL**RRR**QL^{66}	1900.90	1900.98	HBS3 R^{62}, R^{63} and R^{64}	3R + 132 **(HPG) x 3**
6	1844.79	^{130}TQECVF**R**EQFEENW^{144}	1977.15	1976.81	HBS1 R^{137}	1R + 132 **(HPG)**
7	1740.92	^{153}YKHVDTG**RR**YYVAL^{166}	2005.01	2004.96	HBS1 R^{160} and R^{161}	2R + 132 **(HPG)** + 132 **(HPG)**
8	2023.06	^{185}THFLP**R**PVDPDKVPELY^{201}	2155.04	2155.09	HBS1 R^{190}	1R + 132 **(HPG)**
9	2103.07	^{75}EIFPNGTIQGT**R**KDHS**R**F^{92}	2367.14	2367.11	HBS1 R^{86} and R^{91}	2R + 132 **(HPG)** + 132 **(HPG)**
10	2156.143	^{167}NKDGTP**R**EGT**R**TK**R**HQKF^{184}	2420.15	2420.18	HBS1 R^{173}, R^{177} and R^{180}	3R + 132 **(HPG) x 2**

Table 5.5B: FGF9 peptide analysis based on prediction by PeptideMass. The reference number (Ref.no.) of the peptide is column 1. The two columns under "Native FGF9" present the predicted m/z and sequence of each peptide after digestion of FGF9 by chymotrypsin. The bold arginines in the 3rd column ("Sequence") are those identified as labelled with HPG. The first two columns under "FGF9 P&L" present the predicted and observed m/z for FGF9 after selective labelling. FGF9 bound to a heparin affinity column (protection step) was reacted with PGO, following elution, then arginine residues of binding sites exposed to solvent were reacted with HPG (labelling step). The 3rd column identifies the peptide as being considered from sequence alignment in the literature [102] to be in an HBS of FGF9. The final column indicates the modification occurring on the arginine residues of peptides.

	Native FGF9		FGF9 protect and label (P&L)			
Ref .no.	**Predicted native peptide m/z**	**Sequence**	**Predicted modified peptide m/z**	**m/z theoretical**	**Published HBS**.	**Modifications**
1	996.472	^{67}YC**R**TGFHL^{74}	1132.02	1131.50	R^{69}	1R + 134
2	891.53	^{62}**RRR**QLY^{67}	1288.12	1287.59	HBS3 R^{62}, R^{63} and R^{64}	3R + 132 **(HPG) x 3**
3	1165.62	^{105}VSI**R**GVDSGLY^{115}	1297.09	1297.642	HBS1 R^{108}	1R + 132 **(HPG)**
4	1287.58	^{130}TQECVF**R**EQF^{140}	1419.79	1419.61	HBS1 R^{137}	1R + 132 **(HPG)**
5	1131.60	^{154}KHVDTG**RR**Y^{162}	1495.17	1495.69	HBS1 R^{160} and R^{161}	2R + 132 **(HPG)** + 232
6	2103.07	^{75}EIFPNGTIQGT**R**KDHS**R**F^{92}	2367.14	2367.11	HBS1 R^{86} and R^{91}	2R + 132 **(HPG)** + 132 **(HPG)**
7	2156.143	^{167}NKDGTP**R**EGT**R**TK**R**HQKF^{184}	2420.15	2420.18	HBS1 R^{173}, R^{177} and R^{180}	3R + 132 **(HPG) x 2**

Table 5.6A: FGF20 peptide analysis based on prediction by Prospector. The reference number (Ref.no.) of the peptide is column 1. The two columns under "Native FGF20" present the predicted m/z and sequence of each peptide after digestion of FGF20 by chymotrypsin. The bold arginines in the 3rd column ("Sequence") are those identified as labelled with HPG. The first two columns under "FGF20 P&L" present the predicted and observed m/z for FGF20 after selective labelling. FGF20 bound to a heparin affinity column (protection step) was reacted with PGO, following elution, then arginine residues of binding sites exposed to solvent were reacted with HPG (labelling step). The 3rd column identifies the peptide as being considered from sequence alignment in the literature [102] to be in an HBS of FGF20. The final column indicates the modification occurring on the arginine residues of peptides.

	Native FGF20		FGF20 protect and label (P&L)			
Ref. no.	**Predicted native peptide m/z**	**Sequence**	**Predicted modified peptide m/z**	**m/z theoretical**	**Published HBS**.	**Modifications**
1	728.46	^{65}**RRR**QL^{69}	1124.56	1124.53	HBS-3 ^{65}R, R^{66}, R^{67}	3R + 132 **(HPG) x 3**
2	583.27	^{71}C**R**TGF^{75}	715.318	715.2868	HBS-1 R^{72}	1R + 132 **(HPG)**
3	579.29	^{140}**R**EQF^{143}	711.366	711.3096	HBS-1 R^{140}	1R + 132 **(HPG)**
4	1653.80	^{152}SSNIYKHGDTG**RR**Y^{165}	1917.928	1917.851	HBS-1 R^{163} and R^{164}	3R + 132 **(HPG) x 2**
5	1271.77	^{25}LLPPAGE**R**PPLL^{37}	1405.709	1405.796	R^{33}	1R + 134
6	1002.56	^{108}VSI**R**GVDSGL^{117}	1134.511	1134.579	HBS-3 R^{111}	1R + 132 **(HPG)**
7	1519.81	^{157}KHGDTG**RR**YFVAL^{169}	1783.858	1783.855	HBS-1 R^{163} and R^{164}	2R + 132 **(HPG) x 2**
8	2381.29	^{166}VALNKDGTP**R**DGA**R**SK**R**HQK^{186}	2645.293	2645.333	HBS-1 R^{176}, R^{180}, and R^{183}	3R + 132 **(HPG) x 2**
9	2098.10	^{169}NKDGTP**R**DGA**R**SK**R**HQKF^{187}	2362.068	2362.143	HBS-1 R^{176}, R^{180}, and R^{183}	3R + 132 **(HPG) x 2**

Table 5.6B: FGF20 peptide analysis based on prediction by PeptideMass. The reference number (Ref.no.) of the peptide is column 1. The two columns under "Native FGF20" present the predicted m/z and sequence of each peptide after digestion of FGF20 by chymotrypsin. The bold arginines in the 3rd column ("Sequence") are those identified as labelled with HPG. The first two columns under "FGF20 P&L" present the predicted and observed m/z for FGF20 after selective labelling. FGF20 bound to a heparin affinity column (protection step) was reacted with PGO, following elution, then arginine residues of binding sites exposed to solvent were reacted with HPG (labelling step). The 3rd column identifies the peptide as being considered from sequence alignment in the literature [102] to be in an HBS of FGF20. The final column indicates the modification occurring on the arginine residues of peptides.

	Native FGF20		FGF20 protect and label (P&L)			
Ref. no.	**Predicted native peptide m/z**	**Sequence**	**Predicted modified peptide m/z**	**m/z theoretical**	**Published HBS**	**Modifications**
1	728.46	^{65}**RRR**QL^{69}	1124.56	1124.53	HBS-3 ^{65}R, R^{66}, R^{67}	3R + 132 **(HPG) x 3**
2	996.472	^{69}YC**R**TGFHL^{77}	1128.49	1128.49	HBS-1 R^{72}	1R + 132 **(HPG)**
3	1002.558	^{108}VSI**R**GVDSGL^{117}	1134.51	1134.58	HBS-3 R^{111}	HR + 132 **(HPG)**
4	1089.555	KHGDTG**RR**Y^{166}	1353.60	1353.60	HBS-1 R^{163} and R^{164}	2R + 132 **(HPG) x 2**
5	1271.77	^{25}LLPPAGE**R**PPLL^{37}	1405.71	1405.79	R^{33}	1R + 134
6	1519.81	^{157}KHGDTG**RR**YFVAL^{169}	1783.86	1783.86	HBS-1 R^{163} and R^{164}	2R + 132 **(HPG) x 2**
7	1653.80	^{152}SSNIYKHGDTG**RR**Y^{165}	1917.93	1917.85	HBS-1 R^{163} and R^{164}	3R + 132 **(HPG) x 2**
8	1800.878	^{162}SSNIYKHGDTG**RR**YF	2064.86	2064.92	HBS-1 R^{163} and R^{164}	2R + 132 **(HPG) x 2**
9	2381.29	^{166}VALNKDGTP**R**DGA**R**SK**R**HQK^{186}	2895.38	2895.40	HBS-1 R^{176}, R^{180}, and R^{183}	3R + 132 **(HPG) x 2** + 250
10	2381.29	^{166}VALNKDGTP**R**DGA**R**SK**R**HQK^{186}	2761.35	2761.36	HBS-1 R^{176}, R^{180}, and R^{183}	3R + 132 **(HPG) x 2** + 116
11	2665.33	^{71}C**R**TGFHLQILPDGSVQGT**R**QDHSL^{94}	2929.351	2929.37	HBS-1 R^{72} and R^{89}	2R + 132 **(HPG) x 2**
12	1516.85	^{191}LP**R**PVDPE**R**VPEL^{203}	1880.93	1880.92	HBS-1 R^{199} and R^{192}	2R + 132 **(HPG)** + 232

Table 5.7A: FGF16 peptide analysis based on prediction by Prospector. The reference number (Ref.no.) of the peptide is column 1. The two columns under "Native FGF16" present the predicted m/z and sequence of each peptide after digestion of FGF16 by chymotrypsin. The bold arginines in the 3rd column ("Sequence") are those identified as labelled with HPG. The first two columns under "FGF16 P&L" present the predicted and observed m/z for FGF16 after selective labelling. FGF16 bound to a heparin affinity column (protection step) was reacted with PGO, following elution, then arginine residues of binding sites exposed to solvent were reacted with HPG (labelling step). The 3rd column identifies the peptide as being considered from sequence alignment in the literature [102] to be in an HBS of FGF16. The final column indicates the modification occurring on the arginine residues of peptides.

	Native FGF16		**FGF16 protect and label (P&L)**			
Ref. no.	**Predicted native peptide m/z**	**Sequence**	**Predicted modified peptide m/z**	**m/z theoretical**	**Published HBS.**	**Modifications**
1	833.4087	^{67}C**R**TGFHL^{73}	1083.45	1083.45	R^{68}	1R + 250
2	1122.52	^{166}NKDGSP**R**EGY^{175}	1254.52	1254.54	HBS1 R^{172}	1R + 132 **(HPG)**
3	1314.63	^{130}T**R**ECVF**R**EQF^{139}	1578.76	1578.67	HBS3 R^{131} and HBS1 R^{136}	2R + 132 **(HPG) x 2**
4	1430.75	^{124}YGSKKLT**R**ECVF^{135}	1562.78	1562.77	HBS3 R^{131}	1R + 132 **(HPG)**
5	1475.66	^{152}YKHSDSE**R**QYY^{162}	1607.73	1607.68	HBS1 R^{159}	1R + 132 **(HPG)**
6	1485.82	^{176}**R**TK**R**HQKFTHF^{186}	1765.97	1765.96	HBS1 R^{176} and R^{179}	2R + 132 **(HPG) x 2**
7	1568.77	^{162}YVALNKDGSP**R**EGY^{175}	1700.80	1700.79	HBS1 R^{172}	1R + 132 **(HPG)**
8	1669.99	^{97}ISLAVGLISI**R**GVDSGL^{113}	1801.91	1802.01	HBS1 R^{107}	1R + 132 **(HPG)**
9	1833.05	^{97}ISLAVGLISI**R**GVDSGLY^{114}	1965.06	1965.07	HBS1 R^{107}	1R + 132 **(HPG)**
10	2390.21	^{74}EIFPNGTVHGT**R**HDHS**R**FGIL^{94}	2654.23	2654.25	HBS1 R^{131} and HBS3 R^{136}	2R + 132 **(HPG) x 2**

Table 5.7B: FGF16 peptide analysis based on prediction by PeptideMass. The reference number (Ref.no.) of the peptide is column 1. The two columns under "Native FGF16" present the predicted m/z and sequence of each peptide after digestion of FGF16 by chymotrypsin. The bold arginines in the 3rd column ("Sequence") are those identified as labelled with HPG. The first two columns under "FGF16 P&L" present the predicted and observed m/z for FGF16 after selective labelling. FGF16 bound to a heparin affinity column (protection step) was reacted with PGO, following elution, then arginine residues of binding sites exposed to solvent were reacted with HPG (labelling step). The 3rd column identifies the peptide as being considered from sequence alignment in the literature [36], [183] to be in an HBS of FGF16. The final column indicates the modification occurring on the arginine residues of peptides.

	Native FGF16		**FGF16 protect and label (P&L)**			
Ref. no.	**Predicted native peptide m/z**	**Sequence**	**Predicted modified peptide m/z**	**m/z theoretical**	**Published HBS.**	**Modifications**
1	891.53	^{61}**RRR**QLY^{66}	1287.76	1287.59	HBS3 R^{61}, R^{62} and R^{63}	3R + 132 **(HPG) x 3**
2	937.45	^{200}S**R**DLFHY^{206}	1069.66	1069.47	HBS3 R^{90}	1R + 132 **(HPG)**
3	1122.52	^{166}NKDGSP**R**EGY^{175}	1254.52	1254.54	HBS1 R^{172}	1R + 132 **(HPG)**
4	1314.63	^{130}T**R**ECVF**R**EQF^{139}	1578.76	1578.67	HBS3 R^{131} and HBS1 R^{136}	2R + 132 **(HPG) x 2**
5	1393.74	^{118}NE**R**GELYGSKKL^{129}	1525.89	1525.76	R^{120}	1R + 132 **(HPG)**
6	1405.71	^{163}VALNKDGSP**R**EGY^{175}	1537.69	1537.73	HBS1 R^{172}	1R + 132 **(HPG)**
7	1430.75	^{124}YGSKKLT**R**ECVF^{135}	1562.78	1562.77	HBS3 R^{131}	1R + 132 **(HPG)**
8	1475.66	^{152}YKHSDSE**R**QYY^{162}	1607.73	1607.68	HBS1 R^{159}	1R + 132 **(HPG)**
9	1485.82	^{176}**R**TK**R**HQKFTHF^{186}	1765.97	1765.96	HBS1 R^{176} and R^{179}	2R + 132 **(HPG) x 2**
10	1568.77	^{162}YVALNKDGSP**R**EGY^{175}	1700.80	1700.79	HBS1 R^{172}	1R + 132 **(HPG)**
11	1669.99	^{97}ISLAVGLISI**R**GVDSGL^{113}	1801.91	1802.01	HBS1 R^{107}	1R + 132 **(HPG)**

12	1821.97	^{184}THFLP**R**PVDPSKLPSM^{199}	1954.11	1953.99	R^{189}	1R + 132 **(HPG)**
13	1833.05	^{97}ISLAVGLISI**R**GVDSGLY^{114}	1965.06	1965.07	HBS1 R^{107}	1R + 132 **(HPG)**
14	2204.14	^{165}NKDGSP**R**EGY**R**TK**R**HQKF^{182}	2600.10	2600.20	HBS1 R^{172}, R^{176} and R^{179}	3R + 132 **(HPG) x 3**

Table 5.7C: Selective labelling of lysine residues. FGF16 peptide analysis based on prediction by PeptideMass. The reference number of the peptide is followed by the predicted m/z and with the sequence of the peptide following cleavage of native FGF16 by chymotrypsin. The two columns under "Native FGF16" present the predicted m/z and sequences of peptides after chymotrypsin digestion of FGF16. The first two columns under "FGF16 protect and label (P&L)" present the observed and predicted m/z for FGF16 after modifications of lysine residues. The third column is the assignment of the peptide to one of the three HBSs of FGF16 [36], [183]. The final columns indicates the modification occurring on the lysine residues of the peptides.

	Native FGF16		**FGF16 protect and label (P&L)**			
	Predicted native peptide m/z	**Sequence**	**m/z observed**	**m/z theoretical**	**Published HBS**	**Modifications**
1	1138.517	N**K**DGSPREGY	1371.627	1371.75	K^{167}	1K + 232 **(biotin)**
2	1595.781	**K**HSDSERQYYVAL	1827.791	1827.93	K^{153}	1K + 232 **(biotin)**
3	1430.746	YGS**KK**LTRECVF	1895.00	1894.766	K^{127} and K^{128}	2K + 232 **(biotin) x 2**
4	2903.594	RT**K**RHQ**K**FTHFLPRPVDPS**K**LPSM	3603.77	3603.624	K^{178}, K^{182} and K^{195}	3K + 232 **(biotin) x 3**

Table 5.8A: FGF8 peptide analysis based on prediction by Prospector. The reference number (Ref.no.) of the peptide is column 1. The two columns under "Native FGF8" present the predicted m/z and sequence of each peptide after digestion of FGF8 by chymotrypsin. The bold arginines in the 3rd column ("Sequence") are those identified as labelled with HPG. The first two columns under "FGF8 P&L" present the predicted and observed m/z for FGF8 after selective labelling. FGF8 bound to a heparin affinity column (protection step) was reacted with PGO, following elution, then arginine residues of binding sites exposed to solvent were reacted with HPG (labelling step). The 3rd column identifies the peptide as being considered from sequence alignment in the literature [102] to be in an HBS of FGF8. The final column indicates the modification occurring on the arginine residues of peptides.

	Native FGF8		**FGF8 protect and label (P&L)**			
Ref. no.	**Predicted native peptide m/z**	**Sequence**	**Predicted modified m/z**	**m/z theoretical**	**Published HBS.**	**Modifications**
1	1087.6218	^{42}VTDQLS**RR**L^{50}	1351.73	1351.664	HBS2 R^{48} and R^{49}	2R + 132 **(HPG) x 2**
2	1097.5698	^{33}TQHV**R**EQSL^{41}	1229.78	1229.591	HBS2 R^{37}	1R + 132 **(HPG)**
3	1357.7659	^{94}GS**R**V**R**V**R**GAETGL^{106}	1753.67	1753.829	HBS2 R^{96,} R^{98} and R^{100}	3R + 132 **(HPG) x 3**
4	1520.8292	^{94}GS**R**V**R**V**R**GAETGLY^{107}	1916.9	1916.893	HBS2 R^{96,} R^{98} and R^{100}	3R + 132 **(HPG) x 3**
5	1648.7748	^{69}ANK**R**INAMAEDGDPF^{83}	1780.63	1780.796	HBS1 R^{72}	1R + 132 **(HPG)**
6	1786.9275	^{190}**R**FEFLNYPPFT**R**SL^{203}	2283.11	2283.038	R^{190} and R^{201}	3R + 232 + 232
7	1966.0042	^{195}NYPPFT**R**SL**R**GSQ**R**TW^{210}	2464.19	2464.13	R^{201}, R^{204} and R^{208}	3R + 232 + 134+ 134
8	2149.1876	^{51}I**R**TYQLYS**R**TSGKHVQVL^{68}	2497.28	2497.282	HBS-2 R^{52} ; HBS1 R^{59}	2R + 232 + 116
9	2166.1738	^{33}TQHV**R**EQSLVTDQLS**RR**L^{50}	2562.22	2562.245	HBS2 R^{37}, R^{48} and R^{49}	3R + 132 **(HPG) x 3**
10	2589.446	^{154}TRKG**R**P**R**KGSKT**R**QHQ**R**EVHF^{174}	3232.94	3233.03	HBS1 R^{155}, R^{158}, R^{160}, R^{170}	5R + 132 **(HPG)** x 4 + 116
11	2766.3767	^{69}ANK**R**INAMAEDGDPFAKLIVETDTF^{93}	2898.3	2898.398	HBS1 R^{72}	1R + 132 **(HPG)**

Table 5.8B: FGF8 peptide analysis based on prediction by PeptideMass. The reference number (Ref.no.) of the peptide is column 1. The two columns under "Native FGF8" present the predicted m/z and sequence of each peptide after digestion of FGF8 by chymotrypsin. The bold arginines in the 3rd column ("Sequence") are those identified as labelled with HPG. The first two columns under "FGF8 P&L" present the predicted and observed m/z for FGF8 after selective labelling. FGF8 bound to a heparin affinity column (protection step) was reacted with PGO, following elution, then arginine residues of binding sites exposed to solvent were reacted with HPG (labelling step). The 3rd column identifies the peptide as being considered from sequence alignment in the literature [102] to be in an HBS of FGF8. The final column indicates the modification occurring on the arginine residues of peptides.

	Native FGF8		**FGF8 protect and label (P&L)**			
Ref. no	**Predicted native peptide m/z**	**Sequence**	**Predicted modified peptide m/z**	**m/z theoretical**	**Published HBS.**	**Modifications**
1	1357.7659	^{94}GS**R**V**R**V**R**GAETGL^{106}	1753.67	1753.83	HBS2 R^{96,} R^{98} and R^{100}	3R + 132 **(HPG) x 3**
2	1520.8292	^{94}GS**R**V**R**V**R**GAETGLY^{107}	1916.9	1916.89	HBS2 R^{96,} R^{98} and R^{100}	3R + 132 **(HPG) x 3**
3	1963.073	^{176}K**R**LP**R**GHHTTEQSL**R**F^{191}	2677.48	2677.24	R^{177}, R^{180} and R^{190}	3R + 232 + 232 + 250
4	2078.088	^{194}LNYPPFT**R**SL**R**GSQ**R**TW^{210}	2710.71	2711.24	R^{201}, R^{204} and R^{208}	2R + 250 + 250 + 134
5	2166.1738	^{33}TQHV**R**EQSLVTDQLS**RR**L^{50}	2562.22	2562.24	HBS2 R^{37}, R^{48} and R^{49}	3R + 132 **(HPG) x 3**
6	2514.372	^{55}QLYS**R**TSGKHVQVLANK**R**INAM^{76}	2896.83	2896.46	HBS1 R^{59} and R^{72}	2R + 132 **(HPG) + 250**
7	2536.396	^{954}GS**R**V**R**V**R**GAETGLYICMNKKGKL^{116}	2932.84	2932.46	HBS2 R^{96,} R^{98} and R^{100}	3R + 132 **(HPG) x 3**
8	2766.3767	^{69}ANK**R**INAMAEDGDPFAKLIVETDTF^{93}	2898.3	2898.39	HBS1 R^{72}	1R + 132 **(HPG)**

Table 5.8C: Selective labelling of lysine residues. FGF8 peptide analysis based on prediction by PeptideMass. The reference number of the peptide is followed by the predicted m/z and with the sequence of the peptide following cleavage of native FGF8 by chymotrypsin. The two columns under "Native FGF8" present the predicted m/z and sequences of peptides after chymotrypsin digestion of FGF8. The first two columns under "FGF8 protect and label (P&L)" present the observed and predicted m/z for FGF8 after modifications of lysine residues. The third column is the assignment of the peptide to one of the three HBSs of FGF8 [36], [183]. The final columns indicates the modification occurring on the lysine residues of the peptides.

	Native FGF8		**FGF8 protect and label (P&L)**			
Ref .no.	**Predicted native peptide m/z**	**Sequence**	**Predicted modified peptide m/z**	**m/z observed**	**Published HBS.**	**Modifications**
1	1034.585	^{108}ICMN**KK**G**K**L^{116}	1556.11	1556.61	HBS1 K^{111}, K^{112} and K^{114}	3K + 232 **(biotin) x 2** + 42 (acetyl) and oxidation
2	1366.715	^{117}IA**K**SNG**K**G**K**DCVF^{129}	1872.62	1872.7	HBS1 K^{119}, K^{123} and K^{125}	3K + 232 **(biotin) x 2** + 42 (acetyl)
3	1443.69	^{139}TALQNA**K**YEGWY^{150}	1485.70	1485.00	K^{145}	1K + 42 (acetyl)
4	1976.986	^{69}AN**K**RINAMAEDGDPFA**K**L^{86}	2253.50	2252.01	HBS1 K^{71}	2K + 232 **(biotin)** + 42 (acetyl)
5	2499.304	^{175}M**K**RLPRGHHTTEQSLRFEFL^{194}	2737.57	2737.314	HBS2 K^{176}	1K + 232 **(biotin)** and oxidation
6	3117.65	^{150}YMAFTR**K**GRPR**K**GS**K**TRQHQREVHF^{174}	3813.96	3813.68	HBS1 K^{155}, K^{161} and K^{164}	3K + 232 **(biotin) x 3**

Table 5.9A: FGF17 MS analysis based on prediction by Prospector. The reference number (Ref.no.) of the peptide is column 1. The two columns under "Native FGF17" present the predicted m/z and sequence of each peptide after digestion of FGF17 by chymotrypsin. The bold arginines in the 3rd column ("Sequence") are those identified as labelled with HPG. The first two columns under "FGF17 P&L" present the predicted and observed m/z for FGF17 after selective labelling. FGF17 bound to a heparin affinity column (protection step) was reacted with PGO, following elution, then arginine residues of binding sites exposed to solvent were reacted with HPG (labelling step). The 3rd column identifies the peptide as being considered from sequence alignment in the literature [36], [183] to be in an HBS of FGF17. The final column indicates the modification occurring on the arginine residues of peptides.

	Native FGF17		**FGF17 on column**			
Ref.no.	**Predicted native peptide m/z**	**Sequence**	**Predicted modified peptide m/z**	**m/z observed**	**Match HBS**	**Modifications**
1	1078.586	^{106}ICMNK**R**GKL^{116}	1210.607	1210.531	HBS1 R^{113}	1R + 132 **(HPG)**
2	997.4599	^{142}QNARHEGW^{149}	1129.481	1129.43	HBS1 R^{145}	1R + 132 **(HPG)**
3	1654.749	^{34}NQYV**R**DQGAMTDQL^{47}	1786.77	1786.746	R^{38}	1R + 132 **(HPG)**
4	1316.613	^{139}TAFQNA**R**HEGW^{149}	1448.634	1448.643	HBS1 R^{145}	1R + 132 **(HPG)**
5	1107.602	^{48}S**RR**QI**R**EY^{58}	1487.67	1487.735	HBS1 R^{49} and R^{50}	3R + 132 **(HPG) x 2 + 116**
6	2915.478	^{151}MAFT**R**QG**R**P**R**QAS**R**S**R**QNQ**R**EAHF^{174}	3691.598	3691.01	HBS1 R^{155}, R^{158}, R^{160}, R^{164} and R^{166}	6R + 132 **(HPG)** x 5 + 116

Table 5.9B: FGF17 MS analysis based on prediction by PeptideMass. The reference number (Ref.no.) of the peptide is column 1. The two columns under "Native FGF17" present the predicted m/z and sequence of each peptide after digestion of FGF17 by chymotrypsin. The bold arginines in the 3rd column ("Sequence") are those identified as labelled with HPG. The first two columns under "FGF17 P&L" present the predicted and observed m/z for FGF17 after selective labelling. FGF17 bound to a heparin affinity column (protection step) was reacted with PGO, following elution, then arginine residues of binding sites exposed to solvent were reacted with HPG (labelling step). The 3rd column identifies the peptide as being considered from sequence alignment in the literature [36], [183] to be in an HBS of FGF17. The final column indicates the modification occurring on the arginine residues of peptides.

	Native FGF17		**FGF17 on column**			
Ref. no.	**Predicted native peptide m/z**	**Sequence**	**Predicted modified peptide m/z**	**m/z observed**	**Match HBS**	**Modifications**
1	1564.82	^{44}TDQLS**RR**QI**R**EY^{58}	1962.89	1960.91	HBS1 R^{49} and R^{50}	3R + 132 **(HPG) x 3**
2	1316.61	^{139}TAFQNA**R**HEGW^{149}	1448.63	1448.64	HBS1 R^{145}	1R + 132 **(HPG)**
3	1181.54	^{34}NQYV**R**DQGAM^{43}	1313.56	1313.58	R^{38}	1R + 132 **(HPG)** and oxidation
4	2077.22	^{198}VGSAPT**RR**TK**R**T**RR**PQPL^{215}	2957.43	2942.40	R^{204}, R^{205}, R^{208}, R^{210} and R^{211}	5R + 116 + 250 + 250 + 132 **(HPG) x 2**
5	1579.86	^{94}GS**R**V**R**IKGAESEKY^{107}	1843.90	1843.89	HBS1 R^{96} and R^{98}	2R + 132 **(HPG) x 2**
6	2702.41	^{59}S**R**TSGKHVQVTG**RR**ISATAEDGNKF^{83}	3182.52	3182.48	HBS1 R^{60}; R^{71} and R^{72}	3R + 132 **(HPG) + 116 + 116**
7	693.4453	^{175}IK**R**LY^{178}	825.46	825.16	HBS1 R^{177}	1R + 132 **(HPG)**

Table 5.10A: FGF18 MS analysis based on prediction by Prospector. The reference number (Ref.no.) of the peptide is column 1. The two columns under "Native FGF18" present the predicted m/z and sequence of each peptide after digestion of FGF18 by chymotrypsin. The bold arginines in the 3rd column ("Sequence") are those identified as labelled with HPG. The first two columns under "FGF18 P&L" present the predicted and observed m/z for FGF18 after selective labelling. FGF18 bound to a heparin affinity column (protection step) was reacted with PGO, following elution, then arginine residues of binding sites exposed to solvent were reacted with HPG (labelling step). The 3rd column identifies the peptide as being considered from sequence alignment in the literature [36], [183] to be in an HBS of FGF18. The final column indicates the modification occurring on the arginine residues of peptides.

	Native FGF18		**FGF18 on column**			
Ref .no.	**Predicted native peptide m/z**	**Sequence**	**Predicted modified peptide m/z**	**m/z observed**	**Match HBS**	**Modifications**
1	949.51	^{108}CMN**R**KGKL^{116}	1082.54	1082.59	HBS1 R^{112}	1R + 132 **(HPG)**
2	1062.59	^{107}LCMN**R**KGKL^{116}	1294.66	1294.76	HBS1 R^{112}	1R + 232
3	1388.76	^{58}YS**R**TSGKHIQVL^{69}	1520.79	1520.86	HBS1 R^{60}	1R + 132 **(HPG)**
4	1579.79	^{70}G**RR**ISA**R**GEDGDKY^{83}	1828.85	1828.94	R^{71}, R^{72} and R^{76}	3R + 116 + 134
5	1754.94	^{94}GSQV**R**IKGKETEFYL^{108}	1886.96	1886.95	HBS1 R^{98}	1R + 132 **(HPG)**
6	2007.12	^{176}K**R**YPKGQPELQKPFKY^{191}	2139.14	2139.09	HBS1 R^{177}	1R + 132 **(HPG)**
7	3303.725	^{143}SGWYVGFTKKG**R**P**R**KGPKT**R**^{166}	3594.01	3593.87	HBS1 R^{158}, R^{160} and R^{166}	3R + 132 **(HPG) x 2**

Table 5.10B: FGF18 MS analysis based on prediction by PeptideMass. The reference number (Ref.no.) of the peptide is column 1. The two columns under "Native FGF18" present the predicted m/z and sequence of each peptide after digestion of FGF18 by chymotrypsin. The bold arginines in the 3rd column ("Sequence") are those identified as labelled with HPG. The first two columns under "FGF18 P&L" present the predicted and observed m/z for FGF18 after selective labelling. FGF18 bound to a heparin affinity column (protection step) was reacted with PGO, following elution, then arginine residues of binding sites exposed to solvent were reacted with HPG (labelling step). The 3rd column identifies the peptide as being considered from sequence alignment in the literature [36], [183] to be in an HBS of FGF18. The final column indicates the modification occurring on the arginine residues of peptides.

	Native FGF18		**FGF18 on column**			
Ref .no.	**Predicted native peptide m/z**	**Sequence**	**Predicted modified peptide m/z**	**m/z observed**	**Match HBS**	**Modifications**
1	949.51	^{108}CMN**R**KGKL^{116}	1097.53	1097.44	HBS1 R^{112}	1R + 132 **(HPG)** and oxidation
2	1062.59	^{107}LCMN**R**KGKL^{116}	1194.61	1194.60	HBS1 R^{112}	1R + 132 **(HPG)**
3	1346.72	^{175}MK**R**YPKGQPEL^{185}	1478.75	1478.82	HBS1 R^{177}	1R + 132 **(HPG)**
4	1388.76	^{58}YS**R**TSGKHIQVL^{69}	1520.79	1520.86	HBS1 R^{60}	1R + 132 **(HPG)**
5	1715.96	^{176}K**R**YPKGQPELQKPF^{189}	1847.98	1847.98	HBS1 R^{177}	1R + 132 **(HPG)**
6	1754.94	^{94}GSQV**R**IKGKETEFYL^{108}	1886.96	1886.95	HBS1 R^{98}	1R + 132 **(HPG)**
7	1877.09	^{191}TTVTK**R**S**RR**I**R**PTHPA^{207}	2573.30	2573.41	R^{197}, R^{199}, R^{200} and R^{202}	4R + 232 x 4
8	1891.79	^{70}G**RR**ISA**R**GEDGDKYAQL^{86}	2388.07	2387.09	R^{71}, R^{72} and R^{76}	3R + 250 + 134 + 116
9	2810.528	^{151}VGFTKKG**R**P**R**KGPKT**R**ENQQ^{170}	3212.64	3207.60	HBS1 R^{158}, R^{160} and R^{166}	3R + 132 **(HPG) x 2 + 134**

Table 5.11A: FGF7 peptide analysis based on prediction by Prospector. The reference number (Ref.no.) of the peptide is column 1. The two columns under "Native FGF7" present the predicted m/z and sequence of each peptide after digestion of FGF7 by chymotrypsin. The bold arginines in the 3rd column ("Sequence") are those identified as labelled with HPG. The first two columns under "FGF7 P&L" present the predicted and observed m/z for FGF7 after selective labelling. FGF7 bound to a heparin affinity column (protection step) was reacted with PGO, following elution, then arginine residues of binding sites exposed to solvent were reacted with HPG (labelling step). The 3rd column identifies the peptide as being considered from sequence alignment in the literature [36], [183] to be in an HBS of FGF7. The final column indicates the modification occurring on the arginine residues of peptides.

	Native FGF7		**FGF7 protect and label (P&L)**			
Ref.no.	**Predicted native peptide m/z**	**Sequence**	**m/z observed**	**Predicted modified peptide m/z**	**Published HBS.**	**Modifications**
1	872.3719	^{71}C**R**TQWY^{76}	1004.456	1004.393	HBS1 R^{72}	1R + 132 **(HPG)**
2	1003.445	^{70}FC**R**TQWY^{76}	1135.444	1135.467	HBS1 R^{72}	1R + 132 **(HPG)**
3	1132.524	^{70}FC**R**TQWYL^{77}	1264.599	1264.546	HBS1 R^{72}	1R + 132 **(HPG)**
4	2278.245	^{95}L**R**IDK**R**GKVKGTQEMKNNY^{78}	2542.38	2542.33	HBS1 R^{78} and R^{82}	1R + 132 **(HPG) x 2**
5	2399.222	^{65}MEGGDI**R**V**RR**LFC**R**TQWYL^{77}	2795.241	2795.286	HBS3 R^{65}, R^{67} and R^{68}	3R + 132 **(HPG) x 3**
6	2554.359	^{96}NIMEI**R**TVAVGIVAIKGVESEFY^{119}	2686.375	2686.38	HBS1 R^{101}	1R + 132 **(HPG)**
7	2564.229	^{63}DYMEGGDI**R**V**RR**LFC**R**TQWY^{76}	2960.25	2960.292	HBS3 R^{65}, R^{67} and R^{68}	3R + 132 **(HPG) x 3**
8	2706.589	^{165}ALNQKGIPV**R**GKKTKKEQKTAHF^{188}	2838.577	2838.61	HBS1 R^{175}	1R + 132 **(HPG)**

Table 5.11B: FGF7 peptide analysis based on prediction by PeptideMass. The reference number (Ref.no.) of the peptide is column 1. The two columns under "Native FGF7" present the predicted m/z and sequence of each peptide after digestion of FGF7 by chymotrypsin. The bold arginines in the 3rd column ("Sequence") are those identified as labelled with HPG. The first two columns under "FGF7 P&L" present the predicted and observed m/z for FGF7 after selective labelling. FGF7 bound to a heparin affinity column (protection step) was reacted with PGO, following elution, then arginine residues of binding sites exposed to solvent were reacted with HPG (labelling step). The 3rd column identifies the peptide as being considered from sequence alignment in the literature [36], [183] to be in an HBS of FGF7. The final column indicates the modification occurring on the arginine residues of peptides.

	Native FGF7		**FGF7 protect and label (P&L)**			
Ref. no	**Predicted native peptide m/z**	**Sequence**	**m/z observed**	**Predicted modified peptide m/z**	**Published HBS**.	**Modifications**
1	969.46	^{71}C**R**TQWYL^{77}	1001.46	1101.48	HBS1 R^{72}	1R + 132 **(HPG)**
2	1003.45	^{70}FC**R**TQWY^{76}	1135.45	1135.47	HBS1 R^{72}	1R + 132 **(HPG)**
3	1170.67	^{60}EGGDI**R**V**RR**L^{69}	1869.99	1869.92	HBS3 R^{65}, R^{67} and R^{68}	3R + 232 x 3
4	1992.03	^{60}EGGDI**R**V**RR**LFC**R**TQW^{75}	2522.27	2522.12	HBS1 R^{72} HBS3 R^{65}, R^{67} and R^{68}	4R + 132 **(HPG) x 3 + 116**
5	2099.92	^{41}ATNVNCSSPE**R**HT**R**SYDY^{58}	2368.27	2368.01	R^{51} and R^{54}	1R + 134 **x 2**
6	2706.59	^{165}ALNQKGIPV**R**GKKTKKEQKTAHF^{188}	2838.58	2838.61	HBS1 R^{175}	1R + 132 **(HPG)**

Table 5.12A: FGF10 peptide analysis based on prediction by Prospector. The reference number (Ref.no.) of the peptide is column 1. The two columns under "Native FGF10" present the predicted m/z and sequence of each peptide after digestion of FGF10 by chymotrypsin. The bold arginines in the 3rd column ("Sequence") are those identified as labelled with HPG. The first two columns under "FGF10 P&L" present the predicted and observed m/z for FGF10 after selective labelling. FGF10 bound to a heparin affinity column (protection step) was reacted with PGO, following elution, then arginine residues of binding sites exposed to solvent were reacted with HPG (labelling step). The 3rd column identifies the peptide as being considered from sequence alignment in the literature [36], [183] to be in an HBS of FGF10. The final column indicates the modification occurring on the arginine residues of peptides.

	Native FGF10		**FGF10 protect and label (P&L)**			
Ref. no.	**Predicted native peptide m/z**	**Sequence**	**m/z observed**	**Predicted modified peptide m/z**	**Published HBS.**	**Modifications**
1	1157.654	^{74}QGDV**R**W**R**KL^{82}	1523.83	1523.73	R^{78} and R^{80}	2R + 134 +250
2	1333.586	^{168}NWQHNG**R**QMY^{177}	1465.548	1465.607	HBS1 R^{174}	1R + 132 **(HPG)**
3	1349.58	^{68}NWQHNG**R**QMY^{177}	1481.616	1481.602	HBS1 R^{174}	1R + 132 **(HPG)** and oxidation
4	1449.825	$^{\mathbf{80}}$**R**KLFSFTKYFL^{90}	1581.917	1581.846	HBS3 R^{80}	1R + 132 **(HPG)**
5	1537.823	^{74}QGDV**R**W**R**KLFSF^{85}	1805.781	1805.865	R^{78} and R^{80}	2R + 134 *** 2**
6	1638.723	^{166}ASFNWQHNG**R**QMY^{177}	1770.841	1770.744	HBS1 R^{174}	1R + 132 **(HPG)**
7	2367.298	^{180}NGKGAP**RR**GQKT**RR**KNTSAHF^{201}	2763.41	2763.39	HBS1 R^{187}, R^{188}, R^{193} and R^{194}	4R + 132 **(HPG) x 3**

Table 5.12B: FGF10 peptide analysis based on prediction by PeptideMass. The reference number (Ref.no.) of the peptide is column 1. The two columns under "Native FGF10" present the predicted m/z and sequence of each peptide after digestion of FGF10 by chymotrypsin. The bold arginines in the 3rd column ("Sequence") are those identified as labelled with HPG. The first two columns under "FGF10 P&L" present the predicted and observed m/z for FGF10 after selective labelling. FGF10 bound to a heparin affinity column (protection step) was reacted with PGO, following elution, then arginine residues of binding sites exposed to solvent were reacted with HPG (labelling step). The 3rd column identifies the peptide as being considered from sequence alignment in the literature [36], [183] to be in an HBS of FGF10. The final column indicates the modification occurring on the arginine residues of peptides.

	Native FGF10		**FGF10 protect and label (P&L)**			
Ref.no.	**Predicted native peptide m/z**	**Sequence**	**m/z observed**	**Predicted modified peptide m/z**	**Published HBS.**	**Modifications**
1	797.47	^{80}**R**KLFSF^{85}	1031.00	1031.53	R^{80}	1R + 232
2	1157.65	^{74}QGDV**R**W**R**KL^{82}	1523.83	1523.73	R^{78} and R^{80}	2R + 134 +250
3	1332.65	^{170}QHNG**R**QMYVAL^{180}	1464.84	1464.67	HBS1 R^{174}	1R + 132 **(HPG)**
4	1390.68	^{58}SSPSSAG**R**HV**R**SY^{70}	1640.72	1640.73	R^{65} and R^{68}	2R + 116 + 134
5	1491.65	^{166}ASFNWQHNG**R**QM^{176}	1623.70	1623.68	HBS1 R^{174}	1R + 132 **(HPG)**
6	2367.30	^{180}NGKGAP**RR**GQKT**RR**KNTSAHF^{201}	2763.41	2763.39	HBS1 R^{187}, R^{188}, R^{193} and R^{194}	4R + 132 **(HPG) x 3**
7	2809.27	^{147}NNDCKLKE**R**IEENGYNTYASFNW^{169}	2941.26	2941.29	HBS1 R^{155}	1R + 132 **(HPG)**
8	2373.11	^{142}GSKEFNNDCKLKE**R**IEENGY^{161}	2504.84	2505.13	HBS1 R^{155}	1R + 132 **(HPG)**

Table 5.13A: FGF3 peptide analysis based on prediction by Prospector. The reference number (Ref.no.) of the peptide is column 1. The two columns under "Native FGF3" present the predicted m/z and sequence of each peptide after digestion of FGF3 by chymotrypsin. The bold arginines in the 3rd column ("Sequence") are those identified as labelled with HPG. The first two columns under "FGF3 P&L" present the predicted and observed m/z for FGF3 after selective labelling. FGF3 bound to a heparin affinity column (protection step) was reacted with PGO, following elution, then arginine residues of binding sites exposed to solvent were reacted with HPG (labelling step). The 3rd column identifies the peptide as being considered from sequence alignment in the literature [36], [183] to be in an HBS of FGF3. The final column indicates the modification occurring on the arginine residues of peptides.

	Native FGF3		**FGF3 protect and label (P&L)**			
Ref .no.	**Predicted native peptide m/z**	**Sequence**	**m/z observed**	**Predicted modified peptide m/z**	**Published HBS.**	**Modifications**
1	1010.633	^{40}GGAP**RRR**KL^{48}	1274.742	1274.675	HBS3 R^{44}, R^{45} and R^{46}	2R + 132 **(HPG) x 2**
2	1108.604	^{98}AMNK**R**G**R**LY^{106}	1356.616	1356.652	HBS1 R^{102} and R^{104}	2R + 132 **(HPG) + 116**
3	1207.574	^{125}GYNTYAS**R**LY^{134}	1339.681	1339.595	HBS1 R^{132}	1R + 132 **(HPG)**
4	1264.676	^{57}QLHPSG**R**VNGSL^{68}	1396.752	1396.697	HBS1 R^{63}	1R + 132 **(HPG)**
5	1264.676	^{56}QLHPSG**R**VNGSL^{68}	1396.752	1396.697	HBS1 R^{63}	1R + 132 **(HPG)**
6	1351.781	^{168}KT**RR**TQKSSLF^{178}	1715.914	1715.87	HBS1 R^{170} and R^{171}	2R + 132 **(HPG)** + 232
7	1429.814	^{155}VSVNGKG**R**P**RR**GF^{177}	1827.45	1827.88	HBS1 R^{162}, R^{164} and R^{165}	3R + 132 **(HPG) x 2 + 134**
8	1521.843	^{93}SG**R**YLAMNK**R**G**R**L^{105}	1920.01	1919.91	HBS1 R^{102} and R^{104}	3R + 132 **(HPG) x 2 + 134**
9	1668.912	^{92}FSG**R**YLAMNK**R**G**R**L^{105}	1801.001	1800.933	HBS1 R^{102} and R^{104}	1R + 132 **(HPG)**

Table 5.13B: FGF3 peptide analysis based on prediction by PeptideMass. The two columns under "Native FGF3" present the predicted m/z and sequence of each peptide after digestion of FGF3 by chymotrypsin. The bold arginines in the 3rd column ("Sequence") are those identified as labelled with HPG. The first two columns under "FGF3 P&L" present the predicted and observed m/z for FGF3 after selective labelling. FGF3 bound to a heparin affinity column (protection step) was reacted with PGO, following elution, then arginine residues of binding sites exposed to solvent were reacted with HPG (labelling step). The 3rd column identifies the peptide as being considered from sequence alignment in the literature [36], [183] to be in an HBS of FGF3. The final column indicates the modification occurring on the arginine residues of peptides.

	Native FGF3		**FGF3 protect and label (P&L)**			
Ref .no.	**Predicted native peptide m/z**	**Sequence**	**m/z observed**	**Predicted modified peptide m/z**	**Published HBS.**	**Modifications**
1	961.536	^{98}AMNK**R**G**R**L^{106}	1225.35	1225.578	HBS1 R^{102} and R^{104}	2R + 132 **(HPG) x 2**
2	1044.511	^{125}GYNTYAS**R**L^{133}	1176.21	1176.532	HBS1 R^{132}	1R + 132 **(HPG)**
3	1429.813	^{155}VSVNGKG**R**P**RR**GF^{177}	1827.49	1827.87	HBS1 R^{162}, R^{164} and R^{165}	3R + 132 **(HPG) x 2 + 134**
4	1451.681	^{184}DH**R**DHEMV**R**QL^{195}	1715.15	1715.723	R^{186} and R^{192}	2R + 132 **(HPG) x 2**
5	1508.782	^{221}EPSHVQAS**R**LGSQL^{264}	1641.61	1641.803	R^{229}	1R + 134
6	1513.819	^{55}HLQLHPSG**R**VNGSL^{68}	1645.96	1645.84	HBS1 R^{63}	1R + 132 **(HPG)**
7	1533.76	^{179}LP**R**VLDH**R**DHEM^{190}	1801.79	1801.81	R^{181} and R^{186}	2R + 134 x 2
8	1542.788	^{26}**RR**DAGG**R**GGVYEHL^{39}	1926.57	1926.86	R^{26}, R^{27} and R^{32}	3R + 134 + 134 + 116
9	1552.882	^{37}EHLGGAP**RRR**KLY^{48}	1950.34	1951.96	HBS3 R^{44}, R^{45} and R^{46}	2R + 132 **(HPG) x 2 + 134**
10	1561.732	^{112}SAECEFVE**R**IHEL^{124}	1693.15	1693.753	HBS1 R^{120}	1R + 132 **(HPG)**
11	1592.877	^{154}YVSVNGKG**R**P**RR**GF^{177}	1989.04	1988.94	HBS1 R^{162}, R^{164} and R^{165}	3R + 132 **(HPG) x 3**
12	1686.98	^{77}EITAVEVGIVAI**R**GLF^{92}	1819.00	1819.00	HBS1 R^{89}	1R + 132 **(HPG)**

Table 5.14A: FGF22 peptide analysis based on prediction by Prospector. The reference number (Ref.no.) of the peptide is column 1. The two columns under "Native FGF22" present the predicted m/z and sequence of each peptide after digestion of FGF22 by chymotrypsin. The bold arginines in the 3rd column ("Sequence") are those identified as labelled with HPG. The first two columns under "FGF22 P&L" present the predicted and observed m/z for FGF22 after selective labelling. FGF22 bound to a heparin affinity column (protection step) was reacted with PGO, following elution, then arginine residues of binding sites exposed to solvent were reacted with HPG (labelling step). The 3rd column identifies the peptide as being considered from sequence alignment in the literature [36], [183] to be in an HBS of FGF22. The final column indicates the modification occurring on the arginine residues of peptides.

	Native FGF22		**FGF22 protect and label (P&L)**			
Ref. no.	**Predicted native peptide m/z**	**Sequence**	**m/z observed**	**Predicted modified peptide m/z**	**Published HBS.**	**Modifications**
1	925.485	^{57}**R**HGQDSIL^{74}	1040.96	1041.50	R^{67}	1R + 116
2	1483.788	^{54}**R**VDPGG**R**VQGT**R**W^{66}	1747.74	1747.83	HBS1 R^{54} and R^{60}	2R + 132 **(HPG) x 2**
3	1398.742	^{94}YVAMN**RR**G**R**LY^{104}	1794.77	1794.806	HBS1 R^{99}, R^{100} and R^{102}	3R + 132 **(HPG) x 3**
4	2054.159	^{140}LALD**RR**GGP**R**PGG**R**T**RR**Y^{158}	2714.31	2714.21	HBS1 R^{145}, R^{146}, R^{150}, R^{153}, R^{155} and R^{156}	6R + 132 **(HPG) x 5**
5	2054.159	^{140}LALD**RR**GGP**R**PGG**R**T**RR**Y^{158}	2848.13	2846.23	HBS1 R^{145}, R^{146}, R^{150}, R^{153}, R^{155} and R^{156}	6R + 132 **(HPG) x 5 + 134**
6	2146.017	^{116}**R**E**R**IEENGHNTYASQ**R**W^{132}	2674.2	2674.102	HBS1 R^{116}, R^{118} and R^{131}	3R + 132 **(HPG) x 3**
7	2533.303	^{95}VAMN**RR**G**R**LYGS**R**LYTVDC**R**F^{115}	3077.72	3077.66	R^{107} HBS1 R^{99}, R^{100}, R^{102} and R^{114}	5R + 132 **(HPG) x 4** and oxidation
8	2533.303	^{95}VAMN**RR**G**R**LYGS**R**LYTVDC**R**F^{115}	3077.72	3077.66	R^{107} HBS1 R^{99}, R^{100}, R^{102} and R^{114}	5R + 132 **(HPG) x 4** + 250 and oxidation

Table 5.14B: FGF22 peptide analysis based on prediction by PeptideMass. The reference number (Ref.no.) of the peptide is column 1. The two columns under "Native FGF22" present the predicted m/z and sequence of each peptide after digestion of FGF22 by chymotrypsin. The bold arginines in the 3rd column ("Sequence") are those identified as labelled with HPG. The first two columns under "FGF22 P&L" present the predicted and observed m/z for FGF22 after selective labelling. FGF22 bound to a heparin affinity column (protection step) was reacted with PGO, following elution, then arginine residues of binding sites exposed to solvent were reacted with HPG (labelling step). The 3rd column identifies the peptide as being considered from sequence alignment in the literature [36], [183] to be in an HBS of FGF22. The final column indicates the modification occurring on the arginine residues of peptides.

	Native FGF22		**FGF22 protect and label (P&L)**			
Ref .no.	**Predicted native peptide m/z**	**Sequence**	**m/z observed**	**Predicted modified peptide m/z**	**Published HBS.**	**Modifications**
1	925.485	^{57}**R**HGQDSIL^{74}	1040.96	1041.50	R^{67}	1R + 116
2	1398.742	^{94}YVAMN**RR**G**R**LY^{104}	1794.77	1794.80	HBS1 R^{99}, R^{100} and R^{102}	3R + 132 **(HPG) x 3**
3	1596.87	^{53}L**R**VDPGG**R**VQGT**R**W^{76}	1994.99	1994.94	R^{76} ; HBS1 R^{54} and R^{60}	3R + 132 **(HPG) x 2** + 134
4	2871.58	^{144}D**RR**GGP**R**PGG**R**T**RR**YHLSAH^{163}	3522.49	3522.54	HBS1 R^{145}, R^{146}, R^{150}, R^{153}, R^{155} and R^{156}	5R + 132 **(HPG) x 5**
5	1160.65	^{132}**RRR**GQPMFL^{141}	1540.73	1540.71	HBS1 R^{132}, R^{133} and R^{134}	3R + 132 **(HPG) x 2 + 116**
6	1333.712	^{37}EGDV**R**W**RR**LF^{46}	1732.81	1732.78	HBS3 R^{41}, R^{43} and R^{44}	3R + 132 **(HPG) x 2 + 134**
7	1186.644	^{37}EGDV**R**W**RR**L^{45}	1585.75	1585.71	HBS3 R^{41}, R^{43} and R^{44}	3R + 132 **(HPG) x 2 + 134**
8	2447.348	^{75}EI**R**SVHVGVVVIKAVSSGFY^{94}	2579.34	2579.37	HBS3 R^{77}	1R + 132 **(HPG)**

Table 5.14C: Selective labelling of lysine residues. FGF22 peptide analysis based on prediction by PeptideMass. The reference number of the peptide is followed by the predicted m/z and with the sequence of the peptide following cleavage of native FGF22 by chymotrypsin. The two columns under "Native FGF22" present the predicted m/z and sequences of peptides after chymotrypsin digestion of FGF22. The first two columns under "FGF22 protect and label (P&L)" present the observed and predicted m/z for FGF22 after modifications of lysine residues. The third column is the assignment of the peptide to one of the three HBSs of FGF22 [36], [183]. The final column indicates the modification occurring on the lysine residues of the peptides.

	Native FGF22		**FGF22 protect and label (P&L)**			
Ref.no.	**Predicted native peptide m/z**	**Sequence**	**m/z observed**	**Predicted modified peptide m/z**	**Published HBS**.	**Modifications**
1	3199.8	^{75}EIRSVHVGVVVI**K**AVSSGFYVAMNRRGRL^{103}	3431.57	3431.81	HBS1 K^{87}	1K + 232 **(biotin)**
2	2146.202	^{75}EIRSVHVGVVVI**K**AVSSGFY^{94}	2377.15	2378.212	HBS1 K^{87}	1K + 232 **(biotin)**

CHAPTER 6 ELECTROSTATIC INTERACTIONS BETWEEN FIBROBLAST GROWTH FACTORS OF FGF7 SUBFAMILY AND CHONROITIN SULFATE, DEMATAN SULFATE IN SOLUTION

6.1 INTRODUCTION

For a variety of reasons, the protein interactions of heparin and HS attract more interest than those of other GAGs. The commercial availability of heparin means that there is a readily available experimental proxy for HS, including in the form of affinity chromatography supports. However, the relative homogeneity of heparin's sulfation and IdoA content compared to HS does not reflect the diversity of sulfation of HS, which is in at least some instances key for the selectivity of interaction with a protein partner (Section 1.3.3.1). CS and DS are abundant in the extracellular matrix (Section 1.3.2.2). Compared to HS and heparin, the number of known protein partners established for CS or DS is far more limited, which may reflect in part the absence of off the shelf affinity chromatography supports for these GAGs. Examination of GAG preferences for the paracrine FGFs [37] [36] indicates that at least some FGFs do engage CS and/or DS, though apparently less strongly than heparin. This is the case for members of the FGF7 subfamily [37] [36] [127], which bind CS species, including DS. To gain an understanding of whether these CS species may engage proteins through the same binding site(s) as heparin, an "in solution" selective labelling protocol was developed using commercial CS and DS. Importantly, such a protocol would bypass the requirement for an affinity column, which has hindered the wider application of the protect and label strategy to any protein interaction involving basic residues. For example, the analysis of the FGF2-pleiotrophin interaction required the immobilisation of the FGF2, and achieving this in a way that did not affect its interactions with its protein partner presented a major challenge [189].

The FGF7 subfamily comprises FGF3, FGF7, FGF10 and FGF22 (Section 1.2.1). Their binding properties to HS, CS, DS and various modifications of heparin have been studied using differential scanning fluorimetry (DSF) [37][36]. Briefly, DSF provides information on the melting temperature of protein and how this changes upon binding to a GAG. In these studies, whereas FGF-3 interacted with DS but not CS, FGF7 and FGF10 bound both CS and DS, though the former more weakly [37][36]. The in solution selective protect and labelling method was initially developed for lysine residues, due to the simpler analysis of the mass spectrometry data.

6.2 RESULTS

6.2.1 FGF7

The lysine residues on FGF7 contributing to the engagement to heparin identified in a previous study [36] were assigned to HBS1 (K^{170}, K^{177}, K^{178}, K^{180} and K^{181} in the core of the canonical HBS-1, K^{155} of β-strand IX, K^{123} and K^{126} on the loop of β-strands VI-VII) and HBS4 (K^{81} and K^{84}) on the loop of β-strands II-III. There are no lysines in the HBS-3 of FGF7, since it is composed entirely of arginine residues (Section 5.2). Here, the interactions between FGF7 and other sulfated GAGs, CS/DS, were determined using in-solution selective labelling, which bypasses the need for an affinity matrix. First, heparin in solution was used, to enable method development to be benchmarked to the previous work done using a heparin affinity column. This method was then applied to CS and DS (Section 2.7).

6.2.1.1. *The interaction of FGF7 with heparin performed in PB buffer containing 50 mM NaCl*

The selective labelling of lysine residues in solution starts with a protection step, in which the solvent exposed lysine side chains of protein bound to heparin are reacted with NHS-acetate, which yields a single acid, stable product. Following the removal of NHS-acetate and

dissociation of protein and heparin using 3 M NaCl (Section 2.9), any residues engaged with heparin will now be exposed to solvent and is able to react with the added NHS-biotin. Biotinylated peptides are readily purified prior to mass spectrometry [120]. The resulting peptides with information about modification, sequence, and final m/z are presented in Tables 1A and the spectra are in Supplementary Figs 6.1A.

Using the list of peptides generated by Prospector, seven peptides containing the biotinylated lysine residues were identified (Table 1A). A number of biotinylated lysine residues corresponded to ones identified previously [36], but there were differences.

A first difference was on β-strand V, since K^{111} was found to be labelled by NHS-biotin (peptide 6, Table 1A) and so engaged to heparin, whereas when the interaction was performed on the heparin affinity mini column, this lysine was exposed to solvent, hence not involved in heparin binding [36]. The second difference is the status of K^{130}, K^{131} and K^{140}. Peptides 5 and 7 (Table 1A) are sisters, containing two adjacent lysine residues K^{130} and K^{131} on β-strand VII and K^{140} on β-strand VIII. However, whereas peptide 5 possessed the mass shift of three NHS-biotin products meaning that all three lysines were biotinylated and bound to heparin, in peptide 7 showed only two out of three lysine residues reacted with NHS-biotin and one reacted with NHS-acetate, hence was not bound to heparin. Moreover, in peptide 2 (Table 1A) K^{140} was biotinylated so this lysine was engaged to heparin. Importantly, K^{130}, K^{131} and K^{140} were not labelled when the interaction was performed on an affinity mini-column [36], indicating a significant difference in the dynamic of interactions depending on whether the polysaccharide was free in solution or linked to a chromatography support. It could be interpreted that an individual bond between protein and heparin was transient over the time of the protection reaction and its reaction with NHS-acetate prevented re-binding.

Other lysine residues, all assigned to HBS1 [36], were labelled, and so engaged to heparin: K^{123} and K^{126} on the loop of β-strand VI-VII (peptide 3, Table 1A; K^{155} on β-strand IX (peptide 1, Table 1A); the core five lysines of HBS-1, K^{170}, K^{177}, K^{178}, K^{180} and K^{181} (peptide 4, table 1A).

The binding state of K^{81} and K^{84} on the loop of β-strand II-III was not addressed here, since no peptide with those lysine residues in the sequence was identified. This is due to the use of chymotrypsin as protease on modified FGF7 resulting in peptides that are either too short or too long for identification by MALDI-TOF MS.

6.2.1.2. *The interaction of FGF7 with CS performed in PB buffer, 10 mM NaCl*

In DSF assays, the thermal stability of FGF7 was increased by CS than by DS and HS, which in turn were less effective than heparin [36]. This suggests that the binding of FGF7 to CS is weaker than to the other GAGs. Consistent with this interpretation, when the interaction between FGF7 and CS was conducted in PB buffer containing 50 mM NaCl, as was done for heparin (Section 6.2.1.1), all lysines were protected and none were labelled with biotin (Data not shown). This may have been due to the weaker interaction of FGF7 with CS, resulting in the individual bonds being transient over the time of the protection reaction. This has been seen previously, for example, with neuropilin-1 [157]. In the latter case, although the protein has a very high affinity for heparin (~ nM), and it bound to the heparin column, upon application of NHS-acetate it eluted. The interpretation was that individual bonds between protein and heparin were dynamic, and reaction with NHS-acetate prevented re-binding, resulting in the protein 'peeling off' the heparin. The solution in this instance was to reduce the concentration of electrolytes in the mobile phase. The same strategy was used here. FGF7 was, therefore, bound to CS in phosphate buffer containing 10 mM NaCl. Under these conditions not all lysine residues were protected. Following the addition of NaCl to 3 M to dissociate the sugar-protein

complex, the lysine residues previously bound to GAG were now exposed to solvent and reacted with NHS-biotin. The protein was then incubated with chymotrypsin to generate peptides, which were analysed by MALDI-TOF MS. Six peptides were identified when FGF7 was bound to CS in PB buffer containing 10 mM NaCl (Table 1B).

K^{111} on β-strand V reacted with NHS-biotin (peptide 4, Table 1B) and so was bound to CS, which was alike to its status in the heparin "in solution" work (Section 6.2.1.1). This is in contrast to the protection (acetylation of this residues when the reaction was performed on a heparin affinity mini column [36]. Peptide 6 (Table 1B) contains two adjacent lysine residues, K^{130} and K^{131} on β-strand VII, and K^{140} on β-strand VIII. This peptide had a mass shift corresponding to two biotins and one acetate, hence only two out of three lysine residues were bound to CS. Again, though these lysine residues were bound to heparin in solution, they were not when the analysis was performed on a mini-affinity column [36].

The HBS-1 lysine residues [36], K^{123} and K^{126} (peptides 1 and 3, Table 1B), K^{155} (peptide 2, table 1B) and those in the core of HBS-1, K^{170}, K^{177}, K^{178}, K^{180} and K^{181} were all labelled (peptide 6, Table 1B), as found for heparin in solution (6.2.1.1). As for heparin, (Section 6.2.1.1), K^{81} and K^{84} were missing from the analysis.

6.2.1.3. *The interaction of FGF7 with DS performed in PB buffer, 10 mM NaCl*

The list of the five peptides containing biotinylated lysines when FGF7 bound to DS in 10 mM NaCl, PB buffer is in Table 1C. There were differences from what had been observed when FGF7 was bound to a heparin affinity mini-column or in solution and when FGF7 bond CS in solution. Thus, three lysines K^{130}, K^{131} on β-strand VII and K^{140} on β-strand VIII were all biotinylated, since the mass shift of peptide 4 corresponded to that of three biotin products (Table 1C). In contrast only two of K^{130}, K^{131} and K^{140} were found to be biotinylated when FGF7 bound to heparin in solution (Section 6.2.1.1) or to CS (6.2.1.2), whereas when FGF7

bound to a heparin affinity column, these residues were not labelled. This observation indicated that FGF7 binds DS differently to heparin and CS.

The remaining labelled lysines, K^{123} and K^{126} (peptide 3, Table 1C), K^{155} (peptide 1, Table 1C), K^{170}, K^{177}, K^{178}, K^{180} and K^{181} (peptide 5, Table 1C), all in HBS1 [36] were consistent with those identified with CS and heparin in solution and heparin in an affinity column and, as for heparin and CS, K^{81} and K^{84} were not observed.

6.2.1.4. *Location of labelled lysine residues on the surface*

Surprisingly, the pattern of labeled lysines identified by the protect and label experiments with heparin, CS and DS in solution were more similar to each other (Section 6.2.1.1, 6.2.1.2 and 6.2.1.3) than to the experiments using a heparin affinity column [36]. Thus, K^{130}, K^{131} and K^{140} were only labelled when FGF7 bound heparin was in solution, and two of these three lyines were labelled when FGF7 bound CS or DS. It may be that these interactions were only detected because the concentration of electrolytes (50 mM NaCl for heparin and 10 mM NaCl for CS, DS) was below the physiologically relevant concentration (~140 mM) used in the experiments with a heparin affinity column. However, several lines of evidence suggest that the identified interactions are relevant. First, not all lysine amino acids were protected/labelled, indicating that binding remains selective at the lower concentrations of electrolytes. Second, there are lysine residues which were labelled when protein bound to heparin on a chromatography support, but were protected when the interactions were in solution, for example, K^{160} of FGF3 (Section 6.2.2) or three lysine residues, K^{183}, K^{191} and K^{195} in the core area of HBS-1 of FGF10 (Section 6.2.3). Third, whereas heparin and CS bound two of, K^{130}, K^{131} and K^{140}, DS bound all three. Fourth, the location of these lysine residues on the surface of FGF7 is consistent with the proposed sites of interaction, determined by previous work [36] and the arginine protect and label experiments (Section 4.2).

On the surface, K^{140} is adjacent to the core of HBS-1 (Figure 6.1). Thus, though distant in the sequence, K^{140} is on the surface a neighbour of K^{177} and K^{123} (Figure 6.1). The fact that heparin on the affinity column did not engage this residue may due to steric effects caused by the immobilisation of the polysaccharide. In addition, K^{111}, which was also only labelled when FGF7 bound polysaccharides in solution, is adjacent to R^{101} and R^{67} (Figure 6.1), which have been assigned to HBS-3 (Section 5.2). Next to K^{111} are K^{130} and K^{131}, which were similarly only labelled when FGF7 bound to GAG in solution (Sections 6.2.1.1, 6.2.1.2 and 6.2.1.3) (Figure 6.1). Importantly, the addition of these residues into the binding site connects HBS-3, which was isolated previously, to HBS-1 and HBS-4 (Section 5.2). This implies that HBS-3, defined when FGF7 was bound to heparin beads is in fact an extension of HBS-4 or HBS-1; because they are orthogonal on the surface, HBS-1 and HBS-1 remain as independent binding sites, that is they cannot bind a single polysaccharide chain.

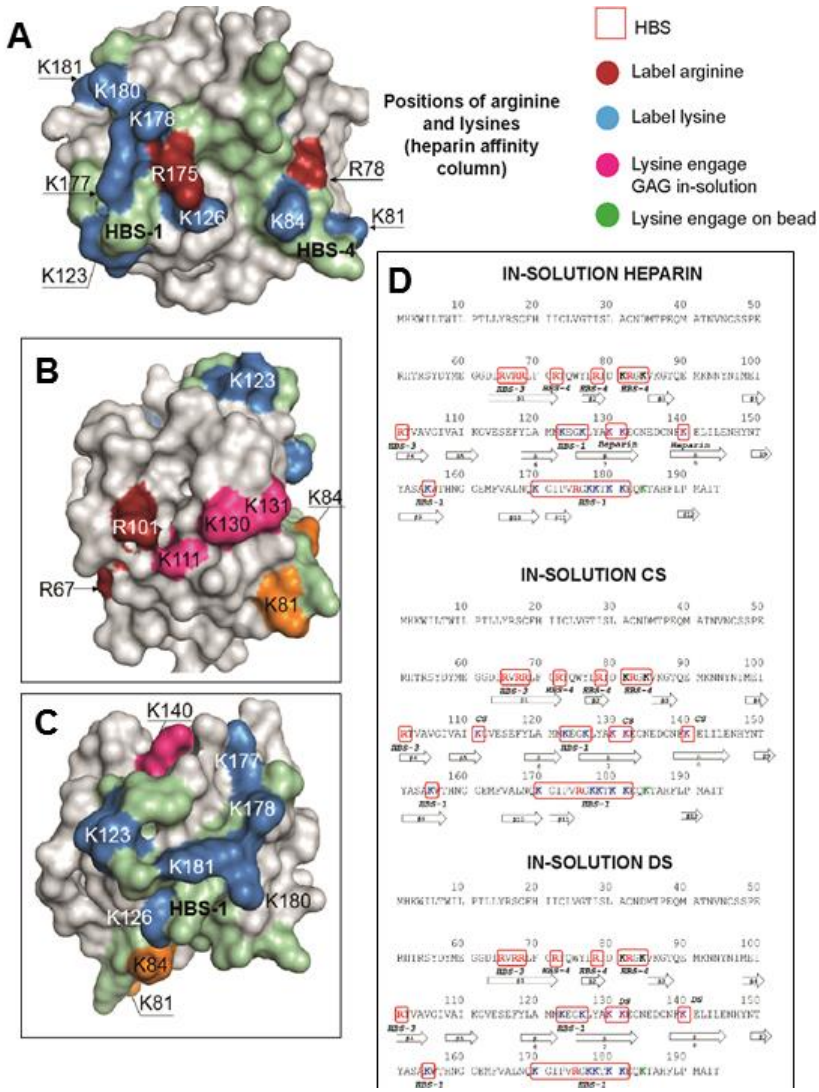


Figure 6.1: Labelled lysine residues in GAG-binding sites of FGF7. The FGF7 structure (PBD code **1qqk**) is shown as a surface. Locations of lysine residues engaging heparin/CS/DS on the sequence. FGF7 sequence, Uniprot code: P21781-1. **A)** K^{81}, K^{84}, K^{123}, K^{126}, R^{175}, K^{177}, K^{178}, K^{180}, K^{181} and K^{184} are indicated. Lysine residues which engage heparin beads on the affinity column are coloured as *blue*. **B and C)** Lysine residues bound to heparin/CS/DS in solution, but not heparin beads are coloured as *pink.* Positions of K^{111}, K^{130}, K^{131} and K^{140} relative to the core of HBS-1 on the surface. **D)** Positions of lysine residues involved in engagement of heparin/CS/DS in solution are indicated on the sequence.

6.2.2 FGF3

The lysine residues involving in the engagement of FGF3 to heparin have been identified by selective labelling on a heparin affinity mini-column, followed by mass spectrometry analysis [37]. They are K^{53} on the loop of β-strand II-III, K^{101} on the loop of β-strand VI-VII, K^{160} and K^{168} in the core of the canonical HBS-1. Here, the interactions between FGF3 and DS were determined using in-solution selective labelling, but not those with CS, because DSF showed that any interaction between FGF3 and CS is too weak to detect or has no effect on the thermal stability of the protein [37].

6.2.2.1 The interaction of FGF3 with heparin

Using the approach used with FGF7 (Section 6.2.1), FGF3 was bound heparin in phosphate buffer containing 50 mM NaCl and NHS-acetate was used for the protection of exposed lysine residues. Following the addition of NaCl to 3 M to dissociate the sugar-protein complex, NHS-biotin was added to react with the lysine residues previously bound to GAG that were now exposed to solvent. The list of the six peptides containing biotinylated lysine residues was provided in Table 2A.

Differences were observed between lysine residues bound to heparin in solution here and the ones identified previously using a heparin affinity column [37]. First, on β-strand I, K^{47} was found to be labelled by NHS-biotin (peptide 3, Table 2A) and so engaged to heparin. In contrast, this lysine reacted with sulfo-NHS-acetate when the interaction between FGF3 and heparin was conducted on a heparin mini-affinity-column [37]. The second difference was the status of K^{160} and K^{174}. Peptide 6 (Table 2A) contains three lysine residues of the canonical HBS-1 [37], K^{160}, K^{168} and K^{174}, showing the mass shift corresponding to two biotin reactions and one acetyl modification. Meanwhile, in peptide 2 (Table 2A) that has two lysine residues, K^{168} and K^{174}, and both were labelled. Consequently, it could be deduced that K^{160} reacted with

NHS-acetate and was not involved in the interaction with heparin. In contrast, when FGF3 bound to heparin on a chromatography support, K^{160} bound to heparin and K^{174} did not [37]. This change of binding status of K^{160} provides evidence that in 10 mM NaCl, phosphate buffer, heparin does not bind indiscriminately to basic residues and so indicates that the observed pattern of interaction is likely to reflect a true binding mode of FGF3 to heparin (Section 6.2.1.4).

The other biotinylated residues are identical to those identified previously using a heparin affinity column [37] is K^{53} on the loop between β-strands I and II (peptide 1, Table 1A), K^{101} on the loop between β-strands VI VII (peptides 4 and 5, Table 1A).

6.2.2.2 The interaction of FGF3 with DS performed in PB buffer 10 mM NaCl

When FGF3 bound to DS, seven peptides were found to contain biotinylated lysine residues (Table 2B). Similar to when FGF3 bound heparin in solution (Section 6.2.2.1), K^{47} on β-strand I was found to engage to DS (peptide 6, Table 2B). In contrast, K^{160} (Section 6.2.2.1 and [37]) was labelled by NHS-biotin (peptide 4, Table 2B), hence was engaged to DS in solution, whereas this residue was biotinylated when FGF3 bound heparin on an affinity column [37], but not heparin in solution (Section 6.2.2.1). Peptide 7 contains three lysine residues K^{160}, K^{168} and K^{174}, but only two were modified by biotin and the other by NHS-acetate (Table 2B). Because K^{160} was already identified as being labelled (peptide 4, Table 2B), along with the evidence from initial work on FGF3 on the affinity column [37], it is likely that K^{168} not K^{174} bound DS in solution. This conclusion is in accord with the labelling pattern of lysines when the interaction between FGF3 and heparin was on an affinity-column [37], but not in solution (Section 6.2.2.1).

K^{53} (peptide 1, Table 2B) and K^{101} (peptides 2, 3 and 5, Table 2B) behaved similarly to when FGF3 bound to heparin in both contexts, on the chromatographic support [37] and in solution (Section 6.2.2.1).

6.2.2.3 Location of labelled lysine residues in FGF3 on the surface

Although not involved in binding to heparin on beads, K^{47} engages to both heparin and DS in solution. On the surface this lysine is adjacent to K^{53} (Fig 6.2) which has been assigned to HBS-1 [37], indicating the potential for the sugar chain to crossover.

Distinct from its state when FGF3 bound to heparin on beads, K^{160} did not engage heparin in solution. This lysine was originally assigned to HBS-1. On the surface, this lysine is located at the end of HBS-1 (Fig 6.2). Hence, the lack of engagement of this lysine to heparin in solution may indicate a shorter heparin chain involved in binding FGF3.

On the other hand, K^{174} is labelled when FGF3 bound to heparin and DS in solution, but not on a heparin mini-column. On the surface, this lysine is nearby the arginine residues of HBS-1 and K^{168} (Fig 6.2). This observation implies the different binding preference of GAG on FGF3 in solution compared to that found on an affinity chromatography column.

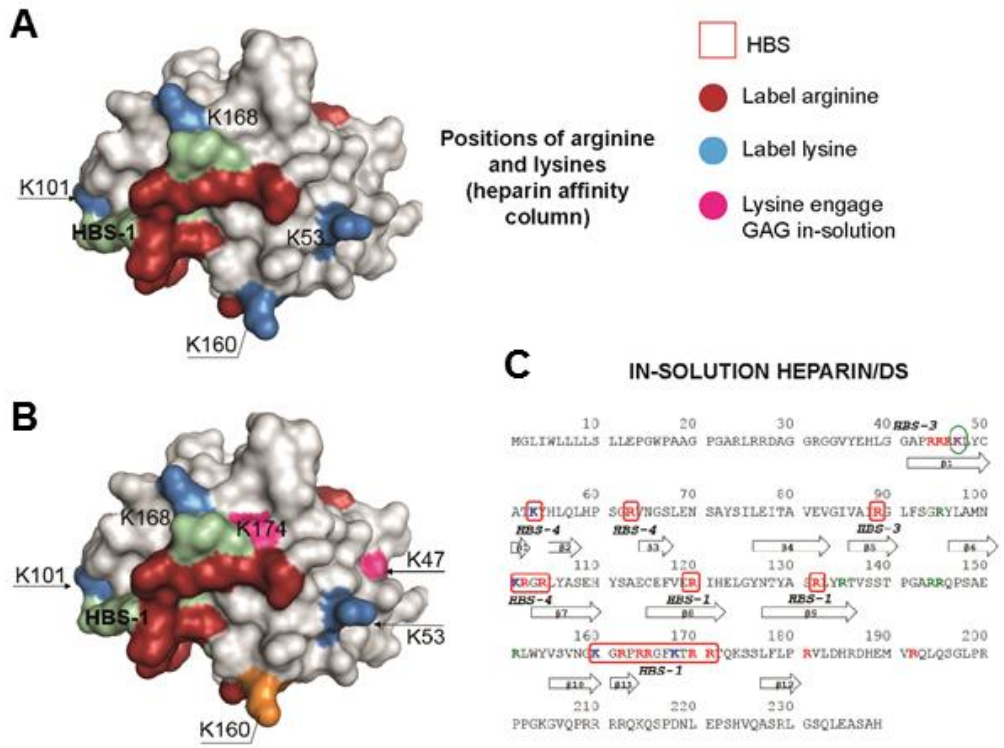


Figure 6.2: Labelled lysine residues in GAG-binding sites of FGF3. The model of FGF3 structure is constructed from the crystal structure of FGF7 (PBD code **1qqk**). Locations of lysine residues engaging heparin/CS/DS on the sequence. FGF3 sequence, Uniprot code: P11487-1. **A)** Lysine residues which engage heparin beads on the affinity column are coloured *blue*. K^{53}, K^{101}, K^{160} and K^{168} were indicated. **B)** Lysine residues bound to heparin/ DS in solution, but not heparin beads are coloured as *pink*. Positions of K^{47} and K^{174} relative to the core of HBS-1 on the surface. **C)** Positions of lysine residues involved in engagement of heparin/DS in solution are indicated on the sequence.

6.2.3 FGF10

The lysine residues involving in the engagement of FGF10 to heparin identified previously [37] are K^{81} on the loop between β-strands I-II; K^{86}, K^{94} and K^{97} on the loop between β-strands II-III; K^{136}, K^{137} and K^{139} on the loop between β-strands VI-VII; K^{183}, K^{191}, and K^{195} in the core of the canonical binding site. It has been showed that FGF10 interacts with both CS or DS [37].

The peptides identified as containing biotinylated lysine residues when FGF10 bound to heparin, CS and DS are in Tables 3A, 3B, 3C.

6.2.3.1 Lysine residues involving in the engagement of FGF10 to heparin

Using the same approach as for FGF7 (Section 6.2.1) and FGF3 (Section 6.2.2), FGF10 was bound heparin in phosphate buffer containing 50 mM NaCl. NHS-acetate and NHS-biotin were then used for the protection of exposed lysine residues and then labelling of lysines involved in binding, respectively. Five peptides containing lysine residues were identified when FGF10 engaged heparin in solution (Table 3A). The differences were observed between lysine residues bound to heparin in solution here and the ones identified previously [37].

A first difference is the status of K^{81}. Peptide 4 (Table 3A) contains two lysine residues K^{81} and K^{86}, but showed a mass shift corresponding to one modification of biotin and one of acetyl, indicating that one lysine was biotinylated and the other was acetylated. Since peptide 1 (Table 3A) had the mass shift of the NHS-biotin reaction product, K^{86} was bound to heparin in solution and K^{81} was acetylated. In contrast, this lysine was defined as part of HBS-3 when FGF10 bound to the heparin affinity column [37]. A second difference is shown by peptide 5 (Table 3A), which has a mass shift corresponding to two products of acetylation, indicating that two lysine residues K^{94} and K^{97} in its sequence were both acetylated. Hence similar to K^{81}, K^{94} and K^{97} were not involved in binding heparin in solution in contrast to when FGF10 bound to heparin in the context of a chromatography support [37].

The other lysine residues shown previously to be engaged to heparin [37] were also observed to be biotinylated when FGF10 bound heparin in solution: K^{136}, K^{137} and K^{139} on the loop between β-strands VI-VII (peptide 3, Table 3A); K^{183}, K^{191} and K^{195} in the core of the canonical HBS1 (peptide 2, Table 3A).

6.2.3.2 Lysine residues involving in the engagement of FGF10 to CS

When FGF10 bound CS in solution in phosphate buffer containing 10 mM NaCl six peptides containing lysine residues were identified as modified by biotin (Table 3B). There were differences observed between lysine residues bound to CS in solution and the ones bound to heparin in solution (Section 6.2.3.1) and on an affinity column [37].

In contrast to what was found when FGF10 bound to heparin in solution (Section 6.2.3.1), but similar to FGF10 bound to a heparin affinity column [37], K^{81} on the loop between β-strands I-II was engaged to CS, because it was biotinylated (peptide 2 and 4, Table 3B). Moreover, FGF10 bound to CS compared to FGF10 bound to heparin in solution and on beads had different labelling in the loop between β-strands VI-VII (K^{136}, K^{137} and K^{139}). Peptide 3 (Table 3B) contains these lysine residues and showed a mass shift corresponding to one biotin and two acetyl groups, so only one of these lysines bound to CS, whereas all three bound to heparin ([37] and Section 6.2.3.1). K^{153} which was not found to bind to heparin (Section 6.2.3.1 and [37]) was biotinylated (peptide 5, Table 3B), indicating that it was engaged to CS in solution. Finally, of three lysine residues, K^{183}, K^{191} and K^{195} in the core area of HBS-1, only one reacted with NHS-biotin and two were acetylated (peptide 6, Table 3B). This indicates a fundamental difference in how FGF10 engages CS compared to heparin. The only residue which retained its binding state compared to heparin (Section 6.2.3.1 and [37]) was K^{86},since peptide 1 showed a modification corresponding to biotinylation (peptide 1, Table 3B).

6.2.3.3 Lysine residues involved in the engagement of FGF10 to DS

FGF10 was then bound to DS in solution and following acetylation, lysine residues involved in binding DS were biotinylated. Six peptides containing biotinylated lysine residues were identified (Table 3C). Again there were differences to the data acquired with a heparin affinity column, and heparin and CS in solution. Interestingly, peptide 6 (Table 3C) showed a mass

shift of two NHS-biotin reaction products, indicating that both lysine residues bound to DS in solution. Peptide 6 encompasses the loop between β-strands III-IV and contains K^{102} and K^{103}, so this identifies a new area of binding sites on the FGF10 surface. As with CS, but different from heparin, K^{81} and K^{86} were found to interact with DS (peptides 1 and 4, Table 3C). Peptide 3 and 4 (table 3C) contains three lysine residues K^{136}, K^{137} and K^{139}, but two were acetylated and only one was biotinylated. Thus, only one out of these three lysines bound to DS, which was similar what was observed with CS (Section 6.2.3.2). Again, for the core of HBS-1, covered by peptide 5 (Table 3C), of K^{183}, K^{191} and K^{195}, one was biotinylated and the other two acetylated as seen with CS (Section 6.2.3.2)

6.2.3.4 Location of labelled lysine residues on the surface

When FGF10 bound to heparin in solution, K^{91}, K^{94} and K^{97}, which were previously assigned to HBS-4 (Section 5.2) were not labelled, indicating that heparin did not engage HBS-4 in solution and only bound to HBS-1 (Fig 6.3). The HBS-3 of FGF10 contains only arginine residues, hence, engagement of HBS-3 could not be addressed.

In the case of CS and DS in solution, one of three K^{136}, K^{137} and K^{139} was labelled, however, since MADLI-TOF only provides the information of the overall mass shift, it could not be concluded which lysine was involved in binding. The positions of these three lysines were shown on the surface (Fig 6.3). K^{139} is adjacent to K^{97}, which was acetylated, so it is less likely to be involved in binding. K^{137} is closer to R^{188} of HBS-1 than K^{136} (Fig 6.3). These observations implied that K^{137} was involved in GAG binding.

K^{153} only bound to CS and this lysine neighbours to R^{155} of HBS-1 (Fig 6.3). K^{102} and K^{103} only bound to DS, but not heparin or CS in solution. These lysine residues are adjacent to K^{86} which was also biotinylated. K^{86} was assigned by the previous work on heparin on beads to

HBS-4 [37]. The addition of K^{102} and K^{103} in fact extended the area of HBS-4, making the T-shape formed by HBS-4 and HBS-1 become larger when protein bound to DS.

Of the three lysine residues of the core of HBS-1, K^{183}, K^{191} and K^{195}, only one was engaged to CS/DS. On the surface, K^{183} is closer to the group of labelled residues including K^{86}, K^{102} and K^{103} so maybe this residue is the one which was biotinylated. The acetylation of two lysine residues in the core of HBS-1 when FGF10 bound to CS/DS in solution indicates a significant difference in how the protein engages these GAGs compared to heparin.

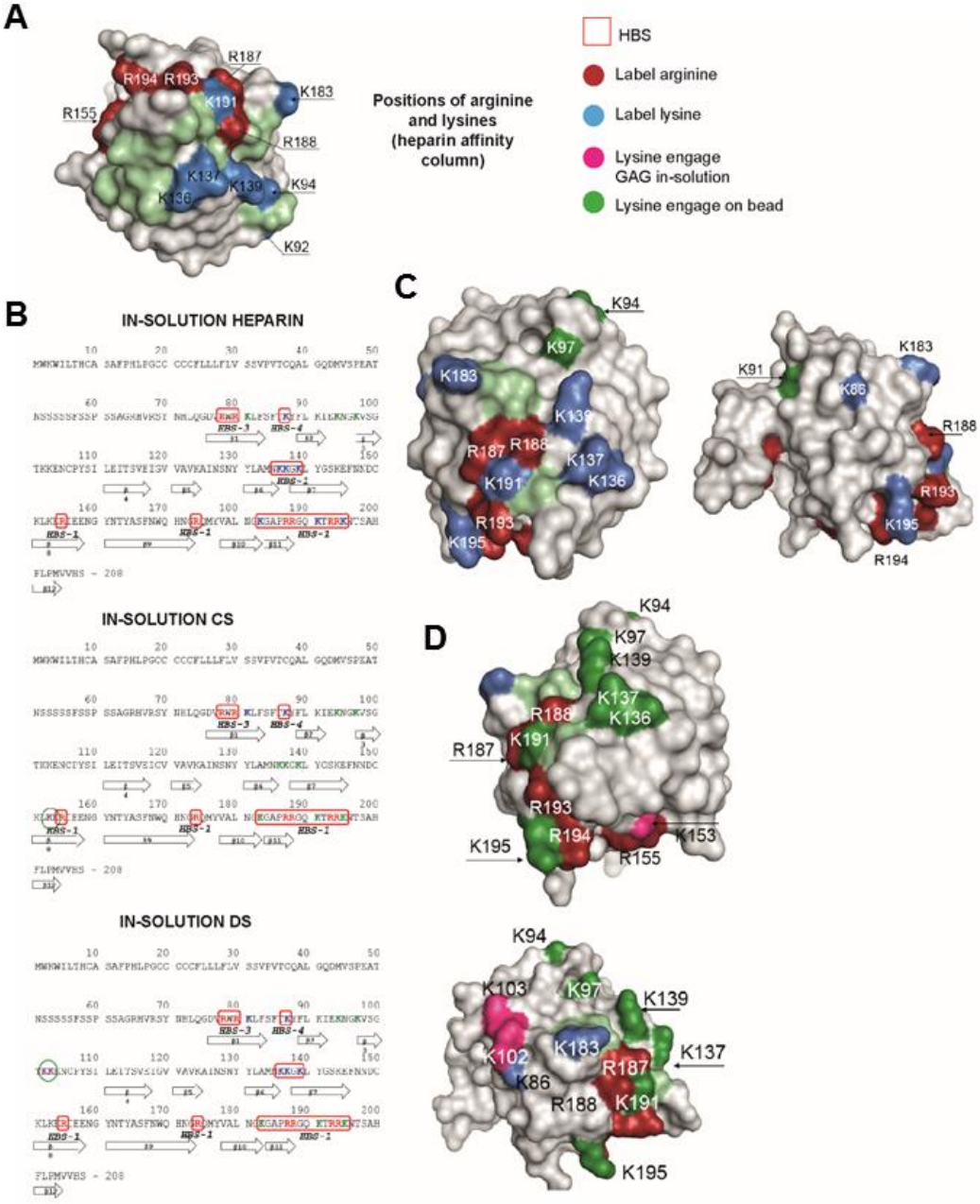


Figure 6.3: Labelled lysine residues in GAG-binding sites of FGF10. The FGF10 structure (PBD code **1nun**) is shown as a surface. Locations of lysine residues engaging heparin/CS/DS on the sequence. FGF10 sequence, Uniprot code: O15520-1. **A)** The information on arginine residues which bound to heparin on the chromatography support (Section 5.2) is also presented. Lysine residues which engage heparin beads on the affinity column are coloured *blue*. **B)** Positions of lysine residues involved in engagement of heparin/CS/DS in solution are indicated

on the sequence. **C and D)** Lysine residues bound to heparin/CS/DS in solution, but not heparin beads are coloured *pink.* Lysine residues engaging heparin, but not bound to heparin/CS/DS in solution are coloured *green.*

6.3 DISCUSSION

The selective labelling in solution opens a new perspective of studying protein - GAG interactions. FGF3, FGF7 and FGF10 all belong to the FGF7 sub-family, and, as demonstrated by previous work on selective labelling of arginine and lysine [37] [36] (Section 5.2) using an affinity column, they share a similar architecture of heparin binding sites, consisting of three independent binding sites, HBS-1, HBS-3 and HBS-4,. However, their interactions with different GAGs in solution are distinct. For FGF7, K^{130}, K^{131} on β-strand VII and K^{140} on β-strand VIII bind to DS in solution, but not to heparin on a chromatography support [36]. This area of binding was not detected when FGF3 or FGF10 bound to GAG in solution. In addition, for FGF3 and FGF7, the lysine residues in the core of HBS-1 also bound GAGs in solution (Section 6.2.1 and Section 6.2.2), however, this was not the case for FGF10, in which only one of these three lysine residues in the core of HBS-1 was biotinylated (Section 6.2.3). In FGF10 K^{102} and K^{103} bound to DS, though not to CS or heparin. These residues are located on the loop between β-strands III-IV, an area that has not been implicated in GAG binding in other members of FGF7 subfamily.

The addition of lysine residues when a FGF of FGF7 subfamily bound to GAG in solution sometimes extended the region of HBS-4, making the T-shaped basic patch formed by HBS-1 and HBS-4 become large, for example, K^{102} and K^{103} of FGF10 when this protein bound to DS in solution (Section 6.2.3). In contrast, the absence of some lysines in binding may alter the binding pattern of GAG to the protein, for instance, K^{160} of FGF3 when it interacts with DS in solution (Section 6.2.2) which seems to shorten the length of binding pattern on HBS-1.

The addition of K^{111}, K^{130} and K^{131} when FGF7 bound to GAGs irrespectively in solution raised an interesting point that the HBS3 and the rest of positively charged surface containing HBS-1 and HBS-4 are connected. Hence, there is no acidic border between them. As raised before a sugar chain cannot bend over to reach HBS-1 and HBS-4 which are orthogonal to each other. Hence, K^{111}, K^{130} and K^{131} should contribute to either HBS-1 or HBS-4. This conclusion may require further analysis, especially the electrostatic potential of FGF7 surface.

The selective labelling of lysines in solution with different GAGs makes it possible to study the interactions of proteins with any GAG, or indeed acidic partner. However, the apparent requirement of a reduced concentration of electrolytes, is a concern, since it may result in artefactual interactions being identified. However, as argued (Section 6.2.1), the fact that the main observation is a shift in labelled lysine residues, rather than simply an increase in these, strongly suggests that any such artefacts are limited. There is, nevertheless, scope for a more detailed examination of the effects of electrolytes and of GAG chain immobilisation on binding of proteins.

One such question to tackle is the effect of reducing electrolytes which may be more complex than simply increasing electrostatic bonds. Regarding that, one major conclusion is that these FGFs have an "adaptable" surface which is able to accommodate different GAGs that present charged groups on different spatial dispositions. This would reflect a functional requirement for such adaptability. These observations may indicate that the evolutionary selectivity impacts differently on the interactions of FGF family and GAGs. Interactions between members of FGF7 family to heparin may be highly conserved but not in relation to other GAGs. This demonstrated the difference in selectivity of FGF10 toward DS compared to CS and heparin which may be underlined by the structural features of DS.

Table 6.1A: FGF7 interaction with heparin, MS analysis. The reference number (Ref.no.) of the peptide is column 1. The two columns under "Native FGF7" present the predicted m/z and sequence of each peptide after digestion of FGF7 by chymotrypsin. The bold lysines in the 3rd column ("Sequence") are those identified as labelled with NHS-biotin. The first two columns under "FGF7 heparin" present the predicted and observed m/z for FGF7 after in-solution selective labelling. FGF7 bound to heparin (protection step) was reacted with NHS-acetate, following dissociation, then lysine residues of binding sites exposed to solvent were reacted with NHS-biotin (labelling step). The 3rd column identifies the peptide as being considered from sequence alignment in the literature [102] to be in an HBS of FGF7. The final column indicates the modification occurring on the lysine residues of peptides.

	Native FGF7		FGF7 heparin			
Ref. no.	**Predicted native peptide m/z**	**Sequence**	**Predicted modified peptide m/z**	**m/z observed**	**Match HBS**	**Modifications**
1	578.29	^{152}ASA**K**W^{156}	810.3033	810.1159	HBS-1 K^{155}	1K + 232 (biotin)
2	615.4076	$^{\mathbf{140}}$**K**ELIL^{144}	847.4176	847.3779	K^{140}	1K + 232 (biotin)
3	1329.69	^{118}YLAMN**K**EG**K**LY^{128}	1793.71	1793.97	HBS-1 K^{123} and K^{126}	2K + 232 (biotin) x 2
4	2423.40	^{168}NQ**K**GIPVRG**KK**T**KK**EQ**K**TAHF^{188}	3623.90	3623.81	HBS-1 K^{170}, R^{175}, K^{177}, K^{178}, K^{180}, K^{181} and K^{184}	6K + 232 (biotin) x 5 + 42 (acetyl)
5	2603.1905	^{128}YA**KK**ECNEDCNF**K**ELILENHY^{145}	3299.9260	3299.2205	K^{130} and K^{131}; K^{140}	3K + 232 (biotin) x 3
6	2667.44	^{96}NIMEIRTVAVGIVAI**K**GVESEFYL^{119}	2900.45	2900.38	K^{111}	1K + 232 (biotin)
7	2981.34	^{128}YA**KK**ECNEDCNF**K**ELILENHYNTY^{148}	3487.37	3486.99	K^{130} and K^{131}; K^{140}	3K + 232 (biotin) x 2 + 42 (acetyl)

Table 6.1B: FGF7 interaction with CS, MS analysis. The reference number (Ref.no.) of the peptide is column 1. The two columns under "Native FGF7" present the predicted m/z and sequence of each peptide after digestion of FGF7 by chymotrypsin. The bold lysines in the 3rd column ("Sequence") are those identified as labelled with NHS-biotin. The first two columns under "FGF7 CS" present the predicted and observed m/z for FGF7 after in-solution selective labelling. FGF7 bound to CS (protection step) was reacted with NHS-acetate, following dissociation, then lysine residues of binding sites exposed to solvent were reacted with NHS-biotin (labelling step). The 3rd column identifies the peptide as being considered from sequence alignment in the literature [102] to be in an HBS of FGF7. The final column indicates the modification occurring on the lysine residues of peptides.

	Native FGF7		FGF7 CS			
Ref. no.	**Predicted native peptide m/z**	**Sequence**	**Predicted modified peptide m/z**	**m/z observed**	**Match HBS**	**Modifications**
1	1069.53	^{120}AMN**K**EG**K**LY^{128}	1533.75	1533.751	HBS-1 K^{123} and K^{126}	2K + 232 (biotin) x 2
2	1499.66	^{145}ENHYNTYASA**K**W^{156}	1733.67	1733.98	HBS-1 K^{155}	1K + 232 (biotin)
3	1329.69	^{118}YLAMN**K**EG**K**LY^{128}	1793.71	1793.97	HBS-1 K^{123} and K^{126}	2K + 232 (biotin) x 2
4	2667.44	^{96}NIMEIRTVAVGIVAI**K**GVESEFYL^{119}	2900.45	2900.38	K^{111}	1K + 232 (biotin)
5	2980.34	^{128}YA**KK**ECNEDCNF**K**ELILENHY^{145}	3485.37	3485.86	K^{130} and K^{131}; K^{140}	3K + 232 (biotin) x 2+ 42 (acetyl)
6	2423.40	^{168}NQ**K**GIPVRG**KK**T**KK**EQ**K**TAHF^{188}	3623.90	3623.81	HBS-1 K^{170}, R^{175}, K^{177}, K^{178}, K^{180}, K^{181} and K^{184}	6K + 232 (biotin) x 5 + 42 (acetyl)

Table 6.1C: FGF7 interaction with DS, MS analysis. The reference number (Ref.no.) of the peptide is column 1. The two columns under "Native FGF7" present the predicted m/z and sequence of each peptide after digestion of FGF7 by chymotrypsin. The bold lysines in the 3rd column ("Sequence") are those identified as labelled with NHS-biotin. The first two columns under "FGF7 DS" present the predicted and observed m/z for FGF7 after in-solution selective labelling. FGF7 bound to DS (protection step) was reacted with NHS-acetate, following dissociation, then lysine residues of binding sites exposed to solvent were reacted with NHS-biotin (labelling step). The 3rd column identifies the peptide as being considered from sequence alignment in the literature [102] to be in an HBS of FGF7. The final column indicates the modification occurring on the lysine residues of peptides.

	Native FGF7		FGF7 DS			
Ref.no.	**Predicted native peptide m/z**	**Sequence**	**Predicted modified peptide m/z**	**m/z observed**	**Match HBS**	**Modifications**
1	578.29	^{152}ASA**K**W^{156}	810.3033	810.1159	HBS-1 K^{155}	1K + 232 (biotin)
2	615.4076	$^{\mathbf{140}}$**K**ELIL^{144}	847.4176	847.3779	K^{140}	1K + 232 (biotin)
3	1019.56	^{119}LAMN**K**EG**K**L^{127}	1499.58	1499.60	HBS-1 K^{123} and K^{126}	2K + 232 (biotin) x 2 and oxidation
4	2059.98	^{128}YA**KK**ECNEDCNF**K**ELIL^{142}	2756.01	2755.87	K^{130} and K^{131}; K^{140}	3K + 232 (biotin) x 3
5	2423.40	^{168}NQ**K**GIPVRG**KK**T**KK**EQ**K**TAHF^{188}	3623.90	3623.81	HBS-1 K^{170}, R^{175}, K^{177}, K^{178}, K^{180}, K^{181} and K^{184}	6K + 232 (biotin) x 5 + 42 (acetyl)

Table 6.2A: FGF3 interaction with heparin, MS analysis. The reference number (Ref.no.) of the peptide is column 1. The two columns under "Native FGF3" present the predicted m/z and sequence of each peptide after digestion of FGF3 by chymotrypsin. The bold lysines in the 3rd column ("Sequence") are those identified as labelled with NHS-biotin. The first two columns under "FGF3 heparin" present the predicted and observed m/z for FGF3 after in-solution selective labelling. FGF3 bound to heparin (protection step) was reacted with NHS-acetate, following dissociation, then lysine residues of binding sites exposed to solvent were reacted with NHS-biotin (labelling step). The 3rd column identifies the peptide as being considered from sequence alignment in the literature [102] to be in an HBS of FGF3. The final column indicates the modification occurring on the lysine residues of peptides.

	Native FGF3		**FGF3 heparin**			
Ref .no.	**Predicted native peptide m/z**	**Sequence**	**Predicted modified peptide m/z**	**m/z observed**	**Match HBS**	**Modifications**
1	835.41	^{50}CAT**K**YHL^{56}	1067.42	1067.58	HBS1 K^{53}	1K + 232 **(biotin)**
2	1204.71	$^{\mathbf{168}}$**K**TRRTQ**K**SSL^{177}	1668.73	1668.92	HBS1 K^{168} and K^{174}	2K + 232 **(biotin) x 2**
3	1173.70	^{40}GGAPRRR**K**L^{48}	1405.71	1405.78	HBS3 K^{47}	1K + 232 **(biotin)**
4	1668.91	^{91}FSGRYLAMN**K**RGRL^{105}	1900.92	1900.88	HBS1 K^{101}	1K + 232 **(biotin)**
5	1700.90	^{93}SGRYLAMN**K**RGRLY^{106}	1932.91	1933.03	HBS1 K^{101}	1K + 232 **(biotin)**
6	2762.58	^{155}VSVNG**K**GRPRRGF**K**TRRTQ**K**SSLF^{172}	3268.61	3268.51	HBS1 K^{160}, K^{168} and K^{174}	3K + 232 **(biotin) x 2 +** 42 (acetyl)

Table 6.2B: FGF3 interaction with DS, MS analysis. The reference number (Ref.no.) of the peptide is column 1. The two columns under "Native FGF3" present the predicted m/z and sequence of each peptide after digestion of FGF3 by chymotrypsin. The bold lysines in the 3rd column ("Sequence") are those identified as labelled with NHS-biotin. The first two columns under "FGF3 DS" present the predicted and observed m/z for FGF3 after in-solution selective labelling. FGF3 bound to DS (protection step) was reacted with NHS-acetate, following dissociation, then lysine residues of binding sites exposed to solvent were reacted with NHS-biotin (labelling step). The 3rd column identifies the peptide as being considered from sequence alignment in the literature [102] to be in an HBS of FGF3. The final column indicates the modification occurring on the lysine residues of peptides.

	Native FGF3		FGF3 DS			
Ref. no.	**Predicted native peptide m/z**	**Sequence**	**Predicted modified peptide m/z**	**m/z observed**	**Match HBS**	**Modifications**
1	998.48	^{49}YCAT**K**YHL^{56}	1230.49	1230.40	HBS1 K^{53}	1K + 232 **(biotin)**
2	1108.60	^{98}AMN**K**RGRLY^{106}	1340.61	1340.74	HBS1 K^{101}	1K + 232 **(biotin)**
3	1521.84	^{93}SGRYLAMN**K**RGRL^{105}	1753.85	1753.77	HBS1 K^{101}	1K + 232 **(biotin)**
4	1592.88	^{154}YVSVNG**K**GRPRRGF^{167}	1866.90	1866.96	HBS1 K^{160}	2K + 232 **(biotin)**
5	1700.90	^{93}SGRYLAMN**K**RGRLY^{106}	1932.91	1933.03	HBS1 K^{101}	1K + 232 **(biotin)**
6	2534.40	^{26}RRDAGGRGGVYEHLGGAPRRR**K**L^{48}	2766.41	2766.35	HBS3 K^{47}	1K + 232 **(biotin)**
7	2762.58	^{155}VSVNG**K**GRPRRGF**K**TRRTQ**K**SSLF^{172}	3268.61	3268.51	HBS1 K^{160}, K^{168} and K^{174}	3K + 232 **(biotin) x 2 +** 42 (acetyl)

Table 6.3A: FGF10 interaction with heparin, MS analysis. The reference number (Ref.no.) of the peptide is column 1. The two columns under "Native FGF10" present the predicted m/z and sequence of each peptide after digestion of FGF10 by chymotrypsin. The bold lysines in the 3rd column ("Sequence") are those identified as labelled with NHS-biotin. The first two columns under "FGF10 heparin" present the predicted and observed m/z for FGF10 after in-solution selective labelling. FGF10 bound to heparin (protection step) was reacted with NHS-acetate, following dissociation, then lysine residues of binding sites exposed to solvent were reacted with NHS-biotin (labelling step). The 3rd column identifies the peptide as being considered from sequence alignment in the literature [102] to be in an HBS of FGF10. The final column indicates the modification occurring on the lysine residues of peptides.

	Native FGF10		FGF10 heparin			
Ref. no.	**Predicted native peptide m/z**	**Sequence**	**Predicted modified peptide m/z**	**m/z observed**	**Match HBS**	**Modifications**
1	939.46	^{83}FSFT**K**YF^{89}	1171.471	1171.612	K^{86}	1K + 232 **(biotin)**
2	2650.487	^{178}VALNG**K**GAPRRGQ**K**TRR**K**NTSAHF^{201}	3349.52	3349.63	HBS-1 K^{183}, K^{191} and K^{195}	3K + 232 **(biotin) x 3**
3	1181.67	^{131}YLAMN**KK**G**K**L^{140}	1889.83	1889.99	HBS-1 K^{136}, K^{137} and K^{139}	3K + 232 **(biotin) x 3 and oxidation**
4	1336.741	^{80}R**K**LFSFT**K**YF^{89}	1798.761	1798.89	K^{81} and K^{86}	2K + 232 **(biotin) x 2**
5	2596.42	^{89}FL**K**IE**K**NG**K**VSGT**KK**ENCPYSIL^{101}	2806.42	2806.40	K^{91}, K^{94}, K^{97}, K^{102} and K^{103}	5K **+ 42 (acetyl) x 5**
6	2373.11	^{142}GS**K**EFNNDC**K**L**K**ERIEENGY^{161}	2647.34	2647.35	K^{144}, K^{151} and K^{153}	3K + 232 **(biotin) + 42 (acetyl)**

Table 6.3B: FGF10 interaction with CS, MS analysis. The reference number (Ref.no.) of the peptide is column 1. The two columns under "Native FGF10" present the predicted m/z and sequence of each peptide after digestion of FGF10 by chymotrypsin. The bold lysines in the 3rd column ("Sequence") are those identified as labelled with NHS-biotin. The first two columns under "FGF10 CS" present the predicted and observed m/z for FGF10 after in-solution selective labelling. FGF10 bound to CS (protection step) was reacted with NHS-acetate, following dissociation, then lysine residues of binding sites exposed to solvent were reacted with NHS-biotin (labelling step). The 3rd column identifies the peptide as being considered from sequence alignment in the literature [102] to be in an HBS of FGF10. The final column indicates the modification occurring on the lysine residues of peptides.

	Native FGF10		FGF10 CS			
Ref.no.	**Predicted native peptide m/z**	**Sequence**	**Predicted modified peptide m/z**	**m/z observed**	**Match HBS**	**Modifications**
1	905.48	^{84}SFT**K**YFL^{90}	1137.49	1137.41	K^{86}	1K + 232 **(biotin)**
2	1903.0086	^{71}NHLQGDVRWR**K**LFSF^{85}	2135.0186	2134.927	K^{81}	1K + 232 **(biotin)**
3	1068.59	^{133}AMN**KK**G**K**LY^{141}	1384.62	1384.77	K^{136}, K^{137} and K^{139}	3K + 232 **(biotin) +** 42 (acetyl) x 2
4	1189.67	^{80}R**K**LFSFT**K**Y^{88}	1653.69	1653.86	K^{81} and K^{86}	2K + 232 **(biotin) x 2**
5	1820.85	$^{\mathbf{153}}$**K**ERIEENGYNTYASF^{167}	2052.86	2052.95	**K^{153}**	1K + 232 **(biotin)**
6	2650.487	^{178}VALNG**K**GAPRRGQ**K**TRR**K**NTSAHF^{201}	2966.51	2966.63	K^{183}, K^{191} and K^{195}	3K + 232 **(biotin) +** 42 (acetyl) x 2
7	2596.42	^{89}FL**K**IE**K**NG**K**VSGT**KK**ENCPYSIL^{101}	2806.42	2806.40	K^{91}, K^{94}, K^{97}, K^{102} and K^{103}	5K **+ 42 (acetyl) x 5**

Table 6.3C: FGF10 interaction with DS, MS analysis. The reference number (Ref.no.) of the peptide is column 1. The two columns under "Native FGF10" present the predicted m/z and sequence of each peptide after digestion of FGF10 by chymotrypsin. The bold lysines in the 3rd column ("Sequence") are those identified as labelled with NHS-biotin. The first two columns under "FGF10 DS" present the predicted and observed m/z for FGF10 after in-solution selective labelling. FGF10 bound to DS (protection step) was reacted with NHS-acetate, following dissociation, then lysine residues of binding sites exposed to solvent were reacted with NHS-biotin (labelling step). The 3rd column identifies the peptide as being considered from sequence alignment in the literature [102] to be in an HBS of FGF10. The final column indicates the modification occurring on the lysine residues of peptides.

	Native FGF10		**FGF10 DS**			
Ref. no.	**Predicted native peptide m/z**	**Sequence**	**Predicted modified peptide m/z**	**m/z observed**	**Match HBS**	**Modifications**
1	905.48	^{84}SFT**K**YFL^{90}	1137.49	1137.41	K^{86}	1K + 232 **(biotin)**
2	1018.61	^{132}LAMN**KK**G**K**L^{140}	1524.64	1524.80	K^{136}, K^{137} and K^{139}	3K + 232 **(biotin) x 2** + 42 (acetyl)
3	1181.67	^{131}YLAMN**KK**G**K**L^{140}	1497.70	1497.73	K^{136}, K^{137} and K^{139}	3K + 232 **(biotin)** + 42 (acetyl) x 2
4	1189.67	^{80}R**K**LFSFT**K**Y^{88}	1653.69	1653.74	K^{81} and K^{86}	2K + 232 **(biotin) x 2**
5	2650.487	^{178}VALNG**K**GAPRRGQ**K**TRR**K**NTSAHF^{201}	2966.51	2966.63	K^{183}, K^{191} and K^{195}	3K + 232 **(biotin)** + 42 (acetyl) x 2
6	2203.01	^{98}NNDC**K**L**K**ERIEENGYNTY^{114}	2667.03	2666.26	K^{102} and K^{103}	2K + 232 **(biotin) x 2**

CHAPTER 7 SELETIVE LABELLING OF LYSINE AND ARGININE RESIDUES FOR NON- FIBROBLAST GROWTH FACTORS PROTEIN-HS4C3 ANTIBODY

7.1 INTRODUCTION

There are over 800 HBPs identified [139] in the context of pancreas, and this number will keep increasing. So far, selective labelling has been applied for a limited number of HBPs in which FGFs were the centre of the interest [120][218] [102]. However, ideally, selective labelling could be able to apply to any electrostatic interactions with an appropriate model. Moreover, it will be useful when there is difficulty in crystalizing the protein. Hence, a question raised then is the capability of the selective labelling of lysine and arginine residues on HBPs rather than FGFs. Moving a step forward from the well-known area of FGFs and HS, among many of proteins having affinity with HS, we got HS4C3 antibody, as gift from Prof Toin H. van Kuppevelt (Netherland) to examine the applicability of the selective labelling.

Single chain variable fragment (scFv) antibodies are synthetic antibodies which adopt the structures of human ones with a heavy and light chain variable regions but being randomly assembled in vitro for certain purpose [188]. HS4C3 is one of them which displays antibodies selected against bovine kidney HS. It binds to HS through a sequence, GRRLKD, which so-called CDR3 region. This sequence of HS4C3 fits into the proposed consensus sequence of glycosaminoglycan binding site XBBXBX (B, basic amino acid residue; X, any amino acid residue) [110] (Section 1.4.3). Moreover, it has been shown that this sequence of HS4C3 is selective to 3-*O*-sulfation in HS [222]. A study measured the binding affinity between the antibody and hexa- to octasaccharide fragment and found that HS4C3 showed weak binding with the sugar fragments with any *N*-sulfated, 2-*O*-and 6-*O*-sulfated but in contrast, it interacted strongly to the oligosaccharide having 3-*O*-sulfated glucosamine residue in the sequence [189].

The previous bodies of work on HS4C3 over-focused on the CDR3 region with little knowledge of the rest of the engagement. Bear in mind that CDR3 is composed of only two arginine residues and one lysine, hence, the length of polysaccharide chain bound to this area is just a trisaccharide or a tetrasaccharide, whereas the length of HS is much longer. From this point of view, the selective labelling of lysine and arginine were applied on HS4C3 to investigate further the binding pattern of this antibody and HS.

7.2 RESULTS

7.2.1 Optimization of the conditions for protect and label on heparin mini-column

The selective labelling of lysine residues on HS4C3 was first conducted following the standard protocol using 150 mM NaCl, PB buffer and Sulfo-NHS acetate as reagent (Figure 7.1A). When antibody (Figure 7.1A - Loading) was loaded onto the heparin mini-column, there were several bands observed in the flow through (FT) and wash fractions (Wash) (Figure 7.1A - LF/T and Wash), however, none of them was HS4C3 indicating that HS4C3 bound as expected (FT and W lanes). After the on-column reaction with 50 mM sulfo-NHS-acetate, the excess reagent was removed (Figure 7.1A NHS F/T) and the column was washed by Na-1 buffer (Figure 7.1A NHS W) before elution with Na-2 buffer (Figure 7.1A Elute). In all fractions, the bands with equivalent molecular weight to HS4C3 were observed indicating the partial elution of the antibody during the reaction. The use of sulfo-NHS-acetate with sulfo groups may cause competition in the ionic binding to heparin, resulting in the dissociation of engaged residues from heparin, their reaction with NHS acetate and consequently the partial elution of HS4C3.

In the next test, we changed the reagent to NHS acetate and repeated the experiment. A large band was observed in the flow through (FT, Figure 7.1B- F/T), however, it was of lower molecular size lower than HS4C3, showing that antibody bound to the heparin column. In the following factions before Elute, there was no band observed. After the antibody was eluted by 2 M NaCl, in elute fraction (Figure 7.1B- Elute), a band with the same molecular weight as

HS4C3 appeared, indicating that under these conditions lysines on the antibody do not dissociate during the reaction and so the antibody remains bound to the column. Note that by changing reactions times and electrolyte concentrations it would be possible potentially to access the local kinetics of interactions of individual residues. This may be of interest in terms of engineering the k_d of the antibody, since reducing the dissociation rate constant will increase the efficacy of washes and so the stringency of immunofluorescence on cells and tissues.

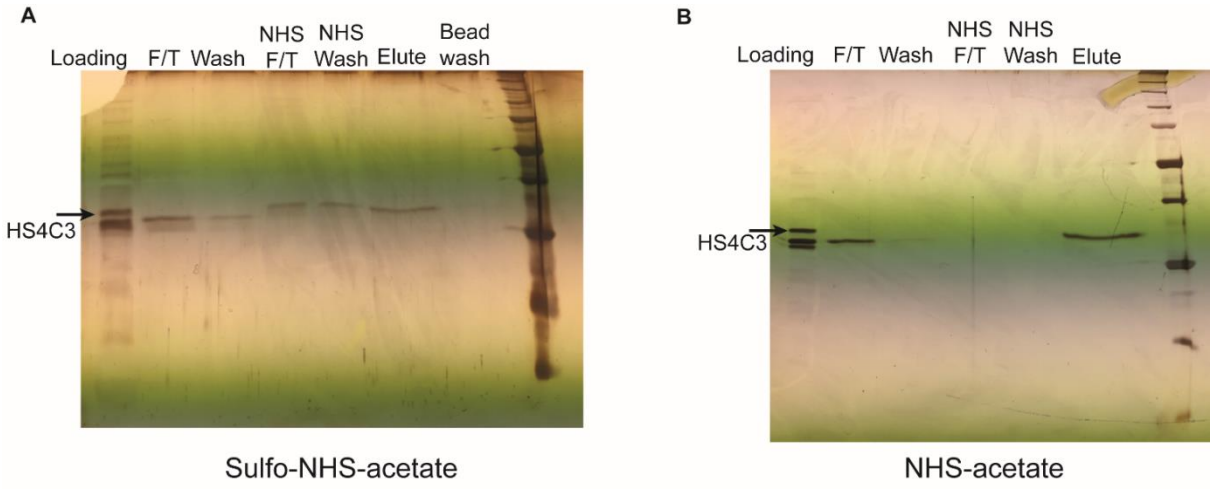


Figure 7.1: Testing conditions for the selective labelling of lysine residues on HS4C3. A) Sulfo-NHS-acetate was used for the protection of exposed lysine residues. Lane loading, the raw material before selective labelling, lane F/T: Flow through fraction after HS4C3 was loaded onto the heparin mini-column, lane wash: wash fraction after Na-1 buffer was loaded, lane NHS/FT: the fraction after PB buffer containing 50 mM sulfo-NHS-acetate was loaded onto the mini-column for the protection, lane NHS wash: wash fraction after Na-1 buffer was loaded, lane Elute: Elute fraction after PB buffer containing 2 M NaCl was loaded onto the mini-column, lane Bead wash: the heparin bead was mixed with SDS and boiled at 96 ºC. **B) NHS-acetate (no sulfo) was used for the protection of exposed lysine residues.** Lane loading, the raw material before selective labelling, lane F/T: Flow through fraction after HS4C3 was loaded onto the heparin mini-column, lane wash: wash fraction after Na-1 buffer was loaded, lane NHS/FT: the fraction after PB buffer containing 50 mM NHS-acetate was

loaded onto the mini-column for the protection, lane NHS wash: wash fraction after Na-1 buffer was loaded, lane Elute: Elute fraction after PB buffer containing 2 M NaCl was loaded onto the mini-column.

7.2.2 Protect and Label strategy for the identification of arginine/lysine in proteins that engage heparin (Section Method 2.7, 2.8 and 2.9)

The peptides generated from native HS4C3 by cleavage with chymotrypsin were predicted by Prospector in the mass range from 500 Da to 4000 Da, with a maximum missed cleavages of 5. The predicted lists were then filtered with the script *PGO-HPG mass predictor* to remove peptides without arginine residues. *PGO-HPG mass predictor* then predicted the mass shifts arising from the reaction of arginine side chains (Table 7.2) and added these to the mass of each peptide. This additional processing step is necessary, because there are three possible reactions products with PGO. All combinations of mass changes were covered. The predicted lists were then matched to lists of observed masses using the *Matchmaker* script with the difference between two peptides set at 0.1-0.5 Da (Section 2.9). The resulting peptides with information about modification, sequence and final m/z are presented in Tables 7.2.

All the matching peaks are present in Supplementary Figure 7.1.

The same process was conducted to lysine with corresponding mass shift for NHS-acetate and NHS-biotin. The resulting peptides with information about modification, sequence and final m/z are presented in Tables 7.1.

7.2.3 Locations of modified arginine residues in the sequence and structure.

The CDR3 region of HS4C3 is indicated by the red box, including three arginine residues and one lysine. All of them are detected by the selective labelling method.

In addition, there are five other arginine, as well as 6 other lysine residues labelled, indicating they too are bonded to heparin. The location of each residue is shown in the sequence (Figure 7.2) as well as on the surface (Figure 7.3A).

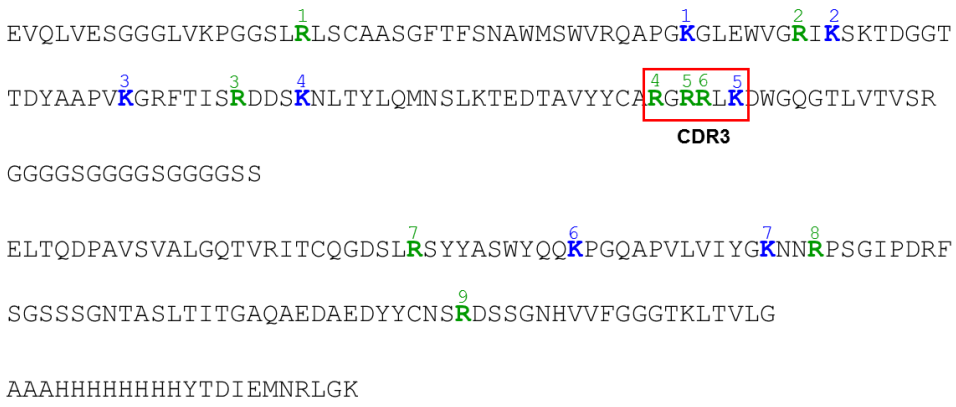


Figure 7.2: The locations of labelled lysine and arginine residues on the sequence of the antibody. The red box locates the CDR3 region of HS4C3 which is shown to bind to heparin/heparan sulfate. Arginine residues that bind to heparin are in *green* and the heparin-engaging lysine residues are in *blu*e.

From the point of view of the surface, CDR3 locates in the central basic area of the antibody (Figure 7.3B). This basic area extends from the light chain through the linker to the heavy chain without any acidic boundaries, indicating that they cooperate to bind a single sugar chain. This claim is support by both algorithms used by Pymol and SwissPDB. Whereas Pymol relies on a single residue charge, SwissPDB viewer considers atomic partial charge to build the surface electrostatic potential. Both models indicate the binding orientation of heparin through the CDR3 (Figure 7.4).

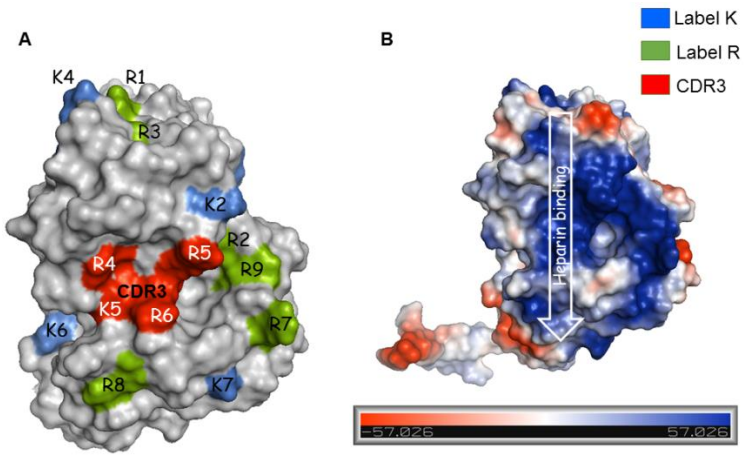


Figure 7.3: The locations of lysine, arginine residues engaging heparin. A) The locations of labelled R, K residues on the surface of HS4C3. **B)** Electrostatic density was built using Pymol, which shows the only charged residues. The highly basic area (*blue* colour) which covers the CDR3 and extends through to some K and R residues of the Light and Heavy chains suggests a potential direction of heparin when bound to the antibody.

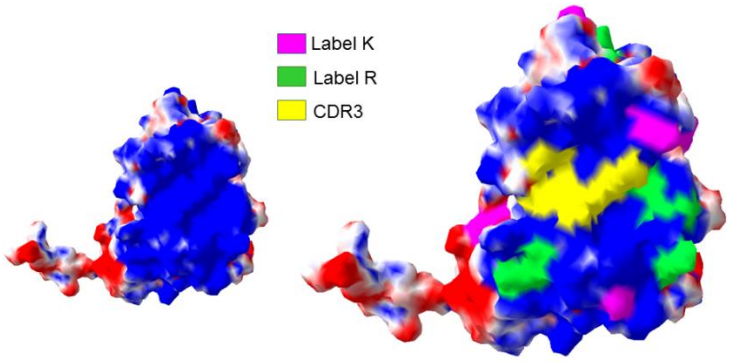


Figure 7.4: Electrostatic density of Antibody surface. Electrostatic density was built using Swiss-PDB View. Superimposed on the electrostatic surface potential map is the location of labelled arginine/lysine residues on the surface, shown as: HPG-modified arginine residues in *pink*; biotin labelled lysine residues in *green; in yellow* are the residues of CDR3

Table 7.1: MS analysis based on prediction by Prospector for the modifications on lysine residues of the antibody. The reference number of the peptide is followed by the predicted m/z and with the sequence of the peptides following cleavage of native antibody by chymotrypsin. The two columns under "Lysine modifications" present the m/z for peptides cleaved from antibody reacted in solution with NHS-acetate when bound to a heparin affinity column (protection step) and then, following elution, reaction of any protected lysine residues with NHS-biotin (labelling step). The final column indicates the modification occurring on the lysine residues of peptides.

	m/z theoretical	**Native Sequence**	**Lysine modifications**	
			m/z observed	**Modifications**
1	904.4523	**K**DWGQGTL	1136.4623	1K + 232 (Biotin)
2	925.5578	VRQAPG**K**GL	1157.5678	1K + 232 (Biotin)
3	1148.591	TISRDDS**K**NL	1380.6006	1K + 232 (Biotin)
4	1228.668	YQQ**K**PGQAPVL	1460.6784	1K + 232 (Biotin)
5	1457.761	G**K**NNRPSGIPDRF	1689.7708	1K + 232 (Biotin)
6	1597.818	VGRI**K**S**K**TDGGTTDY	1871.838	2K + 232 (Biotin) + 42 (acetyl)
7	1660.826	M(Oxidation)SWVRQAPG**K**GLEW	1892.8364	1K + 232 (Biotin)
8	1803.896	SNAWM(Oxidation)SWVRQAPG**K**GL	2035.9059	1K + 232 (Biotin)
9	1879.971	YCARGRRL**K**DWGQGTL	2111.9808	1K + 232 (Biotin)
10	1948.033	ASWYQQ**K**PGQAPVLVIY	2180.0427	1K + 232 (Biotin)
11	2052.012	TFSNAWM(Oxidation)SWVRQAPG**K**GL	2284.022	1K + 232 (Biotin)
12	2111.096	YASWYQQ**K**PGQAPVLVIY	2343.106	1K + 232 (Biotin)
13	2239.183	AAPV**K**GRFTISRDDS**K**NLTY	2513.203	2K + 232 (Biotin) + 42 (acetyl)
14	2424.299	VGRI**K**S**K**TDGGTTDYAAPV**K**GRF	2930.3294	3K + 232 (Biotin) x 2 + 42 (acetyl)

Table 7.2: MS analysis based on prediction by Prospector for the modifications on arginine residues of the antibody. The reference number of the peptide is followed by the predicted m/z and with the sequence of the peptides following cleavage of native antibody by chymotrypsin. The two columns under "Arginine modifications" present the m/z for peptides cleaved from antibody reacted in solution with PGO when bound to a heparin affinity column (protection step) and then, following elution, reaction of any protected arginine residues with HPG (labelling step). The final column indicates the modification occurring on the arginine residues of peptides.

	m/z theoretical	Native Sequence	Arginine modifications	
			m/z observed	Modifications
1	932.4261	**R**SYYASW	1064.4472	1R + HPG
2	1525.785	CA**R**G**RR**LKDW	1657.8068	3R + HPG + PGO (134) x 2
3	1597.818	TIS**R**DDSKNLTYL	1729.839	1R + HPG
4	1260.674	CA**R**G**RR**LKDW	1856.8315	3R + HPG + PGO (232) x 2
5	1946.939	GQTV**R**ITCQGDSL**R**SYY	2194.9862	2R + HPG + PGO (116)
6	2097.977	YCNS**R**DSSGNHVVFGGGTKL	2229.9982	1R + HPG
7	2331.166	KTEDTAVYYCA**R**G**RR**LKDW	2727.2297	3R + HPG x 3
8	2660.371	EVQLVESGGGLVKPGGSL**R**LSCAASGF	2792.3923	1R + HPG
9	2781.3914	VIYGKNN**R**PSGIPDRFSGSSSGNTASL	2913.4125	1R + HPG

7.3 DISCUSSION

The finding from the selective labelling of basic amino acids, lysine, and arginine, reinforced the previous definition of CDR3 on the HS4C3. The labelled lysine and arginine residues located on the electropositive charged surface of the antibody without any acidic border, indicating that a sugar chain may engage them. The previous definition of CDR3 contained R5, R6 and K5. However, this piece of work highly recommended the addition of R4 into the centre of binding, CDR3 (Figure 7.3 and 7.4). Previous work [189] used hexa- to octasaccharide fragment to study the binding properties, indicating that HS4C3 employed more than three basic residues of CDR3, which engage only a trisaccharide, to the binding. Moreover, it opens a chance to see the whole picture of sugar engagement on the antibody rather than focus on the core region of binding.

The engagement between HS4C3 and heparin was performed here at 150 mM NaCl. However, the study on the selectivity of HS4C3 toward 3-*O*-sulfate was in 500 mM NaCl [188]. As discussed in Section 6.2, the concentration of NaCl may affect electrostatic interactions. In the same meaning, it could be seen that the use of sulfo-NHS-acetate interfered the engagement of the antibody and heparin bead (Section 7.2.1). It was because sulfo carries hydorxyl group which result in the competition in the ionic binding.

It has been known that HS4C3 is selective for the 3-*O*-sulfation on the antibody, hence a question risen is which residue of CDR3 engages to 3-*O*-sulfation? Indeed, with the knowledge of the relative positions of lysine and arginine residues on the surface, is that possible to assign the binding of each amino acid to the charged groups on the heparin?

CHAPTER 8 PREDICTIVE MODEL FOR HBS(S) ON HB4C3 USING POSITIVELY CHARGED RESIDUES AND CRYSTAL STRUCTURE OF HEPARIN

8.1 INTRODUCTION

Longstanding questions are how specific and selective protein-glycosaminoglycan interactions are, to what extent can HBSs be predicted, and, if the residues in an HBS are known, what might be the sugar specificity of that protein. It is generally accepted that the spatial disposition of sulfate, the carboxyl and hydroxyl groups on the polysaccharide is important for engaging groups on the protein. However, the extent to which this is selective in the sense that there are differences in the binding structures in the polysaccharide recognised by different proteins that have functional significance is a matter of debate. Moreover, there is, as yet no informatics approach that allows any sort of prediction. For example, the Cardin-Weintraub 'consensus sequence' in a peptide will form an alpha helix, though it is based on the 3_{10} helix of collagen and it is linear, yet many GAG-binding sites in proteins are formed by amino acids distant in sequence, but adjacent in structure [110] .

This thesis and previous work have generated a dataset based on the selective labelling of lysine and arginine residues that interact with heparin in the 15 paracrine fibroblast growth factors [36] [37] (Section 4.2, Section 5.2) and a 'phage display antibody (Section 7.2)). These data can now be exploited to determine whether it may be possible to now develop a predictive model of heparin binding sites in proteins. The data are structure based, in that the geometric positions of the labelled lysines and arginines in the antibody and the FGFs are defined by structural models of the proteins (or ones based on a template model). There are also structural models of heparin oligosaccharides based on NMR experiments [92] and co-crystals of these with proteins [223] [143]. In this first attempt at exploring modelling of protein-GAG interactions, processive manual docking of the charged groups on a heparin trisaccharide to the relative spatial positions in 2-D of the lysine and arginine residues that bind sugar has been

used. A trisaccharide was chosen to try to bypass some of the uncertainty relating to the interpretation of the NMR structure of a heparin decasaccharide [92] (Section 1.4). A 2-D approach was chosen on the grounds that sugar flexibility is sufficient for a 2-D model to apply in 3-D. In future, torsion angles might be used to refine the models.

The first step was to match the potential geometric relations of charged groups on a trisaccharide (GlcNS,6S(0)-IdoA2S(+1)- GlcNS (+2)) derived from 1HPN to the heparin binding lysine and arginine residues in the central part of HBS1 (Fig 8.1A, B). The trisaccharide was then moved stepwise a disaccharide unit towards each end of the binding site, each time keeping the previously determined binding geometry of the overlapping IdoA2S, and then determining consistent interactions of the terminal disaccharide. The distance between sulfate (SO_3) groups of adjacent monosaccharides was measured between S atoms, whereas the distance between sulfate (SO_3) and the carboxyl groups was value of the distance between the respective S and C atoms. This seemed reasonable, since ionic bonds can vary in length and the position of the charge on the sulfate is not fixed to a particular oxygen. On the protein, the distance between basic residues was measured from the nitrogen (N) of the lysine side chain and from the carbon in the guanidino side chain of arginine. There are many possible 'registers' for the starting trisaccharide. Two different pieces of information were used to discriminate between registers and so models: the known selectivity of the protein for particular patterns of sulfation, determined previously by DSF [36] [37] and the requirement for all labelled lysines and arginines to have an interaction with a charged group on the sugar.

8.2 RESULTS

8.2.1. Model for the engagement of heparin dodesaccharide to the antibody HS4C3

8.2.1.1 CDR3 loop

The CDR3 loop is the variable region in the 'phage display library used to identify single chain variable fragment (scFv) antibodies that bind selectively to different structures in GAGs [224]. It is the CDR3 that is considered to impart such selectivity, since the remainder of the scFv is the same in all members of the library. The HS3C4 antibody is selective for glucosamine residues that are both *N*- and 3-*O* sulfated. Since this selectivity is imparted by its CDR3 sequence, then this was used as a starting point. K5, R4, R5 and R6 in HS3C4's CDR3 were identified as binding to heparin (Section 7.1), so these were matched to GlcNS-3S-IdoA2S-GlcNS6S unit (Figure 8.1). The only way to ensure these residues in the CDR3 were forming bonds with charged groups on the trisaccharide was for the face of the sugar, with the *N*-sulfate and 3-*O*-sulfate of GlcN together with 2-*O*-sulfate of IdoA +1 and the 6-*O*-sulfate of GlcNS +2 to engage the CDR3 of the antibody (Figure 8.1C, D). The comparison of the triangles formed by the anionic groups on GlcNS, 0 and IdoA, +1 to the triangles formed by R5, R6 and K5 supports this interpretation. In detail, the distance between the *N*-sulfate of GlcNS and the 2-*O*-sulfate of IdoA, +1, measured between the two S atoms of the two sulfate groups is 0.61 nm (Figure 8.1D), which approximates the distance between R6 and the adjacent side chain of K5, 0.8 nm (Figure 8.1C). The distance between the 3-*O*-sulfate and the *N*-sulfate of GlcNS (0) is 0.46 nm, whereas the R5 to R6 distance is 0.97 nm. However, because the 3-*O*-sulfate is missing from this structure and the size of sulfate group itself is quite massive, implying that the distance between the 3-*O*-sulfate and the *N*-sulfate of GlcNS (0) could vary considerably. The angle of the triangle formed by the 3-*O*-sulfate of GlcNS, the N-sulfate of GlcNS → the 2-*O*-sulfate of IdoA +1 is 124.4º (Figure 8.1D), which matches to the angle of the triangle R5-R6-K5 on the antibody, which is 123.5º (Figure 8.1C). Hence, R5 may engage the 3-*O*-sulfate

of central GlcNS and R6 interacts with *N*-sulfate of GlcNS (0), whereas K5 has ionic interaction with the 2-*O*-sulfate of IdoA+1.

The second comparison was performed between the triangle formed by the *N*-sulfate of GlcNS (0), the 2-*O*-sulfate of IdoA+1, the 6-*O*-sulfate of IdoA +2 and the triangle of R6, K5, R4 on the antibody. There was the already defined edge of 0.61 nm between the *N*-sulfate of GlcNS (0) and the 2-*O*-sulfate of IdoA+1, which was approximate to the edge of 0.8 nm between K5 and R6 on the antibody. The distance between the 2-*O*-sulfate of IdoA+1, the 6-*O*-sulfate of IdoA +2 was 0.61 nm, which was close value to the edge of 0.69 nm between K5 and R4 on the antibody (Figure 8.1D). The angle formed by the *N*-sulfate of GlcNS (0) → the 2-*O*-sulfate of IdoA+1 → the 6-*O*-sulfate of IdoA +2 was 78.9º (Figure 8.1D), which matched the angle of 71.4º between R6 → K5 → R4 on the antibody (Figure 8.1C). Hence, in addition to defined interactions of residues of CDR3 above, R4 may engage the 6-*O*-sulfate of IdoA, +1. The summary of the binding between CDR3 and central trisaccharide derived from the 1HPN dodesaccharide is shown Figure 8.1B.

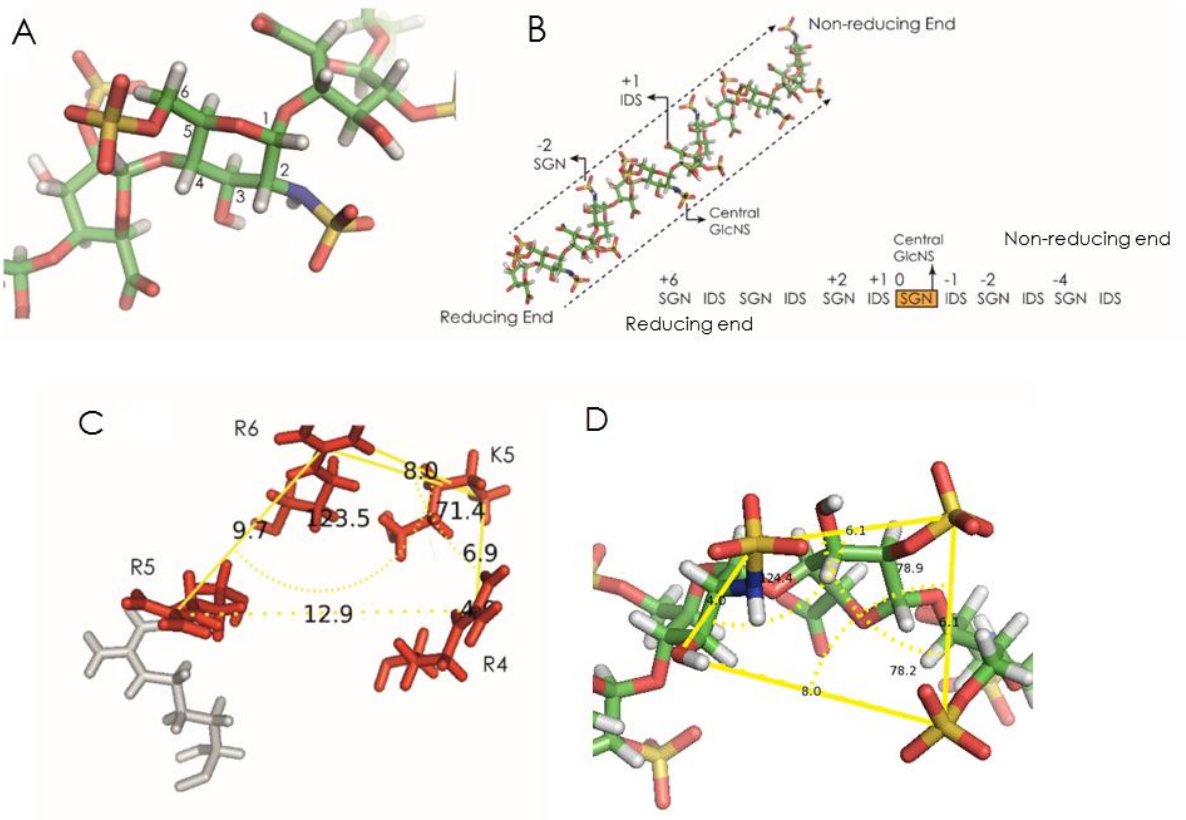


Figure 8.1: The binding of CDR3 to the core of heparin dodesaccharide. A, B) This uses the rigid model of 1HPN. Monosaccharide nomenclature is that used in the 1HPN structure: SGN = GlcNS, 6S; IDS = IdoA, 2S. The core trisaccharide centred on GlcNS, 3S, 6S will be fairly rigid; obviously the geometry of the NS-3S unit is fixed; 6S on the glucosamine exerts a 'push' on the neighbouring IdoA, so ensuring its COO- is in the right orientation for binding. This will increase the association rate constant and likely decrease the dissociation rate constant, so though unlikely to bind, 6S may play an indirect role. **C, D)** Distance and angle measurements for the residues in CDR3 identified as binding heparin, followed by the comparison to those parameters of the heparin trisaccharide.

8.2.1.2 Territory of the antibody binding to the reducing end sugars -1 to -5

The trisaccharide was moved a series of disaccharide in the reducing direction, using the constraints imposed by reducing end iduronate, whose interactions were fixed in the previous step. Then, triangles of groups on the sugar were matched to labelled residues on the antibody, with successive triangles incorporating previously assigned interactions. In this way, the modelling would be consistent with the geometry of a full-length binding oligosaccharide.

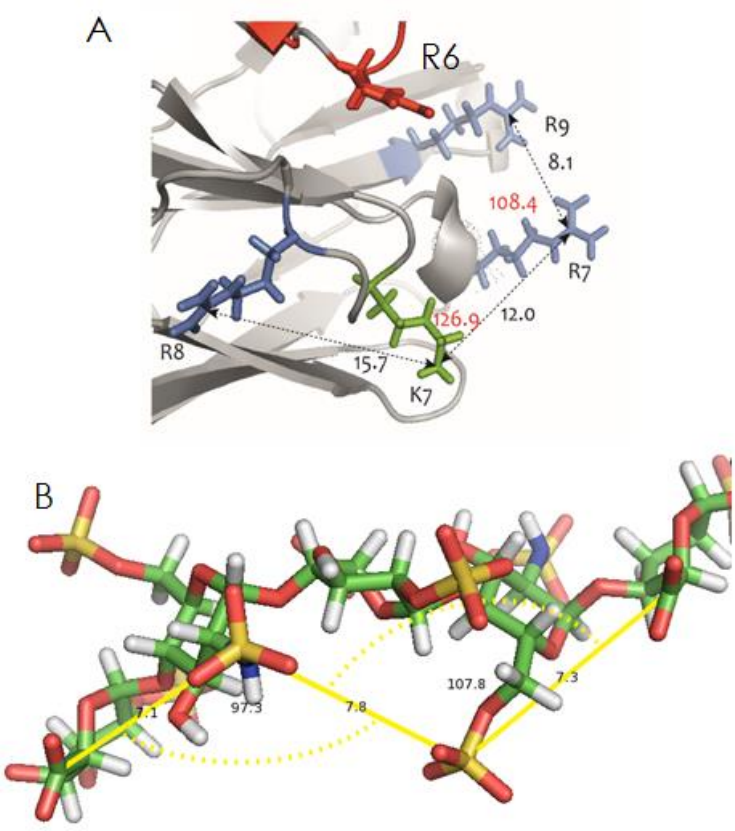


Figure 8.2: Territory of the antibody binding reducing end sugars -1 to -5. A) The distance and angle of the corresponding triangles on the antibody composing of R9, R7, K7, R8 **B)** The dimensions angle of the triangles composing the carboxyl of IdoA (-1), the 6-*O*-sulfate of GlcNS6S (-2), the *N*-sulfate of GlcNS (-4) and the carboxyl of IdoA (-5).

The reducing end starts from IdoA (-1), with the carboxyl is close to the 3-*O*-sulfate of GlcNS (0). Since the 3-*O*-sulfate of GlcNS (0) has been shown above to interact with R5 of CDR3,

the potential binding partner of the carboxyl of IdoA (-1) is R9 (Fig 8.2A). For the flexibility of the 3-*O*-sulfate and the lack of the information, the comparison of triangle is not applied in this case. However, for the neighbourhood of R9 to R6 on the antibody and the carboxyl of IdoA (-1) to the *N*-sulfate of GlcNS (0), this is most likely hypothsis in the perspective of this register on CDR3.

Moving a further to the reducing end there are IdoA (-1), GlcNS6S (-2), IdoA (-3) and GlcNS (-4) (Figure 8.2B). There are two triangles being compared. First, the triangle of the carboxyl (IdoA, -1), the 6-*O*-sulfate (GlcNS, -2), the *N*-sulfate (GlcNS, -4) was compared with the triangle of R9, R7 and K7 on the antibody. The edge of the carboxyl (IdoA, -1) and the 6-*O*-sulfate (GlcNS, -2) is 0.73 nm (Figure 8.2B), which is close in value to the edge of 0.81 between R9 and R7 on HS4C3 (Figure 8.2A). The distance between the 6-*O*-sulfate (GlcNS, -2) and the *N*-sulfate (GlcNS, -4) is 0.78 nm (Figure 8.3A), which with the length of an ionic bond approximates to the distance of 1.2 nm between R7 and K7 (Figure 8.3B). The angle of the carboxyl (IdoA, -1) → the 6-*O*-sulfate (GlcNS, -2) → the *N*-sulfate (GlcNS, -4) on the sugar are 107.8º (Figure 8.2B), which is close to the value of the angle made by R9, R7 and K7, 108.4º (Figure 8.3B). Hence, R9 interacts with the carboxyl (IdoA, -1), R7 engages the 6-*O*-sulfate (GlcNS, -2) and the binding partner of K7 is the *N*-sulfate (GlcNS, -4).

Second, the triangles formed by the 6-*O*-sulfate (GlcNS, -2), the *N*-sulfate (GlcNS, -4) and the carboxyl of IdoA (-5) match that of R7, K7 and R8 (Figure 8.2A, B). The edge of the *N*-sulfate (GlcNS, -4) and the carboxyl (IdoA, -5) is 0.71 nm, which when adding in the length of ionic bonds is equivalent to the distance between K7 and R8 on HS4C3, 1.57 nm (Figure 8.2A). The angle formed by the 6-*O*-sulfate (GlcNS, -2) → the *N*-sulfate (GlcNS, -4) → the carboxyl of IdoA (-5) is 97.2º (Figure 8.2B). The angle formed between R7-K7 to K7-R8, 126.9º (Figure 8.2A). Since the side chain of K8 is relatively more flexible than the side chain of arginine

amino acid and the 6-*O*-sulfate (GlcNS, -2) could rotate considerably, R8 likely interact with the carboxyl group of IdoA, -5.

The engagement of the 2-*O*-sulfate of IdoA (-3) is missing, but for the fact the not all C2 of IdoA would be sulfated. The case proposed above is sensible.

8.2.1.3 Territory of the antibody binding non-reducing end sugars +3 to +6

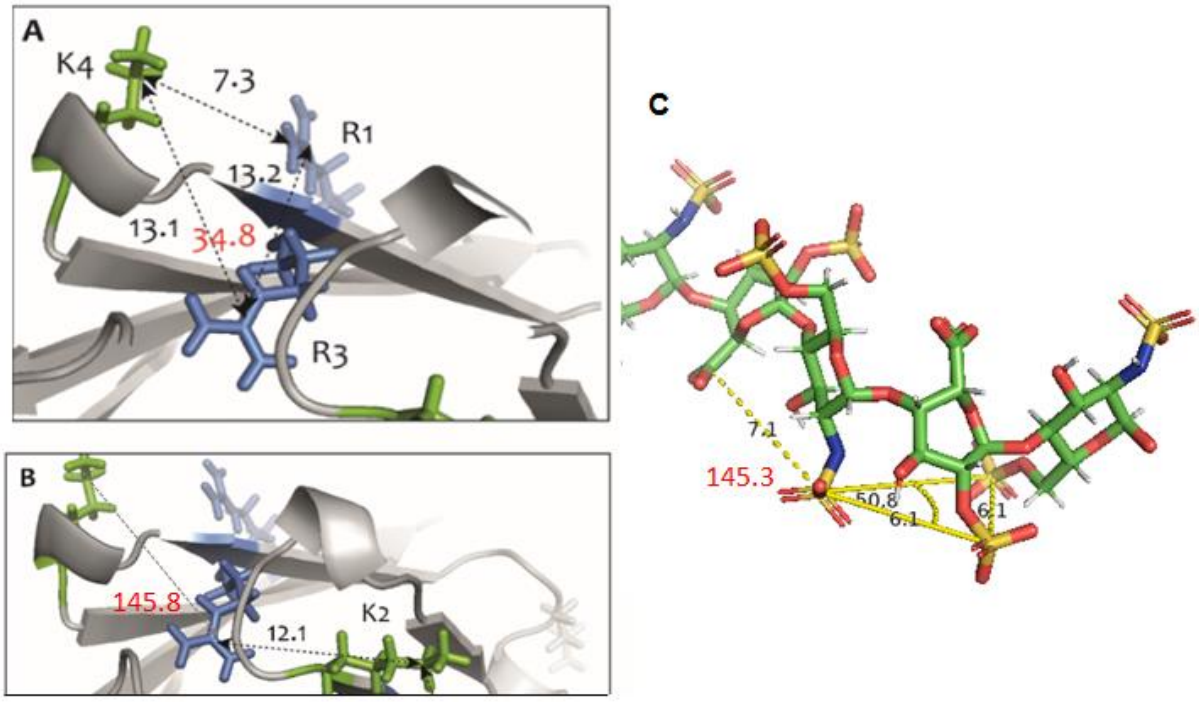


Figure 8.3: Territory of the antibody binding non-reducing end sugars +3 and +6. A, B) The distance and angle of the corresponding triangle on the antibody. **C)** The dimensions and angle of the interested triangles on dodesaccharide.

For K2 is neighbour to R4, hence the carboxyl group of IdoA (+3) likely engages this residue. The next comparison is between the triangle of the carboxyl of IdoA (+3), the *N*-sulfate of GlcNS (+4) and the 2-*O*-sulfate of IdoA (+5) and that of K2, R3, K4 on antibody. The distance between the carboxyl of IdoA (+3) and the *N*-sulfate of GlcNS (+4) is 0.71 nm. This value adding up with the length of ionic bonds which is ~ 0.5 nm is approximate to the distance of 1.21 nm between K2 and R3 on the antibody (Fig 8.3B). The edge formed by the *N*-sulfate of

GlcNS (+4) and the 2-*O*-sulfate of IdoA (+5) is 0.61 nm which in addition with the length of ionic bonds is close to the distance of 1.31 nm between R3 and K4 (Fig 8.3A). The angle formed by the carboxyl of IdoA (+3) → the *N*-sulfate of GlcNS (+4) → the 2-*O*-sulfate of IdoA (+5) is 143.5º (Figure 8.3C) which is close to the value of the angle between K2 → R3 → R1, which is 145.8º (Figure 8.3B). Hence, K2 may engage to the carboxyl of IdoA (+3), the interactor of R3 is the *N*-sulfate of GlcNS (+4) and K4 likely interacts with the 2-*O*-sulfate of IdoA (+5).

The triangle formed by the *N*-sulfate of GlcNS (+4) and the 2-*O*-sulfate of IdoA (+5) and the 6-*O*-sulfate of GlcNS (+6) was compared with the triangle of K2, K4 and R1 on HS4C3. The edge of the *N*-sulfate of GlcNS (+4) and the 2-*O*-sulfate of IdoA (+5) has been assigned above. The, the distace between the 2-*O*-sulfate of IdoA (+5) and the 6-*O*-sulfate of GlcNS (+6) was 0.61 nm which is equivalent to the distace of 0.73 nm between K4 and R1 on the antibody. The angle formed by by the *N*-sulfate of GlcNS (+4) → the 2-*O*-sulfate of IdoA (+5) → the 6-*O*-sulfate of GlcNS (+6) was 78.8 º (Fig 8.3C) which is appoximate to the angle on the antibody formed by K2 → K4 → R1 which is 74.5 º (Fig 8.3A). Those comparisons lead to the conclusion that R1 likely engages to the 6-*O*-sulfate of GlcNS (+6).

The summary of the engagement model is presented in Fig 8.4.

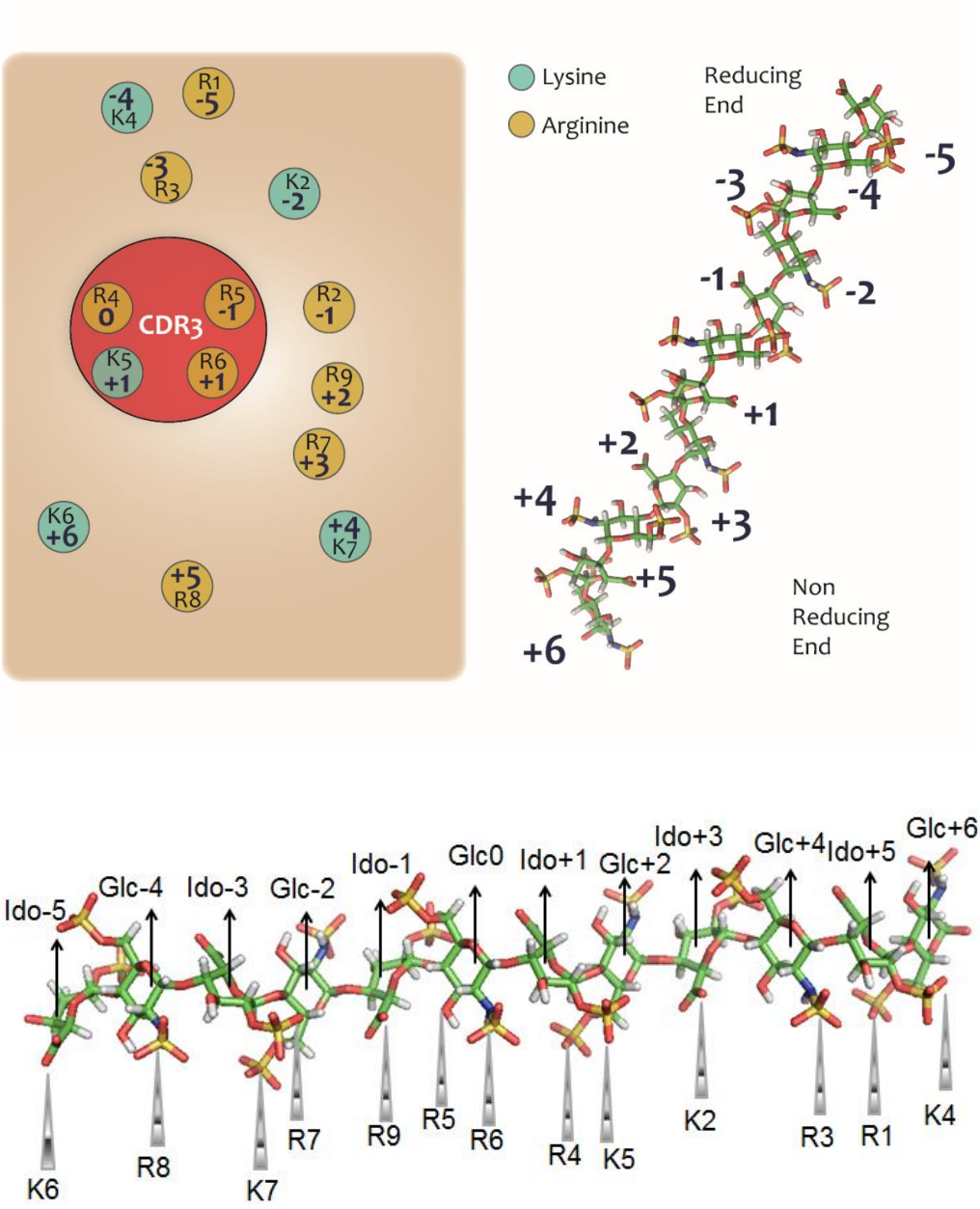


Figure 8.4: Model for binding of labelled lysine and arginine residues on antibody to sugar based on distance.

8.2.1.4 Conclusion

The above model identifies three regions of the antibody that bind the sugar. The CDR3 loop, a reducing end territory and a non-reducing end territory. Whereas antibodies are selected on the basis of CDR3 loop sequence, it is established [225] that efficient binding requires oligosaccharides that are much longer than the CDR3. This earlier work agrees with the results of the protect and label analysis of the antibody, which identified lysine and arginine residues in the CDR3 and ones outside this as binding heparin. Given the importance of the CDR3 in defining antibody selectivity, this was chosen as the starting point for modelling. The selectivity of the antibody for the 3-*O* sulfated glucosamine fixed the orientation of this residue in the trisaccharide and hence that of the flanking IdoA residues. Overall, the approach seems to have been successful, in that it was possible to identify partners on the sugar for all labelled residues in the antibody and to remain within the likely tolerance of ionic and H-bonds.

The model suggests that there is limited scope for increasing selectivity of the antibody by altering the CDR3 loop. However, changes to the reducing and non-reducing end territories could alter binding. On the kinetic front, it should be possible to reduce the dissociation rate constant, which would allow more stringent washing. This might be achieved by swapping lysine for arginine (bearing in mind that the 'reach' of arginine is less). In addition, placing residues in more favourable positions would be expected to increase the delta G of the interaction. Finally, in contrast to most mammalian heparin-binding proteins (Sections 4.2, FGF2, Section 5.2.1.3, FGF5, Section 5.2.3.3, FGF18), there is no insulation by acidic residues on the antibody. The absence of acidic borders might allow variations in sequence selectivity for binding and might in some instances even compensate for a lack of 3-*O* sulfate. Therefore, an unforeseen conclusion is that altering some of the non-binding residues flanking the labelled K and R residues to D or E might be a very effective way to constrain and so tailor the antibody's selectivity. However, to make changes that would increase binding to the product

of a specific HS3ST would require knowledge of the structures that the enzyme sulfates. One target would be the so-called gD recognition structure, where the Glc3S is flanked by IdoA2S [51]–[53]. In contrast, the present antibody, which binds ATIII recognition structures, where interaction comprise the 2-S of one of the flanking IdoA residues and the other flanking uronic acid (IdoA and GlcA being equivalent) is bound via its carboxylic acid.

8.3 DISCUSSION

By fulfilling the map of electrostatic map on each FGF, one may draw the binding path of sulfate or the carboxyl groups of HS on proteins. Some FGFs prefer a long chain HS, some choose short chain HS. The specificity is possible to predict as well, because the space distance and angle of the 6-O-sulfate, 3-O-sulfate, 2-O-sulfate. This is an addition piece of information to allow us to design and alter the binding affinity and binding specificity of protein to GAG on purpose.

The model offers a view on the potential engagement of polysaccharide on protein but not the direct evidence of the binding, hence, it may require an approach to validate. These could include mutagenesis or chemical modifications to specifically address each point of binding. Indeed, there is a weakness of using crystal structure for modelling. First, it is well known that crystal structure is fairy rigid but the interactions in the cell is fairly flexible with the induced conformation changed on both protein and sugar, making it difficult to study using an experimental model. NMR shows the conformation of protein but in free form which varies even more than the crystal structures in this case.

This approach has been applied to the ‘phage’ display antibody HS4C3 (Chapter 7). There are many FGFs for which selectivity toward the sulfation of GAG were established and they will be the future object to the method of modelling initiated here. The eventual automation of this

approach, allied to machine learning, may enable a reasonable tool for the prediction of heparin binding sites to be developed.

CHAPTER 9 GENERAL DISCUSSION AND PERSPECTIVE

9.1 DISCUSSION

The first aim of this thesis was to develop an approach for the identification of the arginine residues in proteins that interact with heparin/HS. The identification of such arginine residues would complement the established method for identifying lysine side chains that form bonds with the polysaccharide [120][36][37]. In this way a complete description of the electrostatic interactions between a protein surface and a sulfated GAG could be obtained. Once a robust method was developed, it was then used to identify heparin-binding arginines in all fifteen paracrine FGFs. This not only provided a further test of the methodology, but also produced a rich and unique dataset on heparin binding surfaces in an evolutionary related set of proteins.

Chapter 4 (Section 4.2) describes the work relating to the first aim, the development and validation of the "selective labelling of arginine residues" method for the identification of arginine residues involved in GAG binding. This approach was inspired by the strategy used to selectively labelling lysine residues with the replacement of the NHS-based reagents (Section 2.7), which are specific primary amines by the dialdehydes PGO and HPG, selective for the guanidine group of arginine. Although the chemical modification of arginine by dialdehydes is not new, it was challenging for several reasons. First, the reaction between the guanidino group of arginine and PGO produces multiple products (Section 4.1). Second, other dialdehydes require the presence of borate buffer to stabilize the final product, but this buffer may react with the vicinal diols of heparin, so interfering with the desired reaction (Section 4.1). Two well-studied FGFs were tested to optimize the conditions of the reaction between PGO/HPG and arginines on proteins. LC MS/MS was the first candidate for the method of analysing the peptide after cleavage by enzymes. The LC MS/MS confirmed the protein was FGF2 (data not shown), however, there was no modification of PGO found. This may be due

to the sensitivity of the PGO reaction product to the acidic conditions of LC-MS/MS sample preparation and the hard ionization of the process [167]. Subsequently, the analysis was changed to MALDI-TOF MS, which employs less acidic conditions and the ionization is relatively softer than in MS/MS. Since the protein introduced to the MALDI-TOF is relatively pure (as confirmed by the prior silver stain SDS-PAGE), the resolution of MALDI-TOF was deemed sufficient for the analysis.

The second aim of this thesis was to identify arginine residues in the HBSs of all fifteen paracrine FGFs (Section 5.2). One of the early challenges was the purification of FGF5, FGF8 and FGF22 (Section 3.2). Previous efforts had largely failed, due to the aggregation of the recombinant protein in inclusion bodies. Hence, N-terminal fusions with Halo-tag were used, as Halo-tag was previously found to be a good solubilisation tag for other FGFs [226]. A range of conditions of expression were tested to identify ones suitable for protein production. For Halo-FGF5, it was found to be best expressed at low temperature for a short time (Section 3.2). This may be due to the reaction of the bacteria to the FGF5 protein. Although not toxic like FGF7 [227], the observation of rapid degradation is consistent with the interpretation that the bacteria will remove this protein if possible. Moreover, even though later being expressed well, Halo-FGF5 produced in bacteria seemed to be prone to aggregation (Section 3.2) as it was eluted at a broad range of NaCl concentrations. This was similar to what was observed with Halo-FGF22, but the underlying reason is not known (Sections 3.2). In contrast, with Halo-FGF8, overnight expression produced the most protein and its chromatography on heparin was simpler (Section 3.2).

In Chapter 5, the successful purification of FGF5, FGF8 and FGF22 as Halo-tag fusions made it possible to complete the analysis of the heparin binding arginines in the fifteen paracrine FGFs, in addition to the identification of lysines in HBS of the FGFs missing from previous work [36][37][120]. The identification of labelled arginine residues in the FGFs was in some

cases supported by other data. For example, X-ray crystallography of FGF9 identified R^{137} which locates on the loop of β-strand VIII and R^{161} on β-strand X as engaged to sulfate anions from the crystallisation buffer [211]. The engagement of R^{161} to heparin is unique for FGF9, because the corresponding residues across other FGFs are not positively charged. There is also indirect evidence for the involvement of R^{134} of FGF1 in heparin binding since in an NMR structure that used inositol hexasulfate as a substitute for heparin it was found to be bound to this sulfated saccharide [200]. Later work using a synthetic heparin hexasaccharide provided direct evidence for the engagement of R^{137} of FGF1 to heparin [33], [201]. In the case of FGF2, a docking model demonstrated that the involvement of R^{90} in heparin binding in which it engages the *N*-sulfate group of glucosamine 5 (GlcNS, -5) at the reducing end of the sugar ligand [185][186].

In Chapters 4 and 5, the results produced a number of surprises. First, even though they belong to the same sub-family, FGF1 and FGF2 are fundamentally different in terms of engagement to heparin. Whereas the basic FGF2 possesses three independent HBSs, the acidic FGF1 is capable of binding to heparin through a long, continuous HBS (Section 4.2). This interpretation is supported by an analysis of the ability of different FGFs to cross-link HS chain brushes. Whereas FGF2 is able to cross-link HS chains in a brush [126] [203], FGF1 cannot [203]. Underlying the differences between FGF1 and FGF2 was the absence of key residues in FGF1 (Sections 4.2 and 5.3.2), which allows a single polysaccharide chain to engage all the heparin binding residues in FGF1.

Second, rather than possessing a secondary HBS-3, as previously thought [102], FGF4 and FGF6 now should be considered to have a single, long HBS (Section 5.2). This again accords with the observation that FGF4 and Halo-tag FGF6 do not cross-link HS chains [203]. An important issue with respect to determining whether a group of basic residues form an independent HBS or not, is the consideration of the atomic details underlying the surface

electrostatic potential. The latter arises from both the side chains of amino acid residues and the peptide backbone itself and it seems reasonable to suggest that only acidic side chains (D and E) can form a barrier to the path of the polysaccharide. For example, although the surface between K^{158} of β-strand VIII of FGF4 and K^{173}, assigned to HBS1 is electronegative, the actual side chain separating them is Y^{172} (Section 5.2), which is not an acidic amino acid, and it should not present a barrier to the polysaccharide chain. The assignment of HBSs in members of the FGF9 subfamily (FGF9, 16 and 20) was consolidated (Section 5.2). Again the issue of the electronegative surface needed to be considered for the assignment of K^{124} in FGF20 to HBS-1, whereas the protection of an arginine in FGF20. There are no data regarding the ability of FGF16 and FGF20 with respect to cross-linking HS chains, but it would appear that this biophysical measurement may be an important piece of evidence to determine whether there are truly independent secondary heparin binding sites in a protein.

As a third surprise, in contrast to other two members of FGF4 subfamily, was the presence of what appeared to be a truly independent secondary HBS in FGF5, which included the residues in the loop of β-strands IX-X (Section 5.2). This area of binding corresponded to HBS-2 of FGF2 (Section 4.2), though by itself, this was deemed insufficient to support the assignment of FGF5 to the FGF4 or the FGF1 subfamily (Section 1.2). Another FGF for which the sub-family classification is still under debate, is FGF3 (Section 1.2). However, in terms of heparin binding properties, this FGF resembles the FGF7 sub-family (Section 5.2.4, and the present data do not support its re-assignment to the FGF4 subfamily.

The data in Chapter 5 enable an examination of heparin binding sites from the viewpoint of each subfamily and so insights into their evolutionary relationships. Within subfamilies, there were differences in the lysine and arginine side chains contributing to each HBS, and in some cases, the presence or absence of what could reasonably be called an independent secondary HBSs (Sections 5.2.1, 5.2.2, 5.2.3). These changes to the configuration of HBSs across

members of an FGF subfamily could be ascribed to relatively small changes in the sequence. The alterations to GAG binding would at the least impact on the diffusion of an FGF between its source and target cells and in the selectivity of an FGF for particular GAG structures. This suggests that this divergence may contribute to the differentiation of the functions of the FGFs, allowing new functions to be acquired following a genome duplication event. The details of the FGFR isoform preferences [228] also show some differences within members of the same subfamily. This would suggest that the functional differentiation of FGFs within subfamilies occurred through relatively modest changes leading to differences in HS and FGFR binding.

There are several examples of changes to the HBS configuration within a subfamily: FGF1 and FGF2; FGF5; FGF18; FGF22. Though the differences between FGF1 and FGF2 are the greatest (one HBS *versus* three, FGF1 is perhaps the most unusual of the FGFs, since it binds to all FGFR isoforms [228]. FGF22 is a unique FGF with only arginine residues in the binding site and the position of a group of arginines on the surface of the protein contributed to the distinct shape of its HBS-1/HBS-4.

Table 9.1: FGF subfamilies with differences in HBS configuration

Subfamily	'Standard' HBSs configuration	HBS configuration	Member with altered HBS configuration	HBS configuration
FGF1 subfamily	FGF2	HBS1, HBS-2, HBS-3	FGF1	HBS-1
FGF4 subfamily	FGF4, FGF6	HBS-1	FGF5	HBS-1, HBS-2
FGF9 subfamily	FGF9, FGF16	HBS-1	FGF20	HBS-1, HBS-3
FGF8 subfamily	FGF8, FGF17	Dogleg HBS-1	FGF18	HBS-1, HBS-2 No Dogleg
FGF7 subfamily	FGF7, FGF3, FGF10	"T" shape, HBS-1, HBS-4	FGF22	"Y" shape, HBS-1, HBS_4

The third purpose of this work was to see if the protect and label strategy could be applied to a GAG in solution, which would overcome the reliance on heparin as the binding partner and allow physiologically relevant GAG chains (Section 1.3.3), to be used. Moreover, this would pave the way for eventually applying the methodology to an extracellular matrix. As discussed (Section 1.3.3.3) heparin and HS are different and other GAGs can interact with at least a subset of so-called HBPs. For example, the interaction between FGF7 and DS has been shown to play a role in wound healing [229]. Hence, a modified technique called "in-solution selective labelling" was developed and demonstrated using commercially available GAGs and a selection of FGFs. Method development was performed with FGF7 and heparin and the results indicated that a more extensive panel of GAGs and proteins should be tested to determine the potential of the approach. Published results analysing FGF-GAG interactions by DSF were used to guide which combinations of FGF and GAG to use [37]. Since the FGF7 subfamily has one of the wider repertoires of GAG partners, this became the focus of the work. While this method has been used in the context of lysine, in the future, the similar definition of arginine residues involved in binding GAGs in solution should be easy to develop using the methods described in Chapter 4. One interesting aspect of this work is the impact of electrolyte concentration on the interaction between GAGs and proteins. DSF measures protein stabilisation, which necessitates an interaction, but it does not provide much insight into binding kinetics, except in the extreme case where the *kd* is so low that over the time of the DSF measurement, binding is essentially irreversible. This has been seen with FGF10, since two populations of FGF10 , bound FGF10 and unbound FGF10, are apparent as the heparin concentration is varied in DSF [230]. The dissociation of complexes of the selected FGFs from CS and DS in PBS occurred on a time scale shorter than the reaction time, so all lysines were protected. Consequently the concentration of NaCl was reduced. Compounding the problem might be the choice of protection agent: NHS acetate will place a carboxylic acid on lysine

residues. Thus, should a lysine be involved in binding GAG dissociate, its reaction would not only prevent re-binding, but may accelerate the dissociation of neighbouring lysine residues for the GAG. Thus, replacing NHS acetate with an NHS containing compound that did not repel GAGs might allow higher NaCl concentrations to be used during the protection step.

The selective labelling of lysine and arginine residues involved in binding sulfated GAGs has applications well beyond the FGFs. For example, the selective labelling of lysine successfully identified lysine residues of PF-4 (platelet factor-4) and PTN (pleiotrophin) engaged to heparin [2]. In the context of this thesis, the antibody HS4C3 (as a gift from Professor Toin van Kuppeveldt), which selectively recognizes heparin and HS structures with a 3-*O* sulfated glucosamine through its CDR3, was successfully analysed (Section 7.2). As a result, a complete set of electrostatic interactions between HS4C3 with heparin was identified, which provides scope for a structure-guided engineering of the selectivity of this antibody for particular 3-*O* sulfated structures, such as those recognised by AT-III and the herpes simplex gD protein. Given that the number of extracellular heparin-binding proteins is now of the order of 830 [139], there is considerable scope for applying selective labelling in this context. Moreover, with the development of the in solution approach, any protein interaction involving lysine or arginine residues can now be probed in this way.

Chapter 8 introduced the idea of modelling the sugar binding to the protein, using constraints from protect and label identification of lysines and arginines involved in binding. However, the model approximates the binding surfaces to 2-D and has several weaknesses. First, the lack of 3-*O*-sulfation on the crystal structure of a heparin oligosaccharide, 1 HPN, makes it difficult to address the case of some proteins which are (or may be) selective for 3-*O*-sulfated glucosamine, such as HS4C3 and FGF7 [222][32]. Second, the structure of 1HPN is rigid, meaning that it does not include the information of the torsion of the pyranose rings along the sugar chain. Third, 1HPN is an approximation of the S-domain of HS rather than NA-NS-S

domains which could be critical for the interaction of a protein with its solution structure. Nonetheless, the approach developed in this thesis would appear to have some merit (Section 8.2). Future developments would be to incorporate the information from selective labelling and DSF analysis of preferred sulfation pattern as constraints into a docking program.

Further work

The success of the selective labelling approach of arginine and lysine on FGFs naturally leads to applying this approach to molecular assemblies involving interactions between proteins and sulfated GAGs, such as the ternary complex of FGF ligand, HS and cognate FGFR. This would complement the several available crystal structures of FGFR, FGF with heparin-derived oligosaccharides, for example [223] [143], and give an insight into the solution complex. The parallel analysis of binary complexes of FGF and FGFR would also enable the confirmation of an overlap or otherwise between the HBS-3 of many FGFs, e.g. FGF2 and FGF7 (Sections 4.2 and 5.2) with FGFR binding sites. More complex would be the interactions of heparin-binding proteins such as FGFs, with extracellular matrix. In the past this could only be analysed through binding assays using radioactive ligand [231]. More recently, advanced microscopy techniques have allowed different insights into these interactions [125][127]. An analysis of the degree to which the lysine and arginine residues found *in vitro* to interact with sulfated GAGs are similarly engaged to these in a matrix would shed light on, for example, the role of secondary HBSs. Thus, binding assays only measure binding and secondary HBSs are not apparent. In microscopy and biophysical measurements, such as the cross-linking of HS chains in a brush [203] [217], there are some correlations between FGF movement and secondary binding sites. For example, FGF1 appears to move shorter distances, but to diffuse faster than FGF2 in the pericellular matrix of rat mammary fibroblasts [127]. However, while a protect and label experiment on extracellular matrix would entail analysis of peptides from endogenous heparin-binding proteins, as well as ones from any exogenously-added protein, it may be difficult to reliably identify labelled peptides directly at a suitable sensitivity in such a complex mixture. For lysine, the biotinylation step allows purification of the peptides, which would simplify the analysis. For arginine, the multiple product of the PGO reaction may preclude such an analysis.

Therefore, one useful development would be a biotin-PGO (or biotin-HPG) in the labelling step, to allow the labelled peptides to be purified before analysis. A potential problem with this is the high concentration of TFA (20% v/v) required for quantitative elution of peptides from Streptactin Sepharose (other forms of biotin binding proteins, such a mono avidin, fail to yield quantitative recovery [120]). The very low pH will cause the product of reaction of PGO/HPG with arginine to degrade. One solution would be to insert a disulphide between the PGO and the biotin (Figure 9.1) so that DTT could be used to cleave S-S and peptides with PGO could be eluted from the column without affecting the PGO-arginine reaction products.

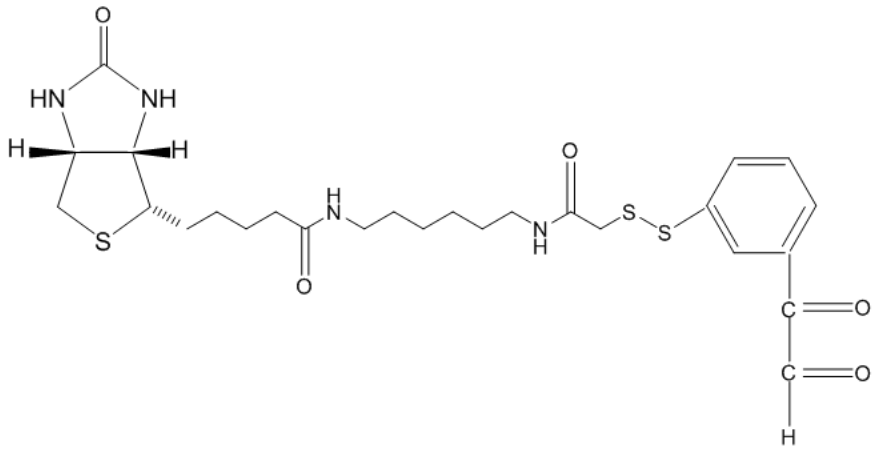


Figure 9.1: The chemical structure of Biotin-S-S-linker-PGO.

As mentioned above, the in solution selective labelling of arginine residues is required to establish the molecular basis of the interaction between protein and GAG (heparin/CS/DS) without the limitation of an affinity chromatography column. The in solution selective labelling of lysine residues revealed a difference in protein bound even when comparing the same GAG, heparin, with beads and in solution. For example, the lysine residues of FGF10 which engage to heparin in solution were not identical to those identified when FGF10 bound to heparin on the affinity mini column. However, due to the lack of data on the arginine residues involved in binding in solution, the definition of binding sites is limited and only partial.

The “T” shaped basic patch of binding formed by HBS-1 and HBS-4 observed in members of FGF7 sub-family may require additional evidence to show that there are more than one GAG chain engaging to protein. The primary test could be done by cross-link which was performed before for FGF10 and found partial cross-link to HS brush. Later, the mutation of the basic amino acids in the joins would enable to separate HBS-1 and HBS-4, hence providing more evidence for this model.

The weakness of identifying peptides by MADLI-TOF is that this technique only able to identify the overall modifications on a peptide. Thus, for peptides containing multiple arginine residues and multiple products, it cannot determine which residues have been modified in a particular way, e.g., reacted with PGO or HPG. The list of peptides where there is uncertainty regarding the labelled arginine is provided in Table 9.1.

	Native FGF6		FGF6 P&L			
Ref. no.	**Predicted native peptide m/z**	**Sequence**	**Predicted modified peptide m/z**	**m/z observed**	**Match HBS**	**Modifications**
1	1288.80	^{80}VGIK**R**Q**RR**LY^{89}	1652.87	1652.96	HBS3 R^{84}	3R + 132 **(HPG)** + 116 + 116
	Native FGF5		**FGF5 P&L**			
2	1316.61	^{157}TDDCKF**R**E**R**F^{166}	1680.69	1680.89	HBS2 R^{163} and R^{165}	2R + 132 **(HPG)** + 232
3	2227.12	^{80}QWSPSG**RR**TGSLYC**R** VGIGF^{99}	2723.25	2723.35	HBS3 R^{86} and R^{87}	3R + 132 **(HPG)** + 132 **(HPG)** + 232
	Native FGF9		**FGF9 protect and label (P&L)**			
4	1131.60	^{154}KHVDTG**RR**Y^{162}	1495.17	1495.69	HBS1 R^{160} and R^{161}	2R + 132 **(HPG)** + 232
	Native FGF20		**FGF20 protect and label (P&L)**			
5	2381.29	^{166}VALNKDGTP**R**DGA**R** SK**R**HQK^{186}	2895.38	2895.40	HBS-1 R^{176}, R^{180}, and R^{183}	3R + 132 **(HPG) x 2 + 250**
6	1516.85	^{191}LP**R**PVDPE**R**VPEL^{203}	1880.93	1880.92	HBS-1 R^{199} and R^{192}	2R + 132 **(HPG) + 232**
	Native FGF8		**FGF8 protect and label (P&L)**			
7	2589.446	^{154}TRKG**R**P**R**KGSKT**R**Q HQ**R**EVHF^{174}	3232.94	3233.03	HBS1 R^{155}, R^{158}, R^{160}, R^{170}	5R + 132 **(HPG) x 4 + 116**
8	2514.372	^{55}QLYS**R**TSGKHVQVLA NK**R**INAM^{76}	2896.83	2896.46	HBS1 R^{59} and R^{72}	2R + 132 **(HPG) + 250**
	Native FGF17		**FGF17 protect and label (P&L)**			
9	1107.602	^{48}S**RR**QI**R**EY^{58}	1487.67	1487.735	HBS1 R^{49} and R^{50}	3R + 132 **(HPG) x 2 + 116**
10	2915.478	^{151}MAFT**R**QG**R**P**R**QAS**R** S**R**QNQ**R**EAHF^{174}	3691.598	3691.01	HBS1 R^{155}, R^{158}, R^{160}, R^{164} and R^{166}	6R + 132 **(HPG)** x 5 + 116
	Native FGF7		**FGF7 protect and label (P&L)**			
11	1992.03	^{60}EGGDI**R**V**RR**LFC**R**TQ W^{75}	2522.27	2522.12	HBS1 R^{72} HBS3 R^{65}, R^{67} and R^{68}	4R + 132 **(HPG) x 3 + 116**
	Native FGF3		**FGF3 protect and label (P&L)**			
12	1108.604	^{98}AMNK**R**G**R**LY^{106}	1356.616	1356.652	HBS1 R^{102} and R^{104}	2R + 132 **(HPG) + 116**
13	1429.814	^{155}VSVNGKG**R**P**RR**GF^{177}	1827.45	1827.88	HBS1 R^{162}, R^{164} and R^{165}	3R + 132 **(HPG) x 2 + 134**
	Native FGF22		**FGF22 protect and label (P&L)**			
14	2054.159	^{140}LALD**RR**GGP**R**PGG**R** T**RR**Y^{158}	2848.13	2846.23	HBS1 R^{145}, R^{146}, R^{150}, R^{153}, R^{155} and R^{156}	6R + 132 **(HPG) x 5 + 134**
15	2533.303	^{95}VAMN**RR**G**R**LYGS**R**L YTVDC**R**F^{115}	3077.72	3077.66	R^{107} HBS1 R^{99}, R^{100}, R^{102} and R^{114}	5R + 132 **(HPG) x 4** + 250 and oxidation

Table 9.1: The list of peptides for further analysis.

The work on modelling samples proteins provides an opportunity for the engineering of antibodies or natural HBPs to alter their selectivity toward GAGs. This would also provide evidence that the modelling was a good approximation and should perhaps be done in tandem with the development of an automated predictive docking system. There are a number of considerations that would be involved in such engineering of a HBS. For example,, arginine binds stronger to sulfate groups than lysine, but it has a shorter side chain. Hence, using arginine rather than lysine would likely decrease the dissociation of the protein from the polysaccharide and, in the context of neighbouring basic residues, potentially enable selection of a 3S-NS unit, as seen in the HS4C3 antibody. In addition, a two dimensional binding surface would likely be more selective than a linear one, as has been proposed from a limited analysis of HBSs [232]. Allied to this point is the shape of the HBS may also impart selectivity, with likely differences arising due to the HBS being linear, possessing a dogleg, or something in between these two extremes.

REFERENCES

[1] C. Popovici, R. Roubin, F. Coulier, and D. Birnbaum, “An evolutionary history of the FGF superfamily,” *BioEssays*, vol. 27, no. 8, pp. 849–857, 2005.

[2] H. P. Makarenkova *et al.*, “Differential interactions of FGFs with heparan sulfate control gradient formation and branching morphogenesis,” *Sci Signal.*, vol. 2, no. 88, pp. 1–21, 2010.

[3] J. W. Crabb *et al.*, “Complete Primary Structure of Prostatropin, a Prostate Epithelial Cell Growth Factor,” *Biochemistry*, vol. 25, no. 18, pp. 4988–4993, 1986.

[4] F. Esch *et al.*, “Primary structure of bovine pituitary basic fibroblast growth factor (FGF) and comparison with the amino-terminal sequence of bovine brain acidic FGF.,” *Proc. Natl. Acad. Sci.*, vol. 82, no. 19, pp. 6507–6511, 2006.

[5] J. A. Smith, P. Winslow, and S. Rudland, “Brain and pituitary fibroblast growth factor activities behave identically on three independent high performance liqid chromatography systems,” *Biochem. Biophys. Res. Commun.*, vol. 119, no. 1, pp. 311–318, 1984.

[6] D. G. Fernig and J. T. Gallagher, “Fibroblast growth factors and their receptors: an information network controlling tissue growth, morphogenesis and repair.,” *Prog. Growth Factor Res.*, vol. 5, no. 4, pp. 353–77, 1994.

[7] N. Itoh, “The Fgf Families in Humans, Mice, and Zebrafish: Their Evolutional Processes and Roles in Development, Metabolism, and Disease,” *Biol. Pharm. Bull.*, vol. 30, no. 10, pp. 1819–1825, 2007.

[8] N. Itoh and D. M. Ornitz, “Functional evolutionary history of the mouse Fgf gene family,” *Dev. Dyn.*, vol. 237, no. 1, pp. 18–27, 2008.

[9] S. Oulion, S. Bertrand, and H. Escriva, “Evolution of the FGF Gene Family,” *Int. J. Evol. Biol.*, vol. 2012, pp. 1–12, 2012.

[10] N. Itoh, H. Ohta, and M. Konishi, “Endocrine FGFs: Evolution, Physiology, Pathophysiology, and Pharmacotherapy,” *Front. Endocrinol. (Lausanne).*, vol. 6, no. September, pp. 1–9, 2015.

[11] D. M. Ornitz and N. Itoh, “The fibroblast growth factor signaling pathway,” *Wiley Interdiscip. Rev. Dev. Biol.*, vol. 4, no. 3, pp. 215–266, 2015.

[12] M. Goldfarb, “Fibroblast growth factor homologous factors: evolution, structure, and function,” *Cytokine Growth Factore Rev.*, vol. 16, no. 2, pp. 215–220, 2011.

[13] G. Mitchell, “Voltage gated sodium channel associated proteins and alternative mechanisms of inactivation and block,” *Cell Mol Life Sci.*, vol. 71, no. 2, pp. 233–236, 2013.

[14] S. Jon and M. Goldfarb, “FGF homologous factors and the Islet brain-2 scaffold protein regulate activation of a stress-activated protein kinase,” *J Biol Chem*, vol. 71, no. 2, pp. 233–236, 2013.

[15] J. A. Abraham *et al.*, “Human basic fibroblast growth factor: nucleotide sequence and genomic organization.,” *EMBO J.*, vol. 5, no. 10, pp. 2523–8, 1986.

[16] C. A. MacArthur *et al.*, “FGF-8 isoforms activate receptor splice forms that are expressed in mesenchymal regions of mouse development,” *Development*, vol. 121, no. 11, pp. 3603–3613, 1995.

[17] D. M. Ornitz and N. Itoh, “Protein family review: Fibroblast growth factors Gene organization and evolutionary history,” *Genome Biol.*, vol. 2, no. reviews, pp. 1–12, 2001.

[18] V. Papaioannou, T. DeChiara, M. Goldfarb, W. Poueymirou, and B. Feldman, “Requirement of FGF-4 for postimplantation mouse development,” *Science (80-.).*, vol. 267, no. 5195, pp. 246–249, 2006.

[19] J. M. Bishop, M. J. Depew, G. R. Martin, J. L. R. Rubenstein, and A. Trumpp, “Cre-mediated gene inactivation demonstrates that FGF8 is required for cell survival and patterning of the first branchial arch,” *Genes Dev.*, vol. 13, no. 23, pp. 3136–3148, 2002.

[20] R. Raballo, J. Rhee, R. Lyn-Cook, J. F. Leckman, M. L. Schwartz, and F. M. Vaccarino, “Basic fibroblast growth factor (Fgf2) is necessary for cell proliferation and neurogenesis in the developing cerebral cortex.,” *J. Neurosci.*, vol. 20, no. 13, pp. 5012–23, 2000.

[21] J. M. Hébert, T. Rosenquist, J. Götz, and G. R. Martin, “FGF5 as a regulator of the hair growth cycle: Evidence from targeted and spontaneous mutations,” *Cell*, vol. 78, no. 6, pp. 1017–1025, 1994.

[22] J. Qiao, R. Uzzo, T. Obara-Ishihara, L. Degenstein, E. Fuchs, and D. Herzlinger, “FGF-7 modulates ureteric bud growth and nephron number in the developing kidney.,” *Development*, vol. 126, pp. 547–554, 1999.
[23] W. S. Simonet *et al.*, “Fgf-10 is required for both limb and lung development and exhibits striking functional similarity to Drosophila branchless,” *Genes Dev.*, vol. 12, no. 20, pp. 3156–3161, 2008.
[24] J. S. Colvin, a C. White, S. J. Pratt, and D. M. Ornitz, “Lung hypoplasia and neonatal death in Fgf9-null mice identify this gene as an essential regulator of lung mesenchyme.,” *Development*, vol. 128, no. 11, pp. 2095–2106, 2001.
[25] N. Ohbayashi *et al.*, “FGF18 is required for normal cell proliferation and differentiation during osteogenesis and chondrogenesis,” *Genes Dev.*, vol. 16, no. 7, pp. 870–879, 2002.
[26] S. L. Mansour, J. M. Goddard, and M. R. Capecchi, “Mice homozygous for a targeted disruption of the proto-oncogene int-2 have developmental defects in the tail and inner ear,” *Development*, vol. 28, pp. 13–28, 1993.
[27] H. Kinoshita *et al.*, “Fgf16 is required for cardiomyocyte proliferation in the mouse embryonic heart,” *Dev. Dyn.*, vol. 237, no. 10, pp. 2947–2954, 2008.
[28] Jae Myoung Suh *et al.*, “Endocrinization of FGF1 produces a neomorphic and potent insulin sensitizer,” *Nature*, vol. 513, no. 7518, pp. 436–439, 2014.
[29] H. Ago, Y. Kitagawa, A. Fujishima, Y. Matsuura, and Y. Katsube, “Crystal structure of basic fibroblast growth factor at 1.6 Ångström resolution,” *J. Biochem.*, vol. 110, no. 3, pp. 360–363, 1991.
[30] X. Zhu *et al.*, “Three-dimensional structures of acidic and basic fibroblast growth factors,” *Science (80-.).*, vol. 251, no. 4989, pp. 90–93, 2006.
[31] B. K. Yeh *et al.*, “Structural basis by which alternative splicing confers specificity in fibroblast growth factor receptors.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 100, no. 5, pp. 2266–71, 2003.
[32] S. Ye *et al.*, “Structural Basis for Interaction of FGF-1 , FGF-2 , and FGF-7 with Different Heparan Sulfate Motifs,” *Biochemistry*, vol. 40, pp. 14429–14439, 2001.
[33] A. Canales *et al.*, “Solution NMR structure of a human FGF-1 monomer, activated by a hexasaccharide heparin-analogue,” *FEBS J.*, vol. 273, no. 20, pp. 4716–4727, 2006.
[34] A. N. Plotnikov *et al.*, “Crystal Structure of Fibroblast Growth Factor 9 Reveals Regions Implicated in Dimerization and Autoinhibition,” *J. Biol. Chem.*, vol. 276, no. 6, pp. 4322–4329, 2001.
[35] T. and W. H. B. Maciag, “The Heparin-Binding (Fibroblast) Growth Factor family of proteins,” *Annu. Rev. Biochem.*, vol. 58, pp. 575–606, 1989.
[36] R. Xu *et al.*, “Diversification of the structural determinants of fibroblast growth factor-heparin interactions: Implications for binding specificity,” *J. Biol. Chem.*, vol. 287, no. 47, pp. 40061–40073, 2012.
[37] Y. Li, C. Sun, E. A. Yates, C. Jiang, M. C. Wilkinson, and D. G. Fernig, “Heparin binding preference and structures in the fibroblast growth factor family parallel their evolutionary diversification,” *Open Biol.*, vol. 6, no. 150275, 2016.
[38] P. Bellosta *et al.*, “Identication of Receptor and Heparin Binding Sites in Fibroblast Growth Factor 4 by Structure-Based Mutagenesis,” *Mol. Cell. Biol.*, vol. 21, no. 17, pp. 5946–5957, 2001.
[39] A. A. Gupta, R. H. Chou, H. Li, L. W. Yang, and C. Yu, “Structural insights into the interaction of human S100B and basic fibroblast growth factor (FGF2): Effects on FGFR1 receptor signaling,” *Biochim. Biophys. Acta - Proteins Proteomics*, vol. 1834, no. 12, pp. 2606–2619, 2013.
[40] R. Z. Florkiewicz, “Human basic fibroblast growth factor gene encodes four polypeptides: Three initiate translation from non-AUG codons,” *Proc. Natl. Acad. Sci.*, vol. 87, no. 5, pp. 2045–2045, 2006.
[41] M. Kaghad *et al.*, “High molecular mass forms of basic fibroblast growth factor are initiated by alternative CUG codons.,” *Proc. Natl. Acad. Sci.*, vol. 86, no. 6, pp. 1836–1840, 2006.
[42] J. Kalinina *et al.*, “Homodimerization controls the fibroblast growth factor 9 subfamily’s receptor binding and heparan sulfate-dependent diffusion in the extracellular matrix,” *Mol. Cell. Biol.*, vol. 29, no. 17, pp. 4663–4678, 2009.

[43] K. Miyakawa, K. Ozawa, T. Uruno, and T. Imamura, “The C-terminal region of fibroblast growth factor-1 is crucial for its biological activity and high level protein expression in mammalian cells.,” *Growth Factors*, vol. 16, no. 3, pp. 191–200, 1999.
[44] N. Itoh and D. M. Ornitz, “Evolution of the Fgf and Fgfr gene families,” *Trends Genet.*, vol. 20, no. 11, pp. 563–569, 2004.
[45] A. Ori, “The heparanome and regulation of cell function: structures, functions and challenges,” *Front. Biosci.*, vol. Volume, no. 13, p. 4309, 2008.
[46] J. Van Den Born, K. Jann, K. J. M. Assmann, U. Lindahl, and J. H. M. Berden, “N-Acetylated Domains in Heparan Sulfates Revealed by a Monoclonal Antibody against the Escherichia coli K5,” *Biochemistry*, vol. 271, no. 37, pp. 22802–22809, 1996.
[47] D. L. Rabenstein, “Heparin and heparan sulfate: structure and function,” *Nat. Prod. Rep.*, vol. 19, no. 3, pp. 312–331, 2002.
[48] J. E. Silbert and G. Sugumaran, “Biosynthesis of Chondroitin / Dermatan Sulfate,” *IUBMB Life*, vol. 54, pp. 177–186, 2002.
[49] J. D. Esko and S. B. Selleck, “ORDER OUT OF CHAOS: Assembly of ligand binding sites in heparan sulfate,” *Annu. Rev. Biochem.*, vol. 71, no. 1, pp. 435–471, 2002.
[50] A. H. and G. B. Magnus Hook, ULF Lindahl, “Biosynthesis of Heparin,” *J. Biol. Chem.*, vol. 250, no. 15, pp. 6065–6071, 1975.
[51] D. Shukla *et al.*, “A novel role for 3-O-sulfated heparan sulfate in herpes simplex virus 1 entry.,” *Cell*, vol. 99, no. 1, pp. 13–22, 1999.
[52] D. Xu, V. Tiwari, G. Xia, C. Clement, D. Shukla, and J. Liu, “Characterization of heparan sulphate 3-O-sulphotransferase isoform 6 and its role in assisting the entry of herpes simplex virus type 1,” *Biochem. J.*, vol. 385, no. 2, pp. 451–459, 2005.
[53] G. Xia, “Heparan sulfate 3-O-sulfotransferase isoform 5 generates both an Antithrombin-binding site and an entry receptor for herpes simplex virus, type 1,” *J. Biol. Chem.*, vol. 277, no. 40, pp. 37912–37919, 2002.
[54] E. Forsberg *et al.*, “Abnormal mast cells in mice deficient in a heparin-synthesizing enzyme.,” *Nature*, vol. 400, no. August, pp. 773–776, 1999.
[55] J. Ledin *et al.*, “Enzymatically active N-deacetylase/N-sulfotransferase-2 is present in liver but does not contribute to heparan sulfate N-sulfation,” *J. Biol. Chem.*, vol. 281, no. 47, pp. 35727–35734, 2006.
[56] T. R. Rudd and E. A. Yates, “A highly efficient tree structure for the biosynthesis of heparan sulfate accounts for the commonly observed disaccharides and suggests a mechanism for domain synthesis,” *Mol. Biosyst.*, vol. 8, no. 5, pp. 1499–1506, 2012.
[57] 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, no. 26, pp. 27239–27245, 2004.
[58] X. Ai, A. T. Do, M. Kusche-Gullberg, U. Lindahl, K. Lu, and C. P. Emerson, “Substrate specificity and domain functions of extracellular heparan sulfate 6-O-endosulfatases, QSulf1 and QSulf2,” *J. Biol. Chem.*, vol. 281, no. 8, pp. 4969–4976, 2006.
[59] G. K. Dhoot, M. K. Gustafsson, X. Ai, W. Sun, D. M. Standiford, and J. Emerson, “Regulation of Wnt signaling and embryo patterning by an extracellular sulfatase,” *Science (80-.).*, vol. 293, no. 5535, pp. 1663–1666, 2001.
[60] F. Gong *et al.*, “Processing of macromolecular heparin by heparanase,” *J. Biol. Chem.*, vol. 278, no. 37, pp. 35152–35158, 2003.
[61] M. Bernfield, Z. Wang, M. L. Fitzgerald, P. W. Park, and G. Murphy, “Shedding of syndecan-1 and -4 ectodomains is regulated by multiple signaling pathways and mediated by a TIMP-3–sensitive metalloproteinase,” *J. Cell Biol.*, vol. 148, no. 4, pp. 811–824, 2002.
[62] J. Kreuger, L. Perez, A. J. Giraldez, and S. M. Cohen, “Opposing activities of Dally-like glypican at high and low levels of wingless morphogen activity,” *Dev. Cell*, vol. 7, no. 4, pp. 503–512, 2004.
[63] J. Kreuger and L. Kjellén, “Heparan sulfate biosynthesis: regulation and variability,” *J. Histochem. Cytochem.*, vol. 60, no. 12, pp. 898–907, 2012.
[64] H. Kitagawa, T. Izumikawa, T. Uyama, and K. Sugahara, “Molecular cloning of a chondroitin polymerizing factor that cooperates with chondroitin synthase for chondroitin polymerization,”

J. Biol. Chem., vol. 278, no. 26, pp. 23666–23671, 2003.
[65] K. Sugahara and S. Yamada, "Structure and function of oversulfated chondroitin sulfate variants. unique sulfation patterns and neuroregulatory activities.," *Trends Glycosci. Glycotechnol.*, vol. 12, no. 67, pp. 321–349, 2011.
[66] K. Sugahara, T. Mikami, T. Uyama, S. Mizuguchi, K. Nomura, and H. Kitagawa, "Recent advances in the structural biology of chondroitin sulfate and dermatan sulfate," *Curr. Opin. Struct. Biol.*, vol. 13, no. 5, pp. 612–620, 2003.
[67] G. C. Epimerization and I. Acid, "Structural diversity of N-sulfated heparan sulfate domains: distinct modes of glucuronyl C5 epimerization, iduronic acid 2-O-sulfation, and glucosamine 6-O-sulfation," *Biochemistry*, vol. 39, pp. 10823–10830, 2000.
[68] H. Kitagawa *et al.*, "A novel pentasaccharide sequence in the oligosaccharides isolated from king crab cartilage chondroitin sulfate k and its differential susceptibility to chondroitinases and hyaluronidase," *Biochemistry*, vol. 2960, no. 96, pp. 3998–4008, 1997.
[69] C. C. I. S. Identical *et al.*, "Biosynthesis of Dermatan Sulfate," *J. Biol. Chem.*, vol. 281, no. 17, pp. 11560–11568, 2006.
[70] J. E. Silbertls and M. E. Palmerg, "Formation of Dermatan Sulfate by cultured human skin fibroblasts," *J. Biol. Chem.*, vol. 261, no. 29, pp. 13397–13400, 1986.
[71] J. E. Silbert and L. S. Freilich, "Biosynthesis of chondroitin sulphate by a Golgi-apparatus-enriched preparation from cultures of mouse mastocytoma cells," *Biochem. J.*, vol. 190, pp. 307–313, 1980.
[72] G. Sugumaran, J. N. Cogburn, and J. E. Silbert, "Simultaneous sulfation of endogenous chondroitin sulfate and chondroitin-derived oligosaccharides," *J. Biol. Chem.*, vol. 261, no. 27, pp. 2–7, 1986.
[73] G. O. Staples, X. Shi, and J. Zaia, "Extended N -Sulfated domains reside at the nonreducing end of heparan sulfate chains," *J. Biol. Chem.*, vol. 285, no. 24, pp. 18336–18343, 2010.
[74] J. R. Couchman *et al.*, "On the roles and regulation of chondroitin sulfate and heparan sulfate in zebrafish pharyngeal cartilage morphogenesis," *J. Biol. Chem.*, vol. 287, no. 40, pp. 33905–33916, 2012.
[75] S. Le Jan *et al.*, "Functional overlap between chondroitin and heparan sulfate proteoglycans during VEGF-induced sprouting angiogenesis," *Arterioscler. Thromb. Vasc. Biol.*, vol. 32, no. 5, pp. 1255–1263, 2012.
[76] J. T. Gallagher and A. Walker, "Molecular distinctions between heparan sulphate and heparin. Analysis of sulphation patterns indicates that heparan sulphate and heparin are separate families of N-sulphated polysaccharides.," *Biochem. J.*, vol. 230, no. 3, pp. 665–74, 1985.
[77] J. L. Dreyfuss, C. V. Regatieri, T. R. Jarrouge, R. P. Cavalheiro, L. O. Sampaio, and H. B. Nader, "Heparan sulfate proteoglycans: structure, protein interactions and cell signaling," *An. Acad. Bras. Cienc.*, vol. 81, no. 3, pp. 409–429, 2009.
[78] J. E. Turnbull and J. T. Gallagher, "Distribution of iduronate 2-sulphate residues in heparan sulphate. Evidence for an ordered polymeric structure.," *Biochem. J.*, vol. 273 (Pt 3, pp. 553–9, 1991.
[79] J. E. Turnbull, J. J. Hopwood, and J. T. Gallagher, "A strategy for rapid sequencing of heparan sulfate and heparin saccharides," *Proc. Natl. Acad. Sci.*, vol. 96, no. 6, pp. 2698–2703, 2002.
[80] J. Bae, U. R. Desai, A. Pervin, E. E. Caldwell, J. M. Weiler, and R. J. Linhardt, "Interaction of heparin with synthetic antithrombin III peptide analogues.," *Biochem. J.*, vol. 301, pp. 121–129, 1994.
[81] M. A. Lemmon *et al.*, "Heparin-induced oligomerization of fgf molecule is responsible for fgf receptor dimerization, activation, and cell proliferation," *Cell*, vol. 79, pp. 1015–1024, 1994.
[82] R. Sadir, E. Forest, and H. Lortat-Jacob, "The heparan sulfate binding sequence of interferon-γ increased the on rate of the interferon-γ-interferon-γ receptor complex formation," *J. Biol. Chem.*, vol. 273, no. 18, pp. 10919–10925, 1998.
[83] C. J. Robinson, B. Mulloy, J. T. Gallagher, and S. E. Stringer, "VEGF165-binding sites within heparan sulfate encompass two highly sulfated domains and can be liberated by K5 lyase," *J. Biol. Chem.*, vol. 281, no. 3, pp. 1731–1740, 2006.
[84] V. Nurcombe, M. D. Ford, J. A. Wildschut, and P. F. Bartlett, "Developmental regulation of neural response to FGF-1 and FGF-2 by heparan sulfate proteoglycan," *Science (80-.).*, vol.

260, no. April, pp. 103–107, 1993.
[85] D. A. V *et al.*, "Differential structural requirements of heparin and heparan sulfate proteoglycans that promote binding of basic fibroblast growth factor to its receptor," *J. Biol. Chem.*, vol. 269, no. 1, pp. 114–121, 1994.
[86] M. Maccaranasg, B. Casufl, and U. Lindahl, "Minimal sequence in heparin / heparan sulfate required for binding of basic fibroblast growth factor," *J. Biol. Chem.*, vol. 268, no. 32, pp. 23898–23905, 1993.
[87] E. Rönnberg, F. R. Melo, and G. Pejler, "Mast Cell Proteoglycans," *J. Histochem. Cytochem.*, vol. 60, no. 12, pp. 950–962, 2012.
[88] M. C. Z. Meneghetti *et al.*, "Heparan sulfate and heparin interactions with proteins," *J. R. Soc. Interface*, vol. 12, no. 110, 2015.
[89] A. S. Brito *et al.*, "A non-hemorrhagic hybrid heparin/heparan sulfate with anticoagulant potential," *Carbohydr. Polym.*, vol. 99, pp. 372–378, 2014.
[90] B. Mulloy, E. Gray, and T. W. Barrowcliffe, "Characterization of unfractionated heparin: Comparison of materials from the last 50 years," *Thromb. Haemost.*, vol. 84, no. 6, pp. 1052–1056, 2000.
[91] S. Khan *et al.*, "The solution structure of heparan sulfate differs from that of heparin: Implications for function," *J. Biol. Chem.*, vol. 288, no. 39, pp. 27737–27751, 2013.
[92] B. Mulloy, M. J. Forster, C. Jones, and D. B. Daviest, "N.m.r. and molecular-modelling studies of the solution conformation of heparin," *Biochem. J.*, vol. 858, no. 293, pp. 849–858, 1993.
[93] D. Mikhailov, J. R. Linhardt, and H. K. Mayo, "NMR solution conformation of heparin-derived hexasaccharide," *Biochem. J.*, vol. 328, no. 1, pp. 51–61, 2015.
[94] M. Zsiška and B. Meyer, "Influence of sulfate and carboxylate groups on the conformation of chondroitin sulfate related disaccharides," *Carbohydr. Res.*, vol. 243, no. 2, pp. 225–258, 1993.
[95] W. C. W. Chen *et al.*, "Sulfation and cation effects on the conformational properties of the glycan backbone of chondroitin sulfate disaccharides," *J Phys Chem B.*, vol. 33, no. 2, pp. 557–573, 2016.
[96] B. Mulloy and M. J. Forster, "Conformation and dynamics of heparin and heparan sulfate," *Glycobiology*, vol. 10, no. 11, pp. 1147–1156, 2000.
[97] A. D. Digabriele *et al.*, "Structure of a heparin-linked biologically active dimer of fibroblast growth factor," *Nature*, vol. 393, no. June, pp. 812–817, 1998.
[98] N. S. Gandhi and R. L. Mancera, "The structure of glycosaminoglycans and their interactions with proteins," *Chem. Biol. Drug Des.*, vol. 72, no. 6, pp. 455–482, 2008.
[99] E. E. Fry *et al.*, "The structure and function of a foot-and-mouth disease virus-oligosaccharide receptor complex," *EMBO J.*, vol. 18, no. 3, pp. 543–554, 1999.
[100] S. Faham, R. E. Hileman, J. R. Fromm, R. J. Linhardt, and D. C. Rees, "Heparin structure and interactions with basic fibroblast growth factor," *Science (80-.).*, vol. 271, no. 5252, pp. 1116–1120, 1996.
[101] A. Almond and J. K. Sheehan, "Glycosaminoglycan conformation: do aqueous molecular dynamics simulations agree with x-ray fiber diffraction," *Glycobiology*, vol. 10, no. 3, pp. 329–338, 2000.
[102] Y. Li *et al.*, "Heparin binding preference and structures in the fibroblast growth factor family parallel their evolutionary diversification," *Open Biol.*, vol. 6, no. 150275, 2016.
[103] I. Capila *et al.*, "Annexin V-heparin oligosaccharide complex suggests heparan sulfate-mediated assembly on cell surfaces," *Structure*, vol. 9, no. 1, pp. 57–64, 2001.
[104] M. Petitou, R. N. Pike, R. W. Carrell, R. Skinner, L. Jin, and J. P. Abrahams, "The anticoagulant activation of antithrombin by heparin," *Proc. Natl. Acad. Sci.*, vol. 94, no. 26, pp. 14683–14688, 2002.
[105] M. Ragazzi, D. R. Ferro, B. Perly, P. Sinaÿ, M. Petitou, and J. Choay, "Conformation of the pentasaccharide corresponding to the binding site of heparin for antithrombin III," *Carbohydr. Res.*, vol. 195, no. 2, pp. 169–185, 1990.
[106] M. Hricovíni, M. Guerrini, and A. Bisio, "Structure of heparin-derived tetrasaccharide complexed to the plasma protein antithrombin derived from NOEs, J-couplings and chemical shifts," *Eur. J. Biochem.*, vol. 261, no. 3, pp. 789–801, 1999.
[107] J. Kreuger, K. Prydz, R. F. Pettersson, U. Lindahl, and M. Salmivirta, "Characterization of

fibroblast growth factor 1 binding heparan sulfate domain.," *Glycobiology*, vol. 9, no. 7, pp. 723–9, 1999.
[108] J. E. Turnbull, D. G. Fernig, Y. Ke, M. C. Wilkinson, and J. T. Gallagher, "Identification of the basic fibroblast growth factor binding sequence in fibroblast heparan sulfate.," *J. Biol. Chem.*, vol. 267, no. 15, pp. 10337–41, 1992.
[109] E. M. Munoz, "Heparin-Binding Domains in Vascular Biology," *Arterioscler. Thromb. Vasc. Biol.*, vol. 24, no. 9, pp. 1549–1557, 2004.
[110] A. D. C. and H. J. R. Weintraub, "Molecular modeling of protein-glycosaminoglycan interactions," *ARTERIOSCLEROSIS*, vol. 9, no. 1, pp. 21–32, 1989.
[111] S. N. Bolten, U. Rinas, and T. Scheper, "Heparin: role in protein purification and substitution with animal-component free material," *Appl. Microbiol. Biotechnol.*, vol. 102, no. 20, pp. 8647–8660, 2018.
[112] H. Margalit, N. Fischer, and S. a Ben-Sasson, "Comparative analysis of structurally defined heparin binding sequences reveals a distinct spatial distribution of basic residues.," *J. Biol. Chem.*, vol. 268, no. 26, pp. 19228–31, 1993.
[113] Z. Johnson *et al.*, "Chemokine inhibition – why, when, where, which and how?," *Biochem. Soc. Trans.*, vol. 32, no. 2, pp. 366–377, 2004.
[114] J. R. Fromm, R. E. Hileman, E. E. Caldwell, J. M. Weiler, and R. J. Linhardt, "Differences in the interaction of heparin with arginine and lysine and the importance of these basic amino acids in the binding of heparin to acidic fibroblast growth factor.," *Arch. Biochem. Biophys.*, vol. 323, no. 2, pp. 279–87, 1995.
[115] B. Musafia, V. Buchner, and D. Arad, "Complex salt bridges in proteins: Statistical analysis of structure and function," *J. Mol. Biol.*, vol. 254, no. 4, pp. 761–770, 1995.
[116] D. Xu, C. J. Tsai, and R. Nussinov, "Hydrogen bonds and salt bridges across protein-protein interfaces," *Protein Eng. Des. Sel.*, vol. 10, no. 9, pp. 999–1012, 2002.
[117] R. G. Pearson, "Hard and Soft Acids and Bases," *J. Am. Chem. Soc.*, vol. 85, no. 22, pp. 3533–3539, 1963.
[118] N. M. Luscombe, R. A. Laskowski, and J. M. Thornton, "Amino acid–base interactions: a three-dimensional analysis of protein–DNA interactions at an atomic level," *Nucleic Acids Res.*, vol. 29, no. 13, pp. 2860–2874, 2001.
[119] A. Nicholls and B. Honig, "Classical electrostatics in biology and chemistry," *Science (80-.).*, vol. 268, no. May, pp. 1144–1149, 1995.
[120] A. Ori, P. Free, J. Courty, M. C. Wilkinson, and D. G. Fernig, "Identification of heparin-binding sites in proteins by selective labeling," *Mol. Cell. Proteomics*, vol. 8, no. 10, pp. 2256–2265, 2009.
[121] L. Kinsella, H. L. Chen, J. A. Smith, P. S. Rudland, and D. G. Fernig, "Interactions of putative heparin-binding domains of basic fibroblast growth factor and its receptor, FGFR-1, with heparin using synthetic peptides," *Glycoconj. J.*, vol. 15, no. 4, pp. 419–422, 1998.
[122] C. J. Dowd, C. L. Cooney, and M. A. Nugent, "Heparan sulfate mediates bFGF transport through basement membrane by diffusion with rapid reversible binding," *J. Biol. Chem.*, vol. 274, no. 8, pp. 5236–5244, 1999.
[123] S. Hiraoka *et al.*, "FGF9 monomer/dimer equilibrium regulates extracellular matrix affinity and tissue diffusion," *Nat Genet.*, vol. 41, no. 3, pp. 289–298, 2009.
[124] T. Y. Belenkaya *et al.*, "Drosophila Dpp morphogen movement is independent of dynamin-mediated endocytosis but regulated by the glypican members of heparan sulfate proteoglycans," *Cell*, vol. 119, no. 2, pp. 231–244, 2004.
[125] L. Duchesne *et al.*, "Transport of fibroblast growth factor 2 in the pericellular matrix is controlled by the spatial distribution of its binding sites in Heparan Sulfate," *PLoS Biol.*, vol. 10, no. 7, p. e1001361, 2012.
[126] E. Migliorini *et al.*, "Cytokines and growth factors cross-link heparan sulfate," *Open Biol.*, 2015.
[127] D. G. Fernig, D. Mason, M. Marcello, Y. Li, C. Sun, and R. Lévy, "Selectivity in glycosaminoglycan binding dictates the distribution and diffusion of fibroblast growth factors in the pericellular matrix," *Open Biol.*, vol. 6, no. 3, p. 150277, 2016.
[128] A. Yayon, M. Klagsbrun, J. D. Esko, P. Leder, and D. M. Ornitz, "Cell-surface, heparin-like molecules are required for binding of basic fibroblast growth-factor to its high-affinity receptor,"

Cell, vol. 64, no. 4, pp. 841–848, 1991.
[129] A. C. Rapraeger, A. Krufka, and B. B. Olwin, "Requirement of heparan sulfate for bFGF-mediated fibroblast growth and myoblast differentiation," *Science (80-.).*, vol. 252, no. 5013, pp. 1705–1708, 1991.
[130] J. Schlessinger *et al.*, "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.
[131] A. Yayon, J. G. Flanagan, P. Leder, E. Levi, C. M. Svahn, and D. M. Ornitz, "Heparin is required for cell-free binding of basic fibroblast growth factor to a soluble receptor and for mitogenesis in whole cells.," *Mol. Cell. Biol.*, vol. 12, no. 1, pp. 240–247, 2015.
[132] R. Raman, G. Venkataraman, S. Ernst, V. Sasisekharan, and R. Sasisekharan, "Structural specificity of heparin binding in the fibroblast growth factor family of proteins.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 100, no. 5, pp. 2357–62, 2003.
[133] A. N. Plotnikov, J. Schlessinger, S. R. Hubbard, and M. Mohammadi, "Structural basis for FGF receptor dimerization and activation.," *Cell*, vol. 98, no. 5, pp. 641–650, 1999.
[134] M. Delehedde *et al.*, "Fibroblast growth factor-2 stimulation of p42/44(MAPK) phosphorylation and IκB degradation is regulated by heparan sulfate/heparin in rat mammary fibroblasts," *J. Biol. Chem.*, vol. 275, no. 43, pp. 33905–33910, 2000.
[135] L. Duchesne, B. Tissot, T. R. Rudd, A. Dell, and D. G. Fernig, "N-Glycosylation of fibroblast growth factor receptor 1 regulates ligand and heparan sulfate co-receptor binding," *J. Biol. Chem.*, vol. 281, no. 37, pp. 27178–27189, 2006.
[136] M. Fannon, K. E. Forsten, and M. A. Nugent, "Potentiation and inhibition of bFGF binding by heparin: A model for regulation of cellular response," *Biochemistry*, vol. 39, no. 6, pp. 1434–1445, 2000.
[137] S. K. Olsen *et al.*, "Structural basis by which alternative splicing modulates the organizer activity of FGF8 in the brain," *Genes Dev.*, vol. 20, no. 2, pp. 185–198, 2006.
[138] A. Ori, M. C. Wilkinson, and D. G. Fernig, "A systems biology approach for the investigation of the heparin/heparan sulfate interactome," *J. Biol. Chem.*, vol. 286, no. 22, pp. 19892–19904, 2011.
[139] Q. M. Nunes *et al.*, "The heparin-binding proteome in murine experimental acute pancreatitis," *Pancreas*, vol. 45, no. 10, p. 1530, 2016.
[140] P. L. Lee, D. E. Johnson, L. S. Cousens, V. a Fried, and L. T. Williams, "Purification and complementary DNA cloning of a receptor for basic fibroblast growth factor.," *Science*, vol. 245, no. 4913, pp. 57–60, 1989.
[141] A. M. Gallagher, P. Wong, E. Szylobryt, W. H. Burgess, M. Jaye, and B. Hampton, "Analysis of Putative Heparin-binding Domains of Fibroblast Growth Factor-1: Using site-directed mutagenesis and peptide analogues," *J. Biol. Chem.*, vol. 270, no. 43, pp. 25805–25811, 2002.
[142] K. H. Mayo *et al.*, "Heparin binding to platelet factor-4. An NMR and site-directed mutagenesis study: arginine residues are crucial for binding," *Biochem J*, vol. 312 (Pt 2, pp. 357–365, 1995.
[143] L. Pellegrini, D. F. Burke, F. von Delft, B. Mulloy, and T. L. Blundell, "Crystal structure of fibroblast growth factor receptor ectodomain bound to ligand and heparin.," *Nature*, vol. 407, no. 6807, pp. 1029–34, 2000.
[144] A. N. Plotnikov, S. R. Hubbard, J. Schlessinger, and M. Mohammadi, "Crystal structures of two FGF-FGFR complexes reveal the determinants of ligand-receptor specificity," *Cell*, vol. 101, no. 4, pp. 413–424, 2000.
[145] K. Uniewicz, A. Ori, R. Xu, Y. Ahmed, D. G. Fernig, and E. Yates, "Differential Scanning Fluorimetry measurement of protein stability changes upon binding to glycosaminoglycans : a rapid screening test for binding specificity," *Anal. Chem.*, vol. 82, no. 9, pp. 1–3, 2010.
[146] R. Sadir, A. Imberty, F. Baleux, and H. Lortat-Jacob, "Heparan sulfate/heparin oligosaccharides protect stromal cell-derived factor-1 (SDF-1)/CXCL12 against proteolysis induced by CD26/dipeptidyl peptidase IV," *J. Biol. Chem.*, vol. 279, no. 42, pp. 43854–43860, 2004.
[147] R. Xu, T. R. Rudd, A. J. Hughes, G. Siligardi, D. G. Fernig, and E. a. Yates, "Analysis of the fibroblast growth factor receptor (FGFR) signalling network with heparin as coreceptor: evidence for the expansion of the core FGFR signalling network," *FEBS J.*, vol. 280, no. 10, pp. 2260–2270, 2013.
[148] 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, no. 2, pp. 265–272, 2015.
[149] N. Keiser, G. Venkataraman, Z. Shriver, and R. Sasisekharan, "Direct isolation and sequencing of specific protein-binding glycosaminoglycans," *Nat. Med.*, vol. 7, no. 1, pp. 123–128, 2001.
[150] S. Ricard-Blum, "Protein–glycosaminoglycan interaction networks: Focus on heparan sulfate," *Perspect. Sci.*, vol. 11, pp. 62–69, 2016.
[151] M. Delehedde, M. Lyon, R. Vidyasagar, T. J. McDonnell, and D. G. Fernig, "Hepatocyte growth factor/scatter factor binds to small heparin-derived oligosaccharides and stimulates the proliferation of human HaCaT keratinocytes," *J. Biol. Chem.*, vol. 277, no. 14, pp. 12456–12462, 2002.
[152] L. S. Jones, B. Yazzie, and C. R. Middaugh, "Polyanions and the Proteome," *Mol. Cell. Proteomics*, vol. 3, no. 8, pp. 746–769, 2004.
[153] B. R. Olsen *et al.*, "Characterization of Endostatin Binding to Heparin and Heparan Sulfate by Surface Plasmon Resonance and Molecular Modeling," *J. Biol. Chem.*, vol. 279, no. 4, pp. 2927–2936, 2004.
[154] N. Sapay, E. Cabannes, M. Petitou, and A. Imberty, "Molecular modeling of the interaction between heparan sulfate and cellular growth factors: Bringing pieces together," *Glycobiology*, vol. 21, no. 9, pp. 1181–1193, 2011.
[155] H. Lortat-Jacob, A. Grosdidier, and A. Imberty, "Structural diversity of heparan sulfate binding domains in chemokines.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 99, no. 10, pp. 1229–1234, 2002.
[156] S. E. Mottarella, D. Beglov, N. Beglova, M. A. Nugent, D. Kozakov, and S. Vajda, "Docking Server for the Identification of Heparin Binding Sites on Proteins," *J. Chem. Inf. Model.*, vol. 54, pp. 2068–2078, 2014.
[157] K. a. Uniewicz, A. Ori, Y. a. Ahmed, E. a. Yates, and D. G. Fernig, "Characterisation of the interaction of neuropilin-1 with heparin and a heparan sulfate mimetic library of heparin-derived sugars," *PeerJ*, vol. 2, p. e461, 2014.
[158] V. L. Mendoza and R. W. Vachet, "Probing Protein Structure by Amino Acid-Specific Covalent Labeling and Mass Spectrometry Vanessa," *Mass Spectrom Rev*, vol. 28, no. 5, pp. 785–815, 2010.
[159] A. Leitner and W. Lindner, "Probing of arginine residues in peptides and proteins using selective tagging and electrospray ionization mass spectrometry," *J. Mass Spectrom.*, vol. 38, no. 8, pp. 891–899, 2003.
[160] K. K. Makinen, "Diketones as photosensitizing agents : application to a-amino acids and enzymes," *Photochem*, vol. 35, pp. 761–765, 1982.
[161] K. Takahashi, "The reaction of phenylglyoxal with arginine residues in proteins," *J. Biol. Chem.*, vol. 243, no. 23, pp. 6171–6179, 1968.
[162] T. Inugami, "Chemical modification of arginine with 1,2-cyclohexanedione," *J. Biol. Chem.*, vol. 240, no. August, pp. 2–5, 1965.
[163] K. Takahashi, "Macromolecules : the reaction of phenylglyoxal with arginine residues in proteins the reaction of phenylglyoxal residues in proteins with arginine," *J. Biol. Chem.*, vol. 243, no. 23, pp. 6171–6179, 1968.
[164] L. Flink, "Phenylglyoxal modification of Cardiac S-l," *J. Biol. Chem.*, vol. 254, no. 24, pp. 12647–12652, 1979.
[165] M. a Vanoni, M. Pilone Simonetta, B. Curti, a Negri, and S. Ronchi, "Phenylglyoxal modification of arginines in mammalian D-amino-acid oxidase.," *Eur. J. Biochem.*, vol. 167, no. 2, pp. 261–7, 1987.
[166] A. J. Poulose and P. E. Kolattukudy, "Enzymatic reduction of phenylglyoxal and 2 , 3-butanedione , two commonly used arginine-modifying reagents , by the ketoacyl reductase domain of fatty acid synthase," *Int. J. Biochem*, vol. 18, no. 9, pp. 807–812, 2001.
[167] Shu-Tong Cheung and Margaret L.Fonda, "Reaction of phenylglyoxal with arginine: the effect of buffers and pH," *Biochem. Biophys. Res. Commun.*, vol. 90, no. 3, pp. 940–947, 1979.
[168] P. B. Pandeswari, V. Sabareesh, and M. A. Vijayalakshmi, "Insights into stoichiometry of arginine modification by phenylglyoxal and 1 , 2-cyclohexanedione probed by LC-ESI-MS," *J. Proteins Proteomics*, vol. 7, no. 4, pp. 323–347, 2016.
[169] M. M. Werber, M. Moldovan, and M. Sokolovsky, "Modification of arginyl residues in porcine

carboxypeptidase B," *Eur. J. Biochem.*, vol. 53, no. 1, pp. 207–216, Apr. 1975.
[170] T. B. Rogers, T. Bqrresen, and R. E. Feeney, "Chemical modification of the arginines in transferrins," vol. 17, no. 6, 1978.
[171] T. D. Wood, Z. Guan, C. L. Borders, L. H. Chen, G. L. Kenyon, and F. W. McLafferty, "Creatine kinase: essential arginine residues at the nucleotide binding site identified by chemical modification and high-resolution tandem mass spectrometry.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 95, no. 7, pp. 3362–5, 1998.
[172] M. O. F. Arginine, "Modification of arginine in proteins by oligomers of 2,3-Butanedione," *Biochemistry*, vol. 9, no. 12, pp. 2433–2439, 1969.
[173] R. R. Lobb, A. M. Stokes, H. A. O. Hill, and J. F. Riordan, "A functional arginine residue in rabbit-muscle aldolase," *Eur. J. Biochem.*, vol. 70, no. 2, pp. 517–522, Nov. 1976.
[174] R. L. . Lundblad, "Chapter 4. The Modification of Arginine," *Chem. Reagents Protein Modif.*, 2004.
[175] L. Angeles, "Identification of functional arginine in ribonuclease a and lysozyme residues," *J. Biol. Chem.*, vol. 259, no. 2, pp. 565–569, 1975.
[176] P. Junkova *et al.*, "Improved approach for the labeling of arginine, glutamic, and aspartic acid side chains in proteins using chromatographic techniques," *J. Liq. Chromatogr. Relat. Technol.*, vol. 36, pp. 1221–1230, 2013.
[177] X. Wu, S. G. Chen, J. M. Petrash, and V. M. Monnier, "Alteration of substrate selectivity through mutation of two arginine residues in the binding site of amadoriase II from Aspergillus sp.," *Biochemistry*, vol. 41, pp. 4453–4458, 2002.
[178] A. Ori, M. C. Wilkinson, and D. G. Fernig, "The heparanome and regulation of cell function: structures, functions and challenges," *Front. Biosci.*, vol. Volume, no. 13, p. 4309, 2008.
[179] S. K. Olsen *et al.*, "Insights into the molecular basis for fibroblast growth factor receptor autoinhibition and ligand-binding promiscuity," *PNAS*, vol. 101, no. 4, pp. 935–940, 2004.
[180] A. Dementiev, M. Petitou, J.-M. Herbert, and P. G. W. Gettins, "The ternary complex of antithrombin–anhydrothrombin–heparin reveals the basis of inhibitor specificity," *Nat. Struct. Mol. Biol.*, vol. 11, no. 9, pp. 863–867, 2004.
[181] S. T. Olson, H. R. Halvorson, and I. Bjork, "Quantitative characterization of the thrombin-heparin interaction: Discrimination between specific and nonspecific binding models," *J. Biol. Chem.*, vol. 266, no. 10, pp. 6342–6352, 1991.
[182] P. Rousselle, R. R. Vivès, E. Crublet, H. Lortat-Jacob, J. Gagnon, and J.-P. Andrieu, "A novel strategy for defining critical amino acid residues involved in protein/glycosaminoglycan interactions," *J. Biol. Chem.*, vol. 279, no. 52, pp. 54327–54333, 2004.
[183] A. Ori, P. Free, J. Courty, M. C. Wilkinson, and D. G. Fernig, "Identification of heparin-binding sites in proteins by selective labeling," *Mol. Cell. Proteomics*, vol. 8, no. 10, pp. 2256–2265, 2009.
[184] F. J. Moy *et al.*, "Properly oriented heparin-decasaccharide-induced dimers are the biologically active form of basic fibroblast growth factor," *Biochemistry*, vol. 36, no. 16, pp. 4782–4791, 1997.
[185] L. D. Thompson, M. W. Pantoliano, and B. A. Springer, "Energetic characterization of the basic fibroblast growth factor-heparin interaction: identification of the heparin binding domain," *Biochemistry*, vol. 33, no. 13, pp. 3831–3840, 1994.
[186] B. A. Springer *et al.*, "Identification and concerted function of 2 receptor-binding surfaces on basic fibroblast growth-factor required for mitogenesis," *J. Biol. Chem.*, vol. 269, no. 43, pp. 26879–26884, 1994.
[187] A. Amara *et al.*, "Stromal cell-derived factor-1 α associates with heparan sulfates through the first β -strand of the chemokine," *J. Biol. Chem.*, vol. 274, no. 34, pp. 23916–23925, 1999.
[188] J. Lee *et al.*, "Structural determinants of heparin-transforming growth factor-β1 interactions and their effects on signaling," *Glycobiology*, vol. 25, no. 12, pp. 1491–1504, 2015.
[189] C. Dos Santos *et al.*, "Proliferation and migration activities of fibroblast growth factor-2 in endothelial cells are modulated by its direct interaction with heparin affin regulatory peptide," *Biochimie*, vol. 107, pp. 350–357, 2014.
[190] K. Takahashi, "Chemistry and metabolism of macromolecules: The reaction of phenylglyoxal with arginine residues in proteins," *J. Biol. Chem.*, vol. 243, no. 23, pp. 6171–6179, 1968.

[191] E. Morkin, I. L. Flink, and Surath K. Banerjee, "Phenylglyoxal modification of Cardiac myosin S-l: Evidence for essential arginine residues at the active sites," *J. Biol. Chem.*, vol. 254, no. 24, pp. 12647–12652, 1979.
[192] M. A. Vanoni, M. P. Simonetta, B. Curti, A. Negri, and S. Ronchi, "Phenylglyoxal modification of arginines in mammalian D-amino-acid oxidase.," *Eur. J. Biochem.*, vol. 167, no. 2, pp. 261–7, 1987.
[193] M. Chevallet, S. Luche, and T. Rabilloud, "Silver staining of proteins in polyacrylamide gels," *Nat. Protoc.*, vol. 1, no. 4, pp. 1852–1858, 2006.
[194] K. Takahashp, "Further Studies on the Reactions of Phenylglyoxal and related reagents with proteins," *J. Biol. Chem.*, vol. 414, no. 81, pp. 403–414, 1977.
[195] O. Eriksson, E. Fontaine, and P. Bernardi, "Chemical modification of arginines by 2,3-butanedione and phenylglyoxal causes closure of the mitochondrial permeability transition pore," *J. Biol. Chem.*, vol. 273, no. 20, pp. 12669–12674, 1998.
[196] A. Baird, D. Schubert, N. Ling, and R. Guillemin, "Receptor- and heparin-binding domains of basic fibroblast growth factor.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 85, no. 7, pp. 2324–2328, 1988.
[197] D. Aviezer, M. Safran, and A. Yayon, "Heparin differentially regulates the interaction of fibroblast growth factor-4 with FGF receptors 1 and 2," *Biochem. Biophys. Res. Commun.*, vol. 263, no. 3, pp. 621–626, 1999.
[198] R. M. Lozano *et al.*, "^{1}H NMR structural characterization of a nonmitogenic, vasodilatory, ischemia-protector and neuromodulatory acidic fibroblast growth factor," *Biochemistry*, vol. 39, no. 17, pp. 4982–4993, 2000.
[199] H. Mach, D. B. Volkin, C. J. Burke, C. R. Middaugh, and L. Mattsson, "Nature of the interaction of heparn with acidic fibroblast growth factor," *Biochemistry*, vol. 32, pp. 5480–5489, 1993.
[200] M. R. and G. G.-G. A. Pineda-Lucena, M. A. Jimezez, J. L. Nieto, J. Santoro, "H-NMR assignment and solution structure of Human acidic Fibroblast growth factor activated by Inositol Hexasulfate." pp. 81–98, 1994.
[201] K. Ogura *et al.*, "Solution Structure of Human Acidic Fibroblast Growth Factor and Interaction with Heparin-Derived Hexasaccharide.," *J. Biomol. NMR*, vol. 13, no. 1, pp. 11–24, 1999.
[202] E. Migliorini *et al.*, "Cytokines and growth factors cross-link heparan sulfate," *Open Biol.*, vol. 5, no. 8, 2015.
[203] Baradji Aïseta, "Interactions of fibroblast growth factors with glycosaminoglycan brushes and the pericellular matrix," 2017.
[204] D. Yan and X. Lin, "Shaping morphogen gradients by proteoglycans.," *Cold Spring Harb. Perspect. Biol.*, vol. 1, no. 3, pp. 1–16, 2009.
[205] D. Krilleke *et al.*, "Molecular mapping and functional characterization of the VEGF164 heparin-binding domain," *J. Biol. Chem.*, vol. 282, no. 38, pp. 28045–28056, 2007.
[206] H. Lortat-Jacob, J. E. Turnbull, and J. A. Grimaud, " Molecular organization of the interferon γ -binding domain in heparan sulphate ," *Biochem. J.*, vol. 310, no. 2, pp. 497–505, 2015.
[207] P. Bellosta, A. Iwahori, A. N. Plotnikov, A. V. Eliseenkova, C. Basilico, and M. Mohammadi, "Identification of Receptor and Heparin Binding Sites in Fibroblast Growth Factor 4 by Structure-Based Mutagenesis," *Mol. Cell. Biol.*, vol. 21, no. 17, pp. 5946–5957, 2002.
[208] T. Rudd, E. Yates, and M. Hricovini, "Spectroscopic and Theoretical Approaches for the Determination of Heparin Saccharide Structure and the Study of Protein-Glycosaminoglycan Complexes in Solution," *Curr. Med. Chem.*, vol. 16, no. 35, pp. 4750–4766, 2009.
[209] T. R. Rudd *et al.*, "Construction and use of a library of bona fide heparins employing 1H NMR and multivariate analysis," *Analyst*, vol. 136, no. 7, pp. 1380–1389, 2011.
[210] Y. C. Li *et al.*, "Interactions that influence the binding of synthetic heparan sulfate based disaccharides to fibroblast growth factor-2," *ACS Chem. Biol.*, vol. 9, no. 8, pp. 1712–1717, 2014.
[211] H. J. Hecht, R. Adar, O. Bogin, B. Hofmann, H. Weich, and A. Yayon, "Structure of fibroblast growth factor 9 shows a symmetric dimer with unique receptor- and heparin-binding interfaces," *Biol. Crystallogr.*, no. April 2001, pp. 378–384, 2017.
[212] Y. Liu *et al.*, "Regulation of Receptor Binding Specificity of FGF9 by an Autoinhibitory Homodimerization Article Regulation of Receptor Binding Specificity of FGF9 by an

Autoinhibitory Homodimerization," *Struct. Des.*, vol. 25, no. 9, pp. 1325-1336.e3, 2017.
[213] S. K. Olsen *et al.*, "Structural basis by which alternative splicing modulates the organizer activity of FGF8 in the brain," pp. 185–198, 2006.
[214] A. Brown, L. E. Adam, and T. L. Blundell, "The crystal structure of fibroblast growth factor 18 (FGF18)," *Protein Cell*, vol. 5, no. 5, pp. 343–347, 2014.
[215] H. P. Makarenkova *et al.*, "Differential interactions of FGFs with heparan sulfate control gradient formation and branching morphogenesis," *Sci. Signal.*, vol. 2, no. 88, pp. 1–11, 2009.
[216] A. A. Belov and M. Mohammadi, "Molecular mechanisms of fibroblast growth factor signaling in physiology and pathology.," *Cold Spring Harb. Perspect. Biol.*, vol. 5, no. 6, pp. 1–24, 2013.
[217] C. Sun, D. G. Fernig, D. Mason, M. Marcello, Y. Li, and R. Lévy, "Selectivity in glycosaminoglycan binding dictates the distribution and diffusion of fibroblast growth factors in the pericellular matrix," *Open Biol.*, vol. 6, no. 3, p. 150277, 2016.
[218] R. Xu *et al.*, "Diversification of the structural determinants of fibroblast growth factor-heparin interactions: Implications for binding specificity," *J. Biol. Chem.*, vol. 287, no. 47, pp. 40061–40073, 2012.
[219] Y. Liu, J. Ma, A. Beenken, L. Srinivasan, A. V. Eliseenkova, and M. Mohammadi, "Regulation of Receptor Binding Specificity of FGF9 by an Autoinhibitory Homodimerization," *Structure*, vol. 25, no. 9, pp. 1325-1336.e3, 2017.
[220] I. Sher, B. K. Yeh, M. Mohammadi, N. Adir, and D. Ron, "Structure-based mutational analyses in FGF7 identify new residues involved in specific interaction with FGFR2IIIb," *FEBS Lett.*, vol. 552, no. 2–3, pp. 150–154, 2003.
[221] Y. Anandappa, J. A. Smith, and C. Wilkinson, "High-level production of human acidic fibroblast growth factor in E. coli cells : inhibition of DNA synthesis in rat mammary fibroblasts at high concentrations of growth factor," *Biochem. Biophys. Res. Commun.*, vol. 171, no. 3, 1990.
[222] G. B. Ten Dam *et al.*, "3-O-sulfated oligosaccharide structures are recognized by anti-heparan sulfate antibody HS4C3," *J. Biol. Chem.*, vol. 281, no. 8, pp. 4654–4662, 2006.
[223] J. Schlessinger *et al.*, "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.
[224] T. H. van Kuppevelt, M. A. B. A. Dennissen, W. J. van Venrooij, R. M. A. Hoet, and J. H. Veerkamp, "Generation and Application of Type-specific Anti-Heparan Sulfate Antibodies Using Phage Display Technology," *J. Biol. Chem.*, vol. 273, no. 21, pp. 12960–12966, 2002.
[225] S. M. Thompson *et al.*, "Heparan Sulfate Phage Display Antibodies Identify Distinct Epitopes with Complex Binding Characteristics," *J. Biol. Chem.*, vol. 284, no. 51, pp. 35621–35631, 2009.
[226] C. Sun, Y. Li, S. E. Taylor, X. Mao, M. C. Wilkinson, and D. G. Fernig, "HaloTag is an effective expression and solubilisation fusion partner for a range of fibroblast growth factors," *PeerJ*, vol. 3, p. e1060, 2015.
[227] D. Ron, D. P. Bottaro, P. W. Finch, D. Morris, J. S. Rubin, and S. A. Aaronson, "Expression of biologically active recombinant keratinocyte growth factor. Structure/function analysis of amino-terminal truncation mutants," *J. Biol. Chem.*, vol. 268, no. 4, pp. 2984–2988, 1993.
[228] M. Mohammadi, D. M. Ornitz, O. A. Ibrahimi, H. Umemori, S. K. Olsen, and X. Zhang, "Receptor Specificity of the Fibroblast Growth Factor Family: The complete mammalian FGF family," *J. Biol. Chem.*, vol. 281, no. 23, pp. 15694–15700, 2006.
[229] J. M. Trowbridge, J. A. Rudisill, D. Ron, and R. L. Gallo, "Dermatan sulfate binds and potentiates activity of keratinocyte growth factor (FGF-7)," *J. Biol. Chem.*, vol. 277, no. 45, pp. 42815–42820, 2002.
[230] C. Sun, Y. Li, E. A. Yates, and D. G. Fernig, "SimpleDSFviewer: a tool to analyse and view differential scanning fluorimetry data for characterising protein thermal stability and interactions," *PeerJ Prepr.*, vol. 3, p. e1937, 2015.
[231] D. G. Fernig, J. A. Smith, and P. S. Rudland, "Appearance of basic fibroblast growth factor receptors upon differentiation of rat mammary epithelial to myoepithelial-like cells in culture," *J. Cell. Physiol.*, vol. 142, no. 1, pp. 108–116, 2005.
[232] A. Sarkar and U. R. Desai, "A simple method for discovering druggable, specific glycosaminoglycan-protein systems. Elucidation of key principles from heparin/heparan sulfate-binding proteins," *PLoS One*, vol. 10, no. 10, 2015.