Order out of complexity – protein structures that interact with heparin
Barbara Mulloy* and Robert J Linhardt†

Many proteins of widely differing functionality and structure are capable of binding heparin. Structural characterisations of the many types of such complexes are being reported in ever-increasing number and at improved resolution. Several crystal structures of complexes formed through the interaction of heparin-derived oligosaccharides with one or more protein partners have been described.

Addresses
*Laboratory for Molecular Structure, National Institute for Biological Standards and Control, Blanche Lane, South Mimms, Potters Bar, Hertfordshire EN6 3QG, UK; e-mail: bmulloy@nibsc.ac.uk.
†Department of Chemistry, Division of Medicinal Chemistry and Department of Chemical and Biochemical Engineering, The University of Iowa, 115 South Grand Avenue, PHAR S328, Iowa City, Iowa 52242-1112, USA; e-mail: robert-linhardt@uiowa.edu.

0959-440X/01/$ – see front matter © 2001 Elsevier Science Ltd. All rights reserved.

Abbreviations
FGF fibroblast growth factor
FGFR fibroblast growth factor receptor
FMDV foot and mouth disease virus
HS heparan sulfate
PDB Protein Data Bank
RANTES regulated on activation normal T-cell expressed and secreted
serpin serine protease inhibitor

Introduction
The glycosaminoglycan heparan sulfate (HS) occurs on the surface of most cells as proteoglycan sidechains [1,2]. This linear polysaccharide consists of simple disaccharide repeats of alternating uronic acid and glucosamine, made complex by variations of substitutions with O- and N-sulfo and N-acetyl groups, as well as by epimerisation of the uronic acid (Figure 1). Heparin, a bulk pharmaceutical, can be described as a highly sulfated type of HS found in mast cell granules. As it is usually the sulfated sequences of HS that interact with proteins, heparin is often used as a model compound for HS.

Without heparin, mast cell granules do not carry their usual range of serine proteases [3*,4*]. Its ability to activate the serine protease inhibitors (serpins), such as antithrombin, also makes heparin a potent anticoagulant. HS also interacts with an extraordinary range of proteins [5*] and has been described as a selective regulator of ligand–receptor encounters [6–8].

Methods for the determination of heparin/HS sequences (e.g. those recently described by Drummond et al. [9] and Shriver et al. [10]) are outside the scope of this review. The concept of specificity, as applied to HS, needs some explanation. Although HS–protein interactions are not exclusively ionic in nature, the large number of strongly acidic sulfate substituents play a dominant role. For many heparin-binding proteins, provided the required sulfo groups are present, extra sulfo groups do not reduce affinity (with occasional exceptions [11]). This is why heparin makes a useful model; its complexes with proteins can be studied in detail before the exact structure of the ‘real’ endogenous HS ligand is elucidated. In crystal structures of the heparin–protein complexes studied so far, the heparin/HS-binding site is occupied by fully sulfated heparin sequences, as in Figure 1a, with the single exception of the antithrombin complex with a synthetic pentasaccharide related to that shown in Figure 1c [12].

There is a great deal of current interest in the study of heparin-binding sites from a wide range and increasing number of proteins. The definition of consensus heparin-binding sites in proteins has been reviewed and extended by Hileman et al. [13]. Although these consensus sequences are not invariably correlated with heparin-binding sites, the concept is useful, especially in the absence of a three-dimensional protein structure.

Where an experimental or a well-founded theoretical three-dimensional structure for the protein of interest is available, it is possible to model heparin–protein complexes using systematic docking calculations. Bitomsky and Wade [14] have suggested a procedure using a disaccharide probe model for heparin, followed by refinement of the model complex using molecular mechanics. Heparin is a well-behaved ligand for this kind of calculation and the site on a protein surface where heparin binds can be identified with good reliability. The conformation of heparin appears to be unusually well defined compared with other polysaccharides: in both solution studies of unbound heparin and crystal structures of complexes, heparin is a ribbon-like molecule with sulfo groups arranged in clusters along opposite sides of the polysaccharide chain [15*]. At the present time, however, the orientation of heparin/HS in its binding site and the exact details of the interactions with proteins cannot be predicted by docking calculations.

The use of NMR spectroscopy, coupled with evidence from site-directed mutagenesis, binding studies and measurements of biological activity, can enable the detailed characterisation of heparin-binding sites on proteins. Such methods have been used recently in the study of the CXC group of chemokines, for example, interleukin-8 [16], platelet factor-4 and growth-related protein-α [17], stromal cell derived factor-1α [18*] and RANTES (regulated on activation normal T-cell expressed and secreted) [19**].
It is now possible to compare heparin-binding sites across this structurally and functionally related group of small proteins [19••].

However, crystallographic structures are the only way, at present, of examining heparin–protein complexes in atomic detail for both partners. So far, the structures of heparin complexes with four contrasting types of protein have been studied: antithrombin, annexin V, the fibroblast growth factors and their receptors, and virus capsid proteins.

**Antithrombin**

So far, the best-characterised protein–heparin complex comprises an unusual heparin pentasaccharide (Figure 1c) bound to the serpin antithrombin, thus increasing its ability to inhibit coagulation. The specificity of the interaction and the resulting conformational change in antithrombin have been described in detail. A useful recent summary of this work is found in the article by Whisstock et al. [20•].

Following the determination of the X-ray structure of antithrombin complexed with a synthetic heparin pentasaccharide [12], recent papers have described the heparin-binding site in more detail. Work with a synthetic heptasaccharide has defined an extended heparin-binding site necessary to accommodate full-length heparin [22].

**Annexin V**

Annexin V is an anticoagulant, calcium- and membrane-binding protein. A recently determined crystal structure of annexin V in complex with heparin tetrasaccharides [23••] reveals two distinct heparin-binding sites. Annexin V has a disc-like shape; one face carries calcium-binding loops and interacts with phospholipid membranes, and the opposite, slightly concave side faces outward. One heparin-binding site (HTS-1) is formed by two of the calcium-binding loops, on the membrane side, and the other (HTS-2), on the outward-facing side, includes the N terminus and nearby loop and helix regions (Figure 2). Contacts between the heparin tetrasaccharide and HTS-2 involve only 0-sulfo groups, with N-sulfo groups oriented away from the protein; at

---

**Figure 1**

The main repeating disaccharides in (a) heparin (R\(^1\), R\(^2\), R\(^3\) = SO\(_3\)) and (b) HS (R\(^1\), R\(^2\) = H, R\(^3\) = Ac), and (c) the naturally occurring pentasaccharide sequence with high affinity for antithrombin. Over 70% of heparin consists of the disaccharide shown in (a). HS contains predominantly the disaccharide shown in (b), with short sequences in which iduronate residues and N- and O-sulfo groups occur: HS binds proteins through these regions.

---

**Figure 2**

Stereo view (from PDB code 1G5N) of annexin V (cyan ribbon) in complex with calcium (yellow spheres) and two heparin tetrasaccharides (stick diagrams) [23••]. The two heparin-binding sites are on opposite faces of annexin; site HTS-1 is on the phospholipid-membrane-binding face and so cannot be occupied when annexin V is bound to a cell surface. The other site, HTS-2, faces outward. Heparin binding at both sites is calcium-dependent. Reproduced with permission from [23••].
site HTS-1, only three sulfo groups — one 2-O-sulfo, one 6-O-sulfo and one N-sulfo — are involved.

### The fibroblast growth factors (small, β-trefoil proteins)

The X-ray structures of fibroblast growth factor-1 (FGF-1) [24] (Figure 3a) and FGF-2 [25] (Figure 3b) in complex with heparin oligosaccharides have been solved; the heparin-binding sites on the two proteins are at the same relative location. The position of the heparin-binding site on FGF-1 has been confirmed by solution NMR studies [26] and by the observation that iodination at Cys131, near the binding site, destroys the heparin affinity [27]. The FGF-1–heparin complex crystallised with two FGF-1 molecules bound to the same heparin decasaccharide, on opposite sides of the polysaccharide chain (Figure 3a).

Keratinocyte growth factor, also known as FGF-7, has a similar fold (β trefoil) to both FGF-1 and FGF-2. Evidence from synthetic peptide studies puts the heparin-binding site in a similar position to those of FGF-1 and FGF-2 [28] (Figure 3c). One of the several possible heparin-binding sites in a recently reported crystal structure of FGF-9 is equivalent to the FGF-1 or FGF-2 site [29] (Figure 3d).

Faham et al. [30], in this series of reviews three years ago, discussed the interactions of heparin and HS with FGF-1 and FGF-2 in some detail. At that time, it was clear that heparin could potentiate the formation of dimeric complexes of FGF with FGF receptors (FGFRs), but it was not obvious whether heparin induced FGF dimerisation, which in turn induced FGFR dimerisation, or whether heparin simultaneously bound to both FGF and FGFR. Several molecular modelling studies were undertaken to examine these questions [31–34]. Additional complication arises from the various forms of FGFR molecule, for which there are four identified genes and several isoforms.

Ternary structures of FGF-1–FGFR2–heparin [35••] and FGF-2–FGFR1–heparin [36••] have been solved, both using a decasaccharide fragment of heparin (Figure 4). Crystal structures of FGF–FGFR complexes in the absence of heparin [37,38] suggested a dimeric arrangement in which a positively charged canyon, formed when FGF binds at the junction between two FGFR subunits, was assumed to be the heparin-binding site. Soaking a heparin decasaccharide into preformed crystals of the dimeric FGF-2–FGFR1 complex [36••] (Figure 4a) located the heparin fragments in the binding canyon and revealed an unexpected stoichiometry. Where a single heparin decasaccharide was expected to occupy the canyon, two decasaccharides were found end-to-end in each; for each decasaccharide, the six nonreducing end monosaccharide units were in contact with protein, interacting with both FGF and FGFR. Essential heparin substituents for binding to FGF-2 were previously shown to include the N-sulfo groups on the glucosamine residues, as well as iduronate carboxylate and 2-O-sulfo groups [25,30]. Additional structural requirements for mitogenesis have been found to include a minimum length of a decamer and at least one 6-O-sulfo group in a study by Pye et al. [39]. The FGF-2–FGFR1–heparin ternary structure would seem to indicate that a hexasaccharide is the minimum length and that all the sulfo and carboxyl substituents are involved, to some extent, in binding to FGF-2 or FGFR-1. However, Pye et al. were working with low sulfated HS oligosaccharides and their samples may not have contained any fully sulfated hexasaccharide sequences.

In contrast, crystallisation of a preformed ternary complex of FGF-1–FGFR2–heparin [35••] (Figure 4b) gave a different
and unexpected geometry, with two 1:1 FGF–FGFR binary complexes associated through a single, asymmetric heparin molecule. Heparin does not occupy the predicted site [37]. The decasaccharide binds to only one of the FGFR molecules in the heteropentamer, but to both FGF-1 molecules. The two FGF-1 molecules sit opposite one another across the heparin chain, rather like as in the heparin–FGF-1 crystal structure [24], except that one FGF-1 copy is rotated through 120°. The only protein–protein contact between the two halves of the dimer is a limited area of interaction between the two FGFR molecules. As well as glucosamine N-sulfo and iduronate 2-O-sulfo groups, glucosamine 6-O-sulfo groups of heparin are involved in binding to FGF-1 [24]; in the FGF-1–FGFR2 ternary structure, all the sulfo and carboxyl groups of the heparin are involved in interactions.

The two ternary structures described above are very different in their overall geometry (Figure 4). They were prepared by different methods and their protein constituents are different. In both of them, however, the heparin–FGF interactions have a stoichiometry and geometry that are similar to those in crystals of heparin–FGF alone. These two structures do not indicate any close prospect of agreement on a single universal architecture for FGF–FGFR–heparin complexes.

Viral proteins
HS acts as a cell surface receptor for many viruses and the interaction between virus surface proteins and heparin has been partially characterised in several cases; pseudorabies glycoprotein C [40] and gp120 of HIV [41] are recent examples. Fry et al. [42••] have determined an X-ray structure of the complex between foot and mouth disease virus (FMDV) and heparin. In this complex, the heparin-binding site consists of contributions from all three major capsid proteins, VP1, VP2 and VP3 (Figure 5). The shallow depression that comprises this site is composed of contributions from strand B1 and a loop of VP3, the C terminus of VP1, and the αB helix of VP2; at the base of the site is a 3_{10} helix in VP3. This structure provides an illustration of the point that heparin-binding sites may not be contiguous in sequence or even composed of a single peptide chain, and that any secondary structural motif may play a part.

Although a heterogeneous sample of low molecular weight heparin was used, a particular sulfated trisaccharide sequence (in which two 6-O-sulfo, N-sulfo glucosamine residues flank a central 2-O-sulfo iduronate) occupied the binding site.
Conclusions
Of the large number of heparin–protein complexes that can be experimentally demonstrated, less than a dozen have been determined to atomic resolution. For a much larger number of interactions, some intermediate characterisation of the complex has been achieved. How far is it possible, at this stage, to derive general principles for protein–heparin interactions? The sheer number and variety (in terms of both structure and function) of heparin-binding proteins make classification a difficult task. Heparin-binding sites show no absolute dependency on specific sequences or protein folds. Some similarities can, however, be seen among heparin-binding sites within families of proteins and, as more and better structures are determined, a degree of order can be expected to emerge from the current apparent complexity. Heparin and HS, along with the other glycosaminoglycans, are old molecules, inhabiting the extracellular spaces of animals from marine invertebrates to mammals [43]. All functional proteins outside the cell, whether involved in structure, growth, communications, adhesion or development, have evolved in the presence of these polysaccharides. Thus, the many types of protein interactions with glycosaminoglycans undoubtedly reflect their important roles in a wide range of physiological and pathophysiological processes.

Update
Recently, the NK1 (N-terminal and first kringle domains) splice variant of hepatocyte growth factor has been crystallised in complex with a heparin oligosaccharide [44*]. The heparin-binding site has residues in common with earlier predictions by Zhou et al. [45] on the basis of NMR evidence.

Acknowledgements
We are grateful to E. Fry for her help with Figure 5 and to MJ Forster for comments on the manuscript. This work was funded in part by a grant from the US National Institutes of Health (HL 52622).

References and recommended reading
Papers of particular interest, published within the annual period of review, have been highlighted as:
• of special interest
•• of outstanding interest
20. A combination of NMR and biochemical techniques was used to identify the heparin-binding site of RANTES. In addition, the paper collates and illustrates published data showing that, for the IL-8-like chemokines, the heparin-binding sites occupy similar positions in the protein fold.
A particularly clear account of the conformational change in antithrombin that is brought about by binding to a synthetic heparin pentasaccharide, including studies of mutant antithrombins.


Structural and functional studies of mutants, the results suggest a model in which cell surface HS might localise or stabilise annexin V within an anticoagulant molecular assembly at the cell surface.