Multiprotein signalling complexes: regional assembly on heparan sulphate

J.T. Gallagher
CRUK Department of Medical Oncology, University of Manchester, Paterson Institute for Cancer Research, Wilmslow Road, Manchester M20 4BX, U.K.

Abstract
Heparan sulphate (HS) is an abundant component of cell surfaces and the extracellular matrix. It binds to a wide variety of peptide growth factors, morphogens, chemokines and extracellular matrix proteins (e.g. fibronectin) and many of these interactions are essential for these effector proteins to transduce signals across the plasma membrane. The unique molecular design and flexibility of HS are essential for its ability to exert control over the cellular response to proteinaceous ligands. The clustering of sulphated sugar residues in a series of complex domains with variable sulphation patterns generates considerable diversity in the molecular fine structure of HS. This diversity reflects a high degree of selectivity in protein recognition and in the assembly of functional multiprotein complexes on the HS polymer chain.

The biosynthesis of HS (heparan sulphate)
HS is a member of the GAG (glycosaminoglycan) family of polysaccharides and as such it has a recognizable repeating disaccharide unit structure of the general form (A–U)n, where A is an amino sugar and U a uronic acid; both monosaccharides can be modified by sulphaion [1] (Figure 1). In HS, the amino sugar, glucosamine, is N-acetylated or N-sulphated and the uronic acid is present as one of two alternative isomers, GlcA or its C-5 epimer IdoA (iduronic acid). HS is initially synthesized as a non-sulphated polymer of GlcNAc1–4GlcAβ1–4 repeats named heparan or N-acetyl-heparanase that is then modified by N- and O-sulphation and conversion of GlcA into IdoA [2]. The transformation of heparan to HS begins with the deacetylation and N-sulphation of clusters of glucosamine residues that sets up the basic domain structure of HS (see below). The enzymes that bring about this key initial modification are dual-specificity enzymes called the NDSTs (N-deacetylase/ N-sulphotransferase) that incorporate both N-deacetylase and N-sulphotransferase activities in a single protein [3]. It is not clear how the NDSTs target particular sections of the heparan chain for modification by N-sulphation. The subsequent conversion of GlcA into IdoA by an HS C-5-epimerase is restricted to the newly formed N-sulphated regions, whereas GlcA remains associated with the unmodified, N-acetyl-rich areas of the chain. Further modifications of HS take the form of ester-linked sulphation (O-sulphates) mainly at C-2 of IdoA and C-6 of the amino sugars [4]. The C-5 epimerase and 2-O-sulphotransferase enzymes are tightly coupled and modify HS mainly in regions where N-sulphated disaccharides form continuous sequences. Both GlcNS (N-sulphated glucosamine) and GlcNAc are substrates for the sulphotransferases that catalyse C-6 O-sulphation but an important constraint is that the modified GlcNAc residues are always adjacent to an N-sulphated disaccharide. The constraint occurs because the 6-OSTs (C-6 O-sulphotransferases) act in and around regions of N-sulphation; the 3-OSTs that effect the final and rare addition of sulphate groups to C-3 of GlcNS are also restricted in their action, targeting mainly regions of high N-sulphation. The 3-OST, 6-OSTs and NDST enzymes each comprise multigene families that differ, sometimes in very subtle ways, in their substrate preferences; the differential expression of these enzymes in embryonic and adult cells probably gives rise to much of the variation in sulphation observed in the HS family [2,5].

The domain structure of HS
The sulphated regions of HS are distributed in a fairly uniform manner along the polymer chain; two distinct sequences, each with repetitive elements, can be recognized in the so-called ‘composite sulphated regions’ or CSRs [6]. The most heavily modified sequences in the CSRs are the S-domains (sulphated domains) composed of contiguous N-sulphated disaccharides with IdoA-2-O-sulphate as the major uronate component [7]. The GlcNS-IdoA-2S repeats extend from 2 to 7/8 disaccharides in length though the majority are at the lower end of this size range. These S-domains are flanked by alternating sequences of N-sulphated and N-acetylated disaccharide units that form T zones (transition zones) between the S-domains and the largely unmodified N-acetylated regions (Figure 2). C-6 sulphation of the amino sugars is present in both S-domains and T zones and, as stated...
The two basic disaccharide repeating units in HS contain N-acetylated or GlcNS and uronic acid (structures a and b respectively). The uronate linked to C-1 of GlcNAc is always GlcA, whereas the C-5 epimer IdoA predominates at C-1 of GlcNS. In the N-sulphated disaccharides, O-sulphation is common at C-2 of IdoA and C-6 of the amino sugar, GlcNAc may also be sulphated at C-6 but only in T zones where it is adjacent to an N-sulphated disaccharide. Other minor but important variations on the illustrated structures are sulphation at C-3 of GlcNS and C-2 of GlcA, and the presence of N-unsubstituted glucosamine (GlcNH$_3^+$).

The 6-O-sulphates are subject to removal by endo-6-sulphatases located on the cell surface. The ‘editing’ function of these enzymes has a profound effect on the binding properties of HS, because 6-sulphates are a common functional group for protein recognition [8,9]. HS also contains a small number of N-unsubstituted GlcN residues that occur in all regions of the chain [10]. The gD (glycoprotein D) virus coat protein of HSV (herpes simplex virus) recognizes S-domain type sequences that contain a single 3-O-sulphated GlcNH$_3^+$ residue and binding to its high-affinity sequence in cell surface HS triggers the fusogenic action of the gD protein that enables the virus to enter cells [11].

**Isolating the bioactive regions**

We have a reasonable idea of the size of the CSRs of HS from the use of the enzyme K5 lyase. This enzyme resides in the tail region of coliphage K5 where it can degrade the protective ‘heparan’ coat of susceptible *Escherichia coli* strain K5 [12]. Recombinant K5 lyase only attacks HS in the non-sulphated regions; N-sulphation blocks the action of the enzyme and as a result the CSRs [i.e. the S-domains in combination with the NA/NS (N-acetylated/N-sulphated) domains] resist enzyme action and are effectively excised from the GAG chain (Figure 2). The CSRs are approx. 14–16 disaccharides in length [6,13]. It seems reasonable to assume that they represent the functional units in HS but so far they have received little attention. It was recently shown that VEGF$_{165}$ (where VEGF is vascular endothelial growth factor) binds efficiently to these regions in endothelial HS, whereas even the largest S-domains from the same polymer fail to meet the binding requirements of a naturally occurring VEGF$_{165}$ dimer [13].

Most studies to date on HS fragments have been carried out on the S-domains that can be isolated with the enzyme heparinase III (or heparitinase) that degrades the N-acetylated and T zone sequences but is inhibited by IdoA$_{2S}$ [14]. Hydrazinolysis and high pH nitrous acid are also an effective way to isolate S-domain type sequences [15]. S-domains are biologically active structures and depending on their length and sulphation they bind efficiently to many growth factors and extracellular matrix proteins. However, *in situ*, their activities are likely to be influenced by the local environment in the polymer chain (see below) and for this reason it will be desirable in future to carry out complementary binding and activity studies with isolated CSRs and with intact HS.

**Solution structures of HS and heparin: implication for protein assemblies**

An appreciation of the conformation of HS saccharides and the disposition of sulphate groups along the sugar backbone is important for elucidating the mode of action of HS in cell regulation and other activities. Although no direct work has been done with HS, the solution structure of heparin as revealed by NMR spectroscopy is that of a well-defined helix with sulphate groups in the repeating disaccharides positioned in an offset manner on both sides of the helical
axis [16]. The S-domains of HS are similar in structure to heparin and therefore they probably adopt a heparin-like conformation. This means that proteins are able to bind on opposite faces of the helix and in this way relatively large protein assemblies can form on a relatively short saccharide sequence. Crystal structures of a 2:2:1 FGF1–FGFR2–heparin complex (where FGF is fibroblast growth factor and FGFR is FGF receptor) published by Pellegrini et al. [17] (and discussed in detail by Harmer [18] in this issue) reveal the importance of the two-sided arrangement of sulphate groups in heparin. In the Pellegrini crystal structure [17], two FGF1 ligands bind on opposite sides of a heparin saccharide in a trans-dimer arrangement with two receptors docked on to each of the FGF1 monomers; one receptor is also in contact with heparin. The heparin fragment size was only five disaccharide units, well within the size range of the S-domains in HS.

Recent solution studies using high-resolution gel chromatography to evaluate the molecular size of FGF–FGFR–heparin complexes suggest that the initial event in FGF signalling, at least in the case of FGF1, may be driven by the co-operative binding of two FGF1 monomers to a single heparin fragment [19]. This ‘dimeric’ form of FGF1 then efficiently recruits two receptors into a highly stable putative signalling assembly. Ligand–receptor interactions were not detectable in the absence of heparin [19], indicating that in addition to creating a bivalent FGF1 ligand (presumed to be in the trans-dimer arrangement described by Pellegrini [17]), the heparin may also induce a conformational change in FGF1 that significantly enhances receptor affinity. Further structural studies are now needed using bioactive S-domains of HS to test the validity of the heparin analogue in the Pellegrini model of FGF signalling. The alternative Schlessinger–Mohammadi symmetrical model of a 2:2:2 FGF2–FGFR1–saccharide complex also used a heparin fragment as the GAG component [20]. In the latter case protein–protein interactions predominate and the role of saccharides is primarily to bind and stabilize each ligand–receptor combination in a highly ordered multicomponent assembly. In contrast, in the Pellegrini model [17], heparin is at the core of the complex and heparin–protein interactions dictate the mode of assembly and architecture of the ligand–receptor pairs.

The potential significance of the domain structure of HS

The question constantly arises of why the N-sulphated GAG on cell surfaces (i.e. HS) has its sulphated sugars arranged in complex clusters in contrast with the uniform, high-density sulphation present in heparin. Perhaps the simplest explanation is that the regulated and differential sulphation patterning of HS allows for discrimination in binding to its many client proteins. This variable patterning is likely to be very important in the regulation of cell function and turnover in adult tissues and in the constantly changing environment of the developing embryo where many critical decisions affecting cell growth and migration are determined by heparin/HS-binding cytokines.

However, the hypervariable sulphation in HS domains has other implications. It has been shown that for some species of mammary cell HS, the polymer chain suppresses the mitogenic potential (i.e. induction of FGF activation) of its constituent S-domains [21]. It is not clear how this suppressive effect is achieved. Moreover, provocative evidence has been published to suggest that the most favoured sites in HS for binding FGF ligands may not correspond to the ‘active sites’ where signalling complexes form [22,23]. If this is the case, it follows that ligands could migrate along the HS chain until they find domains that bind and orientate them in an appropriate manner for efficient receptor recognition. This dynamic concept of mobile ligands sampling a repertoire of binding regions in the HS polymer network around cells provides for an additional level of control on both cell signalling and on other important aspects of HS function, especially its critical role in the formation of morphogen gradients during embryogenesis [5,24].

Concluding remarks

HS is designed to bind effector proteins and to regulate their biological activities. The sulphation patterns are the principal determinant of binding specificity but variations in the uronate isomer and the presence of N-acetylated and N-unsubstituted amines amplify the structural diversity of HS and its potential for highly selective protein interactions. More information is needed on the relationship between the HS fine structure and the expression of genetically distinct forms of the polymer modifying enzymes. At present, we have no clear ideas about how the N-unsubstituted amines are formed nor how the NDST enzymes establish the basic domain structure that so clearly distinguishes HS from heparin and other GAGs. The availability of enzymatic and chemical methods for excision of the sulphated regions of HS will enable close examination of the validity of the heparin analogue in ligand–receptor binding and will also allow the issue of specificity of active site sequences to be addressed, not only in the extensive FGF system but for many other ligand–receptor combinations, for example the VEGF/neuropilin/VEGFR (where VEGFR is VEGF receptor) [13,25] and HGF/Met systems [26,27]. It is intriguing that the structure of HS produced by neural cell precursors undergoes a change in sulphation as the cells become less sensitive to FGF2 and responsive to FGF1 [28]. We can anticipate more studies of this kind on tractable developing systems (e.g. embryonal stem cell cultures) at times when they are known to be influenced by FGFs and other HS-dependent cytokines. It will be interesting to see if during cell differentiation HS species are produced with a distinctive sulphation imprint that finely tunes the signalling response to specific growth factors, morphogens etc.

I thank all my colleagues in the Proteoglycan Group at The University of Manchester for their contributions to the work described here.
and Professor Tom Blundell and Dr Nic Harmer at The University of Cambridge for their invaluable help with the FGF studies. Cancer Research UK generously supported this work.

References

Received 20 December 2005