Cytokines and proteoglycans: an introductory overview

B. Mulloy* and C.C. Rider†

*Laboratory for Molecular Structure, National Institute for Biological Standards and Control, Blanche Lane, South Mimms, Potters Bar, Herts. EN6 3Qc, U.K., and †School of Biological Sciences, Royal Holloway University of London, Egham Hill, Egham, Surrey TW20 0EX, U.K.

Abstract

The defining characteristic of the glycoproteins known as proteoglycans is the presence of O-linked acidic polysaccharides known as GAGs (glycosaminoglycans). The backbone of these linear polysaccharides is a repeating disaccharide, comprising N-acetyl hexosamine alternating with β-d-glucuronic acid, α-l-iduronic acid, or galactose. For some GAGs, partial deacetylation, epimerization of glucuronic acid, and substitution with N- and O-sulphates result in highly complex, heterogeneous structures. The interactions with proteins through which GAGs exert their biological effects depend on the resulting sequences. Some proteins, for example antithrombin, have highly specific sequence requirements for their GAG ligand [in this case heparin or HS (heparan sulphate)]; others, for example the fibroblast growth factors, are less demanding. GAGs, in particular HS, play a role as co-receptors for some cytokines. In addition, HS is thought to be important for the localization of cytokines, acting both as a tissue store and as a mediator of morphogen gradient formation in development. The structural determinants of GAG-cytokine interactions are therefore clearly important to understanding the biology of development, wound healing and the immune system. No single paradigm has been identified for such interactions, and the search for general principles underlying involvement of GAGs in cytokine function is at an early stage.

The structure and biosynthesis of proteoglycans and GAGs (glycosaminoglycans)

Proteoglycans make up a relatively small subset of glycoproteins defined by their possession of long, unbranched and highly acidic chains of polysaccharide called GAG. GAG structures are based on a disaccharide repeat, and there are four classes of GAG, each distinguished by a particular repeating disaccharide (see Table 1). Thus the HS (heparan sulphate) and heparin class of GAG is based on a repeat disaccharide of glucuronic acid-(β1-4)-N-acetylglucosamine-(α1-4). However, during biosynthesis in the Golgi, even while the GAG chain is still being elongated, the incorporated disaccharides are subject to a set of complex modifications (see Figure 1A; previously reviewed in [1]). For the HS/heparin class, the glucuronic residues may be epimerized at C-5 to yield iduronic acid, and sulphated at the 2-O- position. Since neither of these modifications is comprehensive, uronic acid residues in HS/heparin chains are found in all four possible combinations of non-epimerized/epimerized and unsubstituted/2-O-sulphation states. The N-acetylg glucosamines are modified to show even more variability. Here, in what is actually the first step in the biosynthetic modification pathway, the amino group may be deacetylated, and subsequently sulphated. The N-deacetylation and N-sulphation reactions are closely coupled functions of a single bifunctional enzyme, N-acetylg glucosamine N-deacetylase/N-sulphotransferase. Nonetheless, unsubstituted free amino groups do occur, albeit relatively rarely. Thus the amino moiety of the glucosamine can occur in three alternative states: N-acetylated, N-sulphated and free amino. The glucosamine is then variably subjected to sulphation at 6-OH and, rather
Table 1 | The disaccharide repeats of the different classes of GAGs, shown prior to post-incorporation modification

In the case of the sulphated GAGs, trisaccharide linkage sequence at their non-reducing ends of two tandem galactose residues linked to a terminal xylose.

<table>
<thead>
<tr>
<th>GAG class</th>
<th>Basic disaccharide repeat</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hyaluronic acid (hyaluronan)</td>
<td>d-glucuronic acid-(β1-3)-N-acetyl-d-glucosamine (hyaluronan)</td>
</tr>
<tr>
<td>CS/DS</td>
<td>d-glucuronic acid-(β1-3)-N-acetyl-d-galactosamine</td>
</tr>
<tr>
<td>Keratan sulphate</td>
<td>d-galactose-(β1-4)-N-acetyl-d-glucosamine</td>
</tr>
<tr>
<td>HS/heparin</td>
<td>d-glucuronic acid-(β1-4)-N-acetyl-d-glucosamine</td>
</tr>
</tbody>
</table>

Figure 1 | Biosynthetic elaboration of heparin and heparan sulphate sequences

(A) The variable post-incorporation reactions that modify the disaccharide repeat of HS/heparin are catalysed by the following biosynthetic enzymes: NDST, N-deactylase/sulphotransferase; 2OST, 2-O-sulphotransferase; 6OST, 6-O-sulphotransferase. The epimerase acts at C-5 to modify β-d-glucuronic acid (GlcA) to α-L-iduronic acid (IdoA). (B) The tri-sulphated disaccharide typical of heparin. (C) The pentasaccharide structure that confers high affinity for antithrombin on a heparin molecule containing it. The 3-O-sulphate is added to the central glucosamine of this sequence by a 3-O-sulphotransferase which is specific for N-sulphated glucosamine with glucuronic acid at its non-reducing side. Asterisks indicate features of the pentasaccharide essential for affinity to antithrombin.

Heparin and HS

Heparin, widely used as an anticoagulant, is a highly sulphated variant of HS (rich in the trisulphated disaccharide shown in Figure 1B), which is localized in the secretory granules of mast cells, and released during inflammatory degranulation reactions. Because of its ready commercial availability, heparin is widely used as a model agent in the experimental study of interactions which in a physiological context would probably involve HS. HS is expressed in almost all cell types, and by virtue of being attached to various glycoprotein core polypeptides, it is found both on the cell surface and secreted into the extracellular matrix. In general, HS is less sulphated than heparin, but there is considerable variation in the extent of modification, including sulphation, between HS from different tissues. Some HS is highly N- and O-sulphated, and therefore very heparin-like, whereas HS from other sources may have low levels of epimerization and sulphation. Indeed, within a single chain of HS, there are regions of high sulphation, so-called S-domains, interspersed with domains of low modification, N-domains [3]. In heparin and highly sulphated HS, the S-domains are large and encompass the N-domains which become very small. In low sulphated HS, the reverse is true. The existence of this domain patterning is evidence for organization rather than randomness in the biosynthetic machinery of HS/heparin modification.
Is protein–GAG binding specific?

Given that GAGs, especially HS/heparin, are strongly acidic polymers with high charge densities, an important question is the extent to which protein–GAG interactions are specific, rather than the outcomes of non-specific ionic interactions. Clearly in the case of those other highly negatively charged, linear, sugar-based biological polymers, the nucleic acids, a high level of specificity is evident in the binding of restriction endonucleases, transcription factors and other proteins. But how much sequence specificity is involved when proteins bind to GAGs? The nucleic acids as informational macromolecules are synthesized by high-fidelity processes involving the copying of templates. Can GAGs, which are biosynthesized without template sequences, also be informational macromolecules engaged in highly sequence-specific interactions? Or is the extensive sequence variation in HS/heparin merely biosynthetic ‘noise’ with no signal content?

Hard evidence of an exquisite oligosaccharide sequence specificity in a GAG–protein interaction first emerged from the now classic studies of Lindahl et al. [4] on the high-affinity binding of antithrombin III to heparin, a key interaction in the potent anticoagulant activity of heparin. Here, the binding site for antithrombin III is a specific penta-saccharide centred on an unusual 3-O-sulphated N-sulphated glucosamine [4] (Figure 1C). Extensive studies using synthetic oligosaccharides have shown that very little sequence variation is permissible without loss of high-affinity binding. However, to what extent is such specificity typical of protein–GAG interactions in general? The prototypic example of an HS/heparin-binding cytokine is FGF-2 (fibroblast growth factor-2). Although there has been extensive investigation, no unique binding sequence has emerged, as previously reviewed in [1]. High-affinity binding seems only to require a sequence in which glucosamines are N-sulphated and one of the intervening uronic acids is a 2-O-sulphated iduronic acid. Similar outcomes seem to be emerging for other cytokines and proteins, that high-affinity GAG binding may need oligosaccharides with some limited specific structural features, but does not depend on any single unique sequence, or even a very restricted set of sequences. More information in this area will help us decide the extent to which this is the case. In turn this will inform us whether we should view HS/heparin chains as arrays of large numbers of unique binding sites for different cytokines and proteins, or whether, as seems more likely, there could be competition between different proteins for overlapping binding sites.

GAG-binding sites on proteins and cytokines: continuous and discontinuous

Heparin and highly sulphated domains in HS are relatively rigid and straight. Within HS, unsulphated domains are flexible and therefore allow for bends in the chains [5]. Irrespective of sulphate density, HS/heparin chains have a large hydrodynamic volume. It follows that any amino acid side chain able to contact a GAG chain within a binding site will need to be prominently exposed on the surface of the polypeptide, and surface residues lying within even quite large clefts or pockets will be inaccessible to the GAG. It is clear that protein–GAG binding involves a variety of different types of interactions including van der Waals forces, hydrogen bonds and hydrophobic interactions with the carbohydrate backbone. However, since the most prominent feature of HS/heparin chains is their highly acidic sulphate groups, it is unsurprising that the major contribution to binding strength is ionic interaction between these sulphates and the basic side chains of arginine, lysine, and to a lesser extent, histidine. Consideration of current well-studied HS/heparin-binding sites indicates that their major element is therefore a number of such basic residues, typically 4–7, uninterrupted by the presence of acidic side chains. These residues may constitute a single surface basic charge cluster, or linear alignment of basic residues (for example, along one side of an α-helix). The shallow helical conformation of heparin and sulphated regions of HS bears sulphate group clusters with a periodicity of approx. 17 Å (1 Å = 0.1 nm). Thus clusters of basic residues at this distance can also give rise to a combined binding site.

If a suitable cluster of basic residues arises from a single secondary structural element such as an α-helix, the cluster of basic residues will be evident from the primary sequence. Alternatively, as in the case of antithrombin III, key heparin-binding residues may be dispersed throughout the primary sequence, but are brought together by the tertiary folding of the polypeptide. This variety in the nature of HS/heparin-binding sites has been well presented elsewhere [6]. Thus although searching protein sequences to identify basically charged ‘heparin-binding motifs’ may have some success in identifying actual heparin-binding sites, such an approach cannot be entirely reliable, as on the one hand it is liable to miss discontinuous binding sites, and on the other hand it cannot account for the possibility that a highly basic peptide sequence may be on the protein surface but folded in such a way that some of the basic side chains in the putative site are either inaccessible to GAG, or on a different face of the protein from the others.

This brings us to a potential technical pitfall in studying the GAG-binding properties of recombinant cytokines and their receptors: many such proteins are tagged with a terminal polyhistidine sequence for ease of purification. It has been shown that these tags are able to function as artefactual heparin-binding sites [7], a finding which the present authors are able to confirm. Studying the GAG-binding properties of polyhistidine-tagged proteins should therefore be avoided.

Consequences of cytokine–GAG binding

In general, the binding of a cytokine to GAG is thought to have a number of biological consequences. Binding to GAG may protect a cytokine from proteolytic degradation. A particularly striking example of this is the case of IFN-γ (interferon-γ), where heparin binding to a C-terminal motif blocks proteolytic cleavage at this site. Since otherwise such processing of the polypeptide leads to rapid clearance from the circulatory system, the protective effect of heparin considerably increases the bioactivity of IFN-γ in the serum.
A different form of cytokine protection also emerged from early work on FGFs in which it was shown that GAG binding stabilized FGFs against thermal and pH denaturation. This led to the view that FGFs could be secreted into the extracellular matrix, where they would be trapped in a stabilized form in the surrounding tissue microcompartment. This view also brings in a concept of restricting cytokine diffusion, so that the cytokine would only be available to neighbouring cells, leading to a paracrine, or juxtacrine mode activity during tissue growth and repair. Clearly limited cell access to FGFs, which are potent mitogens for many cell types, is required if a well-organized tissue architecture is to be built during organogenesis and maintained by repair thereafter. A similar requirement for localization of activity is seen in the triggering of inflammatory reactions by chemokines, many of which bind to HS/heparin. Here, chemokines are synthesized by tissue cells at sites of pathological damage. By poorly understood mechanisms, the chemokines are transported to and across vascular endothelial cells where they are displayed by HS chains on the luminal surface for recognition by receptors on passing leucocytes [9]. This will then initiate leucocyte activation and migration into the damaged tissue site. In both of these examples then, a key role of HS is to provide a local tissue depot of active cytokine.

These concepts can be extended into other areas of functioning of other cytokines. In embryogenesis and organogenesis, a widely held paradigm is that gradients in morphogen concentration are essential to organize the various axes of cell growth and differentiation necessary to generate body patterning. How can this be achieved by the small, soluble growth factors that several morphogens appear to be? Inherently one would expect such factors to diffuse rapidly on secretion. We know that morphogens such as Hedgehog, Wnt and the BMPs (bone morphogenetic proteins) bind to HS. Such HS binding would restrict diffusion, thereby maintaining concentration gradients for a sufficient time for the morphogenic organization of cell proliferation and differentiation. Recently, however, it has emerged that HS proteoglycans may not only play a role in the persistence of morphogenetic gradients, but may also have a more active role in gradient establishment by transporting the morphogen across fields of cells in the first instance. In Drosophila, the morphogenic gradient of the BMP homologue dpp, decapentaplegic, required for developmental patterning of the wing, depends on two Drosophila homologues of the glypicanc HS proteoglycans [10]. It is notable in this context that in humans, Simpson–Golabi–Behmel syndrome, characterized by abnormal growth, arises from mutations of the glypican-3 gene, and that certain tumours are associated with under- or over-expression of glypican-1 and -3 [11].

**FGF: the prototypical heparin-binding cytokine**

In the now distant era when growth factors, like other proteins, had to be purified for biochemical characterization (rather than being identified through gene sequences and studied in recombinant form), it was found that affinity chromatography on immobilized heparin afforded a high degree of purification in a single step. The realization dawned that this affinity for GAG might reflect physiological behaviour. This led to our current understanding that initial capture on HS chains is essential for cell signalling via FGFR (FGF receptor) polypeptides. Thus, in this context, HS is referred to as being a co-receptor. Our best understanding of the molecular mechanism underlying this co-receptor function arises from X-ray crystallographic studies of heparin–FGF–receptor complexes. One of the crystallographic structures of a heparin–oligosaccharide binding to a dimer of FGF/FGFR pairs shows two oligosaccharides, non-reducing end to non-reducing end, interacting with two identical halves of the cytokine–receptor complex in an entirely symmetrical manner [12]. Another such structure shows instead an asymmetric contact between heparin, one receptor and two cytokine polypeptides in the complex [13]. The heparin chain stabilizes the FGF–FGFR complex, in which a cytokine dimer and two receptor molecules, which need to phosphorylate each other’s cytoplasmic domains, are in close proximity with each other. In other words, the co-receptor function of HS arises from its ability to stabilize the active signalling cytokine–receptor complex.

But can we expect this mechanism to be replicated with other HS/heparin-binding cytokines? As only partly indicated by the cytokines mentioned in this introductory paper and discussed elsewhere by the contributors to this Focused Meeting, a great many cytokines from diverse growth factor families are now known to bind to heparin and HS. That is not to say that all cytokines show such binding, EGF (epidermal growth factor) and NGF (nerve growth factor) are just two well-studied cytokines for which there is no evidence for GAG binding. However, the many cytokines that do show binding have great diversity of structure. For example, several members of the TGF-β (transforming growth factor-β) cytokine family including BMPs, glial cell line-derived neurotrophic factor and two of the isoforms of TGF-β itself bind to HS and heparin [14]. These are examples of cytokines that are already dimerized in circulation, in this case via a disulphide linkage. Clearly, such cytokines do not require the intervention of HS to generate dimers. Moreover, there is considerable diversity in the receptors for heparin-binding cytokines. Often two distinct receptor polypeptide chain types comprise the signalling receptor complex. Indeed the homodimeric nature of FGFR complexes is unusual, so it is unreasonable to expect the paradigm that has emerged for the molecular role of HS in FGF signalling to be fully applicable to other cytokine–receptor systems. It will therefore be necessary to investigate each such system individually. Our understanding of the role and importance of HS in cytokine signalling is currently at an early stage. It will be of great interest to learn, as this field develops, the extent to which there are generalized principles for the involvement of HS in cytokine function, and to what extent each HS–cytokine interaction has evolved independently, and is therefore structurally and functionally distinct.
We thank David McClarence (Royal Holloway) for his critical reading of this paper.

References

Received 17 January 2006