Pure & Appl. Chem., Vol. 69, No. 9, pp. 1897–1902, 1997.
Printed in Great Britain.

Heparan sulfate—a polyanion with multiple messages

Ulf Lindahl

Dept. of Medical and Physiological Chemistry, University of Uppsala, The Biomedical Center, Box 575, S-751 23 Uppsala, Sweden

Abstract: Proteoglycans are composed of sulfate-substituted, negatively charged glycosaminoglycan chains that are covalently linked to proteins. Studies on proteoglycan biosynthesis have been focused on the isolation and molecular cloning of the various enzymes that catalyze this process. Enzymes involved in the biosynthesis of heparin and heparan sulfate include the glycosyltransferases responsible for generating the initial $(GlcA\text{-}GlcNAc)_n$ chains, the GlcNAc N-deacetylase/N-sulfotransferase that introduces N-sulfate groups, the D-GlcA C5-epimerase that generates L-IdoA units, and O-sulfotransferases that sulfate hydroxyl groups in various positions. Restricted polymer modification will lead to the generation of complex saccharide sequences of varied structure. Attempts have been made to define the minimal saccharide sequences required for binding of various proteins of biological interest, including growth factors of the fibroblast growth factor family. It is proposed that many "heparin-binding proteins", with affinity for the predominant structure in the highly sulfated heparin molecule, may bind to distinct, less modified, regions of heparan sulfate chains. These studies are expected to promote our understanding of the regulatory mechanisms behind polysaccharide biosynthesis, and of the physiological roles of proteoglycans. Further, they may provide the basis for the generation of novel drugs.

INTRODUCTION

Heparin is isolated from animal tissues, primarily pig intestine, and is used in the clinic, due to its blood anticoagulant properties, to prevent and treat thromboembolic disease. In the intact tissue it is contained within the connective-tissue type mast cell, where it is stored in cytoplasmic granules. Heparan sulfate (HS), a glycosaminoglycan structurally related to heparin, has a ubiquitous distribution in the extracellular matrix and on cell surfaces. Interactions between HS and proteins are being increasingly implicated in a variety of physiological processes, such as cytokine action, enzyme regulation, cell adhesion etc. (1).

Heparin has the highest negative charge density of any known biological macromolecule, and thus is prone to electrostatic interaction with a variety of proteins. While such interactions, commonly exploited in the purification of "heparin-binding proteins", may show an element of selectivity beyond that expected for simple cation-exchange chromatography, the sheer multitude of proteins involved suggests a relatively nonspecific mode of binding. HS, on the other hand, has a more complex structure than heparin, and appears to be tailored for specific interactions in diverse functional contexts (Fig. 1) (2). These relations will be highlighted in the following discussion, along with an account of the biosynthetic mechanisms involved in the generation of the polysaccharide chains.

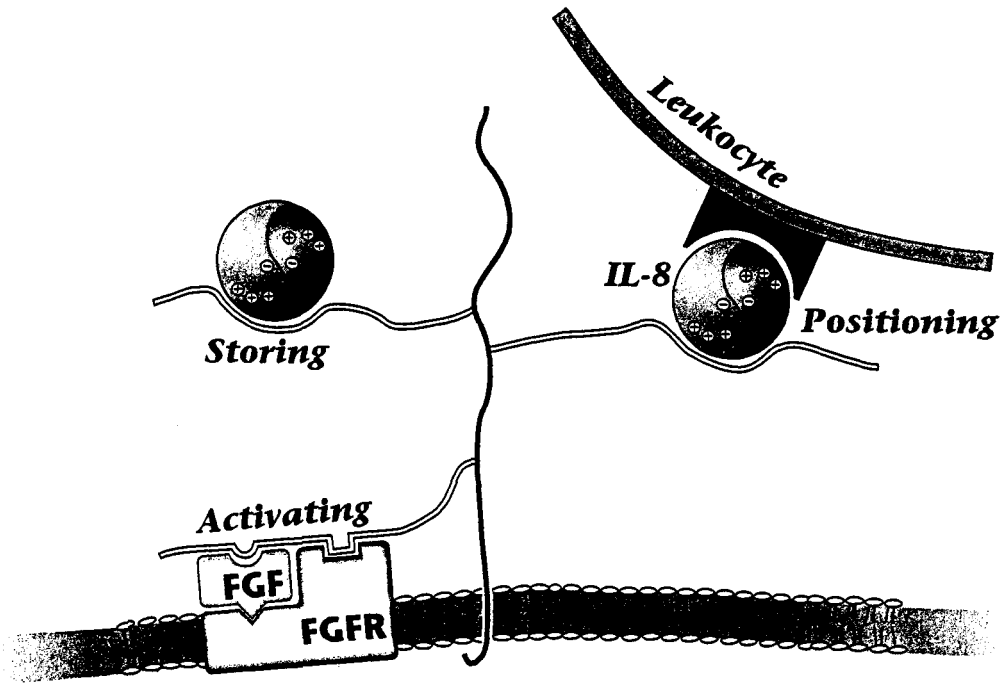

Fig. 1 Schematic representation of a cell-surface proteoglycan, and conceivable functional roles exerted through protein binding to the constituent glycosaminoglycan chains (generally HS). For information regarding the examples illustrated, see *e.g.* ref. 2.

BIOSYNTHESIS OF HEPARIN AND HEPARAN SULFATE

Heparin and HS are both synthesized as proteoglycans, which consist of glycosaminoglycan chains covalently bound to a protein core (for references to previous work, see ref. 2). A single protein, serglycin, has been identified as the core of heparin proteoglycans, whereas HS proteoglycans contain a variety of core structures (3-5). Biosynthesis of either heparin of HS proteoglycans involves the formation of an initial, simple polymer, that is composed of alternating D-glucuronic acid (GlcA) and N-acetyl-D-glucosamine (GlcNAc) units, and is linked to the core protein through a →4GlcAß1→3Galß1→3Galß1→4Xylß→O-serine sequence. The GlcA and GlcNAc units are added, through ß1→4 and α1→4 linkages, respectively, to the nonreducing termini of nascent polysaccharide chains, in a process that appears to be catalyzed by a single enzyme protein (6,7).

The initial polymerization product is modified through a series of reactions which ultimately results in the generation of - IdoA(2-OSO_3) - GlcNSO_3(6-OSO_3) - sequences [where L-iduronic acid (IdoA) is the C5-epimerization product of GlcA], *i.e.* the predominant and most extensively sulfated disaccharide unit of heparin (Figs. 2, 3; refs. 1,2). The modification of the $(GlcA\text{-}GlcNAc)_n$ polymer is initiated by N-deacetylation of GlcNAc units, which is followed by sulfation of the exposed free amino groups; the latter reaction as well as the subsequent O-sulfotransferase steps require 3'-phosphoadenosine-5'-phosphosulfate (PAPS) as a sulfate donor. Following, and dependent on, N-sulfation, GlcA units are C5-epimerized to IdoA units, which are then O-sulfated at C2. Finally, GlcN residues are O-sulfated at C6. Additional O-sulfation, generally to a minor extent, may occur at C3 of GlcN and at C2 of GlcA units. The former of these reactions has been associated with formation of the antithrombin-binding region in heparin and HS (8-10).

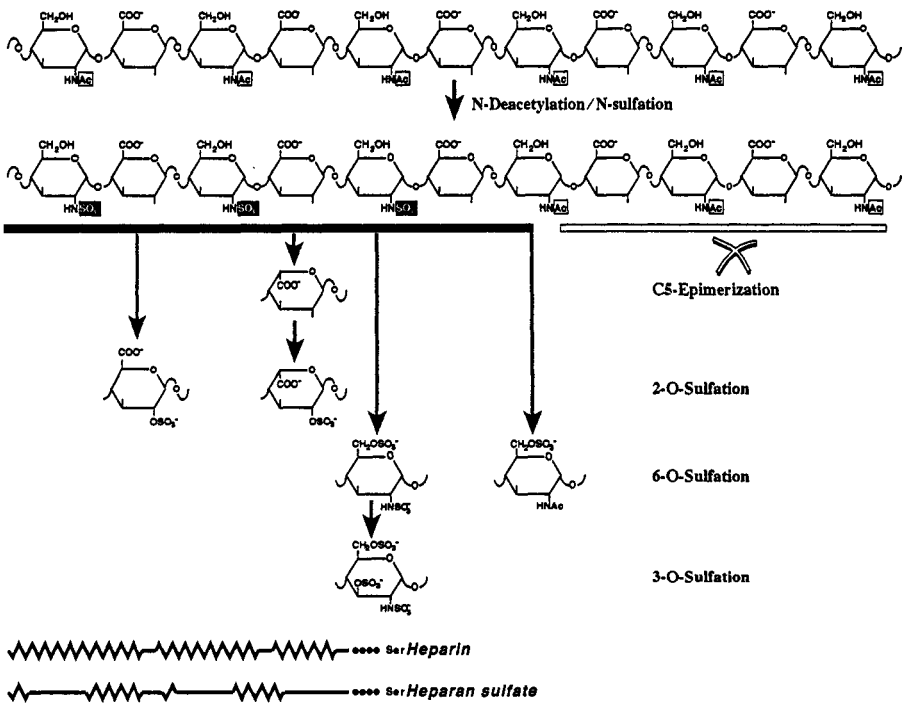

Fig. 2 (A) Polymer-modification reactions in the biosynthesis of heparin and HS. The scheme illustrates the key regulatory role of the first modification reaction, which results in the incorporation of N-sulfate groups. All subsequent modification reactions depend on the presence of N-sulfate groups for substrate recognition. (B) Schematic representation of the domain organization of heparin and HS. Zig-zag line, N-sulfated domain; uninterrupted line, N-acetylated domain.

By and large, the structural properties of HS, as related to those of heparin, can be conceived as the result of restricted polymer modification. The most conspicious restriction of polymer modification is due to incomplete N-deacetylation/N-sulfation. The enzymes that catalyze the C5-epimerization and different O-sulfation reactions all require N-sulfate groups for substrate recognition, and sequences composed of consecutive N-acetylated disaccharide units will be devoid of IdoA and O-sulfate residues (1,11). Such sequences are rare in heparin but are typical for HS chains. Furthermore, sequences composed of alternating N-acetylated and N-sulfated disaccharide units, and with distinctive O-sulfation patterns, are abundant in HS (11,12) but scarce in heparin. The structural distinction between heparin and HS thus heavily relies on analysis of N-substituent patterns, as schematically visualized in Fig. 2. However, even within the N-sulfated regions potential target units will escape modification. Thus, a GlcA residue interspersed between N-sulfated GlcN units may undergo C5-epimerization to IdoA, but may also remain unchanged (Fig. 2). Similarly, 6-O-sulfation of $GlcNSO_3$ units is optional. The mechanism(s) of target selection in polymer modification (within the limits set by the substrate specificities of the corresponding enzymes) are unknown. As a result of this selection process, the structural complexity and heterogeneity of the HS chain will increase through the modification process. As will be further discussed, selectively restricted polymer modification provides the basis for the generation of protein-binding regions of defined structure.

To achieve a deeper understanding of heparin/HS biosynthesis and its regulation, several research groups are currently involved in the molecular characterization of the enzymes involved in the process. Results pertaining to the ~110 kDa GlcNAc N-deacetylase/N-sulfotransferase, another enzyme that catalyzes two distinct reactions, revealed remarkable complexity (13-15). As noted above, the regulation of the corresponding reactions is essential, since the resultant distribution of N-acetyl and N-sulfate groups will control subsequent modification. The N-deacetylase/N-sulfotransferase has been found to

occur in two distinct forms with partly different catalytic properties. One of these forms, isolated from rat liver (13), was associated with the biosynthesis of HS, whereas the other, derived from mouse mastocytoma (14), was implicated with heparin generation. The two enzymes are encoded by transcripts of markedly different size, ~8 kb and ~4 kb, respectively, and presumably have distinct roles in regulating the N-deacetylation/N-sulfation process during heparin/HS biosynthesis. Thus, transfection of a HS-producing cell line with cDNA corresponding to the mast-cell N-deacetylase/N-sulfotransferase induced a dramatic shift of the N-substituent pattern of the HS produced by the cell, toward that typical for heparin (16). Will yet other enzymes involved in heparin/HS assembly occur in multiple forms? How many non-enzyme (auxiliary) proteins are required to generate the functional assembly lines (see *e.g.* ref. 9)?

A current model of heparin/HS biosynthesis features simultaneous elongation and modification of the polysaccharide precursor, in reflection of the demonstrated coupling of the polymerization and N-deacetylation/N-sulfation reactions (17). Saccharide sequences with N-sulfate or residual N-acetyl groups are generated by a glycosyltransferase - N-deacetylase/N-sulfotransferase complex, located at the nonreducing end of the nascent chain which concomitantly undergoes further downstream modification catalyzed by other enzymes (2). The enzymes are visualized as acting in a semi-processive manner along the polymer, in an on-off manner that is related to the selection of target sites discussed above. Clearly, major aspects of this model remain speculative, pending further detailed information regarding the molecular and kinetic properties of the enzymes, the mode of interaction of the enzymes with each other and with the polysaccharide substrate, and the overall organization of the biosynthetic apparatus within the Golgi system.

INTERACTIONS WITH PROTEINS

Binding of heparin/HS sequences to proteins is largely electrostatic in nature, and thus involves positively charged amino-acid residues in the protein components. Attempts to define polypeptide consensus sequences for heparin binding, based on clustered basic amino acids, have been partly contradictory (18); other types of polysaccharide-binding regions, composed of multiple peptide loops (see *e.g.* ref. 19) must also be considered. Conversely, our knowledge regarding the structures of protein-binding domains of polysaccharide chains is limited. One category of such domains would seem to involve "unique" structural components, that are either absent or rare in other regions of the polymers. An example of such a marker component, associated with a defined biological function, is the 3-O-sulfated GlcN unit, which is contained within the antithrombin-binding pentasaccharide sequence of heparin as well as HS, and is essential to the blood anticoagulant activity of the compound (reviewed in ref. 20). Other variable, and generally minor, constituents, hence potentially involved in selective interactions with proteins, include N-unsubstituted GlcN units (21) (see also ref. 22 for relation to L-selectin binding), and 2-O-sulfated GlcA residues. The latter component accounts for as much as 11% of the total N-sulfated disaccharide units of HS from adult human cerebral cortex, but is virtually absent in the corresponding neonatal polysaccharide, as well as in HS preparations from other selected adult tissues (23).

Contrary to proteins such as antithrombin, which requires a unique component for polysaccharide recognition, a large number of "heparin-binding proteins" appear to bind to sequences composed of the major, trisulfated [- IdoA(2-OSO_3) - GlcNSO_3(6-OSO_3) -] disaccharide unit of heparin (see refs. 2,18 for further information). The same disaccharide unit occurs also in most HSs, although in lower abundance. However, the precise requirements for the individual sulfate groups within this sequence are likely to vary from one "heparin-binding" protein to another; whereas some groups are essential for the interaction, others may be redundant, neither contributing to nor interfering with protein binding. The concept is illustrated by structure/function relationships bearing on polysaccharides interacting with members of the fibroblast growth factor (FGF) family. A number of these proteins appear to bind heparin *via* the same sequence of trisulfated disaccharide units. However, the minimal binding sequence for FGF-2 (basic FGF) has

been identified as a pentasaccharide structure in which the essential sulfate groups are limited to a single IdoA 2-O-sulfate and one or two N-sulfate groups (24,25); no GlcN 6-O-sulfate group seems to be required (Fig. 3). Moreover, the available data suggest that binding of different members of the FGF family may require different combinations of sulfate groups, hence different saccharide sequences (18). Such sequences may well be expressed, albeit in "hidden" form, by the same fully sulfated heparin structure. However, they may also occur, differentially expressed and reduced to the minimal essential structures, in glycosaminoglycan chains of the HS type. The generation of such specific protein-binding regions would require selective restriction of biosynthetic polymer modification.

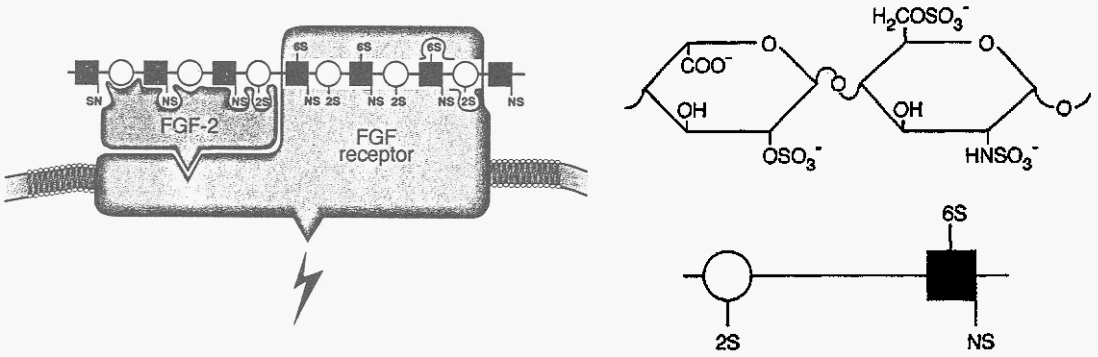

Fig. 3 Scheme illustrating the proposed mode of interaction between HS, FGF-2 and the FGF-2 receptor. The symbols used to denote the various monosaccharide units are defined to the right. The pentasaccharide sequence interacting with FGF-2 is shown to contain a single IdoA 2-O-sulfate group, whereas the receptor-binding sequence contains both IdoA 2-O-sulfate and GlcN 6-O-sulfate groups (the number of 2-O- and 6-O-sulfate groups in the latter domain has not been defined). For additional information see the text and ref. 26.

The generation of multiple protein-binding sites along the same polysaccharide chain introduces further complexity in regulation of polymer modification, particularly if these sites differ from each other. We may again consider interactions involving FGF-2, now extended to include also its cell-surface tyrosine kinase type receptor (26). The minimal heparin/HS fragment required to support FGF-2-induced DNA replication (cell proliferation) was found to consist of a 12-saccharide, thus more than twice the size of the pentasaccharide sequence that binds to the growth factor (see above). One explanation to this phenomenon is that binding of two growth factor molecules to adjacent sites on the polysaccharide chain is needed to promote receptor dimerization, as required for receptor activation (see ref. 27). However, the effects of selectively O-desulfated heparin preparations on FGF-2-dependent cell proliferation pointed to a requirement for both IdoA 2-O- and GlcN 6-O-sulfate groups (26), thus contrary to growth factor binding alone which required 2-O-sulfate groups only (see above). These findings suggested the occurrence of a ternary complex, with FGF-2 and its receptor binding to adjacent, distinct, sites on the same polysaccharide chain. While both of these sites would presumably be accommodated by the same predominant, highly sulfated heparin sequence, they might well be differentially expressed along a HS polymer. In such a case, only HS chains that contain the appropriately spaced growth factor-binding and receptor-binding sequences would promote the FGF-induced cellular response. Indeed, other species containing only one of the sequences, or with incorrectly spaced binding regions, would inhibit the response.

Previous studies indicate that HS preparations derived from different cells and tissues differ in structure, thus in agreement with the concept of differentially expressed protein-binding sites (21,23,28,29). The precise binding characteristics for various proteins need to be established for several reasons, one being that such information will be prerequisite to the generation of carbohydrate-based drugs or glycomimetics. Further, knowledge of such

structure-function relationships may help us to assess the mechanisms in control of the polymer-modification reactions involved in HS biosynthesis. Ultimately, the enzymes catalyzing these reactions, as well as any auxiliary proteins required to support the process, need to be cloned and characterized in detail.

REFERENCES

1. Lindahl, U., Lidholt, K., Spillmann, D., and Kjellén, L. (1994) *Thromb. Res.* **75,** 1-32
2. Salmivirta, M., Lidholt, K., and Lindahl, U. (1996) *FASEB J.* in press
3. Kjellén, L., and Lindahl, U. (1991) *Annu. Rev. Biochem.* **60,** 443-475
4. David, G. (1993) *FASEB J.* **7,** 1023-1030
5. Iozzo, R. V. (1994) *Matrix Biology* **14,** 203-208
6. Lidholt, K., Weinke, J. L., Kiser, C. S., Lugemwa, F. N., Bame, K. J., Cheifetz, S., Massagué, J., Lindahl, U., and Esko, J. D. (1992) *Proc. Natl. Acad. Sci.* **89,** 2267-2271
7. Lind, T., Lindahl, U., and Lidholt, K. (1993) *J. Biol. Chem.* **268,** 20705-20708
8. Kusche, M., Bäckström, G., Riesenfeld, J., Petitou, M., Choay, J., and Lindahl, U. (1988) *J. Biol. Chem.* **263,** 15474-15484
9. Shworak, N. W., Shirakawa, M., Colliec-Jouault, S., Liu, J., Mulligan, R. C., Birinyi, L. K., and Rosenberg, R. D. (1994) *J. Biol. Chem.* **269,** 24941-24952
10. Razi, N., and Lindahl, U. (1995) *J. Biol. Chem.* **270,** 11267-11275
11. Gallagher, J. T., Turnbull, J. E., and Lyon, M. (1992) *Int. J. Biochem.* **24,** 553-560
12. Maccarana, M., Sakura, Y., Tawada, A., Yoshida, K., and Lindahl, U. (1996) *J. Biol. Chem.* in press
13. Hashimoto, Y., Orellana, A., Gil, G., and Hirschberg, C. B. (1992) *J. Biol. Chem.* **267,** 15744-15750
14. Eriksson, I., Sandbäck, D., Ek, B., Lindahl, U., and Kjellén, L. (1994) *J. Biol. Chem.* **269,** 10438-10443
15. Orellana, A., Hirschberg, C. B., Wei, Z., Swiedler, S. J., and Ishihara, M. (1994) *J. Biol. Chem.* **269,** 2270-2276
16. Cheung, W.-F., Eriksson, I., Kusche-Gullberg, M., Lindahl, U., and Kjellén, L. (1996) *Biochemistry* **35,** 5250-5256
17. Lidholt, K., and Lindahl, U. (1992) *Biochem. J.* **287,** 21-29
18. Spillmann, D., and Lindahl, U. (1994) *Curr. Opin. Struct. Biol.* **4,** 677-682
19. van Boeckel, C. A. A., Grootenhuis, P. D. J., and Visser, A. (1994) *Nature struct. Biol.* **1,** 423-425
20. Bourin, M.-C., and Lindahl, U. (1993) *Biochem. J.* **289,** 313-330
21. van den Born, J., Gunnarsson, K., Bakker, M. A. H., Kjellén, L., Kusche-Gullberg, M., Maccarana, M., Berden, J. H. M., and Lindahl, U. (1995) *J. Biol. Chem.* **270,** 31303-31309
22. Norgard-Sumnicht, K., and Varki, A. (1995) *J. Biol. Chem.* **270,** 12012-12024
23. Lindahl, B., Eriksson, L., and Lindahl, U. (1995) *Biochem. J.* **306,** 177-184
24. Maccarana, M., Casu, B., and Lindahl, U. (1993) *J. Biol. Chem.* **268,** 23898-23905
25. Faham, S., Hileman, R. E., Fromm, J. R., Linhardt, R. J., and Rees, D. C. (1996) *Science* **271,** 1116-1120
26. Guimond, S., Maccarana, M., Olwin, B. B., Lindahl, U., and Rapraeger, A. C. (1993) *J. Biol. Chem.* **268,** 23906-23914
27. Heldin, C.-H. (1995) *Cell* **80,** 213-223
28. Lyon, M., Deakin, J. A., and Gallagher, J. T. (1994) *J. Biol. Chem.* **269,** 11208-11215
29. Kato, M., Wang, H., Bernfield, M., Gallagher, J. T., and Turnbull, J. E. (1994) *J. Biol. Chem.* **269,** 18881-18890