Pure Appl. Chem., Vol. 84, No. 1, pp. 65–76, 2012.
http://dx.doi.org/10.1351/PAC-CON-11-10-27

Molecular architecture of heparin and heparan sulfate: Recent developments in solution structural studies*

Barbara Mulloy[1,‡], Sanaullah Khan[2], and Stephen J. Perkins[2]

[1]Laboratory for Molecular Structure, National Institute for Biological Standards and Control, Blanche Lane, South Mimms, Potters Bar, Herts. EN6 3QG, UK; [2]Department of Structural and Molecular Biology, Darwin Building, University College London, Gower Street, London WC1E 6BT, UK

Abstract: The study of the relationship between the complex structures and numerous physiological functions of the glycosaminoglycans (GAGs) heparin and heparan sulfate (HS) has continued to thrive in the past decade. Though it is clear that the monosaccharide sequences of these polysaccharides must determine their ability to modulate the action of growth factors, morphogens, chemokines, cytokines, and many other extracellular proteins, the exact details of this dependence still prove elusive. Sequence determines the 3D structure of GAGs at more than one level; detailed sequences of highly sulfated regions may influence affinity for specific proteins in some cases, but in addition attention has been called to the importance of the length and spacing of these highly sulfated sequences, which are separated by unsulfated domains. Within the sulfated "S-domains", the internal dynamics of the conformationally flexible iduronate pyranose ring have continued to interest NMR spectroscopists and molecular modelers. New studies of the relative degrees of flexibility of sulfated and unsulfated domains lead to an overall model of heparin/HS in which protein-binding, highly sulfated S-domains with well-defined conformations are separated by more flexible NA-domains.

Keywords: analytical ultracentrifugation; heparin; heparan sulfate; NMR spectroscopy; X-ray scattering.

INTRODUCTION

Although the cell-surface glycosaminoglycan (GAG) heparan sulfate (HS) is an unbranched polysaccharide with a simple disaccharide repeating unit, it has a remarkably intricate structure. Like most of the other members of the GAG family, it is initially synthesized as a simple polysaccharide, in this case consisting of alternating *N*-acetyl-α-D-glucosamine (GlcNAc) and β-D-glucuronic acid (GlcA) [1] (Fig. 1A). The subsequent action of a series of post-polymerization enzymes introduces a domain structure of highly sulfated regions known as S-domains [2], consisting of 2-O-sulfated α-L-iduronic acid (IdoA2S) and 2-N-sulfated, 6-O-sulfated α-D-glucosamine (GlcNS6S) separated by unchanged, unsulfated regions (abbreviated to NA-domains [3] as the simpler term A-domain has already been used for the sequence in heparin with high affinity for antithrombin) [4]. Incomplete action of the post-poly-

Pure Appl. Chem.* **84, 1–106 (2012). A collection of invited papers based on presentations at the 16th European Carbohydrate Symposium (Eurocarb-16), Sorrento, Italy, 3–7 July 2011.
‡Corresponding author

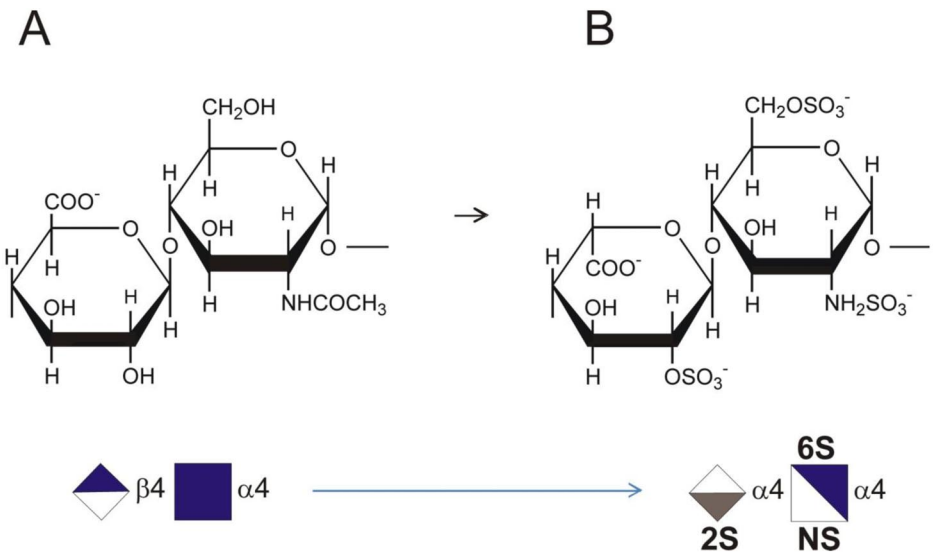

Fig. 1 Biosynthesis of heparin and HS; a simplified pictorial summary, with (below) sequences written in a conventional symbolic format [65]. The structure depicted in A is the disaccharide repeating unit of heparosan, [4-GlcA(β1-4)GlcNAc(α1-]$_n$ the precursor polysaccharide of heparin and HS. After conversion of GlcNAc to GlcNS, the β-D-GlcA residue is epimerized to α-L-IdoA , which may then be sulfated at position 2; position 6 of GlcNS may also then be sulfated. The structure depicted in B is that of the fully sulfated regular repeat unit of heparin, or the central S-domains in HS, [4-IdoA2SO$_3^-$(α1-4)GlcNSO$_3^-$6SO$_3^-$(α1-]$_n$.

merization enzymes results in a complex fine structure within and at the boundaries of the S-domains. As HS is known to be an essential modulator of the action of morphogens in embryonic development [5], its interaction with extracellular proteins is expected to be selective and so dependent on the presence of specific binding motifs in both the protein and HS; but the relationship between HS sequence and specificity for particular proteins is still not solved, and over the past decade has been controversial [6–8].

Part of the solution must surely lie in an understanding of the 3D structures of all the partners (GAGs, signaling proteins, and sometimes their receptors) in the interactions [9–13]. Solid-state crystal structures of proteins with bound heparin-like oligosaccharides have increased in number in the past decade [14] and there are ways of predicting some aspects of protein–HS interactions by theoretical calculations and testing these predictions by means of solution techniques of molecular and structural biology [15]. However, these structures contain only short oligosaccharide fragments of a homogeneous nature, so are not informative about either selectivity and sequence variation, or about the overall shape of the GAG in solution.

Much experimental work in this area uses heparin, a variant of HS found in mast cell granules, rather than cell-surface HS. Heparin consists largely, though not entirely, of the S-domain trisulfated disaccharide sequence, and oligosaccharides prepared from it are quite uniform in structure, which is an advantage to the crystallographer. In addition, it is widely available, as it is used in medicine as an anticoagulant and antithrombotic drug [16].

For linear polysaccharides, the major source of overall flexibility is at the glycosidic linkages between monosaccharide residues, consisting of two bonds (or three for 1 → 6 linked polysaccharides) around which some degree of rotation is possible. There is an extra degree of interest in the detailed dynamics of heparin and HS, as in addition to the possibility of flexibility around linkages between monosaccharide residues, the α-L-iduronate residue (IdoA), formed by epimerization at C5 of β-D-GlcA exists in an equilibrium between more than one pyranose ring shape. This has potential con-

sequences for the overall conformation of the polysaccharide as well as for "induced fit" models of interactions with proteins.

In this review we deal with the conformation and the dynamics of HS at the fine structure level, where NMR spectroscopy and computational chemistry are appropriate techniques (other adventurous spectroscopic techniques for the purpose have recently also been reviewed [17]). In addition, we describe our recent ultracentrifugation and scattering work on the overall solution structures of the sulfated and unsulfated domains of heparin and HS. Taking all these studies together, we can survey the potential sizes and shapes of HS molecules and the various ways in which they can interact with one or more protein partners.

S- AND NA-DOMAIN STRUCTURE OF HEPARIN AND HS

The biosynthesis of heparin/HS is summarized in Fig. 1. Subsequent to the addition of a linkage region to a serine residue in the protein core, the heparosan chain (sequence shown in Fig. 1A) is elongated by addition of alternating GlcA and GlcNAc residues by the glycosyl transferases EXT1 and EXT2 [18]. After that, the action of NDST (*N*-deacetylase-sulfotransferase), C5-epimerase, 2-*O*-sulfotransferase, and 6-*O*-sulfotransferase (in that order) introduce the sulfated regions known as the S-domains [1] (Fig. 1B). Clearly, as all the subsequent enzymes are dependent on the introduction of *N*-sulfamido groups, the action of the NDST enzyme controls the length and the spacing of the S- and NA-domains along the polysaccharide; isoform NDST1 for HS, isoform NDST2 for heparin [19,20]. HS domain patterns, as indicated by disaccharide analysis, vary according to the tissue of origin of HS [21]. HS from liver and brain were found to have the highest S-domain content; a recent study has identified heparin-like HS also in cartilage [22]. The size and degree of sulfation of S-domains has been found to increase towards the nonreducing end of the polysaccharide chain [23,24]; the positioning of protein-binding domains at the distal end of a cell-surface glycan makes functional sense in general, and in particular it is possible that highly sulfated nonreducing termini are involved in growth factor/HS/receptor complexes [25]. Highly sulfated S-domains give rise to the epitope for the phage display antibody NS4F5, associated with activated endothelial cells and amyloid deposits [26].

The size and sequence of S-domains in HS have been studied by excision of the NA-domains using a bacteriophage K5 lyase [27]. This investigation showed that mixed domains, containing both N-acetyl and N-sulfated glucosamine, are associated closely with S-domains, most probably as transition domains between NA- and S-domains. Further variability may be introduced by the action of extracellular Sulf enzymes, which trim away some of the 6-*O*-sulfates [28]; the physiological consequences of this small change are not trivial, and the Sulf enzymes are upregulated in several forms of cancer [29].

CONFORMATION, DYNAMICS, AND FINE STRUCTURE OF THE S-DOMAINS AND MIXED REGIONS

Interactions between HS and the numerous proteins through which it exerts its biological activities take place through the S-domains, possibly involving the transition zones of intermediate sulfation. These highly sulfated, iduronate-containing sequences have therefore been the subject of numerous detailed structural studies. The overall conformation of the alternating sequence of IdoA2S and GlcNS6S is well defined; the structure based on NMR spectroscopy (Fig. 2A) [30] is similar to that of the heparin oligosaccharides complexed with proteins in several crystal structures determined subsequently [14,31]. The systematic removal of sulfates, or replacement of *N*-sulfate with an *N*-acetyl substituent, does not perturb this conformation to any great extent [32]. This means that a rough model of any S-domain sequence can be predicted quite easily. A full turn of the helical structure of the S-domain sequence takes four monosaccharide residues, so that the distribution of sulfates along the polysaccharide chain is two-sided, with clusters of sulfates aligned opposite one another and offset, as in Fig. 2A. The result-

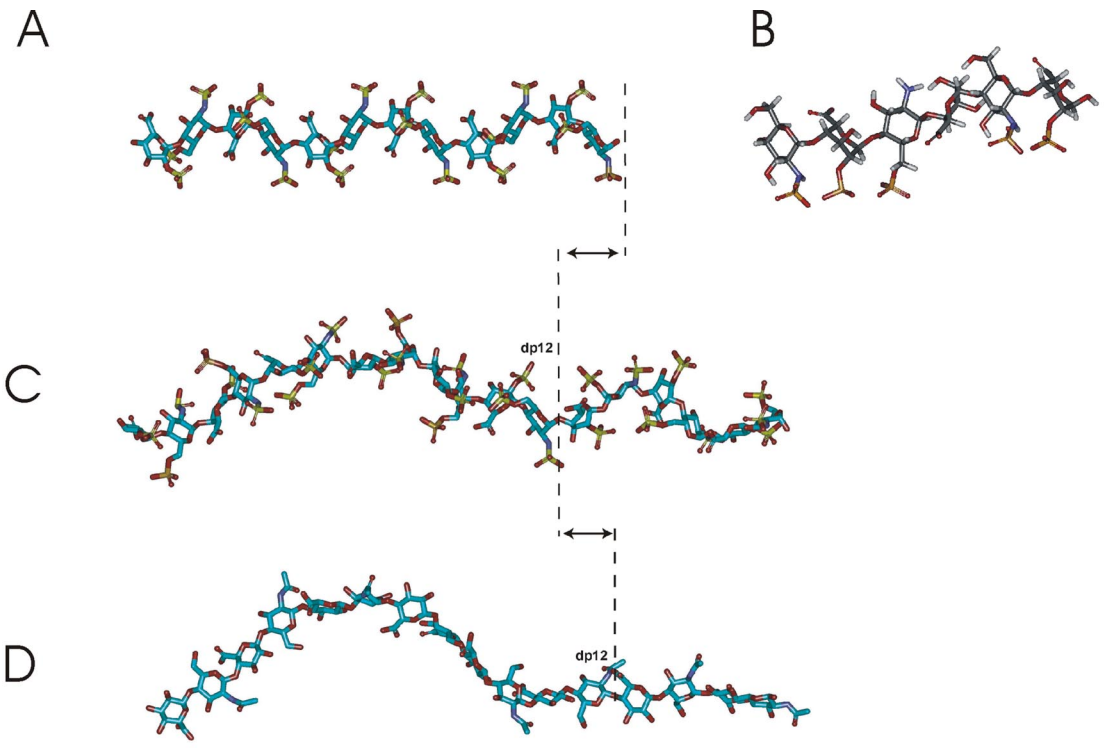

Fig. 2 3D solution structures of heparin and HS: (A) The NMR structure of heparin [30] with IdoA residues in the $^2S_0$ conformation; (B) a synthetic hexasaccharide with sulfate substitution along only one "side" of the chain [38]; (C) X-ray scattering derived structure of heparin [14]; (D) X-ray scattering derived structure of HS [62]. The chain length for 12 monosaccharide residues (dp12) is indicated by a dotted line for A, C, and D. The coloring is conventional, with carbon shown in green, oxygen red, sulfur yellow, nitrogen blue, and hydrogen white.

ing linear structure of the polysaccharide backbone is not symmetrical, but the pattern of negative charge, including sulfates and uronic acid carboxylates, is much the same in both directions. Missing sulfates will destroy this symmetry to produce 3D patterns of charge; but because of the "two-sided" nature of heparin there are always at least two sequences corresponding to any charge pattern. The two-sided nature of sulfate substitution along the heparin chain allows a simplified format for the description of sequences to show something of their 3D structure [7] (Figs. 3A,B). Using this format, Fig. 3C summarizes 21 different pentasaccharides containing the minimum pattern of sulfate substitution, along one side of the heparin chain, which is involved in interactions with one monomer of fibroblast growth factor 1 (FGF-1) in the crystal structure of the complex [7,33]. This level of redundancy, in the sequences that can give rise to a binding epitope for a particular protein, makes it exceptionally difficult to distinguish experimentally between selectivity for a single pattern of sulfate substitution and a completely nonspecific charge-based interaction. The sequence analysis of octasaccharides eluted from immobilized FGF-1 at different salt concentrations [34] produced results compatible with this analysis, though the authors were able to find a sequence-based motif which also distinguished between high- and low-affinity octasaccharides. Heparin is not a peptide; though sequence determines structure, there is no strict one-to-one relationship between sequence and 3D charge pattern.

Kreuger and co-workers [6] have reviewed the literature on the relationship of HS sequence with growth factor and morphogen interactions, and have established that interactions of HS with growth factors clearly do not depend on the presence of a single definable sequence in HS. They conclude that while some interactions with proteins require specific saccharide components, others depend only on

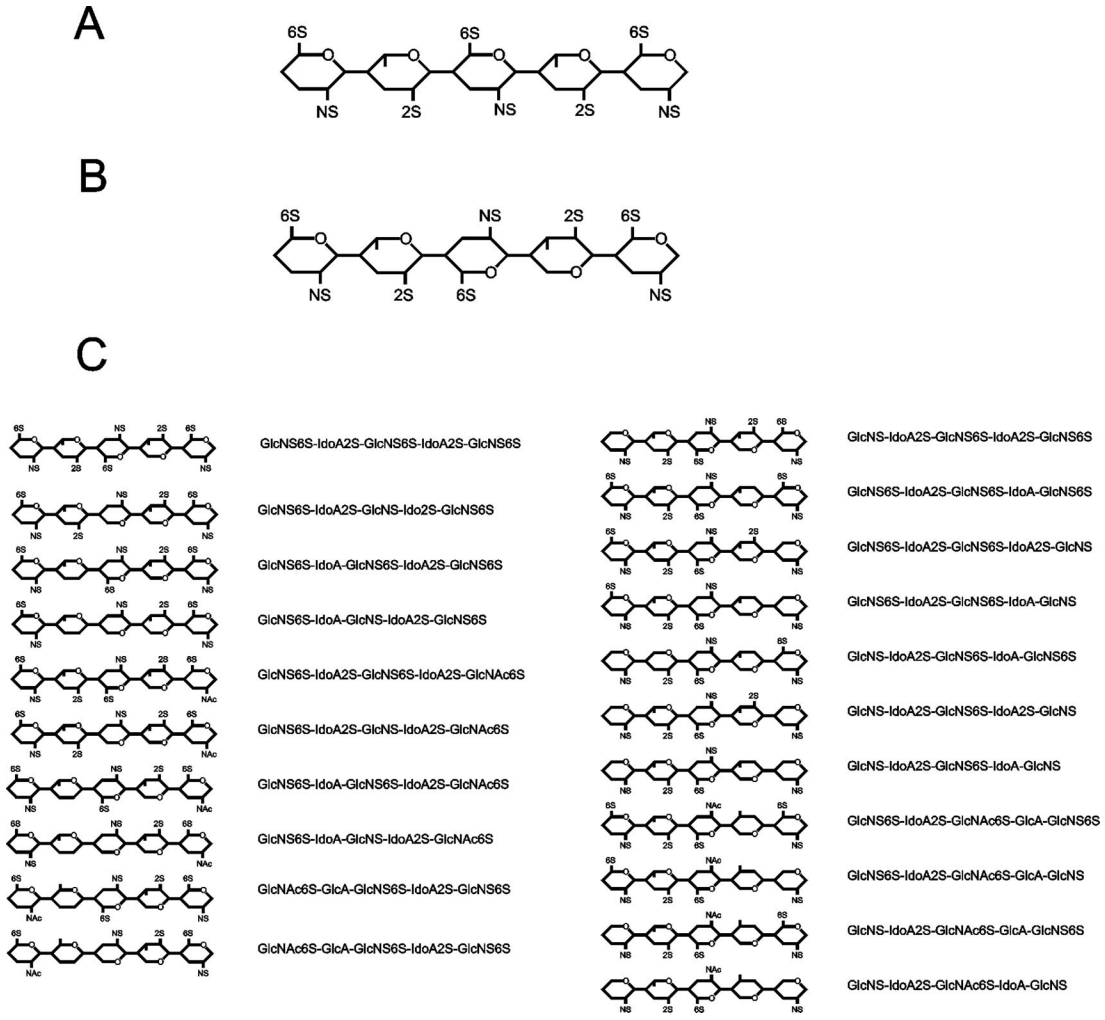

Fig. 3 (A) A pentasaccharide fragment of the structure of the main repeating unit of heparin or HS is a conventional, simplified format; all ring oxygens are shown at the top right of the ring. This representation correctly shows sequence. (B) The same structure in slightly amended format, in which one disaccharide has been inverted, so that the sequence is correctly indicated and an idea of the 3D structure can also be conveyed. (C) 21 pentasaccharide sequences are shown in the same depiction as B, all containing the minimum motif that binds to FGF-1 in the crystal structure of the complex.

"charge distribution". They propose that domain distribution along the chain, and degree of sulfation, are the two dominant structural features governing HS interactions with protein, mentioning the motif of two S-domains separated by one A-domain—the SAS domain. Furthermore, a number of chemically sulfated plant polysaccharides are capable of both stabilizing and supporting cell signaling of growth factors, though their carbohydrate backbones are not similar to those of heparin or HS [35]. The patterns and/or density of charged substituents are more important in this context than the molecular scaffold that carries them.

Based on a simple structure prediction, synthetic oligosaccharides have been prepared with sequences arranged to give "one-sided" distribution of sulfates (Fig. 2B) [36]. These are examples of perturbed heparin structures based ingeniously on what is known about the determinants of protein

binding. The oligosaccharides can stimulate the biological activity of FGF-1 [37], showing that the formation of FGF-1 dimers by binding to opposite sides of the heparin chain, as in the crystal structure of the complex [33], is not necessary. The structural and dynamic properties of an oligosaccharide of this type have been investigated by NMR spectroscopy [38], confirming its expected conformational properties.

The characterization of S-domains as rigid structures is not justified. Although rotation around the glycosidic bonds is inhibited, the IdoA residues do not adopt a single strongly preferred pyranose ring form, but have long been known to exist in an equilibrium between two or more forms [39]. This phenomenon continues to be of interest, and several molecular dynamics (MD) studies have appeared in the last 10 years [40–45]. An MD study on a synthetic S-domain-like hexasaccharide adopted a sophisticated treatment of electrostatics to predict that the $^2S_0$ skew boat form is stable on the nanosecond time scale, with no pseudorotational averaging [44] such as had been predicted in earlier studies [46]. The length of practicable MD runs increases with the availability of computing power. The use of MD for 0.2 μs on iduronate within oligosaccharide sequences has identified potential inter-residue hydrogen bonds [42]. Very long MD runs increase the chances that an equilibrium is reached, and a 3-μs run for IdoA and IdoA2S as monosaccharides has been performed [47], confirming that the conformational exchange takes place on the microsecond time scale. Exact agreement between MD predictions and NMR experimental data was still not achieved; the authors point out that force fields are not yet perfect and long runs will accentuate the effects of their imperfections.

In addition to force-field-based calculations such as MD, optimization of the structures of smaller fragments of heparin—mono- and disacccharides—may be achieved by ab initio calculations. In a series of studies using density functional theory (DFT) calculations at the B3LYP/6-311++G** level of theory, Hricovini and co-workers have obtained values for spin–spin coupling constants in an IdoA2S monosaccharide [48] and other oligosaccharides leading to the characterization of a Karplus curve for $^3J_{HH}$ as a function of torsional angle, which is particularly suited for use with GAGs [49]. Such refinements of the basic tools for NMR analysis are necessary to match the increase in quality of data available from modern NMR spectrometers. Detailed inspection of the geometry of the disaccharide GlcNS6S-IdoA2S after optimization leads to the conclusion that changes in IdoA2S conformation have a modest but significant effect on the pyranose ring of GlcNS6S [50,51].

The question of why the binding properties of IdoA for proteins are so favorable was analyzed in an NMR study on a heparin tetrasaccharide [52]. Using residual dipolar couplings as well as nuclear Overhauser enhancements (NOEs), restrained MD indicated that the overall structure of the oligosaccharide was not affected by the variation in ring form of the internal IdoA residue, a result in good agreement with earlier findings on heparin polysaccharide [32]. This study was conducted in the presence of calcium, rather than for the pure sodium salt as is more usual. The influence of cations on the function of heparin is a relatively little explored area; some studies indicate that cations such as calcium and copper have subtle effects on the conformation of heparin but may considerably affect biological activity [53]. For a heparin disaccharide model, the presence of calcium was predicted to affect both IdoA ring form and glycosidic bond conformation [51].

The physical basis for the relatively restricted rotation around glycosidic bonds in the S-domains is not clearly established yet. Mobli et al. [3] prefer mutual repulsion of negatively charged substituents rather than specific interactions such as hydrogen bonds between residues. However, removal of all sulfates still retains the distinctive heparin pattern of proton–proton NOEs in the NMR spectra of chemically modified heparins [32]. In their MD study of IdoA2S in oligosaccharides, Pol-Fachin and Verli [42] identify hydrogen bonds between IdoA2S and neighboring residues that stabilize the $^2S_0$ conformation and no doubt also the glycosidic linkage conformation.

The heparin–protein interaction with the most current medical and commercial significance also has the best-defined monosaccharide sequence dependence yet identified. This is the pentasaccharide with high affinity for antithrombin. It has been argued that this unusual finding has held back progress in understanding heparin–protein interactions by providing an unproductive, sequence-based paradigm

[8,54]. The key monosaccharide residue is the central, N-sulfated, 3,6-di-O-sulfated glucosamine in the pentasaccharide. A careful analysis of the contribution of the 3-*O*-S group by Richard and co-workers [55] showed that its absence reduces the affinity of the synthetic pentasaccharide by several orders of magnitude. However, the pentasaccharide lacking 3-*O*-sulfate was still capable of activating antithrombin, inducing fluorescence enhancement and inhibition of coagulation factor Xa. Other, more subtle changes in the pentasaccharide structure, introduced by chemical depolymerization to make low-molecular-weight heparin, also reduce affinity for antithrombin and anti-Xa activity [56], explicable by detailed NMR analysis of interactions between the oligosaccharides and antithrombin.

CONFORMATION AND DYNAMICS OF THE UNSULFATED, "NA" DOMAINS

The properties of the NA domains have been investigated using ^{15}N-enriched oligosaccharides prepared from the *E. coli* K5 capsular polysaccharide, which has exactly the same sequence as heparosan (see above) and the unsulfated NA-domains of HS [3] (Fig. 1). This allowed detailed NMR analysis of short oligosaccharides (tetra-, hexa-, and octasaccharides) through heteronuclear filtering, in spite of the heavily overlapped nature of their proton NMR spectra. DOSY (diffusion ordered spectroscopy) revealed a relationship between translational diffusion times and molecular size typical of random coil behavior, implying that internal mobility around the glycosidic bonds may be greater in NA-domains than in S-domains. Analysis of ^{15}N relaxation times making different assumptions about the anisotropy of solution reorientation for the oligosaccharides indicated that the tetra- and hexasaccharides behaved isotropically, but that some degree of anisotropy was demonstrated by the octasaccharide. This anisotropy was also seen in an earlier study of K5 polysaccharide using ^{13}C relaxation [57], which led to the conclusion that internal motions in this sequence are more limited in the NA-domain sequence than in the S-domain sequence.

SOLUTION BEHAVIOR OF THE POLYSACCHARIDES HEPARIN AND HS

GAGs such as heparin and HS can be thought of in more than one way; either as sulfated sequences (S-domains) with differential affinities for protein ligands, set in amongst nonbinding linker sequences (NA-domains); or as linear patterns of negative charge, that is to say as anionic polyelectrolytes [8]. Whichever approach is taken, the solution behavior of the polysaccharide as a whole needs to be studied, and the NMR studies described above make it clear that the two major domain types make different conformational and dynamic contributions. Two other approaches, ultracentrifugation and scattering, have yielded new insights on polysaccharide conformations.

Analytical ultracentrifugation is a dynamic method that studies the sedimentation of macromolecules in solution when they are subjected to a high centrifugal force inside a rotor [58]. At the beginning of the experiment, the macromolecule is present at a uniform concentration throughout the sample cell. During the experiment, the sedimentation of the macromolecule is monitored. The size-distribution analyses of the boundary scans obtained from experiments performed using high rotor speeds result in the sedimentation coefficients (a monitor of shape) and molecular masses of the species that are present. A major advantage of this technique is that sample polydispersity can be monitored. Sedimentation coefficients can now be routinely modeled in terms of previously determined high-resolution structures.

Solution X-ray scattering is a low-resolution structural technique that provides structural information on the size, shape, and interactions of biological macromolecules and other polymer or detergent systems in random orientations in solution. It has been used to determine the structure of proteins, nucleic acids, polysaccharides, and their complexes in solution under near physiological conditions without the need to crystallize the molecules. If constrained modeling is used to fit the scattering curves, the structural resolution is much improved [59]. A major advantage of this technique is its capability to provide structural information on partially or completely disordered systems.

An initial ultracentrifugation and scattering study of an extensive series of unfractionated polydisperse heparin fragments provided macroscopic parameters for these fragments in the absence of molecular modeling [60]. The hydrodynamic behavior of heparin in 0.2 M NaCl was described by a worm-like chain model with a segment length of 9 ± 2 nm that corresponded to a semi-rigid polymer. A more detailed analytical ultracentrifugation and X-ray scattering study [61] was recently applied to heparin [14] using oligosaccharides of defined lengths prepared from heparin by brief depolymerization with heparinase I. Ultracentrifugation revealed that the fragments were monodisperse from the single peaks seen in *c(s)* analyses. Short oligosaccharides behave as almost completely rigid rods, but show an increased propensity to bend at longer lengths. As the NA-domains in the sample were not removed, and heparinase I cleaved heparin within S-domains, there is certainly some NA region present in the longer oligosaccharides (Fig. 4A); it is not possible to say whether this is the source of the flexibility.

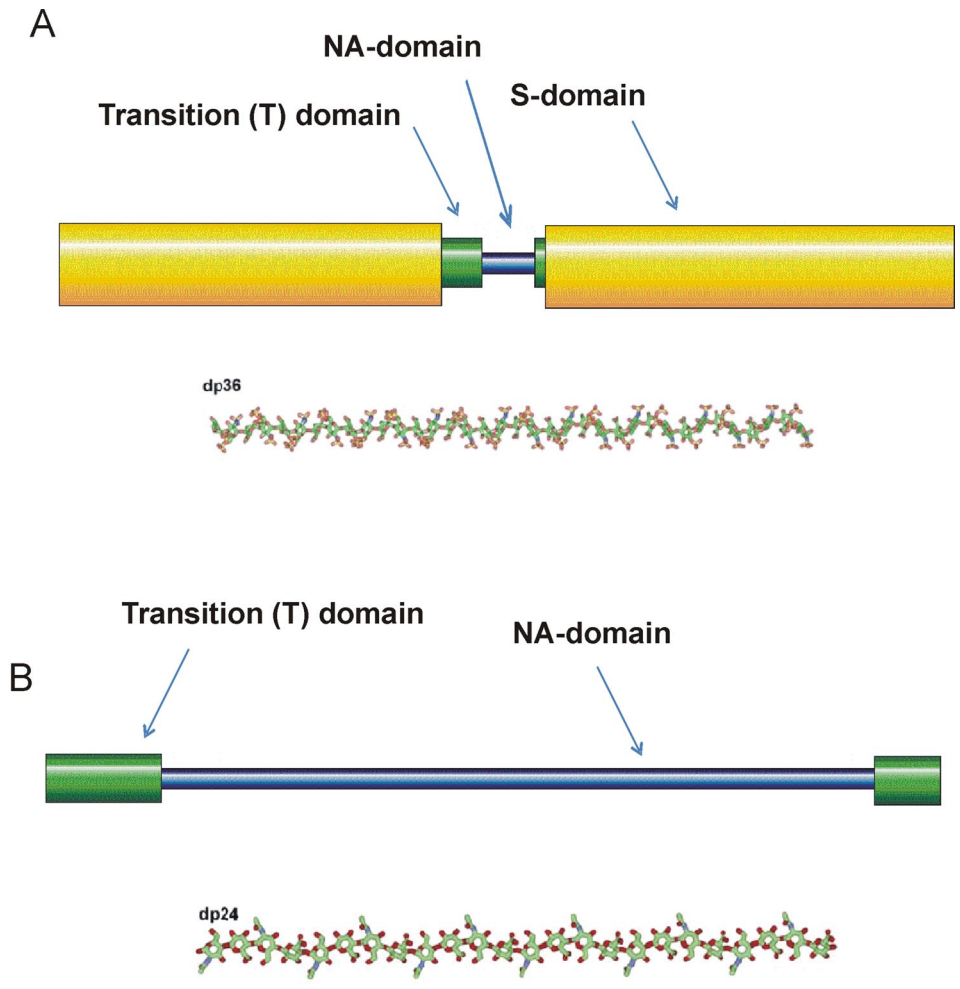

Fig. 4 Cartoon representations for domain structures, and corresponding linear molecular models used as initial structures in restrained molecular modeling [14,62], of fragments prepared for AUC and X-ray scattering studies. In the cartoons, NA domains are depicted as inner, blue rods; S-domains as outer, yellow cylinders; transition domains between the two in green. Coloring for the models is defined in Fig. 2. (A) Heparin partially depolymerized by heparinase I (a dp36 model, the longest studied, is shown), leaving intact small internal NA sequences. (B) HS digested to exhaustion by heparinase I (a dp24 model, the longest studied, is shown), thus removing S-domains and leaving only some transitional domain at the ends of the fragments.

The solution structures for heparin oligosaccharides containing from 6 to 36 monosaccharides were successfully modeled starting from linear structures (Fig. 4B), indicating that these possessed semi-rigid extended structures with slight degrees of bend (Fig. 2C).

HS fragments were also prepared [62], in this case using exhaustive digestion with heparinase I to remove S-domains almost completely. The NMR spectra were consistent with largely NA sequences, with vestiges of transitional sequences at the reducing and nonreducing ends, but no intact S-domain (Fig. 4B). These oligosaccharides also behaved as if they were rigid at short lengths, but deviated from rigid behavior at shorter lengths than the heparin fragments and showed structures that were longer and more bent than the equivalent-sized heparin fragment (Figs. 2C,D). Interestingly, while constrained modeling yielded good scattering fits for HS up to 16 monosaccharides in size, constrained modeling was not able to fit any single HS molecular model for degree of polymerization (dp) of over 16 monosaccharides to the scattering curves. Thus, unlike heparin, larger HS structures appear to show too great a level of flexibility to be modeled as a single structure.

STATIC COORDINATES AND SOLUTION STRUCTURES

Coordinates for an ensemble of modeled structures were deposited in the Protein Data Bank (PDB; <http://www.pdb.org/pdb/home>) for X-ray scattering based models of heparin [14] at several different oligosaccharide lengths; 3IRI.pdb for the 18-mer, 3IRJ.pdb for the 24-mer, 3IRK.pdb for the 30-mer, and 3IRL.pdb for the 36-mer. Compared with the completely straight rods used as starting models (Fig. 4), the models derived from X-ray scattering data are more uneven, containing more or less sharp bends (Figs. 2C,D). PDB-formatted coordinates for 6-mer to 16-mer structures for HS are likewise available [62].

The interpretation of coordinates deposited into the PDB for solution structures is not straightforward. Sets of coordinates in the PDB must of course be wholly compatible with the corresponding experimental data set. However, the dynamics of macromolecules in solution generally, whether peptide- or saccharide-based, means that some or all of the molecule is internally flexible, so a single structure is not a realistic representation. This issue has often been surmounted through depositing an ensemble of structures. In this way, the relative uncertainty of parts of the structure can be indicated, though it is not always clear whether this uncertainty is due to molecular flexibility or to lack of experimental data. For proteins, this usually results in a well-defined set of coordinates for compact, globular domains, with long loops and terminal sequences adopting a widely differing set of conformations. Each individual member of the ensemble represents one possible conformation, which if adopted by every molecule in the sample, would give good agreement with the experimental data set. For the linear polysaccharide heparin, with no globular domains, the individual ensemble members generated by molecular modeling on the basis of X-ray scattering data are highly artificial. They are linear, like the NMR structure, but with occasional more or less sharp bends (Fig. 2C); at those bends, the glycosidic linkages are not at their minimum energy configuration, and so are not stable. Taken as a whole, the ensemble of structures gives a good idea of the range of solution conformations accessible by the polysaccharide. Taken as individual conformations, they must be interpreted with caution. In contrast to the case for globular proteins, the concept of an average heparin or HS conformation derived from the whole ensemble for use in further modeling studies (docking, dynamics, and so on) makes no sense at all, and such an averaged conformation has therefore not been determined or deposited.

There are also cases in which a representative set of conformations has to stand in for a much more complex situation. The solution structure of the fully trisulfated disaccharide repeating unit of heparin (the S-domain structure) was investigated by short-range NOE determinations [30] using a sample of polydisperse heparin. The resulting two dodecasaccharide models in 1HPN.pdb, one with all the IdoA residues in the $^{1}C_{4}$ chair conformation and the other with all the IdoA residues in the $^{2}S_{0}$ form, are not realistic representations of the material studied. Two recent publications have, however, assumed that the experiments were done using a dodecasaccharide sample [8,52].

THE BIG PICTURE: WAYS IN WHICH DIFFERENT HS ARCHITECTURES CAN INFLUENCE PROTEIN BINDING

In conclusion, the model of HS that has emerged from NMR, ultracentrifugation, and scattering, and is now well accepted, is that of long, flexible NA-domains and short, relatively rigid S-domains. This seems appropriate for a cell surface molecule that needs to find and interact with small proteins in the extracellular environment. This combination of flexible linkers connected to highly sulfated, protein-binding motifs can sweep out more conformational space, can reach between distant heparin binding sites, and can "go round corners" to access binding sites on different faces of a protein multimer. On the other hand, in the structure of heparin, a combination of large S-domains with smaller NA-domains is likely to provide regularity of structure, appropriate to close packing with basic proteases in the mast cell granule [63].

Interactions in which HS is required to bind simultaneously to two protein molecules are readily envisaged for the model of HS in which S-domains are separated by more flexible, less highly sulfated NA link regions. The ternary complex between heparin (or HS), antithrombin, and thrombin is a well-studied example. Oligomerization of some small proteins may be enhanced by HS, and the "SAS" motif fits comfortably with a number of HS-dimer complexes. A review by Imberty et al. [15] illustrated a number of examples. For several of these models, the relatively rigid, continuous S-domain structure of heparin does not seem optimal; the presence of a flexible NA linker allows the GAG structure to bend around the protein oligomer to access two or more heparin binding sites.

The search for single heparin and HS sequences with specific affinity for proteins is unlikely to be successful in many, perhaps most cases. The 3D motifs recognized by proteins can be formed by several different sequences (Fig. 3C). Because the overwhelming contribution to heparin–protein interactions is charge-based, it is not possible to distinguish from sequence-based reasoning between interactions with protein that depend on a selective complementarity of charged surfaces or the simple, nonspecific attraction of opposite charges. Over the past decade or so, attention has turned away from the exploration of fine structure of heparin and HS, given that sequence-based strategies in the search for the molecular basis of HS–protein interactions have proved inconclusive. A recent review article has summed this up expressively, bringing back to general attention the advantages of a rigorous treatment of GAGs as polyelectrolytes [8]. The depiction of protein–HS interactions as being based solely on short-range, pairwise directional interactions between atoms is rightly criticized in this work. Such an approach may seem to be superficially precise, but in fact is likely to underestimate contributions of electrostatics to binding, and underplay the influence of entropic rather than enthalpic factors.

The reductionist approach to the design of drugs based on the structure of heparin/HS has been successful in the case of the antithrombin-mediated anticoagulant action of heparin, which is dependent on an oligosaccharide sequence in which little variability is tolerated, so allowing the design of a synthetic mimic. Some current work takes a different approach, looking for ways to reproduce heparin structures in a controlled way, amenable to licensing and regulatory requirements, but retaining some of the strengths of complexity [64]. Molecular size, spacing, and extent of S-domains, and fine structure within and around S-domains, will all play a part in the spectra of physiological activities displayed by these new heparinoids.

REFERENCES

1. K. Sugahara, H. Kitagawa. *IUBMB Life* **54**, 163 (2002).
2. C. L. Merry, M. Lyon, J. A. Deakin, J. J. Hopwood, J. T. Gallagher. *J. Biol. Chem*. **274**, 18455 (1999).
3. M. Mobli, M. Nilsson, A. Almond. *Glycoconjugate J.* **25**, 401 (2008).
4. D. R. Al, R. Wagenvoord, G. W. van Dedem, S. Beguin, H. C. Hemker. *J. Thromb. Haemost.* **1**, 907 (2003).

5. K. L. Kramer. *Semin. Cell Dev. Biol.* **21**, 631 (2010).
6. J. Kreuger, D. Spillmann, J. P. Li, U. Lindahl. *J. Cell Biol.* **174**, 323 (2006).
7. B. Mulloy. *An. Acad. Bras. Cienc.* **77**, 651 (2005).
8. E. Seyrek, P. Dubin. *Adv. Colloid Interface Sci.* **158**, 119 (2010).
9. N. S. Gandhi, R. L. Mancera. *Chem. Biol. Drug Des.* **72**, 455 (2008).
10. M. A. Skidmore, S. E. Guimond, T. R. Rudd, D. G. Fernig, J. E. Turnbull, E. A. Yates. *Connect. Tissue Res.* **49**, 140 (2008).
11. D. R. Coombe, W. C. Kett. *Cell Mol. Life Sci.* **62**, 410 (2005).
12. B. Mulloy, M. J. Forster. *Glycobiology* **10**, 1147 (2000).
13. I. Capila, R. J. Linhardt. *Angew. Chem., Int. Ed.* **41**, 391 (2002).
14. S. Khan, J. Gor, B. Mulloy, S. J. Perkins. *J. Mol. Biol.* **395**, 504 (2010).
15. A. Imberty, H. Lortat-Jacob, S. Perez. *Carbohydr. Res.* **342**, 430 (2007).
16. E. Gray, B. Mulloy, T. W. Barrowcliffe. *Thromb. Haemost.* **99**, 807 (2008).
17. T. R. Rudd, M. A. Skidmore, M. Guerrini, M. Hricovini, A. K. Powell, G. Siligardi, E. A. Yates. *Curr. Opin. Struct. Biol.* **20**, 567 (2010).
18. S. Nadanaka, H. Kitagawa. *J. Biochem.* **144**, 7 (2008).
19. J. Sheng, R. Liu, Y. Xu, J. Liu. *J. Biol. Chem.* **286**, 19768 (2011).
20. P. Carlsson, J. Presto, D. Spillmann, U. Lindahl, L. Kjellen. *J. Biol. Chem.* **283**, 20008 (2008).
21. M. Warda, T. Toida, F. Zhang, P. Sun, E. Munoz, J. Xie, R. J. Linhardt. *Glycoconjugate J.* **23**, 555 (2006).
22. A. Parra, N. Veraldi, M. Locatelli, M. Fini, L. Martini, G. Torri, L. Sangiorgi, A. Bisio. *Glycobiology* (2012). In press.
23. G. O. Staples, X. Shi, J. Zaia. *J. Biol. Chem.* **285**, 18336 (2010).
24. M. Lyon, J. A. Deakin, J. T. Gallagher. *J. Biol. Chem.* **269**, 11208 (1994).
25. J. Schlessinger, A. N. Plotnikov, O. A. Ibrahimi, A. V. Eliseenkova, B. K. Yeh, A. Yayon, R. J. Linhardt, M. Mohammadi. *Mol. Cell* **6**, 743 (2000).
26. N. C. Smits, S. Kurup, A. L. Rops, G. B. ten Dam, L. F. Massuger, T. Hafmans, J. E. Turnbull, D. Spillmann, J. P. Li, S. J. Kennel, J. S. Wall, N. W. Shworak, P. N. Dekhuijzen, J. van der Vlag, T. H. van Kuppevelt. *J. Biol. Chem.* **285**, 41143 (2010).
27. K. J. Murphy, C. L. Merry, M. Lyon, J. E. Thompson, I. S. Roberts, J. T. Gallagher. *J. Biol. Chem.* **279**, 27239 (2004).
28. W. C. Lamanna, I. Kalus, M. Padva, R. J. Baldwin, C. L. Merry, T. Dierks. *J. Biotechnol.* **129**, 290 (2007).
29. S. D. Rosen, H. Lemjabbar-Alaoui. *Exp. Opin. Ther. Targets* **14**, 935 (2010).
30. B. Mulloy, M. J. Forster, C. Jones, D. B. Davies. *Biochem. J.* **293** (Pt. 3), 849 (1993).
31. B. Mulloy, R. J. Linhardt. *Curr. Opin. Struct. Biol.* **11**, 623 (2001).
32. B. Mulloy, M. J. Forster, C. Jones, A. F. Drake, E. A. Johnson, D. B. Davies. *Carbohydr. Res.* **255**, 1 (1994).
33. A. D. DiGabriele, I. Lax, D. I. Chen, C. M. Svahn, M. Jaye, J. Schlessinger, W. A. Hendrickson. *Nature* **393**, 812 (1998).
34. J. Kreuger, M. Salmivirta, L. Sturiale, G. Gimenez-Gallego, U. Lindahl. *J. Biol. Chem.* **276**, 30744 (2001).
35. T. R. Rudd, K. A. Uniewicz, A. Ori, S. E. Guimond, M. A. Skidmore, D. Gaudesi, R. Xu, J. E. Turnbull, M. Guerrini, G. Torri, G. Siligardi, M. C. Wilkinson, D. G. Fernig, E. A. Yates. *Org. Biomol. Chem*. **8**, 5390 (2010).
36. R. Lucas, J. Angulo, P. M. Nieto, M. Martin-Lomas. *Org. Biomol. Chem*. **1**, 2253 (2003).
37. J. Angulo, R. Ojeda, J. L. de Paz, R. Lucas, P. M. Nieto, R. M. Lozano, M. Redondo-Horcajo, G. Gimenez-Gallego, M. Martin-Lomas. *ChemBioChem* **5**, 55 (2004).
38. J. Angulo, M. Hricovini, M. Gairi, M. Guerrini, J. L. de Paz, R. Ojeda, M. Martin-Lomas, P. M. Nieto. *Glycobiology* **15**, 1008 (2005).

39. G. Gatti, B. Casu, A. S. Perlin. *Biochem. Biophys. Res. Commun*. **85**, 14 (1978).
40. C. F. Becker, J. A. Guimaraes, H. Verli. *Carbohydr. Res.* **340**, 1499 (2005).
41. H. Verli, J. A. Guimaraes. *Carbohydr. Res.* **339**, 281 (2004).
42. L. Pol-Fachin, H. Verli. *Carbohydr. Res.* **343**, 1435 (2008).
43. B. M. Sattelle, S. U. Hansen, J. Gardiner, A. Almond. *J. Am. Chem. Soc.* **132**, 13132 (2010).
44. J. Angulo, P. M. Nieto, M. Martin-Lomas. *Chem. Commun.* 1512 (2003).
45. N. S. Gandhi, R. L. Mancera. *Carbohydr. Res.* **345**, 689 (2010).
46. M. Forster, B. Mulloy. *Biopolymers* **33**, 575 (1993).
47. B. M. Sattelle, S. U. Hansen, J. Gardiner, A. Almond. *J. Am. Chem. Soc.* **132**, 13132 (2010).
48. M. Hricovini. *Carbohydr. Res.* **341**, 2575 (2006).
49. M. Hricovini, F. Bizik. *Carbohydr. Res.* **342**, 779 (2007).
50. M. Hricovini, E. Scholtzova, F. Bizik. *Carbohydr. Res.* **342**, 1350 (2007).
51. M. Hricovini. *J. Phys. Chem. B* **115**, 1503 (2011).
52. L. Jin, M. Hricovini, J. A. Deakin, M. Lyon, D. Uhrin. *Glycobiology* **19**, 1185 (2009).
53. T. R. Rudd, S. E. Guimond, M. A. Skidmore, L. Duchesne, M. Guerrini, G. Torri, C. Cosentino, A. Brown, D. T. Clarke, J. E. Turnbull, D. G. Fernig, E. A. Yates. *Glycobiology* **17**, 983 (2007).
54. E. Seyrek, P. L. Dubin, J. Henriksen. *Biopolymers* **86**, 249 (2007).
55. B. Richard, R. Swanson, S. T. Olson. *J. Biol. Chem.* **284**, 27054 (2009).
56. M. Guerrini, S. Elli, D. Gaudesi, G. Torri, B. Casu, P. Mourier, F. Herman, C. Boudier, M. Lorenz, C. Viskov. *J. Med. Chem*. **53**, 8030 (2010).
57. M. Hricovini, M. Guerrini, G. Torri, B. Casu. *Carbohydr. Res.* **300**, 69 (1997).
58. S. J. Perkins, R. Nan, K. Li, S. Khan, Y. Abe. *Methods* **54**, 181 (2011).
59. S. J. Perkins, A. I. Okemefuna, R. Nan, K. Li, A. Bonner. J. R. Soc. *Interface* **6**, Suppl. 5, S679 (2009).
60. G. Pavlov, S. Finet, K. Tatarenko, E. Korneeva, C. Ebel. *Eur. Biophys. J.* **32**, 437 (2003).
61. S. J. Perkins, R. Nan, K. Li, S. Khan, Y. Abe. *Methods* **54**, 181 (2011).
62. S. Khan, E. Rodriguez, R. Patel, J. Gor, B. Mulloy, S. J. Perkins. *J. Biol. Chem.* **286**, 24842 (2011).
63. R. L. Stevens, R. Adachi. *Immunol. Rev.* **217**, 155 (2007).
64. R. Liu, Y. Xu, M. Chen, M. Weiwer, X. Zhou, A. S. Bridges, P. L. DeAngelis, Q. Zhang, R. J. Linhardt, J. Liu. *J. Biol. Chem.* **285**, 34240 (2010).
65. A. Varki, R. D. Cummings, J. D. Esko, H. H. Freeze, P. Stanley, J. D. Marth, C. R. Bertozzi, G. W. Hart, M. E. Etzler. *Proteomics* **9**, 5398 (2009).

