Crystal Structures of the Heparan Sulfate-binding Domain of Follistatin

INSIGHTS INTO LIGAND BINDING*

Received for publication, November 5, 2002, and in revised form, July 14, 2003
Published, JBC Papers in Press, July 16, 2003, DOI 10.1074/jbc.M211284200

C. Axel Innis‡ and Marko Hyvönen§

From the Department of Biochemistry, University of Cambridge, 80 Tennis Court Road, Cambridge CB2 1GA, United Kingdom

Follistatin associates with transforming growth factor-β-like growth factors such as activin or bone morphogenetic proteins to form an inactive complex, thereby regulating processes as diverse as embryonic development and cell secretion. Although an interaction between heparan sulfate chains present at the cell surface and follistatin has been recorded, the impact of this binding reaction on the follistatin-mediated inhibition of transforming growth factor-β-like signaling remains unclear. To gain a structural insight into this interaction, we have solved the crystal structure of the presumed heparan sulfate-binding domain of follistatin, both alone and in complex with the small heparin analogs sucrose octasulfate and D-myoinositol hexasulfate. In addition, we have confirmed the binding of the sucrose octasulfate and D-myoinositol hexasulfate molecules to this follistatin domain and determined the association constants and stoichiometries of both interactions in solution using isothermal titration calorimetry. Overall, our results shed light upon the structure of this follistatin domain and reveal a novel conformation for a hinge region connecting epidermal growth factor-like and Kazal-like subdomains compared with the follistatin-like domain found in the extracellular matrix protein BM-40. Moreover, the crystallographic analysis of the two protein-ligand complexes mentioned above leads us to propose a potential location for the heparan sulfate-binding site on the surface of follistatin and to suggest the involvement of residues Asn80 and Arg96 in such a follistatin-heparin interaction.

The secreted polypeptide follistatin regulates several signaling pathways in a cell- and tissue-specific manner, largely through its ability to inactivate transforming growth factor-β-like growth factor molecules (1, 2). Although first identified as a factor capable of countering the inducing effects of activin on follicle-stimulating hormone secretion (3, 4), follistatin has now been isolated and characterized from a variety of tissues and organisms, where it has been shown to take part in processes such as cell growth, differentiation, and secretion (5–7). Moreover, the importance of this polypeptide has been demonstrated in follistatin-null mice, which can survive birth but die from multiple skeletal and cutaneous abnormalities within a few hours of delivery (8).

Follistatin isoforms sharing the same N terminus but differing in the length of their polypeptide chain and in their glycosylation patterns coexist in vivo, with molecular masses ranging from 31 to 39 kDa (4, 9). Indeed, alternative splicing of the follistatin gene yields the Fs-315 and Fs-288 isoforms, featuring 315 and 288 residues, respectively. An additional variant of follistatin, Fs-303, is thought to arise from the proteolytic cleavage of the Fs-315 C terminus and is likely to be the predominant species in vivo (3, 10, 11). Despite its relatively limited size, follistatin contains a total of 36 cysteine residues, believed to be arranged into nonoverlapping sets of disulfide bridges corresponding to four autonomous folding units (Fig. 1). The first of these units, which we call Fs0, comprises the 63 N-terminal residues of the mature polypeptide and bears no sequence similarity with any other protein of known structure. In contrast, the rest of the follistatin chain appears to fold into a series of three consecutive 70–74-residue-long structural repeats, referred to as Fs1, Fs2, and Fs3, which display homology to the follistatin-like domain of the extracellular matrix protein BM-40 (12) and are also found in several other extracellular matrix proteins, such as agrin (13), tenomreulin (14), and complement proteins C6 and C7 (15). Each of these repeats can further be divided into two subregions: a ~30-residue-long fragment reminiscent of EGF-like modules and a ~40-residue domain homologous to Kazal protease inhibitor domains. In both Fs-315 and Fs-303, the last follistatin repeat is flanked on its C-terminal side by a short, glutamate-rich tail.

A closer look at the primary structures for the three follistatin repeats reveals striking differences in the values of their theoretical isoelectric point (pI), which ranges from basic for Fs1 (pI 8.9) to neutral for Fs2 (pI 6.7) and acidic for Fs3 (pI 4.8). Coupled to the observation that follistatin interacts with the heparan sulfate chains of cell surface proteoglycans (16, 17), these pI values suggest a model in which a heparin-binding site is located within the most basic of the follistatin repeats, Fs1. Interestingly, Fs-288 has a greater affinity for heparan sulfate than Fs-315, suggesting that the acidic C terminus of the latter

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EBI Data Bank with accession number(s)204173. The atomic coordinates and structure factors (code 1LR7, 1LR8, and 1LR9) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

‡ Supported by the Biotechnology and Biological Sciences Research Council (BBSRC), the Cambridge European Trust and the Sackler Fund. Present address: National Centre for Biological Sciences, Tata Institute of Fundamental Research, UAS-GVK Campus, Bellary Rd., Bangalore 560065, India.

§ Recipient of a BBSRC David Phillips research fellowship. To whom correspondence should be addressed. Tel.: 44-1223-766-031; Fax: 44-1223-766-002; E-mail: marko@cryst.bioc.cam.ac.uk.

1 The abbreviations used are: EGF, epidermal growth factor; Ins68S, D-myoinositol hexasulfate; ITC, isothermal titration calorimetry; r.m.s.d., root mean squared deviation; SOS, sucrose octasulfate.

This paper is available on line at http://www.jbc.org
can mask the heparin-binding site of follistatin by mimicking the structure and electrostatic properties of sulfated glycans (18). Although the significance of a follistatin-heparin interaction remains relatively obscure, the affinity of Fs-288 for activin appears to be affected by heparin (19). In addition, Fs-288 has been shown to stimulate the endocytotic degradation of activin in a time- and dose-dependent manner by virtue of its association with sulfated proteoglycans present on the cell surface (18).

To understand the mechanism of follistatin action, we have used multiple anomalous dispersion to solve the three-dimensional structure of the Fs1 domain, both alone and in complex with the small heparan sulfate analogs sucrose octasulfate (SOS) and D-xylose-inositol hexasulfate (Ins6S). Moreover, we have investigated the binding of these two molecules to Fs1 through the use of isothermal titration calorimetry (ITC), thus revealing the tight nature of both interactions. Overall, these experiments reveal the location of a potential heparin-binding site on Fs1 and highlight the importance of specific polar groups in its association with sulfated proteoglycans. The results presented in this paper provide us with a more detailed view of follistatin-heparin interactions and give clues about the nature of the interaction with heparan sulfate.

EXPERIMENTAL PROCEDURES

All of the chemicals used were of research grade or higher and were purchased from BDH, Difco, Fluka, Melford Laboratories, or Sigma, unless specified otherwise. Ins6S and SOS, both as potassium salts, were purchased from Sigma and Toronto Research Chemicals, respectively. A-Sulfsalt, and 2-O-Methyl-d-threonine (2OMe-D-Thr) were obtained from Chem-Impex. 18 mercaptoundecanol was purchased from BDH, Difco, Fluka, Melford Laboratories, or Sigma, unless specified otherwise. Ins6S and SOS, both as potassium salts, were purchased from Sigma and Toronto Research Chemicals, respectively. A-Sulfate and 2-O-Methyl-d-threonine (2OMe-D-Thr) were obtained from Chem-Impex. 18 mercaptoundecanol was purchased from Sigma.

Preparation of Recombinant Protein—The coding sequence for Fs1, comprising residues 64-136 of the mature follistatin gene product fused to a 5-AUG codon, was amplified using a standard PCR protocol and inserted into the multiple cloning sequence of the T7-based expression vector pBAT4 (21). Expression of soluble recombinant protein was subsequently carried out in an Escherichia coli BL21(DE3) host and was induced at 37 °C with 400 μM isopropyl-1-thio-β-D-galactopyranoside. Soluble Fs1 was then purified as follows: (i) cation exchange chromography on an SP-Sepharose matrix (Amersham Biosciences), using a 0.0-1.0 M NaCl gradient; (ii) separation in a Superdex 75 (HiLoad 16/60) size exclusion column (Amersham Biosciences); (iii) reversed phase HPLC in a Hypersil C8 column, using a 5–100% acetonitrile gradient; (iv) lyophilization and subsequent resublimation in water. Selenomethionyl-Fs1 was prepared according to the metallic inhibitor protocol described by Doublie (22) and purified in a fashion similar to native Fs1.

Protein Crystalization—Crystals of Fs1, Fs1/Ins6S, and Fs1/SOS were obtained by vapor diffusion using the hanging drop method. In all cases, equal volumes of the mother liquor and of a solution containing 1.8 mM (15 mg/ml) protein and 2 mM ligand were mixed and left to equilibrate for a day or 2 at 18 °C. The final, optimized crystalization conditions were: 30–35% PEG 4000, 0.4 mM magnesium chloride, 0.1 M Tris-HCl (pH 8.5) for Fs1 crystals; 15–25% PEG 8000, 0.2–0.6 M magnesium acetate, 0.1 M sodium cacodylate (pH 6.5) for Fs1/Ins6S crystals; and 30–35% 2-propanol, 0.6–0.7 M ammonium acetate, 0.1 M Tris-HCl (pH 8.5) for Fs1/SOS crystals. Cryoprotection of all three crystal types was achieved through the gradual addition of 20% glycerol to the mother liquor.

Phasing and Refinement—The structure of Fs1/SOS was determined by multiple anomalous dispersion, using a single crystal of selenomethionyl-Fs1 in complex with SOS. A three-wavelength multiple anomalous dispersion experiment was performed at beamline BM14 of the European Synchrotron Radiation Facility (ESRF, Grenoble, France). Diffraction images were integrated, and reflection intensities were merged using the HKL2000 package (23). The position of one of the two selenium atoms present in the selenomethionyl-Fs1/SOS asymmetric unit was determined using SnB (24) and improved by solvent flattening with Solomon (26), using data collected from a single native Fs1/SOS crystal. Phases were calculated with Sharp (25) and improved by solvent flattening with Solomon (26), using a solvent content of 35%. The resulting 2.5 Å electron density map was clearly interpretable, and an initial model could be built into the density using the program O (27). Phases were then extended to 1.5 Å with ARP/wARP (28) using data collected from a single native Fs1/SOS crystal at the ESRF beamline ID-14-4. The structure was subsequently optimized through several rounds of restrained refinement against native amplitudes, treating isotropic B factors individually for all atoms and applying a standard bulk solvent model. Refinement was carried out both in CNS (29) and Refmac 5.0 (30), with manual rebuilding and adjustments in O. Water addition was performed using the program ARP (28) followed by visual inspection and editing in O.

Once a suitable model of Fs1 was obtained, its atomic coordinates were used to compute initial phases for the Fs1/Ins6S and Fs1 structures. Molecular replacement and rigid body refinement were performed in CNS, using the native data sets collected for Fs1/Ins6S and
The Heparan Sulfate-binding Domain of Follistatin

Table I

Crystallographic statistics for the Fs1, Fs1/SOS, and Fs1/Ins6S structures

| Fs1 Native | Fs1/Ins6S Native | Fs1/SOS |
|------------|-----------------|---------|
| Diffraction data | | |
| Cell dimensions (space group: P2₁2₁2₁, 1 molecule/asymmetric unit) | | |
| a (Å) | 21.55 | 21.59 | 21.66 |
| b (Å) | 43.73 | 38.15 | 38.48 |
| c (Å) | 77.38 | 78.50 | 78.54 |
| Volume (Å³) | 114,816 | 114,852 | 114,878 |
| Wavelength (Å) | 0.9785 | 0.9787 | 0.8660 |
| Resolution (outer shell) (Å) | 32.8–2.50 | 39.2–2.50 | 39.2–2.50 |
| Reflections (R_re = set) | 2,639 (126) | 9,913 (499) |
| Non-hydrogen protein atoms | 523 | 538 | 528 |
| Ligand sulfates | 5 | 5 | 5 |
| Waters | 28 | 6 | 78 |
| Rcryst | 0.220 | 0.221 | 0.189 |
| Rfree | 0.258 | 0.251 | 0.222 |
| Average B factor for all atoms (Å²) | 22.76 | 24.85 | 20.24 |
| Bond angle r.m.s.d. (°) | 1.65 | 1.27 | 1.94 |
| Bond distance r.m.s.d. (Å) | 0.019 | 0.026 | 0.021 |

Data were processed with the program Denzo (23) and scaled with Scalepack. (23) In the case of the MAD data sets, a more recent distribution was used. Phasing and refinement statistics were obtained from Sharp (25) and Refmac-5.0 (30), respectively.

* Data from high (1.5 Å) and low (3.0 Å) resolution data sets were merged to collect as many high resolution reflections as possible without causing pixel saturation for the low resolution reactions.
* Three of these groups comprise an additional carbon atom attached to the sulfate (CO₃SO₃⁻).
* This includes two sulphate groups refined with half occupancy.
* Calculated using Engh and Huber bond and angle parameters (36).

Fs1 at the ESRF beamlines ID29 and ID-14-4, respectively. Refinement was then carried out as described for Fs1/SOS. In all cases, the quality of the final models was assessed using the programs Procheck (34) and WHATCHECK (32, 33). Superimposition of the final Fs1 models with their homologs and root mean squared deviation (r.m.s.d.) calculations were carried out using the program LEGNAN (34).

Isothermal Titration Calorimetry of Protein-Ligand Interactions—Purified Fs1 was dialyzed extensively against a buffer containing 0.1 M Tris-HCl (pH 8.5) and diluted down to a final working concentration of 0.025 mM, as determined by absorbance measurements using the Gill and von Hippel (35) theoretical extinction coefficient of Fs1 at 280 nm (ε₂₈₀ = 8,850 M⁻¹ cm⁻¹). SOS and Ins6S in powder form were weighed out accurately and solubilized in the dialysis buffer to the final concentrations of 0.2 and 0.4 mM, respectively. Calorimetric titrations were performed using these solutions in a MicroCal VP-ITC microcalorimeter. The sample cell of the calorimeter was filled completely with the Fs1 solution up to a volume of 1.4 ml, and the system was allowed to equilibrate thermally at 25 °C. 10-μl pulses of the ligand solution were injected into the sample at 3-min intervals. Raw ITC data were integrated using the Microcal Origin software, background heats from ligand to buffer titrations were subtracted, and the corrected heats from the binding reaction were used to derive values for the stoichiometry of the binding (n), the association constant (Kₐ), the apparent enthalpy of binding (ΔH) and the change in Gibbs free energy occurring upon binding (ΔG).

As controls, the ligands were titrated into the dialysis buffer without protein, and buffer alone was injected into the protein solution. Possible interaction of potassium (the counter ion for both Ins6S and SOS) with FS1 domain was also studied but found to have no additional effect compared with the buffer alone control.

RESULTS

Fs1 Structure—A brief overview of the three Fs1 crystallographic models described here is given in Table I. All structures display normal bond angle and length variability as defined by Engh and Huber (36), and the majority of φ/ψ dihedral angle combinations occupy the most favored regions of the Ramachandran plot (37). Moreover, given that the Cα atoms for the whole three Fs1 structures can be superimposed with a r.m.s.d. of 0.2 Å for Fs1/SOS and Fs1/Ins6S, and 0.9 Å for Fs1/SOS and Fs1, we shall use the 1.5 Å resolution model of Fs1 in complex with SOS to describe the overall architecture of the Fs1 domain (Fig. 2). Relevant differences among the three structures will be highlighted where appropriate.

The Fs1 fold consists of two distinct subdomains, as expected from sequence comparisons with the follistatin-like domain of BM-40, with whom Fs1 shares 31.5% sequence identity. The N-terminal moiety of Fs1 (residues 64–89) is reminiscent of EGF-like modules, as indicated by an identical disulfide connectivity. This subdomain comprises a stretch of residues in an extended conformation (64–74) held in place by two disulfide linkages to strands β1 (75–79) and β2 (85–89). Refined temperature factors for all three structures are highest for this N-terminal domain and for the extended coil in particular, a phenomenon also observed in the BM-40 structure. As one might expect, this is apparent in the electron density for this...
region in all three structures, which is at its weakest for the type II \( \beta \)-turn (72–75) preceding strand \( \beta \)-1. In addition, conformational differences between the polypeptide backbones of the refined models are clearly visible for loop 64–74.

The C-terminal moiety of Fs1 (89–136) has a fold similar to that of Kazal protease inhibitor domains, consisting of a small three-stranded antiparallel \( \beta \)-sheet packed against an \( \alpha \)-helix. Strand \( \beta \)-3 (102–104) is connected to \( \beta \)-4 (108–110) by a type I \( \beta \)-turn, immediately followed by helix \( \alpha \)-1 (113–122) and strand \( \beta \)-5 (129–132). The lowest temperature factors for the Fs1 structure are seen in this small \( \beta \)-3-\( \beta \)-4-\( \alpha \)-1 core, which is centered on a conserved tyrosine residue (Tyr110). A long stretch of loosely packed amino acids (89–101) is attached to helix \( \alpha \)-1 via two disulfide bridges (Cys89–Cys121 (bond 3) and Cys93–Cys114 (bond 4)). The greatest backbone differences between the liganded and unliganded forms of Fs1 are seen within this region and can mainly be attributed to an alternative conformation of disulfide bond 4. This places the sulfur atom of Cys93 for the superimposed liganded and unliganded structures ≈ 3 Å apart and, as we shall see, is likely to be induced by ligand binding. In the case of Fs1/SOS and Fs1/Ins6S, a short \( 3_{10} \)-helix is also visible between residues 93 and 95. This element of secondary structure is absent from the unliganded structure because of the more extended conformation of the protein backbone for this loop. Finally, a disulfide bond between Cys103 and Cys135 (bond 5) connects strand \( \beta \)-3 to the C terminus of the Fs1 domain, thus completing the follistatin fold.

Structural Homologs of Fs1—The search for structural homologs of Fs1 using the DALI server (38) failed to yield more than three significant hits, none of which was the follistatin-like domain of BM-40. Instead, Kazal domains from the double-domain protease inhibitor rhodniin (39) (PDB code: 1tbr, resi-
of their two subdomains (Fig. 3). If we superimpose the Kazal-like subdomain of Fs1 could be superimposed onto the above Kazal domains and the C-terminal fragment of the BM-40 follistatin-like domain (PDB code: 1bmo) (Fig. 3). Indeed, the greater part of the coil joining residues 90 and 101 in Fs1 follows a different trajectory in its homologs, despite being invariably attached to the rest of the Kazal-like domain through a pair of conserved disulfide bridges (disulfides 3 and 4 in Fs1). Interestingly, the fifth cysteine of the Kazal inhibitor domains does not align in the primary structure with its structural counterparts in Fs1 and BM-40 (Fig. 1). Thus, it appears that the degree of sequence and structural conservation between the Kazal inhibitors and “follistatin-like” molecules does not extend beyond the disulfide bond connectivity for this coil region. This is emphasized further by the apparent lack of tight packing between the coil region and the core of the Kazal-like subdomain and through differences in the spacing of the two cysine.

If we omit the poorly conserved coil region from our comparison, the percentage sequence identity between Fs1 and its various homologs is of the order of 42–47% with 35–36 residues from the three Kazal domains, compared with only 32.4% with the corresponding 35 residues from the C-terminal half of the BM-40 follistatin-like domain. Moreover, the distribution of these values is matched by the Cβ r.m.s.d. calculated for these same fragments relative to Fs1, which lies in the range of 0.45–1.2 Å for the Kazal inhibitors and is greater than 1.2 Å for BM-40. On the whole, these figures argue against the rather arbitrary distinction that is currently made between Kazal and follistatin-like domains such as that seen in BM-40. On the other hand, the presence of the N-terminal EGF-like hairpin subdomain in both Fs1 and BM-40 indicates that these two molecules might share a common ancestor, which may have arisen from ancestral Kazal-type and EGF-like molecules through a domain fusion event. As can be seen in the structure-based alignment in Fig. 1, the four N-terminal cysteines of the follistatin domains and of BM-40 align perfectly with their counterparts in the blood coagulation factor VIIa EGF-like module (42), suggesting a common ancestry.

Interestingly, the SMART data base (43) breaks down each follistatin repeat into two, treating the EGF-like moiety as a distinct follistatin (FOLN) domain. Although separating these two subdomains probably makes sense from the point of view of sequence conservation, we cannot rule out that the succession of EGF-like and Kazal-like subdomains might actually offer a wider range of functional capability compared with the individual subdomains. Such a fusion might indeed confer the rigidity necessary to build long rod-like structures from individual modular repeats, as seen in the case of agrin (44). The hinge region linking the two subdomains may additionally exert some control over the shape of the molecule, allowing specific conformational changes to take place in response to environmental stimuli. As we shall see next, contacts between the two subdomains are scarce, suggesting that minor changes in the hinge region might lead to a drastic shift in the relative orientation of the two subdomains. As far as follistatin itself is concerned, a hinge motion in any of the three repeats could occur during the binding of activin or heparan sulfate, resulting in cooperativity. This, however, has yet to be observed.

**Comparison with the BM-40 Follistatin-like Domain**—Perhaps not so surprisingly, the most striking difference between the Fs1 and BM-40 structures lies in the relative arrangement of their two subdomains (Fig. 3). If we superimpose the Kazal-like moieties of these two structures, their respective EGF-like subdomains appear to be bent and axially rotated relative to one another so that the angle formed by Cys89 from Fs1 (which superimposes reasonably well with Cys75 from BM-40) and the distal tips of the Fs1 and BM-40 β-hairpins is close to 45°. Despite the limited contacts between the two halves of these molecules, two factors likely to be responsible for the differences observed can be identified.

First, the long loop between residues 89 and 100 in Fs1 follows a completely different trajectory to the equivalent albeit longer loop between residues 78 and 92 in BM-40. Because most of the contacts between the EGF-like subdomain and the Kazal-like subdomain appear to involve this loop, a shift in the position of the polypeptide chain for this region will most likely have repercussions on the rest of the structure. In this respect, the main chain hydrogen bond between the carbonyl group of Gly74 and the amide group of Ala90 in Fs1 (equivalent to Gly63 and Gln79 in BM-40, respectively) appears to play a considerable part in stabilizing the conformation of the hinge region. In fact, this is seen to a lesser extent between the liganded and unliganded forms of Fs1, where ligand binding and altered crystal contacts induce conformational differences in loop 89–100, ultimately leading to a small but noticeable shift in the arrangement of residues 64–75. In the case of the various Fs1 crystals, however, these conformational changes fail to impact on the relative orientation of the whole EGF-like subdomain.

A second factor likely to play a role in the differences recorded between the intersubdomain angles of Fs1 and BM-40 is the presence of a salt bridge between Lys75 and Glu116 of BM-40, which undoubtedly contributes to locking the two subdomains into place. In the Fs1 structure, no such linkage is possible because these two residues are replaced with Lys75 and Gln114, respectively. Even if the charge of residue 116 were retained, its side chain would be too far and too poorly oriented to interact with Lys75.
Interactions between Fs1 and the Small Heparin Analogs SOS and Ins6S—Although building the polypeptide chain of Fs1 did not present any major difficulty, the tracing of the SOS and Ins6S molecules could not be completed. In both cases, large peaks of electron density were visible in the $2F_{\text{obs}}-F_{\text{calc}}$ map contoured at 1.0 vol%, corresponding to ordered sulfate groups interacting with the presumed heparin-binding site of Fs1 (Fig. 4). With the exception of the occasional density for neighboring carbon atoms, no connection could be established between the different sulfates observed, suggesting that the SOS and Ins6S molecules were either partially disordered or bound in a variety of conformations. Models featuring the ligands bound in alternative orientations were refined for both structures, each time allowing various combinations of sulfates to be fitted into all of the visible electron density peaks. However, the lack of connecting density made it impossible to justify one particular model over another, and refinement of both the SOS and Ins6S structures had to be stopped once individual sulfate groups had been fitted into appropriate regions of the electron density and the values of $R_{\text{cryst}}$ and $R_{\text{free}}$ had converged. In the case of Fs1/SOS this led to modeling six sulfate groups, two of which were refined at half-occupancy to account for reduced density and to prevent steric clashes. Five sulfate groups could be modeled for the Fs1/Ins6S structure; three of these were built with an additional carbon atom attached to them, thereby forming a COSO$_4$ pseudo group.

Despite being incomplete, these structures nevertheless allow us to see that both heparin analogs interact primarily with residues belonging to the putative P$^{\text{7-13}}$GKKCRMNKKPKP$^{\text{86}}$ heparin-binding sequence identified by Wang and co-workers (45) (Fig. 5, A–C). In the Fs1/SOS structure, the ligand makes four hydrogen bonds with this presumed heparin-binding site. These occur between the side chain amide of Asn$^{\text{80}}$ and sulfate S1 (3.13 Å), the guanidino group of Arg$^{\text{86}}$ and the half-occupancy sulfate S4b (2.40 Å), the ε-amino group of Arg$^{\text{86}}$ and the S2 sulfate (2.90 Å), and finally, between the backbone amide nitrogen of Lys$^{\text{81}}$ and sulfate S5 (3.03 Å). Interestingly, the side chains of residues Lys$^{\text{81}}$ and Lys$^{\text{84}}$ are not disordered in this structure, but yet they fail to establish any significant contacts with the Ins6S molecule. As a result, the major visible electrostatic component in the protein-ligand interaction once again appears to be Arg$^{\text{86}}$, which is linked to the I1 sulfate via a salt bridge. Despite its involvement in both of the interactions described here, Arg$^{\text{86}}$ adopts radically different side chain conformations after the C$_{\text{y}}$ atom, with the guanidino group pointing into opposite directions and making contacts with differently positioned sulfates.

Additional interactions are also seen between the heparin analogs and a symmetry-related Fs1 molecule (Fig. 5, D–F). In both the Fs1/SOS and Fs1/Ins6S structures, the ligands make contacts with the Kazal-like subdomain of an adjacent protein molecule in the crystal and occupy a shallow pocket centered on...
residue Leu$^{117}$. In the case of Fs1/SOS, the S4b sulfate forms a hydrogen bond with the side chain amide of Asn$^{95}$ (3.08 Å), and the S1 sulfate interacts with the guanidino group of Arg$^{120}$ through a salt bridge. The interaction between the Kazal subdomain of Fs1 and SOS is further reinforced through a hydrogen bond linking the backbone amide of residue Asp$^{92}$ and sulfate S4b (3.20 Å). In the Fs1/Ins6S structure, the only close ionic interactions between the ligand and the corresponding symmetry-related Fs1 molecule are a couple of salt bridges between the I1 and I4 sulfates and Arg$^{120}$. In addition, I1 makes a clear van der Waals contact with Leu$^{117}$.

By acting as bridges between the heparin-binding site of one Fs1 molecule and the Kazal-like domain of its symmetry-related counterpart, both ligand molecules disrupt the crystal contacts that exist in the unliganded Fs1 structure and at the same time create new, possibly stronger ones. Ultimately, this is reflected in the relative orientation of the symmetry-related Fs1 molecules, leading to a higher solvent content and to a 5 Å increase in the length of the $b$ axis for the unliganded Fs1 crystallographic unit cell relative to the other two structures (Table I). Looking at Fig. 5, we see that the EGF-like subdomain interacts with the Kazal-like subdomain of another protein molecule via two salt bridges in the unliganded Fs1 structure (Arg$^{96}$-Glu$^{113}$ and Lys$^{82}$-Glu$^{128}$). The addition of SOS or Ins6S disrupts both of these interactions but creates all of the protein-ligand contacts described above, as well as a few new protein-protein contacts. The latter include a hydrogen bond between Lys$^{75}$ and Asn$^{95}$ in the Fs1/SOS structure, or two salt bridges (Lys$^{84}$-Glu$^{113}$ and Arg$^{96}$-Glu$^{113}$) and a hydrogen bond between the backbone amide of Cys$^{87}$ and the carbonyl oxygen of Asn$^{95}$ (2.89 Å) in the case of Fs1/Ins6S.

Finding the ligands between the two symmetry-related molecules raises a question of possible oligomerization of Fs1 in the presence of a heparan sulfate ligand, as has been seen with other heparan sulfate-binding proteins (47). Because the ligand does not lie on a 2-fold symmetry axis, these interactions would promote aggregation, or higher order oligomerization, rather than dimerization. It is therefore unlikely that the contacts

**Figure 6. Raw and normalized ITC data.** The recorded calorimetric signals of Fs1 titrations with SOS and Ins6S in 0.1 M Tris-HCl (pH 8.5) along with data from titrations of the ligands to buffer (shifted up by 0.1 μcal/s for clarity) are shown on the left, and the fitted normalized data are displayed on the right. The association constant ($K$), stoichiometry ($N$), enthalpy ($ΔH$), and entropy ($ΔS$) of binding are shown next to each curve for all binding events. The protein concentration in the sample cell was 0.025 mM, and ligand concentrations in the syringes were 0.2 mM for SOS and 0.4 mM for Ins6S. Each of the spikes on the left side graphs corresponds to a 10-μl injection of ligand into the calorimeter cell.

![Graph 1](image1.png)

**Graph 1:** SOS Binding to Fs1

- **Time (s)**
- **μcal/s**

![Graph 2](image2.png)

**Graph 2:** Ins6S binding to Fs1

- **Time (s)**
- **μcal/s**

![Graph 3](image3.png)

**Graph 3:** Fitted normalized data

- **Molar ratio**
- **kcal/mol**
- **ΔH = -5953 kcal/mol**
- **ΔS = 16.36 cal/K.mol**
- **$N = 0.8884$**
- **$K = 8.7 \times 10^7$ M$^{-1}$**
- **Site 1:**
  - **ΔH = -4746 kcal/mol**
  - **ΔS = 20.00 cal/K.mol**
  - **$N = 0.7308$**
  - **$K = 7.1 \times 10^7$ M$^{-1}$**

![Graph 4](image4.png)

**Graph 4:** Fitted normalized data

- **Molar ratio**
- **kcal/mol**
- **ΔH = -2049 kcal/mol**
- **ΔS = 16.71 cal/K.mol**
- **$N = 0.8764$**
- **$K = 1.4 \times 10^5$ M$^{-1}$**
between the ligands and symmetry-related protein molecules mimic possible physiological contacts between heparan sulfate-linked Fs1 domains.

A comparison of the Fs1/SOS and Fs1/Ins6S structures reveals that the positions of only two sulfate groups are reasonably well conserved between the two. To begin with, the sulfur atoms of S2 and I2 are only 0.17 Å apart when superimposed, although the relative orientation of the sulfate groups differs slightly. Having a conserved sulfate at this position is understandable in the case of the SOS structure, given that S2 forms a hydrogen bond with the ε-nitrogen of Arg86 and presumably also interacts electrostatically with its guanidino group. On the other hand, the Arg86 side chain adopts an entirely different conformation in the Fs1/Ins6S structure, thereby preventing any hydrogen bonding with I2 and restricting van der Waals packing to very limited contacts between a sulfate oxygen and the Cα and Cβ atoms of Arg86.

Another pair of sulfates occupies a reasonably conserved position in the two liganded structures is S1/I1, with both groups lying only 1.34 Å apart in the two superimposed structures. Although this separation is greater than that observed for the S2/I2 pair, it is still sufficiently small to think of this sulfate position as conserved. In addition, retaining such a functional group at this location can be explained by the fact that both S1 and I1 make similar contacts with the side chain amide of Asn80. Slight differences in the observed positions of these two sulfates are most likely to be caused by the additional contacts made between Arg86 and I2 in the Fs1/Ins6S structure.

**ITC of Protein-Ligand Interactions—**Titrations with both SOS and Ins6S yielded good quality ITC profiles (Fig. 6). Blank titrations of the ligands to the buffer are shown in the same figure, and these heats were subtracted from the ligand to protein titration before fitting the data. As can be seen, residual heats at the ends of titrations equal those of the control experiments.

Based on the crystallographic results discussed above, SOS and Ins6S are expected to associate with the highly basic Fs1 domain through their negatively charged sulfate groups, in what should constitute a strong though perhaps not entirely specific interaction. From a thermodynamic point of view, ITC experiments aiming to determine the binding affinity of these two heparin analogs for the Fs1 domain agree with these predictions.

SOS titration could be fitted to a single binding site model, and the results indicate an association constant greater than $8.7 \times 10^7$ M$^{-1}$ ($K_d = 11$ nm) and stoichiometry of 0.89. Surprisingly the Ins6S titration showed a delayed saturation compared with SOS and could best be fitted to a model of two independent binding sites. The higher affinity binding event was calculated to have an association constant of $7.0 \times 10^7$ M$^{-1}$ ($K_d = 10$ nm) and stoichiometry of 0.88, whereas the second site has a more moderate affinity with $K_d$ of $1.4 \times 10^5$ M$^{-1}$ ($K_d = 7$ μM) and slightly lower occupancy of 0.73. We believe the first binding event corresponds to the site seen in the crystal structures and is where the SOS binds as well. Ins6S is smaller and more charged compared with SOS, and it is possible that a second ligand can bind close to the first at the tip of the domain, but we cannot exclude a possibility that the second Ins6S molecule binds to another heparan sulfate-binding epitope on the domain. As a point of comparison, the binding of Ins6S to fibroblast growth factor was studied by ITC and yielded values in the order of $10^5$−$10^7$ M$^{-1}$ depending on the ionic strength of the sample (48).

The difference in the stoichiometry could be explained by two Ins6S molecules binding close to each other, whereas another SOS (being a larger molecule) could not be accommodated. Highly similar association constants and free energies both for SOS and high affinity Ins6S binding suggest that they bind to the same site, and we think that this is the position seen in the crystal structures. The fact that we only see a single binding event for SOS would argue against a completely separate epitope to which the second Ins6S binds. The positioning and spacing of the sulfates in the Ins6S complex structure would support the model of a single ligand in this site, but a possibility of a weakly bound ligand next to the high affinity site in solution cannot be excluded.

**DISCUSSION**

In this paper we have presented the structure of Fs1, the putative heparan sulfate-binding domain of follistatin, both alone and in complex with the small heparin analogs SOS and Ins6S. Although we were able to collect high resolution data sets from Fs1/SOS and Fs1/Ins6S crystals, the electron density for the ligands was poor, and neither of these molecules could be fully traced, most likely an indication that multiple modes of binding coexist within the two complexes. Nevertheless, locating six sulfate groups from SOS (two of which were refined at half-occupancy) and five sulfate groups from Ins6S allowed us to map a potential heparan sulfate-binding site on the surface of this follistatin domain and highlight the importance of residues Asn80 and Arg86 in the Fs1 interaction with sulfated compounds. In addition, the location of two sulfate groups was shown to be conserved between the two structures, indicating that sulfates might also be found at these positions in follistatin-heparin or follistatin-heparan sulfate complexes. Because clear electron density was only visible for a limited number of ligand sulfates in the two structures, one might assume that those groups that could be built occupy key locations on the surface of the heparan sulfate-binding site. The two analogs used in this study may therefore prove to be valuable analytical tools for the identification of residues that are likely to be involved in the binding of the follistatin ligand, heparan sulfate.

Of course, the structure of a more physiologically relevant Fs1-heparin or Fs1-heparan sulfate complex might be more straightforward to interpret, and obtaining diffraction-quality crystals of such complexes therefore remains one of the prime objectives in the structural study of follistatin. Before this can be done, however, the optimal length of heparin required to interact with Fs1 must be determined. It seems unlikely that the heparin-binding site of Fs1 would stretch from the primary epitope identified on the surface of the EGF-like subdomain to the secondary epitope revealed by crystal symmetry on the Kazal subdomain. Indeed, the heparin molecule would have to wrap itself around the Fs1 domain to interact with both epitopes, a highly improbable scenario if we consider the inherent rigidity of this sulfated glycan.

If we assume that sulfate groups from the natural ligand of follistatin occupy the two conserved sulfate positions seen in the SOS and Ins6S complexes, two potential ligand-binding scenarios arise. In the first case, the heparan sulfate chain binds to the Fs1 domain diagonally, over a region spanning the primary epitope as well as positively charged Kazal-like domain residues distinct from those forming the secondary epitope. Alternatively, both conserved sulfate positions could be occupied by a heparan sulfate chain placed across the tip of the EGF-like module, but additional contacts with the Fs1 domain would be unlikely. In the case of a diagonal mode of binding, at least four, if not five, disaccharide units would be required to allow interactions with both parts of the molecule. On the other hand, perpendicular binding across the EGF-like subdomain might only require one or two disaccharide units.
In this study, crystallographic work on Fs1 was complemented through the use of ITC, which provided an efficient means of quantifying the affinities of the interactions between Fs1 and SOS/Ins6S. Although the nanomolar $K_d$ values obtained by this method initially seem to contradict the crystallographic data, the high sulfation level of the ligands used suggests that specificity at the functional group (sulfate) level can be achieved in multiple different orientations of the ligand. If one orientation only could bind with high affinity, we would expect to see a clear electron density for it in our structures. All contacts between the ligand and the protein in these structures occur through the sulfate groups, and we postulate that specificity of binding is retained as long as any given set of correctly positioned sulfates is able to interact with the Fs1 presumed heparin-binding site. Because of the larger number of sulfate groups present on the SOS and Ins6S ligands, several combinations of sulfates are likely to satisfy these spatial restraints, thereby accounting for the multiple modes of binding deduced from the corresponding crystal structures.

A second molecule of Ins6S appears to bind to the domain in solution with 3 orders of magnitude lower affinity. However, we cannot tell whether this binding occurs close to the ligand site identified in the crystal structures and which we believe to be the high affinity site, or in a separate, as yet unidentified site.

Work on follistatin must now be pursued on two levels. First, an in-depth functional analysis of the interaction between follistatin and heparan sulfate must be carried out to enhance our knowledge of its role in the follistatin-mediated inhibition of transforming growth factor-β-like signaling pathways. Second, structural work must now shift its focus onto a more physiologically relevant binding partner for follistatin, preferably using the follistatin molecule as a whole. Heparin disaccharides are readily available commercially, but obtaining similar amounts of homogeneous heparin octa-, deca-, or dodecasaccharide preparations represents a considerable challenge, mainly because of the relative lag in the development of heparin purification techniques compared with protein or nucleic acid separation methods. Testing the binding of such molecules to follistatin by ITC may therefore present a convenient way of identifying the correct heparin length, degree of sulfation, and stoichiometry required for crystallization.

Acknowledgments—We thank Prof. John Gurdon (Wellcome Trust/Cancer Research UK Institute, Cambridge, UK) for providing the rat follistatin clone, Prof. Alan Cooper and Margaret Nutley at the BBSRC/ EPSRC Microlamelometry Service (Department of Chemistry, University of Glasgow, UK) for help with ITC, and the ESRF staff for beamline support. We also thank Prof. Tom Blundell for useful discussions on follistatin and for providing some of the necessary resources for performing this work.

REFERENCES

1. Phillips, D. J., and Kretser, D. M. (1998) Front. Neuroendocrinol. 19, 287–322
2. de Winter, J. P., ten Dijke, P., de Vries, C. J. M., van Achterberg, T. A. E., van Eijndhoven-van Raaij, A. J. M. (1996) Mol. Endocrinol. 10, 953–963
3. Ciccarelli, F., Copley, R. R., Ponting, C. P., and Bork, P. (2002) Nucleic Acids Res. 30, 2443–2448
4. Denzer, A. H., Schiffman, D., Schonbrun, N. A., Kaiman, A., Engel, J., and Ruegg, M. A. (1998) EMBO J. 17, 335–343
5. Wang, Q., Keutmann, H. T., Schneyer, A. L., and Sluss, P. M. (2000) Endocrinology 141, 3183–3193
6. Letunic, I., Goodstadt, L., Dickens, N. J., Doerkis, T., Schultz, J., Mott, R., Ceroni, N., Copley, R. R., Whittington, R. P., and Bork, P. (2002) Nucleic Acids Res. 30, 244–244
7. Gao, D., Blundell, T. L., Schonbrun, N. A., Kaiman, A., Engel, J., and Ruegg, M. A. (1998) EMBO J. 17, 335–343
8. Wang, Q., Keutmann, H. T., Schneyer, A. L., and Sluss, P. M. (2000) Endocrinology 141, 3183–3193
9. deletion in proof
10. deletion in proof
11. deletion in proof
12. deletion in proof
13. deletion in proof
14. deletion in proof
15. deletion in proof
16. deletion in proof
17. deletion in proof
18. deletion in proof
19. deletion in proof
20. deletion in proof
21. deletion in proof
22. deletion in proof
23. deletion in proof
24. deletion in proof
25. deletion in proof
26. deletion in proof
27. deletion in proof
28. deletion in proof
29. deletion in proof
30. deletion in proof
31. deletion in proof
32. deletion in proof
33. deletion in proof
34. deletion in proof
35. deletion in proof
36. deletion in proof
37. deletion in proof
38. deletion in proof
39. deletion in proof
40. deletion in proof
41. deletion in proof
42. deletion in proof
43. deletion in proof
44. deletion in proof
45. deletion in proof
46. deletion in proof
47. deletion in proof
48. deletion in proof
49. deletion in proof
50. deletion in proof

4. Hynes, R. O., and Blumberg, B. M. (1985) Science 228, 123–128
5. Sidman, R. L., and Scolnick, E. M. (1984) Science 225, 123–128
6. de Winter, J. P., ten Dijke, P., de Vries, C. J. M., van Achterberg, T. A. E., van Eijndhoven-van Raaij, A. J. M. (1996) Mol. Endocrinol. 10, 953–963
7. Ciccarelli, F., Copley, R. R., Ponting, C. P., and Bork, P. (2002) Nucleic Acids Res. 30, 2443–2448
8. Denzer, A. H., Schiffman, D., Schonbrun, N. A., Kaiman, A., Engel, J., and Ruegg, M. A. (1998) EMBO J. 17, 335–343
9. Wang, Q., Keutmann, H. T., Schneyer, A. L., and Sluss, P. M. (2000) Endocrinology 141, 3183–3193
10. deletion in proof
11. deletion in proof
12. deletion in proof
13. deletion in proof
14. deletion in proof
15. deletion in proof
16. deletion in proof
17. deletion in proof
18. deletion in proof
19. deletion in proof
20. deletion in proof
21. deletion in proof
22. deletion in proof
23. deletion in proof
24. deletion in proof
25. deletion in proof
26. deletion in proof
27. deletion in proof
28. deletion in proof
29. deletion in proof
30. deletion in proof
31. deletion in proof
32. deletion in proof
33. deletion in proof
34. deletion in proof
35. deletion in proof
36. deletion in proof
37. deletion in proof
38. deletion in proof
39. deletion in proof
40. deletion in proof
41. deletion in proof
42. deletion in proof
43. deletion in proof
44. deletion in proof
45. deletion in proof
46. deletion in proof
47. deletion in proof
48. deletion in proof
49. deletion in proof
50. deletion in proof