Published in final edited form as:

Biochemistry. 2018 April 17; 57(15): 2189–2199. doi:10.1021/acs.biochem.8b00241.

Structural Characterization of a Heparan Sulfate Pentamer Interacting with LAR-Ig1-2

Qi Gao¹, Jeong-Yeh Yang¹, Kelley W. Moremen¹, John G. Flanagan², and James H. Prestegard^{1,*}

¹Complex Carbohydrate Research Center, University of Georgia, Athens GA 30602

²Department of Cell Biology and Program in Neuroscience, Harvard Medical School, Boston, MA 02115

Abstract

Leukocyte common antigen-related (LAR) protein is one of the type IIa receptor protein tyrosine phosphatases (RPTPs), which are important for signal transduction in biological processes including axon growth and regeneration. Glycosaminoglycan chains, including heparan sulfate (HS) and chondroitin sulfate (CS), act as ligands that regulate LAR signaling. Here, we report the structural characterization of the first two immunoglobulin domains (Ig1-2) of LAR interacting with an HS pentasaccharide (GlcNS6S-GlcA-GlcNS3,6S-IdoA2S-GlcNS6S-OME, fondaparinux) using multiple solution-based NMR methods. In the course of the study, we extended an assignment strategy useful for sparsely labeled proteins expressed in mammalian cell culture supplemented with a single type of isotopically enriched amino acid (¹⁵N-lysine in this case) by including paramagnetic perturbations to NMR resonances. The folded two domain structure for LAR-Ig1-2 seen in previous crystal structures has been validated in solution using residual dipolar coupling data, and a combination of chemical shift perturbation on titration of LAR-Ig1-2 with fondaparinux, saturation transfer difference (STD) spectra and transferred nuclear Overhauser effect (trNOEs) have been employed in the docking program, HADDOCK, to generate models for the LAR-fondaparinux complex. These models are further analyzed by post-processing energetic analysis to identify key binding interactions. In addition to providing insight into the ligand interaction mechanisms of type IIa RPTPs and the origin of opposing effects of CS and HS ligands, these results may assist in future design of therapeutic compounds for nervous system repair.

Introduction

LAR, or leukocyte common antigen-related protein, is a type IIa receptor protein tyrosine phosphatase (RPTP), a class of proteins shown to bind a variety of cell surface proteins or soluble ligands and to be involved in cell-cell or cell-matrix contacts¹. LAR and the other two members of the type IIa family in mammals, RPTP σ and RPTP δ , regulate biological

*correspondence: jpresteg@ccrc.uga.edu.

Supporting Information

Supporting Information is available free of charge on the ACS publications website at DOI: Included are four supplementary figures and seven supplementary tables.

events that include axon extension during neural development, synaptic organization, and axon regeneration following nerve injury²⁻¹¹. Among the ligands that can interact with type IIa RPTPs and regulate downstream signaling in neurons are heparan sulfate proteoglycans (HSPGs) and chondroitin sulfate proteoglycans (CSPGs)^{7, 12}. Interestingly, these ligands have opposite effects on axon growth and synaptic function: HSPGs are believed to promote clustering of LAR resulting in axon extension, whereas CSPGs are believed to interfere with protein clustering, leading to inhibition of axon growth and regeneration^{4, 7, 12}. An atomic level structural characterization of the interaction of well-defined segments from the glycosaminoglycans (GAG) chains of these proteoglycans with LAR would help to understand the mechanisms of ligand-regulated signaling, as well as providing a basis for future structure-based design of therapeutic compounds to promote axon regeneration following nerve injury. Here we provide a step toward that understanding by presenting a model for the interaction of the N-terminal two domains of LAR (LAR-Ig1-2) with a well-defined heparan sulfate (HS) pentamer, the anticoagulant fondaparinux (GlcNS6S-GlcA-GlcNS6S3S-IdoA2S-GlcNS6S-OME). Our approach builds on an existing crystal structure for the LAR two-domain fragment¹² using a solution NMR approach designed to work with challenging proteins and protein complexes that can be difficult to express in the bacterial hosts required for uniform metabolic labeling with NMR active isotopes¹³. The approach includes a resonance assignment strategy applicable to proteins expressed in mammalian (HEK293F) cells and sparsely labeled with ¹⁵N in all occurrences of a single amino acid type, use of ligand induced changes in chemical shifts of these resonances to locate a binding site on the protein, residual dipolar couplings (RDCs) to confirm retention of protein structure in solution, and transferred nuclear Overhauser effects (trNOEs) and saturation transfer difference (STD) spectra to determine binding epitopes and conformation of the bound ligand, respectively. The model, produced with a ligand docking program and analyzed with molecular dynamics simulations, shows an extended binding site consistent with electrostatic interactions previously identified using HS mimics and mutagenesis.

Structurally, type IIa RPTPs, including LAR, share a very similar domain architecture, containing three immunoglobulin-like (Ig) domains, followed by nine fibronectin type II (FN) units, a single transmembrane helix and two intracellular phosphotyrosine-specific phosphatase domains². Earlier crystallography and site-directed mutagenesis studies have suggested that the first Ig domain (Ig-1) is crucial for glycosaminoglycan binding and the first two Ig domains (Ig 1-2) are a minimum structural requirement for interaction with the postsynaptic ligand TrkC³. Each of the first two domains has a single disulfide bond, which likely contributes to stability, and the first domain has a canonical N-glycosylation site. Both factors suggest that functional expression would require a host that employs appropriate folding chaperones and glycosylation machinery. The availability of an Ig1-2 crystal structure¹² not only facilitates the NMR resonance assignment process, but also allows a direct comparison with structurally sensitive data from our study. Compared with the remaining Ig domains, Ig1 and Ig2 are believed to display the least inter-domain orientation flexibility³. As depicted in Figure 1A, the domains display a unique folded structure in the crystal. It is of interest to see if this structure is preserved in solution.

Significant efforts have also been directed at characterizing the interaction between RPTP proteins and glycosaminoglycans^{14, 15}. For example, among different chondroitin sulfate

isoforms (e.g. CS-A, CS-C and CS-E), CS-E, which is more highly sulfated, shows a preference for interaction with RPTP σ and inhibits the signaling pathways more strongly¹⁶. For most members of the RPTP family, heparan sulfate interactions increase with the length of HS oligomers up to about ten residues¹², but these oligomers come from digests of long chain heparin or HS, and do not yield data on preference for any specific sulfation pattern. There is also significant mutagenesis data identifying the general region in which GAGs bind^{1, 2, 12}. Interestingly, both oligomers from CS and heparin digests appear to share this binding region^{2, 12}. In terms of structural data on well-defined ligands, only the crystal structure of LAR a complex with sucrose octasulfate exists. Sucrose octasulfate differs significantly from a true HS oligomer; it has just two sugar residues, neither one occurring in HS, and the placement of sulfate groups is not a close mimic of naturally occurring placements in HS. Therefore, generating a structure showing a specific interaction between LAR and a well-defined HS or CS oligomer would provide structural insight and an improved understanding of signaling mechanisms. Here, we focus on fondaparinux, a well-defined therapeutic agent. It is more representative of HS, sharing a repeating disaccharide of GlcNS and IdoA or GlcA, as well as a number of sulfation sites that can be found in HS. The structure of fondaparinux, along with the numbering system used in this communication is shown in Figure 1B. In addition to the structural and biological insight gained from studying the interaction of fondaparinux with LAR-Ig1-2, the methods developed in the course of study should be extendable to other well-defined HS and CS oligomers.

Proteins that are well expressed in mammalian cells (the LAR used for crystal studies was expressed in HEK293 cells) present special problems for NMR investigation in that NMR active isotopes must be supplied in the form of isotopically labeled amino acids. If triple resonance methods are used to provide resonance assignments, the expense and difficulty of protein production is greatly increased by the requirement for uniform labeling of all amino acids with both ¹³C and ¹⁵N. Also, for proteins the size of LAR-Ig1-2, perdeuteration is often desirable as this reduces line widths and improves resolution, but mammalian cells do not tolerate high levels of deuterium in metabolic substrates. Previously we have used a sparse labeling strategy which uses a single, or small number of, isotopically labeled amino acids^{17, 18}. This reduces costs and improves resolution by reducing the numbers of resonances that need to be resolved. The strategy does require an alternative resonance assignment approach. To address this we developed a program (ASSIGN_SLP) that uses a genetic algorithm to search for an optimal assignment of heteronuclear single quantum coherence (HSQC) crosspeaks to specific labeled sites by comparing predicted and experimental values for chemical shifts, residual dipolar couplings (RDCs) and nuclear Overhauser effects (NOEs)^{18–23}. We use this procedure here, and also supplement data types with crosspeak intensity loss due to paramagnetic relaxation enhancement (PRE) from a lanthanide binding peptide introduced into the protein construct^{24, 25}.

Our structural information comes from a variety of sources. RDCs provide information on protein structure in terms of the average orientation (in the frame of the medium used for partial alignment) of bonds between pairs of nuclei, such as the ¹H-¹⁵N pair in the amide bond of a labeled site. When labeled sites are sparse this cannot produce a full protein structure, but it can show consistency with proposed models such as the highly bent model of LAR seen in the crystal structure. Perturbation of crosspeak positions in HSQC spectra on

addition of ligand also helps to identify the ligand binding site. Information on the bound structure of a ligand comes from NMR experiments based on observation of proton resonances from the ligand, such as saturation transfer difference (STD) spectroscopy and transferred nuclear Overhauser effects (trNOEs). We combine these data using the docking program, HADDOCK²⁶, to generate a model for the LAR-Ig1-2 – fondaparinux complex. The output of HADDOCK is a cluster of possible structures. We provide a more detailed analysis of members of this cluster using a generalized Born surface accessibility (GBSA) estimation of free energies of binding from short molecular dynamics trajectories on the complexes. A per-residue decomposition of these energies then points to interactions that can guide a search for potential LAR ligands having higher affinity.

Materials and Methods

Materials

¹⁵N-Lys and deuterium oxide were purchased from Cambridge Isotope Laboratories (Tewksbury, MA). All other chemicals, including the HS oligomer, fondaparinux, sodium salt, were purchased from MilliporeSigma (St. Louis, MO) unless otherwise stated.

Protein expression and purification

The genes encoding the signal sequence and Ig1-2 domains of LAR (receptor-type tyrosine-protein phosphatase F, PTPRF, UniProt P10586, residues 1-231, designated here as LAR-Ig1-2) and a construct containing a lanthanide-binding peptide loop inserted following residue 70 (designated LAR-loop-Ig1-2) were synthesized with human codon optimization (GeneArt gene synthesis, Thermo Fisher Scientific, Waltham MA) including an upstream Kozak sequence and flanking restrictions sites. The DNA fragments for both sequences were cloned into a mammalian expression vector analogous to the Gateway pGec2-DEST²⁷, which includes downstream fusion sequences comprising a TEV protease cleavage site, superfolder GFP domain, an AviTag, and an 8xHis-tag. In the present expression constructs the synthetic LAR sequences were inserted through use of restriction sites between the CMV promoter and the TEV recognition sequence in the vector rather than by Gateway recombination in the pGec2-DEST vector²⁷ to yield the fusion protein coding sequence shown in Figure S1 of Supporting Information.

For metabolic labeling of the HEK293S (GnTI-) cell cultures with [¹⁵N] lysine, cells were transfected as previously described^{27, 28} and 16h after transfection the medium was exchanged for FreeStyle™ 293 expression medium (Thermo Fisher Scientific) supplemented with 150mg/liter of [¹⁵N] lysine 98% (Cambridge Isotope Laboratories, Andover, MA) and 2.2mM valproic acid (MilliporeSigma). The recombinant protein was harvested after 6 days of growth and was purified from the culture supernatant using Ni²⁺-NTA chromatography and concentrated to ~1 mg/mL. The resulting protein preparation was digested with recombinant TEV protease to cleave between LAR and GFP and recombinant endoglycosidase F1 (EndoF1) to cleave any glycans to a single GlcNAc residue. The preparation was then subjected to Ni²⁺-NTA chromatography a second time to remove the GFP fusion tag, TEV protease, and EndoF1, each of which contain a His tag. Mass spectrometry analysis of trypsin digested peptides unexpectedly showed only an unmodified

peptide containing the consensus site for *N*-glycosylation, and no GlcNAc containing versions of this peptide, indicating that this site was not glycosylated during recombinant expression. The final protein yield was 12 mg/L. Average ^{15}N labeling efficiency for lysines was 77%, as determined by analyzing the isotopic envelope of mass spectrometric peaks of tryptic peptides containing lysine residues.

NMR spectroscopy

All NMR spectroscopy was carried out at 18.8 T or 14.0 T on Agilent instruments with DD2 consoles and 5 mm cryogenically cooled triple resonance probes. NMR protein samples were 220 μM in 10% D_2O buffer containing 10 mM MES and 100 mM NaCl at pH 6.0 for ^{15}N HSQC titrations, 3D ^{15}N -filtered NOE, and RDC experiments. The protein-ligand complex samples for STD and trNOE experiments were in 100% D_2O buffer containing 20 mM sodium phosphate and 100 mM NaCl at pH 6.5, at a protein ligand ratio of 1:30 with a ligand concentration of 1.2 mM. LAR-loop-Ig1-2 samples contained lanthanides at lanthanide:protein ratios slightly less than 1:1.

NMR experiments were standard Biopack experiments conducted at 25 °C. For protein resonance assignments and structure validation, HSQC-NOESY experiments (pulse sequence: gnoesyNfhsqcA) were carried out using a mixing time of 40 ms. Two sets of RDCs were measured on protein samples containing either 12.5 mg/mL Pf1 phage (ASLA biotech) or 4.2 % PEG (C12E5/hexanol) bicelle using a pulse sequence in which cross-peaks in HSQC spectra are modulated by J+D coupling in the ^{15}N dimension ²⁹. To assess possible dimerization of the protein, residue specific rotational correlation times for the protein were obtained from an SCT-CCR experiment using the same sample as the titration experiment ³⁰. Data were collected at 7 delay times from 15 to 90 ms and fit to mono-exponential decays for each of the 13 crosspeaks. Averages of these values were used to assess the extent of dimerization. To identify protein crosspeaks perturbed by fondaparinux and determine a binding constant ^1H - ^{15}N HSQC spectra (pulse sequence: gNfhsqc) were acquired on a 170 μM LAR-Ig1-2 sample at concentrations of fondaparinux from 50 μM to 550 μM with steps of 50 μM from 0 μM to 150 μM and steps of 100 μM from 150 μM to 550 μM . The dissociation constant was determined by fitting the chemical shift perturbation as a function of concentration for each labeled residue as previously described ¹⁸.

The ligand proton resonances were assigned by acquiring and analyzing ^1H proton, ^1H - ^1H TOCSY, ^1H - ^1H NOESY, ^{13}C - ^1H HSQC and ^{13}C - ^1H HMBC spectra; assignments prove consistent with previous publications ³¹. For characterization of bound ligand geometry trNOE experiments were acquired using the pulse sequence, gnoesyNfhsqcA, with a mixing time of 90 ms. STD experiments were acquired with the pulse sequence, dpfgse_satxfer, irradiating the samples at both -2.5 and 9.5 ppm and increasing saturation times from 1 to 4 s in steps of 1 s, with interleaved irradiation at 30 ppm to create difference spectra. Double difference spectra were created by conducting the same STD experiments on samples without ligand and subtracting the spectra from spectra produced with ligand. All the NMR data were processed with NMRPipe ³² and analyzed with SPARKY ³³.

Assignment of sparsely labeled LAR

The assignment of each labeled Lys residue in LAR-Ig1-2 was achieved using the program ASSIGN_SLP¹⁹. The input experimental data included amide ¹H and ¹⁵N chemical shifts, two sets of RDCs and ¹⁵N-filtered NOEs. NOE peak lists generated in SPARKY were converted to vectors representing columns of the 3D set using a script contained in the program package. These were compared to predicted vectors based on PPM1 predictions for chemical shift³⁴ and intensities based on 1/r⁶ distance measurements on the crystal structure (2YD5)¹². For RDCs experimental data were compared to predictions from singular value decomposition of equations relating crystal structure coordinates to RDCs, and for chemical shifts experimental data were compared to PPM1 predictions. In order to improve crosspeak assignments, LAR with the lanthanide binding loop loaded with Gd³⁺ was used to separate labeled sites into a nearby group (n sites) and a distant group (N-n sites) based on paramagnetic broadening and loss of crosspeak intensity. There is less potential degeneracy of data when groups are smaller ((N-n)! x n! vs N!), making the assignment more robust. Different snapshots (600 ns, 800 ns and 1000 ns) from an MD simulation of LAR-Ig1-2 containing the lanthanide binding loop were utilized to identify 5 sites with the closest approach to the Gd³⁺ ion, leaving groups of 5 and 8. Assignment to these groups was enforced by adding penalties to the genetic algorithm objective function in proportion to elements of a user-specified constraint matrix in which elements denoting assignment to a correct group have value zero and elements denoting assignment to an incorrect group have value one (proportionality constant 10 in this case). In principle, the efficiency of the calculation could be improved by running the ASSIGN_SLP separately on the two groups. However, the use of RDC data requires the determination of 5 order parameters from at least 6 pieces of data. The size of one of our groups is too small to take this route. Other details of the genetic algorithm implementation have been described previously³⁵. Briefly, a combination of mutation and crossover rates (0.2, 0.4, 0.6 and 0.8 for both mutation and crossover) was used to allow a thorough search of assignment space, and all the assignments with a score under 8 were collected as possible solutions and saved for analysis.

Computational Docking

The docking program HADDOCK²⁶ was used to generate the molecular models of the LAR-Ig1-2-fondaparinux complex. The crystal structure of human LAR-Ig1-2 (PDB 2YD5) was used as the input protein structure. Two structures of fondaparinux with the 2S IdoA group in either ¹C₄ or ²S₀ ring forms were generated using the GLYCAM web server³⁶. Ambiguous interaction restraints were set for the protein residues identified as involved in binding by chemical shift perturbation in the NMR titration experiments and similar restraints were set for residues in the ligand based on STD data. Inter-proton distance restraints within the ligand were set based on distances calculated from trNOE data using a 1/r⁶ dependence of NOE intensity and a reference distance from the intra-residue ¹H-²H pair of the central GlcNAc residue. The upper and lower limits were set by adding or subtracting 0.3 Å to the calculated distance. During the flexible docking part of the routine, the ligand was set to be fully flexible and strands of the protein containing identified binding residues were specified as semi-flexible. The detailed docking process has been described previously¹⁸. In the end, 20 top scoring models with rms NOE distance deviations less than 0.5 Å and the lowest energies were obtained.

MD simulation, binding free energy calculation and per-residue decomposition

In order to provide pdb structures of loop containing LAR-Ig1-2, a molecular dynamics (MD) trajectory was carried out using the SANDER module of AMBER 14³⁷ and the ff14SB force field³⁸. The atomic coordinates were initially obtained from pdb 2YD5 with the lanthanide binding loop modeled within CHIMERA³⁹. The protein was solvated by a cubic box of TIP3P water. The MD simulation lasted for 1 μ s after 2000 steps of minimization followed by 400 ps of heating. Shorter MD trajectories were initiated with the top 4 ligand-bound HADDOCK structures using the AMBER 14 package. In addition to the ff14SB³⁸ peptide force field, parameters from the GLYCAM_06j-1⁴⁰ force field were used to support carbohydrate simulation. These MD simulations lasted for 50 ns after minimization, heating and density equilibration. The molecular mechanics generalized Born surface area (MM-GBSA)^{41, 42} method was used to estimate the free energy of binding and this was followed by per-residue decomposition. Detailed parameterization has been described elsewhere¹⁸.

Results

Assignment of LAR-Ig1-2 HSQC crosspeaks

Figure 2 shows superimposed ¹⁵N-¹H HSQC spectra of the ¹⁵N-Lys labeled LAR-Ig1-2, with a lanthanide binding loop carrying a diamagnetic ion (blue), and without that loop (red). Several options for loop insertion had been explored, however, insertion in the loop normally connecting β -strands E and F (between residues G70 and K71) was the only position giving acceptable levels of expression, retention of LAR structure, and useful lanthanide binding affinity. Thirteen crosspeaks are observed in the LAR-loop spectrum; there are 13 lysine sites in both constructs, but we see only 12 crosspeaks in the non-loop spectrum. Peaks that are absent or of reduced intensity often reflect exchange broadening due to internal motion at intermediate timescales near these sites. The weak peak labeled 3N-H for the loop-containing construct is likely the one missing in the non-loop structure.

Except for crosspeaks 5, 10, and 11, which move upon loop insertion, preservation of crosspeak positions on comparison of the non-loop and loop-containing samples verifies the retention of the basic structural features of LAR in the loop-containing construct. There are three lysines within two residues of the loop insertion, and it is not surprising that crosspeaks from these sites would shift. The affinity for fondaparinux also proves to be greatly reduced in the loop construct. This is also not surprising as the three lysines close to the loop are among the residues previously suggested to be involved in HS binding⁴³. Hence, we have used the loop construct only for crosspeak assignment purposes and for verification of the in-solution inter-domain structure of the LAR two domain construct and will return to the non-loop construct for structure determination of a fondaparinux-LAR complex.

The LAR-loop construct was selected to conduct initial resonance assignments partly considering the sharper lines and the more dispersed chemical shifts. This dispersion is particularly critical for the proton dimension as collection of ¹⁵N-filtered NOESY spectra were collected in a pseudo 2D fashion relying on dispersion in this dimension. To carry out

assignments we used the program ASSIGN_SLP which requires multiple NMR observables¹⁹. Detailed procedures can be found in the experimental section. Predicted and experimental values for chemical shifts and NOEs are given in Table S1 and Table S2 of Supporting Information. Experimental RDCs are given in Table S3.

Our usual assignment strategy¹⁹, which is based on minimizing scores dependent on deviations between predicted and experimental RDCs, chemical shifts and NOE data, was supplemented with penalties that constrain assignment of paramagnetically perturbed crosspeaks to the sites closest to a Gd³⁺ ion in the lanthanide binding loop of the LAR-loop construct. The superimposed spectra of lysine labeled LAR-loop with and without Gd³⁺ are shown in Figure 3. From the overlaid spectra we can tell that the intensities of crosspeaks 2, 3, 5, 10, and 11 decrease the most. Distances between the Gd³⁺ ion and labeled sites in an MD simulation of the loop-containing construct suggest the 5 residues giving rise to these crosspeaks are K68, K69, K71, K72 and K121. Combined scores were generated for trial assignments of the 13 crosspeaks to the 13 sites and a search for scores below that based on estimated errors in predicted and experimental data was made with the aid of a genetic algorithm.

The results of the assignment procedure using the 1000 ns frame of the MD simulation as an input protein structure are presented in terms of a histogram showing the frequency of assignment of each crosspeak (measurement) to a particular residue in Figure 4. In the list of acceptable assignments some crosspeaks will be consistently assigned to the same site. Previous work using a test set of small proteins has indicated that assignment of a crosspeak to the same site more than 50% of the time corresponds roughly to a 95% confidence limit¹⁹. In the histogram shown, six of the assignments meet this criterion. By doing the assignment on two additional frames (600ns and 800ns) two additional high confidence assignments are found and three more assignments can be made on the basis of being the most favorable assignment using two or more of the frames. The remaining two assignments can then be made by elimination. These assignments are summarized in Table 1.

Inter-domain geometry of LAR-Ig1-2

Once crosspeaks are assigned, we can assess the validity of the folded geometry directly by looking at Q factors for RDC data⁴⁴. These factors are ratios of the root-mean square deviation (RMSD) between measured and calculated RDCs and the root-mean square of the measurements. Figure 5 shows correlation plots comparing measured and predicted RDCs for phage and bicelle data. Regions near the loop insertion are expected to have significant internal motion, hence the residues within two amino acids of the loop insertion are excluded (K69, K71 and K72). K68 is also close (three amino acids away), however, its N-H vector does not move significantly during the MD simulation. Therefore, K68 is included. Ten pieces of RDC data remain and these were used in the final calculation. Q factors of 0.19 (phage) and 0.43 (PEG) were obtained. A Q factor of 0.4 (based on a larger set of RDC values) has been suggested to reflect a structure comparable to an X-ray structure of approximately 2.5 Å resolution⁴⁵. Hence, we believe these numbers support the contention that the folded LAR structure is preserved in solution.

Comparison of Q factors back-calculated using the folded LAR structure to those back-calculated using an alternative structure provide another means of assessing the significance of these measurements. The Ig1-2 domains of LAR are structurally similar to the Ig1-2 domains of Robo1, but Robo1 has a more extended structure⁴⁶. So, adjusting LAR-Ig1-2 inter-domain geometry to superimpose with Robo1-Ig1-2 would yield a suitable trial structure. Using this trial structure, Q factors obtained for phage and bicelle data are now 0.43 and 0.88. The increase confirms that, despite the small number of RDCs, the data are sensitive to inter-domain geometry, and the folded LAR-Ig1-2 structure is the best representation.

Protein binding site identified by NMR titration

Direct information on the residues making up a ligand binding site, as well as disassociation constants (K_d), can be extracted from chemical shift perturbations on addition of ligand to LAR-Ig1-2. The overlaid 2D ¹⁵N-¹H HSQC spectra of LAR-Ig1-2 with increasing concentrations of fondaparinux are shown in Figure 6A. The maximum chemical shift perturbation of each lysine site is plotted in Figure 6B. Four residues, belonging to peaks 3, 5, 10 and 11, display substantial chemical shift perturbation (> 0.05 ppm) upon the titration with fondaparinux. Fitting the normalized average of the shifts for these peaks, a dissociation constant (K_d) of 60 ± 24 μM is determined. The data and best fit line are shown in Figure S3.

The limiting shifts for perturbed crosspeaks can also be used as a qualitative indicator of residues involved in fondaparinux binding. Crosspeaks 3, 5, 10 and 11 have been assigned to Lys 69, 68, 72 and 71. Previous literature has also reported these four residues to be crucial for binding to heparin or chondroitin sulfate by site direct mutagenesis of the homologous protein RPTPσ^{7, 43}. The crystal structures of LAR with sucrose octasulfate has also pointed to these four residues¹². Proximity to these residues will be used as a constraint in generating a model for a fondaparinux-LAR complex.

Ligand binding epitopes from STD experiments

Binding epitopes on a ligand can be qualitatively identified by saturation transfer difference (STD) NMR^{18, 47, 48}. The saturation of a resonance of a protein proton can be transferred to that of a bound ligand proton in a 1/r⁶ fashion, where r is the distance between the two protons. If the ligand returns back to the free solution state in a short time compared to its T₁ relaxation time, selective decreases in intensity of resonances in the solution spectrum of the ligand occur, even when free ligand is in large excess. Therefore, the binding epitope of ligand can be qualitatively identified. Ligand resonances dominate normal STD spectra, first because one normally works at a high ligand to protein ratio (50:1 or 100:1) and then because filtering of broad protein lines eliminates most residual protein signals. For LAR-Ig1-2 some complexities arise because of its relatively small size (22 kDa). STD effects are roughly proportional to protein molecular weight, making STD signals weaker and interference from protein background more problematic for small proteins. Also, had the protein carried its expected glycans, these more mobile moieties would have added interfering resonances in the same region as fondaparinux resonances. Therefore, we chose to work at a 30:1 ratio of fondaparinux to LAR-Ig1-2 to make STD signals stronger. To

further suppress interfering resonances we also utilized a double difference STD experiment (STDD) in which an additional control STD experiment is run in the absence of ligand and subtracted from the experiment with ligand present. Data were collected applying saturating rf at both -2.5 ppm and 9.5 ppm, as spin diffusion among protein protons is also less efficient in smaller proteins and one can expect some STD differences depending on whether phenyl protons (near 9.5 ppm) or methyl protons (near -2.5) are irradiated.

Both data sets showed the strongest STDD signal for resonances assigned to protons BH3 or BH4 and additional strong signals for AH2 and AH5 in the 9.5 ppm and -2.5 sets respectively (the first letter being a ring designation as in Figure 1B and protons being numbered from the anomeric proton as 1). These signals suggest that the non-reducing end of fondaparinux is in most intimate contact with the protein surface. There is, however, a strong signal from the reducing end O-methyl group in the 9.5 ppm set. Because the longer methyl relaxation times and sharp resonance makes this group more susceptible to experimental artifacts, restraints derived from this signal will be omitted in docking procedures. In addition to the strongly perturbed resonances there are significant STDD effects observed for many other resonances. Some of these additional effects arise from spin diffusion among sets of proximate protons in the ligand, especially when binding is tight and release from the protein is slow. All measureable STDD values are presented in Figure S4 of the supplement. However, only the strongest STD signals (more than 50%, omitting the O-Me signal) have been implemented as ambiguous restraints for use in the HADDOCK program to be described below.

Ligand bound conformation from trNOE experiments

Transferred nuclear Overhauser effects (trNOEs) provide structural information on the bound conformations of ligands even when free ligand is in excess. This is the result of their dependence on the effective rotational correlation time which approaches that of the protein in a ligand-protein complex. In our case this is approximately a factor of 25 times that of the free ligand. In addition, we use the change in sign of the NOE from negative to positive near correlation times of 0.2 ns to our advantage by choosing a temperature at which contributions from free ligands in solution are near zero (~ 30 °C at 800 MHz for fondaparinux). Even with averaging over significant free-state populations, observed trNOEs can then be interpreted in terms of a $1/r^6$ dependence on inter-proton distances of the bound state. We use a distance of 2.5 Å between H2 and H4 of the GlcNS6S residue at the non-reducing terminus as a reference distance when converting trNOE intensity to distance.

Heparan sulfate glycosaminoglycans display a significant level of conformational variability, primarily due to differences in glycosidic linkage torsion angles and the multiple glycan ring conformations for IdoA residues. In Table 2, we report trNOE-based distances between pairs of nuclei affected either by changes in glycosidic torsion angles or ring conformations. The distances of the same pairs of protons determined from NOEs in the free state are included for comparison. There are some significant differences. For instance, the measurable transglycosidic distances between CH3 and DH1 are 0.62 Å longer in the bound state (3.05 Å) than that in the free state 2.43 Å. Similarly, a trNOE observed between DH1 and EH4 gives a calculated distance of 3.14 Å, which is significantly longer than the free state

distance (2.71 Å). These two trNOE distances measured from a bound fondaparinux suggest it adopts a preferred conformation which is different from the dominant free state conformation, particularly with respect to the glycosidic linkages between GlcNS3,6S (residue C) and IdoA2S (residue D), and IdoA2S (residue D) and GlcNS6S (residue E). Besides the transglycosidic distance, another piece of useful information comes from distances between protons on the IdoA2S ring. The distance between IdoA2S D H1 and H2 and D H1 and D H3 are 3.16 Å and 3.27 Å respectively in the free state. These suggest a mixture of a ¹C₄ chair and ²S₀ skew-boat conformation. The skew-boat conformation is known to be particularly favorable in solution when a 3-O-sulfate is attached to the preceding GlcNAc residue ⁴⁹. A trNOE between IdoA2S D H1 and H3 in the bound state gives a 2.72 Å distance, which is shorter than that of the free state (3.27 Å), indicating a much higher population of the skew-boat conformer and perhaps even some sampling of a true boat conformer. These trNOE data provide another source of structural information that can be applied as restraints in the docking process.

Oligomerization state calculated from rotational correlation times

A previous study showed that heparin oligosaccharides were able to induce oligomerization of RPTPσ ⁵⁰. In that study, heparin dp8 was the minimum length of a heparin oligosaccharide to promote RPTPσ oligomerization, whereas dp4 and dp6 did not. In our study a correlation time (τ_c) of 8.4 ns was determined from a cross-correlation based experiment on an LAR-fondaparinux complex (see table S5 for data). The molecular weight of LAR is 22 kDa giving estimated τ_cs of 11ns and 22ns for the monomer and dimeric states, respectively at 25°C. Hence, there is no indication of dimer formation under our experimental condition. Fondaparinux is 5 residues long, so while we cannot make conclusions about the effect of longer HS oligomers, our results are consistent with a model where polymers 6 saccharide units in length are capable of binding to a protein monomer, but are not capable of cross-linking and thereby oligomerizing subunits of the protein.

Complex modeling by computational docking

High Ambiguity Driven biomolecular DOCKing (HADDOCK) is a versatile software package for biomolecule docking that allows use of a variety of biophysical information ²⁶. Here we combine all the restraints deduced from different types of NMR measurements, including the chemical shift perturbation data, STD-based information and trNOE measurements, to produce a structural model of the LAR-fondaparinux complex. Details concerning the specific docking process are described in the Materials and Methods section. Although we have clear evidence for a conformation near that of a ²S₀-skew boat for the central IdoA residue in the bound form, two sets of HADDOCK runs were conducted starting with skew-boat and chair conformations separately. The energy minimized input structures for IdoA were made using the GLYCAM web tool ³⁶. The overall free energies of binding taken from a molecular mechanics-generalized Born surface area (MM-GBSA) calculation conducted on a 50 ns MD simulation of docked structures were in fact significantly more negative for the skewed-boat structures (−39.34 kcal/mol for the best chair structure versus −43.71kcal/mol for the average of the best three skew-boat structures – see Table S6 for details). Hence, we focus on the skew-boat structures in the following discussion. Three structures with the lowest HADDOCK scores and no NOE violations

more than 0.5 Å are presented in Figure 7A. These also have the lowest molecular energies. Comparing the docked structures, we find that most of the structural differences arise from the last two sugars (IdoA2S and the reducing end GlcNS6S, residues D and E). These interact with multiple lysines and arginines on the binding loop. A single structural representation of the protein residues having close contacts with the ligand is illustrated in Figure 7B.

Free energy calculation and analysis

Molecular mechanics-generalized Born surface area (MM-GBSA)⁴¹ calculations were conducted to identify residues making major contributions to a free energy of binding. Even though docking finds the same binding pocket, the top 5 docked structures with IdoA2S in the ²S₀ conformation were not well aligned and showed three types of positioning, therefore 3 structures representing different bound states were chosen for analysis of binding energies (all within +/- 3.8 kcal). The complete energy analysis is shown in Table S7. Because the conformational and vibrational entropy terms are excluded from the calculation and conformations for both protein and ligand are assumed to be the same in the dissociated state as in the associated state, there is typically an overestimation of binding energies. However, the protein part of the overestimation, which is associated with not allowing the protein to adjust conformation and interact optimally with both solvent and counter-ions in the free state, is the same for all ligands, and will not have large effects on ordering of per-residue contributions to interaction energies. Errors coming from solvent interactions with the free ligand may also be moderate because most surfaces of the ligand are solvent exposed, regardless of conformation. As we will see below, the analysis provides some valuable insight into interactions that drive binding and provide selectivity.

Discussion

It is useful to first compare our docked models with the crystal structure of LAR complexed with sucrose octasulfate (pdb 2YD8)¹². In this structure the ligand is within 1 Å van der Waals contact with three residues of the protein: K68, K69 and R77. In our study, docked low energy structures of fondaparinux come within 1 Å of van der Waals contact with residues K68, K69, K71, S74, S75, R77 and R100. The three residues identified in the sucrose octasulfate complex structure are in this set. Five of the residues in contact with fondaparinux are positively charged. Fondaparinux is a highly negatively charged ligand. There are 8 different sulfate groups and two carboxylate groups distributed to both sides of the ligand, giving many opportunities for electrostatic interactions with positively charged protein residues. Even though sucrose octasulfate is not a close structural mimic of HS, it also has a high level of sulfation and this is likely responsible for the overlap in interacting residues. Examining the protein structure, we find that all of the binding site contact residues for fondaparinux are located on loops. Significant motional flexibility of these loops as well as the flexibility of lysine and arginine sidechains likely contributes to the distribution of structures seen in Figure 7A.

It is also useful for us to compare the results for LAR-fondaparinux to those from our previous study of an HS tetramer binding to the axon guidance molecule, Robo1¹⁸. A

binding constant of 45 μM was obtained for the interaction of Robo1 and an HS tetramer which only contains 5 sulfates, and two carboxylates. Given that fondaparinux is a much more highly sulfated GAG and more positive protein residues are involved, a binding constant of 60 μM may suggest that fondaparinux is not the most ideal ligand for LAR interactions. MM-GBSA free energies and an examination of the ligand portion of the decomposition of these energies may provide some insight into which groups are more important, and in the future could guide the identification or synthesis of higher-affinity ligands.

We treated the 2-O-sulfate 3-O-sulfate and 6-O-sulfates as separate residues in this decomposition so their contributions can be explicitly examined. These sulfates have strong interactions with the protein dominated by electrostatic terms (see Table S7). Interaction energy contributions from the sugar rings also come primarily from electrostatic terms. The high electrostatic energy can, however, be compensated by high desolvation energies and even result in a positive total binding energy contribution in some cases. A similar phenomenon was observed in our study of Robo1 interacting with an HS tetramer. To be more specific, among the 5 O-sulfates in fondaparinux (three 6-O-sulfates, one 3-O-sulfate and one 2-O-sulfate), only the 3-O-sulfate on the middle GlcNS3,6S shows an average favorable total energy (-3.45 kcal/mol). It is worth noting that 3-O-sulfation is rare in native HS⁵¹, suggesting a role in possible targeting to specific regions of HS chains. All of the other O-sulfates, although having very large electrostatic contributions, have this offset by a higher desolvation penalty (polar plus non-polar solvation terms). This doesn't mean that electrostatic interactions between negatively charged ligand groups and positively charged protein residues are unimportant. If they didn't exist, desolvation penalties would prevent any significant binding. It does mean that other energy contributions are often important drivers of the interaction. Among the sugar rings, which contain either a negatively charged carboxylate in the case of GlcA and IdoA or additional N-sulfates in the case of the two GlcNS6S residues and the GlcNS3,6S residue, the non-reducing end GlcNS6S (residue A) shows the most favorable energy of interaction. About a third of this comes from a favorable electrostatic interaction that out-weighs desolvation effects and the remainder comes from favorable van der Waals interactions. The dominant interactions for residue A are indicated with dotted lines in Figure 7B.

It is also worth considering what changes in ligand structure might enhance interactions. Areas where potential electrostatic interactions can compensate for desolvation of additional sulfate groups are easiest to recognize. Examining the current structure, one finds that K68 is well positioned to interact with a 2-O-sulfate if the GlcA residue near the non-reducing end were replaced with IdoA2S. Extending the oligomer at the non-reducing end with an additional IdoA2S may also add favorable contacts with R97. Some additional modeling ligands with these additions can be used to justify experimental investigation.

In summary, biomolecular NMR measurements combined with high ambiguity driven docking has provided a detailed model for the interaction of a well-defined heparan sulfate pentasaccharide (fondaparinux) with the first two Ig domains of human LAR. Compared with a crystal structure where LAR is co-crystallized with sucrose octasulfate, the structure produces a set of molecular interactions that are likely to be more closely representative of

physiologically-relevant ligands such as heparan and chondroitin sulfate polymers. In addition to providing a mechanistic understanding of the type IIa RPTPs, including their signaling mechanisms, structural studies of the atomic basis of the ligand interactions of these receptors may allow future rationally-designed strategies to identify or synthesize therapeutic compounds to promote nervous system repair and regeneration. The methodology developed here, particularly that for the assignment of spectra from sparsely labeled proteins, also opens application to a large number of other proteins that are best expressed in mammalian cells.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding

This work was supported by a grant from the National Institutes of Health, P41GM103390. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

The authors thank Laura Morris and Dr. David Thieker for advice on MD simulations and MM-GBSA methods.

References

1. Tonks NK. Protein tyrosine phosphatases: from genes, to function, to disease. *Nat Rev Mol Cell Bio.* 2006; 7:833–846. [PubMed: 17057753]
2. Coles CH, Jones EY, Aricescu AR. Extracellular regulation of type IIa receptor protein tyrosine phosphatases: mechanistic insights from structural analyses. *Semin Cell Dev Biol.* 2015; 37:98–107. [PubMed: 25234613]
3. Coles CH, Mitakidis N, Zhang P, Elegheert J, Lu WX, Stoker AW, Nakagawa T, Craig AM, Jones EY, Aricescu AR. Structural basis for extracellular cis and trans RPTP sigma signal competition in synaptogenesis. *Nat Commun.* 2014; 5209:1–12.
4. Fisher D, Xing B, Dill J, Li H, Hoang HH, Zhao ZZ, Yang XL, Bachoo R, Cannon S, Longo FM, Sheng M, Silver J, Li SX. Leukocyte Common Antigen-Related Phosphatase Is a Functional Receptor for Chondroitin Sulfate Proteoglycan Axon Growth Inhibitors. *J Neurosci.* 2011; 31:14051–14066. [PubMed: 21976490]
5. Dunah AW, Hueske E, Wyszynski M, Hoogenraad CC, Jaworski J, Pak DT, Simonetta A, Liu G, Sheng M. LAR receptor protein tyrosine phosphatases in the development and maintenance of excitatory synapses. *Nature neuroscience.* 2005; 8:458–467. [PubMed: 15750591]
6. Ohtake Y, Li SX. Molecular mechanisms of scar-sourced axon growth inhibitors. *Brain Research.* 2015; 1619:22–35. [PubMed: 25192646]
7. Shen YJ, Tenney AP, Busch SA, Horn KP, Cuascut FX, Liu K, He ZG, Silver J, Flanagan JG. PTP sigma Is a Receptor for Chondroitin Sulfate Proteoglycan, an Inhibitor of Neural Regeneration. *Science.* 2009; 326:592–596. [PubMed: 19833921]
8. Stoker AW. RPTPs in axons, synapses and neurology. *Semin Cell Dev Biol.* 2015; 37:90–97. [PubMed: 25234542]
9. Takahashi H, Craig AM. Protein tyrosine phosphatases PTP δ , PTP σ , and LAR: presynaptic hubs for synapse organization. *Trends in neurosciences.* 2013; 36:522–534. [PubMed: 23835198]
10. Van der Zee C, Man TY, Van Lieshout EMM, Van der Heijden I, Van Bree M, Hendriks W. Delayed peripheral nerve regeneration and central nervous system collateral sprouting in leucocyte common antigen-related protein tyrosine phosphatase-deficient mice. *European Journal of Neuroscience.* 2003; 17:991–1005. [PubMed: 12653975]

11. Xie YM, Yeo TT, Zhang C, Yang T, Tisi MA, Massa SM, Longo FM. The leukocyte common antigen-related protein tyrosine phosphatase receptor regulates regenerative neurite outgrowth in vivo. *J Neurosci*. 2001; 21:5130–5138. [PubMed: 11438588]
12. Coles CH, Shen Y, Tenney AP, Siebold C, Sutton GC, Lu W, Gallagher JT, Jones EY, Flanagan JG, Aricescu AR. Proteoglycan-specific molecular switch for RPTP σ clustering and neuronal extension. *Science*. 2011; 332:484–488. [PubMed: 21454754]
13. Prestegard JH, Agard DA, Moremen KW, Lavery LA, Morris LC, Pederson K. Sparse labeling of proteins: structural characterization from long range constraints. *J Magn Reson*. 2014; 241:32–40. [PubMed: 24656078]
14. Capila I, Linhardt RJ. Heparin–protein interactions. *Angewandte Chemie International Edition*. 2002; 41:390–412.
15. Raman R, Sasisekharan V, Sasisekharan R. Structural insights into biological roles of protein-glycosaminoglycan interactions. *Chemistry & biology*. 2005; 12:267–277. [PubMed: 15797210]
16. Brown JM, Xia J, Zhuang BQ, Cho KS, Rogers CJ, Gama CI, Rawat M, Tully SE, Uetani N, Mason DE, Tremblay ML, Peters EC, Habuchi O, Chen DF, Hsieh-Wilson LC. A sulfated carbohydrate epitope inhibits axon regeneration after injury. *P Natl Acad Sci USA*. 2012; 109:4768–4773.
17. Zhuo Y, Yang JY, Moremen KW, Prestegard JH. Glycosylation Alters Dimerization Properties of a Cell-surface Signaling Protein, Carcinoembryonic Antigen-related Cell Adhesion Molecule 1 (CEACAM1). *J Biol Chem*. 2016; 291:20085–20095. [PubMed: 27471271]
18. Gao Q, Chen CY, Zong C, Wang S, Ramiah A, Prabhakar P, Morris LC, Boons GJ, Moremen KW, Prestegard JH. Structural Aspects of Heparan Sulfate Binding to Robo1-Ig1-2. *ACS Chem Biol*. 2016; 11:3106–3113. [PubMed: 27653286]
19. Gao Q, Chalmers GR, Moremen KW, Prestegard JH. NMR assignments of sparsely labeled proteins using a genetic algorithm. *Journal of biomolecular NMR*. 2017; 67:283–294. [PubMed: 28289927]
20. Pederson K, Chalmers GR, Gao Q, Elnatan D, Ramelot TA, Ma LC, Montelione GT, Kennedy MA, Agard DA, Prestegard JH. NMR characterization of HtpG, the E-coli Hsp90, using sparse labeling with C-13-methyl alanine. *Journal of biomolecular NMR*. 2017; 68:225–236. [PubMed: 28653216]
21. Prestegard JH, Sahu SC, Nkari WK, Morris LC, Live D, Gruta C. Chemical shift prediction for denatured proteins. *Journal of biomolecular NMR*. 2013; 55:201–209. [PubMed: 23297019]
22. Feng L, Lee HS, Prestegard JH. NMR resonance assignments for sparsely ¹⁵N labeled proteins. *Journal of biomolecular NMR*. 2007; 38:213–219. [PubMed: 17487550]
23. Nkari WK, Prestegard JH. NMR Resonance Assignments of Sparsely Labeled Proteins: Amide Proton Exchange Correlations in Native and Denatured States. *J Am Chem Soc*. 2009; 131:5344–5349. [PubMed: 19317468]
24. Barthelmes K, Reynolds AM, Peisach E, Jonker HRA, DeNunzio NJ, Allen KN, Imperiali B, Schwalbe H. Engineering Encodable Lanthanide-Binding Tags into Loop Regions of Proteins. *J Am Chem Soc*. 2011; 133:808–819. [PubMed: 21182275]
25. Gao Q, Chen CY, Zong C, Wang S, Ramiah A, Prabhakar P, Morris LC, Boons GJ, Moremen KW, Prestegard JH. Structural Aspects of Heparan Sulfate Binding to Robo1-Ig1-2. *Acs Chemical Biology*. 2016; 11:3106–3113. [PubMed: 27653286]
26. Dominguez C, Boelens R, Bonvin AMJJ. HADDOCK: A protein-protein docking approach based on biochemical or biophysical information. *J Am Chem Soc*. 2003; 125:1731–1737. [PubMed: 12580598]
27. Moremen KW, Ramiah A, Stuart M, Steel J, Meng L, Forouhar F, Moniz HA, Gahlay G, Gao ZW, Chapla D, Wang S, Yang JY, Prabhakar PK, Johnson R, dela Rosa M, Geisler C, Nairn AV, Seetharaman J, Wu SC, Tong L, Gilbert HJ, LaBaer J, Jarvis DL. Expression system for structural and functional studies of human glycosylation enzymes. *Nature Chemical Biology*. 2018; 14:156–162. [PubMed: 29251719]
28. Subedi GP, Johnson RW, Moniz HA, Moremen KW, Barb A. High Yield Expression of Recombinant Human Proteins with the Transient Transfection of HEK293 Cells in Suspension. *JoVE (Journal of Visualized Experiments)*. 2015:e53568–e53568. [PubMed: 26779721]

29. Tjandra N, Grzesiek S, Bax A. Magnetic field dependence of nitrogen-proton J splittings in N-15-enriched human ubiquitin resulting from relaxation interference and residual dipolar coupling. *J Am Chem Soc.* 1996; 118:6264–6272.
30. Liu YZ, Prestegard JH. Direct measurement of dipole-dipole/CSA cross-correlated relaxation by a constant-time experiment. *J Magn Reson.* 2008; 193:23–31. [PubMed: 18406649]
31. Mazak K, Beecher CN, Kraszni M, Larive CK. The interaction of enoxaparin and fondaparinux with calcium. *Carbohydrate Research.* 2014; 384:13–19. [PubMed: 24334236]
32. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. Nmrpipe - a Multidimensional Spectral Processing System Based on Unix Pipes. *Journal of biomolecular NMR.* 1995; 6:277–293. [PubMed: 8520220]
33. Goddard, T., Kneller, D. SPARKY 3. Vol. 15. University of California; San Francisco: 2004.
34. Li DW, Bruschweiler R. PPM_One: a static protein structure based chemical shift predictor. *Journal of biomolecular NMR.* 2015; 62:403–409. [PubMed: 26091586]
35. Qi Gao GRC, Moremen Kelley W, Prestegard James H. NMR Assignments of Sparsely Labeled Proteins Using a Genetic Algorithm. submitted.
36. Singh A, Tessier MB, Pederson K, Wang XC, Venot AP, Boons GJ, Prestegard JH, Woods RJ. Extension and validation of the GLYCAM force field parameters for modeling glycosaminoglycans. *Canadian Journal of Chemistry.* 2016; 94:927–935. [PubMed: 28603292]
37. Case D, Babin V, Berryman J, Betz R, Cai Q, Cerutti D, Cheatham T III, Darden T, Duke R, Gohlke H. Amber. 2014:14.
38. Case D, VBTTBBR, Cai Q, Cerutti D, Cheatham T III, Darden T, Duke R, Gohlke H, Goetz A, Gusarov S. The FF14SB force field. *AMBER.* 2014; 14:29–31.
39. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF chimera - A visualization system for exploratory research and analysis. *J Comput Chem.* 2004; 25:1605–1612. [PubMed: 15264254]
40. Kirschner KN, Yongye AB, Tschampel SM, Gonzalez-Outeirino J, Daniels CR, Foley BL, Woods RJ. GLYCAM06: A generalizable Biomolecular force field. *Carbohydrates. J Comput Chem.* 2008; 29:622–655. [PubMed: 17849372]
41. Kollman PA, Massova I, Reyes C, Kuhn B, Huo SH, Chong L, Lee M, Lee T, Duan Y, Wang W, Donini O, Cieplak P, Srinivasan J, Case DA, Cheatham TE. Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. *Accounts Chem Res.* 2000; 33:889–897.
42. Hadden JA, Tessier MB, Fadda E, Woods RJ. Calculating binding free energies for protein-carbohydrate complexes. *Method Mol Biol.* 2015; 1273:431–465.
43. Aricescu AR, McKinnell IW, Halfter W, Stoker AW. Heparan sulfate proteoglycans are ligands for receptor protein tyrosine phosphatase sigma. *Mol Cell Biol.* 2002; 22:1881–1892. [PubMed: 11865065]
44. Lipsitz RS, Tjandra N. Residual dipolar couplings in NMR structure analysis. *Annu Rev Bioph Biom.* 2004; 33:387–413.
45. Bax A. Weak alignment offers new NMR opportunities to study protein structure and dynamics. *Protein Sci.* 2003; 12:1–16. [PubMed: 12493823]
46. Morlot C, Thielens NM, Ravelli RBG, Hemrika W, Romijn RA, Gros P, Cusack S, McCarthy AA. Structural insights into the Slit-Robo complex. *P Natl Acad Sci USA.* 2007; 104:14923–14928.
47. Bhunia A, Bhattacharjya S, Chatterjee S. Applications of saturation transfer difference NMR in biological systems. *Drug discovery today.* 2012; 17:505–513. [PubMed: 22210119]
48. Pederson K, Mitchell DA, Prestegard JH. Structural Characterization of the DC-SIGN-Lewis(X) Complex. *Biochemistry-Us.* 2014; 53:5700–5709.
49. Hsieh PH, Thieker DF, Guerrini M, Woods RJ, Liu J. Uncovering the Relationship between Sulphation Patterns and Conformation of Iduronic Acid in Heparan Sulphate. *Sci Rep-Uk.* 2016; 29602:1–8.
50. Coles CH, Shen YJ, Tenney AP, Siebold C, Sutton GC, Lu WX, Gallagher JT, Jones EY, Flanagan JG, Aricescu AR. Proteoglycan-Specific Molecular Switch for RPTP sigma Clustering and Neuronal Extension. *Science.* 2011; 332:484–488. [PubMed: 21454754]

51. Thacker BE, Xu D, Lawrence R, Esko JD. Heparan sulfate 3-O-sulfation: a rare modification in search of a function. *Matrix Biol.* 2014; 35:60–72. [PubMed: 24361527]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 1.

(A) The LAR Ig1-2 structure is shown as a ribbon diagram with Lys residues labeled in blue.

(B) Structure of the heparan sulfate pentasaccharide used in this study with sulfate groups labeled in red.

Figure 2.
2D ^{15}N - ^1H HSQC spectra of (red) ^{15}N Lys labeled LAR and (blue) ^{15}N Lys labeled LAR-loop. Each peak in the LAR-loop spectrum is labeled with a peak number as a reference for the following assignment.

Figure 3. Superposition of ^{15}N - ^1H HSQC spectra of ^{15}N -Lys labeled LAR Ig1-2 (red) and LAR Ig1-2 engineered with a lanthanide binding peptide loaded with Gd^{3+} (blue)

Figure 4. Histogram of frequency of assignment of crosspeaks (measurements) to specific residues using the 1000 ns frame as an input structure.

Figure 5. Correlation plot of experimental RDCs and back-calculated RDCs for (A) phage and (B) PEG bicelle media. The crystal structure, 2YD5 was used for back-calculation.

Figure 6.

(A) Superimposed HSQC spectra of 170 μM LAR-Ig1-2 with increasing concentrations of fondaparinux from 0 μM to 550 μM color coded from red to purple. (B) Maximum chemical shift perturbation of each Lys residue is plotted against the crosspeak number.

Figure 7. HADDOCK structures with the highest score and lowest energy for the LAR-fondaparinux complex, starting with IdoA2S in the 2S_0 -skew boat conformation. (A) Top three structures are shown superimposed (colored cyan, blue and purple). The A residue at the non-reducing terminus of each is colored black. (B) Expanded view of the binding pocket for the LAR-fondaparinux HADDOCK structure with the lowest energy. While IdoA2S started in a 2S_0 conformation, it is more boat-like in this structure. Nearby interacting protein residues are shown as stick structures with their labels. For the A residue of fondaparinux contacts to protein residues at distances less than 4 Å (heavy atom to heavy atom) are shown as dotted lines.

Table 1

Crosspeak Assignment of ^{15}N -Lys LAR-loop.

Cross-peak	1	2	3	4	5	6	7	8	9	10	11	12	13
^1H ppm	9.46	9.15	9.11	8.85	8.74	8.69	8.67	8.59	8.51	8.26	8.05	7.98	7.96
^{15}N ppm	123.1	121.6	128.1	124.8	125.2	127.8	121.3	123.9	123.7	122.3	120.4	122.1	123.6
Residue	K144	K121 [*]	K69 [*]	K32	K68 [#]	K185 [*]	K170 [#]	K61 [*]	K204 [*]	K72 [#]	K71 [*]	K37 [*]	K148 [*]

^{*} indicates definitive assignment using one or more frames.

[#] indicates less than definitive assignment, but most favorable in two or more frames

Table 2

Interproton distances of fondaparinux derived in free and bound states from NOE and trNOE data. Error estimates are based on the signal to noise ratio of individual peaks.

Nuclei pair	Free ligand (Å)	Error (Å)	trNOE (Å)	Error (Å)
AH1-BH4	2.62	0.02	2.86	0.21
CH1-DH4	2.41	0.02	2.65	0.20
CH3-DH1	2.43	0.11	3.05	0.25
DH1-DH2	3.16	0.01	N/A	N/A
DH1-DH3	3.27	0.01	2.72	0.19
DH1-EH4	2.71	0.02	3.14	0.25
DH1-EH61	2.85	0.02	2.91	0.23
DH1-EH62	3.22	0.03	2.49	0.19
O-Me H-EH1	2.96	0.02	3.22	0.20