Deciphering the structural attributes of protein–heparan sulfate interactions using chemo-enzymatic approaches and NMR spectroscopy

Aurélie Préchoux, Jean-Pierre Simorre, Hugues Lortat-Jacob, Cédric Laguri

Glycobiology, 2021, 1–8
doi: 10.1093/glycob/cwab012
Regular Manuscripts

Structural Biology

Deciphering the structural attributes of protein–heparan sulfate interactions using chemo-enzymatic approaches and NMR spectroscopy

Aurélie Préchoux, Jean-Pierre Simorre, Hugues Lortat-Jacob* and Cédric Laguri*

Université Grenoble Alpes, CNRS, CEA, IBS, F-38000 Grenoble, France

*To whom correspondence should be addressed: Tel.: (+33) 457428577, (+33) 457428533; Fax: (+33) 476501890; e-mail: cedric.laguri@ibs.fr; e-mail: hugues.lortat-jacob@ibs.fr

Received 18 November 2020; Revised 18 December 2020; Editorial Decision 28 January 2021; Accepted 28 January 2021

Abstract

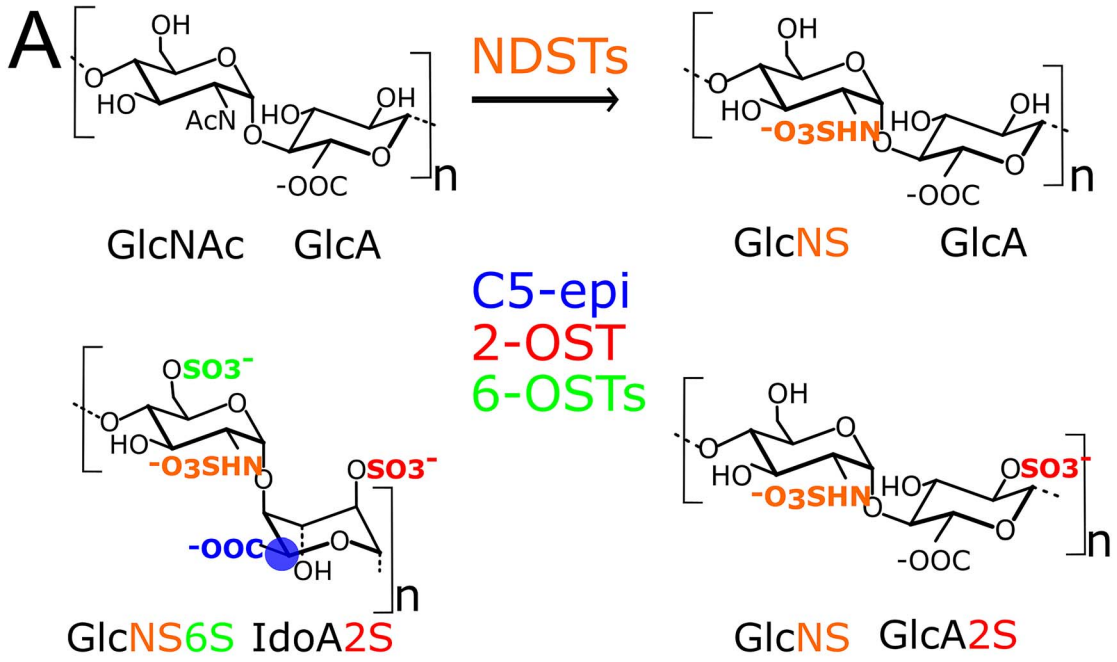
Heparan sulfates (HS) is a polysaccharide found at the cell surface, where it mediates interactions with hundreds of proteins and regulates major pathophysiological processes. HS is highly heterogeneous and structurally complex and examples that define their structure–activity relationships remain limited. Here, in order to characterize a protein–HS interface and define the corresponding saccharide-binding domain, we present a chemo-enzymatic approach that generates ^{13}C-labeled HS-based oligosaccharide structures. Nuclear magnetic resonance (NMR) spectroscopy, which efficiently discriminates between important or redundant chemical groups in the oligosaccharides, is employed to characterize these molecules alone and in interaction with proteins. Using chemokines as model system, docking based on NMR data on both proteins and oligosaccharides enable the identification of the structural determinant involved in the complex. This study shows that both the position of the sulfo groups along the chain and their mode of presentation, rather than their overall number, are key determinant and further points out the usefulness of these ^{13}C-labeled oligosaccharides in obtaining detailed structural information on HS–protein complexes.

Key words: chemokine, glycosaminoglycans, NMR

Introduction

Glycosaminoglycans are at the heart of many biological systems and regulatory mechanisms (Sarrazin et al. 2011). Heparan sulfate (HS), in particular, affect the conformation of hundreds of proteins to which they bind, promoting signaling protein–receptor recognition or controlling in time and space the stability and distribution of their ligands (Kjellén and Lindahl 2018). HS have a unique level of structural complexity. It consists of a repeating disaccharide unit (Figure 1), *N*-acetyl glucosamine (GlcNAc) and glucuronic acid (GlcA), which undergoes during biosynthesis, in a cell type-dependent manner, a series of distinct sulfation and epimerization reactions catalyzed by Golgi-localized enzymes (Kreuger and Kjellén 2012). These reactions, which do not occur uniformly along the polymer, are believed to confer to the polysaccharide distinct and biologically relevant docking sites for proteins. With only few exceptions, however, the structural principles of protein–HS recognition are still unclear, and the question of which type of HS sequence are recognized by proteins remains central (Meneghetti et al. 2015).

Progresses in the field have been hampered by different levels of conceptual and technological challenges that remain to be solved: (1) the generation of libraries of structurally defined oligosaccharides covering a sufficient large sequence variety to probe the interactions, (2) the structural characterization of these molecules that are highly flexible and dynamic, (3) the acquisition of atomic scale details on



B

	1	2	3	4	5	6	sulfates
dp4[1]	ΔHexA	GlcNS	**IdoA2S**	GlcNS			3
dp4[2]	ΔHexA	**GlcNS6S**	**IdoA2S**	GlcNS			4
dp4[3]	ΔHexA	**GlcNS6S**	GlcA	GlcNS			3
dp4[4]	ΔHexA	**GlcNS6S**	GlcA	**GlcNS6S**			4
dp6[1]	ΔHexA	**GlcNS6S**	GlcA	GlcNS	GlcA	GlcNS	4
dp6[2]	ΔHexA	GlcNS	**GlcA2S**	GlcNS	**IdoA2S**	GlcNS	5
dp6[3]	ΔHexA	**GlcNS6S/GlcNS**	**GlcA/IdoA/IdoA2S**	**GlcNS6S**	**GlcA/IdoA/IdoA2S**	**GlcNS/GlcNS6S**	>5
dp6[4]	ΔHexA	**GlcNS6S**	**IdoA2S**	**GlcNS6S**	**IdoA2S**	GlcNS	7
dp6[5]	ΔHexA	GlcNS	**IdoA2S**	GlcNS	**IdoA2S**	GlcNS	5

Fig. 1. (**A**) Heparan sulfate modifications. N-deacetylation, N-sulfation step and examples of modifications produced by C5-epi and 2- or 6-OST enzymes. (**B**) Table summarizing the tetra- and hexasaccharides chemo-enzymatically produced and purified and their sequence determined by NMR, including dp6[3] that remained heterogeneous.

HS–proteins complexes that frequently feature inherent plasticity and could be of low affinity and (4) the likelihood that the interaction is governed more by a combination of charge distribution and conformational flexibility, rather than solely by the sequence per se itself.

Results

Generation of ^{13}C-labeled HS oligosaccharides

Here, we developed two related strategies to engineer HS oligosaccharides from ^{13}C–heparosan purified from the capsule of *Escherichia coli* 010:K5/H4 grown in ^{13}C-labeled glucose-containing medium and chemically N-deacetylated and re-N-sulfated (Supplementary Figure S1). In the first approach, we combined the capacity of the C5-epimerase (C5-epi) and 2-*O*-sulfotransferase (OST) enzymes to generate extended GlcNS-IdoA2S repeats when they function as a complex, and the specificity of Heparinase III to generate ΔHexA-(GlcNS-IdoA2S/GlcA2S)$_n$-GlcNS-resistant domains, which were size fractionated, modified with 6-OST-1, and then purified by high-performance liquid chromatography (HPLC) (Préchoux et al. 2015). The second approach consists first in generating size-defined (GlcNS-GlcA)$_n$ repeats through controlled Heparinase II digestion of N-sulfated heparosan, followed by sequential modification by the C5-epi/2-OST and 6-OST-1 enzymes and purification by HPLC (Supplementary Figure S1). We isolated four tetrasaccharides and five hexasaccharides with unique and defined structures (Figure 1), except for one structurally heterogeneous hexasaccharide (dp6[3]). Oligosaccharides compositions were assessed by ^{1}H–^{13}C nuclear magnetic resonance (NMR) (Supplementary Figure S2), using the ^{1}H and ^{13}C chemical shifts of the pyranose ring that are sensitive on its modifications (*O*-sulfate substituents or epimerization) and on the modifications of the neighboring sugar.

Interaction of CXCL12 chemokines with HS tetrasaccharides

As a model system to investigate their structure–binding activity relationships, we used two related chemokines, chemokine ligand 12 (CXCL12α) and CXCL12γ, whose interactions with HS have

been functionally characterized (Sadir et al. 2001; Laguri et al. 2007; Rueda et al. 2008; Connell et al. 2016). They display distinct affinity for heparin, used as an HS proxy in binding studies, of 93 and 0.9 nM. CXCL12α and CXCL12γ occur from alternative splicing of the same gene, possess the same first 68 amino acids, with a 30 amino acids long C-terminal unstructured and highly basic extension for the γ isoform. CXCL12γ possesses two HS binding sites, one strictly required, found in the core structured domain of the protein shared with CXCL12α, and the other one in the C-terminus, which, by enhancing the half-life of the complexes, contributes to a strong retention on cell surfaces (Laguri et al. 2007).

The interaction of four oligosaccharides ($dp4^{1-4}$) was tested with adding increasing protein concentration (CXCL12γ, CXCL12α) to the oligosaccharide up to 4:1 protein:oligosaccharide molar ratios while monitoring the sugars resonances by ^{13}C–^{1}H correlation NMR experiments. Changes in the environment of the oligosaccharides nuclei due to protein interaction are reflected by a change in their chemical shift, termed hereafter chemical shift perturbation (CSP).

Interaction with $dp4^1$ that contains an IdoA2S in position 3 with CXCL12α and CXCL12γ shows CSPs around the IdoA2S residue, reflecting protein binding at this position (Figure 2 and Supplementary Figure S3). Addition of a 6-*O*-sulfate to $dp4^1$, leading to $dp4^2$, did not change CSP compared with $dp4^1$ for any protein, suggesting that this sulfate group is not important for binding, compared with the IdoA2S group.

Binding to $dp4^1$ can be compared with an oligosaccharide of different structure, $dp4^3$ (with a 6-sulfate at position 2 and no IdoA2S), but with the same overall number of sulfate groups. No significant CSP can be observed, suggesting poor binding to this oligosaccharide. Addition of one 6-*O*-sulfate group on the reducing end glucosamine of $dp4^3$, leading to $dp4^4$, significantly increases CSP and thus binding around the additional 6-*O*-sulfate, especially for CXCL12γ (Figure 3 and Supplementary Figure S3).

Modeling of hexasaccharides–protein complexes using NMR data on both partners

Using tetrasaccharides, it was thus possible to observe significant differences in the binding of proteins with respect to overall sulfate numbers and to additions of 6-*O*-sulfates at a specific position. Nevertheless, tetrasaccharides have usually a low affinity for proteins and are shorter than the usual optimal size for binding (dp6–dp8). We thus investigated CXCL12 binding to two defined hexasaccharides, featuring a single difference so that structure–binding activity relationships can be easily identified.

An IdoA2S-containing dp6 ($dp6^5$), with a ΔHexA-(GlcNS-IdoA2S)2-GlcNS structure was first tested in interaction with CXCL12α and CXCL12γ. The hexasaccharides showed mostly CSP on IdoA2S residues and, as for $dp4^1$, mostly on the reducing end IdoA2S (Figure 3 and Supplementary Figure S4), pointing out the importance of this position for the binding process. Because the proteins used in these assays were also ^{15}N-labeled, structural information could be also recorded at the protein level. ^{15}N–^{1}H NMR spectra of both proteins showed significant CSPs upon $dp6^5$ interaction (Supplementary Figure S5) located around residues K24 and R41 with additional CSPs in the basic C-terminus for CXCL12γ, as previously shown with other oligosaccharides (Laguri et al. 2007, 2011).

In order to generate a 3D model of the complexes, based on these data, the 3D structure of CXCL12γ was first solved by NMR (protein data bank (PDB):6EHZ, Figure 3, Supplementary Figure S6 and Supplementary Table SI). CXCL12γ is monomeric in solution (Supplementary Figure S5) and its structure is identical to CXCL12α, except for the orientation of the C-terminal α-helix (~25°) and its disordered C-terminal extension. HADDOCK procedure (Dominguez et al. 2003) was next used to calculate a model of the protein–glycan complexes for both (^{15}N–^{1}H)-labeled CXCL12 isoforms. It combines use of residues on both protein (Supplementary Figure S5) and (^{13}C–^{1}H)-labeled ligand (Figure 3 and Supplementary Figures S4 and S5) that are perturbed by the interaction as input and docks the molecules by creating ambiguous interaction restraints between these residues to form the complex (Supplementary Table SII). Complexes obtained are then refined in explicit solvent and sorted in clusters (Supplementary Table SIII). CXCL12α exists as a monomer–dimer equilibrium in solution (Veldkamp 2005; Laguri et al. 2011) that is shifted towards dimer in presence of HS oligosaccharides, and the dimeric structure was thus used for modeling the complexes. On the other hand, CXCL12γ is monomeric, as demonstrated by analytical ultracentrifugation (Supplementary Figure S5D), and its NMR spectra in interaction with oligosaccharides do not suggest dimerization upon interaction. The monomeric form was thus considered for docking. Furthermore, CXCL12γ has two HS binding sites in the protein core and in the disordered C-terminus. Oligosaccharide binding to the chemokine core domain was considered only, for docking, in order to compare with CXCL12α.

The best models are shown in Figure 3. The orientation of the dp6 with respect to the protein is slightly different due to the dimeric and monomeric natures of CXCL12α and CXCL12γ, respectively, which create different binding surfaces. As expected from the CSP observed on the proteins (Supplementary Figure S6) as well as from previous NMR and mutational studies, the same residues of both proteins are involved in HS binding, primarily on β1 strand (Sadir et al. 2001; Laguri et al. 2007, 2011). The reducing end IdoA2S, which shows high CSP in $dp6^5$, is located at the same position in CXCL12α and CXCL12γ, in contact with K24, R41 and K27 (Supplementary Figure S7). The NMR data, associated with the models of the two complexes, clearly underline the importance of the reducing end IdoA2S that contacts residues K24 and K27, which are essential for HS binding (Sadir et al. 2001).

Influence of epimerization on binding to CXCL12α

We next investigated another hexasaccharide produced ($dp6^2$), presenting a GlcA2S instead of an IdoA2S at position 3. It offers a unique opportunity to observe the influence of the epimerization state at that position on CXCL12α binding. CSPs pattern upon CXCL12α addition to $dp6^2$ is similar to that of $dp6^5$, but also showed additional perturbation on the GlcNS preceding the second GlcA (Supplementary Figure S4). Docking of $dp6^2$ to CXCL12α following the same docking protocol allows comparison of the two complexes. Models containing $dp6^2$ present more extensive intermolecular contacts than with $dp6^5$, better physical energy and convergence of solutions, indicating an overall better model than with $dp6^5$ (Supplementary Figure S7 and Supplementary Table SIII). The 2-*O*-sulfate position in IdoA2S, which points toward the solvent, is flipped by 180° in GlcA2S so that, together with the sulfo groups of the consecutive GlcNS, it is inserted into the groove formed at the CXCL12α dimer interface (Figure 4 and Supplementary Figure S7), interacting with the K24, K27 and R41 residues of one monomer. Such reorientation did not affect the reducing end IdoA2S at position 5, which still interacts with K24, K27 and R41 on the other monomer. The ability of this oligosaccharide to extensively bind the two K24,

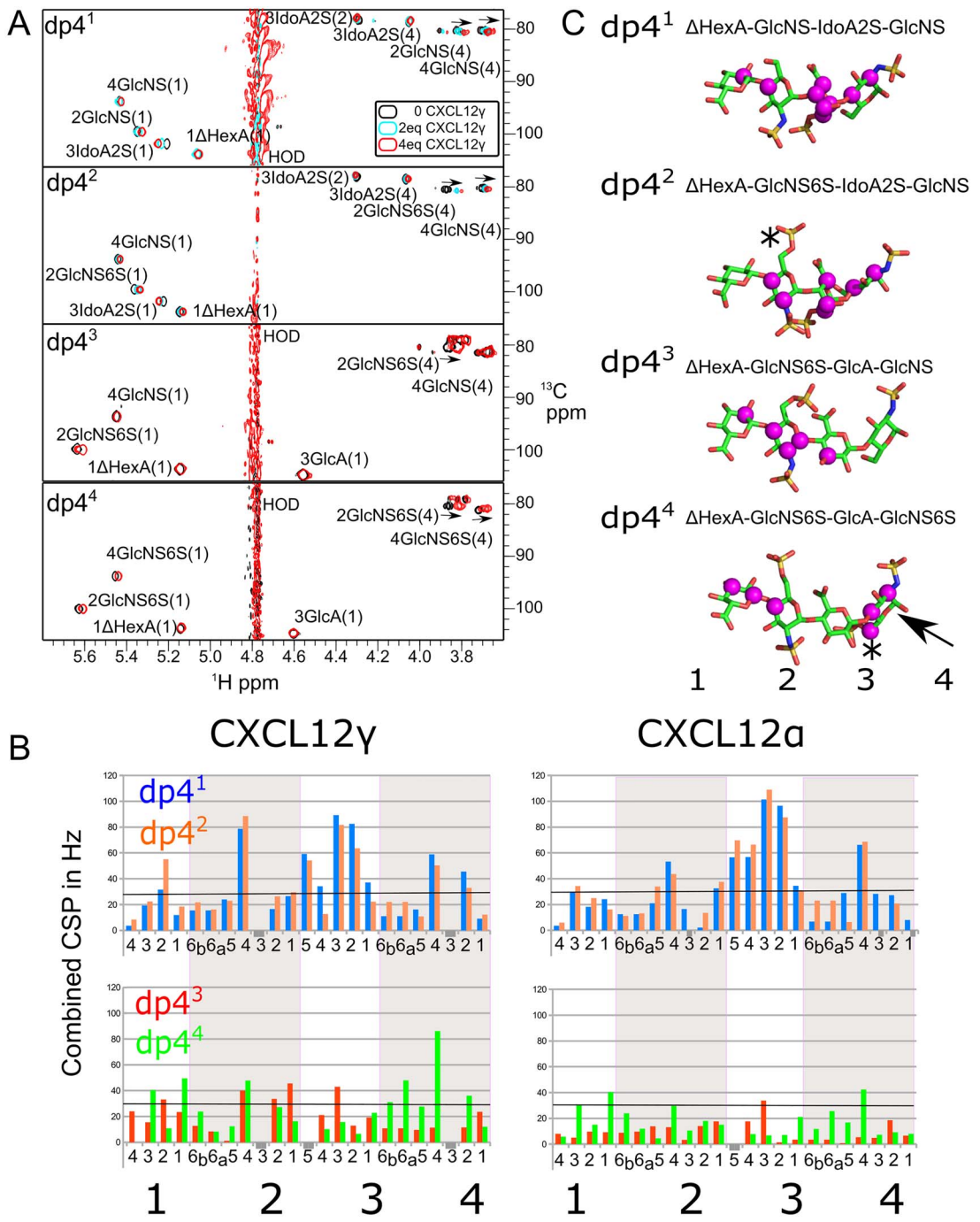


Fig. 2. Interaction of four different tetrasaccharides with CXCL12γ. (**A**) NMR spectra of dp4$^{1-4}$ upon CXCL12γ addition. (**B**) Chemical shift variations of the tetrasaccharides ^{13}C–^{1}H groups upon protein interaction (molar ratio protein:oligosaccharide: 4:1). A fixed threshold for all experiments (horizontal black line) was set corresponding to the standard deviation of the experiment showing the most extensive chemical shifts (CXCL12γ + dp4^{1}). (**C**) ^{1}H–^{13}C CSPs of the oligosaccharides visualized on the oligosaccharides structures in magenta, with an asterisk showing 6-sulfates and the arrow showing additional CSP induced by the 6-sulfate presence.

K27, R41 sites located at the CXCL12α dimer interface explains it forms a better complex than dp6[5].

Discussion

The increased availability of libraries of HS oligosaccharides is key to the understanding of protein–HS recognition (Hsieh et al. 2014; Lu et al. 2018). Here, we can investigate at the molecular level the influence of both the position of a sulfate group and the epimerization level in binding, thanks to ^{13}C-labeled oligosaccharides displaying different modification patterns. CXCL12 core domain, either in CXCL12α or CXCL12γ, recognizes primarily an IdoA2S residue penultimate to the reducing end (in dp6[2], dp6[5], dp4[1] and dp4[2]) that establishes contacts with residues important for HS binding (K24,

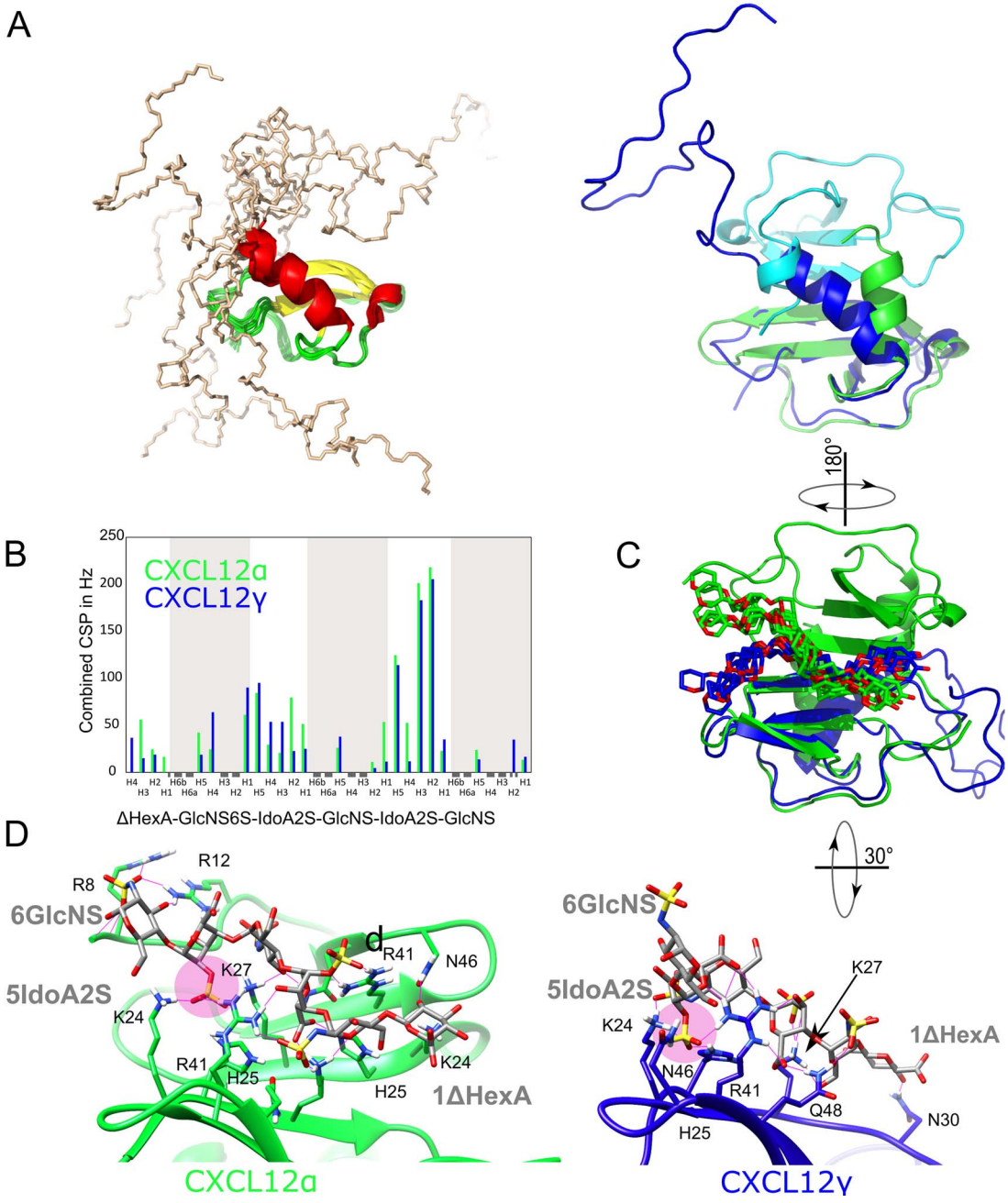


Fig. 3. Docking of dp6[5] with CXCL12α and CXCL12γ. (**A**) Left: Overlay of 10 structures of CXCL12γ determined by NMR (red helices, yellow β-sheet, green loops and wheat disordered C-terminal extension). Right: Superimposition of CXCL12γ (dark blue) with CXCL12α dimer (monomers in green and cyan). (**B**) ^{1}H–^{13}C CSP of dp6[5] upon interaction with CXCL12α and CXCL12γ. (**C**) Four best models of dp6[5] docked onto overlaid CXCL12α (green) and CXCL12γ (blue). Only the carbons and oxygens of the pyranose ring are shown for clarity. (**D**) CXCL12α and CXCL12γ in the same orientation showing intermolecular interactions with reducing end IdoA2S (magenta).

R41). Addition of a 6-sulfate group in the preceding glucosamine residue (position 4, Figure 4, see also dp4[2]) does not increase binding as it would point to the solvent in the model (Figure 4, bottom). Similarly, the IdoA2S sulfo group (position 3) in dp6[5] points towards the solvent and does not contribute to binding. In dp6[2], change of the IdoA2S to GlcA2S at this position inverts the conformation of the sugar and enables insertion of the 2-*O*-sulfate as well as the preceding GlcNS sulfo groups into a cluster of basic amino acids, leading to a better interaction energy compared with dp6[5], although both hexasaccharides have the same number of sulfate groups.

The model suggests that GlcA2S-GlcNS-IdoA2S-GlcNS motif might possess a good specificity for CXCL12α and could initiate rational optimization of this basic motif. The GlcNS (position 4 in Figure 4) does not establish contacts in the complex and might be replaced by a GlcNAc to reduce interactions with other HS-binding proteins, and a GlcNS6S at position 2 might provide additional interactions (with R41, Figure 4). Such an oligosaccharide could represent a good starting point for designing a specific inhibitor of CXCL12–CXCR4 axis, for possible applications in cancer research (Xue et al. 2017).

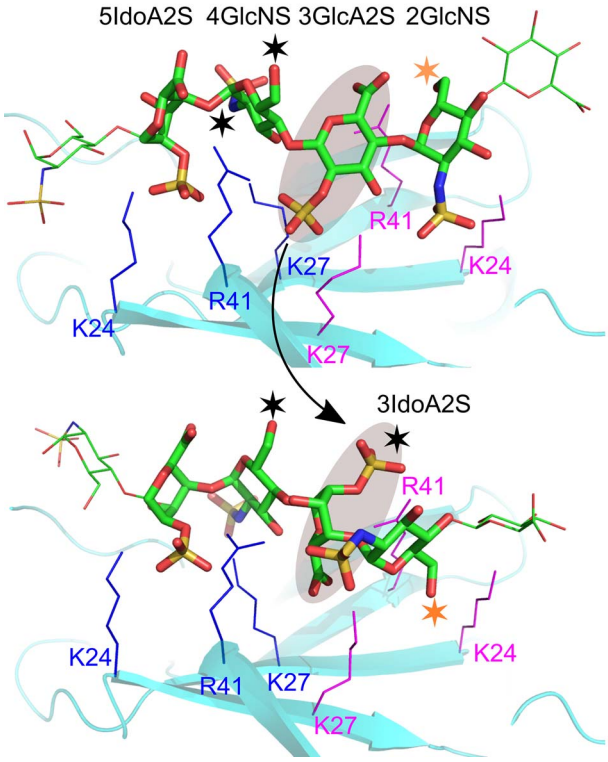


Fig. 4. Effect of epimerization on CXCL12α binding. GlcA2S in dp6[2] (top) allows insertion of three sulfate groups into the basic groove, compared with dp6[5] (bottom). Black and orange asterisks represent unnecessary or advantageous sulfo groups, respectively. Note that the reducing end sugar is on the left side in that figure.

The methodology we described here could be applied to many functionally important HS-binding proteins (cytokines, growth factors, etc.). It was also tested with the CXCL12 homolog in *Danio rerio*, CXCL12A (Supplementary Figure S8) (Boldajipour et al. 2011). On the contrary to CXCL12α or CXCL12γ, interaction with dp4[1–4] showed no preference for IdoA2S residue, but rather for dp4[4] containing two GlcNS6S, and a poor interaction with dp6[5], further illustrating the importance of structural investigations in the field of protein–HS interaction. Our observations support the postulate that the 3D arrangement of the oligosaccharide (Kjellén and Lindahl 2018) and hence the way it presents it sulfate groups to a given protein motif (flat motif, grooved motif), rather than the overall number of sulfo groups, is key to the recognition process.

Material and methods

Preparation of proteins

Mus musculus CXCL12α, CXCL12γ and *D. rerio* stromal cell derived factor (SDF1A) were expressed ^{15}N-labeled and purified as described (Laguri et al. 2007, 2011; Boldajipour et al. 2011). Human maltose binding protein (MBP)-C5-epi and MBP-2-OSTs, His6-Aryl sulfotransferase IV were expressed and purified as previously described (Préchoux et al. 2015). *M. musculus* 6-OST1 gene (corresponding to residues 53-401) was cloned between NdeI and BamHI restriction sites into a pMal-C5e plasmid (New England Biolabs). MBP-6OST1 is expressed in *E. coli* Origami B cells (NEB) containing the pGro7 plasmid (Takara) coding for the GroEL protein. Transformed cells were grown in Luria-Bertani (LB) medium until an OD600 of 0.6 and protein production was induced with 0.5 mM isopropylthiogalactoside (IPTG) and 1 mg/mL arabinose at 22°C Overnight. Purification by amylose affinity was identical to C5-epi purification.

Heparosan purification and chemo-enzymatic modifications

^{13}C-labeled heparosan is purified from *E. coli* K5 grown in M9 minimal medium supplemented with ^{13}C glucose as previously described (Laguri et al. 2011; Préchoux et al. 2015). N-sulfation is achieved by successive N-deacetylation and N-resulfation as described. Enzymatic reactions were performed at 25°C with 1 mg/mL of poly- or oligosaccharides in 25 mM 2-morpholinoethanesulfonic acid (MES), 5 mM $CaCl_2$, pH 7 buffer in presence of the 3′-phosphoadénosine 5′-phosphosulfate (PAPS) regeneration system (100 µM (Adenosine 3′,5′-diphosphate)(Merck) PAP, 10 mM potassium 4-nitrophenyl sulfate (PNPS) and 15 µg/mL of aryl sulfotransferase aryl sulfotransferase (AST) IV). For C5-epimerization 2-O-sulfation steps, C_5-epi and 2OST were added at 30 and 80 µg/mL, respectively, until a level of IdoA2S content of about 60% was reached, a condition that also gives rise to a low amount of GlcA2S. 6-O-sulfation step was applied with 180 µg/mL of protein in the reaction mixture. Enzymatic reactions were typically incubated for up to 4 days. Poly- and oligosaccharides were desalted using two G25 (GE Healthcare) in-line Hitrap G25 desalting columns equilibrated in Milli Q water. Desalting was done before and after enzymatic reactions and followed by freeze-drying. Oligosaccharides are purified according to their size on two consecutive superdex peptide 10/300GL columns (GE Healthcare) equilibrated in phosphate-buffered saline buffer and according to their charge on analytical Propac-PA1 anion-exchange HPLC column (Thermo Scientific) equilibrated in HPLC grade water at pH 3.5 and eluted with a gradient of NaCl up to 2 M, pooled desalted and freeze-dried.

NMR spectroscopy

NMR experiments were recorded on an 850 MHz Bruker Avance III spectrometer equipped with a sample changer and a TCI 1.7 mm cryoprobe. Experiments were processed with Topspin3.2 and analyzed with ccpnmr. For assignment, 2.5 mg/mL samples of each oligosaccharide were prepared and its composition was assessed by 2D ^{13}C-edited ^{1}H–^{1}H total correlated spectroscopy (TOCSY) experiments. Oligosaccharides at 50 µM in 20 mM NaH_2PO_4, 150 mM NaCl and 0.5 mM ethylenediaminetetraacetic acid (EDTA), pH 6 buffer at 25°C were titrated with increasing protein quantities in the same buffer in 10 or 100% D_2O. ^{13}C–^{1}H heteronuclear single quantum coherence spectroscopy (HSQC) and ^{15}N–^{1}H band-selective optimized flip angle short transient (SOFAST) heteronuclear multi quantum coherence spectroscopy (HMQC) were recorded for monitoring of sugars and proteins, respectively. Combined heteronuclear chemical shifts are expressed in hertz according to the expression $\delta\text{Hz} = \sqrt{\left(\left(\delta^1\text{H}\right)^2 + \left(\delta X \times \frac{\gamma X}{\gamma 1\text{H}}\right)^2\right)}$, with X being either ^{13}C or ^{15}N and γ the gyromagnetic ratio. HCCONH, CCONH, HCCH-TOCSY, methyl–methyl specific nuclear overhauser effect spectroscopy (NOESY), ^{13}C-edited ^{1}H–^{1}H HSQC NOESY, ^{15}N-edited ^{1}H–^{1}H NOESY (150 ms mixing time) experiments used for side chain assignment and structure determination were recorded on 600 and 700 MHz spectrometers equipped with 5 mm TCI

cryoprobes on ^{13}C–^{15}N CXCL12γ at 1 mM in 20 mM NaH_2PO_4 pH 5.7 at 30°C.

NMR structure of CXCL12γ

CXCL12γ 3D structure was determined with Aria2 (Rieping et al. 2007) software using standard assignment and annealing protocols. The disordered C-terminal extension (residues 70–98) assignments were excluded from the calculation to avoid wrong NOESY peak assignments. The final 20 best structures refined in explicit water were deposited in the PDB under accession number 6EHZ and NMR resonances in the biological magnetic resonance bank (BMRB) (34176).

HADDOCK modeling of HS–protein complexes

Oligosaccharides structures are generated as already reported (Schanda et al. 2014) with the GLYCAN tool (haddock.science.uu.nl/enmr/services/GLYCANS/) with phi and psi angle of the glycosidic bonds according to values published for heparin (Khan et al. 2013) using crystallography and NMR system (CNS) (Brunger et al. 1998) and refined in water. The structure generated was submitted to a simulated annealing protocol using CNS, and the 10 best energy structures were retained for the docking protocol. The structure including residues 7–67 of CXCL12α (2NWG) refined in explicit water (Laguri et al. 2011) and the structure of CXCL12γ determined in this work PDB:6EHZ were used for docking. Residues defined as active were selected if CSP was higher than one standard deviation for all CSP in the case of sugars and higher than two times the standard deviation for proteins (Supplementary Table SII). Initial rigid body docking generated 10,000 structures, 2500 structures went through initial minimization and 200 structures for final minimization in explicit water. Structures were clustered with a 3 Å root-mean-square deviation (RMSD) cutoff and five structures minimum by cluster. Best HADDOCK energy cluster was retained and statistics presented in Supplementary Table SIII.

Supplementary data

Supplementary data for this article is available online at http://glycob.oxfordjournals.org/.

Acknowledgments

The authors thank S. Gil-Caballero and A. Favier for assistance with NMR, P. Gans for recording NMR experiments on CXCL12γ, and C. Ebel and A. Leroy for AUC experiments. This work used the platforms of the Grenoble Instruct Center (ISBG; UMS 3518 CNRS-CEA-UGA-EMBL) with support from FRISBI (ANR-10-INSB-05-02) and GRAL (ANR-10-LABX-49-01) within the Grenoble Partnership for Structural Biology (PSB).

Conflict of interest

The authors declare no conflict of interest.

Funding

Agence Nationale de la Recherche (grant number ANR-12-JSV5-0006-01).

Abbreviations

C5-epi, C5-epimerase; CSP, chemical shift perturbation; CXC, chemokine ligand 12; GlcA, glucuronic acid; GlcNAc, *N*-acetyl glucosamine; HPLC, high-performance liquid chromatography; HS, heparan sulfate; NMR, nuclear magnetic resonance, OST, *O*-sulfotransferase.

References

Boldajipour B, Doitsidou M, Tarbashevich K, Laguri C, Yu SR, Ries J, Dumstrei K, Thelen S, Dorries J, Messerschmidt E-M *et al*. 2011. Cxcl12 evolution - subfunctionalization of a ligand through altered interaction with the chemokine receptor. *Development*. 138:2909–2914.

Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS *et al*. 1998. Crystallography and NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr Sect D Biol Crystallogr*. 54: 905–921.

Connell BJ, Sadir R, Baleux F, Laguri C, Kleman JP, Luo L, Arenzana-Seisdedos F, Lortat-Jacob H. 2016. Heparan sulfate differentially controls CXCL12a- and CXCL12g-mediated cell migration through differential presentation to their receptor CXCR4. *Sci Signal*. 9:1–12.

Dominguez C, Boelens R, Bonvin AMJJ. 2003. HADDOCK: A protein-protein docking approach based on biochemical or biophysical information. *J Am Chem Soc*. 125:1731–1737.

Hsieh PH, Xu Y, Keire DA, Liu J. 2014. Chemoenzymatic synthesis and structural characterization of 2-O-sulfated glucuronic acid-containing heparan sulfate hexasaccharides. *Glycobiology*. 24: 681–692.

Khan S, Fung KW, Rodriguez E, Patel R, Gor J, Mulloy B, Perkins SJ. 2013. The solution structure of heparan sulfate differs from that of heparin: Implications for function. *J Biol Chem*. 288:27737–27751.

Kjellén L, Lindahl U. 2018. Specificity of glycosaminoglycan–protein interactions. *Curr Opin Struct Biol*. 50:101–108.

Kreuger J, Kjellén L. 2012. Heparan sulfate biosynthesis: Regulation and variability. *J Histochem Cytochem*. 60:898–907.

Laguri C, Sadir R, Rueda P, Baleux F, Gans P, Arenzana-Seisdedos F, Lortat-Jacob H. 2007. The novel CXCL12γ isoform encodes an unstructured cationic domain which regulates bioactivity and interaction with both glycosaminoglycans and CXCR4. *PLoS One*. 2:e1110.

Laguri C, Sapay N, Simorre JP, Brutscher B, Imberty A, Gans P, Lortat-Jacob H. 2011. 13C-labeled heparan sulfate analogue as a tool to study protein/heparan sulfate interactions by NMR spectroscopy: Application to the CXCL12α chemokine. *J Am Chem Soc*. 133: 9642–9645.

Lu W, Zong C, Chopra P, Pepi LE, Xu Y, Amster IJ, Liu J, Boons G-J. 2018. Controlled chemoenzymatic synthesis of heparan sulfate oligosaccharides. *Angew Chem Int Ed Engl*. 57:5340–5344.

Meneghetti MCZ, Hughes AJ, Rudd TR, Nader HB, Powell AK, Yates EA, Lima MA. 2015. Heparan sulfate and heparin interactions with proteins. *J R Soc Interface*. 12.

Préchoux A, Halimi C, Simorre JP, Lortat-Jacob H, Laguri C. 2015. C5-epimerase and 2-O sulfotransferase associate in vitro to generate contiguous epimerized and 2- O -sulfated heparan sulfate domains. *ACS Chem Biol*. 10:1064–1071.

Rieping W, Bardiaux B, Bernard A, Malliavin TE, Nilges M. 2007. ARIA2: Automated NOE assignment and data integration in NMR structure calculation. *Bioinformatics*. 23:381–382.

Rueda P, Balabanian K, Lagane B, Staropoli I, Chow K, Levoye A, Laguri C, Sadir R, Delaunay T, Izquierdo E *et al*. 2008. The CXCL12γ chemokine displays unprecedented structural and functional properties that make it a paradigm of chemoattractant proteins. *PLoS One*. 3:e2543.

Sadir R, Baleux F, Grosdidier A, Imberty A, Lortat-Jacob H. 2001. Characterization of the stromal cell-derived factor-1α-heparin complex. *J Biol Chem*. 276:8288–8296.

Sarrazin S, Lamanna WC, Esko JD. 2011. Heparan sulfate proteoglycans. *Cold Spring Harb Perspect Biol*. 3:1–33.

Schanda P, Triboulet S, Laguri C, Bougault CM, Ayala I, Callon M, Arthur M, Simorre JP. 2014. Atomic model of a cell-wall cross-linking enzyme in complex with an intact bacterial peptidoglycan. *J Am Chem Soc*. 136:17852–17860.

Veldkamp CT. 2005. The monomer-dimer equilibrium of stromal cell-derived factor-1 (CXCL 12) is altered by pH, phosphate, sulfate, and heparin. *Protein Sci*. 14:1071–1081.

Xue L-J, Mao X-B, Ren L-L, Chu X-Y. 2017. Inhibition of CXCL12/CXCR4 axis as a potential targeted therapy of advanced gastric carcinoma. *Cancer Med*. 6:1424–1436.