Sequence analysis and domain motifs in the porcine skin decorin glycosaminoglycan chain*

Xue Zhao,1,2 Bo Yang,2 Kemal Solakylidirim,2 Eun Ji Joo,2 Toshihiko Toida,3 Kyohei Higashi,3 Robert J. Linhardt,2,3,4,5 Lingyun Li2

1College of Food Science and Technology, Ocean University of China, Qingdao P. R. China 266003, 2Department of Chemistry and Chemical Biology, 3Department of Chemical and Biological Engineering, 4Department of Biomedical Engineering, Center for Biotechnology and Interdisciplinary Studies, Rensselaer Polytechnic Institute, Troy, NY 12180, USA, 5Graduate School of Pharmaceutical Sciences, Chiba University, 1-8-1 Inohana, Chuo-ku, Chiba 260-8675 Japan.

*Running title: Sequencing decorin glycosaminoglycan

To Whom Correspondence should be addressed: Lingyun Li, Department of Chemistry and Chemical Biology, Center for Biotechnology and Interdisciplinary Studies, Rensselaer Polytechnic Institute, Biotechnology Building 4211; Troy, NY, 12180, USA, Tel: (518) 276-2927; Fax: (518) 276-3405; Email: lil12@rpi.edu

Keywords: decorin, chondroitin sulfate, dermatan sulfate, glycosaminoglycan, sequence, motif, mass spectrometry, chondroitin lyase

Background: Decorin’s GAG is important but its structural motifs and sequences are unknown.

Results: FT-MS has mapped domains and a small GAG chain was sequenced.

Conclusions: GAG chain structure varies depending on origin but motif structure appears relatively consistent and may be comprised of a small number of sequences.

Significance: Understanding GAG structure may lead to understanding its biosynthesis and biological functions.

SUMMARY

Decorin proteoglycan is comprised of a core protein containing a single O-linked dermatan sulfate/chondroitin sulfate glycosaminoglycan (GAG) chain. While the sequence of the decorin core protein is determined by the gene encoding its structure, the structure of its GAG chain is determined in the Golgi. The recent application of modern mass spectrometry (MS) to bikunin, a far simpler chondroitin sulfate proteoglycans, suggests that it has a single or small number of defined sequences. On this basis, a similar approach to sequence porcine skin decorin’s much larger and more structurally complex dermatan sulfate/chondroitin sulfate GAG chain was undertaken. This approach resulted in information on the consistency/variability of its linkage region at the GAG chain’s reducing end, its iduronic acid-rich domain, glucuronic acid-rich domain, and non-reducing end. A general motif for the porcine skin decorin GAG chain was established. A single small decorin GAG chain was sequenced using MS/MS analysis. The data obtained in the study suggests that the decorin GAG chain has a small or a limited number of sequences.

Proteoglycans (PGs) are biomolecular glycoconjugates comprised of one or more glycosaminoglycan (GAG) chains with a molecular weights (MW) in the range 5,000 to 40,000, covalently attached to a protein core (1). PGs make up a large proportion of the extracellular matrix, performing a variety of roles in the excretory system, respiratory system, circulatory system, and skeletal system, and are involved in multisystem diseases of aging and cancer (2). The GAG chains of PGs are responsible for many of the biological functions of PGs, and subtle variations in the GAG structure can have pronounced effect on the organism physiology and pathophysiology (1,3). Recent studies in our laboratory have focused on the application of mass spectrometry (MS), in
combination with other analytical methods, to sequence (4) the chains of the small of biologically important small chondroitin sulfate (CS) PG, bikunin, (5,6) and small dermatan sulfate (DS) PG, decorin (7).

Decorin is one of the simplest cellular or pericellular matrix PGs belonging to the small leucine-rich proteoglycan (SLRP) family (8). Decorin consists of a protein core containing multiple leucine repeats with flanking cysteine-rich disulfide domains and a single chondroitin sulfate (CS) or dermatan sulfate (DS) GAG chain. The GAG attachment site in human decorin is located near its N-terminus (Ser34), and several N-linked oligosaccharide attachment sites are located on Asn211, Asn262, and Asn303 (9). Decorin has a high level of homology across mammalian species, for example porcine and human decorin are ~90% homologous. Decorin GAG varies from CS-type (glucuronic acid (GlcA)-rich) to DS-type (iduronic acid (IdoA) rich) depending on the tissue in which it is biosynthesized (2, 7, 10-12). In the endoplasmic reticulum xylose is added to the Ser residue of the core protein’s GAGylation site, followed by the sequential addition of two galactose residues and glucuronate residue in the cis-Golgi (13-19). The resulting tetrasaccharide linker is extended through the action of glycosyltransferases and modified as it transits the Golgi by 4-O-sulfotransferase (OST) (4 isoforms), 6-OST (2 isoforms), 2-OST (1 isoform), 4,6-OST (1 isoform) and C5-epimerase (2 isoforms), which complete the decorin GAG biosynthesis (13-20). The disaccharide composition of the resulting CS/DS GAG chain is often determined through exhaustive treatment with chondroitin lyases (21). This results in disaccharides comprised of ΔUA (4-deoxy-a-L-threo-hex-4-enopyranosyl uronic acid) 1,3-linked to N-acetylgalactosamine (GlcNAc) that have 8 distinct sulfation patterns, 0S, 2S, 4S, 6S, 2S4S, 2S6S, 4S6S and 2S4S6S (22).

Decorin is particularly rich in tissues such as skin (23, 24), tendons (25, 26), and cornea (27) where it plays a critical role in decorating or organizing collagen fibrils (28, 29). Both the core protein and GAG chain play critical but distinct roles in fibrillar organization (25, 28). The role of the GAG chain is somewhat less clear as they lie on the surface of the fibrils perpendicular to their long axis and may inhibit the lateral interaction of fibrils, resisting compression and regulating fibril diameter (23, 28). GAG stiffness, related to the structure and content of AC-/B-type domains (conformationally flexible IdoA residues (30) are present in B-type domains) may play a critical role in the biomechanical function of these tissues (23, 26, 28). Indeed, a genetic mutation of human galactosyltransferase-1, which reduces both the GAG chain occupancy in decorin and the IdoA content of decorin GAG, results in connective tissue disorders (24). Furthermore, C5-epimerase-1 deficient mice showed an altered distribution of IdoA, resulting in an altered collagen structure leading to skin fragility (19, 31).

In addition to its role in the proper assembly of collagen fibrils, fibrillogenesis (28,29), decorin plays a role in interactions with cytokines, and growth factors, such as tumor necrosis factor (TNF), and fibroblast growth factor (FGF)-2. Growth factor receptors, such as epidermal growth factor receptor (EGFR), insulin like growth factor receptor 1 (IGF-1R), and heparoyte growth factor receptor (Met) are also regulated by decorin (10, 31). Thus, decorin acts as a pan-tyrosine kinase receptor inhibitor (31) regulating cell growth, wound healing, angiogenesis, axon regeneration, and neural stem cell proliferation (10). Decorin acts as a “guardian for the matrix” in preventing tumor growth (32, 33). While much of decorin’s activity is associated with its core protein (32), its CS/DS chain certainly also contributes to decorin biology (2, 34). Rare (<10%) and more highly sulfated GAG domains containing disulfated and potentially even trisulfated CS/DS disaccharides (19, 35, 36) represent structural features important in protein-GAG interactions (2). Decorin GAG may also inhibit blood coagulation through its interaction with von Willebrand factor (37) and heparin cofactor II (8, 34, 38).

A detailed knowledge of the decorin PG structure is necessary to fully understand its structure-function relationship (39). While the sequence and the sites of post-translational modifications of decorin core protein have been determined (8) and its crystal structure has been solved (40), the GAG sequence analysis remains a challenge, in part due to the lack of sequencing tools capable of handling the tremendous structural complexity of this polysaccharide. Structural characterization of large complex polysaccharides, such as decorin GAG, relies on enzymology combined with an array of analytical techniques.

Recombinantly expressed decorins have been subjected to structural analysis (7, 13). CHO cell-expressed recombinant human decorin GAG
chains have several linkage region hexasaccharides (arising from the chain’s reducing end (RE)), the majority containing the sequence ΔUA β1-3GalNAc β1-4GlcA β1-3Gal β1-4Xyl/Xyl-ol, (where Gal is galactopyranose and Xyl/Xyl-ol are xylose and its NaBH₄ reduced form xylitol) 12% of which were unsulfated and 60% contained GalNAc4S residue (13). It’s GAG chain was comprised of 88% 4S, 8% 0S, and 4% 2S4S (13). Domain analysis of human decorin from cultured fibroblasts showed it to be rich in 4S and 0S disaccharide units with small amounts of 2S4S located at the chains non-reducing end (NRE) (35, 36, 41). Domain mapping of HEK cell-expressed recombinant human decorin PG was expressed as a reducing end (NRE) (35, 36, 41). Domain mapping strategy to a much more structurally complex target, porcine skin decorin, and the 4S6S in the AC-type domain near the NRE (7). The disaccharide composition of the GAG chains was 63% 4S, 23% 6S, 12% 0S, 1% 2S6S, 1% 2S4S and 0.4% 4S6S, with the 2S6S in the B-type domain and the 4S6S in the AC-type domain near the NRE (7).

In previous studies we performed domain mapping, on the structurally simpler bikunin PG with a single CS GAG chain comprised of a linkage region and ~16 disaccharides (0S and 4S and no IdoA), analyzing the composition of intact GAG chains using Fourier-transform (FT) MS analysis (5). Recently, FT-MS/MS has been used in a top-down approach to successfully determine the sequence of the intact bikunin CS GAG chains (6). In the current study, we have applied a similar domain mapping strategy to a much more structurally complex target, porcine skin decorin, and have used FT-MS and MS/MS to begin to elucidate the structure of its GAG chains.

**EXPERIMENTAL PROCEDURES**

Actinase E (EC 3.4.24.4) was from Kaken Biochemicals (Tokyo, Japan). Chondroitin sulfate A (CS-A) was purchased from Celsus Laboratories (Cincinnati, OH). Recombinant human decorin PG was expressed as a polyhistidine fusion protein in a stably transfected HEK (human embryonic kidney)-293-EBNA cell line as previously described (42). Endolytic chondroitinase ABC from Proteus vulgaris (EC 4.2.2.4), endolytic chondroitinase AC-1 from Flavobacterium heparinum (EC 4.2.2.5), exo- and no IdoA), analyzing the composition of intact GAG chains using Fourier-transform (FT) MS analysis (5). Recently, FT-MS/MS has been used in a top-down approach to successfully determine the sequence of the intact bikunin CS GAG chains (6). In the current study, we have applied a similar domain mapping strategy to a much more structurally complex target, porcine skin decorin, and have used FT-MS and MS/MS to begin to elucidate the structure of its GAG chains.

**EXPERIMENTAL PROCEDURES**

Actinase E (EC 3.4.24.4) was from Kaken Biochemicals (Tokyo, Japan). Chondroitin sulfate A (CS-A) was purchased from Celsus Laboratories (Cincinnati, OH). Recombinant human decorin PG was expressed as a polyhistidine fusion protein in a stably transfected HEK (human embryonic kidney)-293-EBNA cell line as previously described (42). Endolytic chondroitinase ABC from Proteus vulgaris (EC 4.2.2.4), endolytic chondroitinase AC-1 from Flavobacterium heparinum (EC 4.2.2.5), exo- and no IdoA), analyzing the composition of intact GAG chains using Fourier-transform (FT) MS analysis (5). Recently, FT-MS/MS has been used in a top-down approach to successfully determine the sequence of the intact bikunin CS GAG chains (6). In the current study, we have applied a similar domain mapping strategy to a much more structurally complex target, porcine skin decorin, and have used FT-MS and MS/MS to begin to elucidate the structure of its GAG chains.

**EXPERIMENTAL PROCEDURES**

Actinase E (EC 3.4.24.4) was from Kaken Biochemicals (Tokyo, Japan). Chondroitin sulfate A (CS-A) was purchased from Celsus Laboratories (Cincinnati, OH). Recombinant human decorin PG was expressed as a polyhistidine fusion protein in a stably transfected HEK (human embryonic kidney)-293-EBNA cell line as previously described (42). Endolytic chondroitinase ABC from Proteus vulgaris (EC 4.2.2.4), endolytic chondroitinase AC-1 from Flavobacterium heparinum (EC 4.2.2.5), exo-
was added and the flask swirled at 4°C for 18 h, with changes at 1 h and 3.5 h. The skin pieces were collected in a mesh and pressed to remove the salt solution before being placed in buffer (1 L of 1 M sodium citrate at pH 3.5) to which PMSF had been added as above. The flask was swirled at 4°C for 18 h, at which time the material was separated into solid and supernatant fractions by passing the preparation through a fine steel mesh. The solid residue was taken once again through a second and third extract. Supernatant collected above was put into 5 dialysis bags each containing almost 200 mL extract, and dialyzed against three changes of 5 L of water. The dialyzed extract was put through a 0.45 mm filter, set out in 200 mL aliquots in sterile containers some was set aside for immediate analysis by gel electrophoresis, while the remainder was freeze-dried and retained as the decorin PG fraction (Figure 1A).

Preparation of decorin GAG. Decorin PG fraction (~100 mg) was treated with actinase E (1 g) in 500 mL pH 8 sodium phosphate buffer at 55 °C for 10 h to afford decorin peptidoglycosaminoglycan. The GAG component of decorin was released by base-catalyzed β-elimination under reducing conditions. Decorin peptidoglycosaminoglycan was dissolved in a 0.2 M NaOH solution containing 1% NaBH₄. The reaction was allowed to proceed overnight at 4 °C and neutralized with 1 M hydrochloric acid. The remaining peptides were precipitated by addition of 20% perchloric acid. The supernatant obtained by centrifuge at 1500 g was collected and dialyzed against water for 16 h and freeze-dried, resulting in 50-60 mg of decorin GAG.

¹H-NMR spectroscopy. Decorin GAG (5 mg) was dissolved in 0.4 mL of [¹H] H₂O (99.96%); freeze-dried three times from [¹H] H₂O, and re-dissolved in 0.4 mL of [¹H] H₂O for one-dimensional NMR. NMR was recorded on a Bruker Avance II 600 MHz spectrometer equipped with a cryogenically cooled HCN probe with a z-axis gradient. All NMR experiments were recorded at 40 (313 K). One-dimensional ¹H-NMR spectra were recorded with 128 scans with a spectral width of 12 kHz and an acquisition time set to 2.7 s.

Size exclusion chromatography. Size exclusion chromatography (SEC) was performed on decorin GAG (20 μg) using TSK-GEL G4000PWx1 SEC column with a sample injection volume of 20 μL and a flow rate of 0.6 mL/min on an apparatus composed of a Shimadzu LC-10Ai pump, a Shimadzu CBM-20A controller and a Shimadzu RID-10A refractive index detector. The mobile phase consisted of 0.1 M NaNO₃. The column was maintained at 40 °C with an Eppendorf column heater during the chromatography. The SEC chromatograms were recorded with the LCsolution version 1.25 software and analyzed with its “GPC Postrun” function. For molecular weight determination, hyaluronic acid standards of different molecular weights (30.6 kDa, 43.8 kDa, 78.7 kDa and 130.2 kDa) were used as calibrants for the standard curve.

Disaccharide analysis of decorin GAG. Decorin GAG (20 μg) was dissolved in 1 μL 0.5 mM NH₄HCO₃ solution and exhaustively treated with 30-mU chondroitinase ABC and 30-mU chondroitinase AC-2 by digesting at 37°C for 10 h. The depolymerization reaction products were labeled by reductive amination with AMAC and subjected to disaccharides analysis by LC-MS (22).

Disaccharide analysis by RP-HPLC-MS. LC-MS analyses were performed on an Agilent 1200 LC/MSD instrument (Agilent Technologies, Inc. Wilmington, DE) equipped with a 6300 ion-trap. The column used was a Poroshell 150 C18 column (2.1 × 100 mm, EC-2.7 μm, Agilent, USA) at 45°C. For dual ammonium acetate and methanol gradient, eluent A was 80 mM ammonium acetate solution and eluent B was methanol. Solution A and 12% solution B was flowed (120 μL/min) through the column for 15 min followed by linear gradients 12-15% solution B from15 to 30 min, 15-30% solution B from 30 to 60 min and 30%-100% solution B from 60 to 62 min.

Preparation of decorin GAG domains. Decorin GAG (200 μg) was exhaustively treated three-times with 10 m-units of chondroitinase AC-1 (endolyase) in 50 mM ammonium bicarbonate buffer (pH 8) at 37°C for 10 h to obtain DS-type domains. Similarly, decorin GAG (200 μg) was exhaustively treated three-times with or 2.5 m-units of chondroitin ase AC-1 (endolyase) in 50 mM ammonium bicarbonate buffer (pH 8) at 37°C for 10 h to obtain AC-type domains. The B-type and AC-type domains were subjected to isocratic gel electrophoresis on a 15% acrylamide gel and disaccharides analysis by LC-MS. The B-type and AC-type domains were subjected to isocratic gel electrophoresis on a 15% acrylamide gel and disaccharides analysis by LC-MS. The B-type and AC-type domains were subjected to isocratic gel electrophoresis on a 15% acrylamide gel and disaccharides analysis by LC-MS. The B-type and AC-type domains were subjected to isocratic gel electrophoresis on a 15% acrylamide gel and disaccharides analysis by LC-MS.
solution and 5% T stacking gel monomer solution. CS-A partially digested by chondroitinase ABC was used as molecular markers. The mini-gels were subjected to electrophoresis at a constant 200 V for 30 min and visualized with 0.5% (w/v) alcian blue in 2% (v/v) aqueous acetic acid solution. MM analysis was performed with the aid of UNSCANIT software (Silk Scientific) using the logarithmic relationship between the GAG molecular weight and its migration distance. 

**Continuous elution PAGE fractionation of B-type domains.** B-type domains (4 mg) were loaded on 15% preparative polyacrylamide gel and fractionated by continuous-elution electrophoresis on a Model 491 Prep cell with a 28-mm internal diameter gel tube (BioRad, Hercules, CA) (45). Electrode running buffer was 1 M glycine and 0.2 M Tris (pH 9 without adjustment), and the lower chamber and elution buffer chamber were filled with resolving buffer. A peristaltic pump was set to 1 mL/min. The gel was subjected to electrophoresis at a constant power of 12 W for 6 h. The fraction was collected 1 mL each tubes. After freeze-dried, each tube was dissolved in 100 μL water and 5 μL was loaded on an analytical 15% gel with 5 μL of 50% sucrose for PAGE analysis. 

**HILIC LC-FT MS and MS/MS characterization of decorin domain structures.** LC-separation and FTMS analysis relied on similar conditions to those used for the analysis of intact chains of low molecular weight heparin (46,47). Briefly, a Luna HILIC column (2.0 × 150 mm, 200 Å, Phenomenex, Torrance, CA) was used to separate the partially or fully digested decorin GAG. Mobile phase A was 5 mM ammonium acetate prepared with HPLC grade water. Mobile B was 5 mM ammonium acetate prepared in 98% HPLC grade acetonitrile with 2% of HPLC grade water. After injection of 8.0 μL decorin GAG (1.0 μg/μL) through an Agilent 1200 auto-sampler, HPLC binary pump was used to deliver the gradient from 10% A to 35% A over 40 min at a flow rate of 150 μL/min. The LC column was directly connected online to the standard ESI source of LTQ-Orbitrap XL FT MS (Thermo Fisher Scientific, San-Jose, CA). The source parameters for FT-MS detection were optimized using Arixtra® to minimize the in-source fragmentation and sulfate loss and maximize the signal/noise in the negative-ion mode. The optimized parameters, used to prevent in-source fragmentation, included a spray voltage of 4.2 kV, a capillary voltage of −40 V, a tube lens voltage of −50 V, a capillary temperature of 275 °C, a sheath flow rate of 30, and an auxiliary gas flow rate of 6. External calibration of mass spectra routinely produced a mass accuracy of better than 3 ppm. All FT mass spectra were acquired at a resolution 60,000 with 400-2000 Da mass range. 

**Bioinformatics.** Charge deconvolution was either performed manually with electronic spreadsheets or auto-processed by DeconTools software (web source from PNNL at OMICS.PNL.GOV). Reducing end and non-reducing end structural assignments were performed by automatic processing using GlycReSoft 1.0 software developed at Boston University School of Medicine (http://code.google.com/p/glycresoft/downloads/list) (47). For automatic processing, GlycReSoft 1.0 parameters were set as: Minimum Abundance, 1.0; Minimum Number of Scans, 1; Molecular Weight Lower Boundary, 500 Da; Molecular Weight Upper Boundary, 10000 Da; Mass Shift, ammonium; Match Error (E ± M), 5.0 ppm; Grouping Error (E ± G), 80 ppm; Adduct Tolerance (E ± A), 5.0 ppm. For reducing end (including linkage), middle chain and non-reducing end oligosaccharide identification, theoretical database was generated by GlycReSoft 1.0 using the following parameters: A, ΔHexA = 0 (non-reducing end), 1 (middle chain) or linker (Xyl-Gal-Gal-GlcA); B, HexA = 0 to 12; C, HexNAc = A + B - 1 to A + B + 1; D, Ac = 0; E, SO3 = B to A + B+ (C × 2) + 1 - D; Modification, Adduct = ammonium from 0-5. 

**RESULTS**

From 500 gm of wet porcine skin ~100 mg of decorin proteoglycans of high purity was obtained (see Figure 1A). Proteolysis with actinase E afforded decorin peptidoglycan which was converted to 50-60 mg of decorin GAGs by reductive β-elimination. One-dimensional ^1^H-NMR analysis confirmed the purity the porcine skin decorin GAG to be >90% (see Figure 1C). The ^1^H-NMR spectrum of the decorin GAG was next used to determine the ratio of GlcA to IdoA. Integration of the anomeric (H1) signals of IdoA and GlcA were observed at 4.95 and 4.55 ppm, respectively, afforded a 3:1 ratio of IdoA:GlcA. 

Next, the molecular weight properties of porcine skin decorin GAG was analyzed using SEC (Figure 1B). The values for decorin GAG weight average (Mn = 22.3 kDa) and number average (Mn = 20.9 kDa) molecular weights were
derived by extrapolating the linear standard curve obtained using the available hyaluronan standards log (molecular weight) as a function of elution time (using \( y = 0.252 x + 7.66, r^2 = 0.999 \)). This corresponds to an average chain length of 46 disaccharides (based on the one-sulfate/disaccharide repeating unit of 459 Da) and a range of chain sizes from 16 disaccharides (7.1 kDa) to 98 disaccharides (44.8 kDa) with dispersity (Mn/Mw = 1.07).

ESI-FT-MS was next used to examine intact decorin GAG chains using the same approach that had been successful for the analysis of the less highly sulfated and much smaller bikunin CS GAG chains (6). GAG chains were first fractionated by preparative continuous elution PAGE (data not shown) and then analyzed by ESI-FT-MS. This approach had limited success providing molecular weight data on the smaller (dp<42) decorin GAG chains. All of the intact chains analyzed by ESI-FT-MS contained a mass signature consistent with the presence of linkage region terminated with xylitol at the reducing end and most contained an odd number of saccharide units consistent with a GalNAc residue at the chain’s NRE.

Disaccharide analysis was next performed on porcine skin decorin GAG following exhaustive depolymerization with chondroitinase ABC and AC-2 using LC-MS analysis (Figure S1 and Table S2). The disaccharide composition of porcine skin decorin GAG showed the expected major was 4S (87.3%) with additional minor disaccharides including: 6S (5.5%), 0S (5%) and 2S4S (2.2%).

Domain mapping with a partial digestion approach was next undertaken by treating porcine skin decorin GAG with chondroitinase ABC (Figure S2), by chondroitinase AC-1 (cutting AC-domains) to afford B-type domains (Figure S3) or by chondroitinase B (cutting B-domains) to afford AC-type domains (Figure S4). 1H-NMR had shown a 3:1 ratio IdoA:GlcA (see Figure 1C), corresponding to a 3:1 ratio of B-type:AC-type domains. Domain mapping involving exhaustive digestion by chondroitinase B or by chondroitinase AC-1, followed by PAGE analysis, was performed on porcine skin decorin (Figure 2 inset, lanes 2 and 3) to visualize the AC-type and B-type domains, respectively. In contrast, domain mapping of recombinant human decorin, prepared in HEK cells and previously studied in our laboratory (7), shows quite different distributions of AC-type and B-type domains (lanes 4 and 5 in Figure 2 inset). Quantification of the AC-type and B-type domain was performed by integrating the gel band intensities in Figure 2 and plotting as weight and mole percentages as a function of chain size in dp (Figure 2A & B). Disaccharide analysis of AC-type domains was mainly composed of 4S (89.8%) with minor amounts of 2S4S (10.2%) and the B-type domain was composed only of 4S (79.9%) with minor amounts of 6S (12.1%) and 0S (8%) (Table S2). The B-domains were fractionated by preparative continuous PAGE and domains ranging from dp6 to dp40 were collected.

HILIC LS-MS was next applied to analyze the domain structure present at the RE of decorin GAG chains. Digestion with chondroitinase ABC (see Figure S2 for digestion kinetics) converted the AC-type and B-type domains into disaccharide and tetrasaccharide products and afforded intact linkage region as hexasaccharides (Figure 3A). Digestion with chondroitinase B (see Figure S4 for digestion kinetics) converted the B-type domain primarily into disaccharide products and afforded intact AC-type domains of dp6-dp22 and linkage regions distributed through these fractions (Figure 3B). Digestion with chondroitinase AC1 (see Figure S3 for digestion kinetics) converted the AC-type domain into disaccharide products (see Table S2) and afforded intact B-type domains of dp6-dp40 and dp4 linkage regions (Figure 3C).

Next, our attention turned to understanding the structural diversity of the linkage regions prepared from porcine skin decorin. MS analysis of porcine skin decorin GAG exhaustively digested with chondroitinase ABC showed multiple hexasaccharides containing linkage region (Figure 3A, shaded peaks). Using HILIC, five hexasaccharides were obtained and separated by HILIC chromatography. Their composition was determined from their accurate masses observed in FT MS analysis (mass accuracy within 5ppm) (Figure 4A). Sequencing by MS/MS showed that these hexasaccharides each contained a linkage region tetrasaccharide at their RE coupled to a near-linkage region disaccharide (Figures 4C and Figures S5-7). The three most abundant hexasaccharides (linker-1, -2 and -3), representing ~90% of the linkage region, contained xylitol residues (corresponding to Xyl) at their RE. The two minor hexasaccharides (linker-4 and -5) had a xylitol-P residues (corresponding to Xyl-2-phosphate) at their RE (Figure 4).

The data presented on the oligosaccharide distribution presented in Figures 3 and 4 are based on total ion chromatogram (TIC) and are only...
semi-quantitative because of differences in ionization efficiencies of oligosaccharides of different size and charge. We next normalized the integrated values for ultraviolet (UV) detected peaks (at 232 nm) to TIC peaks to obtain a more quantitative interpretation of the structure of porcine skin decorin GAG. Oligosaccharides from dp4 to dp20, prepared by partial chondroitinase AC-1 treatment of CS-A, were fractionated by BioGel P-10 column with UV detection and used as standards. In this way HILIC-LC-MS TIC peak areas could be correlated to their UV peak areas and their corresponding to molar concentrations (Figure S8). These correction factors were then used to quantify oligosaccharides measured in the TIC. Oligosaccharides from dp2-12 gave comparable TIC response, while higher oligosaccharides, dp14-18 and above, required increasingly larger correction factors (Figure S9). Using these correction factors the relative abundance could be estimated of oligosaccharides arising from the RE (containing the linkage region) and the NRE (containing a terminal N-acetylgalactosamine or saturated uronic acid residue) (Figure 5). In Figure 5A the relative abundance of the different reducing ends clearly demonstrates that chondroitinase AC-1 generates primarily an unsulfated tetrasaccharide (dp4 + 0 sulfo groups) and chondroitinase ABC generates primarily three hexasaccharides (dp6 + 0, 1, or 2 sulfo groups) and chondroitinase B generates a range of larger oligosaccharides (dp6 to dp20) comprised of an unsulfated tetrasaccharide linkage region followed by 1 to 8 monosulfated glucuronic acid-containing disaccharides. Oligosaccharides arising from the NRE of the porcine skin decorin GAG chain by treatment with the same chondroitinases were quantified using this method (Figure 5B). Chondroitinase ABC releases only odd oligosaccharides, primarily a trisaccharide (with 1, 2 or 3 sulfo groups) and to a lesser degree pentasaccharides (with 2, 3, 4 or 5 sulfo groups). Chondroitinase B releases primarily tetrasaccharide (with 2 sulfo groups), or pentasaccharide (with 2, 3 or 4 sulfo groups) and much lower amounts of dp5, dp7, dp9 dp11 and dp13 oligosaccharides. Chondroitinase AC1 releases nearly uniform amounts of both odd and even oligosaccharides ranging in size from dp3 to dp13. In summary, all three enzymes demonstrated a preference for releasing of odd chains (ending in GalNAc) that were especially rich in B-type domains containing disulfated disaccharide units. A summary of these quantitative data can be found in Table 1.

Analysis of the internal domains was next undertaken. The AC-type domains account for ~25% of the chains by 3H-NMR and ~32% of the chains by integration of the scanned PAGE data (Figure 2B and inset). Based on HILIC LC-MS data on dp6-dp12 AC-type oligosaccharides, only about ~30% of the AC-type domains are internal with ~70% of these domains containing the linkage region at the chains reducing end (Figure S10). Thus, only ~10% of the internal chain domains are AC-type domains. These AC-type domains are relatively small in length dp4-dp20. In contrast, the B-type domains are usually found in the middle of the chain and account for ~90% of the internal chain domain. These domains can be quite large and range from dp6 to dp40. The B-type domains consist almost entirely of di4S units as judged by FTMS (Figure S11) and from disaccharide analysis (Table S2).

Finally, we selected a single small decorin GAG chain, with a molecular weight of 6903.130 Da, prepared by preparative gel electrophoresis (45) for sequencing by ESI-FT-MS/MS. The parent ion [M–11H]+, m/z=625.547 Da, selected for MS/MS analysis showed a single sequence consistent with the structure shown in Figure 6.

**DISCUSSION**

Skin is known to be one of the tissues with the greatest amount of decorin. Decorin GAG was prepared from porcine skin in multi-milligram amounts from by scaling up and modifying a procedure used to isolate small amounts of human skin decorin (28). NMR analysis of decorin GAG chains confirmed that was pure and consisted primarily dermatan sulfate B-type chains with an IdoA:GlcA ratio of 3:1. This is in contrast to recombinant human decorin GAG chains expressed in CHO cells (11) and HEK cells (7) that are much richer in CS-type chain structures with IdoA:GlcA ratios of 1.5:1 and 1:5, respectively (Table 1). Moreover, MS analysis of the smaller chains clearly demonstrated that most of the chains had been isolated intact, each with a reduced xylose residue at their RE (Table S1). The porcine skin decorin had an Mw = 22.3 kDa corresponding to an average chain length of 46 disaccharides (ranging of chain sizes from 16 to 98 disaccharides). Thus, the porcine skin decorin GAG chains were slightly smaller than the recombinant human decorin with an average chain length of 50 disaccharides (Table 1). The overall
Sequencing decorin glycosaminoglycan
disaccharide composition of the porcine skin decorin GAG chains was somewhat less complex than the HEK cell human decorin as it lacked the minor 2S4S, 4S6S disaccharides and was relatively rich in di4S (87.3%) (Table 1). These data made porcine skin decorin GAG a good candidate for domain mapping and potentially for GAG chain sequencing.

Domain mapping relied on the exhaustive treatment of GAG chains with three endolytic lyases (43, 44), chondroitinase ABC (affording the RE and NRE chain ends), chondroitinase AC-1 (affording B-type domains) and chondroitinase B (affording AC-type domains). Previous studies on HEK cell recombinant human decorin suggested that the easiest domains to characterize are the linkage region (GlcA-Gal-Gal-Xyl/Xylol (after NaBH$_4$ reduction) at the RE and the chain NRE (either a saturated uronic acid in a chain containing an even number of saccharide units or hexosamine residue in a chain with an odd number of saccharide units), as these domains both possess mass signatures. Exhaustive chondroitinase ABC treatment shows that there is some RE heterogeneity as evidenced by five different linkage region-containing hexasaccharides containing three unique linkage region tetrasaccharides (Figure 3). Three major linkage region/near linkage region hexasaccharides vary on their number (0, 1, or 2) of sulfo groups and the two minor linkage regions contain a unique Xylol2P residue (14). The RE domain was further extended by treating chains with chondroitinase AC and B selective for B-type and AC-type domains, respectively (Figure 5A). Selective chondroitinase treatments demonstrated that there were two types of domains immediately adjacent to the linkage region, a short (0-4 residues) AC-type domains or long (avg. of 23 residues) B-type domains (Figure 5B). NRE chain analysis using a similar approach demonstrated that most of the porcine decorin GAG chains contained an odd number of saccharide units (Table 1) and chains were terminated in N-acetylgalactosamine that were especially rich in 4S6S units and demonstrated the presence of both AC-type and B-type domains close to the NRE. The internal portion of the chain primarily consists of a large B-type domain of repeating di4S composition. A general motif could be constructed for each chain size and is shown in Figure 7. In a short chain, such as the one sequenced using FT-MS/MS (Figure 6) the chain is consistent with the motif as read from the RE with a common with linkage region domain with a Gal4S but no 2-phospho group on xylose, a near linkage region 4S disaccharide, an short RE domain, an internal IdoA-rich domain of 4S disaccharides that abruptly capped with an NRE terminal GalNAc4S, giving it an odd number of saccharide units. As chains become larger, NRE domains are added containing GlcA-rich 4S and 6S repeats as well as IdoA-rich 2,4S disaccharides. These larger chains would also contain longer internal domains of IdoA-rich 4S or 0S. The sequenced chain shown in Figure 6 contains an odd number of saccharide residues, a common disulfated linkage region with no phosphate, a short AC-type domain adjacent to the linkage region, and a continuous B-type domain of repeating di4S composition terminated at the NRE with GalNAc4S. We anticipate that as longer decorin GAG chains are sequenced we will begin to see the appearance of additional AC-type domains at the NRE and IdoA-rich 4S6S disulfated disaccharides. Based on our successful sequencing of bikunin we believe that there is an excellent chance that decorin contains an unique or a small number of sequence variants that have a well defined RE and extended NRE repeating domains of variable lengths. These data continue to suggest that GAGs have a deterministic single or small number of sequences.

REFERENCES
Sequencing decorin glycosaminoglycan


Sequencing decorin glycosaminoglycan


**FOOTNOTES**

* This work was supported by grants from the National Institutes of Heath GM38060, HL096972, ES 020903 to RJL and by a grant from the China Scholarship Council (XZ).

**FIGURE LEGENDS**

Figure 1. Intact decorin characterization. A. Analysis of porcine skin decorin by SDS PAGE. Gel stained with alcin blue. B. SEC analysis of porcine skin decorin GAG. The column was calibrated using hyaluronan oligosaccharide standards of defined molecular weight. C. 1H-NMR analysis of porcine skin decorin GAG performed in D2O at 800 MHz. The ratio of the IdoA-H1 and GlcA-H1 were calculated as 3:1.

Figure 2. Domain mapping of decorin. Insert: PAGE (15% acrylamide) analysis of chondroitinase-treated decorin GAGs. Lane 1, chondroitin sulfate A standard partially (~30% completion) treated with chondroitinase AC-1 with oligosaccharide size dp6 through dp38 labeled on the left; Lane 2, AC-type domains prepared by chondroitinase B-treatment of porcine skin decorin GAG. Lane 3, B-type domains prepared by chondroitinase AC-treatment of porcine skin decorin GAG. Lane 4, AC-type domains prepared by chondroitinase B-treatment of HEK cell decorin GAG; Lane 5, B-type domains prepared by chondroitinase ACI-treatment of HEK cell decorin GAG. The PAGE-gel analysis of porcine skin decorin (lanes 2 and 3) was next digitized by UN-SCAN-IT and the oligosaccharide composition calculated. A. The wt % composition of porcine skin decorin GAG is shown with the AC1-type domains (open bars) and the B-type domains (filled bars). B. The mol % composition of porcine skin decorin GAG is shown with the AC1-type domains (open bars) and the B-type domains (filled bars).

Figure 3. TIC of HILIC LC-MS analysis of porcine skin decorin GAG domains (shaded area are oligosaccharides from linkage domains). A. chondroitinase ABC treatment to afford AC-type and B-type domains products (see Table S2) and intact linkage regions; B. chondroitinase B treatment to afford B-type domain products (see Table S2) and intact AC-type domains and linkage regions; and C. chondroitinase AC-1 treatment to afford AC-type domain products (see Table S2) and intact B-type domains and linkage regions.

Figure 4. HILIC LC-MS analysis of porcine skin decorin linkage region. A. EIC (extracted ion chromatography, mass accuracy within 5 ppm) showing HILIC separation, integration and structures of the 5 linkage regions. Accurate mass can distinguish phosphorylation (monoisotopic mass: 79.96633 Da) from sulfation(monoisotopic mass:: 79.95682 Da). B. Relative abundance of the 5 linkage regions. C. FT MS/MS fragments example of linker-4 to confirm the phosphate location.

Figure 5. Normalized HILIC LC-MS data for the A. reducing ends and B. non-reducing ends of the porcine skin decorin GAG chains (inserts indicate the major structure motif at the reducing end and non-reducing end). Blue bars correspond to chondroitinase ABC treated, green bars correspond to chondroitinase AC-1 treated, and red bars correspond to chondroitinase B treated.

Figure 6. FT-MS/MS analysis of a short decorin GAG chain (dp31+14S, M = 6903.13, m/z = 626.547 Da).

Figure 7. Proposed structures of porcine skin decorin GAG chains.

**TABLE LEGENDS**

**TABLE 1.** Comparison of GAG chains from different decorin samples. nr, not reported
Figure 1

A
Lane A, decorin  
Lane B, decorin + chondroitin ABC  
Lane C, chondroitin ABC

B

C

Figure 2

A

B

Normalized percentage (by weight)

Normalized percentage (by Mol)
Figure 3
Sequencing decorin glycosaminoglycan

Figure 4

A

Linker-1
RT: 29.78

Linker-2
RT: 30.44

Linker-3
RT: 31.57

Linker-4
RT: 32.97

Linker-5
RT: 33.83

Relative Abundance

Time (min)

B

Relative Abundance

Linker

C

Y5

Y4

Y3

Y2

B2

B3

M-P

M-P-S

Relative Abundance

m/z
Figure 5

A

B
Figure 6
Figure 7

Sequencing decorin glycosaminoglycan
## Table 1

<table>
<thead>
<tr>
<th>Property</th>
<th>Porcine skin</th>
<th>HEK human recombinant (7)</th>
<th>CHO human recombinant (11)</th>
<th>Human fibroblast (31-33, 41,44)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Molecular mass</td>
<td>Mn 20.9 kDa</td>
<td>Mn 22 kDa</td>
<td>nr</td>
<td>nr</td>
</tr>
<tr>
<td></td>
<td>Mw 22.3 kDa</td>
<td>Mw 30 kDa</td>
<td></td>
<td></td>
</tr>
<tr>
<td>dp-high, low, avg</td>
<td>296, 32, 92</td>
<td>nr, nr, 100</td>
<td>nr</td>
<td>nr</td>
</tr>
<tr>
<td>IdoA:GlcA</td>
<td>3:1</td>
<td>1:5</td>
<td>nr</td>
<td>1.5:1</td>
</tr>
<tr>
<td>Sulfation level</td>
<td>0.98 S/di</td>
<td>0.9 S/di</td>
<td>nr</td>
<td>nr</td>
</tr>
<tr>
<td>Disaccharide composition</td>
<td>88% 4S, 5.5% 6S, 5% 0S, 2.2% 2S4S</td>
<td>63% 4S, 23% 6S, 12% 0S, 1% 2S6S, 1% 2S4S, 0.4% 4S6S</td>
<td>88% 4S, 8% 0S, 4% 2S4S</td>
<td>nr</td>
</tr>
<tr>
<td>Odd/even chains</td>
<td>4:1</td>
<td>nr</td>
<td>nr</td>
<td>nr</td>
</tr>
<tr>
<td>Motifs (avg.)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linkage regions</td>
<td>three tetras, major w Gal, minor w Gal4S, very minor w Xyl2P</td>
<td>nr</td>
<td>two tetras, major w Gal, minor with Gal4S</td>
<td>nr</td>
</tr>
<tr>
<td>Near linkage region</td>
<td>one 4S or 0S AC-type disaccharide</td>
<td>nr</td>
<td>one 4S, 0S or 6S AC-type disaccharide</td>
<td>nr</td>
</tr>
<tr>
<td>RE domain</td>
<td>two 4S AC-type disaccharides</td>
<td>seven 4S disaccharides AC-type, 4S6S possible</td>
<td>nr</td>
<td>AC-type disaccharides</td>
</tr>
<tr>
<td>Internal domain</td>
<td>thirty 4S or 0S B-type disaccharides</td>
<td>twelve 4S or 0S B-type disaccharides</td>
<td>nr</td>
<td>B-type disaccharides in blocks of ≥4</td>
</tr>
<tr>
<td>NRE domain</td>
<td>five 4S, two 6S, one 2S4S AC-type disaccharides</td>
<td>twenty-two AC-type disaccharides, rich in 4S6S</td>
<td>nr</td>
<td>nr</td>
</tr>
</tbody>
</table>
Sequence analysis and domain motifs in the porcine skin decorin glycosaminoglycan chain.

Xue Zhao, Bo Yang, Kemal Solakyildirim, Eun Ji Joo, Toshihiko Toida, Kyohei Higashi, Robert J. Linhardt, and Lingyun Li

Dr. Solakyildirim’s last name was misspelled. The correct spelling is shown above.
Sequence analysis and domain motifs in the porcine skin decorin glycosaminoglycan chain
Xue Zhao, Bo Yang, Kemal Solakyildirim, Eun Ji Joo, Toshihiko Toida, Kyohei Higashi, Robert J. Linhardt and Lingyun Li

J. Biol. Chem. published online February 19, 2013

Access the most updated version of this article at doi: 10.1074/jbc.M112.437236

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

Supplemental material:
http://www.jbc.org/content/suppl/2013/02/21/M112.437236.DC1

This article cites 0 references, 0 of which can be accessed free at
http://www.jbc.org/content/early/2013/02/19/jbc.M112.437236.full.html#ref-list-1