JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
Article
Cite This: J. Am. Chem. Soc. 2017, 139, 16986-16995
pubs.acs.org/JACS

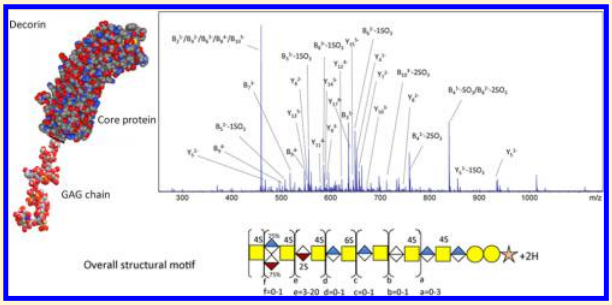
Sequencing the Dermatan Sulfate Chain of Decorin

Yanlei Yu,[†,‡,○] Jiana Duan,[#,○] Franklin E. Leach, III,[#] Toshihiko Toida,[▽] Kyohei Higashi,[▽] Hong Zhang,[†] Fuming Zhang,[‡] I. Jonathan Amster,[*,#] and Robert J. Linhardt[*,‡,§,∥,⊥]

[†]School of Food Science and Biotechnology, Zhejiang Gongshang University, Hangzhou, Zhejiang 310018, China
[‡]Department of Chemistry and Chemical Biology, Center for Biotechnology and Interdisciplinary Studies, [§]Department of Biology, [∥]Department of Biomedical Engineering, and [⊥]Department of Chemical and Biological Engineering, Rensselaer Polytechnic Institute, Troy, New York 12180, United States
[#]Department of Chemistry, University of Georgia, Athens, Georgia United States
[▽]Graduate School of Pharmaceutical Sciences, Chiba University, Chiba, Japan

Supporting Information

ABSTRACT: Glycomics represents one of the last frontiers and most challenging in omic analysis. Glycosylation occurs in the endoplasmic reticulum and the Golgi organelle and its control is neither well-understood nor predictable based on proteomic or genomic analysis. One of the most structurally complex classes of glycoconjugates is the proteoglycans (PGs) and their glycosaminoglycan (GAG) side chains. Previously, our laboratory solved the structure of the chondroitin sulfate chain of the bikunin PG. The current study examines the much more complex structure of the dermatan sulfate GAG chain of decorin PG. By utilizing sophisticated separation methods followed by compositional analysis, domain mapping, and tandem mass spectrometry coupled with analysis by a modified genetic algorithm approach, the structural motif for the decorin dermatan sulfate chain was determined. This represents the second example of a GAG with a prominent structural motif, suggesting that the structural variability of this class of glycoconjugates is somewhat simpler than had been expected.



INTRODUCTION

The recent and rapid progress in genomics and proteomics has not been matched in glycomics, corresponding to the largest portion of the metabolome.[1] One reason for the slow progress in establishing glycomes of various organisms is the complexities associated with the structural characterization and sequencing of complex carbohydrates.[2] Genomic sequencing relies on amplification methods that can also be applied to the proteomic sequencing as the result of the one gene to one protein paradigm.[3] Unfortunately, post-translational modification of the proteome in the case of glycosylation involves non-template-driven biosynthesis beginning in the endoplasmic reticulum (ER), continuing in the Golgi, and completing in catabolic remodeling outside the cell.[4] Thus, structural characterization and sequencing need to be carried out using conventional analytical chemistry relying primarily on separation and spectroscopy methods.

Proteoglycans (PGs) are among the most structurally complex glycoconjugates and are polydisperse, microheterogeneous mixtures having average molecular mass ranging from 25 to 2500 kDa.[5,6] These PG glycoconjugates are biosynthesized in three steps (Figure 1). The first is the template-driven synthesis of the core protein in the rough ER, the second is the installation of tetrasaccharide linkage regions, on specific serine residues of the core protein, and the third is transit through the Golgi and extension of the glycosaminoglycan (GAG) polysaccharide chains and the structural modification of their saccharide residues through epimerization and sulfation [Figure S1, Supporting Information (SI)].[7−9] The structural complexity of PGs are associated with (1) the occupancy of GAGylation sites; (2) the type of GAG chains, i.e., chondroitin sulfate, dermatan sulfate, heparan sulfate, etc.; (3) the length of each GAG chain, i.e., the degree of polymerization (dp) or number of saccharide residues; and (4) the structure or sequence of each individual GAG chain. The simplest 25 kDa PG, bikunin, has a single GAGylation site occupied by a ∼6 kDa chondroitin/chondroitin-4-sulfate GAG chain, of dp27−39, with a single well-defined sequence motif.[10] In contrast, one of the more complex PGs, aggrecan, has a molecular weight of 2500 kDa, has up to 160 GAGylation sites occupied by either ∼100 chondroitin/chondroitin-4 and/or chondroitin-6-sulfate GAG chains of ∼dp80, and ∼60 keratan/keratan sulfate (6-*O*-sulfo-galactose and/or 6-*O*-sulfo-*N*-acetylglucosamine) GAG chains of ∼dp40,[11] with still unknown sequence motifs.

Ongoing research in our laboratory has focused on the study of the structural glycomics of PGs. We began with the simplest PG, bikunin,[12] a serine protease inhibitor with an important

Received: September 27, 2017
Published: November 7, 2017

DOI: 10.1021/jacs.7b10164
J. Am. Chem. Soc. 2017, 139, 16986−16995

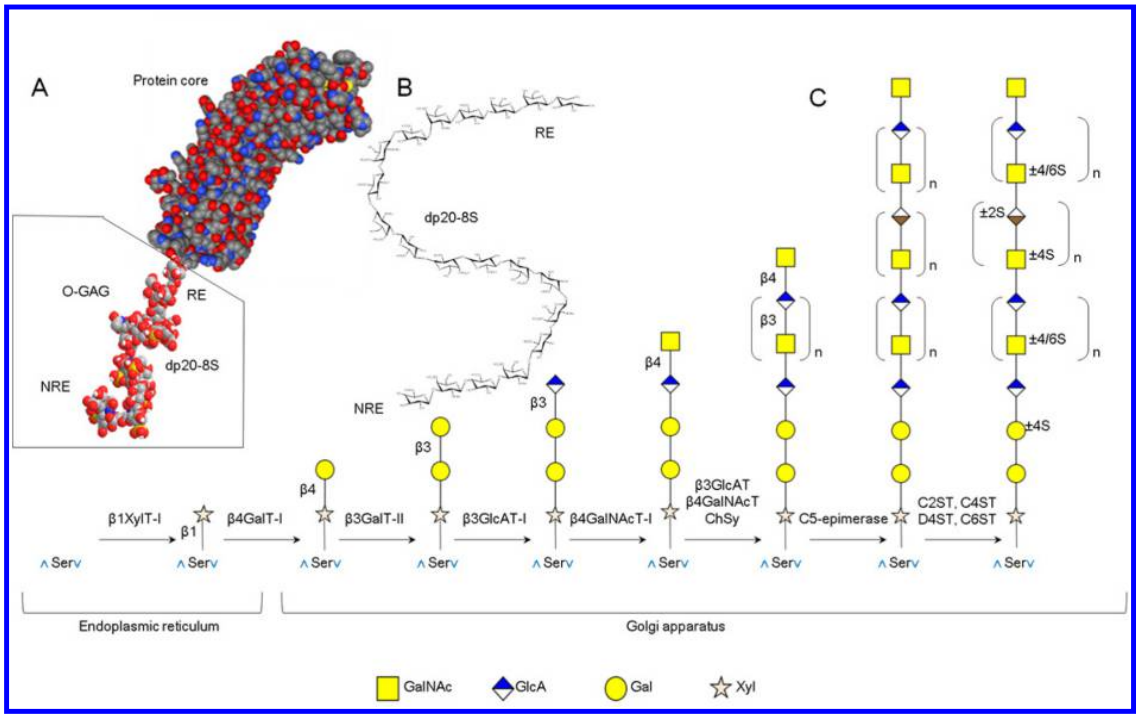


Figure 1. Modeled structure and biosynthetic pathway of decorin glycosaminoglycan. (A) Space-filling structure of decorin PG, with the decorin core protein from PDB (1XCD). Carbons (gray), hydrogens (white), oxygens (red), nitrogens (blue), and sulfurs (yellow) are shown. The O-linked GAG chain (dp20−8S) is shown with the reducing end (RE) and nonreducing end (NRE). (B) Chemical structure of GAG chain dp20−8S with a tetrasaccharide linkage region (GlcA-Gal-Gal-Xyl) at the RE. (C) Biosynthetic pathway for chondroitin sulfate/dermatan sulfate GAG. The GAG on a serine residue of the core protein, is synthesized in a pathway that begins in the endoplasmic reticulum and concludes in the Golgi apparatus. The biosynthetic enzymes are *β*1XylT-I, *β*-xylosyl transferase I; *β*4GalT-I, *β*-4-galactosyl transferase I; *β*3GalT-II, *β*-3-galactosyl transferase II, *β*3GlcAT-I, *β*-3-glucuronosyl transferase I; *β*4GalNAcT-I, *β*-4-*N*-acetyl galactosaminyl transferase I; *β*3GlcAT, *β*-3-glucuronosyl transferase; *β*4GalNAcT, *β*-4-*N*-acetyl galactosaminyl transferase; ChSy, chondroitin synthases; C5-epimerase; C2ST, 2-*O*-sulfotransferases, C4ST, chondroitin 4-*O*-sulfotransferases, D4ST, dermatan 4-*O*-sulfotransferases; and C6ST, 6-*O*-sulfotransferases.

role in inflammation,[13] potentially having ∼10^{11} GAG sequences, and demonstrated that it had a singular sequence motif (Figure S2, SI).[10] Over the past decade our laboratory[14,15] and others[16−19] have laid the biochemical groundwork to examine the structure and sequence of the next simplest PG, decorin. These studies included disaccharide compositional analysis, linkage region variability analysis, and domain mapping. The decorin GAG chain structure elucidated in these studies represents a composite average and does not necessarily correspond to an actual sequence nor does it provide any information on sequence variability (i.e., the number of potential sequences that are actually present). Decorin with its single GAGylation site, occupied by a 4-*O*- and/or 2-*O*- and/or 6-*O*-sulfo dermatan/chondroitin dp14−40 GAG chain (Figure S3, SI), potentially has ∼10^{18} GAG sequences (Table 1). This large number of sequence permutations is due to both the eight different disaccharide structures, comprised of 1→3-linked L-iduronic acid (IdoA) [with and without 2-*O*-sulfo (2S) groups] and D-glucuronic acid (GlcA) and 1→4-linked *N*-acetyl-D-galactosamine (GalNAc) (with/without 4S and/or 6S), and decorin's requisite polydispersity.[14,15] Thus, we anticipated that our efforts for the direct sequencing of the decorin GAG chain would be roughly a million times more difficult than our first successful sequencing of bikunin.[10]

Decorin is the simplest pericellular PG member belonging to the small leucine-rich proteoglycan (SLRP) family. It was named on the basis of its property of "decorating" collagen fibrils in the skin and tendons and its control of fibrillogenesis.[20−23] Decorin has been called the guardian of the extracellular matrix (ECM) resulting from its role as a pan-inhibitor of tyrosine kinase signaling, and in this role decorin displays a large ECM interactome of importance in controlling tumor growth, angiogenesis, and autophagy.[24] While many of decorin's biological activities are associated with its banana-shaped core protein, others, particularly its ability to modulate matrix maturation and its interaction with ECM enzymes, such as metalloproteinases and growth factors, require its GAG chain.[23] The GAG chains of endothelial decorin are also important participants in human infection by the tick born spirochete *Borrelia bergdorferi*, which causes Lyme disease, by serving as a receptor for its surface proteins, facilitating tissue colonization.[25,26] The flexible dermatan sulfate chain of decorin PG,[27] particularly the highly sulfated domains rich in iduronic acid, are important features for protein interaction associated with biological activity.[28]

The current study examines porcine skin decorin, available in multigram quantities, and undertakes its extensive fractionation to prepare a less heterogeneous mixture of decorin chains with an average level of sulfation but with a relatively small chain length (low dp), as determined by analytical polyacrylamide gel electrophoresis (PAGE). The disaccharide composition, linkage region structures, and domain structures of these chains were determined by mass spectrometry. The accurate mass measure-

DOI: 10.1021/jacs.7b10164
J. Am. Chem. Soc. 2017, 139, 16986−16995

Table 1. Showing the Number of Possible Permutations for a Decorin Glycan of a Specific Chain Length with Various Levels of Restrictions Set

		possible permutations			
dp[a]	disaccharides[b]	total possible[c]	limited number of SO_3[d]	disaccharide analysis restrictions[e]	tandem MS restriction[f]
14	7	2.10×10^6	2.77×10^4	4.12×10^2	2
15	7	8.39×10^6	1.43×10^5	1.85×10^3	26
16	8	1.68×10^7	1.88×10^5	2.11×10^3	27
17	8	6.71×10^7	9.54×10^5	4.46×10^3	29
18	9	1.34×10^8	1.26×10^6	5.04×10^3	30
19	9	5.37×10^8	6.33×10^6	2.16×10^4	61
20	10	1.07×10^9	8.48×10^6	2.50×10^4	63
21	10	4.29×10^9	4.19×10^7	5.33×10^4	67
22	11	8.59×10^9	5.67×10^7	6.12×10^4	69
23	11	3.44×10^{10}	2.78×10^8	2.54×10^5	6.38×10^2
24	12	6.87×10^{10}	3.79×10^8	2.96×10^5	6.74×10^2
25	12	2.75×10^{11}	1.84×10^9	6.35×10^5	7.49×10^2
26	13	5.50×10^{11}	2.52×10^9	7.37×10^5	7.88×10^2
27	13	2.20×10^{12}	1.22×10^{10}	2.98×10^6	1.61×10^3
28	14	4.40×10^{12}	1.68×10^{10}	3.51×10^6	1.69×10^3
29	14	1.76×10^{13}	8.07×10^{10}	7.56×10^6	1.86×10^3
30	15	3.52×10^{13}	1.12×10^{11}	8.84×10^6	1.95×10^3
31	15	1.41×10^{14}	5.35×10^{11}	3.51×10^7	1.64×10^4
32	16	2.81×10^{14}	7.45×10^{11}	4.16×10^7	1.75×10^4
33	16	1.13×10^{15}	3.54×10^{12}	8.98×10^7	1.98×10^4
34	17	2.25×10^{15}	4.96×10^{12}	1.06×10^8	2.10×10^4
35	17	9.01×10^{15}	2.35×10^{13}	4.14×10^8	4.32×10^4
36	18	1.80×10^{16}	3.30×10^{13}	4.93×10^8	4.58×10^4
37	18	7.21×10^{16}	1.56×10^{14}	1.07×10^9	5.13×10^4
38	19	1.44×10^{17}	2.20×10^{14}	1.26×10^9	5.42×10^4
39	19	5.76×10^{17}	1.03×10^{15}	4.90×10^9	4.29×10^5
40	20	1.15×10^{18}	1.46×10^{15}	5.85×10^9	4.60×10^5
	totals	1.98×10^{18}	2.94×10^{15}	1.43×10^{10}	1.17×10^6

[a]Degree of polymerization. [b]Number of disaccharides of specified dp. [c]Number of possible permutations for SO_3 modifications for a specified dp assuming all possible dermatan sulfate modifications (0S, 2S, 4S, 6S, 2S4S, 2S6S, 4S6S, 2S4S, and 2S4S6S). The total possible number of permutations was calculated as a summation of a binomial combination using n-choose-k $[n!/k!(n-k)!]$, where n is the total number of saccharide units and k is the number of modifications, with k ranging from 0 to 3 times the number of disaccharides. [d]The limited number of permutations based on composition assignments done with MS1 (FT-ICR-MS and Orbitrap-MS) at high mass accuracy. A limited number of SO_3 modifications are possible with respect to chain length. The number of possible SO_3 is reduced to the number of dissachardies/2 − 4 to number of dissachardies/2 − 2. The n-choose-k calculations are still used but with a significantly reduced range in k. [e]Disaccharide analysis restrictions limit the possible sulfo group modifications on any disaccharide to four possible combinations: 0S, 2S4S, 6S, 4S. [f]Tandem MS analysis restrictions, where MS^2 reveals the prevalence of a characteristic, multicharged peak that suggests a single sulfate modification per disaccharide from the nonreducing end that extends 3−10 disaccharide units (for dp14−40).

ment of GAG chains by Fourier transform mass spectrometry (FTMS) enabled the determination of polymerization and sulfation extent. Collisional dissociation tandem mass spectrometry was used to determine the pattern of sulfo groups through the decorin GAG chain and afforded a prominent sequence motif.

RESULTS AND DISCUSSION

Linkage Region. The decorin GAG chain is biosynthesized on serine 34 near the N-terminus of its core protein (Figure 1A). The linkage region tetrasaccharide of porcine skin decorin, →4) GlcA (1→3) galactose (Gal) (±4S) (1→3) Gal (1→4) xylose (Xyl) [±2 phospho (P)] (1→, assembled in the ER is variable with ∼30% of the chains contain a 4-S-Gal and ∼5% of the chains contain a 2-P-Xyl, (Figures S4 and S5, SI).[15] Chain extension on this linkage region tetrasaccharide occurs in the Golgi compartment, resulting in a linear polysaccharide chain having 14−40 saccharide residues (Figure 1B). C5-Epimerization, converting GlcA to IdoA, and the introduction of sulfo groups to the 4-, 6-, and 2-positions afford the mature decorin dermatan sulfate GAG chain (Figure 1C). Not all modification steps are complete, which results in considerable structural hetereogeneity. Moreover, enzyme selectivity and/or another unknown control factor(s) results in the formation of structural domains within the mature decorin dermatan sulfate GAG chains. The GAG chains are typically released from the core protein through β-elimination under reducing conditions so that each chain carries a xylitol (Xyt) at its reducing end (Figure 1B). The released GAG chains present such a complex mixture of polymer lengths, sulfation levels, and domains that it is not possible to apply current mass spectrometry methods to determine their sequence. A fractionation approach was designed to prepare a representative set of chains of relatively short chain length but with average sulfation density for sequence analysis (Figure 2). Ten kilograms of porcine skin (wet weight) afforded 2 g of decorin PG, 1 g of decorin GAG

DOI: 10.1021/jacs.7b10164
J. Am. Chem. Soc. 2017, 139, 16986−16995

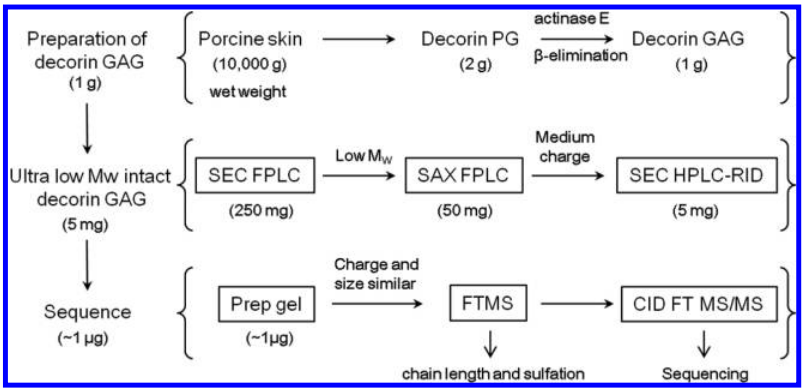


Figure 2. Decorin flowchart for solving the structure. From 10 000 g of wet porcine skin, 2 g of decorin proteoglycans of high purity was obtained. Proteolysis with actinase E afforded decorin peptidoglycan that was converted to 1 g of decorin GAGs by reductive β-elimination. Size exclusion chromatography was applied to obtain low molecular weight decorin, and strong anion exchange was applied to get medium charge fractions. SEC with a refractive index detector on HPLC was applied to online separate decorin fractions. Continued elution preparative PAGE was next done to get fractions of similar size and charge for FT and FT-MS/MS analysis.

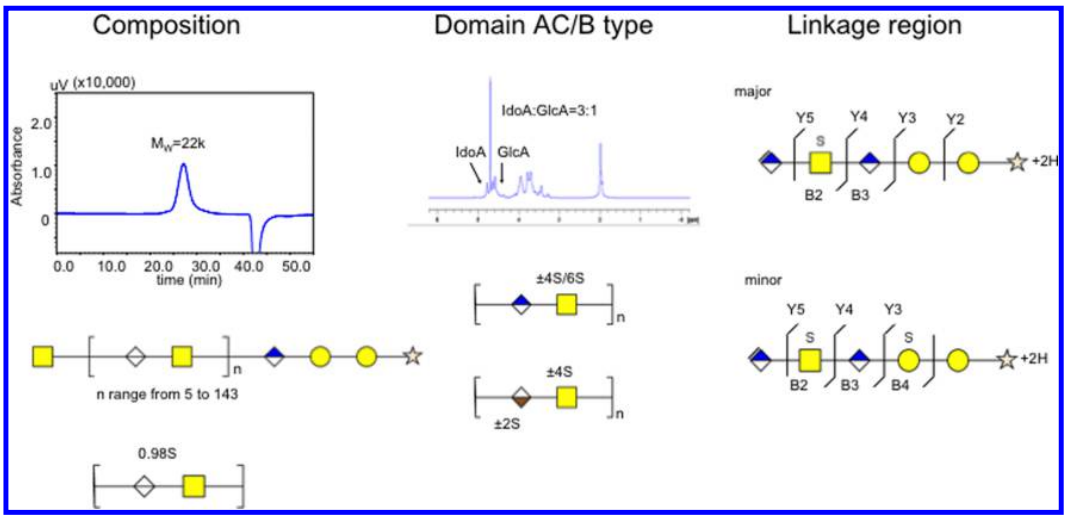
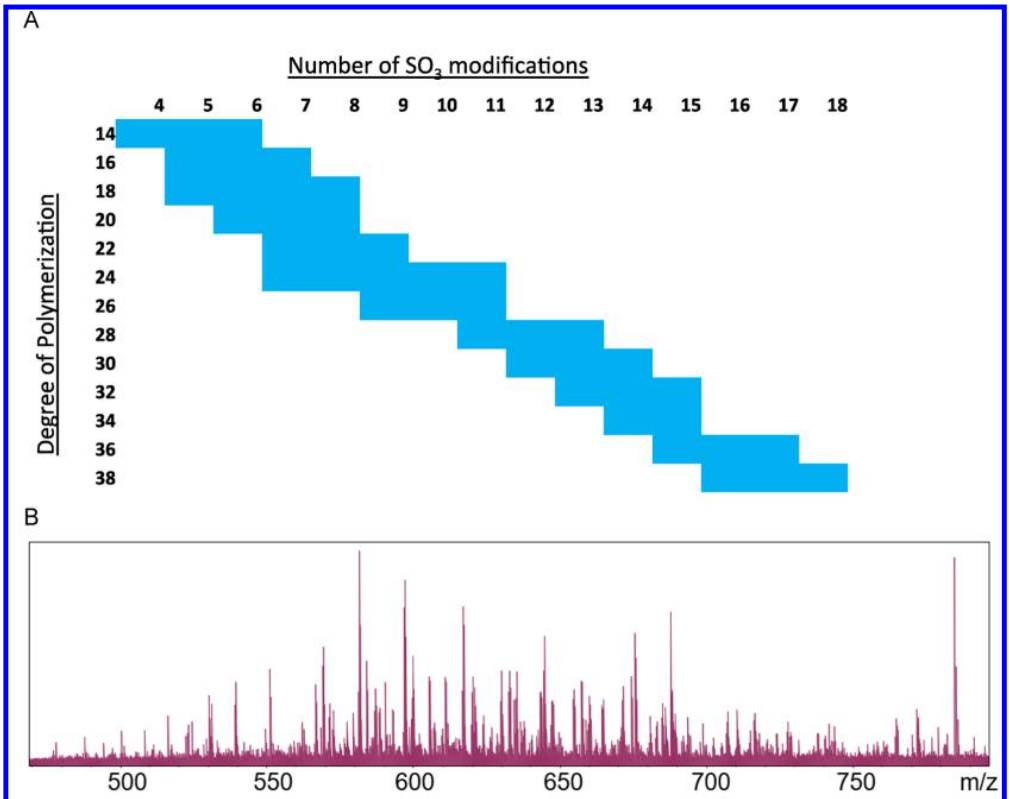


Figure 3. Decorin GAG compositional, domain mapping, and linkage analysis. Decorin GAG molecular weight was determined by GPC−HPLC. Disaccharides analysis showed that the average sulfation per disaccharide was 0.98 and the chain length ranged from dp14 to dp290, with the average being dp92. The ration of the IdoA-H1 and GlcA-H1 was calculated as 3:1 on the basis of ^{1}H NMR. Disaccharides analysis of AC-type and B-type domains revealed that 2S4S only exists on a B-type domain, and 6S and 0S only exist on an AC-type domain. Completed digestion of decorin GAG afforded the major linkage region GlcA-GalNAc4S-GlcA-Gal-Gal-Xyt and the minor linkage region GlcA-GalNAc4S-GlcA-Gal4S-Gal-Xyt.

chains, and ultimately 1 μg of size-uniform and charge-uniform GAG chains for sequencing.

Composition and Domain Mapping. The initial mixture of decorin GAG chains released from decorin PG had an average molecular mass of 22 kDa, a chain length ranging from dp14 to dp290 based on the lowest and highest molecular weight acquired from GPC, and an average sulfation density of 0.98 sulfo groups/disaccharide repeating unit (Figure 3). Disaccharide compositional analysis, comprising chondroitin lyase catalyzed depolymerization of the decorin GAG to unsaturated disaccharides (with a nonreducing terminal ΔUA, deoxy-β-L-threo-hex-4-enopyranosiduronic acid) and HPLC−MS analysis, afforded a composition of 96.8 mol % ΔUA(1→3)GalNAc4S, 2.0 mol % ΔUA2S(1→3)GalNAc4S, 1.0 mol % ΔUA(1→3)GalNAc6S, and 0.2 mol % ΔUA(1→3)GalNAc. Next, we examined the AC/B domain structure of the decorin GAG chain using NMR spectroscopy, which showed an IdoA/IdoA2S:GlcA ratio of 3:1 (Figure 3). The B domain contains IdoA or IdoA2S residues and is susceptible to treatment with endolytic chondroitin B lyase, while the AC domain is susceptible to treatment with endolytic chondroitin AC lyase (Figures S6−S8, SI).[29−31] Thus, exhaustive treatment of decorin GAG with chondroitin B lyase and chondroitin AC lyase and recovery of intact chains of reduced size affords AC and B domains, respectively. Disaccharide analysis shows that the AC domains exclusively contain ΔUA(1→3)GalNAc4S, ΔUA(1→3)GalNAc6S, and ΔUA(1→3)GalNAc at 93.2, 6.0, and 0.8 mol %, respectively, and the B domain exclusively contains only ΔUA(1→3)GalNAc4S at 95.3 mol % and ΔUA2S(1→3)GalNAc4S at 4.7 mol % (Figure S8, SI). This analysis is consistent with the presence of 2-*O*-sulfo groups present only on IdoA residues and only in B domains. As expected, the B domains were often >dp10 and were on average ~3 times longer than the AC domains, which were most frequently dp6 (Figure S7, SI). Linkage region analysis confirmed the presence of two major reduced structures from

DOI: 10.1021/jacs.7b10164
J. Am. Chem. Soc. 2017, 139, 16986−16995



Composition	M/Z	Z	ME	Composition	M/Z	Z	ME
dp10-3S	675.4622	3	-1.28	dp20-7S	706.4574	6	-1.83
dp11-4S	769.8082	3	-0.38	dp20-7S	605.3913	7	-1.41
dp12-3S	801.8335	3	-0.06	dp20-8S	616.8147	7	-1.42
dp12-4S	621.1122	4	-0.64	dp20-8S (1 Na)	619.9536	7	-1.91
dp13-4S	671.8813	4	-1.74	dp22-6S	659.4364	7	1.05
dp13-5S	553.2949	5	-1.80	dp22-8S	670.972	7	-1.93
dp14-4S	715.8899	4	-0.84	dp22-8S	586.975	8	-1.23
dp14-5S	588.5015	5	-1.31	dp22-8S (1Na)	674.1138	7	0.28
dp16-5S	664.3242	5	-0.59	dp22-9S	596.9692	8	-0.81
dp16-6S	566.7611	6	-1.64	dp22-9S (1Na)	599.7176	8	-0.76
dp16-6S	680.3156	5	-0.54	dp24-7S	713.7093	7	-0.57
dp18-5S	616.6203	6	-1.41	dp24-9S	572.651	9	-1.59
dp18-5S	740.1459	5	-1.31	dp24-9S	644.359	8	-0.38
dp18-6S	629.9466	6	-1.18	dp24-10S	581.5353	9	-1.63
dp18-7S	551.2323	7	-1.86	dp24-10S	654.3536	8	-1.58
dp18-7S (1Na)	646.9367	6	-0.68	dp26-9S	614.7744	9	-1.58
dp18-8S	562.6549	7	-1.49	dp26-10S	623.6583	9	-1.94
dp20-6S	693.1323	6	-0.36	dp26-10S	561.1931	10	0.59

Figure 4. Complexity of decorin GAG chain mixtures over a range of chain sizes. (A) Accurate mass measurement in the MS^1 using FT-ICR MS and Orbitrap-MS makes composition assignment possible. Compositions calculated from aggregated MS^1 data show a limit in the overall number of SO_3 modifications that can exist for a specific dp. (B) Example of composition assignments for decorin fraction 39 using FT-ICR MS. All compositions are within 2 ppm mass error.

DOI: 10.1021/jacs.7b10164
J. Am. Chem. Soc. 2017, 139, 16986−16995

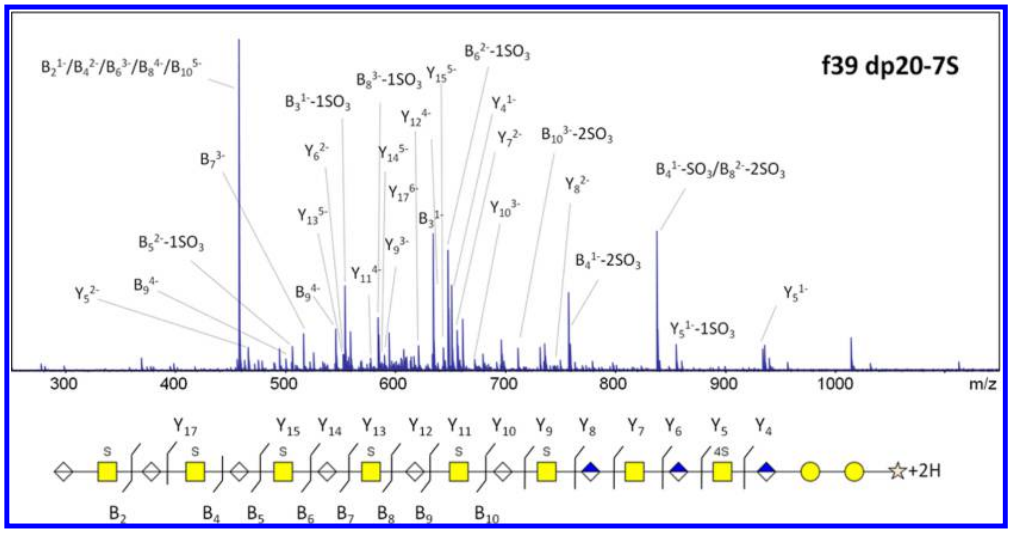


Figure 5. Tandem MS analysis of decorin GAG chain dp20−7S^{7-}. The structure was determined using in-house GAG algorithm software, and corresponding MS^2 cleavages are shown.

the linkage region domain, →4) GlcA (1→3) Gal (±4S) (1→3) Gal (1→4) Xyt, comprising >90 mol % of the decorin chains (Figure 3).

Simplifying Mixture Complexity by Fractionation. Our experience in MS sequencing the much less complex bikunin GAG chain suggested that while a sulfation level of 0.98 *O*-sulfo groups/disaccharide repeating unit of the decorin GAG was a tractable problem, the large chain length up to dp290 (Figure S3, SI) would challenge the limits of current MS capabilities. Moreover, the presence of more than 10−20 prominent molecular ions in a fraction makes the selection of one having the appropriate intensity for MS/MS analysis problematic. We set out to fractionate the decorin GAG chains to obtain relatively homogeneous fractions of chains of average sulfation level but of sizes <dp36 to generate a sample set suitable for mass spectrometry. Decorin GAG (1 g) of M_w (avg) 22 kDa (Figure 3) was first fractionated by size exclusion chromatography (SEC)−fast performance liquid chromatography (FPLC), and approximately the last third of the eluting sample, corresponding to M_w (avg) 19 kDa, was collected (Figure 2). Next, 250 mg of this fraction was applied to strong anion exchange (SAX)−FPLC, and the center third of the peak was collected and subjected to SEC−HPLC to again enrich 5 mg of the small chains (Figure S9, SI). Finally, preparative PAGE was applied to obtain 135 fractions that were analyzed by analytical PAGE (Figure S10, SI) of microgram quantities of fractions ranging in estimated size from 3 to 32 kDa (Table S1, SI). Orbitrap FTMS analysis was then undertaken on selected PAGE fractions ranging from 38 (M_w ∼ 4.7 kDa) to 60 (M_w ∼ 8.7 kDa) to determine the intact mass of fraction components and to assess the suitability of these and neighboring fractions for MS/MS sequencing (Figures S11−S18 and Tables S2−S9, SI). Accurate masses were obtained for 57 molecular ions ranging from dp20 to dp44 and carrying 8−20 sulfo groups with only a small amount of sodium adduction observed. The complexity of these fractions with 10−20 molecular ions detected in each and sufficient S/N suggested that these fractions would be suitable for sequencing by MS/MS.

Sequencing. FT-ICR MS was applied to neighboring PAGE fraction 39. Initial MS analysis provided molecular compositions that were then selected for sequencing by collisional dissociation. The dp range observed in fraction 39 (dp14−26) by FT-ICR MS (Figure 4A) was slightly wider than had been observed in fraction 38 (dp16−24) by Orbitrap FTMS (Figure S11, Table S2, SI). At the MS level, molecular ions corresponding to 25 unique compositions were detected and four were found suitable for further MS/MS analysis [Figure 4B and Figure S19−S20 and Table S10−S11 (SI)]. Ions were not selected for MS/MS analysis for several reasons, including (a) a lack of clean quadrupole mass selection, where more than one composition was isolated; (b) low intensity precursor ions that did not provide fragment ions above the limit of detection; or (c) depletion of the same fraction from multiple MS experiments. A sample annotated spectrum of decorin MS/MS is shown in Figure 5, where CID-FT-ICR−MS/MS of the molecular ion m/z 605.3919, corresponding to dp20−7S^{7-}, resulted in a spectrum rich in glycosidic bond cleavages suitable for sequence determination. Additional sequence analysis of other tandem MS spectra from fraction 39 are available in the Supporting Information. Complete composition and tandem MS analyses of fractions 35 and 51 are also provided (Figures S21−S32 and Tables S12−S23, SI).

Matching adjacent glycosidic fragments ($B_n + B_{n+1}$, $Y_n + Y_{n+1}$, etc.,) would differ by 176.0321 Da for hexuronic acid and 203.0793 Da for *N*-acetylgalactosamine, as observed previously in bikunin glycan analysis.[10] Nonmatching adjacent glycosidic fragments (e.g., $B_n + C_{n+1}$ or $C_n + B_{n+1}$, $Y_n + Z_{n+1}$, etc.) differ by an additional ±18.0106 Da. Sulfo group modifications were assigned to sugar residues based on the addition of 79.9568 Da between adjacent sets of glycosidic fragments. On the basis of disaccharide analysis, sulfo modifications exist as only 2-*O*-sulfo on hexuronic acids but as 4-*O*- or 6-*O*-sulfo groups on *N*-acetylgalactosamine. A 4-O-modification was determined for the reducing end linker region (GlcA-GalNAc4S-GlcA-Gal-Gal-Xyt), but additional differentiation of 4-*O*- and 6-*O*-sulfo groups depended on the presence of diagnostic cross-ring fragmentations. Ion activation by collision-induced dissociation (CID) was capable of breaking single bonds between residues, providing glycosidic fragments, but lacked sufficient energy to break multiple bonds across the hexose sugar residue, yielding

DOI: 10.1021/jacs.7b10164
J. Am. Chem. Soc. 2017, 139, 16986−16995

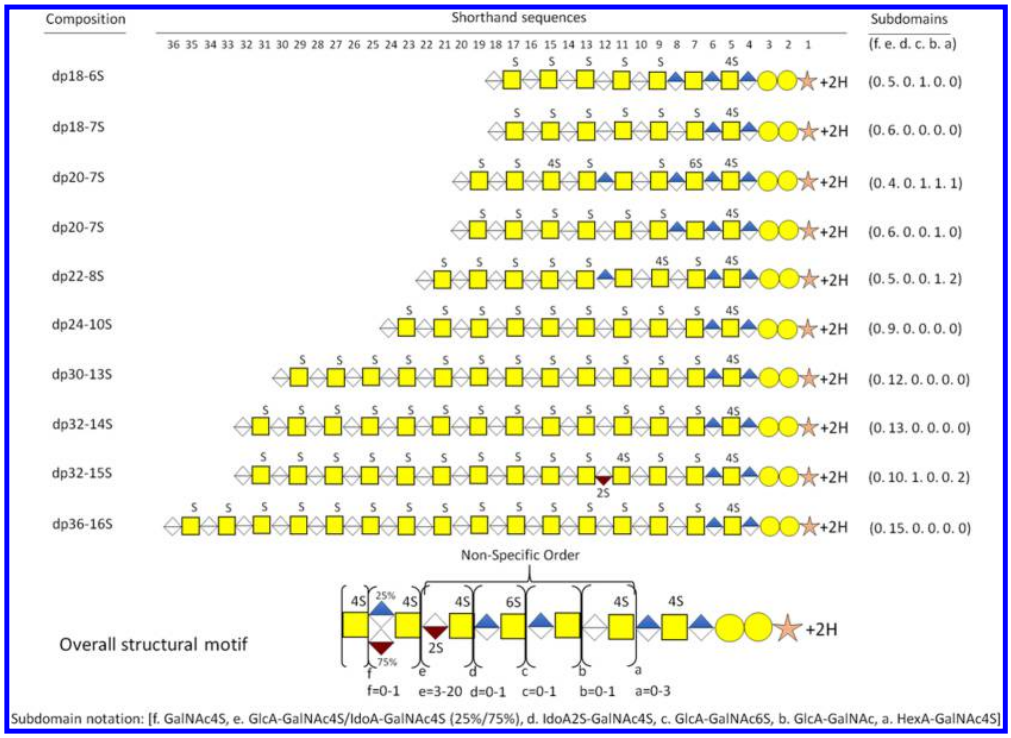


Figure 6. Motif for decorin GAG chains from dp14 to dp40. Disaccharide analysis reveals a relatively small fraction of disaccharides having modifications that are not 4S on the GalNAc. MS^2 shows that the majority of these GalNAc4S exist near the nonreducing end. Structural variability occurs primarily near the linker region. IdoA2S-GalNAc4S and GlcA-GalNAc are both observed in the MS^2 spectra but never more than one instance per chain. 6-O-Sulfation of the GalNAc is more difficult to observe using the CID ion activation method, as differentiation from the 4-O-sulfation requires specific cross-ring cleavages. Electron-based methods have proved to be more useful but have yet to be successful on highly heterogeneous samples such as decorin.

no diagnostic cross-ring fragments to localize sulfo moeties, except in rare cases. One rare case is shown in Figure S23 and Table S14 (SI), where cross-ring fragments are used to differentiate 4-*O*- and 6-*O*-sulfo modifications. The presence of the $^{1,4}X_6$ fragment ion with two sulfo groups sandwiched between the Y_6 ion containing one sulfo-group and the Y_7 ion containing two sulfo groups is evidence of a 6-O-modification. Regions where cross-ring fragments were identified (both diagnostic and nondiagnostic) are provided in the Supporting Information.

Assigned decorin structures show a consistent run of sequential glycosidic fragments from both the reducing and nonreducing end. Assigned glycosidic fragments with the highest signal intensity and greatest occurrence were B and Y ions. Low-abundance C fragments (<5%) were observed in nearly all spectra as well, while Z fragment ions were rarely assigned. The intense peak at m/z 458.06 observed in every tandem MS of decorin, regardless of fraction or composition, matches the mass of a B-fragment ion for a singly sulfated disaccharide residue or a polymer chain of the same residue, $(\text{HexA-GalNAc-S-H})_n^{n-}$. Examination of the isotopic distribution shows the possibility of multiple charge states from −1 to −11 in some spectra, suggesting a repetitive 1-sulfo group per disaccharide chain for up to 22 sugar residues. Comparison of this isotope distribution between MS/MS spectra shows the −1 charge state as dominant for shorter chains (dp18), whereas the −4 charge state is most abundant for the highest chain. Loss of the sulfo modification was observed most commonly for glycosidic B-fragments (B_1−B_8) but typically appeared at intensities of 3−22% of the fully modified counterpart. The C-5 stereochemistry of acidic residues (IdoA vs GlcA) could not be determined with our current tandem mass spectrometry method. Work from Zaia et al.[32,33] suggests a correlation between C-5 epimerization and the abundance of $^{0,2}X_n$ and Y_n ions for CS/DS but lacks definitive measurements for application in long-chain glycans such as decorin. Differentiation of the C-5 uronic acid stereochemistry in CS/DS is possible using electron detachment dissociation (EDD), based on a diagnostic ratio of fragment ion relative intensity that was determined using principal component analysis (PCA).[34−36] Current work is focused on extending the diagnostic ratios approach to longer oligomers. While future efforts will be required to establish the definitive sequence for porcine skin decorin GAG chains, a major structural motif can be derived on the basis of both biochemical analysis and MS sequencing (Figure 6).

Nonreducing End Motif. Sequence analysis shows homogeneity in the nonreducing end of porcine decorin GAG. The HexA-GalNAc-S disaccharide motif is consistently repeated at the nonreducing end with no evidence of alternative structures. Although information to determine C-5 uronic sugar stereochemistry is lacking, an intense series of B fragments and/or C fragments is present in all decorin tandem MS for the HexA-GalNAc-S repeating unit. The overall percentage of HexA-GalNAc-4S disaccharide as determined by disaccharide analysis increases with respect to chain length. These findings are reflected in the tandem MS results, with the total number of

DOI: 10.1021/jacs.7b10164
J. Am. Chem. Soc. 2017, 139, 16986−16995

observed sequential B-ions that favor the HexA-GalNAc-S pattern increasing when chains are longer.

Reducing End Motif. Observable variations in sulfation patterns of different sequences occur primarily at the reducing end. HexA-GalNAc6S, IdoA2S-GalNAc4S, and GlcA-GalNAc disaccharide variants occur infrequently, as suggested by disaccharide analysis, and are minor components in the overall sequence. Sequences derived from tandem MS data show that these alternatives occur from 0 to 2 times per chain. Sequential series of Y fragments suggest variability in the region closest to the reducing end after the GlcA-GalNAc4S-GlcA-Gal-Gal-Xyt linker region. An unmodified disaccharide unit exists 1−4 disaccharide units after the linker region. The IdoA2S-GalNAc4S modification exists in a similar region. Cross-ring fragments that validate the presence of 6-*O*-sulfo modification on the GalNAc (hence a GlcA-GalNAc6S unit) are only observed near the reducing end. Sequential Y-ions near the reducing end exist from 2 to 30% relative ion intensity but exhibit no common intensity-dependent pattern (unlike B-ions at the nonreducing end). The reducing end is the region where all variations to the HexA-GalNAc-S pattern exist but are observed in no specific order.

■ CONCLUSIONS

Similar to the bikunin GAG chain, the porcine skin decorin GAG chain appears to also show a major structural motif with no variation within a short region comprising the six saccharide units at the chain's reducing end (Figure 6). The next 12 saccharides (residues 7−18) show subtle variability in both sulfation and uronic acid epimers. The remaining saccharides (residues 18−30) extending to the nonreducing end of the chain are enriched (∼75%) in →4) IdoA (1→3) GalNAc4S (1→ repeating units. These flexible IdoA-rich domains are believed to be responsible for much of the protein−GAG interaction associated with decorins GAG-chain-mediated activities. In summary, the heterogeneity in the decorin GAG chain structure is infrequent and occurs as a minor percentage of the overall sequence. These results should simplify future structure−activity relationship studies on decorin PGs.

It remains to be seen if the relatively invariant structures of the chondroitin sulfate PG bikunin and dermatan sulfate PG decorin extend to the heparan sulfate family. Heparan sulfates contain *N*-sulfo groups and both uronic acid epimers, so they are much more structurally complex. We estimate the theoretical structural variability of a single heparan sulfate chain occupying a specific site within a core protein to be $>10^{24}$. The current study on decorin demonstrates that MS analysis is possible on GAG chains of dp20−44 with 8−20 sulfo groups/GAG chain. Improved MS technology will be required for the longer and more highly sulfated heparan sulfate GAG chains. Moreover, the routine determination of the C-5 uronic acid stereochemistry and the potential lability of heparan sulfate's *N*-sulfo groups will require extensive exploration. While progress continues to be made in GAG sequencing, additional improvements in separation technology and mass spectrometry will be required to sequence this incredibly complex family of biomacromolecules.

■ EXPERIMENTAL SECTION

Materials. All chemicals and reagents were obtained from commercial sources and used as received unless otherwise specified. Chondroitin sulfate lyase ABC from *Proteus vulgaris* was expressed in our laboratory using an *Escherichia coli* strain provided by Prof. Miroslaw Cygler (College of Medicine, University of Saskatchewan), and chondroitin sulfate lyase II from *Arthrobacter aurescens* was expressed in *E. coli*.[37]

Preparation of Decorin Glycosaminoglycan. Decorin proteoglycan was purified from porcine skin.[15] Decorin PG fraction was proteolyzed by a 5% (w/w) actinase E digestion at pH 8.0 in 50 mM Tris-HCl in sodium acetate. The enzymatic reaction proceeded at 55 °C for 24 h and was then isolated from the digestion mixture by a strong anion exchange spin column. Spin columns were pre-equilibrated with 8 M urea containing 2% (w/v) CHAPS [3-((3-cholamidopropyl)dimethylammonio)-1-propanesulfonic acid] and centrifuged at 500*g* for 5 min. The bound pG was washed once with 8 M urea containing 2% (w/v) CHAPS and three times with 50 mM NaCl. Decorin pG was eluted with 2 M NaCl, desalted using a 3 kDa MWCO (molecular weight cutoff) centrifugal filter, and lyophilized. The GAG component of decorin was released by base-catalyzed β-elimination under reducing conditions. Samples were dissolved in 0.2 M NaOH solution containing 1% $NaBH_4$. The reaction was allowed to proceed overnight at 4 °C and neutralized with 1 M hydrochloric acid. The resulting GAG mixture was purified using a 3 kDa MWCO centrifugal filter.

Linkage Region Analysis of Decorin Glycosaminoglycan. Approximately 200 μg of decorin glycosaminoglycan was digested completely using 200 mU chondroitin sulfate lyase ABC at 37 °C for 18 h, and the digested sample was then purified by a 3 kDa MWCO spin column filter to isolate reducing ends and lyophilized for LC−MS analysis.

Isolation and Preparation of Low Molecular Weight and Medium Charge GAG Fractions. Size exclusion chromatography (SEC) was performed on an ÄKTApurifier fast protein liquid chromatograph (GE Healthcare Bio-Science) using prepacked Superdex S75 column with a sample injection volume of 200 μL and a flow rate of 0.5 mL/min. The mobile phase consisted of 0.2 M ammonium bicarbonate. The fraction collector (Frac 920) was set to 2 min in conjunction-accumulated fractions. The GAG concentration in each fraction was determined by the microcarbozole assay. Strong anion exchange chromatography (SAX) was performed on an ÄKTApurifier using a Hiprep Q HP 16/10 column with a sample injection volume of 25 mL and a flow rate of 3 mL/min. The mobile phase was sodium chloride and water, and a gradient wash from 0 to 2 M sodium chloride in 15 column volumes was applied. Further SEC fractionation was performed on a HPLC (Shimadazu) using Superdex increase 75 10/300 GL (GE Healthcare) with a refractive index detector.

Fractionation of Decorin Glycosaminoglycan by Continuous Elution PAGE. A gel of 10 cm column height with 4 mL of 15% total acrylamide monomer resolving solution was allowed to polymerize overnight with 4 μL of TEMED and 12 μL of 10% (w/v) ammonium persulfate and was cast in a Mini Prep column with a 7 mm internal diameter (Bio-Rad). Above the polymerized resolving gel, 1 mL of 5% total acrylamide monomer stacking gel was cast. An aliquot of 1 mg of purified decorin glycosaminoglycan was loaded in a solution of 10 μg/mL (w/v) Phenol Red and 25% (w/v) sucrose. Electrophoresis was performed for 8 h at a constant power of 1 W with a peristaltic pump (Econo pump, Bio-Rad) set to 0.08 mL/min and fraction collector (Model 2110, Bio-Rad) set to 3 min in conjunction-accumulated separating fractions from the Mini Prep cell (Bio-Rad). Buffer salts from electrophoresis for each fraction were removed by a strong anion exchange column (High-Capacity Mini-Q, Satorius) and thoroughly desalted by a 3 kDa spin column (Millipore) washed with LC-grade water. The extent of separation was visualized by 15% total acrylamide monomer solution using native mini-slab PAGE stained with Alcian Blue, and the molecular weight of the fractionated GAG was estimated by PAGE densitometry against heparin ladder using UN-SCANIT (Silk Scientific).

Orbitrap FTMS Analysis of Decorin Glycosaminoglycan. Glycosaminoglycans were analyzed in the negative-ion mode by electrospray ionization on a Thermo Scientific LTQ Orbitrap XL FT mass spectrometer with a standard, factory-installed ion source (Thermo Scientific). External calibration of mass spectra produced a mass accuracy of <3 ppm. Samples were dissolved in 50% aqueous

DOI: 10.1021/jacs.7b10164
J. Am. Chem. Soc. 2017, 139, 16986−16995

methanol with 0.1% formic acid and were delivered by an Agilent 1200 nano-LC pump at a flow rate of 50 μL/min. Mass spectra were acquired at a resolution of 60 000 and detection range of m/z 400−2000, and the charge deconvolution was performed manually with electronic spreadsheets. Acquisition parameters used to prevent in-source fragmentation included the following: spray voltage, −4.2 kV; capillary voltage, −15 V; tube lens voltage, −100 V; capillary temperature, 250 °C; sheath flow rate, 25; and auxiliary gas flow rate, 5. For collision-induced dissociation (CID) MS/MS of the linkage region, parent ions were fragmented with a specified collision energy of 35 V.

FT-ICR MS Analysis of Decorin Glycosaminoglycan. Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR MS) experiments were performed in negative-ion mode with a 9.4 T Bruker Apex Ultra QeFTMS (Bruker Daltonics) fitted with an Apollo II dual source. Solutions of each decorin glycosaminoglycan fraction were introduced at a concentration of ∼0.001 mg/mL in 50% aqueous MeOH. The sample solutions were infused at a rate of 7−14 μL/h and were ionized by nanoelectrospray using a pulled fused silica tip (model FS360-75-15-d-20, New Objective). Compositions from mass spectra where the monoisotopic peak was observed were assigned with a mass accuracy of 6 ppm or better after external calibration. Isotope packets where monoisotopic peaks were not observed directly were extrapolated mathematically and assigned compositions with a mass accuracy of 10 ppm or better. Tandem MS experiments were performed by mass selection of precursor ions and activation by CID in the hexapole collision cell of the Apex instrument with specified collision energies ranging from 8 to 20 V. A total of 24−48 scans were signal-averaged for each tandem MS experiment with average transient lengths of 0.75 s and 200 000 mass resolving power at m/z 400. Glycosidic bond cleavages were assigned with a mass accuracy of less than 5 ppm with in-house-developed software described below.

Automated Tandem MS Data Analysis. The overall search space of viable structures for each chain length is shown in Table 1. Analysis of over 1 million structures by manual examination is impractical. In-house software developed in the MATLAB (Mathworks) coding environment is used to automate the interpretation of tandem MS data. Modification and chain lengths are provided as user inputs after composition determination by accurate mass measurement. The software employs a genetic algorithm in combination with the sulfate modification limitations based on disaccharide analysis to drastically reduce overall search space of possible structures for unknown glycans. Tandem mass spectra of unknown structures are compared against theoretical structures using a closeness-of-fit function based on the number of matching glycosidic fragments, fragment intensities, and overall depth of sequence coverage. Differentiation between 4-*O*- and 6-*O*-sulfo group modifications occurs when diagnostic cross-ring fragmentation is observed. Sulfo groups without diagnostic cross-ring fragments are not assigned a specific position and intentionally left ambiguous with a notation of S in annotated structures. Charge-state matching and isotope-pattern recognition were verified first using an automated charge-spacing module and then by manual examination when interpreting FT-ICR-MS/MS data. A more comprehensive and detailed explanation of the algorithm and the software is provided in the doctoral thesis of author J.D.[38]

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.7b10164.

Figures S1−S32 and Table S1−S23 provide separation, mass spectral, and NMR data (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

*jamster@uga.edu

*linhar@rpi.edu

ORCID

Fuming Zhang: 0000-0003-2803-3704

I. Jonathan Amster: 0000-0001-7523-5144

Robert J. Linhardt: 0000-0003-2219-5833

Author Contributions

○Y.Y. and J.D. contributed equally to this work.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by grants from the US National Institutes of Health (grants HL096972, HL136271, HL62244, HL094463, GM38060, and GM103390).

■ REFERENCES

(1) Hart, G. W.; Copeland, R. J. *Cell* **2010**, *143*, 672.

(2) Pilobello, K. T.; Mahal, L. K. *Curr. Opin. Chem. Biol.* **2007**, *11*, 300.

(3) Tyers, M.; Mann, M. *Nature* **2003**, *422*, 193.

(4) Neelamegham, S.; Mahal, L. K. *Curr. Opin. Struct. Biol.* **2016**, *40*, 145.

(5) Vetr, H.; Gerhard, W. *Biol. Chem. Hoppe-Seyler* **1990**, *371*, 1185.

(6) Hascall, V. C.; Sajdera, S. W. *J. Biol. Chem.* **1970**, *245*, 4920.

(7) Kitagawa, H.; Oyama, M.; Masayama, K.; Yamaguchi, Y.; Sugahara, K. *Glycobiology* **1997**, *7*, 1175.

(8) Fransson, L. A.; Belting, M.; Jönsson, M.; Mani, K.; Moses, J.; Oldberg, A. *Matrix Biol.* **2000**, *19*, 367.

(9) Malmström, A.; Bartolini, B.; Thelin, M. A.; Pacheco, B.; Maccarana, M. *J. Histochem. Cytochem.* **2012**, *60*, 916.

(10) Ly, M.; Leach, F. E., III.; Laremore, T. N.; Toida, T.; Amster, I. J.; Linhardt, R. J. *Nat. Chem. Biol.* **2011**, *7*, 827.

(11) Kiani, C.; Chen, L.; Wu, Y. J.; Yee, A. J.; Yang, B. B. *Cell Res.* **2002**, *12*, 19.

(12) Chi, L.; Wolff, J. J.; Laremore, T. N.; Restaino, O. F.; Xie, J.; Schiraldi, C.; Toida, T.; Amster, I. J.; Linhardt, R. J. *J. Am. Chem. Soc.* **2008**, *130*, 2617.

(13) Lamkin, E.; Cheng, G.; Calabro, A.; Hascall, V. C.; Joo, E. J.; Li, L.; Linhardt, R. J.; Lauer, M. E. *J. Biol. Chem.* **2015**, *290*, 5156.

(14) Laremore, T. N.; Ly, M.; Zhang, Z.; Solakyildirim, K.; McCallum, S. A.; Owens, R. T.; Linhardt, R. J. *Biochem. J.* **2010**, *431*, 199.

(15) Zhao, X.; Yang, B.; Solakylidirim, K.; Joo, E. J.; Toida, T.; Higashi, K.; Linhardt, R. J.; Li, L. *J. Biol. Chem.* **2013**, *288*, 9226.

(16) Toyoda, H.; Kobayashi, S.; Sakamoto, S.; Toida, T.; Imanari, T. *Biol. Pharm. Bull.* **1993**, *16*, 945.

(17) Yamada, S.; Oyama, M.; Kinugasa, H.; Nakagawa, T.; Kawasaki, T.; Nagasawa, S.; Khoo, K. H.; Morris, H. R.; Dell, A.; Sugahara, K. *Glycobiology* **1995**, *5*, 335.

(18) Zamfir, A.; Seidler, D. G.; Kresse, H.; Peter-Katalinic. *Glycobiology* **2003**, *13*, 733.

(19) Seo, N. S.; Hocking, A. M.; Höök, M.; McQuillan, D. J. *J. Biol. Chem.* **2005**, *280*, 42774.

(20) Hocking, A. M.; Shinomura, T.; McQuillan, D. J. *Matrix Biol.* **1998**, *17*, 1.

(21) Seidler, D. G.; Dreier, R. *IUBMB Life* **2008**, *60*, 729.

(22) Järveläinen, H.; Sainio, A.; Wight, T. N. *Matrix Biol.* **2015**, *43*, 15.

(23) Gubbiotti, M. A.; Vallet, S. D.; Ricard-Blum, S.; Iozzo, R. V. *Matrix Biol.* **2016**, *55*, 7.

(24) Schaefer, L.; Tredup, C.; Gubbiotti, M. A.; Iozzo, R. V. *FEBS J.* **2017**, *284*, 10.

(25) Lin, Y. P.; Osburne, M. S.; Pereira, M. J.; Coburn, J.; Leong, M. Glycosaminoglycan Binding by the Lyme Disease Spirochete is a Determinant of Tissue Tropism and Disease. In *Glycobiology and Human Diseases*; CRC Press, 2016; p 86.

(26) Lin, Y.; Li, L.; Zhang, F.; Linhardt, R. J. *Microbiology* **2017**, DOI: 10.1099/mic.0.000571.

(27) Linhardt, R. J.; Hileman, R. E. *Gen. Pharmacol.* **1995**, *26*, 443.
(28) Malavaki, C.; Mizumoto, S.; Karamanos, N.; Sugahara, K. *Connect. Tissue Res.* **2008**, *49*, 133.
(29) Pojasek, K.; Shriver, Z.; Kiley, P.; Venkataraman, G.; Sasisekharan, R. *Biochem. Biophys. Res. Commun.* **2001**, *286*, 343.
(30) Linhardt, R. J. *Current Protocols in Molecular Biology*; Wiley, 2001; Chapter 17, Unit 17.13B.
(31) He, W.; Zhu, Y.; Shirke, A.; Sun, X.; Liu, J.; Gross, R. A.; Koffas, M. A. G.; Linhardt, R. J.; Li, M. *Appl. Microbiol. Biotechnol.* **2017**, *101*, 6919.
(32) Zaia, J.; Li, X. Q.; Chan, S. Y.; Costello, C. E. *J. Am. Soc. Mass Spectrom.* **2003**, *14*, 1270.
(33) Bielik, A. M.; Zaia. *Int. J. Mass Spectrom.* **2011**, *305*, 131.
(34) Leach, F. E.; Ly, M.; Laremore, T. N.; Wolff, J. J.; Perlow, J.; Linhardt, R. J.; Amster, I. J. *J. Am. Soc. Mass Spectrom.* **2012**, *23*, 1488.
(35) Kailemia, M. J.; Patel, A. B.; Johnson, D. T.; Li, L.; Linhardt, R. J.; Amster, I. J. *Eur. Mass Spectrom.* **2015**, *21*, 275.
(36) Agyekum, I.; Zong, C. L.; Boons, G. J.; Amster, I. J. *J. Am. Soc. Mass Spectrom.* **2017**, *28*, 1741.
(37) Williams, A.; He, W.; Cress, B. F.; Liu, X.; Alexandria, J.; Yoshizawa, H.; Nishimura, K.; Toida, T.; Koffas, M.; Linhardt, R. J. *Biotechnol. J.* **2017**, *12*, 1700239.
(38) Duan, J. Doctoral Thesis, Department of Chemistry, University of Georgia, 2018.

DOI: 10.1021/jacs.7b10164
J. Am. Chem. Soc. 2017, 139, 16986−16995