

# The proteoglycan bikunin has a defined sequence

Mellisa Ly<sup>1</sup>, Franklin E Leach III<sup>2</sup>, Tatiana N Laremore<sup>1</sup>, Toshihiko Toida<sup>3</sup>, I Jonathan Amster<sup>2</sup> & Robert J Linhardt<sup>1,4-6\*</sup>

**Proteoglycans are complex glycoconjugates that regulate critical biological pathways in all higher organisms. Bikunin, the simplest proteoglycan, with a single glycosaminoglycan chain, is a serine protease inhibitor used to treat acute pancreatitis. Unlike nucleic acids and proteins, whose synthesis is template driven, Golgi-synthesized glycosaminoglycans are not believed to have predictable or deterministic sequences. Bikunin peptidoglycosaminoglycans were prepared and fractionated to obtain a collection of size-similar and charge-similar chains. Fourier transform mass spectral analysis identified a small number of parent molecular ions corresponding to monocompositional peptidoglycosaminoglycans. Fragmentation using collision-induced dissociation unexpectedly afforded a single sequence for each monocompositional parent ion, unequivocally demonstrating the presence of a defined sequence. The biosynthetic pathway common to all proteoglycans suggests that even more structurally complex proteoglycans, such as heparan sulfate, may have defined sequences, requiring a readjustment in the understanding of information storage in complex glycans.**

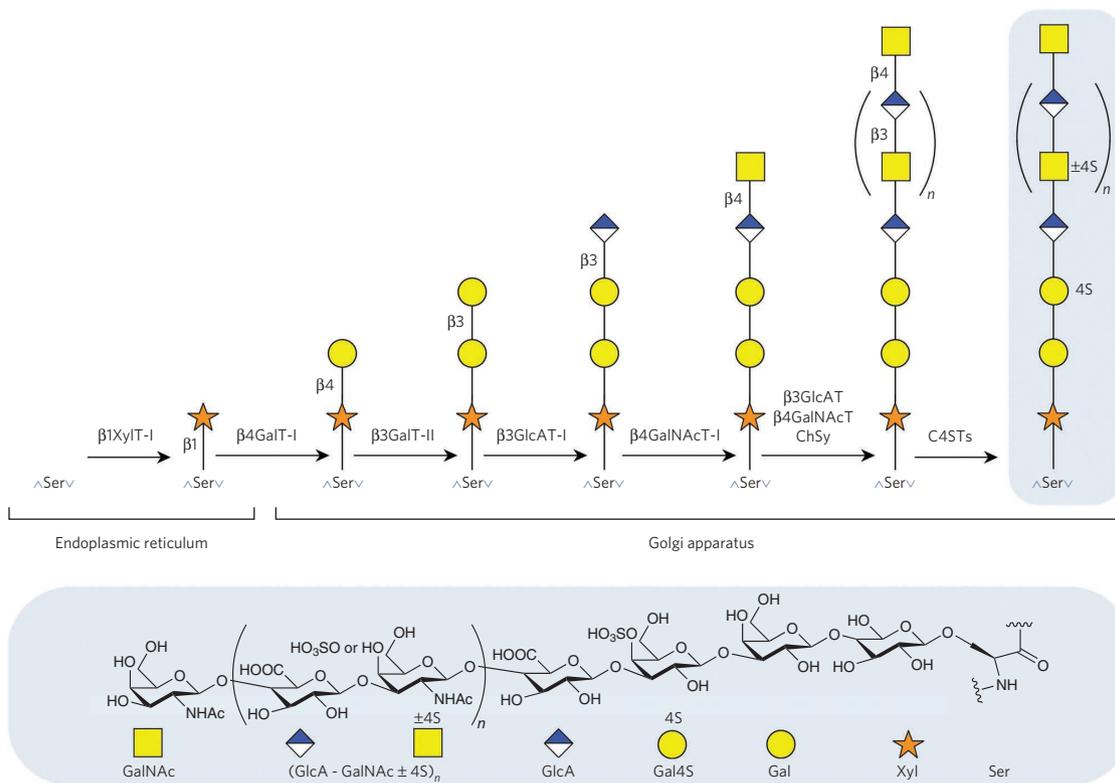
The human genome, proteome and glycome, the latter being the major component of the metabolome, have come under increased scrutiny in the development of new therapeutic strategies, particularly the next generation of biopharmaceuticals<sup>1</sup>. Solving the chemical structure of glycan components of biomolecules is also critical in answering fundamental questions in basic biology relevant to the translational science of medicine. The question of whether glycans have a predictable or deterministic sequence is currently disputed<sup>2,3</sup>. Although the biosynthetic mechanism resulting in the introduction of sequence into nucleic acids and proteins, which involves template-driven biosynthesis, is well understood, the introduction of sequence into glycans within the endoplasmic reticulum and Golgi is not. Although enzymes can recognize remote sequences that impart domains, the organized biosynthesis of the multiple domains within a full glycosaminoglycan (GAG) chain may be beyond enzyme specificity. Despite all that is known about glycan biosynthesis, the current understanding is still insufficient to infer sequence or even to suggest that glycans possess sequence. Indeed, it is valid to question why a glycan would even need a definable sequence. Many glycans, such as glycogen and starch<sup>4</sup>, which are responsible for storing energy in plants and animals, or alginate and cellulose<sup>5</sup>, which are responsible for providing structure to aquatic and terrestrial plants, have relatively simple physiological roles. In such cases, the specific arrangement of branches, linear sequence and chain length may not be required, as multiple structures might fill these biological roles.

Some glycans, however, have a much more complex or even a dominant role in biology. Proteoglycans, for example, are glycoconjugates containing GAG side chains that are intricately modified during biosynthesis in the endoplasmic reticulum and Golgi (Fig. 1)<sup>6-8</sup>. There are four families of proteoglycans: hyaluronan, keratan sulfate, chondroitin sulfate (including dermatan sulfate) and heparan sulfate (including heparin). With the exception of hyaluronan, these consist of one or more linear, sulfated polysaccharide chains (GAGs) covalently linked to a core protein<sup>2</sup>. The GAG chains of these proteoglycans are involved in complex biological roles such as signal transduction (that is, assembling and

activating growth factor–growth factor receptor complexes)<sup>9</sup> or pathway regulation (that is, the coagulation and complement cascades)<sup>10</sup>. These complex roles suggest sequence-dependent or sequence-specific mechanisms. Despite the suggestion that these glycans carry out sequence-dependent roles, only a handful of specific GAG sequences have a well-understood structure-activity relationship; the prototypical example is the heparin antithrombin-binding pentasaccharide sequence, which is critical for heparin's anticoagulant activity<sup>11,12</sup>.

If proteoglycans have defined sequence, determined during their biosynthesis, then one might ask why it has been so difficult to determine GAG chain sequence. There are several reasons for the slow progress of this field, including the lack of sufficient quantities of pure proteoglycans, the multiple sequences possible for the multiple GAG chains often present on a single core protein, the difficulties in purifying a single GAG chain for sequencing and the difficulties in determining GAG sequence. In an effort to sequence a proteoglycan, we have carefully selected the bikunin proteoglycan as our target. Bikunin, a member of the kunin family of serine protease inhibitors<sup>13-15</sup>, is a therapeutically relevant proteoglycan that is used in Japan as a drug for the treatment of acute pancreatitis<sup>16</sup>. As a result, bikunin is available at a high level of purity in multi-milligram quantities. Bikunin has the simplest chemical structure of any proteoglycan, with a single site for O-linked modification by a GAG chain, located at Ser10 in its 16-kDa core protein<sup>17-20</sup>. Though the protein component of bikunin is well characterized<sup>21,22</sup>, its GAG chain structure is heterogeneous and has received less attention because of the technical difficulties associated with GAG analysis. Although the bikunin GAG chain is quite short, it is very heterogeneous in size and composition, with 27–39 saccharide residues and a molecular mass ( $M_R$ ) ranging from 5,505 Da to 7,102 Da<sup>23</sup>. In addition, enzymatic analysis shows that the bikunin GAG chains contain single-uronic-acid stereochemistry (glucuronic acid), sulfo groups at only the 4 position of its galactosamine residue and no *N*-sulfo group or *N*-acetyl group variability, which is common in the GAG chains of the more structurally complex members of the heparan sulfate proteoglycan family.

<sup>1</sup>Department of Chemistry and Chemical Biology, Rensselaer Polytechnic Institute, Troy, New York, USA. <sup>2</sup>Department of Chemistry, University of Georgia, Athens, Georgia, USA. <sup>3</sup>Graduate School of Pharmaceutical Sciences, Chiba University, Chiba, Japan. <sup>4</sup>Department of Chemical and Biological Engineering, Rensselaer Polytechnic Institute, Troy, New York, USA. <sup>5</sup>Department of Biology, Rensselaer Polytechnic Institute, Troy, New York, USA. <sup>6</sup>Department of Biomedical Engineering, Center for Biotechnology and Interdisciplinary Studies, Rensselaer Polytechnic Institute, Troy, New York, USA. \*e-mail: linhar@rpi.edu



**Figure 1 | Biosynthetic pathway for chondroitin sulfate A 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:  $\beta$ 1XylT-I,  $\beta$ -xylosyl transferase I;  $\beta$ 4GalT-I,  $\beta$ -4-galactosyl transferase I;  $\beta$ 3GalT-II,  $\beta$ -3-galactosyl transferase II;  $\beta$ 3GlcAT-I,  $\beta$ -3-glucuronosyl transferase I;  $\beta$ 4GalNAcT-I,  $\beta$ -4-N-acetyl galactosaminyl transferase I;  $\beta$ 3GlcAT,  $\beta$ -3-glucuronosyl transferase;  $\beta$ 4GalNAcT,  $\beta$ -4-N-acetyl galactosaminyl transferase; ChSy, chondroitin synthases; C4STs, galactosyl 4-O-sulfo transferase and N-acetyl galactosaminyl-4-O-sulfotransferase.

Several critical issues, including GAG-chain release from the core protein, GAG-chain recovery and purification and identification of an appropriate sequencing strategy must be addressed to successfully sequence bikunin. GAG chains linked via an O-glycosidic bond are most commonly released from the core protein through  $\beta$ -elimination<sup>24</sup>, but this method involves harsh conditions that can result in GAG-chain modification, complicating sequencing. Proteolysis, a mild method to recover peptidoglycosaminoglycan from a proteoglycan, also poses risks. If proteolysis is incomplete, the mixture complexity can be increased by variability in the length of residual core peptide<sup>23</sup>. Next, the recovery and purification of GAG or peptidoglycosaminoglycan should require as few steps as possible so as not to lose sample or bias the mixture, but it should have sufficient resolution to obtain one or a small number of individual chain lengths and compositions for sequencing. Finally, the analytical approach used for sequencing should be rapid and definitive, and ideally it would use widely available technology.

Methods for GAG sequencing reported to date have relied on various bottom-up approaches, for example, the reassembly of fragments of GAGs, like pieces of a puzzle, into motifs or domains<sup>25–29</sup>. Short sequences of GAG oligosaccharides have been read using a combination of enzymes and chemical-chain scission<sup>30–33</sup>. Tandem MS of oligosaccharides has led to an understanding of the importance of charge state in providing sequence-meaningful product ions<sup>27–29,34–36</sup>.

The present work demonstrates a new top-down glycomics approach that takes advantage of mass spectrometry to completely characterize the GAG-peptidoglycosaminoglycan chains of a proteoglycan. The presence of ordered domains on both ends of the bikunin GAG chain, shown by previous studies, reduces the number of possible sequences<sup>23,37,38</sup>. In the current study, bikunin

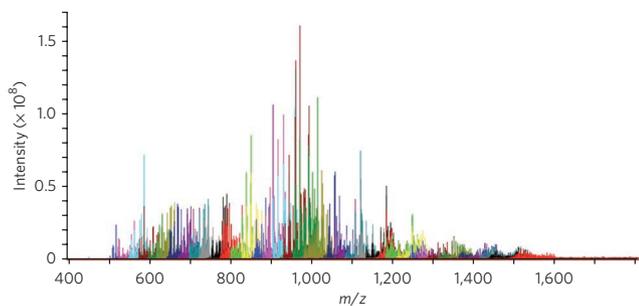
peptidoglycosaminoglycan is recovered from the therapeutic urinary bikunin proteoglycan and fractionated by continuous elution preparative PAGE to obtain a number of simple mixtures of size-similar and charge-similar peptidoglycosaminoglycans that are sequenced in a top-down approach using Fourier transform (FT) and FT ion cyclotron resonance (ICR) MS with collision-induced dissociation (CID) fragmentation.

## RESULTS

### Preparation of peptidoglycosaminoglycan and linkage region

Complete digestion of the bikunin proteoglycan core protein was accomplished by exhaustive treatment with actinase E. The complete chondroitin sulfate lyase ABC digestion of the resulting bikunin peptidoglycosaminoglycan afforded several products: a single product corresponding to the linkage region at the reducing end, the  $\Delta_{4,5}$  unsaturated disaccharides and tetrasaccharides from the middle of the chain, and a saturated trisaccharide from the chain's nonreducing end. Centrifugal spin membrane (nominal molecular weight cut-off (MWCO) 10 kDa) separation afforded a retentate containing a single reducing-end hexasaccharide, O-glycosidically linked to a serine residue ( $m/z$  628.1118). Low collision-energy (25 eV) activation applied to this abundant reducing-end ion using FT-MS produced fragmentation with no sulfo-group loss and identified the linkage region as  $\Delta$ UA-GalNAc4S-GlcA-Gal4S-Gal-Xyl-Ser (Xyl, xylose) (**Supplementary Results, Supplementary Fig. 1**). Thus, actinase E proteolysis of bikunin proteoglycan afforded a relatively simple mixture of peptidoglycosaminoglycan chains.

The weight percentage of the peptidoglycosaminoglycan prepared from bikunin proteoglycan after dialysis was ~30%. The poly-disperse bikunin peptidoglycosaminoglycan mixture had a number average molecular mass ( $M_N$ ) of 7.0 kDa and a weight-average



**Figure 2 | FT-ICR-MS analysis of polydisperse and heterogeneous bikunin peptidoglycosaminoglycan.** ‘Quad-windowed’ mass spectra are depicted with individual narrow  $m/z$  acquisitions in unique colors.

molecular mass ( $M_w$ ) of 7.7 kDa based on PAGE using fractionated chondroitin sulfate-A standards (Supplementary Fig. 2).

### FT-ICR-MS for peptidoglycosaminoglycan identification

Previously, attempts to characterize the complex mixture of molecular compositions derived from bikunin at the MS level have provided minimal information because of heterogeneity in the extent of sulfation and degree of polymerization and extensive Na-H exchange. Windows of 50  $m/z$  were selected with a quadrupole mass filter, introduced into the FT-ICR mass analyzer and combined to generate a composite mass spectrum and provide sufficient signal for MS analysis<sup>23</sup>. This technique results in an improved spectrum for a mixture of intact bikunin peptidoglycosaminoglycan chains (Fig. 2), although acquisition without quadrupole windowing can also be achieved owing to advances in purification<sup>39,40</sup> and reduction of Na-H heterogeneity using formic acid (0.1 vol. %) in the ESI solvent<sup>41</sup>. Analysis of the quadrupole-windowed MS results in 47 molecular compositions, as listed in Supplementary Table 1.

### Peptidoglycosaminoglycan fractionation

Continuous elution PAGE was next used to separate bikunin into size-similar and charge-similar fractions. Bikunin peptidoglycosaminoglycan fractions of 1.4 ml eluting from preparative electrophoresis were labeled as f1–f200, beginning with the ion front that coincided with phenol red dye. Initial examination by mini-slab PAGE with alcian blue followed by silver staining showed that peptidoglycosaminoglycan was first detected in f45. Fractions ( $\geq f50$ ) containing sufficient amounts of peptidoglycosaminoglycan for MS analysis were then analyzed on 15% mini-slab PAGE together with molecular-weight standards. Peptidoglycosaminoglycans with  $M_R$  ranging from 5.37 kDa to 9.77 kDa were assigned based on the electrophoresis results of f50–f117 (Supplementary Fig. 2, Supplementary Table 2). Uronic acid determination showed 13–52  $\mu\text{g}$  bikunin peptidoglycosaminoglycan per fraction, which corresponded to ~1.4–7.3 nmol bikunin peptidoglycosaminoglycan per fraction (Supplementary Fig. 3, Supplementary Table 3). The major peptidoglycosaminoglycan (degree of polymerization (dp) 35-5-Ser) represented ~2 mol % of the total. The composition is designated degree of polymerization (dp) and sulfo group number (that is, dp35-5-Ser is 35 saccharide units with five sulfo groups O-glycosidically linked to a serine residue).

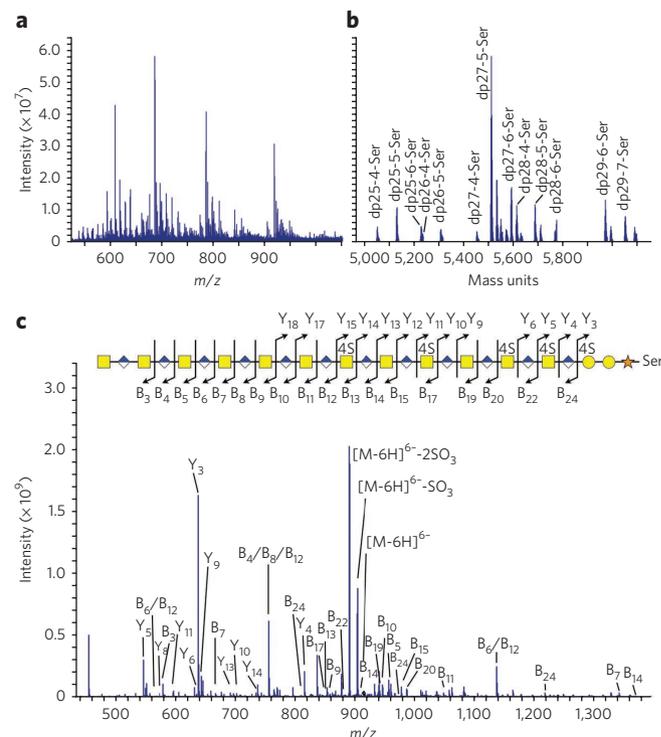
### Accurate mass of peptidoglycosaminoglycan fractions

FT-MS identified a wide range of intact peptidoglycosaminoglycans with degrees of polymerization from 23 to 55, all of which were O-linked to one serine residue and included four to nine sulfo groups per chain (Supplementary Tables 4–14). Nearly the same range was seen on two different mass spectrometers in adjacent fractions. Chains with both even and odd numbers of saccharide units were observed. Odd-numbered chains were much more abundant with

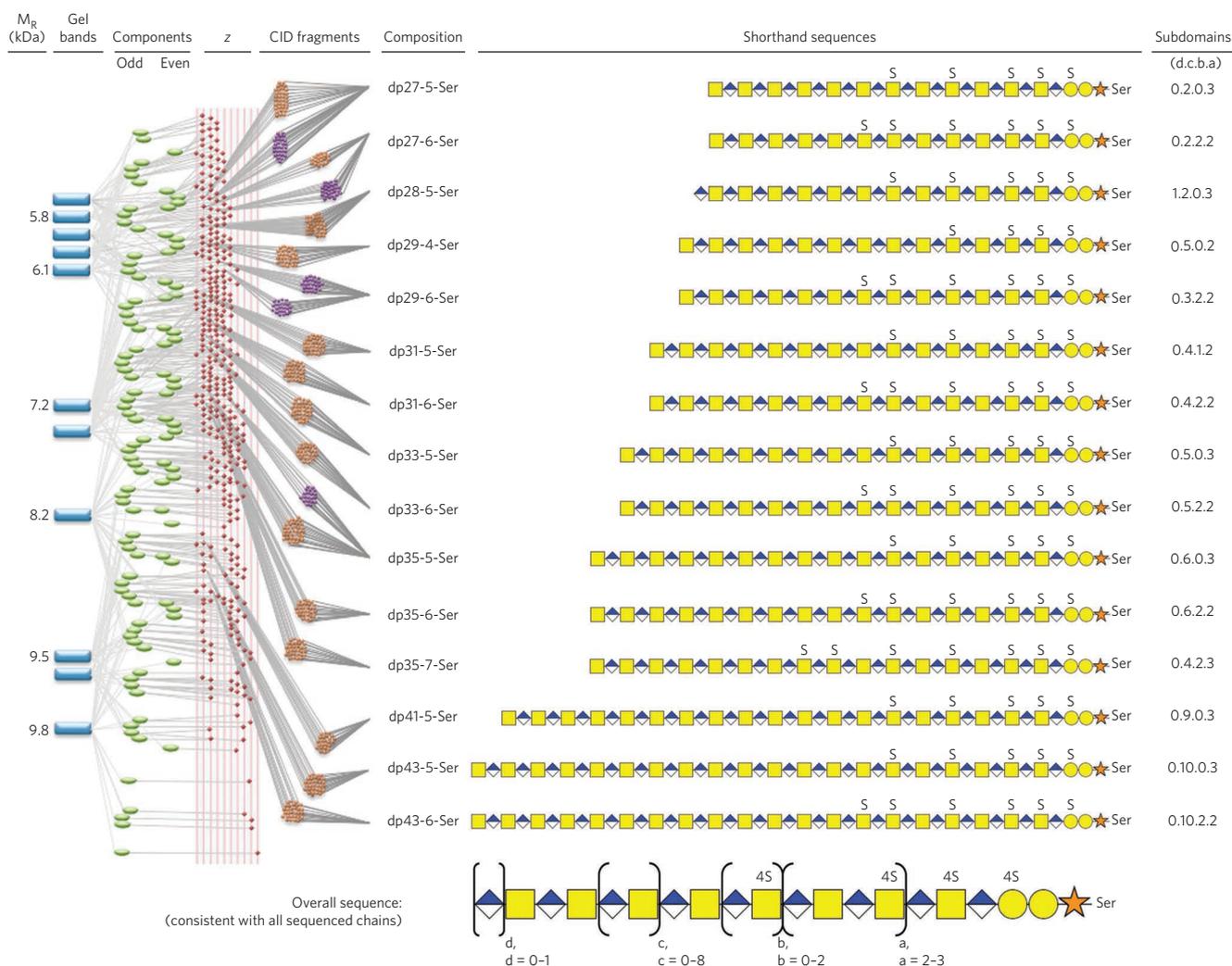
respect to relative parent-ion abundance. Compositional assignments were made with high confidence based on monoisotopic peaks and the isotopic envelope patterns. Different adjacent fractions, such as those corresponding to 5.80 kDa  $M_R$  by PAGE and 5.87 kDa  $M_R$  by PAGE, were analyzed by FT-MS and FT-ICR-MS (Fig. 3a), deconvoluted manually (Fig. 3b) and compared to statistically derived isotope patterns. The fraction corresponding to 5.80 kDa showed 18 components ranging from dp23-4-Ser to dp30-6-Ser (Supplementary Table 5), whereas the fraction corresponding to 5.87 kDa showed 20 components ranging from dp23-4-Ser to dp30-7-Ser. Much overlap was observed between adjacent fractions. Identified components of bikunin peptidoglycosaminoglycan are listed (Supplementary Tables 4–14) and their spectra are shown (Supplementary Figs. 4–13). Single components were identified with one to up to seven charge states, ranging from  $z = 5$ –14, dependent on the chain size of the peptidoglycosaminoglycan analyzed.

### Sequencing with CID-FT & CID-FT-ICR-MS/MS

CID experiments performed on FT-MS allowed unambiguous sequencing of dp27-5-Ser, dp27-6-Ser, dp29-6-Ser and dp35-5-Ser. CID data from FT-ICR-MS (Fig. 3c) afforded improved sensitivity and coverage for unambiguous sequencing of dp27-5-Ser, dp27-6-Ser, dp28-5-Ser, dp29-4-Ser, dp31-5-Ser, dp31-6-Ser, dp33-5-Ser, dp33-6-Ser, dp35-5-Ser, dp35-6-Ser, dp35-7-Ser, dp43-5-Ser and dp43-6-Ser. With 19–37 fragments per CID spectrum, we could sequence 15 odd chains having 4, 5, 6 and 7 sulfo groups ranging in chain length from 25–43 saccharide units and an even chain of dp28-5-Ser (Supplementary Figs. 14–33 and Supplementary Tables 15–35). An example of a peptidoglycosaminoglycan of single composition with a single charge state sequenced on both types of FT mass spectrometers is dp27-5-Ser (Fig. 3c and Supplementary Fig. 14). The range of charge states used for CID experiments, which



**Figure 3 | FT-ICR-MS analysis of a bikunin peptidoglycosaminoglycan fraction.** (a) FT-ICR negative-ion mass spectrum of 5.80-kDa  $M_R$  fraction by PAGE with 18 isobars and 63 parent ions. (b) Deconvolution of spectrum a. (c) CID-FT-ICR-MS/MS spectra of parent-ion  $m/z = 917.38$  ( $z = 6$ ) and annotated fragment-ions providing sequence with dp27-5-Ser fragmentation pattern assigned from spectrum.



**Figure 4 | Bikunin sequencing flow chart.** Flow chart reads from left to right. The  $M_R$  (kDa) determined based on PAGE of fractions (blue rectangles represent gel bands) of bikunin in peptidoglycosaminoglycan prepared by continuous elution PAGE is shown. The deconvoluted MS obtained using FT-ICR-MS affords the mass of 3–5 odd and even components (green ovals) observed in each bikunin peptidoglycosaminoglycan fraction is shown. Each MS spectrum showed multiple charge states ( $z$  values) shown as red diamonds from which parent ions were selected for MS/MS giving CID fragments by analysis on FT (purple circles) or FT-ICR (brown circles). A shorthand sequence for each chain is shown with a, b, c and d subdomain repeats indicated by numbers (that is, 0.2.0.3 for  $d = 0$ ,  $c = 2$ ,  $b = 0$ ,  $a = 3$ ). The overall sequence of bikunin chondroitin sulfate-A peptidoglycosaminoglycan shown at the bottom is consistent with all determined sequences.

produced full sequence information, was  $z = 6–9$ . GAG chains with the same composition were observed in different fractions and different charge states, yielding redundant sequence data. On larger and more highly sulfated GAG chains it was difficult to obtain meaningful fragmentation data, especially when relying on the LTQ Orbitrap FT-MS. Sulfo group placement on peptidoglycosaminoglycans with different GAG chain lengths and the same number of sulfo groups afforded exactly the same sulfo-group placement for all sequenced chains, with 4–7 sulfo groups. All bikunin chains containing 5 sulfo groups had the same sulfo-group pattern, as did all chains containing 6 sulfo groups. Plasma bikunin peptidoglycosaminoglycan<sup>42</sup> was also examined using CID-FT-ICR-MS/MS, and the presence of the same sequence motif suggests that the observed bikunin sequence was not altered or modified during filtration by the kidney or through the action of catabolic enzymes such as  $\beta$ -glucuronidases or sulfatases (Supplementary Fig. 33, Supplementary Table 35). The number of possible sequences calculated for bikunin GAG chains decreased as experimental constraints increased (Supplementary Table 36), and the sequence of bikunin GAG was determined (Fig. 4 and Supplementary Table 37).

## DISCUSSION

Bikunin is a physiologically and pharmacologically important proteoglycan that is found both in human plasma and urine<sup>13–16</sup>. From a structural standpoint, bikunin is the simplest proteoglycan consisting of a 143-amino-acid core protein substituted at Asn45 with an N-linked complex biantennary glycan and substituted at Ser10 with a single O-linked chondroitin 4-sulfate chain with 23–55 saccharide units<sup>17–20</sup>. Based only on the size and composition of this GAG chain, there are 210 billion sequence possibilities (Supplementary Table 36). Previous studies from our laboratory<sup>23,42</sup> and others<sup>37,38</sup> have found fixed structural motifs at both the reducing and nonreducing ends of the bikunin GAG chains, observations that greatly decrease the expected number of sequences to 43 million (Supplementary Table 36). In the current study we attempted to completely sequence a single bikunin GAG chain using tandem MS to establish whether there were motifs present within the center of the GAG chain that might further decrease the large number of expected bikunin sequences.

High-resolution separation was required to prepare a simple mixture containing a single prominent bikunin GAG chain for

sequencing. Preliminary studies<sup>39,40,42</sup> suggested that continuous elution preparative PAGE, offering the highest resolving power currently available, might permit the purification and recovery of the multi-microgram quantities of sample enriched in one to several GAG chains with a single composition (a fixed chain length and fixed number of sulfo groups). PAGE fractionation can be performed on the proteoglycan, peptidoglycosaminoglycan or GAG. We chose to focus on the peptidoglycosaminoglycan, as it offered an advantage over the proteoglycan, whose core protein might reduce the dominant physical-chemical feature of the GAG chain, and over the GAG chain, whose preparation requires harsh  $\beta$ -elimination conditions that run the risk of introducing structurally complicating artifacts.

Next, we selected the optimal approach for GAG-chain sequencing. Our previous efforts involved a combination of bottom-up and top-down glycomics approaches that relied on the application of polysaccharide lyases to fragment chains and used the oligosaccharide sequences to discover an overall chain sequence<sup>23,26</sup>. There are limitations to this approach: not all the requisite enzymes are available to determine sequence<sup>26</sup>; the available enzymes do not have strict specificity<sup>26,43</sup>; most of the available enzymes are endolytic, making their use in direct sequencing difficult<sup>43,44</sup>; and the exolytic polysaccharide lyases are not purely exolytic and are capable of jumping resistant sites, making their use in sequencing difficult<sup>43</sup>. ESI-MS offers some potential advantages for GAG sequencing but also has limitations, including difficulties in analyzing full-length or highly sulfated GAG chains<sup>2</sup>; the special care required in ESI solvent selection and sample preparation because Na-H exchange can introduce additional spectral heterogeneity beyond composition (chain length and number of sulfo groups)<sup>41</sup>; the facile loss of sulfo groups, which complicates sequencing by giving fragmentation that is not sequence informative<sup>45,46</sup>; and the insufficient resolution and sensitivity of MS for determination of small amounts of sample or samples consisting of even simple mixtures<sup>23</sup>. We settled on an approach that uses optimized solvent, containing dilute formic acid to remove cations, which simplifies spectra and forms negative ions with high charge states ( $z$  values) and thus suppresses sulfo-group loss<sup>35,41</sup>. The direct infusion of simple mixtures containing bikunin peptidoglycosaminoglycan compositions (~1 nmol) on FT-MS showed 18–34 compositions with high  $z$  values, suggesting that CID might give sequence-informative fragmentation. These high  $z$ -values could have been anticipated on the basis of the modest number of sulfo groups present and the relatively large number of carboxyl groups associated with the polysaccharide analyte. In the past, more highly sulfated and shorter oligosaccharides have afforded lower  $z$  values with the CID spectra showing substantial loss of sulfo groups<sup>35</sup>. Parent ions were selected, and the CID of each showed a surprisingly simple spectrum suggesting that a single composition represented a single sequence. Comparison of the sequences of individual parent ions afforded a simple pattern or sequence motif (**Fig. 3b,c**). Thus, on the basis of our sampling of bikunin peptidoglycosaminoglycan we anticipate that each of the ~150 compositions has a single sequence. Results from CID-FT-ICR-MS/MS provided enhanced mass accuracy, resolving power and sensitivity compared to CID-FT-MS/MS (requiring ~1 pmol of bikunin peptidoglycosaminoglycan PAGE fraction). CID-FT-ICR-MS/MS and CID-FT-MS/MS of the same precursor ion were different but gave identical sequences. To our knowledge, this manuscript is the first report of the concurrent use of two types of FT-MS instruments, and the excellent data provided by the FT-MS, previously thought to be obtainable only by sophisticated FT-ICR-MS instruments, suggests that this sequencing approach will be generally available to the glycobiology community on more widely available spectrometers.

Bikunin peptidoglycosaminoglycan analysis pointed to the greater abundance of odd-numbered chains, which occurs probably because their formation naturally occurs during biosynthesis. We initially hypothesized that the prominence of GalNAc-terminated

odd chains resulted from the removal of GlcA from the nonreducing end by  $\beta$ -D-glucuronidase, however, this is unlikely as plasma bikunin shows even- and odd-chain distribution identical to that of urinary bikunin<sup>42</sup>. A previous study, whose suggestion that the nonreducing end was monosulfated was based on the recovery of a saturated monosulfated trisaccharide following enzymatic digestion<sup>23</sup>, apparently missed the major saturated nonsulfated trisaccharide formed from the nonreducing end of the odd-numbered chondroitin sulfate chain.

The biosynthetic assembly of bikunin's chondroitin sulfate chain occurs within the Golgi complex, and bikunin is secreted by hepatocytes. In chondroitin sulfate biosynthesis, sugar residues are sequentially attached by specific glycosyl transferases to first form the linkage region (**Fig. 1**), then GAG elongation occurs with independent and alternating addition of GalNAc and GlcA, and sulfonation occurs during elongation. The kinetics of the biosynthetic events may perhaps also have a role in defining domain structure, and catabolic or anabolic processing may be important in stopping biosynthesis. It seems that there is some means whereby defined glycan domain structures are introduced, despite the apparent absence of an instructive template. Two ordered domains were identified in the bikunin GAG sequence: a sulfated domain near the reducing end and a nonsulfated domain at the nonreducing end. In the sulfated domain, 12 residues adjacent to the tetrasaccharide linkage region showed a single sequence motif. Five residues, which constitute a motif, follow this domain but with variable sulfo-group occupancy. The nonsulfated domain at the nonreducing end varied in length from ~6–22 residues.

One biological function of the nonsulfated domain is that it binds protein heavy chains (HC1, HC2) through ester linkages. HC1 and HC2 covalently linked to bikunin bind inter- $\alpha$ -inhibitor protein to form a serine proteinase inhibitor complex. Plasma and urine concentrations of free and complexed bikunin are related to its anti-inflammatory activity. The domain structure of the bikunin GAG may be critical for biological switching of covalently bound HC1 and HC2 from bikunin chondroitin sulfate to hyaluronan<sup>15,17,18,47</sup>.

Chain length and molecular mass do not seem to be the major limitations in the application of FT-MS for sequencing GAGs. We have analyzed single chains of up to 80 saccharide units<sup>40</sup>. Instead, increasing molecular mass and sulfation level leads to more compositions and greater challenges in fractionation. In continuous elution PAGE of bikunin peptidoglycosaminoglycan we observe 10–30 compositions in each band. When we select a single composition for sequencing by MS/MS, it might represent ~10 mol % of the mixture. Moreover, as the number of sulfo groups increases it becomes difficult to obtain molecular ions in which the charge state is greater than the number of sulfo groups, resulting in a reduced number of sequence-informative fragments. Improved fractionation and FT-ICR-MS/MS methods may allow the application of these methods for more variable and highly sulfated GAGs in the future.

To put these results in perspective, one must consider how improbable is it that a proteoglycan has a single sequence motif when there are so many possible sequence possibilities (**Supplementary Table 36**), despite the fact that there is no known biosynthetic mechanism that could explain how this sequence is installed. The common biosynthetic pathway for all proteoglycan families suggests that other defined sequences may even be present in the more structurally complex proteoglycans, such as heparan sulfate. These findings clearly require a readjustment of our understanding of sequence and information storage in glycans that have complex structure.

## METHODS

Preparation and linkage region analysis of bikunin peptidoglycosaminoglycan. Pharmaceutical-grade bikunin proteoglycan (Mochida Pharmaceuticals) was purified from excipients and buffer salts by dialysis against distilled water using

30-kDa molecular weight centrifugal devices (Millipore) and then was lyophilized. The proteoglycan (7 mg) was then proteolyzed by a 5% (w/w) actinase E (EC 3.4.24.4, Kaken Biochemicals) digestion at pH 7.5 in 50 mM Tris-HCl in sodium acetate (Sigma). The enzymatic reaction proceeded at 45 °C for 18 h and was then isolated from the digestion mixture by strong-anion exchange spin column chromatography (High-Capacity Maxi-Q, Sartorius). Anion exchange spin columns were pre-equilibrated with 50 mM sodium chloride, and after the digestion mixture was brought up to 50 mM, the samples were bound by centrifugation at 500g. Washed twice with 50 mM sodium chloride, the peptidoglycosaminoglycan was eluted with 1.5 M NaCl, desalted using a YM-10 kDa MWCO centrifugal filter (Millipore) with deionized water washes and then lyophilized. The actinase E digestion and purification were repeated. Approximately 50 µg of bikunin peptidoglycosaminoglycan was digested completely using 250 mU chondroitin sulfate lyase ABC from *Proteus vulgaris* (EC 4.2.2.4, Seikagaku America) at 37 °C for 18 h and was then purified by a 10-kDa-MWCO spin column filter to isolate reducing ends.

**Fractionation of bikunin peptidoglycosaminoglycan by continuous elution PAGE.** Electrophoresis resolving buffer, electrode running buffer and total acrylamide monomer solutions, including resolving and stacking gel solutions, were prepared as previously described<sup>39</sup>. A gel of 10-cm column height using 15% total acrylamide monomer resolving solution was allowed to polymerize overnight with 40 µl TEMED and 200 µl 10% (w/v) ammonium persulfate and was cast in a 37-mm diameter column (Bio-Rad). Above the polymerized resolving gel column, 4 ml of 5% total acrylamide monomer stacking gel was cast. An aliquot of 2 mg purified bikunin peptidoglycosaminoglycan was loaded in a solution of 10 µg ml<sup>-1</sup> (w/v) phenol red and 25% (w/v) sucrose. Electrophoresis was performed for 8 h at a constant power of 12 W, with a peristaltic pump (Econo Pump, Bio-Rad) set to 0.7 ml min<sup>-1</sup> and fraction collector (Model 2110, Bio-Rad) set to 2 min in conjunction-accumulated separating fractions from the Model 491 Prep cell (Bio-Rad). Buffer salts from electrophoresis for each fraction were removed by strong anion exchange (Medium-Capacity Mini-Q, Sartorius) and thoroughly desalted by LC-grade water washes with Microcon YM-10 centrifugal filters (Millipore). Carbazole assay<sup>48</sup> quantified the amount of bikunin peptidoglycosaminoglycan per fraction based on a standard curve of chondroitin sulfate A (Celsus). The extent of separation was visualized by 15% total acrylamide monomer solution using native mini-slab PAGE stained with Alcian blue, and molecular mass distribution properties ( $M_R$ ,  $M_N$  and  $M_W$ ) of the fractionated and unfractionated peptidoglycosaminoglycans were estimated on PAGE densitometry against identified bikunin standards using UN-SCANIT (Silk Scientific)<sup>40</sup>.

**ESI-LTQ-Orbitrap-FT-MS analysis of bikunin peptidoglycosaminoglycan.** All peptidoglycosaminoglycans were analyzed in the negative-ion mode by electrospray ionization FT-MS 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 methanol with 0.1% formic acid and were delivered by an Agilent 1200 nano-LC pump at a flow rate of 20 µl min<sup>-1</sup>. The purified bikunin peptidoglycosaminoglycans at volumes between 0.5 µl and 5 µl were directly infused through an Agilent 1200 autosampler. Mass spectra were acquired at a resolution between 30,000 and 60,000, and the charge deconvolution was performed manually with electronic spreadsheets. Acquisition parameters used to prevent in-source fragmentation included spray voltage 3–4.2 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 MS analysis of the linkage-region peptidoglycosaminoglycans and intact peptidoglycosaminoglycans, precursor ions were isolated and activated in the linear ion trap for 800 µs before injection into the orbitrap for mass analysis. Parent ions were fragmented by low-energy collision (25 eV to 65 eV) with a 3 *m/z* mass window. Acquired data were averaged to produce a single mass spectrum per fraction sample in MS and MS/MS. Fragments and peptidoglycosaminoglycans were identified manually using Glycoworkbench<sup>49</sup>. CID spectra are annotated using the Domon and Costello nomenclature with sulfo group loss shown by roman numerals<sup>50</sup>.

**Nano-ESI-FT-ICR-MS analysis of bikunin peptidoglycosaminoglycan.** FT-ICR-MS experiments were performed in negative-ion mode with a 9.4T acrylamide Bruker Apex Ultra QeFTMS (Bruker) fitted with an Apollo II dual source. The sample solutions were infused at a rate of 5–10 µl per hour and were ionized by nanoelectrospray using a pulled fused silica-tip model FS360-75-15-D-20 (New Objective). Solutions of each bikunin peptidoglycosaminoglycan fraction were introduced at a concentration of ~0.1 mg ml<sup>-1</sup> in 50:50:0.1 methanol/H<sub>2</sub>O/FA (Sigma) to minimize Na-H heterogeneity<sup>41</sup>.

During mixture analysis for urinary bikunin peptidoglycosaminoglycan, a quadrupole mass filter window of 50 *m/z* was advanced across the mass range of interest (400–2000 *m/z*)<sup>23</sup>. For each window, a mass spectrum was acquired that consisted of 24 signal-averaged acquisitions. For each mass spectrum, 1M points were acquired, padded with one zero fill and apodized using a sinebell window. Independent windowed spectra were then combined to generate a composite mass spectrum. Mass spectra for purified fractions were acquired without the use of the quadrupole mass filter. For each mass

spectrum, 24 acquisitions consisting of 1M points were signal averaged, padded with one zero fill and apodized using a sinebell window.

For tandem mass spectrometry experiments, mass-selected precursor ions were activated by CID in the hexapole collision cell of the Apex instrument with collision energy of 16–28 eV. The 36–48 acquisitions were signal averaged for each tandem mass spectrum during the analysis of urinary peptidoglycosaminoglycan chains. Because of limited material during analysis of plasma bikunin peptidoglycosaminoglycan chains, 148 acquisitions were signal averaged. External calibration of mass spectra produced mass accuracy of 1–2 ppm. Internal calibration was also performed using confidently assigned glycosidic bond cleavage products as internal calibrants, providing mass accuracy of <1 ppm. Product-ion assignment and annotation were identical to that used in FTMS<sup>49,50</sup>.

See **Supplementary Results** for additional mass spectra, sequence information and quantification of bikunin peptidoglycosaminoglycan chains.

Received 6 June 2011; accepted 30 July 2011;  
published online 9 October 2011

## References

- Feero, W.G., Gutmacher, A.E. & Collins, F.S. Genomic medicine—an updated primer. *N. Engl. J. Med.* **362**, 2001–2011 (2010).
- Ly, M., Laremore, T.N. & Linhardt, R.J. Proteoglycomics: recent progress and future challenges. *OMICS* **14**, 389–399 (2010).
- Kreuger, J., Spillmann, D., Li, J. & Lindahl, U. Interactions between heparan sulfate and proteins: the concept of specificity. *J. Cell Biol.* **174**, 323–327 (2006).
- Ball, S. *et al.* From glycogen to amylopectin: a model for the biogenesis of the plant starch granule. *Cell* **86**, 349–352 (1996).
- Perez, S. & Mazeau, K. Conformations, structures, and morphologies of celluloses. in *Polysaccharides: Structural diversity and functional versatility* 2nd edn (ed. Dumitriu, S.) 41–68 (Marcel Dekker, 1998).
- Esko, J.D. & Selleck, S.B. Order out of chaos: assembly of ligand binding sites in heparan sulfate. *Annu. Rev. Biochem.* **71**, 435–471 (2002).
- Silbert, J.E. & Sugumar, G. Biosynthesis of chondroitin/dermatan sulfate. *IUBMB Life* **54**, 177–186 (2002).
- Nairn, A.V. *et al.* Glycomics of proteoglycan biosynthesis in murine embryonic stem cell differentiation. *J. Proteome Res.* **6**, 4374–4387 (2007).
- Couchman, J.R. Transmembrane signaling proteoglycans. *Annu. Rev. Cell Dev. Biol.* **26**, 89–114 (2010).
- Linhardt, R.J. & Toida, T. Role of glycosaminoglycans in cellular communication. *Acc. Chem. Res.* **37**, 431–438 (2004).
- Atha, D.H., Lormeau, J.C., Petitou, M., Rosenberg, R.D. & Choay, J. Contribution of 3-O- and 6-O-sulfated glucosamine residues in the heparin-induced conformational change in antithrombin III. *Biochemistry* **26**, 6454–6461 (1987).
- Petitou, M. & van Boeckel, C.A. A synthetic antithrombin III binding pentasaccharide is now a drug! What comes next? *Angew. Chem. Int. Edn Engl.* **43**, 3118–3133 (2004).
- Fries, E. & Blom, A.M. Bikunin—not just a plasma proteinase inhibitor. *Int. J. Biochem. Cell Biol.* **32**, 125–137 (2000).
- Fries, E. & Kaczmarczyk, A. Inter-alpha-inhibitor, hyaluronan and inflammation. *Acta Biochim. Pol.* **50**, 735–742 (2003).
- Zhuo, L., Salustri, A. & Kimata, K. A physiological function of serum proteoglycan bikunin: the chondroitin sulfate moiety plays a central role. *Glycoconj. J.* **19**, 241–247 (2002).
- Michalski, C. *et al.* Preparation and properties of a therapeutic inter-alpha-trypsin inhibitor concentrate from human plasma. *Vox Sang.* **67**, 329–336 (1994).
- Engchild, J.J. *et al.* Chondroitin 4-sulfate covalently cross-links the chains of the human blood protein pre-alpha-inhibitor. *J. Biol. Chem.* **266**, 747–751 (1991).
- Morelle, W. *et al.* Chondroitin sulphate covalently cross-links the three polypeptide chains of inter-alpha-trypsin inhibitor. *Eur. J. Biochem.* **221**, 881–888 (1994).
- Zhuo, L., Hascall, V.C. & Kimata, K. Inter-alpha-trypsin inhibitor, a covalent protein-glycosaminoglycan-protein complex. *J. Biol. Chem.* **279**, 38079–38082 (2004).
- Engchild, J.J. *et al.* Organization of the inter-alpha-inhibitor heavy chains on the chondroitin sulfate originating from Ser(10) of bikunin: posttranslational modification of IalphaI-derived bikunin. *Biochemistry* **38**, 11804–11813 (1999).
- Josic, D. *et al.* Proteomic characterization of inter-alpha inhibitor proteins from human plasma. *Proteomics* **6**, 2874–2885 (2006).
- Delaria, K.A. *et al.* Characterization of placental bikunin, a novel human serine protease inhibitor. *J. Biol. Chem.* **272**, 12209–12214 (1997).
- Chi, L. *et al.* Structural analysis of bikunin glycosaminoglycan. *J. Am. Chem. Soc.* **130**, 2617–2625 (2008).
- Conrad, H.E. Beta-elimination for release of O-linked glycosaminoglycans from proteoglycans. *Curr. Protoc. Mol. Biol.* **17**, 15.1–15.3 (2001).
- Venkataraman, G., Shriver, Z., Raman, R. & Sasisekharan, R. Sequencing complex polysaccharides. *Science* **286**, 537–542 (1999).
- Laremore, T.N. *et al.* Domain structure elucidation of human decorin glycosaminoglycans. *Biochem. J.* **431**, 199–205 (2010).

27. Wolff, J.J., Amster, I.J., Chi, L. & Linhardt, R.J. Electron detachment dissociation of glycosaminoglycan tetrasaccharides. *J. Am. Soc. Mass Spectrom.* **18**, 234–244 (2007).
28. Wolff, J.J., Laremore, T.N., Aslam, H., Linhardt, R.J. & Amster, I.J. Electron-induced dissociation of glycosaminoglycan tetrasaccharides. *J. Am. Soc. Mass Spectrom.* **19**, 1449–1458 (2008).
29. Wolff, J.J. *et al.* Negative electron transfer dissociation of glycosaminoglycans. *Anal. Chem.* **82**, 3460–3466 (2010).
30. Turnbull, J.E., Hopwood, J.J. & Gallagher, J.T. A strategy for rapid sequencing of heparan sulfate and heparin saccharides. *Proc. Natl. Acad. Sci. USA* **96**, 2698–2703 (1999).
31. Merry, C.L.R., Lyon, M., Deakin, J.A., Hopwood, J.J. & Gallagher, J.T. Highly sensitive sequencing of the sulfated domains of heparan sulfate. *J. Biol. Chem.* **274**, 18455–18462 (1999).
32. Turnbull, J.E. & Gallagher, J.T. Sequence analysis of heparan sulphate indicates defined location of *N*-sulphated glucosamine and iduronate 2-sulphate residues proximal to the protein linkage region. *Biochem. J.* **277**, 297–303 (1991).
33. Liu, J., Desai, U.R., Han, X.-J., Toida, T. & Linhardt, R.J. Strategy for the sequence analysis of heparin. *Glycobiology* **5**, 765–774 (1995).
34. Zaia, J., Li, X.Q., Chan, S.Y. & Costello, C.E. Tandem mass spectrometric strategies for determination of sulfation positions and uronic acid epimerization in chondroitin sulfate oligosaccharides. *J. Am. Soc. Mass Spectrom.* **14**, 1270–1281 (2003).
35. McClellan, J.E., Costello, C.E., O'Connor, P.B. & Zaia, J. Influence of charge state on product ion mass spectra and the determination of 4S/6S sulfation sequence of chondroitin sulfate oligosaccharides. *Anal. Chem.* **74**, 3760–3771 (2002).
36. Hitchcock, A.M., Yates, K.E., Costello, C.E. & Zaia, J. Comparative glycomics of connective tissue glycosaminoglycans. *Proteomics* **8**, 1384–1397 (2008).
37. Toyoda, H., Kobayashi, S., Sakamoto, S., Toida, T. & Imanari, T. Structural analysis of a low-sulfated chondroitin sulfate chain in human urinary trypsin inhibitor. *Biol. Pharm. Bull.* **16**, 945–947 (1993).
38. Yamada, S. *et al.* The sulphated carbohydrate-protein linkage region isolated from chondroitin 4-sulphate chains of inter-alpha-trypsin inhibitor in human plasma. *Glycobiology* **5**, 335–341 (1995).
39. Laremore, T.N., Ly, M., Solakyildirim, K., Zagorevski, D.V. & Linhardt, R.J. High-resolution preparative separation of glycosaminoglycan oligosaccharides by polyacrylamide gel electrophoresis. *Anal. Biochem.* **401**, 236–241 (2010).
40. Ly, M. *et al.* Analysis of *E. coli* K5 capsular polysaccharide heparosan. *Anal. Bioanal. Chem.* **399**, 737–745 (2011).
41. Wolff, J.J., Laremore, T.N., Busch, A.M., Linhardt, R.J. & Amster, I.J. Influence of charge state and sodium cationization on the electron detachment dissociation and infrared multiphoton dissociation of glycosaminoglycan oligosaccharides. *J. Am. Soc. Mass Spectrom.* **19**, 790–798 (2008).
42. Laremore, T.N., Leach, F.E. III, Amster, I.J. & Linhardt, R.J. Electrospray ionization Fourier transform mass spectrometric analysis of intact bikunin glycosaminoglycan from normal human plasma. *Int. J. Mass. Spectrom.* **305**, 109–115 (2011).
43. Ly, M. *Glycosaminoglycan Sequencing of Proteoglycan*. PhD thesis, Rensselaer Polytechnic Institute (2011).
44. Gu, K., Liu, J., Pervin, A. & Linhardt, R.J. Comparison of the activity of two chondroitin AC lyases on dermatan sulfate. *Carbohydr. Res.* **244**, 369–377 (1993).
45. Zaia, J. Principles of mass spectrometry of glycosaminoglycans. *J. Biomacromol. Mass. Spec.* **1**, 3–36 (2005).
46. Gunay, N.S., Tadano-Aritomi, K., Toida, T., Ishizuka, I. & Linhardt, R.J. Evaluation of counterions for electrospray ionization mass spectral analysis of a highly sulfated carbohydrate, sucrose octasulfate. *Anal. Chem.* **75**, 3226–3231 (2003).
47. Capon, C., Mizon, C., Lemoine, J., Rodié-Talbère, P. & Mizon, J. In acute inflammation, the chondroitin-4 sulphate carried by bikunin is not only longer; it is also undersulphated. *Biochimie* **85**, 101–107 (2003).
48. Bitter, T. & Muir, H.M. A modified uronic acid carbazole reaction. *Anal. Biochem.* **4**, 330–334 (1962).
49. Ceroni, A. *et al.* GlycoWorkbench: a tool for the computer-assisted annotation of mass spectra of glycans. *J. Proteome Res.* **7**, 1650–1659 (2008).
50. Domon, B. & Costello, C. A systematic nomenclature for carbohydrate fragmentations in FAB-MS/MS spectra of glycoconjugates. *Glycoconj. J.* **5**, 397–409 (1988).

### Acknowledgments

The authors thank D. Zagorevski for his expertise in the proteomics core at Rensselaer Polytechnic Institute and the US National Institutes of Health for support (GM38060).

### Author contributions

M.L. and F.E.L. III contributed experiments and data interpretation. T.N.L. contributed the fractions for analysis and assisted in writing. T.T. contributed the bikunin and assisted in writing. R.J.L. and I.J.A. contributed experimental planning, result interpretation and wrote the paper.

### Competing financial interests

The authors declare no competing financial interests.

### Additional information

Supplementary information is available online at <http://www.nature.com/naturechemicalbiology/>. Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>. Correspondence and requests for materials should be addressed to R.J.L.