

# Sequencing of 3-O sulfate containing heparin decasaccharides with a partial antithrombin III binding site

Zachary Shriver\*, Rahul Raman\*, Ganesh Venkataraman<sup>†</sup>, Katherine Drummond<sup>‡</sup>, Jeremy Turnbull<sup>‡</sup>, Toshihiko Toida<sup>§</sup>, Robert Linhardt<sup>¶</sup>, Klaus Biemann<sup>||</sup>, and Ram Sasisekharan<sup>\*,\*\*\*††</sup>

\*Division of Bioengineering and Environmental Health, <sup>†</sup>Harvard-Massachusetts Institute of Technology Division of Health Sciences and Technology, <sup>||</sup>Department of Chemistry, and <sup>\*\*</sup>Center for Biomedical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139; <sup>‡</sup>School of Biosciences, University of Birmingham, Edgbaston, Birmingham B15 2TT, United Kingdom; <sup>¶</sup>Department of Chemistry, Division of Medicinal and Natural Products Chemistry, and Department of Chemical and Biochemical Engineering, University of Iowa, Iowa City, IA 52242; and <sup>§</sup>Faculty of Pharmaceutical Sciences, Chiba University, Chiba 263, Japan

Contributed by Klaus Biemann, June 22, 2000

Heparin- and heparan sulfate-like glycosaminoglycans (HLGAGs) represent an important class of molecules that interact with and modulate the activity of growth factors, enzymes, and morphogens. Of the many biological functions for this class of molecules, one of its most important functions is its interaction with antithrombin III (AT-III). AT-III binding to a specific heparin pentasaccharide sequence, containing an unusual 3-O sulfate on a N-sulfated, 6-O sulfated glucosamine, increases 1,000-fold AT-III's ability to inhibit specific proteases in the coagulation cascade. In this manner, HLGAGs play an important biological and pharmacological role in the modulation of blood clotting. Recently, a sequencing methodology was developed to further structure-function relationships of this important class of molecules. This methodology combines a property-encoded nomenclature scheme to handle the large information content (properties) of HLGAGs, with matrix-assisted laser desorption ionization MS and enzymatic and chemical degradation as experimental constraints to rapidly sequence picomole quantities of HLGAG oligosaccharides. Using the above property-encoded nomenclature-matrix-assisted laser desorption ionization approach, we found that the sequence of the decasaccharide used in this study is  $\Delta U_{25}H_{NS,6S}I_{2S}H_{NS,6S}I_{2S}H_{NS,6S}I_{2S}H_{NAC,6S}GH_{NS,3S,6S}$  ( $\pm$ DDD4-7). We confirmed our results by using integral glycan sequencing and one-dimensional proton NMR. Furthermore, we show that this approach is flexible and is able to derive sequence information on an oligosaccharide mixture. Thus, this methodology will make possible both the analysis of other unusual sequences in HLGAGs with important biological activity as well as provide the basis for the structural analysis of these pharmacologically important group of heparin/heparan sulfates.

Heparin- and heparan sulfate-like glycosaminoglycans (HLGAGs), present both at the cell surface and in the extracellular matrix, are a group of complex polysaccharides that are variable in length, consisting of a disaccharide repeat unit composed of glucosamine and an uronic acid (either iduronic or glucuronic acid). The high degree of complexity for HLGAGs arises not only from their polydispersity and the possibility of two different uronic acid components, but also from differential modification at four positions of the disaccharide unit. Three positions, namely, C2 of the uronic acid and the C3, C6 positions of the glucosamine can be O-sulfated. In addition, C2 of the glucosamine can be N-acetylated or N-sulfated. Together, these modifications theoretically could lead to 32 possible disaccharide units, making HLGAGs potentially more information dense than either DNA (four bases) or proteins (20 aa). This enormity of possible structural variants allows HLGAGs to be involved in a large number of diverse biological processes, including angiogenesis (1), embryogenesis (2-4), and the formation of  $\beta$ -fibrils in Alzheimer's disease (5, 6).

This structural diversity is one of the factors that has made it difficult to study sequence-function relationships for HLGAGs. Chemical synthesis of defined oligosaccharides has been used in

studying the relative contribution to high-affinity antithrombin III (AT-III) binding of specific modifications in the pentasaccharide sequence (7). However, such synthetic methods are complex and have not been widely applied to the study of other biological sequences. An alternative approach involving affinity fractionation of HLGAGs with proteins of interest and subsequent characterization has provided some overall information regarding sulfation patterns that determine affinity (8-10). However, until recently there have been no effective methods to sequence these saccharides.

Here, we outline the sequence determination of an AT-III-fractionated decasaccharide (AT-10) by using a property-encoded nomenclature (PEN)/matrix-assisted laser desorption ionization (MALDI)-MS scheme, a sequencing methodology described earlier (11). Integral glycan sequencing (IGS) (12) and proton NMR (<sup>1</sup>H NMR) analysis of the decasaccharide are consistent with the results of PEN-MALDI sequencing. The flexibility of this sequencing strategy also is demonstrated by the fact that we can derive sequence information for contaminating oligosaccharides, if present. Sequencing of a chemically complex AT-III-fractionated saccharide (including the rare 3-O-sulfation of glucosamine) establishes a methodology that can be extended to the analysis of other HLGAG oligosaccharides of interest, for example those with growth factor binding properties. A straightforward sequencing methodology for these types of sequences should further enable structure-function studies of this important class of molecules.

## Materials and Methods

**Materials.** The decasaccharide AT-10 is the same saccharide used in previous studies (13, 14). Oligosaccharides were dissolved in deionized water at concentrations of 10-35  $\mu$ M. Heparinase I-III from *Flavobacterium heparinum* were purified as described (11). The exoenzymes  $\alpha$ -L-iduronate 2-O sulfatase,  $\alpha$ -L-iduronidase,  $\beta$ -D-glucuronidase, and N-acetylglucosamine-6-sulfatase were purchased from Oxford Glycosciences, Rosedale, NY. A 40% aqueous solution of sodium nitrite was purchased

Abbreviations: HLGAG, heparin- and heparan sulfate-like glycosaminoglycan; AT-III, antithrombin III; AT-10, AT-III-fractionated decasaccharide; IGS, integral glycan sequencing; PEN, property encoded nomenclature; MALDI, matrix-assisted laser desorption ionization; CE, capillary electrophoresis;  $\Delta U$ ,  $\Delta^{4,5}$  uronic acid; 2S, 3S, and 6S, 2-O, 3-O, and 6-O sulfation, respectively; NS and NAC, N-sulfation and N-acetylation of the glucosamine, respectively.

See commentary on page 10301.

<sup>††</sup>To whom reprint requests should be addressed at: Massachusetts Institute of Technology, 77 Massachusetts Avenue, Building 16-561, Cambridge, MA 02139. E-mail: rams@mit.edu.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

from Aldrich Chemical, Metuchen, NJ. Disaccharide standards for compositional analysis were purchased from Sigma.

**Compositional Analysis.** Compositional analysis of oligosaccharides was completed by exhaustive digest of a 30  $\mu\text{M}$  sample of AT-10 followed by capillary electrophoresis (CE) as described (15). Briefly, to 1 nmol of oligosaccharide was added 200 nM of heparinases I, II, and III in 25 mM sodium acetate, 100 mM NaCl, 5 mM calcium acetate buffer, pH 7.0. The reaction was allowed to proceed at 30°C overnight and then analyzed by CE in reverse polarity with a running buffer of 50 mM Tris/phosphate/10  $\mu\text{M}$  dextran sulfate, pH 2.5.

**Digests.** Heparinase I digestions were designated either short or exhaustive. For short digestion, 50 nM heparinase I was incubated with the substrate for 10 min before analysis. Exhaustive digestions were completed with 200 nM enzyme overnight. Enzyme reactions were performed by adding 1  $\mu\text{l}$  of enzyme solution in a buffer containing 10  $\mu\text{M}$  ovalbumin, 1  $\mu\text{M}$  dextran sulfate, 5 mM calcium acetate, and 10 mM ethylenediamine buffer at pH 7.0 to 4  $\mu\text{l}$  of aqueous substrate solution; digestion was allowed to proceed at room temperature as described (11, 15). Partial nitrous acid cleavage was completed by using a modification of published procedures (12). Exoenzyme digests were completed either simultaneously or sequentially. Final enzyme concentrations were in the range of 20 to 40 milliunits/ml and digestion was carried out at 37°C.

**MS.** Mass spectral analyses were carried out on a PerSeptive Biosystems (Framingham, MA) Voyager Elite reflectron time-of-flight instrument in the linear mode with delayed extraction. Samples from digests were prepared by removing 0.5  $\mu\text{l}$  of the reaction mixture and adding it to 4.5  $\mu\text{l}$  of matrix solution (12 mg/ml caffeic acid in 30% acetonitrile) that contained a 2-fold molar excess of the basic peptide (RG)<sub>19</sub>R [calculated mass of the (M + H)<sup>+</sup> ion = 4226.8]. Addition of the basic peptide to specifically chelate HLGAG oligosaccharides and mass spectral collection parameters allowed for direct sample analysis without need for sample repurification (15). Samples were spotted on the target, and mass spectra were collected by using parameters outlined previously (15). Observed in each mass spectrum were the (M + H)<sup>+</sup> ions of the basic peptide and the (M + H)<sup>+</sup> ion of a 1:1 peptide/saccharide complex, and the mass of the saccharide was determined by subtracting the measured  $m/z$  value of the (M + H)<sup>+</sup> ion of the peptide from that of the 1:1 complex (16). All spectra on a plate were calibrated externally by using a standard of (RG)<sub>19</sub>R and its complex with a nitrous acid-derived hexasaccharide of the sequence I<sub>25</sub>H<sub>NS,6S</sub>I<sub>25</sub>H<sub>NS,6S</sub>I<sub>25</sub>Man<sub>6S</sub> (calculated mass of 1655.4) under identical instrumental parameters. This methodology requires sufficient sulfation of the saccharide to ensure efficient complexation. As such, small, undersulfated saccharides (i.e., mono- and disaccharides) are not observed with this methodology (16).

**IGS.** IGS using electrophoretic separation was carried out as described (12). Partial nitrous acid cleavage conditions were modified by using 25 mM HCl and 2.5 mM sodium nitrite and stop time points of 5, 10, 20, 30, 120, and 240 min.

**<sup>1</sup>H NMR Spectroscopy.** <sup>1</sup>H NMR spectroscopy was performed by using the conditions described previously (17). AT-10 was subjected to ion-exchange chromatography to remove paramagnetic impurities. A column (1 cm × 10 cm) of AG 50W-X8 (Bio-Rad) was converted into sodium form by treatment with 5 ml of 0.1 M NaOH and washed with water for 12 h before use. The sample for NMR experiments was applied to the column, eluted with 20 ml of water, and freeze-dried. The sample ( $\approx$  1 mg) then was freeze-dried three times from 99.8% D<sub>2</sub>O (Merck) and dissolved in 0.5 ml of 100% D<sub>2</sub>O (Aldrich) for NMR spectroscopy in a 5-mm tube. One-dimensional <sup>1</sup>H NMR spectroscopy of AT-10

was performed on a JEOL GSX 500A spectrometer equipped with a 5-mm field gradient tunable probe at 298 K.

## Results and Discussion

**Introduction to Sequencing Methodology.** Recently a MALDI-MS technique enabling the determination of the mass of HLGAG complex oligosaccharides (from di to decasaccharides) to an accuracy of better than  $\pm$  1 Da was developed (16, 18). Because of the accuracy of the resulting molecular mass measurement of the individual HLGAGs, a unique assignment for both the length of a fragment and the number and kind of substituents is theoretically possible, provided the oligosaccharide is a tetrade-casaccharide or smaller (11). In addition, MALDI-MS can detect oligosaccharide fragments generated upon enzymatic or chemical degradation of an oligosaccharide (13–15). Finally, the sensitivity of MALDI-MS is such that as little as 100 femtomoles of material can be readily detected.

In addition to the MALDI-MS experimental technique, a PEN system for representing the 32 disaccharide units using hexadecimal coding was developed (11). The development of PEN is necessary to handle the large information content (properties) of HLGAGs. Each of the hexadecimal numbers is derived based on an internal logic that manifests itself in terms of the distribution of sulfates on a particular building block unit and is not randomly assigned simply to identify each disaccharide unit. This system is critical for HLGAGs in that it enables the rapid manipulation of sequences using simple mathematical or binary operations, hence providing a handle on the large information content of complex polysaccharides. In addition, the inherent structural diversity in HLGAGs necessarily arises from the property differences (location of charged sulfate and acetate groups), thereby making PEN a natural assignment scheme for HLGAGs. This system is in direct contrast to the alphabetic codes used to represent the nucleotides of DNA and amino acids of proteins that serve as mere identifiers and do not code for any information (properties) and do not capture the chemical heterogeneity of these biopolymers.

The hexadecimal coding system comprises the alphanumerals 0–9 and A–F. Because the disaccharide unit has four positions, namely 2-O, N, 3-O, and 6-O that can be modified, it is straightforward to assign each of the four binary digit positions of the hexadecimal code to one of these chemical positions. In addition, because there are only two modifications possible at each position (2-O, 3-O, and 6-O can either be sulfated or free, and the N position can be sulfated or acetylated), the use of a binary system captures these modifications as simple on or off states. [There are some rare HLGAG sequences with an unsubstituted N position, which can be accounted for in the PEN system by adding extra bits. However, in our studies, initial experiments including compositional analysis (see below) did not show the presence of free amine containing disaccharides.] For example, if, within a given disaccharide unit, the 2-O position is sulfated, then it is assigned the binary value of 1. Conversely, if the 2-O position on a given disaccharide is unsulfated, then it is assigned the binary value of 0. To identify a disaccharide with an alphanumeric character, the four binary positions have been assigned in the following manner: the 2-O position was assigned the leftmost binary position, followed by the 6-O, 3-O, and N position in that order.

To code for the isomeric state of the uronic acid (i.e., iduronic vs. glucuronic acid), we designated disaccharide units as + or –. In this way, it is possible to assign the positive hexadecimal codes to iduronic acid-containing units and the negative hexadecimal codes to glucuronic acid-containing units. Thus, disaccharide units with the same hexadecimal code but opposite signs possess the same sulfation pattern, differing only in the isomeric state of the uronic acid. Table 1 outlines the use of PEN for the disaccharide units present in this study.

Thus, the strategy for the sequence assignment of HLGAG oligosaccharides by PEN-MALDI essentially involves the fol-

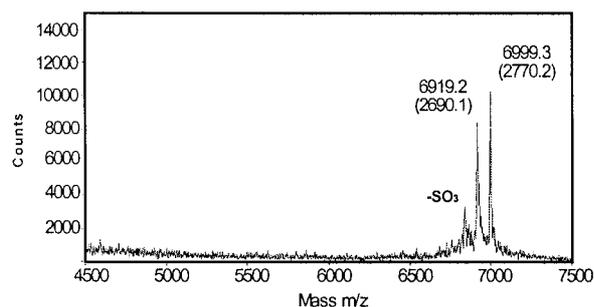
**Table 1. Derivation of PEN for disaccharide units used in this study**

I/G	2X	6X	3X	NX	Hex	Disacc	Mass
0	0	1	0	0	4	I-H <sub>NAC,6S</sub>	459.4
0	0	1	0	1	5	I-H <sub>NS,6S</sub>	497.4
0	1	1	0	1	D	I <sub>2S</sub> -H <sub>NS,6S</sub>	577.5
1	0	1	0	1	-5	G-H <sub>NS,6S</sub>	497.4
1	0	1	1	1	-7	G-H <sub>NS,3S,6S</sub>	577.5

The hexadecimal code derived for the disaccharide units occurring in AT-10. Column 1 is the binary position that codes for the isomeric state of the uronic acid. Columns 2–5 code for the modifications at the 2-O, 6-O, 3-O, and N positions of the disaccharide unit. Column 6 shows the hexadecimal codes represented by the binary digits in columns 2–5. Column 7 shows the disaccharide unit represented by the code in column 6. Column 8 shows the calculated theoretical masses of the disaccharide unit present internally in a sequence. For chemical or enzymatic modifications to these disaccharides, the following nomenclature is used: uronic acid with a  $\Delta^{4,5}$  unsaturated linkage ( $\Delta U$ ) =  $\pm$ ; reducing end disaccharide unit with a mass tag =  $\dagger$ ; disaccharide unit with a five-membered anhydromannose ring =  $\cdot$ .

lowing steps. First, MALDI-MS of the intact oligosaccharide is used to assign the length as well as the total number of sulfates and acetates present in the oligosaccharide. Compositional analysis then is used to determine the number and type of disaccharide building blocks. With this information, a master list is constructed of all possible sequences that contain those disaccharide units. In this manner, no sequences are excluded from the analysis, no matter how unusual a given sequence may be. The mass of oligosaccharide fragments generated from enzymatic digestion or chemical degradation are applied as experimental constraints and sequences that do not satisfy these constraints are eliminated. In an iterative manner, moving from experimental constraints to the ever-decreasing master list of possible sequences, one can rapidly arrive at a unique sequence solution by using a minimum of material. Importantly, multiple pathways, using separate experimental constraints, can be used to converge on a sequence, ensuring assignment accuracy.

**Analysis of AT-10.** AT-10 and all oligosaccharides derived from it either upon enzymatic or chemical treatment are detected with MALDI-MS as noncovalent complexes with the basic peptide



**Fig. 1.** MALDI mass spectrum of the protonated complex of AT-10 with (RG)<sub>19</sub>R.

(RG)<sub>19</sub>R (15, 16). Using this methodology, two species are observed, a  $(M + H)^+$  ion of (RG)<sub>19</sub>R and a  $(M + H)^+$  ion for the saccharide/peptide complex. The molecular mass of an oligosaccharide is obtained by subtracting the  $(M + H)^+$  value of the peptide from the  $(M + H)^+$  value of the 1:1 saccharide/peptide complex. Table 2 lists all fragments observed in this study, their calculated and experimentally derived mass values, and the deduced structure of the fragments after sequence assignment of AT-10. Fig. 1 shows that the major component of AT-10 has a  $m/z$  value of 6999.3. When the  $m/z$  value of the protonated peptide is subtracted, the experimental value for the mass of this oligosaccharide is found to be 2770.2, which can uniquely be assigned to a deca-saccharide with 13 sulfates and one acetate group. The mass spectrum of AT-10 indicates the presence of another species (hereafter referred to as the contaminant) of mass 2690.1 (after subtraction of the peptide contribution), corresponding to an oligosaccharide with 12 sulfates and one acetate group.

Compositional analysis using CE indicates the presence of four disaccharide building blocks, corresponding to  $\Delta U_{2S}\text{-H}_{NS,6S}$  ( $\pm D$ ),  $\Delta U\text{-H}_{NAC,6S}$  ( $\pm 4$ ),  $\Delta U\text{-H}_{NS,3S,6S}$  ( $\pm 7$ ), and  $\Delta U\text{-H}_{NS,6S}$  ( $\pm 5$ ), in the relative ratio of 2.90:1.00:1.05:0.15, respectively, where  $\Delta U$ , is a  $\Delta^{4,5}$  uronic acid; 2S, 3S, and 6S, are 2-O, 3-O, or 6-O sulfation, respectively; and NS and NAc are N-sulfation and N-acetylation of the glucosamine, respectively. Thus, compositional analysis of this sample confirms that there are two species, one major ( $\approx 85\%$ ) and one minor ( $\approx 15\%$ ). Accordingly, AT-10

**Table 2.  $m/z$  values for the peaks in the mass spectra and their deduced structures**

Complex $(M + H)^+$	Saccharide (observed)	Deduced structure	Mass (calculated)
6999.3	2770.2	$\Delta U_{2S}\text{H}_{NS,6S}\text{I}_{2S}\text{H}_{NS,6S}\text{I}\text{H}_{NAC,6S}\text{GH}_{NS,3S,6S}$ (Fig. 1)	2769.3
6919.2	2690.1	$\Delta U_{2S}\text{H}_{NS,6S}\text{I}/\text{GH}_{NS,6S}\text{I}_{2S}\text{H}_{NS,6S}\text{I}\text{H}_{NAC,6S}\text{GH}_{NS,3S,6S}$ (Fig. 1)	2689.2
6899.6	2673.0	$\Delta U_{2S}\text{H}_{NS,6S}\text{I}_{2S}\text{H}_{NS,6S}\text{I}_{2S}\text{H}_{NS,6S}\text{I}\text{H}_{NAC,6S}\text{GMan}_{3S,6S}$ (Fig. 6)	2672.2
6435.8	2209.2	$\text{I}_{2S}\text{H}_{NS,6S}\text{I}_{2S}\text{H}_{NS,6S}\text{I}\text{H}_{NAC,6S}\text{GH}_{NS,3S,6S}$ (Fig. 6)	2209.8
6419.7	2192.2	$\Delta U_{2S}\text{H}_{NS,6S}\text{I}_{2S}\text{H}_{NS,6S}\text{I}\text{H}_{NAC,6S}\text{GH}_{NS,3S,6S}$ (Fig. 2A)	2191.8
6339.8	2113.2	$\text{I}_{2S}\text{H}_{NS,6S}\text{I}_{2S}\text{H}_{NS,6S}\text{I}\text{H}_{NAC,6S}\text{GMan}_{3S,6S}$ (Fig. 6)	2112.7
5899.9	1671.4	$\Delta U_{2S}\text{H}_{NS,6S}\text{I}\text{H}_{NAC,6S}\text{GH}_{NS,3S,6S}$ (mass tagged, Fig. 2C)	1670.4
5859.8	1633.2	$\text{I}_{2S}\text{H}_{NS,6S}\text{I}\text{H}_{NAC,6S}\text{GH}_{NS,3S,6S}$ (Fig. 6)	1632.3
5842.1, 5842.2, 5843.6	1614.6, 1614.4, 1615.1	$\Delta U_{2S}\text{H}_{NS,6S}\text{I}\text{H}_{NAC,6S}\text{GH}_{NS,3S,6S}$ (Fig. 2A–C)	1614.3
5383.1, 5382.5	1155.6, 1154.0	$\Delta U_{2S}\text{H}_{NS,6S}\text{I}_{2S}\text{H}_{NS,6S}$ (Fig. 2A and C)	1154.9
5301.7	1073.9	$\Delta U_{2S}\text{H}_{NS,6S}\text{I}/\text{GH}_{NS,6S}$ (Fig. 2B)	1074.9
5284.5	1057.9	$\Delta U_{2S}\text{H}_{NS,6S}\text{I}_{2S}\text{Man}_{6S}$ (Fig. 6)	1057.8
5241.5	1013.8	$\text{I}\text{H}_{NAC,6S}\text{GMan}_{3S,6S}$ (mass tagged, Fig. 4)	1013.9
5186.5	958.8	$\text{I}\text{H}_{NAC,6S}\text{GMan}_{3S,6S}$ (Fig. 4)	957.8
5007.8	780.8	$\text{H}_{NAC,6S}\text{GMan}_{3S,6S}$ (Fig. 4)	780.7
4805.2, 4805.3, 4805.2	577.7, 577.5, 576.7	$\Delta U_{2S}\text{H}_{NS,6S}$ (Fig. 2A–C)	577.5

Shown in column 1 is the  $m/z$  value of the protonated 1:1 complex of the saccharide and the basic peptide (RG)<sub>19</sub>R. Column 2 shows the observed mass of the saccharide obtained by subtracting the value of the protonated peptide observed in the spectrum from the protonated 1:1 complex. The deduced chemical structures of the saccharides for the corresponding peaks in the mass spectra are shown in column 3. Shown in column 4 are the theoretical masses calculated for the deduced structures. Note that the observed mass (column 2) is always within  $\pm 1$  Da of the calculated mass (column 4). d represents the semicarbazide mass tag ( $\Delta = 56.1$  Da).

must be a decasaccharide made of the building blocks  $\Delta U_{2S}-H_{NS,6S}$  ( $\pm D$ ),  $\Delta U-H_{NAc,6S}$  ( $\pm 4$ ), and  $\Delta U-H_{NS,3S,6S}$  ( $\pm 7$ ) in a ratio of 3:1:1. Together, the CE and MALDI-MS data were used to construct a master list of possible sequences for AT-10. We found that 320 sequences can account for both the CE and MS data (Fig. 3). These 320 sequences constitute the master list from which sequences were eliminated based on experimental constraints until convergence at a single solution (see below).

In addition, the compositional analysis confirmed that there is a contaminant present that is structurally similar to AT-10, except for the presence of  $\Delta U-H_{NS,6S}$  ( $\pm 5$ ). The composition of the contaminant can be determined to be  $\Delta U_{2S}-H_{NS,6S}$  ( $\pm D$ ),  $\Delta U-H_{NAc,6S}$  ( $\pm 4$ ),  $\Delta U-H_{NS,3S,6S}$  ( $\pm 7$ ), and  $\Delta U-H_{NS,6S}$  ( $\pm 5$ ) in the relative ratio of 2:1:1:1 from successive subtraction, based on the CE data.

Having constructed the master list of sequence possibilities, a combination of PEN-MALDI, IGS (12), and NMR analysis was used to sequence AT-10 and then PEN-MALDI was used to analyze the sequence of the contaminant.

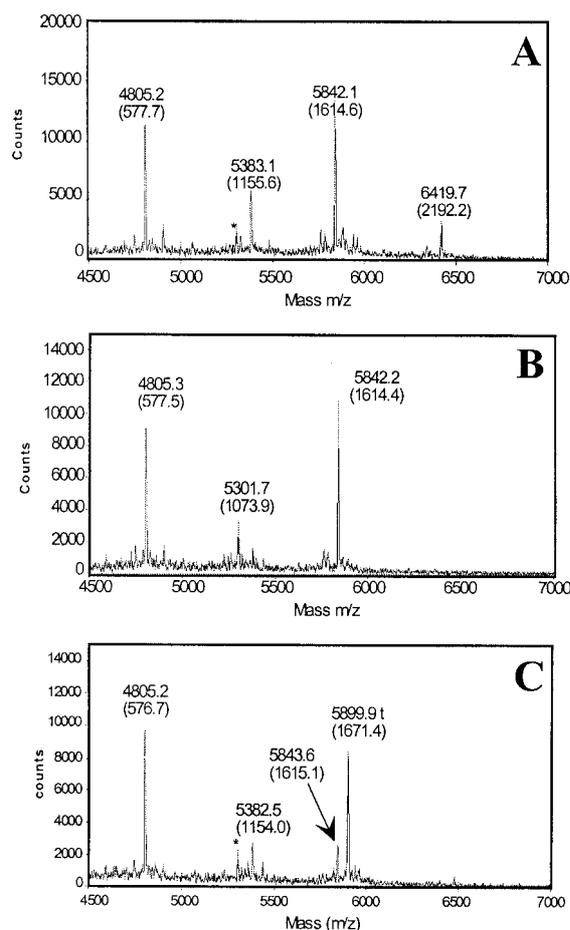
**MALDI-MS Sequencing of AT-10.** From the list of 320 possible sequences generated from the composition data, we used a series of experimental constraints, including the use of heparinase I and nitrous acid, respectively, to assign the sequence of AT-10.

Short (incomplete) digestion of AT-10 with heparinase I resulted in five fragments of molecular mass 577.7, 1073.9, 1155.6, 1614.6, and 2192.2 (Fig. 2A). The fragment with mass 577.7 corresponds to  $\pm D$ . The 1155.6 fragment corresponds to a hexasulfated tetrasaccharide, which has to have one of the following structures:  $\pm DD$ ,  $\pm D-D$ ,  $\pm D7$ ,  $\pm D-7$ ,  $\pm 7D$ , or  $\pm 7-D$ . The fragment with 1614.6 corresponds to a heptasulfated monoacetylated hexasaccharide and the fragment with a mass of 2192.2 corresponds to a decasulfated monoacetylated octasaccharide. The last peak at 1073.9 could be assigned unambiguously to the contaminant (see below for analysis). When the list of 320 sequences was searched for these fragments formed by simulated heparinase I digestion, the list was reduced to 52 sequences (Fig. 3).

The rate of substrate cleavage by heparinase I is size-dependent (19). To identify all 2-O sulfated iduronate-containing linkages in AT-10, it was treated with heparinase I under conditions that resulted in complete cleavage of all susceptible linkages. Under these conditions, the hexasaccharide and tetrasaccharide, from the contaminant, remained intact (Fig. 2B). However, the hexasulfated tetrasaccharide (mass of 1155.6 from Fig. 2A) was cleaved. Thus, this saccharide has the sequence  $\pm DD$ . Of the 52 possible sequence assignments for AT-10, only 28 can satisfy the heparinase I exhaustive digest data.

Next, AT-10 was treated with semicarbazide to yield a semicarbazone at the anomeric position. In this fashion a mass tag ( $\Delta = 56.1$ ) was introduced to differentiate fragments derived from the reducing end as opposed to the nonreducing end. Treatment of tagged AT-10 with heparinase I yielded five fragments (Fig. 2C). From a comparison of heparinase I-treated underivatized AT-10 (Fig. 2A), the heptasulfated monoacetylated hexasaccharide appears tagged and thus must be derived from the reducing end of AT-10. Application of this constraint to the 28 remaining sequences eliminated all but 12 of them (Fig. 3).

Inspection of the 12 remaining sequences in Fig. 3 indicates that the primary difference is in the identity of the reducing-end hexasaccharide. Therefore, through nitrous acid degradation of AT-10 and judicious use of exoenzymes (see *Materials and Methods*), the sequence of the reducing end tetrasaccharide was determined. First, tagged AT-10 was exhaustively treated with nitrous acid and two species were readily detectable (Fig. 4). The first, with a mass of 1013.8, corresponds to a tagged anhydromannose tetrasaccharide with four sulfates and one acetate. The other with a mass of 958.8 corresponds to the same tetrasaccharide that is untagged. Both could be assigned to one of the following sequences:  $\pm 47$ ,  $\pm 4-7$ ,  $\pm 4-D$ . Thus, based on this information,

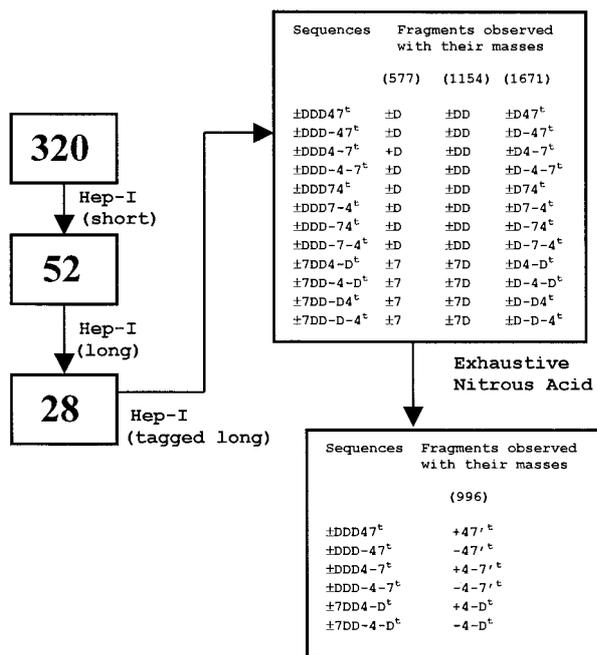


**Fig. 2.** Heparinase treatment of AT-10. Those species assignable to the contaminant are marked (\*). (A) Incomplete heparinase I treatment of AT-10. Under the conditions used in this study, heparinase I cleaves a glycosidic linkage containing an  $I_{2S}$ . (B) MALDI mass spectrum of AT-10 fragments from exhaustive digestion with heparinase I. (C) MALDI mass spectrum of tagged AT-10 treated with heparinase I shows five fragments: one with molecular mass of 576.7 Da (assignable to  $\pm D$ ), two tetrasaccharides with molecular mass of 1073.9 (\*) and 1154.0 Da, and a mass tagged hexasaccharide with a molecular mass of 1671.4 (mass of 1615.3 plus the mass tag of 56.1). Because the coupling efficiency was  $\approx 90\%$ , also seen is unlabeled hexasaccharide (mass of 1615.1).

half of the possible sequences could be eliminated, leaving only six possible sequence solutions for AT-10 (Fig. 3).

To assign uniquely the isomeric state of the two disaccharide units at the reducing end of AT-10, the following experimental constraints were used: the exhaustive nitrous acid digest was incubated with the exolytic enzyme  $\alpha$ -iduronidase that specifically clips the iduronic acid at the nonreducing end. A shift in the spectrum by 178.0 confirmed the uronic acid as 4 and its isomeric state as +, i.e.,  $IH_{NAc,6S}$  or + 4. Only three sequences could give the observed fragments, namely,  $\pm 7DD4-D$ ,  $\pm DDD4-7$ ,  $\pm DDD47$ .

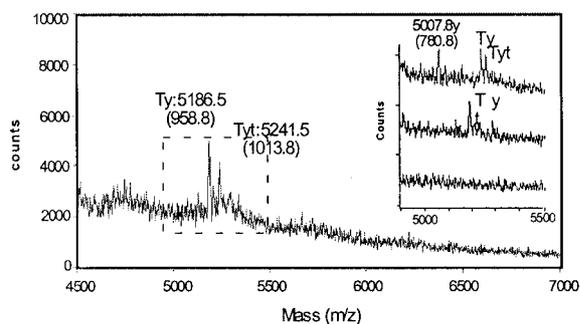
To distinguish among these last three alternatives and to identify the reducing end disaccharide, the iduronidase-treated product first was treated with 6-O sulfatase and N-deacetylase to remove the hexosamine, leaving only the reducing end disaccharide. Treatment of this sample with  $\beta$ -glucuronidase resulted in degradation to monosaccharides. This identified the reducing end disaccharide as -7 ( $G-H_{NS,3S,6S}$ ). Thus the deduced sequence of the AT-III fractionated decasaccharide is  $\pm DDD4-7$  ( $\Delta U_{2S}H_{NS,6S}I_{2S}H_{NS,6S}I_{2S}H_{NS,6S}IH_{NAc,6S}GH_{NS,3S,6S}$ ). Of note is the fact that this sequence does not agree with the sequence



**Fig. 3.** Convergence of the AT-10 sequence using a stepwise strategy. Based on the compositional analysis data that indicated the presence of three  $\pm$ Ds, one  $\pm$ 4 and one  $\pm$ 7 disaccharide units for the deca-saccharide, 320 possible sequences can be constructed, considering all possible permutations. The number of possible sequences equals  ${}^5C_3$  (arrange 3 Ds in five positions) multiplied by  ${}^2C_1$  (arrange the four in the remaining two positions)  $\times 2^4$  (to account for the epimeric state of the uronic acid at each of the four internal positions; the nonreducing end uronic acid contains a  $\Delta^{4,5}$  bond from heparinase degradation) =  $10 \times 2 \times 16 = 320$ . Application of experimental constraints to eliminate prospective solutions from the master list of 320 possible sequences results in convergence to the final sequence. Shown in the boxes at the left is the number of sequences that satisfy the experimental constraints. The boxes on the right show the sequences that satisfy the experimental constraints along with the possible fragments formed for the masses shown in parentheses at the top.

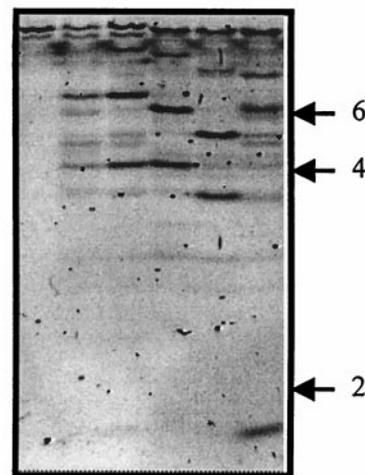
assignment for a deca-saccharide produced in an identical manner (20). Therefore, we sought to confirm our sequencing assignment by using other analytical methodologies.

**IGS Sequencing of AT-10.** AT-10 also was sequenced by using the recently established technique IGS, which uses an electrophoretic separation of saccharides tagged at the reducing end



**Fig. 4.** Exhaustive degradation of AT-10 with nitrous acid. Nitrous acid cleaves at H<sub>NS</sub> residues, leaving behind an anhydromannose ( $\Delta = 97.1$  Da). (Inset) The mass spectrum of the degradation profile when the sample was treated with iduronidase (Top), glucosamine 6-O sulfatase (Middle), and glucuronidase (Bottom) in that order. Ty = tetrasaccharide with an anhydromannose at the reducing end; t = semicarbazide tagged.

	1	2	3	4	5	6
pHNO <sub>2</sub>	-	+	+	+	+	+
Gase	-	-	+	-	-	+
I <sub>2</sub> Sase	-	-	-	+	+	+
Idase	-	-	-	-	+	+
G <sub>6</sub> Sase	-	-	-	-	-	+



**Fig. 5.** IGS of AT-10 using electrophoretic separation. AT-10 was tagged with the fluorophore 2-aminobenzoic acid, subjected to partial nitrous acid cleavage, and digested with combinations of exoenzymes as described in *Materials and Methods*. The resulting products were separated on a 16-cm 33% Tris-acetate gel and imaged on a UV transilluminator. The positions of standard di-, tetra-, and hexasaccharides were as shown.

with a fluorophore. Partial nitrous acid cleavage and exoenzyme digestion of the saccharide produces a ladder from which the sequence can be determined. In accord with the PEN-MALDI data, electrophoretic analysis of the fluorophore-tagged sample produced a single major deca-saccharide, but in addition, the smaller contaminant was also evident (Fig. 5). The products of partial nitrous acid cleavage were deca-, octa-, hexa-, and tetrasaccharides, with no disaccharide products observed (Fig. 5). This result defines the positions of all of the NS and NAC moieties, with just one N-acetylated disaccharide in the position proximate to the reducing end. Gel shifts caused by treatment of these products with different combinations of exoenzymes demonstrated iduronate residues in three positions, two of which were 2-O-sulfated, and the presence in three positions of 6-O-sulfated glucosamine residues. The nonreducing end was clearly 2-O-sulfated, but confirmation of the presence a 6-O-sulfate on the nonreducing end glucosamine residue, and details of the sulfation pattern on the reducing end monosaccharide were not obtained in this analysis. These data define the structure of the AT-10 as  $\Delta U_{2S}H_{NS,\pm 6S}I_{2S}H_{NS,6S}I_{2S}H_{NS,6S}IH_{NAC,6S}GH_{NAC/NS,\pm 3S,\pm 6S}$ . These data, derived from an independent sequencing approach, are entirely consistent with the PEN-MALDI analysis.

**Interpretation of NMR Spectrum of AT-10.** The NMR spectrum of AT-10 is shown in Fig. 6. Consistent with our analysis, the small signals of  $\alpha$ H-1,  $\alpha$ H-2, and  $\alpha$ H-3 of H<sub>NS,3S,6S</sub> [at 5.45, 3.45 and 4.50 parts per million (ppm), respectively] allow us to assign the reducing end monosaccharide unit as H<sub>NS,3S,6S</sub>. In this case, the anomeric proton of the reducing end H<sub>NS3S6S</sub> residue must be split into  $\alpha$  and  $\beta$  configurations. The  $\alpha$  configuration of the anomeric proton of the H<sub>NS</sub> residue is dominant ( $\approx 95\%$ ) in protic solvents, such as deuterium oxide, based on the anomeric effect. Furthermore, the presence of two I<sub>2S</sub> moieties could be detected. Interestingly, the anomeric signals of the two I<sub>2S</sub>, which usually resonate around 5.20

