Very-high-field n.m.r. studies of bovine lung heparan sulphate tetrasaccharides produced by nitrous acid deaminative cleavage

Determination of saccharide sequence, uronate composition and degrees of sulphation

Paul N. SANDERSON,* Thomas N. HUCKERBY† and Ian A. NIEDUSZYNSKI*‡

*Department of Biological Sciences, University of Lancaster, Bailrigg, Lancaster LA1 4YQ, and †Department of Chemistry, University of Lancaster, Bailrigg, Lancaster LA1 4YA, U.K.

(Received 16 May 1984/Accepted 18 June 1984)

Tetrasaccharides with the general structure UA–GlcNAc–GlcUA–aManOH (where UA represents uronate, GlcNAc *N*-acetylglucosamine, GlcUA glucuronate and aManOH anhydromannitol) were prepared from low-sulphated heparan sulphates of bovine lung origin by complete nitrous acid deaminative cleavage followed by reduction and fractionated by gel filtration. Ion-exchange chromatography of the tetrasaccharides yielded three major fractions in approximate yields of 37%, 45% and 14%. These were shown to be non-, mono- and di-sulphated respectively. Complete structural characterization of the tetrasaccharide fractions by quantitative high-field n.m.r. spectroscopy showed that each fraction contained only two discrete species and led to the following observations. (1) All of the uronate residues in the tetrasaccharides (and in larger oligosaccharides) are unsulphated, and hence sulphated iduronate [IdUA(2SO₃)] must occur exclusively within –GlcNSO₃–IdUA(2SO₃)–GlcNSO₃– sequences (where GlcNSO₃ represents *N*-sulphoglucosamine) in the parent polymers. (2) The GlcNAc residues in the tetrasaccharides are more highly C-6-*O*-sulphated than are the aManOH residues, and furthermore sulphation on the aManOH appears to occur only where the GlcNAc is also sulphated. (3) Where the GlcNAc is C-6-*O*-sulphated, iduronate is the major non-reducing terminal residue, whereas glucuronate predominates in this position if the GlcNAc is unsulphated. The quantitative data obtained are used to determine the degree of C-6-*O*-sulphation of glucosamine residues in specific sequences within the parent heparan sulphates.

Heparan sulphates are glycosaminoglycans that are found as proteoglycan components on the surfaces of possibly all mammalian cells (Kraemer, 1971). Structurally, heparan sulphates are the most heterogeneous of the glycosaminoglycans, containing both 'heparin-like' [–IdUA(2SO₃)–GlcNSO₃(6SO₃)–]_n and 'heparan' [–GlcUA–GlcNAc]_n block structures, together with regions of intermediate substitution (Linker & Hovingh, 1975; Cifonelli & King, 1977).

Abbreviations used: UA, uronic acid; GlcUA, β-D-glucuronic acid; IdUA, α-L-iduronic acid; GlcNAc, 2-acetamido-2-deoxy-α-D-glucose (*N*-acetylglucosamine); GlcNSO₃, 2-sulphamido-2-deoxy-α-D-glucose (*N*-sulphoglucosamine); aManOH, 2,5-anhydro-D-mannitol; SO₃, sulphate ester.

‡ To whom requests for reprints should be addressed.

Heparan sulphate chains have been shown to self-associate (Fransson *et al.*, 1980a). The interaction may be specific and is strongest between chains of similar composition (Fransson & Havsmark, 1982). Segments of the heparan sulphate chains containing both iduronate and glucuronate are believed to comprise the contact zones, and, in particular, sequences of the form –GlcNSO₃–IdUA–GlcNAc–GlcUA–GlcNSO₃– have been shown to be relatively common in self-associating species (Fransson *et al.*, 1980b; Fransson, 1981). The contact zones may be located in their correct interacting orientations by intervening [–GlcUA–GlcNAc–]_n sequences (Fransson, 1981). Heparan sulphates from virus-transformed cells have a smaller degree of sulphation (Winterbourne & Mora, 1981) and have been shown to have a

decreased affinity for specific cell-surface receptors (Winterbourne, 1982) compared with heparan sulphates from normal cells.

Treatment of heparan sulphates with nitrous acid at low pH selectivity cleaves the glycosidic linkages of *N*-sulphoglucosamine residues within the polymers (Shively & Conrad, 1976), resulting in a mixture of oligosaccharide fragments. These oligosaccharides can be more readily characterized than can the parent polymers. Disaccharides obtained in this manner from heparan sulphates have been characterized by chromatographic techniques (Jacobsson *et al.*, 1979; Delaney & Conrad, 1983). Heparan sulphate oligosaccharides have also been characterized by the sequential use of exo-enzymes (Linker, 1979).

A ^{13}C -n.m.r. study of the methylene resonances from a graded series of heparan sulphate oligosaccharides showing the extent and residue positions of C-6-*O*-sulphation has previously been reported (Sanderson *et al.*, 1983). In the present study the heparan sulphate tetrasaccharides were further fractionated by ion-exchange chromatography into non-, mono- and di-sulphated fractions, from which quantitative high-field ^{13}C - and ^1H -n.m.r. spectra were obtained. Analysis of the spectra afforded a complete sequence analysis of these heparan sulphate tetrasaccharides.

Experimental

Materials

Sephadex G-15 was purchased from Sigma Chemical Co. (Poole, Dorset, U.K.). DE-52 DEAE-cellulose anion-exchange resin was from Whatman (Maidstone, Kent, U.K.). Other materials used were obtained as before (Sanderson *et al.*, 1983).

General methods

Preparation of the heparan sulphate parent material, subsequent complete deaminative cleavage, and gel filtration of the oligosaccharides produced, have been described previously (Sanderson *et al.*, 1983).

Fractionation of the heparan sulphate tetrasaccharides

The heparan sulphate tetrasaccharides were subjected to analytical high-voltage paper electrophoresis on Whatman 3MM paper at a constant current of 100mA for 1 h in 1.6M-formic acid, pH 1.7. Strips 0.5cm wide were assayed for ^3H radioactivity by scintillation counting.

Analytical-scale fractionation of the tetrasaccharides by ion-exchange chromatography was achieved on a 1 ml column of DEAE-cellulose (DE-52) eluted with a 20ml linear gradient of 0–0.3M-

NaCl at 4ml/h. This procedure was scaled up for preparative-scale fractionation to an 8cm $\times$ 2.6cm DEAE-cellulose column eluted with a 1-litre linear gradient of 0–0.2M-NaCl at 48ml/h.

The pooled fractions were concentrated and de-salted on a Sephadex G-15 column (size, 87cm $\times$ 1.4cm; eluent, distilled water; flow rate, 10ml/h).

N.m.r. spectroscopy

The de-salted samples were buffered to pH 7 with phosphate. 3-Trimethylsilyl[$^2\text{H}_4$]propionic acid sodium salt was added as internal reference (Huckerby, 1983). The samples (17–60mg) were exchanged several times with $^2\text{H}_2\text{O}$ and finally dissolved in 0.5ml of $^2\text{H}_2\text{O}$.

Spectra were determined on a Bruker WH400 instrument with a 5mm variable-temperature probe at 60°C, operating at 400.14 MHz for ^1H and 100.61 MHz for ^{13}C .

Values for ^{13}C spin-lattice relaxation times (T_1) were determined for the unfractionated tetrasaccharide mixture by using the standard inversion-recovery pulse sequence. The longest T_1 value was found to be 1.1s. Fully relaxed spectra were then obtained by the use of a 5s delay after each acquisition. By using these conditions spectra were obtained in which the area of a peak represents the number of carbon atoms in a distinct molecular environment. Quantitative data were therefore obtained by comparison with the peak areas of 'unit-carbon' atoms present in each sample (e.g. the unit-carbon resonances at 82–89p.p.m. from aManOH ring carbon atoms).

Results

Determination of degrees of sulphation in the parent polymers by ^{13}C n.m.r.

The overall degrees and position of sulphation in the parent heparan sulphates were determined from the 100MHz ^{13}C -n.m.r. spectrum (Fig. 1). Comparison of the total peak areas at $\delta \sim 60$ p.p.m. (from C-2 of GlcNSO₃) to those at $\delta \sim 56$ p.p.m. (from C-2 of GlcNAc) showed that approx. 46% of the glucosamine residues were *N*-sulphated. There was no evidence of any *N*-unsubstituted glucosamine residues (i.e. no significant signals at $\delta \sim 57.5$ p.p.m.). The ratio of the area of the resonance at $\delta \sim 63$ p.p.m. to that at $\delta \sim 69$ p.p.m. (from unsulphated and sulphated methylene carbon atoms respectively) indicates that approx. 30% of the total glucosamine residues are C-6-*O*-sulphated. Sulphated iduronate accounts for approx. 19% of the total uronate residues, as determined from the relative areas of the signals at $\delta \sim 102$ p.p.m. to those at $\delta 104$ –106p.p.m.

Fig. 1. 100 MHz ^{13}C -n.m.r. spectrum of the parent heparan sulphate preparation

The unsulphated uronate anomeric resonances (at δ 104–106 p.p.m.) appear downfield of the anomeric resonances from IdUA(2SO_3) (δ ~102 p.p.m.). Methylene resonances appear at δ 62–71 p.p.m. Signals from C-2 of glucosamine residues are at δ ~60 p.p.m. (GlcNSO₃) and δ ~56 p.p.m. (GlcNAc). Carbonyl resonances are at δ 177–178 p.p.m., and methyl resonances are at δ ~25 p.p.m. The broad hump seen at δ ~112 p.p.m. is an artifact produced by the probe.

Characterization of the heparan sulphate tetrasaccharides

The heparan sulphate tetrasaccharide mixture yielded three major fractions, as determined both by analytical ion-exchange chromatography (Fig. 2) and by analytical high-voltage electrophoresis. These fractions, designated I, II and III (see Fig. 2), accounted for 37%, 45% and 14% respectively of the total ^3H radiolabel. Both fractionation procedures gave rise to a minor component that contained 4% of the total ^3H radioactivity and was eluted from the DEAE-cellulose column in 0.5M-NaCl, but migrated more slowly than the other components on electrophoresis.

For ease of reference the residues within the tetrasaccharide sequence are numbered as follows: residue 1 is the non-reducing terminal uronate, residue 2 is the GlcNAc, residue 3 is the internal GlcUA and residue 4 is the terminal aManOH.

The ^{13}C anomeric (i.e. C-1) resonance from an IdUA(2SO_3) residue appears at $\delta \sim 102$ p.p.m., approx. 2 p.p.m. upfield of the anomeric resonances from unsulphated uronate residues in oligosaccharides of this type (Kosakai & Yosizawa, 1981). The absence of any significant resonances at $\delta \sim 102$ p.p.m. in the spectra from the tetrasaccharide fractions indicates that residue 1 in these tetrasaccharides is an unsulphated uronate. Consequently, sulphation within these tetrasaccharides must be confined to the GlcNAc and aManOH residues.

The methylene region of the ^{13}C -n.m.r. spectra (δ 62–71 p.p.m.) clearly indicates the degree and location of C-6-*O*-sulphation within these tetrasaccharides. This region is shown in Fig. 3 and the relevant chemical shifts are listed in Table 1. The absence of any signals in the region δ 68–71 p.p.m. in the spectrum from fraction I indicates that this

fraction is completely unsulphated. The two resonances (whose combined area is equivalent to a unit-carbon) at δ 62–63 p.p.m. in the fraction I spectrum arise from C-6 of the unsulphated GlcNAc (Sanderson *et al.*, 1983) and are absent from the spectra of both fractions II and III. However, these spectra contain additional resonances at $\delta \sim 69$ p.p.m., which are assigned to C-6 of a GlcNAc(6SO_3) residue. The two unit-carbon resonances at $\delta \sim 64$ p.p.m. in the spectra of fractions I and II are from the methylene carbon atoms 1 and 6 of an unsulphated aManOH residue. In the spectrum from fraction III a unit-carbon signal is present at δ 70.47 p.p.m., which is assigned to C-6 of an aManOH(6SO_3) residue. The single peak at δ 63.75 p.p.m. is consequently assigned to C-1 of an aManOH(6SO_3) residue.

From these data it can be seen that fraction I is completely unsulphated, fraction II is monosulphated (exclusively on the GlcNAc) and fraction III is disulphated (on both the GlcNAc and aManOH residues).

The resonances originating from GlcNAc C-1, C-2 and C-6 in each of the tetrasaccharide spectra all have a bimodal character. The relative proportions of the two components of each bimodal resonance are listed in Table 1, and are shown to be approximately constant within each spectrum. The only variation in the molecular environment of the GlcNAc in a given fraction arises from the possibility of residue 1 being GlcUA or IdUA. Consequently, the bimodal nature of the GlcNAc resonances must reflect the uronate distribution at residue 1. This GlcUA/IdUA ratio, which was not directly deducible from the ^{13}C -n.m.r. spectra, was determined by reference to the ^1H -n.m.r. spectra.

In the 400 MHz-n.m.r. spectra of these fractions (not shown), signals from the anomeric protons of glucosamine and iduronate residues appear downfield from the main block of signals (Kosakai & Yosizawa, 1981). On taking the area of the GlcNAc [or GlcNAc(6SO_3)] anomeric signal as unity, the relative proportion of IdUA in each fraction can be determined. This information is shown in Table 2 and indicates that the IdUA content is much greater in fractions II and III (i.e. where the GlcNAc is C-6-*O*-sulphated) than it is in fraction I.

By using the ^1H data from Table 2 the two components of the ^{13}C GlcNAc resonances can be assigned. Since the downfield (i.e. left-hand) component is the smaller of the bimodal GlcNAc resonances in the spectrum from fraction I, and is the major component for both fractions II and III, this peak must arise from the appropriate carbon atom in a GlcNAc [or GlcNAc(6SO_3)] residue that is adjacent to an IdUA at residue 1. Consequently, the upfield peak of the bimodal resonances must

Fig. 2. Analytical ion-exchange fractionation of heparan sulphate tetrasaccharides

For experimental conditions refer to the text. Fractions I, II and III were subsequently shown to be non-, mono- and di-sulphated respectively. ●, ^3H radioactivity; -----, NaCl gradient.

Fig. 3. 100 MHz ^{13}C -n.m.r. spectra of heparan sulphate tetrasaccharide fractions: upfield main-ring resonances (δ 55–71 p.p.m.)

GlcNAc C-2 (δ ~56 p.p.m.) and C-6 (CH_2OH δ 62–63 p.p.m. and CH_2OSO_3 δ ~69 p.p.m.) bimodal resonances and aManOH C-1 and C-6 methylene resonances are shown for the heparan sulphate tetrasaccharide fractions: (a) fraction I; (b) fraction II; (c) fraction III. Refer to Table 1 for chemical-shift listings.

represent the respective GlcNAc carbon atom 'sensing' and adjacent terminal GlcUA residue.

By averaging the quantitative ^{13}C - and ^1H -n.m.r. results (from Tables 1 and 2) an approximate GlcUA/IdUA ratio for residue 1 can be obtained. This ratio is 60:40, 36:64 and 33:67 for fractions I, II and III respectively. Within these tetrasaccharide sequences, therefore, GlcUA is the

major component at residue 1 where the GlcNAc is unsulphated. However, where the GlcNAc is C-6-*O*-sulphated, IdUA predominates at residue 1.

The relative proportions of the various tetrasaccharide sequences, in which a GlcNAc occurs between two GlcNSO₃ glucosamine residues in the parent heparan sulphate polymers, can be determined from the above n.m.r. analysis of the tetra-

Table 1. 100 MHz ^{13}C -n.m.r. chemical-shift assignments for methylene and identified GlcNAc resonances from the heparan sulphate tetrasaccharide fractions

Percentage of signal refers to the relative areas of the two components of the bimodal resonances from GlcNAc C-1, C-2 and C-6 in each spectrum.

Carbon assignment		Heparan sulphate tetrasaccharide fraction					
		I		II		III	
		δ (p.p.m.)	% of signal	δ (p.p.m.)	% of signal	δ (p.p.m.)	% of signal
Methylene carbon atoms							
GlcNAc	C-6	62.66	39	—	—	—	—
GlcNAc	C-6	62.50	61	—	—	—	—
GlcNAc(6SO ₃)	C-6	—	—	68.96	61	68.96	65
GlcNAc(6SO ₃)	C-6	—	—	68.71	39	68.73	35
aManOH	C-1	64.14	63.90	64.14	—	—	—
aManOH	C-6			63.91	—	—	—
aManOH(6SO ₃)	C-1	—	—	—	—	63.75	—
aManOH(6SO ₃)	C-6	—	—	—	—	70.47	—
Other GlcNAc carbon atoms							
GlcNAc	C-1	99.59	43	99.63	63	99.51	71
GlcNAc	C-1	99.47	57	99.46	37	99.35	29
GlcNAc	C-2	56.46	39	56.34	67	56.33	62
GlcNAc	C-2	56.15	61	56.00	33	55.99	38

Table 2. 400 MHz ^1H -n.m.r. chemical-shift assignments and relative peak areas for glucosamine and iduronate anomeric protons, showing the iduronate content of the heparan sulphate tetrasaccharide fractions

For the values given as unit area, the area of the peaks from the glucosamine anomeric proton is taken as unity and the relative area of the iduronate anomeric signals to this value is calculated to determine the proportion of iduronate at the non-reducing terminal of each heparan sulphate tetrasaccharide fraction.

Proton assignment		Heparan sulphate tetrasaccharide fraction					
		I		II		III	
		δ (p.p.m.)	Unit area	δ (p.p.m.)	Unit area	δ (p.p.m.)	Unit area
GlcNAc	H-1	5.382	1.00	—	—	—	—
GlcNAc	H-1	5.374		—	—	—	—
GlcNAc(6SO ₃)	H-1	—	—	5.416	1.00	5.430	1.00
GlcNAc(6SO ₃)	H-1	—	—	5.399		5.415	
IdUA	H-1	4.844	0.37	4.941	0.66	4.935	0.70

saccharide fractions. This information is summarized in Table 3.

Assignment of the anomeric resonances

The GlcUA/IdUA ratio at residue 1 in these tetrasaccharides was derived from the relative areas of the anomeric signals in the ^1H -n.m.r. spectra. Equivalent information could be derived from the anomeric region of the ^{13}C -n.m.r. spectra if complete assignment of this region can be made. By examination of the anomeric region of the ^{13}C -n.m.r. spectra of the tetrasaccharide fractions (Fig. 4), and by using the previously determined

structures for each of the fractions, such assignment is now possible. In the ^{13}C -n.m.r. spectrum of the unfractionated tetrasaccharide mixture, unsulphated uronate anomeric carbon atoms give rise to five major resonances (at δ 103–106 p.p.m., shown in Fig. 5a), in contrast with the bimodal character of the glucosamine anomeric carbon atoms (at δ ~99.5 p.p.m.) described previously. These unsulphated uronate anomeric carbon resonances are now assigned below.

In the ^{13}C -n.m.r. spectrum from fraction I, the three major uronate anomeric signals observed (see Fig. 4a) represent the three different molecular

Table 3. *Distribution of uronate residues and sulphation in the vicinity of 'isolated' GlcNAc residues in the parent heparan sulphate preparation*

The data were calculated from analysis of the quantitative n.m.r. spectra of the heparan sulphate tetrasaccharide fractions.

Residue no. ...	Heparan sulphate tetrasaccharide sequence*					Frequency of occurrence (% of total tetrasaccharide pool)
	(0)	1	2	3	4	
	GlcUA	–GlcNAc	–GlcUA	–aManOH		22
	IdUA	–GlcNAc	–GlcUA	–aManOH		15
	GlcUA	–GlcNAc(6SO ₃)	–GlcUA	–aManOH		16
	IdUA	–GlcNAc(6SO ₃)	–GlcUA	–aManOH		29
	GlcUA	–GlcNAc(6SO ₃)	–GlcUA	–aManOH(6SO ₃)		5
	IdUA	–GlcNAc(6SO ₃)	–GlcUA	–aManOH(6SO ₃)		9

* In the parent heparan sulphate polymers residue 0 and residue 4 were both GlcNSO₃ residues. The distribution of 6-*O*-sulphate at residue 0 cannot be assessed from this investigation.

Fig. 4. 100 MHz ¹³C-n.m.r. spectra of heparan sulphate tetrasaccharide fractions: anomeric resonances (δ 99–106 p.p.m.) Glucosamine (δ 99–100 p.p.m.) and unsulphated uronate (δ 104–106 p.p.m.) anomeric resonances are shown for the heparan sulphate tetrasaccharide fractions: (a) fraction I; (b) fraction II; (c) fraction III.

environments of the uronate anomeric carbon atoms in this fraction. The largest signal (at δ 104.75 p.p.m.) is assigned to the internal GlcUA (at residue 3), which is linked across its glycosidic

linkage to an unsulphated aManOH. The two remaining signals arise from residue 1. The larger of these (at δ 105.12 p.p.m.) is assigned to GlcUA C-1, since GlcUA predominates at residue 1 in

Fig. 5. 100 MHz ^{13}C -n.m.r. spectra of heparan sulphate oligosaccharides: anomeric resonances (δ 98–106 p.p.m.) Glucosamine (δ 99–100 p.p.m.) and unsulphated uronate (δ 103–106 p.p.m.) anomeric resonances are shown for the following oligosaccharide fractions: (a) tetrasaccharide; (b) hexasaccharide; (c) octasaccharide; (d) deca-saccharide. Note the absence of signals at δ ~102 p.p.m. from the anomeric carbon atom of IdUA(2SO_3) residues.

fraction I. The signal at this chemical shift position becomes the major signal in spectra from larger oligosaccharides in this series (see Fig. 5), and therefore, in general terms, represents C-1 of a GlcUA residue linked through its glycosidic linkage to an unsulphated GlcNAc. The remaining uronate anomeric signal (at δ 104.09 p.p.m.) in the fraction I spectrum (Fig. 4a) is therefore assigned to C-1 of IdUA at residue 1, adjacent to an unsulphated GlcNAc.

Only two major components (representing approx. 1.3 and 0.7 unit-carbons respectively) are observed in the uronate anomeric region of the spectrum from fraction II (Fig. 4b). The larger of the two (at δ 104.72 p.p.m.) is assigned to C-1 of GlcUA at residue 3 (as for fraction I). This resonance coincides with the resonance representing C-1 of the terminal GlcUA (residue 1) linked to GlcNAc(6SO₃). The remaining peak (at δ 104.44 p.p.m.) is assigned to C-1 of IdUA at residue 1 [i.e. linked to a GlcNAc(6SO₃)]. The signal at this position (δ 104.44 p.p.m.) in the fraction III spectrum (Fig. 4c) is similarly assigned to the terminal IdUA residue. The smallest of the three major uronate anomeric signals (at δ 104.77 p.p.m.) in the fraction III spectrum (Fig. 4c) is assigned to C-1 of GlcUA at residue 1 (as for fraction II). The remaining, and largest, of the fraction III uronate signals (at δ 104.94 p.p.m.) is consequently assigned to C-1 of the GlcUA at residue 3 [i.e. linked via its glycosidic linkage to an aManOH(6SO₃) residue].

The resonances representing the anomeric carbon atom of unsulphated iduronate residues therefore appear upfield of the GlcUA C-1 resonances, as shown in Fig. 4. On the basis of the assignments made for the tetrasaccharide fractions, the anomeric region of the ¹³C-n.m.r. spectra of the larger oligosaccharides (shown in Fig. 5) may also be assigned.

Overall, the ¹³C-n.m.r. spectra presented here provide a quantitative representation of the components in a given fraction. It should be noted however, that the uronic acid anomeric signals generally have a larger unit area than expected by comparison with the GlcNAc anomeric signals. This characteristic has been consistently observed in many spectra of oligosaccharides and polymers of this type. The reasons for this phenomenon are not yet understood; they may be related to the striking differences observed (Huckerby & Nieduszynski, 1982) for spin-lattice relaxation times at iduronate and hexosamine carbon atoms in a wide range of heparinoid polymers. These differences are also found for the heparan sulphate oligosaccharides described here (T. N. Huckerby, I. A. Nieduszynski & P. N. Sanderson, unpublished work).

Discussion

Complete nitrous acid digestion of the heparan sulphate preparation yielded oligosaccharides which were derived from sequences of the form -GlcNSO₃-UA-[GlcNAc-GlcUA]_n-GlcNSO₃-, where $n \geq 0$ in the parent polymers. The distribution in the preparation of the differently sized oligosaccharides, determined from the gel-filtration profile as assayed by orcinol, is as follows: $n = 1$, 24%; $n = 2$, 13%; $n = 3$, 8%; $n = 4$, 6%; $n = 5$, 5%; $n > 5$, 15%. This oligosaccharide distribution accounts for approx. 22% of the total glucosamine residues being present as GlcNSO₃ within -GlcNAc-GlcUA-GlcNSO₃- sequences. The ¹³C-n.m.r. analysis of the parent polymer preparation indicates that approx. 46% of the glucosamine residues are GlcNSO₃. Consequently, disaccharide units (where $n = 0$) represent approx. 24% of the total heparan sulphate material.

The absence of IdUA(2SO₃) anomeric resonances (at $\delta \sim 102$ p.p.m.) in the ¹³C-n.m.r. spectra of the tetrasaccharides and of larger oligosaccharides (Fig. 5) indicates that these oligosaccharides contain no IdUA(2SO₃) and, furthermore, shows that the IdUA(2SO₃) in the parent heparan sulphates (see Fig. 1) is found only within -GlcNSO₃-IdUA(2SO₃)-GlcNSO₃- sequences. Consequently, IdUA(2SO₃) is confined to the disaccharide fraction of this oligosaccharide preparation, and indeed, since IdUA(2SO₃) accounts for approx. 19% of the total uronate in the parent polymers, approx. 80% of the disaccharide uronate must be IdUA(2SO₃). These results are consistent with those obtained by Jacobsson *et al.* (1979) and of Delaney & Conrad (1983), both groups having shown that IdUA(2SO₃) is the principal uronate component of disaccharides obtained from heparan sulphates by nitrous acid digestion.

Approx. 30% of the glucosamine residues in the parent heparan sulphate preparation were C-6-O-sulphated. The distribution of these sulphate substituents can be calculated from the relative yields of the differently sized oligosaccharides, together with their C-6-O-sulphate contents and residue positions, as previously determined (Sanderson *et al.*, 1983). Such calculations indicate that approx. 43% of the total C-6-O-sulphate occurs on GlcNAc residues (uniquely within -GlcNSO₃-UA-GlcNAc- sequences) and approx. 7% is found on GlcNSO₃ residues within -GlcNAc(6X)-GlcUA-GlcNSO₃- sequences (where X is either OH or SO₃) in the parent heparan sulphates. The remaining 50% of the total C-6-O-sulphate must derive from -GlcNSO₃(6X)-UA-GlcNSO₃(6SO₃)- sequences (where X may be OH or SO₃, and cannot be determined from this investigation) within the heparan sulphate poly-

mers, and consequently must be present on the disaccharide aManOH residues. Approx. 62% of the disaccharides from the nitrous acid digestion must therefore contain an aManOH(6SO₃) residue, again consistent with the findings obtained by Jacobsson *et al.* (1979).

The biosynthesis of heparan sulphate is believed to resemble that of heparin (Riesenfeld *et al.*, 1982). In the following discussion it is assumed that all the discrete heparan sulphates that are present in the preparation studied here, have arisen from a single biosynthetic pathway. The substrate specificity of the C-6-*O*-sulphotransferases involved in the polymer-modification process of heparin biosynthesis has been investigated by Jacobsson & Lindahl (1980). A GlcNSO₃ residue present to the reducing side of another GlcNSO₃ residue within the polymer (i.e. the underlined GlcNSO₃ residue in –GlcNSO₃–UA–GlcNSO₃– sequences) was found to be a particularly efficient acceptor of *O*-sulphate, and furthermore it was suggested that the GlcNAc within –GlcNSO₃–UA–GlcNAc– sequences might also be an acceptor. It has also been reported that C-6-*O*-sulphation of GlcNAc residues in oligosaccharides derived from heparan sulphates by nitrous acid digestion appeared to be restricted to those GlcNAc residues that occurred in –GlcNSO₃–UA–GlcNAc– sequences (Sanderson *et al.*, 1983).

From the results presented in this paper, the degree of glucosamine C-6-*O*-sulphation within various sequences in the heparan sulphate polymers can be calculated. This information is presented in Table 4. The frequency of C-6-*O*-sulphation in sequences (b), (c) and (e) is derived from the analysis of the tetrasaccharides described above, whereas the degree of sulphation of sequence (f) is obtained from the calculated

disaccharide composition. Sequence (a) represents a GlcNAc in a 'heparan'-type sequence, within which GlcNAc C-6-*O*-sulphation has not been detected. The ratio of GlcUA to IdUA at the non-reducing terminal of the hexa- and larger oligosaccharides of this series is believed to be constant, as is the degree of C-6-*O*-sulphation on the adjacent GlcNAc (Sanderson *et al.*, 1983). Although no evidence was found for the presence of GlcNSO₃(6SO₃) in sequences of the form (d) in the tetrasaccharide preparation, such sequences may be present at a <10% level in larger oligosaccharides of this series (Sanderson *et al.*, 1983), and have also been shown to represent approx. 15% of the tetrasaccharides from a more highly sulphated heparan sulphate preparation (Linker, 1979). By using the degree of C-6-*O*-sulphation of specific glucosamine residues (listed in Table 4) and the frequency of occurrence of such sequences within the polymers, the gross degree of C-6-*O*-sulphation in the parent heparan sulphates can be calculated and is found to be approx. 32%. This is consistent with the n.m.r. estimate of a 30% level of C-6-*O*-sulphation in the parent material. Thus the frequency of C-6-*O*-sulphation of glucosamine residues within the sequences listed in Table 4 presents a representative description of glucosamine sulphation in such sequences within this parent heparan sulphate preparation.

The uronate composition of these heparan sulphate tetrasaccharides has also been determined in the present study. From biosynthetic considerations, the internal uronate (residue 3) can only be glucuronate, as it has a GlcNAc residue on its non-reducing side (Jacobsson *et al.*, 1984). Also, there is no evidence from the ¹³C-n.m.r. spectra of any environmental variation in the vicinity of the aManOH residues, such as that observed in the

Table 4. Frequency of C-6-*O*-sulphation of glucosamine residues occurring within various sequences in the parent heparan sulphate preparation

The target residue for C-6-*O*-sulphation is underlined in each case. These sequences represent only the immediate residues on the non-reducing side of the target glucosamine residue, as identified by analysis of the oligosaccharide products of nitrous acid digestion. In the parent heparan sulphates, the residues on the reducing side of the target glucosamine residue will presumably also be required for substrate recognition by the C-6-*O*-sulphotransferase. For GlcNSO₃(6X) residues in these sequences X may be either OH or SO₃. The nature of X cannot be determined from this investigation, since the glycosidic linkage of this residue is the site of nitrous acid cleavage.

Disaccharide sequence structure on the non-reducing side of the site of glucosamine C-6- <i>O</i> -sulphation				Degree of glucosamine C-6- <i>O</i> -sulphation in these sequences within the parent heparan sulphates (%)
(a)	– GlcNAc	– GlcUA	– <u>GlcNAc</u> –	n.d.*
(b)	– GlcNSO ₃ (6X)	– GlcUA	– <u>GlcNAc</u> –	48
(c)	– GlcNSO ₃ (6X)	– IdUA	– <u>GlcNAc</u> –	72
(d)	– GlcNAc	– GlcUA	– <u>GlcNSO₃</u> –	<10
(e)	– GlcNAc(6SO ₃)	– GlcUA	– <u>GlcNSO₃</u> –	24
(f)	– GlcNSO ₃ (6X)	– UA	– <u>GlcNSO₃</u> –	62

* C-6-*O*-sulphation of GlcNAc residues in sequence (a) structures has not been detected.

bimodal character of the GlcNAc resonances that arises from the GlcUA/IdUA distribution at residue 1. The data presented here show that the non-reducing terminal uronate is unsulphated, in agreement with the findings reported by Linker (1979). In addition, the GlcUA/IdUA ratio at residue 1 is shown to be correlated with sulphation at C-6 of the adjacent GlcNAc. If the GlcNAc at residue 2 is sulphated, the major component at residue 1 is IdUA, whereas GlcUA predominates where the GlcNAc is unsulphated. As epimerization is believed to occur before C-6-*O*-sulphation (Jacobsson & Lindahl, 1980), it can be deduced from these findings that the GlcNAc in -GlcNSO₃-IdUA-GlcNAc- sequences is a favoured substrate for the C-6-*O*-sulphotransferase. It should be noted that GlcNSO₃ residues in -IdUA(2SO₃)-GlcNSO₃- sequences are very good substrates for the C-6-*O*-sulphotransferase (Jacobsson & Lindahl, 1980); however, in the heparan sulphate structures described in the present paper, -IdUA(2SO₃)-GlcNAc- sequences do not occur, and hence an IdUA(2SO₃) residue in this position cannot be required for substrate recognition by the GlcNAc C-6-*O*-sulphotransferase.

Tetrasaccharides obtained from heparan sulphates by nitrous acid digestion have also been studied by Linker (1979), by using exo-enzymic degradation techniques. In this previous work, no unsulphated tetrasaccharides were identified, and monosulphated tetrasaccharides containing aManOH(6SO₃) residues were found to represent approx. 15% of the total tetrasaccharide preparation. The discrepancies between the frequency of occurrence of individual tetrasaccharides in this present study and that by Linker (1979) may derive from the different analytical techniques employed, and also, perhaps more importantly, from the use of a more highly sulphated heparan sulphate parent material in the earlier work.

These tetrasaccharides represent almost a quarter of the total heparan sulphate parent material, and thus sequence analysis of the tetrasaccharides provides a greater understanding of heparan sulphate sequence structure as a whole. As demonstrated in the present paper, high-field n.m.r. spectroscopy is a valuable technique for the determination of sulphate and uronate contents and locations, and consequently is of great use in the quantitative analysis of heparan sulphate sequence structures.

It is noteworthy that tetrasaccharide sequences of the form described in the present paper are relatively common in self-associating heparan sulphate species (Fransson, 1981). In addition, the sequence -IdUA-GlcNAc(6SO₃)-GlcUA-, which is found in approx. 40% of the tetrasaccharides in this preparation, also comprises

part of the antithrombin-binding sequence of heparin (Lindahl *et al.*, 1981). The C-6-*O*-sulphate group on the GlcNAc within this sequence has further been shown to be of critical importance to the interaction between heparin and antithrombin (Lindahl *et al.*, 1983).

We thank the Science and Engineering Research Council for a research assistantship and Glaxo Operations Ltd. for a C.A.S.E. studentship (to P. N. S.). Support from the Science and Engineering Research Council for use of their 400MHz n.m.r. facility is gratefully acknowledged. We are indebted to Dr. B. E. Mann and Dr. C. Spencer for assistance and valuable discussions, and to Dr. J. T. Gallagher and Dr. I. N. Hampson for the high-voltage electrophoresis experimentation and for helpful suggestions.

References

- Cifonelli, J. A. & King, J. (1977) *Biochemistry* **16**, 2137-2141
- Delaney, S. R. & Conrad, H. E. (1983) *Biochem. J.* **209**, 315-322
- Fransson, L.-Å. (1981) *Eur. J. Biochem.* **120**, 251-255
- Fransson, L.-Å. & Havsmark, B. (1982) *Carbohydr. Res.* **110**, 135-144
- Fransson, L.-Å., Nieduszynski, I. A. & Sheehan, J. K. (1980a) *Biochim. Biophys. Acta* **630**, 287-300
- Fransson, L.-Å., Havsmark, B., Nieduszynski, I. A. & Huckerby, T. N. (1980b) *Biochim. Biophys. Acta* **633**, 95-104
- Huckerby, T. N. (1983) *Org. Magn. Reson.* **21**, 67-70
- Huckerby, T. N. & Nieduszynski, I. A. (1982) *Int. J. Biol. Macromol.* **4**, 269-274
- Jacobsson, I. & Lindahl, U. (1980) *J. Biol. Chem.* **255**, 5094-5100
- Jacobsson, I., Höök, M., Pettersson, I., Lindahl, U., Larm, O., Wirén, E. & von Figura, K. (1979) *Biochem. J.* **179**, 77-87
- Jacobsson, I., Lindahl, U., Jensen, J. W., Rodén, L., Prihar, H. & Feingold, D. S. (1984) *J. Biol. Chem.* **259**, 1056-1063
- Kosakai, M. & Yosizawa, Z. (1981) *J. Biochem. (Tokyo)* **89**, 1933-1944
- Kraemer, P. M. (1971) *Biochemistry* **10**, 1437-1445
- Lindahl, U., Thunberg, L., Bäckstrom, G. & Riesenfeld, J. (1981) *Biochem. Soc. Trans.* **9**, 499-501
- Lindahl, U., Bäckstrom, G. & Thunberg, L. (1983) *J. Biol. Chem.* **258**, 9826-9830
- Linker, A. (1979) *Biochem. J.* **183**, 711-720
- Linker, A. & Hovingh, P. (1975) *Biochim. Biophys. Acta* **385**, 324-333
- Riesenfeld, J., Höök, M. & Lindahl, U. (1982) *J. Biol. Chem.* **257**, 7050-7055
- Sanderson, P. N., Nieduszynski, I. A. & Huckerby, T. N. (1983) *Biochem. J.* **211**, 677-682
- Shively, J. E. & Conrad, H. E. (1976) *Biochemistry* **15**, 3932-3942
- Winterbourne, D. J. (1982) *Biosci. Rep.* **2**, 1009-1015
- Winterbourne, D. J. & Mora, P. T. (1981) *J. Biol. Chem.* **256**, 4310-4320