Oligosaccharide Structure and Amino Acid Sequence of the Major Glycopeptides of Mature Human β-Hexosaminidase†

Brian F. O'Dowd,‡,§ Dale A. Cumming,‖ Roy A. Gravel,‡,‖ and Don Mahuran*,‡,§

Research Institute, Hospital for Sick Children, 555 University Avenue, Toronto, Ontario, Canada M5G 1X8, and Department of Clinical Biochemistry and Department of Medical Genetics, University of Toronto, Toronto, Ontario, Canada M5S 1A8

Received November 12, 1987; Revised Manuscript Received January 29, 1988

ABSTRACT: Human β-hexosaminidase (EC 3.2.1.52) is a lysosomal enzyme that hydrolyzes terminal *N*-acetylhexosamines from GM_2 ganglioside, oligosaccharides, and other carbohydrate-containing macromolecules. There are two major forms of hexosaminidase: hexosaminidase A, with the structure $\alpha(\beta_a\beta_b)$, and hexosaminidase B, $2(\beta_a\beta_b)$. Like other lysosomal proteins, hexosaminidase is targeted to its destination via glycosylation and processing in the rough endoplasmic reticulum and Golgi apparatus. Phosphorylation of specific mannose residues allows binding of the protein to the phosphomannosyl receptor and transfer to the lysosome. In order to define the structure and placement of the oligosaccharides in mature hexosaminidase and thus identify candidate mannose 6-phosphate recipient sites, the major tryptic/chymotryptic glycopeptides from each isozyme were purified by reverse-phase high-performance liquid chromatography. Two major concanavalin A binding glycopeptides, localized to the β_b chain, and one non concanavalin A binding glycopeptide, localized to the β_a chain, were found associated with the β-subunit in both hexosaminidase A and hexosaminidase B. A single major concanavalin A binding glycopeptide was found to be associated with the α subunit of hexosaminidase A. The oligosaccharide structures were determined by nuclear magnetic resonance spectrometry. Two of them, the α and one of the β_b glycans, contained a Man_3-$GlcNAc_2$ structure, while the remaining one on the β_b chain was composed of a mixture of Man_{5-7}-$GlcNAc_2$ glycans. The unique glycopeptide associated with the β_a chain contained a single GlcNAc residue. Thus, all three mature polypeptides comprising the α and β subunits of hexosaminidase contain carbohydrate, the structures of which have the appearance of being partially degraded in the lysosome. In the α chain we found only one possible site for in vivo phosphorylation. In the β it is unclear if only one or all three of the sites could have contained phosphate. However, mature placental hexosaminidase A and B can be rephosphorylated in vitro. This requires the presence of an oligosaccharide containing an α1,2-linked mannose residue. Only the single Man_{6-7} (of the Man_{5-7}-$GlcNAc_2$ glycans) containing site on the β_b chain retains this type of residue. Therefore, this site may act as the sole in vitro substrate in both of the mature isozymes for the phosphotransferase.

β-Hexosaminidase (EC 3.2.1.52) is a lysosomal hydrolase. There are two major isozymes in normal human tissues, hexosaminidase A and B. Hexosaminidase A is composed of two subunits, an α subunit, consisting of a single polypeptide chain, and a β subunit containing two peptides, β_a and β_b, derived by cleavage of a common pro-β precursor. Its structure is $\alpha(\beta_a\beta_b)$. Hexosaminidase B, on the other hand, contains two β subunits, $2(\beta_a\beta_b)$. Both enzymes are able to hydrolyze many of the same natural and artificial substrates; however, only hexosaminidase A, in combination with a specific activator protein, can hydrolyze the terminal *N*-acetylgalactosamine residue from GM_2 ganglioside[1] [reviewed in Mahuran et al. (1985)].

Like other lysosomal, secretory, and plasma membrane proteins, the hexosaminidase isozymes are glycoproteins and must be specifically targeted to their destination by the cell. Much has been learned about the mechanism that the cell employs to identify glycoproteins destined for the lysosome [reviewed in Goldberg et al. (1984) and von Figura and Hasilik (1986)]. At least one oligosaccharide, contained on lysosomal enzymes, must become phosphorylated by the addition of a GlcNAc-phosphate residue to carbon 6 of an α1,2-linked mannose residue (Couso et al., 1986). This reaction is catalyzed by the enzyme *N*-acetylglucosaminylphosphotransferase, situated in the cis Golgi [reviewed in Kornfeld and Kornfeld (1985)]. The terminal GlcNAc residue is subsequently removed by *N*-acetylglucosamine-1-phosphodiester-*N*-acetylglucosaminidase, exposing the mannose 6-phosphate recognition site on the protein. Lysosomal proteins achieve compartmentalization through binding of the phosphomannosyl moiety to the phosphomannosyl receptor and translocation of the enzyme–receptor complex to the lysosome (Natowitz et al., 1983).

Several lines of evidence have demonstrated that hexosaminidase is transported via the phosphomannosyl receptor pathway. First, both the pro-α and the pro-β polypeptide chains of hexosaminidase have been shown to contain a phosphorylated high-mannose oligosaccharide(s) and are thus in vivo substrates for the phosphotransferase enzyme (Hasilik & Neufeld, 1980). Second, mature hexosaminidase which has

† Supported by Program Grant PG-4 from the Medical Research Council of Canada.

* Author to whom correspondence should be addressed at the Hospital for Sick Children.

‡ Hospital for Sick Children.

§ Department of Clinical Biochemistry, University of Toronto.

‖ Department of Medical Genetics, University of Toronto.

[1] Abbreviations: GM_2, GalNAcβ(1–4)[NANa(2–3)–]Galβ(1–4)-Glc-ceramide; GlcNAc or Gn, *N*-acetylglucosamine; Man, mannose; Con A, concanavilin A; TFA, trifluoroacetic acid; NMR, nuclear magnetic resonance; RP-HPLC, reverse-phase high-performance liquid chromatography; Endo H, endoglycosidase H.

0006-2960/88/0427-5216$01.50/0

lost its original phosphate through degradation in the lysosome can be rephosphorylated in vitro with a partially purified preparation of the phosphotransferase (Goldberg et al., 1984). This reaction occurred in preference to the phosphorylation of glycoproteins not destined for the lysosome and suggested the existence of a unique protein structural component that functions as a recognition site for the phosphotransferase. Finally, in phosphotransferase-deficient fibroblasts from patients with I-cell disease (Reitman et al., 1981), hexosaminidase, along with numerous other lysosomal enzymes, fails to be phosphorylated and is secreted by cells rather than incorporated into lysosomes (Hasilik & Neufeld, 1980).

To begin to understand the details of the biosynthesis and the mechanism directing hexosaminidase to the lysosome, it is first necessary to define the location and structure of the attached oligosaccharides on the mature isozymes. This would allow identification of candidate mannose 6-phosphate recipient sites in the pro forms and focus the search for peptide-associated recognition markers (Lang et al., 1984) to nearby sequences. We have previously shown that the mature α and β_b polypeptides bind Con A lectin, indicative of the presence of attached high mannose type oligosaccharides (O'Dowd et al., 1985). We now report the location and structure of these oligosaccharides and localize a single GlcNAc residue to the β_a polypeptide. In a previous paper, we described the amino terminus of the α, β_a, and β_b polypeptides of hexosaminidase (Mahuran et al., 1988). Together these data provide a near complete description of the primary structure of mature hexosaminidase A and B.

Materials and Methods

Isolation of the Glycopeptides from Hexosaminidase A and B. Hexosaminidase A and B were purified 8000-fold from human placenta (Mahuran & Lowden, 1980). The basic method for the isolation of Con A binding tryptic glycopeptides derived from hexosaminidase A and B by reverse-phase HPLC has been described (O'Dowd et al., 1985). To improve the yields of the glycopeptides, siliconized collection tubes (Prosil-28, Speciality Chemicals) were used throughout the procedure. Two lots of hexosaminidase B (30 mg, i.e., 600 nmol of the β subunit, each) and hexosaminidase A (20 mg, i.e., 200 nmol of the α and 200 nmol of the β subunits per lot) from separate enzyme preparations were precipitated from concentrated stocks by the addition of 4 volumes of acetone. The protein pellets were denatured in 7 M guanidine hydrochloride, reduced, and alkylated, and the protein was reprecipitated with 10 volumes of ethanol–glacial acetic acid, 99:1. In order to obtain sufficiently small glycopeptides that would decrease the likelihood of specifically losing one due to a lack of solubility, each protein pellet, suspended in 15 mL of 0.1 M NH_4HCO_3, was digested with an excess of trypsin [ratio of 50:1 (mg/mg) protein:trypsin, DPCC treated from Sigma), for an extended period of time (15 h) at 37 °C. Due to the presence of residual chymotryptic activity in the trypsin preparation, these reaction conditions were sufficiently severe as to cause the hydrolysis of most of the major and some of the minor chymotryptic sites as well as all of the tryptic sites (Smyth, 1970) in the proteins. The digests of the hexosaminidase isozymes were lyophilized and the peptide fragments separated from the residual protease by molecular sieve HPLC as previously described (O'Dowd et al., 1985).

One lot of lyophilized tryptic/chymotryptic peptides was loaded onto the Con A column, and the column was washed with several hundred column volumes of buffer. Con A binding glycopeptides were released by rotating the tube overnight with a 15% solution of methyl α-mannoside. The resin was subsequently centrifuged and the supernatant removed. The solution was lyophilized, and the glycopeptides and mannoside were dissolved in 0.1% TFA in water and passed through a C18 Sep-pack (supplied by Waters). The C18 Sep-pack was washed (0.1% TFA; 12 mL) to remove the unbound mannoside from the bound glycopeptides. The glycopeptides were eluted (2 × 4 mL) with 2-propanol–acetonitrile (1:2). The solution was lyophilized and the residue dissolved in 0.1% TFA (1 mL). RP-HPLC separation (5 × 200-μl aliquots) of the glycopeptides was performed on a 5-μm ODS, 25-cm reverse-phase Altex column.

The above procedure was repeated with the second lot of enzyme except that a 3-μm ODS, 7.5-cm Altex column was used for the final separation. Further purification and/or analytical separations after digestion with Endo H (Genzyme) or glycopeptidase F (N-glycanase, Genzyme) (see below) were performed on either a 5-μm Ultrasphere ODS reverse-phase Altex column (Beckman), 4.6 × 250 mm; a 3-μm Ultrasphere ODS, 4.6 × 75 mm Altex column (Beckman); or a 5-μm Sepralyte ODS reverse-phase Jones column, 4.6 × 250 mm (Analytichem International), as noted under Results and in the figure legends. The glycopeptides were eluted with a gradient of increasing amounts of acetonitrile/2-propanol (2:1) containing 0.06% TFA, and the peptides were detected at 214 nm (O'Dowd et al., 1985).

A third lot of a similar amount of each isozyme was digested, as above, and each of the total digests was separated by RP-HPLC on a Jones Sperisorb 5-μm ODS 25-cm column. Fractions were collected so that no fraction contained more than four closely eluting peptide peaks. Glucosamine analysis was carried out on each of the Con A binding glycopeptides and each of the fractions from the total digest to confirm the presence of oligosaccharide (O'Dowd et al., 1985). Fractions from the total digest that proved to be GlcNAc positive were repurified by reverse-phase and molecular sieve HPLC (Korneluk et al., 1986), and each peak was retested for GlcNAc.

^1H NMR Spectroscopy of the Con A Binding Oligosaccharides from Hexosaminidase. Glycopeptides were prepared for ^1H NMR spectroscopy by repeated dissolution–lyophilization in D_2O (99.8 atom % D, Merck-Sharp-Dohme Canada, Ltd.) after concentration in a 0.1-mL cylindrical microcell (Wilmad Glass Co., Buena, NJ) with a speed-vac concentrator (Savant Instruments). After final filling of the microcell with D_2O (99.96 atom % D) under an argon atmosphere, the microcell was sealed and inserted in a standard 5-mm NMR tube. NMR spectra were recorded on a Nicolet NT-360 spectrometer or Bruker AM-500 spectrometer located at the Centre for the Determination of Carbohydrate Structure, University of Toronto. Procedures for the 360-MHz ^1H NMR spectroscopy have been described (O'Dowd et al., 1985). The Bruker AM-500 spectrometer was operated in the Fourier-transform mode at 50 or 70 °C under the following conditions: a 16K data set was acquired with a 72° (8 μs) observe pulse after 4 s of presaturation of the HDO peak with the homonuclear decoupler (gated off acquisition). Full quadrature phase cycling and a 5000-Hz sweep width were employed and approximately 8000 transients accumulated. Prior to transformation to the frequency domain, the accumulated free-induction decay was apodized by Gaussian multiplication and the data set zero-filled to 32K. Chemical shifts were measured relative to that of acetone at 2.225 ppm.

The amount, in nanomoles, of oligosaccharide present in the microcell was determined from the observed signal to noise ratio and the number of accrued accumulations (prior to

Table I: Summary of β-Hexosaminidase Peptides Containing Putative Glycosylation Sites

	amino acid sequence [deduced (cDNA); found (isolated peptide)]	GlcNAc	sensitivity[g] Endo H	Glyco F	Con A
	Mature α Subunit				
(1) cDNA	LVVSVVTPGCNQLPTLESVENYT				
peptide A4[a]	VVSVVTPGCN	+	−	nd	+
(2) cDNA	FFINKTEIEDFPR[b]				
peptide	not found				−
(3) cDNA	NPSLNNTYEFM[c]				
peptide	not found				−
	Mature $\beta_a\beta_b$ Subunit				
β_b chain					
(1) cDNA	YISHSPNSTAGPTA[d]				
peptide	not found				
(2) cDNA	QSECDAFPNISSD				
peptide A3[a]	SECDAFP-IS	+	−	±	+
peptide B3[a]	SECDAFP-IS	+	−	±	+
(3) cDNA	FTINESTIIDSPR[b]				
peptide A1[a]	TI-ESTIIDS	+	+	+	+
peptide B1[a]	TI-ESTIIDS	+	+	+	+
peptide A2[a]	TI-ESTIIDS	+	−	+	+
peptide B2[a]	TI-ESTIIDS	+	−	nd	+
β_a chain					
(4) cDNA	KLDSFGPINPTLNTTYSFL[c]				
peptide[a]	LDSFGPINPT[e]	+	−	nd	−
N-terminus	(K)LDSFGPINPTL-TTYSFL[e]				
(5) cDNA	YVDATNLTPR				
peptide	VDATNLTPR[f]	nd	nd	nd	−

[a] The amino acids found in the first ten "turns" of the sequenator from the RP-HPLC isolated glycopeptide. [b] First coincident, putative glycosylation sites in the aligned primary structures of the α and β chains. [c] Second coincident, putative glycosylation sites in the aligned primary structures of the α and β chains. [d] This peptide is not present in the mature protein (Figure 2) (Mahuran et al., 1988). [e] The single non Con A binding glycopeptide from hexosaminidase B compared with the partial N-terminus of the β_a chain from hexosaminidase A (K residue not found) and B (K residue present) (Mahuran et al., 1988). [f] Previously isolated peptide (Korneluk et al., 1986). [g] Abbreviations: nd, not determined; Glyco F, glycopeptidase F; (−) in amino acid sequence, no amino acid was detected in this turn of the sequenator (indicates the presence of an attached oligosaccharide); (+), (±), and (−) in the sensitivity columns, good, poor, and no, respectively.

Fourier transformation) for each sample as compared to those of standards of known concentrations.

Endoglycosidase Treatment of the Glycopeptides from Hexosaminidase. Positive evidence that the oligosaccharides attached to isolated glycopeptides from hexosaminidase A or B were substrates for the specific glycosidase enzymes was taken to be the detection of an increase in the retention time of the treated (deglycosylated) peptides versus those of the original glycopeptides on RP-HPLC, reflecting an increased hydrophobicity due to the removal of the hydrophilic oligosaccharide. In some cases this was confirmed by amino acid sequencing. The GlcNAc-positive peaks from the HPLC from hexosaminidase A and B (dissolved in D_2O) were recovered from the NMR microcell, and 5-nmol aliquots were lyophilized and incubated with either (1) Endo H (6 μL of a 74 μg/mL stock) in 0.1 M sodium citrate, pH 5.5, at 37 °C overnight (total reaction volume 40 μL), or (2) glycopeptidase F (2 μL of a 5 unit/18 μL stock) incubated at room temperature overnight in 0.1 M NH_4HCO_3 buffer, pH 8.6 (total reaction volume 40 μL). The digestion mixtures were diluted in 0.1% TFA to 250 μL and the peptides separated by RP-HPLC. Each peak was hand collected as it emerged from the detector and lyophilized for amino acid sequencing.

Amino Acid Sequence Analysis of the Con A Binding Glycopeptides. The sequence of the first 10 amino acid residues from each of the original and/or deglycosylated glycopeptides (2–10 nmol) recovered from the reverse-phase column was determined in a Beckman 890c sequenator. Phenylthiohydantoin derivatives of the amino acids were identified by HPLC (Peckett et al., 1984). In some cases the presence of the attached oligosaccharide or residual GlcNAc residue after Endo H treatment on the predicted Asn of the putative glycosylation site was confirmed by the occurrence of a "blank" in the corresponding sequenator "turn". When glycopeptidase F was used, the predicted Asn residue was converted to Asp.

Results

Analysis of the Con A Binding from Hexosaminidase B. Initially, we examined the glycopeptides of hexosaminidase B, since it contains only β subunits. The Con A binding glycopeptides were resolved by separation on RP-HPLC with an Altex 5-μm, 25-cm ODS column. Two major glucosamine-containing glycopeptide peaks, designated B1 (≃75 nmol as determined by NMR, 12% yield) and B3 (≃50 nmol as determined by NMR, 8% yield), and one minor glycopeptide peak, B2 (≃5 nmol, estimated from OD_{214} with B1 as a standard), were identified (Figure 1A). The sequence of the first 10 amino-terminal amino acid residues from B1, B2, and B3, recovered from the reverse-phase column, was determined and compared to the deduced amino acid sequence of the hexosaminidase B cDNA clone (Table I). The amino acid sequences of B1 and B2 were identical and correspond, with a perfect match, to residues 188–197 of the deduced amino acid sequence. The sequence of the second peptide, B3, was similarly localized to a single region, residues 136–145 on the cDNA clone, 20 residues from the previously determined mature amino terminus of the β_b polypeptide chain (Mahuran et al., 1988) (Figure 2).

Two of the glycopeptides, B1 and B3, were available in sufficient quantities for structural analysis of the attached glycans by ^1H NMR spectroscopy at 360 MHz. A preliminary characterization of B1 has been made by ^1H NMR carried out at 70 °C (O'Dowd et al., 1985). The availability of larger quantities of B1 now allowed us to obtain spectra at 23 °C, a temperature which provides a better resolution of chemical shifts (Figure 3). For example, this would allow us to resolve

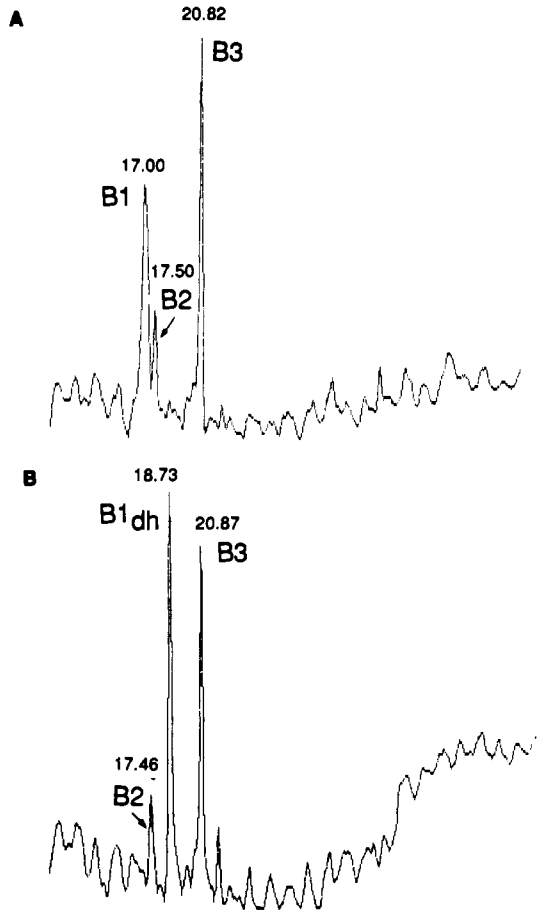

FIGURE 1: (A) Elution profile of RP-HPLC (Altex 5 μm, 25 cm) separation of the Con A binding glycopeptides, B1, B2 and B3, from tryptic/chymotryptic digests of hexosaminidase B. (B) The above hexosaminidase B glycopeptides were incubated with Endo H and rechromatographed on the RP-HPLC. The time of elution (in minutes) of each peak is indicated. Peaks were detected by their UV absorption at 214 nm, and the largest was automatically scaled to 100%. $B1_{dh}$ indicates the deglycosylated (based on its increased retention time) form of B1.

the 6^{i2}-H1 and 3^{t}-H1 signals which overlap in spectra obtained at 70 °C. The 6^{i2}-H1 signal is important for identifying mannoses in α1,2 linkages. As observed earlier (O'Dowd et al., 1985), anomeric hydrogen signals were obtained for glycopeptide B1, consistent with a glycan mixture composed principally of Man_5-$GlcNAc_2$-Asn (Figure 4A) and Man_6-$GlcNAc_2$-Asn (Figure 4B). The former structure is indicated by Man(α1,6)-H1 signals at 4.900 (6^{t}) and 4.870 ppm (6^{i}) and a Man(α1,3)-H1 signal at 5.090 ppm (3^{t}) (Figure 3). The Man_6-$GlcNAc_2$-Asn structure is indicated by a Man 3^{i2}-H1 signal at 5.350 ppm and a Man 2^{t}-H1 signal at 5.050 ppm. However, in addition, a new signal (6^{i2}, Figure 3) was observed at 5.150 ppm, a chemical shift ascribable to the anomeric hydrogen of a 2-O-substituted Man(α1,6) residue (Carver et al., 1981; Takahashi et al., 1983). This observation could be accounted for by the addition of a Man(α1,2) residue to the 6^{t} residue of Man_5-$GlcNAc_2$-Asn, resulting in a second Man_6 isomer (Figure 4C), or to the 6^{t} residue of Man_6-$GlcNAc_2$-Asn, resulting in a Man_7-$GlcNAc_2$-Asn structure (Figure 4D). The two possibilities could not be resolved by ^1H NMR spectroscopy. The chemical shifts for the glycan mixture and those of the authentic high-mannose oligosaccharides are given in Table II. Glycopeptide B3 was also investigated by ^1H NMR. Figure 5A shows a partial ^1H NMR spectrum of B3 obtained at 23 °C. Only three diagnostic signals are clearly discernible. A 9-Hz doublet is evident at 5.026 ppm and is

Table II: Chemical Shifts (ppm) of B1 Glycans at 23 °C

residue	Hex B	Δ[a]	M_5Gn_2-Asn	M_6Gn_2-Asn	M_7Gn_2-Asn
3^{t}	5.090	0.005	5.095	5.090	5.090
3^{i2}	5.350	0.002		5.348	5.352
6^{t}	4.900	0.010	4.910	4.910	
6^{i}	4.870	0.005	4.875	4.868	4.870
6^{i2}	5.150	0.000		5.150	5.150
2^{t}	5.050	0.001		5.049	5.049

[a] Largest chemical shift difference between the appropriate Hex B glycopeptide signal and the analogous signal from either ovalbumin (Carver et al., 1981) or cathepsin D (Takahashi et al., 1983) in Man_5-$GlcNAc_2$-Asn, Man_6-$GlcNAc_2$-Asn, or Man_7-$GlcNAc_2$-Asn.

Table III: Chemical Shifts (ppm) of Glycopeptide B3 (Figure 5) and an Authentic Man_3-$GlcNAc_2$-Asn-Linked Glycopeptide at 23 °C

residue	Hex B	Δ[a]	M_3Gn_2-Asn
4^{i}	4.770	0.005	4.775
3^{t}	5.098	0.000	5.098
6^{t}	4.912	0.003	4.915
Asn-Gn	5.026	0.002	5.028
core-Gn	4.600	0.020	4.620
core-GnNHAc	2.076		
Asn-GnNHAc	2.007		

[a] Largest chemical shift difference between the appropriate B3 glycopeptide signal and the analogous signal of a standard Man_3-$GlcNAc_2$-Asn glycopeptide from cathepsin D (Takahashi et al., 1983).

ascribable to a GlcNAc residue linked to the side chain of Asn. The other two signals at 5.098 (3^{t}) and 4.912 ppm (6^{t}) are characteristic of unsubstituted Man(α1,3) and Man(α1,6) anomeric hydrogen, respectively (Takahashi et al., 1983; Grey et al., 1982). The absence of any other mannose anomeric signals and the presence of only two high-field *N*-(acetylmethyl) signals at 2.076 and 2.007 ppm (ascribable to the core and Asn-linked *N*-acetylglucosamine residues of an N-linked glycan, data not shown) lead us to conclude that a single glycan structure is found at this glycosylation site, the Man_3-$GlcNAc_2$-Asn structure shown in Figure 5B. The presence of the Man(β1,4)-GlcNAc (β1,-) disaccharide moiety is presumptive upon the observed chemical shifts of the other residues and their agreement with the known chemical shifts of this glycan. The observed chemical shifts for the glycopeptide and those of the authentic glycopeptide from cathepsin D are given in Table III.

The Con A binding glycopeptides, B1, B2, and B3 (Figure 1A), were incubated with Endo H (Figure 1B). The RP-HPLC glycopeptide peak corresponding to B1, which had previously eluted at a time of 17.00 min, was no longer present, and a new peak representing the deglycosylated form of B1, $B1_{dh}$ (retaining a single GlcNAc residue), eluted at a time of 18.73 min (Figure 1B), reflecting the loss of its hydrophilic oligosaccharide moiety. This observation was confirmed by finding identical amino acid sequences for the peptide contained in the $B1_{dh}$ peak and that in the untreated B1 peak. The retention times of glycopeptides B2 and B3 were not affected by treatment with Endo H (Figure 1B). This finding for B3 is consistent with the observation that $Man_{\leq 3}$-$GlcNAc_2$-Asn glycopeptides are not substrates for Endo H (Tai et al., 1977), confirming the finding by NMR that B3 lacks any larger high-mannose oligosaccharide structures.

Analysis of the Con A Binding Glycopeptides from Hexosaminidase A. We next examined the Con A-Sepharose-binding glycopeptides of hexosaminidase A. They were separated by RP-HPLC using the Altex 3-μm, 7.5-cm ODS column, and the glycopeptides of hexosaminidase B were run on this column for comparison. Four glucosamine-positive peaks, labeled A1, A2, A3, and A4 (Figure 6A), were resolved

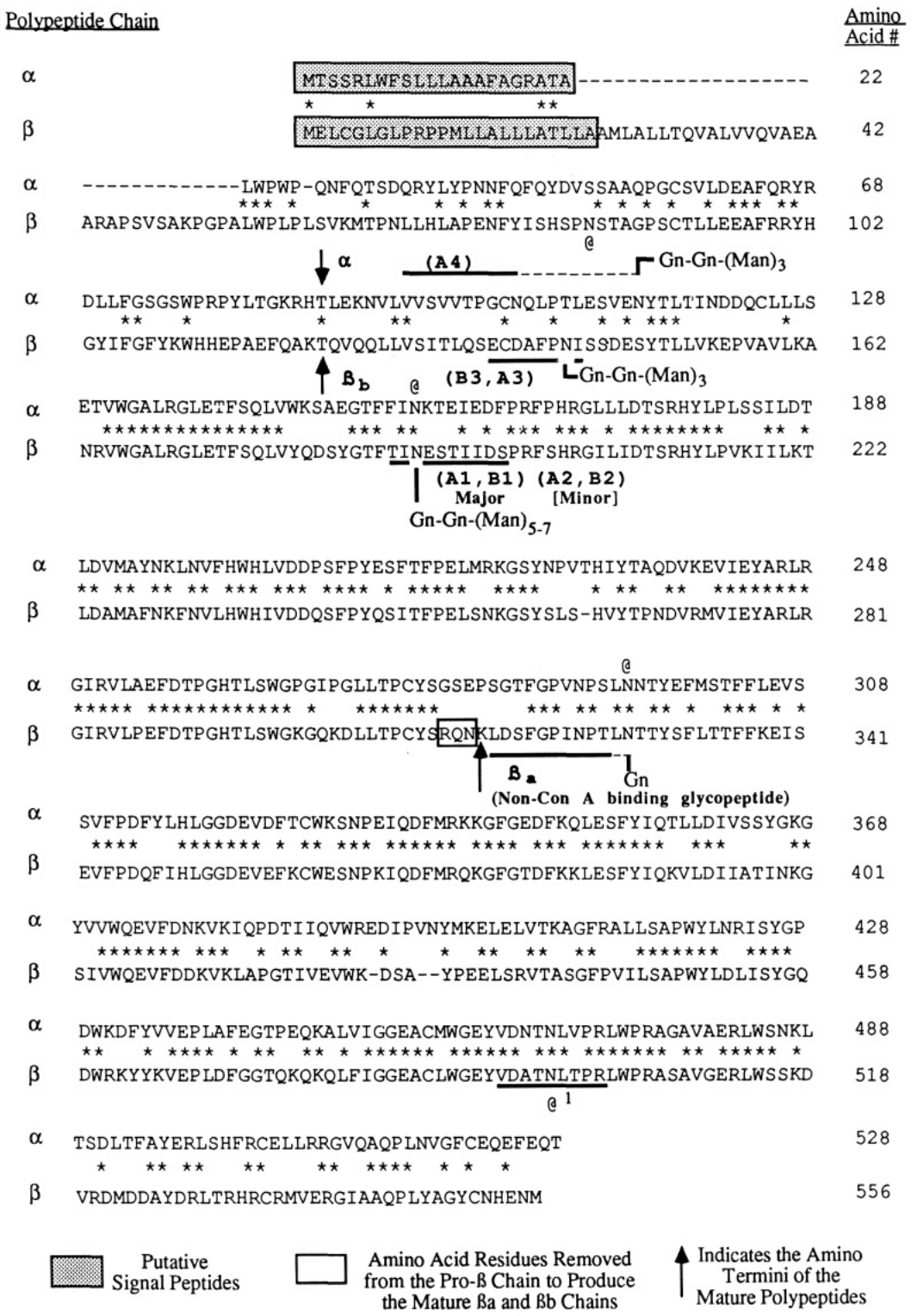

FIGURE 2: Deduced primary structures of the α and β prepro polypeptide chains of hexosaminidase. Homologies between the two primary structures are indicated by (*). The localization and glycan structures of the major glycopeptides (A1, B1; A3, B3; A4) and the non Con A binding β_a glycopeptide are indicated. The minor Con A binding glycopeptides, A2 and B2, whose glycan structure was not determined, were localized to the same position as A1 and B1. The positions of the putative signal peptides, the mature amino termini of each chain, and the internal hydrolysis site in the pro-β chain that gives rise to the two mature β_b and β_a chains are indicated (Mahuran et al., 1988).

from hexosaminidase A. Two other small peaks eluting earlier in the gradient (designated a and b in Figure 6A) were not GlcNAc positive. Amino acid sequencing of these peaks revealed that while they were authentic α chain peptides, they did not contain a putative N-linked oligosaccharide attachment site.

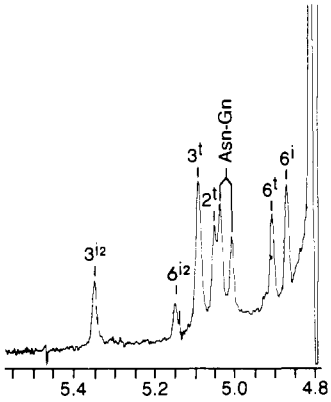

FIGURE 3: Partial 360-MHz ^{1}H NMR spectrum of the Con A binding tryptic/chymotryptic glycopeptide, B1, from hexosaminidase B (Figure 1). The glycopeptide spectrum was recorded at 23 °C as described under Materials and Methods. Nomenclature for the various monosaccharide residues and their anomeric signals is that of Carver et al. (1981).

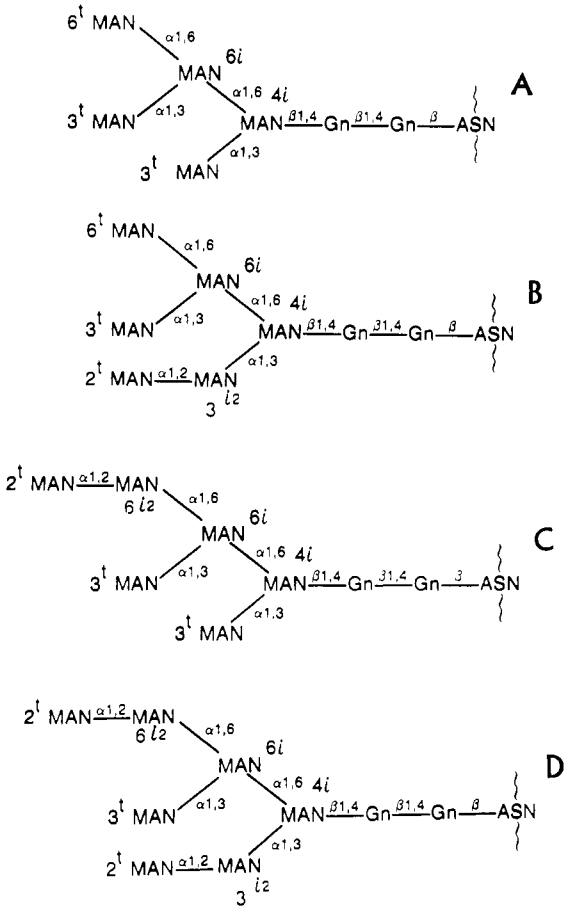

FIGURE 4: Structures of the glycans on the tryptic/chymotryptic glycopeptide B1 from hexosaminidase B and A1 from hexosaminidase A.

By inspection of the HPLC profile, it appeared that several of the glycopeptides of hexosaminidase A correspond to those seen in hexosaminidase B (compare panels A and B of Figure 6). Thus, glycopeptides A1 and B1 show similar retention times, as do A2 and B2, and A3 and B3. A4 was unique to the hexosaminidase A preparation. Amino acid sequencing (first 10 residues) of the amino terminus of the associated peptides (Table I) identified the attachment sites of the glycopeptides in the primary sequence of the α or β subunits. The partial peptide sequences of A1 and B1 (as well as A2 and B2, see below) were identical, indicating that they are attached to the same site in the β subunit in both hexosaminidases A

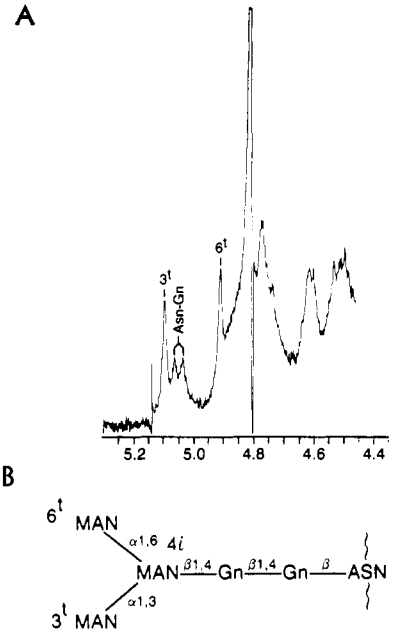

FIGURE 5: (A) Partial 360-MHz ^{1}H NMR spectrum of the Con A binding tryptic/chymotryptic glycopeptide, B3, from hexosaminidase B (Figure 1). The glycopeptide spectrum was recorded at 23 °C as described under Materials and Methods. Nomenclature for the various monosaccharide residues and their anomeric signals is that of Carver et al. (1981). (B) Structures of the glycans of the Con A binding glycopeptide, B3, isolated from hexaminidase B.

Table IV: Chemical Shifts (ppm) of Glycopeptide A1 and Standard Ovalbumin Glycans at 70 °C

residue	Hex A	Δ[a]	M_5Gn_2-Asn[b]	M_6Gn_2-Asn[b]	M_7Gn_2[c]
3^{i2}	5.338	0.003		5.336	5.341
3^t	5.127	0.009	5.124	5.126	5.118
2^t	5.058, 5.067	0.002		5.069	5.064
Asn-Gn	5.023	0.045	5.068	5.068	
6^t	4.920	0.002	4.918	4.920	
6^i	4.881	0.003	4.880	4.883	4.884
4^i	4.768	0.004	4.772	4.765	4.773
core-Gn	4.629	0.009	4.636	4.638	
core-GnNHAc	2.065	0.002	2.063	2.063	
Asn-GnNHAc	2.023	0.002	2.021	2.022	

[a] Largest chemical shift difference between the appropriate A1 glycopeptide signal and the analogous signal from ovalbumin in either Man_5-$GlcNAc_2$-Asn or Man_6-$GlcNAc_2$-Asn, whichever is larger. [b] Shift values of ovalbumin glycan standards from Carver et al. (1981). [c] Man_7-$GlcNAc_2$ Endo H released oligosaccharide (Hakimi et al., 1981).

and B. Similarly, A3 and B3 were localized to a single putative glycosylation site and are, therefore, also β subunit glycopeptides. Only A4 showed a unique N-terminus, and it was localized within the primary sequence of the α subunit (Figure 2).

The ^{1}H NMR spectroscopy of the hexosaminidase A glycopeptides was carried out with a 500-MHz instrument at 50 °C (Figure 7) and 70 °C (spectra not shown, but data in Table IV are given for 70 °C due to a lack of published standards at 50 °C). The higher temperatures and the use of the 500-MHz instrument allowed resolution of spectra for the smaller amounts of material available from the hexosaminidase A digests. As in glycopeptide B1, the set of observed chemical shifts for glycopeptide A1 (≃15 nmol, 8% yield) (Table IV) is consistent with the presence of a mixture of high mannose type structures. Man_5-$GlcNAc_2$-Asn (Figure 3A) is again the dominant glycan present with a mixture of either Man_6-

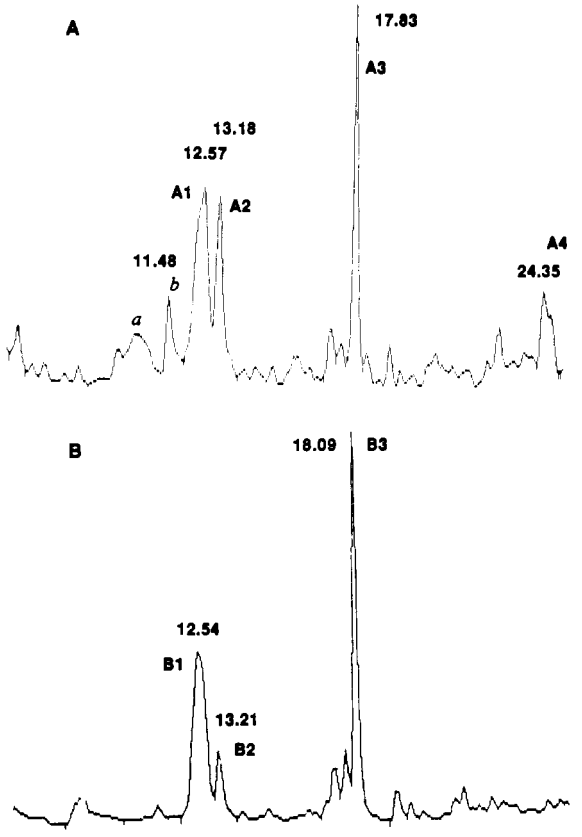

FIGURE 6: Preparative RP-HPLC separation of the concanavalin A binding glycopeptides from tryptic/chymotryptic digests of (A) hexosaminidase A and (B) hexosaminidase B (separated on the Altex Ultrasphere, ODS 3 μm, 7.5 cm). The time of elution (in minutes) of each peak is indicated. Peaks were detected by their UV absorption at 214 nm, and the largest was automatically scaled to 100%. The amino acid sequences of the labeled peaks are shown in Table I, and their location in the primary structure of the polypeptides is shown in Figure 2.

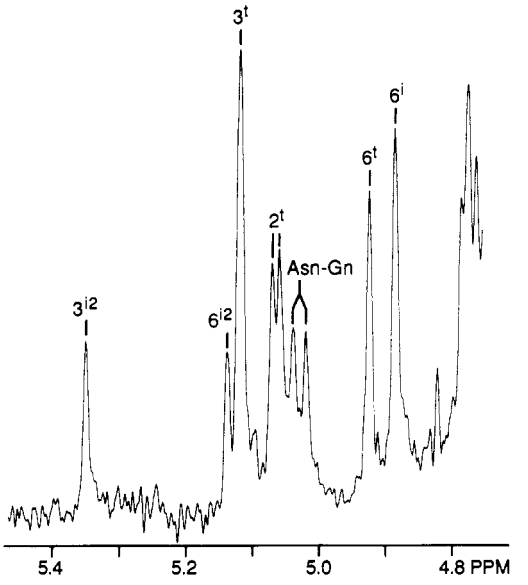

FIGURE 7: Partial 500-MHz ^{1}H NMR spectrum of the Con A binding tryptic/chymotryptic glycopeptide, A1, from hexosaminidase A (Figure 6A). The glycopeptide spectrum was recorded at 50 °C as described under Materials and Methods. Nomenclature for the various monosaccharide residues and their anomeric signals is that of Carver et al. (1981). The chemical shifts indicate the presence of one Man_5 oligosaccharide, two isomeric Man_6, and a Man_7 (structures outlined in Figure 4).

$GlcNAc_2$ structural isomers (Figure 3B,C) or Man_6 isomers plus a single Man_7 isomer (Figure 3D). For the Man_5-$GlcNA_2$-Asn structure, the two Man(α1,6)-H1 signals are observed at 70 °C at 4.920 (6^t) and 4.881 ppm (6^i), while the two Man(α1,3)-H1 signals are at 5.127 ppm (3^t, double intensity) (these signals can be identified at approximately the same positions in the spectrum taken at 50 °C in Figure 7). While the core-GlcNAc H1 signal cannot be discerned due to spectral overlap with peptidyl signals, the 9-Hz Asn-GlcNAc H1 doublet is clearly evident at 5.023 ppm (Figure 7). The presence of terminal Man(α1,2) residues in either Man_6 or Man_7 structures can be deduced from the characteristic downfield displacement of 2-O-substituted Man H1 signals, i.e., substitution of the 3^t or 6^t mannose residues. Substitution at the 3^t position produces the 3^{i2}-H1 signal which is displaced downfield from 5.127 (unsubstituted, 3^t) to 5.338 ppm (substituted) at 70 °C (Table IV; also observable at 50 °C, Figure 7). Substitution at the 6^t position gives the 6^{i2}-H1 signal at 5.137 ppm, detected only at 50 °C, displaced from 4.923 (Figure 7). In addition, the H1 signals of the terminal Man(α1,2) residues, 2^t, are observed as a doublet, reflecting the two different structures (compare structures B and D of Figure 4) at 5.067 and 5.058 ppm (70 °C, Table IV). Finally, the identification of the glycans of A1 as only of the high-mannose type is further substantiated by the observation of only two high-field *N*-(acetylmethyl) signals at 2.065 and 2.023 ppm (70 °C, Table IV). We therefore conclude that glycopeptides, A1 and B1 are identical, corresponding to a conserved glycosylation site and glycan within the β subunits of hexosaminidase A and B, respectively.

Insufficient amounts of the minor A2 glycopeptide were available for investigation by ^{1}H NMR; however, a partial characterization of the attached oligosaccharides was attempted by chemical and enzymatic means. A mixture containing predominantly glycopeptide A2 with some residual A1, initially isolated on the Altex Ultrasphere ODS 3-μm, 7.5-cm reverse-phase column, was rechromatographed on the Jones Sperisorb ODS 5-μm, 25-cm reverse-phase column. This column resolved the A2 peak into two smaller peaks at 20.44 (labeled A2 in Figure 8-1) and 22.84 min (labeled α). The amino acid sequence of the 22.84-min peak was determined and mapped to the α chain, although it did not contain a possible oligosaccharide attachment site (data not shown). Aliquots of the mixture (in Figure 8-1) were treated with two endoglycosidase enzymes, Endo H (cleaves core GlcNAc–GlcNAc glycosidic bonds of high-mannose structures larger than Man_3) and glycopeptidase F (removes the entire oligosaccharide by hydrolysis of the GlcNAc–Asn bonds of most N-linked oligosaccharides) (Figure 8, panels 2 and 3). Endo H produced a loss of the residual A1 peak (19.83 min in Figure 8-1) and gain of a new sharp peak labeled $A1_{dh}$ (21.93 min, Figure 8-2). A similar shift was observed for Endo H treated and untreated glycopeptide B1 of hexosaminidase B (Figure 1). Glycopeptide A2 from hexosaminidase A, however, was not affected by Endo H (retention of peak at 20.32 min, Figure 8-2). Also, as expected, the retention time of the unglycosylated α peak remained essentially unchanged (22.73 min in panel 2). The increase in separation between $A1_{dh}$ and A2 allowed A2 to be isolated in a pure form for amino acid sequencing. The minor glycopeptide A2 (from Figure 8-2) was sequenced and found to have the same amino acid sequence as A1 (and B1; Table I).

Glycopeptides A2 and A1 (from Figure 8-1) were deglycosylated by glycopeptidase F as indicated by the loss of their corresponding peaks at retention times of 20.44 and 19.83 min, respectively (Figure 8, compare panel 1 with panel 3). In this case, due to the removal of the complete oligosaccharide

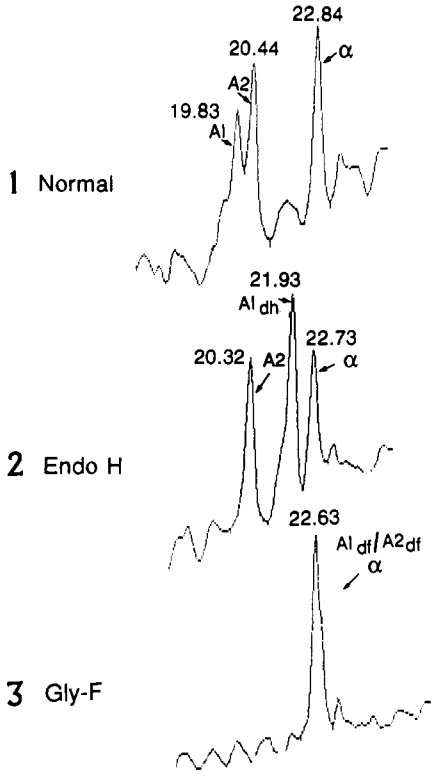

FIGURE 8: (1) Analytical sample of a mixture containing predominantly glycopeptides A2 and some residual A1, purified on an Altex 3-μm RP-HPLC column (Figure 6), was rechromatographed on a Jones 5-μm 25-cm RP-HPLC column. Peaks were detected by their UV absorption at 214 nm, and the largest peak was automatically scaled to 100%. In addition to the two glycopeptide peaks of A1 and A2 observed in the preparative separation (Figure 6), a third peak containing an unglycosylated α peptide eluted with a retention time of 22.84 min. (2) Incubation of an aliquot of the mixture shown in panel 1 with Endo H. Peak A1 was deglycosylated by the action of Endo H ($A1_{dh}$ represents glycopeptide A1 which retains a single GlcNAc-Asn residue). The position of glycopeptide A2 remained unchanged. (3) Incubation of the glycopeptides A1 and A2 with glycopeptidase F. Both A1 and A2 were fully deglycosylated by the action of glycopeptidase F ($A1_{df}$ and $A2_{df}$ represent the deglycosylated form of glycopeptides A1 and A2), and both coelute with the unglycosylated α peptide.

both peptides now coeluted at 22.63 min with the contaminating unglycosylated α peptide. Thus, glycopeptides A1 and A2, containing identical peptides, differ in their sensitivity to Endo H although their glycans are equally removed by glycopeptidase F. These data indicate that A1 and A2 contain different oligosaccharide moieties.

It remained that A2 and B2, with similar retention times by RP-HPLC, had significantly different peak heights (using the heights of A1 and B1 as a standard of reference) in the initial analyses (Figure 6). Therefore, a combined fraction of the Con A binding glycopeptides from hexosaminidase A (A1, A2, and A3) isolated from the Altex Ultrasphere ODS 3-μm column, was rechromatographed on the Jones Sperisorb ODS 5-μm reverse-phase column to examine peak heights of the original material. The resulting chromatograph demonstrates that the unglycosylated α peptide peak (revealed in Figure 8-1) was contributing to the height of the peak A2 in Figure 6 when the original Altex column was used for the preparative separation (data not shown). Thus, peak A2 was present in approximately the same relatively minor amounts as peak B2. Therefore, these data suggest that the minor glycopeptides A2 and B2 are present with identical structures and frequency on the β subunits of hexosaminidase A and B and that they correspond to the major glycopeptides A1 and

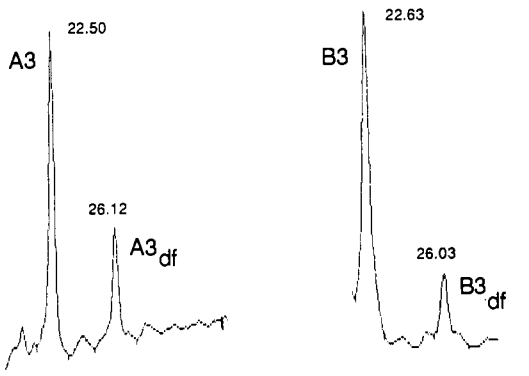

FIGURE 9: Samples of HPLC-purified glycopeptides B3 and A3 (Figure 6) were incubated with glycopeptidase F and rechromatographed on the Jones Sperisorb ODS 5-μm, 25-cm RP-HPLC column. Although they appear resistant to total hydrolysis, a small amount of deglycosylate peptide, $B3_{df}$ and $A3_{df}$, was generated. The time of elution (in minutes) of each peak is indicated. Peaks were detected by their UV absorption at 214 nm, and the largest peak was automatically scaled to 100%.

B1 with slight modifications in their glycan structure.

The structure of the N-linked glycan of A3 was determined by comparison with that of glycopeptide B3. These two glycopeptides have the same amino acid sequence (Table I), and their oligosaccharides share several characteristics: (1) While both bind to concanavalin A (Figure 6), neither are substrates for Endo H (Figure 1B, data not shown for A3). (2) Both oligosaccharides are only slowly removed by the action of glycopeptidase F (Figure 9, $B3_{df}$ and $A3_{df}$). This was confirmed by amino acid sequencing of $B3_{df}$ and $A3_{df}$. Due to the action of glycopeptidase F, the "blank" in the sequenator turn corresponding to the oligosaccharide-linked Asn residue in the untreated samples was converted to a carbohydrate-free Asp residue in the treated samples; the remainder of the sequences was identical (data not shown). (3) The elution times of the two glycopeptides on two different RP-HPLC columns are very similar (Figures 6 and 9, A3 and B3). (4) The NMR spectral characteristics of A3 (≃10 nmol, 5% yield), while not conclusive in themselves, are consistent with a Man_3-$GlcNAc_2$-Asn structure. The 500-MHz ^1H NMR spectrum obtained at 70 °C showed two presumptive mannose anomeric signals at 5.124 and 4.920 ppm (peak area 1:1) (data not shown). The chemical shifts of these two signals are in complete agreement with those observed for the observed 3^t and 6^t anomeric protons of authentic Man_3-$GlcNAc_2$-Asn at 70 °C (Grey et al., 1982). These two shifts are not consistent with any reported "complex"-type N-linked structure. Further, the absence of any additional anomeric signals downfield of 5.142 ppm precludes the presence of larger high-mannose oligosaccharides. Unfortunately, the presence of extensive spectral overlap with peptidyl ^1H signals did not allow for a more definitive identification by NMR. Nevertheless, these four observations together show that A3 and B3 are identical glycopeptides and that they likely bear a Man_3-$GlcNAc_2$ oligosaccharide.

The final glycopeptide, A4 (≃10 nmol), was also examined by a comparison of ^1H NMR and endoglycosidase analysis. It produced similar spectral and chemical evidence to that obtained for glycopeptides A3 and B3 (data not shown), indicative of a Man_3-$GlcNAc_2$ N-linked oligosaccharide. Its peptide sequence, residues 95–104 on the α chain (Table I and Figure 2), made it the only glycopeptide isolated that was associated with the α subunit of hexosaminidase A.

Analysis of a Non Con A Binding Glycopeptide from Hexosaminidase B. In order to identify any non Con A

Table V: Summary of the Location and Predicted Secondary Structure of the Putative Glycosylation Sites in β-Hexosaminidase

residue no.[a]	glycopeptide isolated	residues (Figure 2) from the mature N-terminus	secondary structure[b] α helix	β sheet	β turn	random coil
		Mature α Subunit				
(1) 114	A4	26	−56	72	61	117
(2) 156	none[c]	68	107	−100	60	40
(3) 294	none[d]	206	−152	−89	180	133
		Mature $\beta_a\beta_b$ Subunit				
β_b chain						
(1) 84	none	−38[e]	−222	−21	−70	161
(2) 142	A3, B3	20	−136	−14	82	113
(3) 190	A1, B1, A2, B2[c]	68	−90	79	51	77
β_a chain						
(4) 327	non Con A binding[d]	13	−218	60	199	167
(5) 497	none	82	−171	12	102	139

[a] Distance, in residues, the putative glycosylation site, N, is from the putative initiation site (residue 1 in Figure 2). [b] Probability that the *i*th residue (expressed as an average of the three amino acids comprising the putative glycosylation site) has the given secondary structure, from a matrix of statistically derived numbers given by Garnier et al. (1978). [c] First coincident, putative glycosylation sites in the aligned primary structures of the α and β chains. [d] Second coincident, putative glycosylation sites in the aligned primary structures of the α and β chains. [e] This peptide is not present in the mature protein (Figure 2).

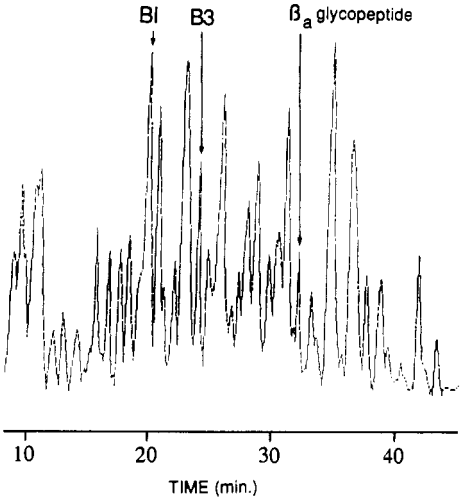

FIGURE 10: RP-HPLC separation of a total tryptic/chymotryptic digest of hexosaminidase B (Jones Sperisorb ODS 5-μm, 25-cm RP-HPLC column). Glucosamine-positive peaks are indicated. The identity of the peaks labeled B1 andd B3 was confirmed by amino acid sequencing. The peak labeled β_a glycopeptide contained a unique sequence which was localized to the β_a region of the prepro-β chain (Figure 2).

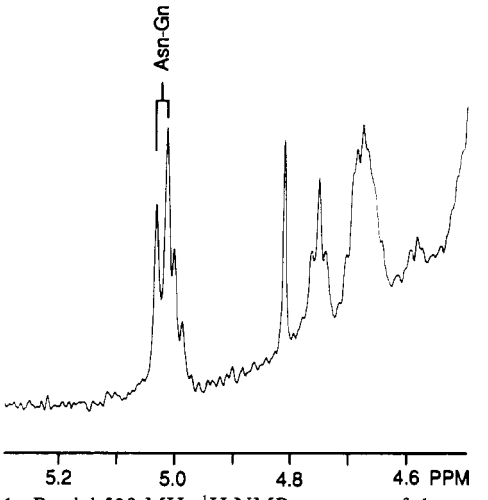

FIGURE 11: Partial 500-MHz ^{1}H NMR spectrum of the non Con A binding β_a glycopeptide from hexosaminidase B (Figure 10). The glycopeptide spectrum was recorded at 70 °C as described under Materials and Methods. The chemical shifts indicate the presence of a single GlcNAc residue.

binding glycopeptides, total tryptic/chymotryptic digests of hexosaminidase A and B were prepared and examined for GlcNAc-positive peaks after RP-HPLC. Two GlcNAc-positive glycopeptide peaks were isolated from hexosaminidase A and three from hexosaminidase B. Two pairs of these included the major Con A binding glycopeptides A1 and A3 from hexosaminidase A (data not shown) and B1 and B3 from hexosaminidase B (Figure 10, peaks labeled B1 and B3). In addition a small, GlcNAc-positive peak (≃20 nmol) which was not sensitive to Endo H treatment (data not shown) was isolated from the hexosaminidase B digest (Figure 10, β_a glycopeptide).

Amino acid sequencing revealed a single sequence which was localized to positions 316–325 on the deduced sequence of the prepro-β polypeptide chain. This corresponds to the Asn residue of a putative glycosylation site 13 residues removed from the mature amino terminus of the β_a polypeptide chain (Mahuran et al., 1988) (Figure 2). The 500-MHz ^{1}H-NMR spectrum obtained at 70 °C showed a doublet at 5.031 ppm with a 10.2-Hz coupling constant, consistent with the anomeric ^{1}H of a GlcNAc-Asn residue (Figure 11). The corresponding *N*-(acetylmethyl) signal appeared as a singlet at 1.996 ppm (data not shown). These data correlate with signals at 5.057 and 2.019 ppm, respectively, found for the single GlcNAc residue in a glycopeptide from cathepsin B (Takahashi et al., 1984). Thus, we conclude that hexosaminidase B contains a single GlcNAc residue at amino acid residue 315 (prepro-β numbering sequence) in the β_a peptide of the mature β subunit.

Discussion

Oligosaccharides were identified at one of three putative glycosylation sites in the α subunit and at three of five sites on the β subunit of hexosaminidase (Figure 2). These were resolved as four glycopeptide peaks on RP-HPLC from hexosaminidase A (peaks A1–A4) and three peaks (B1–B3) from hexosaminidase B after purification by Con A–Sepharose chromatography, a procedure that enriches high mannose type glycopeptides. One additional glycopeptide was isolated as a glucosamine-positive peak on RP-HPLC after a tryptic digestion of hexosaminidase B. The following summarizes the characterization of the eight isolated glycopeptides.

Amino acid sequencing of the purified glycopeptides revealed that A3 and B3 had identical peptide sequences and were localized to the first available Asn-X-Thr(Ser) glycosylation sites from the N-terminus of the mature β_b polypeptide (Tables I and V; Figure 2). Similarly, the amino acid se-

quences of the A1 and A2, and B1 and B2, glycopeptides were identical and were localized to the next available glycosylation site in the β_b chain. Thus, glycopeptides A1–A3 and B1–B3 are all associated with the β subunit of hexosaminidase A and B. The single hexosaminidase A specific glycopeptide, A4, was localized near the mature amino terminus of the α chain, also at its first available site (Figure 2).

The oligosaccharides contained on A1 were found to be identical with those on B1 by NMR. They consist of Man_{5-7}-$GlcNAc_2$ glycans in both isozymes. The oligosaccharide on B3, from hexosaminidase B, was shown by NMR to contain a single Man_3-$GlcNAc_2$ glycan. Although insufficient quantities of its A3 counterpart were available for conclusive structural characterization by NMR, the 500-MHz ^1H NMR spectrum obtained was consistent with a Man_3-$GlcNAc_2$-Asn structure. Also the chemical shifts were inconsistent with any reported complex-type or larger high-mannose oligosaccharides. Furthermore, the similar elution times from two different RP-HPLC columns (Figures 6 and 9) and the observed glycosidase sensitivities (Table I) indicate that glycopeptide A3 contains the same oligosaccharide moiety as B3.

The minor Con A binding peptides, A2 and B2, isolated from hexosaminidase A and B, respectively, showed longer retention times on RP-HPLC than the major A1 and B1 glycopeptides. Since the amino acid sequences of all four peptides were identical, it was left that the oligosaccharides on A2 and B2 must contain some structural differences from those identified in the major glycopeptides. Since A2 and B2 themselves had similar retention times on RP-HPLC, it is likely that they contain the same oligosaccharide.

The major non Con A binding glycopeptide isolated from hexosaminidase B was found by NMR to contain a single GlcNAc residue. We were unable to isolate a similar GlcNAc-containing glycopeptide from the hexosaminidase A digest, due to the smaller absolute sample size and the increased complexity of the elution pattern resulting from the combination of peptides from both the α and β subunits. However, the occurrence of a sequenator "blank" at the predicted Asn residue was found when the N-terminal sequence of the combined β_a and β_b polypeptide chains from hexosaminidase A was determined. Furthermore, the identical SDS–PAGE pattern of the reduced β_a chain from either isozyme was observed both before and after glycopeptidase F digestion (Mahuran et al., 1988). This suggests that the carbohydrate is also present in the mature β subunit from hexosaminidase A. Thus, the β subunits from either hexosaminidase A or hexosaminidase B contain identical carbohydrate moieties.

The single GlcNAc residue comprising the carbohydrate on the mature β_a chain has not previously been found on any lysosomal enzyme except cathepsin B, whose localization to the lysosome is uncertain (Takahashi et al., 1984). This may indicate the presence of an endoglycosidase in the lysosome.

Only one Con A binding glycopeptide, A4, was mapped to the α chain of hexosaminidase A. Partial characterization by NMR and its lack of sensitivity to Endo H digestion strongly suggest that it contains an N-linked Man_3-$GlcNAc_2$ glycan.

Thus, the eight glycopeptides characterized in this study were assigned to four of the seven eligible glycosylation sites found on the mature α and β subunits. The only remaining site on the mature β subunit was excluded on the basis of peptide sequence data. A previously isolated tryptic peptide containing this site included the underivatized, i.e., unglycosylated, Asn residue (Korneluk et al., 1986; peptide designated B-3-H2). We have, therefore, accounted for all of the putative glycosylation sites on the mature β subunit. There remains a single site in the procomponent of the prepro-β sequence (Figure 2) which could not be examined in this study that could contain an attached oligosaccharide.

Additional non Con A binding, i.e., complex type, oligosaccharides could yet be found on the two remaining sites in the α chain. However, one of these sites is an unlikely candidate for glycosylation due to secondary structure constraints. Most N-glycosylated Asn residues are located in regions of the protein that favor the formation of β-turns or loops (Kornfeld & Kornfeld, 1985). The secondary structure predictions calculated by the method of Garnier et al. (1978) strongly favor an α helix for the first of these regions, residue 56, in the α chain. In contrast, the homologous site in the β subunit is glycosylated (Man_{5-7}-$GlcNAc_2$), and its structure is predicted to be a β sheet or random coil, consistent with the hypothesis (Table V). The Garnier algorithm does suggest a β turn for the second α chain site, also opposite a glycosylation site on the β subunit (GlcNAc). We have yet to sequence the relevant peptide.

Only the second available glycosylation site in the mature β subunit in either hexosaminidase A or hexosaminidase B contained a mixture of oligosaccharide structures (Man_{5-7}-$GlcNAc_2$ in peptides A1 and B1 and uncharacterized minor structures in A2 and B2). There are precedents for such heterogeneity of oligosaccharide structures at a single attachment site. For example, similar results have been obtained for the lysosomal enzyme cathepsin D. Both of its attachment sites contain several high-mannose structural variants (Takahashi et al., 1983). One explanation for these findings is that multiple structural variants are a reflection of progressive glycolytic degradation occurring with lysosomes. This must certainly be the case for the Man_3-$GlcNAc_2$ glycan found in glycopeptides B3 (and A3), and A4, as well as the single GlcNAc residue found in the β_a glycopeptide, since no known processing step within the Golgi could produce a Man_3-$GlcNAc_2$ or GlNAc-Asn structure. These data do not represent artifacts of purification because (1) the same glycopeptides were isolated from three enzyme preparations each from 10–15 placentas and (2) identical glycan structures from these glycopeptides were found on the β subunit of both hexosaminidase A and B. Thus, we propose that this "processing" represents the rapid removal of nonessential structural domains exposed to the catabolic milieu of the lysosome. This can explain not only carbohydrate degradation but the amino-terminal processing of the pro-α and pro-β polypeptides we have described (Mahuran et al., 1988).

It has been shown that mature placental hexosaminidase A and B can be phosphorylated in vitro (Goldberg et al., 1984) and that the transfer of the phosphate group requires the presence of an α1,2-linked mannose residue (Couso et al., 1986). Thus, neither the Man_5 or Man_3 structures we have found in the β subunit nor the Man_3 structure on α could explain the in vitro phosphorylation of hexosaminidase. This leaves only the Man_6 or Man_7 structures found on glycopeptides B1 and A1 (present along with the Man_5). Since these were mapped to a single glycosylation site on the β_b chain, we suggest that this site on the β subunit is solely responsible for the in vitro phosphorylation of hexosaminidase A and B. This argument does not extend to in vivo phosphorylation. Because of the evident glycolytic processing of some of these sites in the lysosome, we are unable to predict the structures generated during passage through the Golgi. For example, the α chain has been shown to be phosphorylated

in vivo (Hasilik & Neufeld, 1980). This must occur at the site on glycopeptide A4 which, after lysosomal processing, contains the Man_3 structure that is not a substrate for the phosphotransferase. Thus, all three β subunit and the single α subunit glycosylation sites remain candidates for in vivo phosphorylation.

A comparison of the amino acid sequences near the glycosylation sites in hexosaminidase A and B with those near glycosylation sites in previously published lysosomal protein sequences (Fukushima et al., 1985; San Segundo et al., 1985; Calhoun et al., 1985; Faust et al., 1985; Nishimura et al., 1986; Oshima et al., 1987) failed to reveal any obvious consensus sequence that could be associated with phosphorylation. This suggests that the three-dimensional structures of lysosomal enzymes may play a role in the recognition of selected oligosaccharides by the phosphotransferase. Nevertheless, given the simple and highly redundant consensus sequence for N-linked glycosylation [Asn-X-Ser(Thr); where X can be any amino acid except aspartic acid or proline (Kornfeld & Kornfeld, 1985)], the existence of a similar type of highly redundant consensus sequence (dependent on the local secondary structure of the polypeptide chain for recognition) for phosphorylation cannot be ruled out.

Added in Proof

Recently Little et al. (1988) and Stirling et al. (1988) experimentally determined the sites of signal peptide cleavage in the pre-α and pre-β chains, respectively. The cleavage site in the pre-α chain is located between residues 22 (Ala) and 23 (Leu) as predicted in Figure 2. However, the cleavage site in the pre-β chain is located between residues 42 (Ala) and 43 (Ala) rather than residues 24 and 25 as we had predicted in Figure 2. This suggests that the third (Met, residue 26 in Figure 2) rather than the first in-frame ATG (Met, residue 1 in Figure 2) is the initiation site for synthesis of the prepro-β polypeptide chain.

Acknowledgments

We thank C. Hew, M. Whiteside, and S. Joshi for determining the amino acid sequence of the glycopeptides. We also thank J. Carver and A. Grey for their help in obtaining and interpreting the NMR spectra.

Registry No. A, 38784-68-8; B, 39114-02-8; C, 85484-03-3; D, 74424-57-0; β-acetylhexosaminidase, 9027-52-5.

References

Calhoun, D. H., Bishop, D. F., Bernstein, H. S., Quinn, M., Hantzopoulos, P., & Desnick, R. J. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 7364–7368.

Carver, J., Grey, H., Winnik, F., Hakimi, J., Ceccarini, C., & Atkinson, P. H. (1981) *Biochemistry 20*, 6600–6606.

Couso, R., Lang, L., Roberts, R. M., & Kornfeld, S. (1986) *J. Biol. Chem. 261*, 6326–6331.

Faust, P. L., Kornfeld, S., & Chirgwin, J. M. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 4910–4914.

Fukushima, H., deWet, J. R., & OBrien, J. S. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 1262–1265.

Garnier, J., Osguthorpe, D. J., & Robson, B. (1978) *J. Mol. Biol. 120*, 97–120.

Goldberg, D., Gabel, C., & Kornfeld, S. (1984) in *Lysosomes in Pathology and Biology* (Dingle, J. T., Dean, R. T., & Sly, W., Eds.) pp 45–62, Elsevier, North-Holland, New York.

Grey, A. A., Narasimhan, S., Brisson, J.-R., Schachter, H., & Carver, J. P. (1982) *Can. J. Biochem. 60*, 1123–1131.

Hakimi, J., Carver, J., & Atkinson, P. H. (1981) *Biochemistry 20*, 7314–7319.

Hasilik, A., & Neufeld, E. F. (1980) *J. Biol. Chem. 255*, 4946–4960.

Korneluk, R. G., Mahuran, D. J., Neote, K., Klavins, M. H., O'Dowd, B. F., Tropak, M., Willard, H. F., Anderson, M.-J., Lowden, J. A., & Gravel, R. A. (1986) *J. Biol. Chem. 261*, 8407–8413.

Kornfeld, R., & Kornfeld, S. (1985) *Annu. Rev. Biochem. 54*, 631–664.

Lang, L., Reitman, M., Tang, J., Roberts, R. M., & Kornfeld, S. (1984) *J. Biol. Chem. 259*, 14663–14671.

Little et al. (1988) *J. Biol. Chem. 263*, 4288–4292.

Mahuran, D. J., & Lowden, J. A. (1980) *Can. J. Biochem. 58*, 287–294.

Mahuran, D., Novak, A., & Lowden, J. A. (1985) *Isozymes: Curr. Top. Biol. Med. Res. 12*, 229–288.

Mahuran, D. J., Neote, K., Klavins, M. H., Leung, A., & Gravel, R. A. (1988) *J. Biol. Chem. 263*, 4612–4618.

Natowitz, M., Hallett, D. W., Frier, C., Chi, M., Shlesinger, P. H., & Baenziger, J. U. (1983) *J. Cell Biol. 96*, 915–919.

Nishimura, Y., Rosenfeld, M. G., Kreibich, G., Gubler, U., Sabatini, D. D., Adesnik, M., & Andy, R. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 7292–7296.

O'Dowd, B. F., Mahuran, D., Cumming, D., & Lowden, J. A. (1985) *Can. J. Biochem. Cell Biol. 63*, 723–729.

Oshima, A., Kyle, J. W., Miller, R. D., Hoffman, J. F., Powell, P. P., Grubb, J. H., Sly, W. S., Tropak, M., Guise, K. S., & Gravel, R. A. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 685–689.

Peckett, M., Scott, G., Davies, P., Wang, N., Jacki, S., & Hew, C. (1984) *Eur. J. Biochem. 143*, 55–58.

Reitman, M. L., Varki, A., & Kornfeld, S. (1981) *J. Clin. Invest. 67*, 1574–1579.

San Segundo, B., Chan, S. J., & Steiner, D. F. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 2320–2324.

Smyth, D. G. (1970) *Methods Enzymol. 11*, 214–227.

Stirling et al. (1988) *FEBS Lett. 231*, 47–50.

Tai, T., Yamashita, K., & Kobata, A. (1977) *Biochem. Biophys. Res. Commun. 78*, 434–441.

Takahashi, T., Schmidt, P. G., & Tang, J. (1983) *J. Biol. Chem. 258*, 2819–2830.

Takahashi, T., Schmidt, P. G., & Tang, J. (1984) *J. Biol. Chem. 259*, 6059–6062.

Von Figura, K., & Hasilik, A. (1986) *Annu. Rev. Biochem. 55*, 167–193.