Structural elucidation of O-linked glycopeptides by high energy collision-induced dissociation

ARTICLE *in* JOURNAL OF THE AMERICAN SOCIETY FOR MASS SPECTROMETRY · APRIL 1996

Impact Factor: 2.95 · DOI: 10.1016/1044-0305(95)00682-6 · Source: PubMed

5 AUTHORS, INCLUDING:

Alma L Burlingame
University of California, San Francisco
627 PUBLICATIONS **26,404** CITATIONS

SEE PROFILE

Structural Elucidation of *O*-Linked Glycopeptides by High Energy Collision-Induced Dissociation

K. F. Medzihradszky, B. L. Gillece-Castro,* R. R. Townsend, and A. L. Burlingame

Department of Pharmaceutical Chemistry and Mass Spectrometry Facility, University of California San Francisco, California, USA

M. R. Hardy

Immunogen, Norwood, Massachusetts, USA

O-linked glycopeptides that bear a GalNAc core with and without the presence of sialic acid have been analyzed by high energy collision-induced dissociation (CID). We show that the CID spectra from the glycosylated precursor ions contain sufficient information to identify the peptide sequence and to determine the glycosylated site(s). Asialo *O*-linked glycopeptides, previously prepared from a tryptic digest of bovine fetuin were studied. One of the glycopeptides contained only a single Hex (hexose)-HexNAc (*N*-acetylhexosamine) substitution at Thr^{262}, whereas the other exhibited Hex-HexNAc moieties at both Thr^{262} and Ser^{264}. In addition, sialo and asialo fetuin glycopeptides from a pronase digest were derivatized with *t*-butoxycarbonyl-tyrosine, and characterized by high energy CID analysis. The presence of a Galβ(1,3)GalNAc core structure at Ser^{264} was confirmed by using the substrate specificity of endo-α-*N*-acetylgalactosaminidase. These studies revealed the presence of a β-galactosidase specific for β(1,4) linkages in the endo-α-*N*-acetylgalactosaminidase preparation employed. Finally, the relative stability of *N*- and *O*-glycosyl bonds to high energy CID is addressed based upon comparison of the behavior of a synthetic *N*-linked glycopeptide with analogous *O*-linked structures. *(J Am Soc Mass Spectrom 1996, 7, 319–328)*

Glycosylation at specific Asn, Ser, or Thr residues is one of the most commonly encountered co- and posttranslational modifications of proteins. Two fundamental questions must be addressed in the structural elucidation of protein glycosylation, namely, the nature of the structures of the oligosaccharides present and their site(s) of attachment to the polypeptide chain. Over the last few years mass spectrometry has played an increasingly important role in facilitation of the structural characterization of protein glycosylation [1]. Progress has been more rapid in the case of studies of the *N*-linked class for a variety of reasons, such as in hand knowledge of tripeptide consensus sequences, the relative ease of generation of *N*-linked glycopeptides by proteolysis, and the availability of enzymes such as endo-H or PNGase F to carry out enzymatic deglycosylation without destruction of either peptide or oligosaccharide. Microbore high-performance liquid chromatography (HPLC)–electrospray mass spectrometry has become the method of choice to detect and map glycopeptides [2, 3]. Although the detection of *O*-linked glycosylation can be achieved readily by collisional activation and glycosyl-moiety-specific oxonium ion monitoring of electrospray ionized species, the detailed structural characterization of *O*-linked glycopeptides still represents a considerable challenge. Some of the problems include the fact that there is no known consensus sequence for *O*-linked glycosylation (discussed in Medzihradszky, K. F.; Reinders, L.; Townsend, R. R.; Burlingame, A. L., in preparation), that several contiguous sites may be occupied without intervening proteolytic cleavage sites, and that the release of the carbohydrate units cannot be accomplished while preserving the structural integrity of the peptide. The only endoglycosidase available for *O*-linked structures releases just a single *O*-linked disaccharide structure, namely, Galβ(1,3)GalNAc [4]. The molecular weight of *O*-

Address reprint requests to Professor A. L. Burlingame, Department of Pharmaceutical Chemistry, University of California, San Francisco, San Francisco, CA 94143-0446.

* Present address: Genentech, Inc., 460 Point San Bruno Boulevard, South San Francisco, CA 94080.

1044–0305/96/$15.00
SSDI 1044-0305(95)00682-6

Received July 28, 1995
Revised October 23, 1995
Accepted November 1, 1995

linked glycopeptides may be determined by a variety of mass spectrometric techniques [2, 5–7], and knowledge of the peptide sequence enables the composition of the carbohydrate to be inferred. In addition, location of the occupied site may be established by observation of a blank cycle during Edman degradation. In a modified Edman method, the PTH-glycoamino acids even have been characterized by chromatographic retention times [8]. The oligosaccharide(s) itself may be isolated for structural study by base-catalyzed elimination. The former methods reveal nothing about the structural nature of the glycosylation, whereas the latter method destroys the peptide and site information.

To circumvent these difficulties, this laboratory has turned to high-performance tandem mass spectrometry for determination of the sites of *O*-glycosylation for glycopeptides isolated from the recombinant growth hormone PDGF and the protein bovine fetuin [9, 10].

In this report we extend our previous studies [9, 10] to high energy collision-induced dissociation (CID) analysis of larger glycopeptides that bear sialic acid (Neu5Ac; SA in figures). In particular, we provide evidence to establish that the CID spectra from the glycosylated precursor ions contain sufficient information both to identify the peptide sequence and to determine the glycosylated site(s). Taken together with the specificity of endo-α-*N*-acetylgalactosaminidase, the presence of the Galβ(1,3)GalNAc core at a specific serine residue in bovine fetuin has been confirmed. In addition, these studies revealed the presence of a β-galactosidase specific for β(1,4) linkages in the endo-α-*N*-acetylgalactosaminidase preparation employed. Finally, the relative stability of *N*- and *O*-glycosyl bonds to high energy CID was investigated by comparison of the fragmentation behavior of a synthetic *N*-linked glycopeptide with analogous *O*-linked structures.

Experimental

Preparation of Asialo Tryptic Glycopeptides (Method 1)

Glycopeptides were prepared from bovine fetuin (Gibco, Grand Island, NY) as previously described [11]. Briefly, reduced and alkylated fetuin was digested with *N*-tosylphenylalanine chloromethyl ketone (TPCK)-treated trypsin for 24 h at 37 °C in 0.2-M Tris buffer (pH 8.2). The digest was fractionated on a Bio-Gel P-10 column (BioRad, Hercules, CA), the peaks (monitored at 215 nm) were pooled and lyophilized, and the carbohydrate-containing fractions were treated with *Arthrobacter ureafaciens* neuraminidase. After removal of free sialic acid by size exclusion on a G25 column, the glycopeptides were separated by using (diethylamino)ethyl (DEAE)–Sephacel chromatography. Six peak fractions were pooled separately and further purified by using reversed-phase high-performance liquid chromatography (RP-HPLC) on an octadecylsilica column, as previously detailed [11].

Preparation of Sialylated t-Butoxycarbonyl-Tyrosine-Labeled O-Linked Glycopeptides (Method 2)

One gram of Gibco (Grand Island, NY; Spiro method, lot 21P8255) fetuin, was dissolved in 20 mL of 0.1-M hepes buffer, which contained 10-mM $CaCl_2$ (pH 7.9). The solution was incubated for 72 h at 37 °C with daily additions of pronase (20 mg) (Calbiochem, San Diego, CA). Insoluble material was removed from the digest by centrifugation (10 min at 10,000 rpm), and the entire supernatant was applied to a Sephadex G50 column (2.5 × 200 cm) equilibrated in pyridine-acetate buffer. Glycopeptides [~ 35 μmol as galactosamine ($GalNH_2$)] from G50 peak pools B and C were each dissolved in 1.5 mL of dimethyl sulfoxide (DMSO; Aldrich, Milwaukee, WI; HPLC grade). To this solution was added 100 μmol of triethylamine (Pierce, Sequanal grade) and 100 μmol of *t*-butoxycarbonyl (tBOC)-Tyr-2,4,5-trichlorophenyl ester (Sigma Chemical Co., St. Louis, MO). The reaction mixture was kept at room temperature for 5 h and then the DMSO was removed by lyophilization. The tBOC-Tyr-modified glycopeptide was extracted into 1.5 mL of water with sonication, and the water-insoluble material was pelleted by centrifugation. The pellet was then extracted with 2 × 1 mL of water and centrifuged again. The supernatants were combined and applied to a column of BioGel P6 [1.5 × 90 cm] in 0.1-M ammonium acetate. Two-milliliter fractions were collected. Aliquots of $A_{280\ nm}$ positive fractions were hydrolyzed for 4 h at 100 °C with 6-N HCl [12] and assayed for amino sugar content. The three $GalNH_2$-containing peak fractions were pooled and lyophilized, and stock solutions were prepared for HPLC. Preparative RP-HPLC was carried out on an octadecylsilica column (4.6 × 250 mm, Supelcosil LC18-DB, Supelco, Bellefonte, PA) by using the apparatus previously described [13]. All solvents were HPLC grade. Glycopeptides were isolated by repeated injections (1–2 mg in 100 μL per injection) of the stock solutions. The peak fractions from the multiple runs were pooled and lyophilized to dryness.

Preparation of Asialo t-Butoxycarbonyl-Tyrosine-Labeled O-Linked Glycopeptides (Method 3)

Bovine fetuin (Sigma, type III) was digested with pronase as previously described [13]. Glycopeptides were prepared by chromatography on a Sephadex G50 column, as described for Method 2. The glycopeptides (from ~ 1 g of protein) were then desialylated with neuraminidase from *Arthrobacter ureafaciens* (1 IU) in acetate buffer (0.5 M, pH 5.0) for 24 h at 37 °C. The digest was purified on a Sephadex G50 column (2.5 × 200 cm) that was equilibrated in pyridine acetate (pH 4.7). The carbohydrate-containing fractions were pooled, lyophilized, and redigested with neu-

raminidase in acetate buffer (0.5 M, pH 5.0) for 24 h at 37 °C. The pH of this digest was increased to 7.6 with hepes and made 5 mM in $CaCl_2$ before digestion with pronase (10 mg) for 24 h at 37 °C, and then was purified by Sephadex G50 chromatography as previously described. The lyophilized glycopeptides were dissolved in 40 mL of 10-mM Tris buffer (pH 7.6) and applied to a DEAE-Sephacel column (2.5 × 23 cm). The glycopeptides were eluted with 740 mL of 10-mM Tris buffer (pH 7.6) followed by a linear gradient of NaCl to a limit concentration of 50 mM (total elution volume of 1 L). The column eluent was monitored at 215 nm and the absorbing fractions were assayed for carbohydrate by using a modified phenol-sulfuric acid method [14]. The nine carbohydrate-containing peaks also were analyzed for neutral sugars by using a borate ion exchange method [15]. The ratio of Man to Gal in all but one fraction was found to be near unity (0.81–0.92), whereas the ratio in the second eluting carbohydrate peak fraction was 0.28, which suggests that it was enriched for the *O*-linked glycopeptides. After lyophilization, this fraction was dissolved in DMSO and derivatized with tBOC-Tyr-2,4,5-trichlorophenyl ester as previously described [13]. The glycopeptides were then purified on a BioGel P-6 column (1.5 × 70 cm) that was equilibrated in 0.1-M acetic acid. Three peaks were found that absorbed at 280 nm and contained amino sugars. Two fractions contained $GalNH_2$, whereas the first eluting peak contained only glucosamine ($GlcNH_2$).

O-Glycanase Digestion

Approximately 10 nmol of a glycopeptide (P-2) was incubated with 5 mU of endo-α-*N*-acetylgalactosaminidase (Boehringer-Mannheim, Indianapolis, IN) in 50-mM sodium-phosphate buffer (pH 7.3) at 37 °C for 24 h. The digestion products were purified by RP-HPLC on a C-18 column (4.6 × 250 mm; Vydac, Separations Group, Hesperia, CA). Solvent A was 0.1% trifluoroacetic acid (TFA) in water. Solvent B was 0.08% TFA in acetonitrile. The column was equilibrated in solvent A, and a linear gradient (1% per minute) was started 5 min after the injection.

Mass Spectrometry

High energy CID spectra were recorded on a Kratos Concept IIHH tandem mass spectrometer (Kratos Analytical, Ramsey, NJ) of EBEB geometry, equipped with a liquid secondary ion mass spectrometry (LSIMS) source, a coolable probe, and a photodiode multichannel array detector capable of monitoring 4% of the mass range simultaneously [16]. The collision gas was He and its pressure was set to reduce the precursor ion abundance by 70%. The collision cell was floated at 4 kV. Glycopeptides (300 pmol to 1 nmol) were loaded into 1 μL of liquid matrix (thioglycerol:glycerol 1:1 with 1% TFA) for the analyses.

Results and Discussion

Figures 1 and 2 (see [17] for nomenclature) show the CID spectra of two asialo glycopeptides (MH^+ at m/z 1615.6 and 1980.8, respectively) from a tryptic digest of bovine fetuin prepared by Method 1. In both spectra the most abundant ions are fragments formed by sequential losses of carbohydrate residues from the molecular ion. Figure 1 shows ions at m/z 1453 and 1250 that indicate losses of hexose (Hex) and Hex-HexNAc (*N*-acetylhexosamine; HNAc in Figures 7–9) moieties from the smaller glycopeptide. The larger glycopeptide exhibits similar losses due to the sequential losses of Hex, HexNAc, Hex, and HexNAc residues, respectively (at m/z 1818, 1615, 1453, and 1250 in Figure 2). However, an oxonium ion that represents nonreducing terminal sequences is at m/z 366 in both spectra (Figures 1 and 2), which indicates the attachment of disaccharyl units (Hex-HexNAc) in both glycopeptides. The smaller molecular weight species bears one such carbohydrate structure, whereas there are two of these disaccharide moieties attached to the larger glycopeptide (Figure 2), as previously indicated by the second HexHexNAc loss. In addition, these data reveal that the peptides themselves must both correspond to a protonated molecular weight of 1250 u [10]. Seven different sequences from carboxamidomethylated (CAM) bovine fetuin match this molecular weight: peptides [55–65] PTGEVYDIEID (1250.6), [140–151] SRVVHAVEVALA (1250.7), [146–157] VAVALATFNAES (1250.6), [211–221] FC*KGSVIQKAL (1250.7), [258–272] AAGPTPSAAGPPVAS (1250.6), [314–326] GKTPIVGQPSIPG (1250.7), and [315–327]

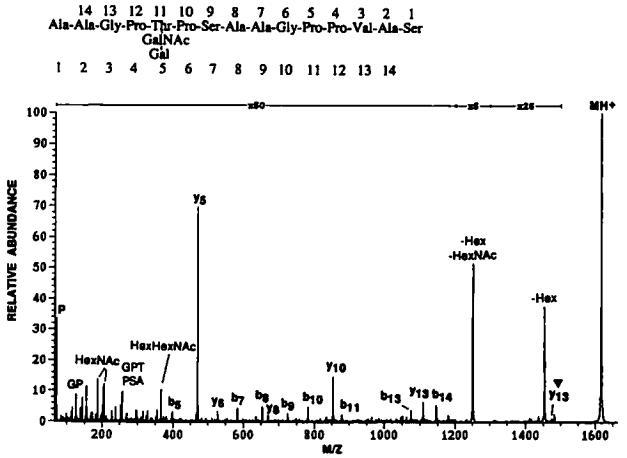

Figure 1. High energy CID spectrum of an asialo fetuin glycopeptide isolated from a tryptic digest. $MH^+ = 1615.6$. The Y-type carbohydrate fragments (nomenclature, see [17]) are labeled with the corresponding sugar unit losses. The B-type (and related via water loss) carbohydrate ions are labeled with the sugar units that comprise the oxonium ion. This CID spectrum does not reveal the identity of the hexose or *N*-acetyl hexosamine units. The peptide fragments are labeled according to the nomenclature [19], even if they were gas-phase deglycosylated (b_5–b_{14}). Only y_{13}, indicated by a triangle, shows the presence of the oligosaccharide.

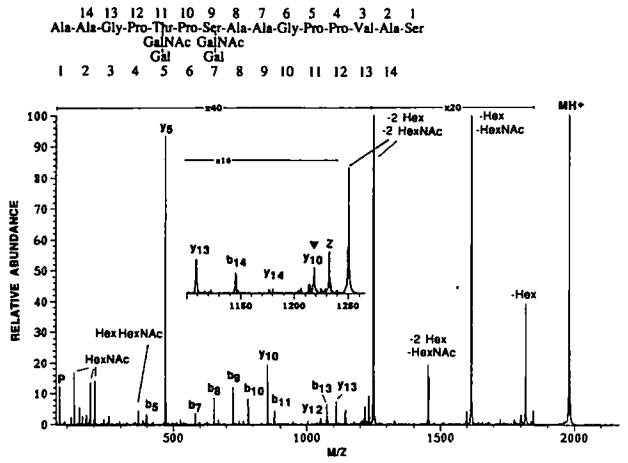

Figure 2. High energy CID spectrum of an asialo fetuin glycopeptide isolated from a tryptic digest. $MH^+ = 1980.8$. Fragment ions are labeled as in Figure 1. Fragment y_{10} (indicated in the inset) shows the presence of the oligosaccharide on the Ser(7) residue.

KTPIVGQPSIPGG (1250.7) [18]. None of these sequences would be expected to arise from specific tryptic cleavages. However, both CID spectra contain sufficient information in other ion series to establish the correct amino acid sequences of these glycopeptides. Comparison of the fragment ions observed with those predicted for all the preceding possible isobaric sequences (MacBioSpec, Perkin-Elmer-Sciex) revealed that both of these glycopeptides have the same amino acid sequence, Ala^{258} to Ser^{272}. Most of the peptide fragment ions, even from the oligosaccharide-containing part of the peptide, were not observed with carbohydrate attached (for example, $\mathbf{b}_{7-13}$ ions) (nomenclature according to [19]). Gas-phase deglycosylation is apparently a favored process. However, ions that bore carbohydrate structures were detected in low abundance. In the CID spectrum of the smaller glycopeptide (Figure 1), the peptide sequence ions observed excluded modification on either of the serine residues. The first ion that displays the presence of carbohydrates was $\mathbf{y}_{13}$ (formed via peptide bond cleavage between the Ala^{259} and Gly^{260} residues, with charge retention on the C-terminal fragment) at m/z 1416, with a Hex-HexNAc unit, which suggests that the sole Thr is glycosylated in this species. In the case of the diglycosylated species (Figure 2), a shifted peptide sequence ion at m/z 1218 is present that corresponds to the $\mathbf{y}_{10}$ + 365 (HexHexNAc) fragment (cleavage between Thr and Pro), as well as its deglycosylated analog (loss of 365 u) $\mathbf{y}_{10}$ ion at m/z 853. The presence of these ions indicates that either the Ser^{264} or the Ser^{272} residue bears a Hex-HexNAc structure. Because sequence ions that contain only Ser^{272} displayed no evidence of modification (for example, $\mathbf{y}_5$), glycosylation of Ser^{264} was established. These two glycopeptides were subjected to Edman degradation (Table 1). Ser^{272} was found unmodified in both cases, whereas blank cycles were observed at position 5 for the smaller glycopeptide and at positions 5 and 7 for the other species, which thus confirmed the foregoing attachment site assignments.

Table 1. Edman degradation of fetuin tryptic *O*-linked glycopeptides[a]

Cycle	RT 1-2 $MH^+ = 1615.6$	RT 1-1 $MH^+ = 1980.8$
1	Ala (110 pmol)[b]	Ala (150 pmol)[b]
2	Ala	Ala
3	Gly	Gly
4	Pro	Pro
5	X[2]	X
6	Pro	Pro
7	Ser	X
8	Ala	Ala
9	Ala	Ala
10	Gly	Gly
11	Pro	Pro
12	Pro?	Pro
13	Val	Val
14	Ala	Ala
15	X	Ser
16	X	X

[a] X indicates a blank cycle.
[b] Initial yield at alanine is shown in parentheses.

Multiple *O*-glycosylation sites typically are located in peptide sequences for which specific enzymes are not available. Thus, to obtain *O*-linked glycopeptides that represent individual glycosylation sites from bovine fetuin, exhaustive digestion with pronase (Methods 2 and 3) was employed. The sialylated *O*-linked pronase glycopeptides were partially separated from the *N*-linked glycopeptides by using Sephadex G50 chromatography. Three hexose-containing pools were prepared—A, B, and C as indicated in Figure 3—and analyzed for amino sugars. Only pools B and C contained $GalNH_2$, and thus contained the *O*-linked glycopeptides. Because tyrosylation of glycopeptides that contain only a few amino acids markedly improved their separation by RP-HPLC [13], pools B and C were derivatized with tBOC-Tyr-2,4,5-trichlorophenyl ester prior to further purification on an octadecylsilica column. BioGel P-6 chromatography was used after derivatization to remove residual *N*-linked glycopeptides (data not shown). Two amino sugar-containing peaks were obtained after BioGel P-6 chromatography of the G50B pool: one with predominantly $GlcNH_2$ that contains the *N*-linked glycopeptides and the other with $GlcNH_2$ and $GalNH_2$. The latter was further fractionated (designated G50B-2). Figure 4 shows the RP-HPLC chromatogram of G50B-2. The G50C pool gave two hexosamine-containing peaks upon BioGel P-6 chromatography: one with a trace of $GalNH_2$ and $GlcNH_2$ and the other, G50C-2, that contains > 90% of the $GalNH_2$. The G50C-2 pool also was fractionated by RP-HPLC.

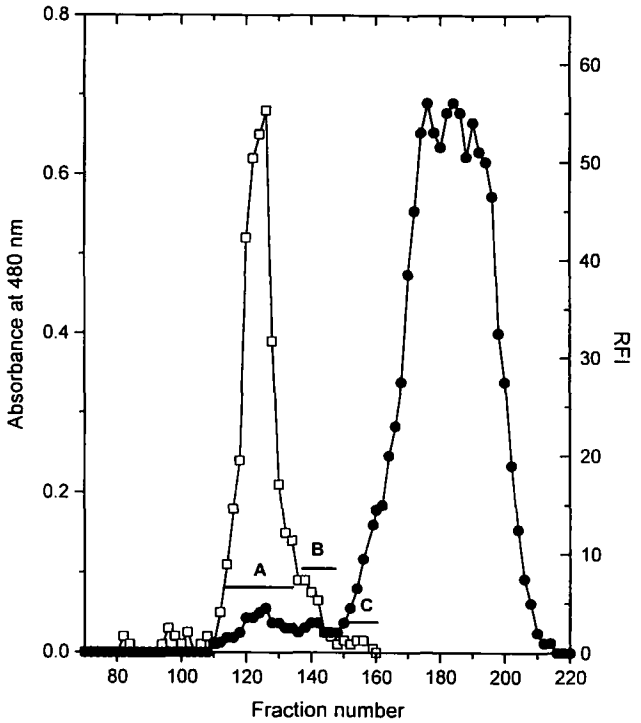

Figure 3. Sephadex G50 chromatography of fetuin pronase digest. Fetuin was digested with pronase as described for Method 2 in Experimental. The soluble portion of the digest was applied to a Sephadex G50 column (2.5 × 200 cm) that was equilibrated in pyridine-acetate buffer (pH 4.7), and 5.3-mL fractions were collected. Neutral hexose content (squares) was assayed by a phenol-sulfuric acid method [14]. Primary amino groups (circles) that indicate the presence of peptides were determined by the relative fluorescence intensity (RFI) of samples analyzed by a fluorescamine assay [20]. Three glycopeptide pools designated as G50A (fractions 113-133), G50B (136-145), and G50C (148-160), were prepared as indicated by the bars.

Asialo glycopeptides were prepared after exhaustive pronase digestion and enzymatic desialylation (Method 3). DEAE-Sephacel chromatography was found to produce a single peak fraction that was enriched in *O*-linked glycopeptides. After derivatization with tBOC-Tyr-2,4,5-trichlorophenyl ester, these glycopeptides were further purified by using BioGel P6 chromatography as shown in Figure 5. Pool P1 contained only $GlcNH_2$ and was judged to contain only the *N*-linked glycopeptides. Peak fraction P2 contained $GlcNH_2$ and $GalNH_2$ in a ratio of 0.8:1, respectively, and only $GalNH_2$ was found in pool P3. RP-HPLC of P3 gave a single peak, whereas P2 showed two other minor peaks (~ 30% of the total peak area).

The fetuin pronase glycopeptides were subjected to LSIMS analysis. Pool G50B-2 yielded three components with molecular weights 2175.2, 1865.9, and 2426.5 u. Pool G50C-2 contained four glycopeptides with molecular weights 1392.7, 1376.7, 1263.6, and 1277.6 u. Two glycopeptides were detected in fraction P-2 with molecular weights 1886.9 and 1958.0 u, whereas fraction P-3 gave a molecular ion at m/z 1522.8. The molecular ions observed and the corresponding structures are summarized in Table 2. The glycopeptide structures were determined by high energy CID as follows.

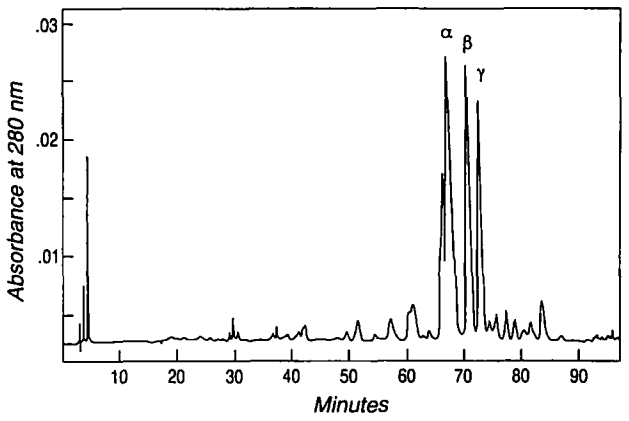

Figure 4. Preparative RP-HPLC of tBOC-Tyr-modified *O*-linked glycopeptides (G50B-2). Pronase glycopeptides were prepared and fractionated according to Method 2 as described in Experimental. The G50B-2 pool was applied to a RP-HPLC column (Supelcosil LC18DB, 4.6 × 250 mm) equilibrated in 50-mM ammonium acetate. The glycopeptides were fractionated by using the following gradients of acetonitrile: 0-15 min, 0%; 45 min, 3%; 75 min, 14%; 95 min, 14%; 120 min, 20%; 130 min, 20%.

The CID spectrum of a sialylated pronase glycopeptide (G50B-2β) is shown in Figure 6. Sequential losses from the molecular ion (MH^+ at m/z 1866.9) gave fragments at m/z 1575 (-Neu5Ac), 1413 (-Hex), 1122 (-Neu5Ac), and 919 (-HexNAc), indicative of a tetrasaccharide structure. The ion pairs that are 100 u apart (e.g., at masses 1575, 1475; 1413, 1313, etc.) are present due to the facile cleavage of the N-terminal tBOC group. Oxonium type fragments that arise from nonreducing termini are observed at m/z 292 from Neu5Ac, at m/z 495 from Neu5Ac-HexNAc, at m/z

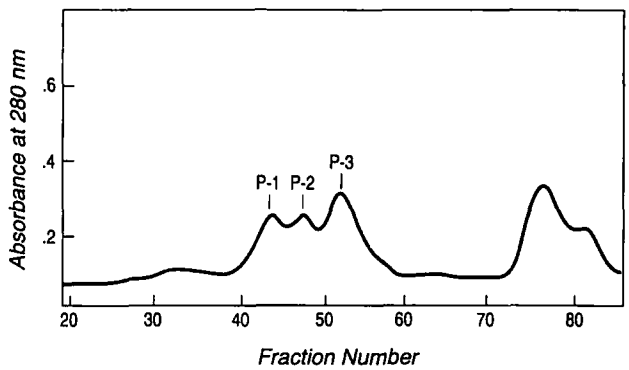

Figure 5. BioGel P6 chromatography of desialylated fetuin pronase glycopeptides. Pronase glycopeptides were prepared by Method 3 and were fractionated by anion-exchange chromatography as described in Experimental. The single peak from the DEAE column, which was enriched in $GalNH_2$, was derivatized with tBOC-Tyr-2,4,5-trichlorophenyl ester and fractionated on a column of BioGel P6 (1.5 × 70 cm) in 0.1-M acetic acid. The column effluent was monitored for UV absorbance at 280 nm. The three designated peaks, P-1, P-2, and P-3, were pooled and analyzed for amino sugars.

Table 2. Fetuin tBOC-Tyr-derivatized *O*-linked glycopeptides[a]

Glycopeptide	MH^+	Structure[b]
G50B-2 α	2176.2	^{260}Gly-Pro-(Neu5AcGalGalNAc)Thr-Pro-(Neu5AcGalGalNAc)Ser-Ala-Ala^{266}
G50B-2 β	1866.9	^{251}Ala-Pro-[Neu5AcGal(Neu5Ac)GalNAc]Ser-Ala-Val-Pro-Asp^{257}
G50B-2 γ	2427.5	^{260}Gly-Pro-(Neu5AcGalGalNAc)Thr-Pro-(Neu5AcGalGalNAc)Ser-Ala-Ala-Gly-Pro-Pro^{269}
G50C-2 α	1393.7	^{250}Glu-Ala-Pro-(Neu5AcGalGalNAc)Ser-Ala^{254}
G50C-2 β	1377.7	^{260}Gly-Pro-(Neu5AcGalGalNAc)Thr-Pro-Ser^{264}
G50C-2 γ	1264.6	^{251}Ala-Pro-(Neu5AcGalGalNAc)Ser-Ala^{254}
G50C-2 δ	1278.6	^{276}Gly-Pro-(Neu5AcGalGalNAc)Ser-Val^{279}
P-2	1887.9	^{260}Gly-Pro-[GalGlcNAc(Gal)GalNAc]Thr-Pro-(GalGalNAc)Ser-Ala^{265}
	1959.0	^{260}Gly-Pro-[GalGlcNAc(Gal)GalNAc]Thr-Pro-(GalGalNAc)Ser-Ala-Ala^{266}
P-3	1522.8	^{260}Gly-Pro-(GalGalNAc)Thr-Pro-(GalGalNAc)Ser-Ala^{265}

[a] Prepared by Methods 2 and 3 as described in Experimental.
[b] Without the N-terminal tBOC-Tyr.

657 from Neu5Ac-(Hex-HexNAc), and for the intact tetrasaccharide group at m/z 948. These ions, taken together with the presence of ions that correspond to losses of analogous carbohydrate moieties observed from the molecular ion previously described, establish the structure as Neu5Ac-Hex-(Neu5Ac)-HexNAc. The oxonium ion (m/z 366), which would be formed by the cleavage of both sialic acids, is virtually absent. Although this CID spectrum is dominated by ions formed via carbohydrate fragmentation, the identity of the peptide also was determined. By subtraction of the Tyr residue weight from the deglycosylated, deprotected peptide fragment (at m/z 819), an MH^+ value for the native peptide was calculated to be 656 u. A search of the CAM fetuin sequence gave four matches, two that contain Ser^{253} (^{251}APSAVPD^{257} or ^{252}PSAVPDA^{258} with 656.3 u as MH^+) and two that contain Ser^{278} (^{274}VVVGPSV^{279} or ^{275}VVGPSVV^{280} with 656.4 u as MH^+). Based on the peptide fragments observed, the sequence was determined to be ^{251}APSAVPD^{257}. An abundant glycopeptide fragment ($\mathbf{y}_6$) due to a cleavage at the C-terminus of Pro^{252} and bearing the complete Neu5Ac-Hex-(Neu5Ac)-HexNAc structure was detected at m/z 1532. The corresponding gas phase partially and completely deglycosylated fragments at m/z 1241 ($\mathbf{y}_6$-Neu5Ac), 1079 ($\mathbf{y}_6$-Neu5Ac-Hex), and 585 ($\mathbf{y}_6$ without the carbohydrate) also were observed. Again, fragments formed via elimination of both sialic acids were not detected. Similarly, glycopeptide fragment ion $\mathbf{b}_6$ (not labeled) with the tetrasaccharide was detected at m/z 1536, as well as sequential losses of Neu5Ac (at m/z 1245), Neu5Ac-Hex (at m/z 1083), and the tetrasaccharide from this ion (at m/z 589).

The CID spectrum of a sialylated pronase glycopeptide (MH^+ = 1377.7) with two possible glycosylation sites is presented in Figure 7. As seen in the preceding spectra, all the abundant ions in this spectrum are due to glycosidic bond cleavages or to cleavage of the

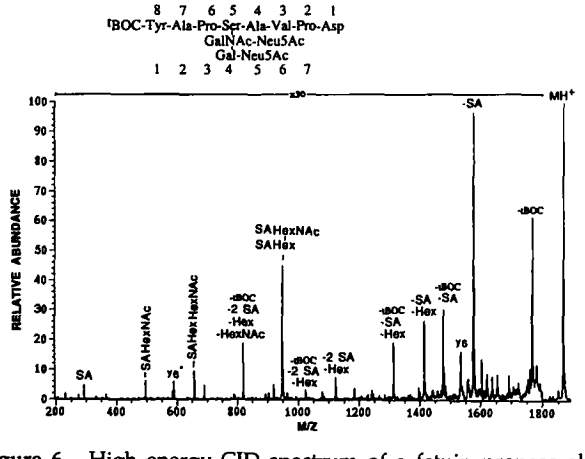

Figure 6. High energy CID spectrum of a fetuin pronase glycopeptide derivatized with tBOC-Tyr (G50B-2β). MH^+ = 1866.9. The carbohydrate fragments are labeled as in Figure 1. Fragment $\mathbf{y}_6$ is formed via the peptide bond cleavage between the Ala and Pro residue, with charge retention on the C-terminus, and it bears the carbohydrate structure. Fragment $\mathbf{y}_6^*$ is the corresponding gas-phase deglycosylated ion.

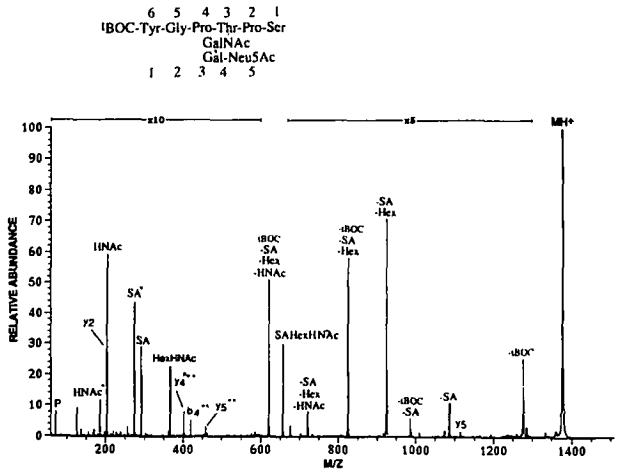

Figure 7. High energy CID spectrum of a fetuin pronase glycopeptide derivatized with tBOC-Tyr (G50C-2β). MH^+ = 1377.7. The carbohydrate fragments are labeled as in Figure 1. The peptide fragments labeled with asterisks represent the gas-phase deglycosylated and deprotected ($\mathbf{b}_4^{**}$) fragments, whereas $\mathbf{y}_5$ bears the oligosaccharide.

N-terminal protecting group. Ions at m/z 204 and 292 indicate the presence of HexNAc and Neu5Ac, respectively. More prominent ions due to water elimination from the oxonium ions are observed at m/z 186 and 274. Peaks at masses 366 and 657 are the oxonium ions for a Hex-HexNAc and a Neu5Ac-Hex-HexNAc structure. The sequence of the carbohydrate units can be confirmed by the sequential 291-u (at m/z 1086), 162-u (at m/z 924), and 203-u (at m/z 721) losses from the molecular ion, which correspond to a Neu5Ac, a Hex, and an HexNAc, respectively. Ion pairs 100 u apart (621-721, 824-924, 986-1086, 1277-1377) are present due to the cleavage of the tBOC group from the N-terminus of the peptide. The molecular weight of the underivatized peptide can be determined readily as shown earlier, and the calculation gave 458 u. Comparison of this mass with the CAM fetuin amino acid sequence (MacBioSpec) shows there are five isobaric candidates: ^{260}GPTPS^{264}, ^{274}VVGPS^{278}, ^{275}VGPSV^{279}, ^{276}GPSVV^{280}, and ^{315}KTPI^{318}. Comparison of the computer-predicted sequence ions with those observed indicates both ^{260}GPTPS^{264} and ^{276}GPSVV^{280} are candidates, because $\mathbf{y}_4$ at m/z 401 and $\mathbf{y}_5$ at m/z 458 are possible for both sequences. Ions $\mathbf{y}_2$ at m/z 203 and $\mathbf{b}_4$ (without the tBOC group) at m/z 419 establish the ^{260}GPTPS^{264} sequence. No glycosylated $\mathbf{y}_2$ fragments were detected. Thus, the site of the carbohydrate attachment is at Thr^{262}.

A hexasaccharide Neu5Acα(2,3)Galβ(1,4)GlcNAc-β(1,6)[Neu5Acα(2,3)Galβ(1,3)]-GalNAc has been described in bovine fetuin, but its peptide location has not been determined [20]. The desialylated tBOC-Tyr-glycopeptide, P-2 (MH^+ at m/z 1887.9; Method 3), was analyzed by high energy CID and was found to contain only the asialo hexasaccharide at Thr^{262} and only the disaccharide at Ser^{264} [21]. The peptide sequence was determined as tBOC-Tyr-^{260}GPTPSA^{265}. To verify this conclusion, the glycopeptide was digested with endo-α-*N*-acetylgalactosaminidase. This enzyme cleaves only Galβ(1,3)GalNAcα structures from glycopeptides. The expected digestion product of this glycopeptide should have a molecular weight 365 u lower (-Gal-GalNAc) than the untreated glycopeptide. Instead, two new compounds were found after RP-HPLC purification from this digest with molecular weights 162 and 568 u lower than the undigested glycopeptide, determined by LSIMS. These mass differences correspond to a Hex and to a $Hex_2HexNAc$ unit. Figure 8 shows the high energy CID spectrum of the glycopeptide with MH^+ 162 u less than the undigested glycopeptide (at m/z 1725.5). The spectrum is dominated by fragments due to the carbohydrate losses from the molecular ion at m/z 1563 (-Hex), 1522 (-HexNAc), 1360 (-HexHexNAc), 1198 (-$Hex_2HexNAc$), 1157 (-$HexHexNAc_2$), 995 (-$Hex_2HexNAc_2$), and 792 (-$Hex_2HexNAc_3$). The ion pairs 100 u apart are present due to the cleavage of the N-terminal tBOC-group. The loss of only a HexNAc from the molecular ion shows that a terminal HexNAc

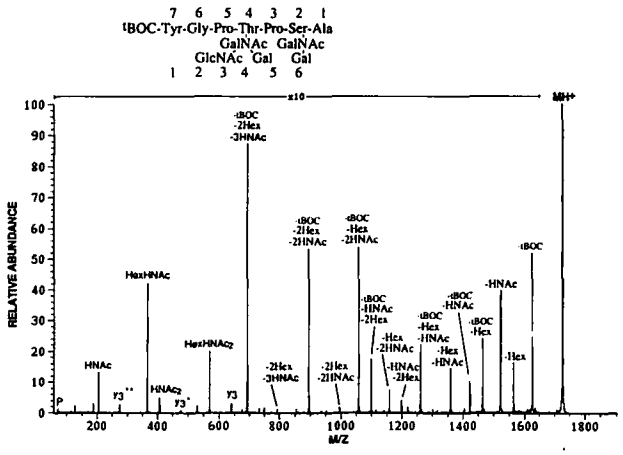

Figure 8. High energy CID spectrum of an asialo pronase fetuin glycopeptide after *O*-glycanase digestion. All the pronase glycopeptides were derivatized with tBOC-Tyr-2,4,5-trichlorophenyl ester as described in Experimental. The glycopeptide that contains the tetra- and disaccharide (P-2) was further digested with *O*-glycanase. Two new glycopeptides were isolated from the digest. MH^+ for this glycopeptide was at m/z 1725.5. The carbohydrate fragment ions are labeled as in Figure 1. Ion $\mathbf{y}_3$ is a peptide fragment formed via peptide bond cleavage between the Thr and Pro residues and with charge retention at the C-terminus, which bears the disaccharide. Ions $\mathbf{y}_3^*$ and $\mathbf{y}_3^{**}$ are the corresponding partially or fully deglycosylated fragments.

residue has been produced in the endo-α-*N*-acetylgalactosaminidase digestion. The largest nonreducing terminal oxonium ion is at m/z 569, which corresponds to a $HexHexNAc_2$ composition, which suggests the presence of a di- and a trisaccharide (Figure 8). If the intact tetrasaccharide were present, an abundant ion would be seen at mass 731 (Townsend, R. R.; Medzihradszky, K. F.; Burlingame, A. L., in preparation). Thus, because a terminal HexNAc residue has to be present (see preceding text) it can be deduced that the Gal residue that was attached to the GlcNAc in the tetrasaccharide was cleaved. Otherwise, no HexNAc loss from the molecular ion could have been observed in the presence of a trisaccharide. This residue was apparently removed by a β-galactosidase contaminant of the endo-α-*N*-acetylgalactosidaminidase. The enzyme preparation used for this digest was isolated from *Streptococcus pneumoniae*, which is known to contain a β-galactosidase more specific for β(1,4) linkage [22]. Among the *O*-linked carbohydrate structures described for bovine fetuin, only the hexasaccharide contains a Galβ(1,4) linkage (linked to the GlcNAc residue). Glycopeptide fragment $\mathbf{y}_3$ ions, which indicate the presence of a disaccharide on Ser^{264}, also were detected at m/z 639 and 477 (-Hex), as was observed for the undigested glycopeptide (Townsend, R. R.; Medzihradszky, K. F.; Burlingame, A. L., in preparation). Figure 9 shows the CID spectrum of the other digestion product with MH^+ at m/z 1360.5. Fragment ions due to sequential carbohydrate losses and due to the N-terminal tBOC group cleavages are at m/z 1198 (-Hex), 1098 (-tBOC, -Hex), 1157 (-HexNAc), 1057

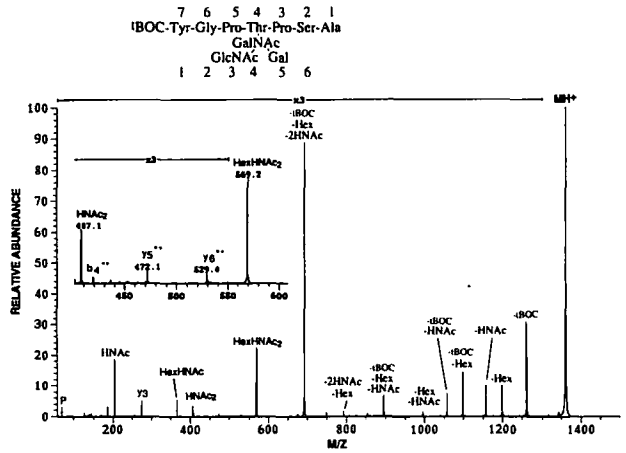

Figure 9. High energy CID spectrum of an asialo pronase fetuin glycopeptide after *O*-glycanase digestion. $MH^+ = 1360.5$. This glycopeptide was prepared as described in Figure 8 and the fragments are labeled as in Figure 8. The inset shows the lack of glycosylated y_3 ion at m/z 477, which was present in the diglycosylated glycopeptide (see Figure 8). The asterisks in the peptide fragment labels indicate gas-phase deglycosylation and the loss of the tBOC-group (b_4^{**}).

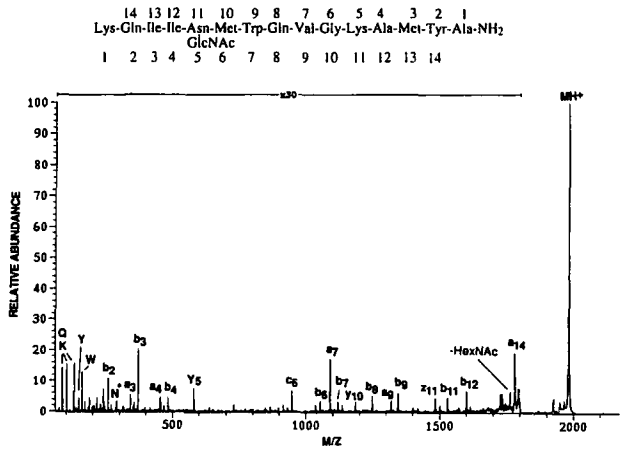

Figure 10. High energy CID spectrum of a synthetic glycopeptide with an Asn-linked *N*-acetylglucosamine residue. $MH^+ = 1983$. N* indicates the immonium ion that corresponds to the modified Asn residue. The peptide fragments are labeled according to the nomenclature in [19]. Gas-phase deglycosylation of the fragments was not observed.

(-tBOC, -HexNAc), 995 (-HexHexNAc), 895 (-tBOC, -HexHexNAc), 792 (-HexHexNAc$_2$), and 692 (-tBOC, -HexHexNAc$_2$). Oxonium ions were detected at m/z 204, 366, 407, and 569, which indicate the presence of a trisaccharide. This trisaccharide consists of a HexNAc and Hex linked to a HexNAc, consistent with the observation of Hex or HexNAc losses from the molecular ion. The glycopeptide y_3 fragments that were observed prior to the *O*-glycanase treatment (Townsend, R. R.; Medzihradszky, K. F.; Burlingame, A. L., in preparation) and in the diglycosylated species (Figure 8) at m/z 639 and 477 were not detected for this glycopeptide (Figure 9). The lack of these fragment ions (see inset in Figure 9) suggests that Ser^{264} is not occupied. In addition, very weak y_5 fragments were observed with the complete carbohydrate structure at m/z 1040 and partially or fully deglycosylated at m/z 978 (-Hex), 675 (-Hex-HexNAc), and 472, which indicates the presence of a GlcNAc(Gal)GalNAc oligosaccharide at Thr^{262} (not labeled).

We determined whether the glycopeptides with a GlcNAc residue linked to the amide-*N* of Asn show a similar fragmentation pattern. The CID spectrum of a synthetic *N*-linked glycopeptide, Lys-Gln-Ile-Ile-(GlcNAc) Asn-Met-Trp-Gln-Val-Gly-Lys-Ala-Met-Tyr-Ala-NH_2, is shown in Figure 10. This synthetic glycopeptide does not show the consensus sequence for *N*-glycosylation; it contains a Trp instead of Thr, Ser, or Cys. Aside from this difference, similar glycopeptides would result from the digestion of glycopeptides that bear a high mannose oligosaccharide with endoglycosidases [23]. The CID fragmentation of this *N*-linked glycopeptide is significantly different from that observed for the *O*-linked glycopeptides. Most of the peptide sequence fragments are shifted as expected and do not show the carbohydrate moiety elimination, the dominant fragmentation process characteristic of *O*-linked glycopeptides. At m/z 290 an immonium ion for the modified Asn was detected. Both N- and C-terminal fragment ions show the 203-u shift due to the presence of the *N*-acetylglucosamine at the Asn. For example $\mathbf{b}_5$ is observed at m/z 800, $\mathbf{b}_6$ at m/z 931, $\mathbf{c}_6$ at m/z 948, $\mathbf{a}_7$ at m/z 1089, etc., and $\mathbf{y}_{11}$ was observed at m/z 1500. Synthetic *O*-linked glycopeptides of identical sequence and bearing either an *N*-acetylglucosamine or *N*-acetylgalactosamine residue showed no differences in the fragmentation (data not shown). This difference apparently reflects the relative stability of the *N*-linked glycopeptides to gas-phase deglycosylation, and not the effects of the different carbohydrate units (GlcNAc versus GalNAc).

Mass spectrometric analysis of protein *O*-glycosylation usually has been limited to the released permethylated oligosaccharides [24]. However, to locate the modified amino acids and to address site heterogeneity, both the peptides and the oligosaccharides attached to them have to be studied as a unit. Molecular weight measurement alone will not establish the identity of the glycopeptide; additional methods are necessary to reveal the amino acid sequence, the structure of the carbohydrate(s), and the glycosylation site(s). Tandem mass spectrometry has enabled us to choose a single component (mass value) from a mixture and to obtain considerable structural information directly from the CID spectrum of the intact *O*-linked glycopeptide. Such analyses have been used successfully for the structure elucidation of arabinosyl hydroxyproline residue-containing *O*-linked glycopeptides from an unfractionated pronase digest of a cell wall glycoprotein from the gymnosperm of Douglas fir [25]. Whether the detection and discrimination of isomeric

isobaric components is possible or not will depend upon whether there are distinct fragments observed in a given CID spectrum that are attributable to each isomer [26]. Agarwala et al. [27] have reported high energy CID spectra of an *O*-linked glycopeptide from human factor IX. However, the precursor ions selected may represent LSIMS-generated fragments, comprised of a mixture of two differently sialylated structures that were reflected in the carbohydrate fragments observed. To analyze glycopeptides in surface sputtering methods such as LSIMS or fast-atom bombardment, derivatization of the glycopeptides has been used to achieve greater hydrophobicity and thus prevent the suppression of the hydrophilic glycopeptides by more surface active peptides [13, 28, 29]. Dell and co-workers used a mixed anhydride of TFA and propionic acid to modify the amino and hydroxyl groups in glycopeptides. Although this derivatization yielded a very complex mixture of differently acylated glycopeptides, increased detection sensitivity was observed. However, because amino acids as well as the oligosaccharides are modified during this procedure, molecular weight determination of the original glycopeptide can be difficult, if not impossible. In addition, due to the complexity of the resulting mixture, the relative abundance of the ions is low—not suitable for tandem mass spectrometric experiments. As another approach, the amino terminus of the glycopeptide can be modified with a hydrophobic residue without affecting the carbohydrate structure. N-terminal derivatization with tBOC-Tyr can increase the mass spectrometric detection sensitivity by a factor of 10 [10]. The other major difficulty in the CID analysis of *O*-linked glycopeptides is the greater susceptibility of the glycosidic bond to fragmentation. Facile elimination of the carbohydrate unit from the peptide results in oligosaccharide-bearing peptide fragments of low abundance. It has been reported that **y** ions that have a Pro residue as their N-terminus are usually very abundant due to their relatively higher stability. All the examples presented here have a Pro residue that precedes the glycosylated amino acid(s), and indeed in most, if not all cases the corresponding **y** ions provided the necessary information to establish the site of modification. The C-terminal fragments are more likely to retain the carbohydrates than the ions with charge retention at the N-terminus (see Figures 1 and 2). Thus, the lack of Pro residues or C-terminal basic amino acids that can lead to preferential charge retention at the N-terminus may prevent the determination of the glycosylation sites [28]. In low energy collisions the difference between the peptide and glycosidic bond seems to be even more significant—usually no information concerning the attachment site(s) can be obtained [2], although preferential C-terminal charge retention, that is, Arg, His, or Lys residues at the C-terminus in connection with Pro residues, may lead to carbohydrate-bearing fragment formation [30].

Conclusions

Interpretation of high energy CID spectra of *O*-linked glycopeptides directly provides structural information on the carbohydrate(s) present and in many cases the site(s) of their attachment from the shifts in peptide sequence ions in a single analysis. From the presence of oxonium ions and analogous carbohydrate unit losses from the molecular ion together the oligosaccharide composition and sequence may be established. Once this is accomplished, further mass spectrometric studies that employ enzymatic or chemical degradation of the peptide chain and/or the oligosaccharide structure(s) present can be employed to establish unambiguous structure. For example, the use of neuraminidases with different substrate specificities (β-galactosidase, *O*-glycanase) will reveal particular aspects of the carbohydrate structure. The identity of such modified peptides can be established by a search of the known protein sequence for molecular weight matches (without the oligosaccharide units) and comparison of the fragment ions observed with those expected from a given amino acid sequence [31]. The carbohydrate attachment site usually can be delineated from the fragments observed in the high energy CID mass spectra. In cases where this is not possible, usually the Edman degradation will provide the necessary information.

Although in *O*-linked glycopeptides gas-phase deglycosylation seems to be the favored fragmentation process, *N*-linked glycopeptides with a single *N*-acetyl-glucosamine possess a relatively stable linkage between the carbohydrate unit and the Asn residue.

Acknowledgments

The authors would like to thank F. C. Walls for his technical assistance and F. R. Masiarz for the Edman data. We are indebted to L. Otvos, Jr. (Wistar Institute, Philadelphia) for supplying the synthetic glycopeptides. This work was supported by grants from the National Center for Research Resources, National Institutes of Health (Grant RR01614), and the National Science Foundation (Grant 8700766).

References

1. Settineri, C. A.; Burlingame, A. L. In *Carbohydrate Analysis*; El Rassi, Z., Ed.; Elsevier: Amsterdam, 1995; pp 447–514.
2. Carr, S. A.; Huddleston, M. J.; Bean, M. F. *Protein Sci.* **1993**, *2*, 183–196.
3. Schindler, P. A.; Settineri, C. A.; Collet, X.; Fielding, C. J.; Burlingame, A. L. *Protein Sci.* **1995**, *4*, 791–803.
4. Umemoto, J.; Bhanvanadan, V. P.; Davidson, E. A. *J. Biol. Chem.* **1977**, *252*, 8609–8614.
5. Harris, R. J.; van Halbeek, H.; Glushka, J.; Basa, L. J.; Ling, V. T.; Smith, K. J.; Spellman, M. W. *Biochemistry* **1993**, *32*, 6539–6547.
6. Hayes, G. R.; Enns, C. A.; Lucas, J. J. *Glycobiology* **1992**, *2*, 355–359.
7. Andersen, J. S.; Sogaard, M.; Svensson, B.; Roepstorff, P. *Biol. Mass Spectrom.* **1994**, *23*, 547–554.

8. Pisano, A.; Redmond, J. W.; Williams, K. L.; Gooley, A. A. *Glycobiology* **1993,** *3,* 429–435.
9. Settineri, C. A.; Medzihradszky, K. F.; Masiarz, F. R.; Chu, C.; George-Nascimento, C.; Burlingame, A. L. *Biomed. Environ. Mass Spectrom.* **1990,** *19,* 665–676.
10. Medzihradszky, K. F.; Gillece-Castro, B. L.; Settineri, C. A.; Townsend, R. R.; Masiarz, F. R.; Burlingame, A. L. *Biomed. Environ. Mass Spectrom.* **1990,** *19,* 777–781.
11. Townsend, R. R.; Alai, M.; Hardy, M. R.; Fenselau, C. C. *Anal. Biochem.* **1988,** *171,* 180–191.
12. Hardy, M. R.; Townsend, R. R.; Lee, Y. C. *Anal. Biochem.* **1988,** *170,* 54–62.
13. Townsend, R. R.; Hardy, M. R.; Wong, T. C.; Lee, Y. C. *Biochemistry* **1986,** *25,* 5716–5725.
14. McKelvy, J. F.; Lee, Y. C. *Arch. Biochem. Biophys.* **1969,** *130,* 629–635.
15. Lee, Y. C. *Methods Enzymol.* **1972,** *28,* 63–72.
16. Walls, F. C.; Baldwin, M. A.; Falick, A. M.; Gibson, B. W.; Kaur, S.; Maltby, D. A.; Gillece-Castro, B. L.; Medzihradszky, K. F.; Evans, S.; Burlingame, A. L. In *Biological Mass Spectrometry;* Burlingame, A. L.; McCloskey, J. A., Eds.; Elsevier: Amsterdam, 1990; pp 197–216.
17. Domon, B.; Costello, C. E. *Glycoconjugate J.* **1988,** *5,* 397–409.
18. Dwigielewska, K. M.; Brown, W. M.; Casey, S. J.; Christie, D. L.; Foreman, R. C.; Hill, R. M.; Saunders, N. R. *J. Biol. Chem.* **1990,** *265,* 4354–4357.
19. Biemann, K. *Methods Enzymol.* **1990,** *193,* 886–887.
20. Edge, A. S. B.; Spiro, R. G. *J. Biol. Chem.* **1987,** *262,* 16135–16141.
21. Townsend, R. R.; Medzihradszky, K. F.; Hardy, M. R.; Rohrer, J.; Burlingame, A. L. Poster Presentation at the 11th International Symposium on Glycoconjugates, Toronto, 1991.
22. Jacob, G. S.; Scudder, P. *Methods Enzymol.* **1994,** *230,* 280–299.
23. Tarentino, A. L.; Plummer, T. H., Jr. *Methods Enzymol.* **1994,** *230,* 44–57.
24. Dell, A.; Reason, A. J.; Khoo, K. H.; Panico, M.; McDowell, R. A.; Morris, H. R. *Methods Enzymol.* **1994,** *230,* 108–132.
25. Kieliszewski, M. J.; O'Neill, M.; Leykam, J.; Orlando, R. *J. Biol. Chem.* **1995,** *270,* 2541–2549.
26. Ding, A.; Zia-Amirhosseini, P.; McDonagh, A. F.; Burlingame, A. L.; Benet, L. Z. *Drug. Metab. Dispos.* **1994,** *23,* 369–376.
27. Agarwala, K. L.; Kawabata, S.; Takao, T.; Murata, H.; Shimonishi, Y.; Nishimura, H.; Iwanaga, S. *Biochemistry* **1994,** *33,* 5167–5171.
28. Reason, A. J.; Blench, I. P.; Haltiwanger, R. S.; Hart, G. W.; Morris, H. R.; Panico, M.; Dell, A. *Glycobiology* **1991,** *1,* 585–594.
29. Reason, A. J.; Morris, H. R.; Panico, M.; Marais, R.; Treisman, R. H.; Haltiwanger, R. S.; Hart, G. W.; Kelly, W. G.; Dell, A. *J. Biol. Chem.* **1992,** *267,* 16911–16921.
30. Linsley, K. B.; Chan, S. Y.; Chan, S.; Reinhold, B. B.; Lisi, P. J.; Reinhold, V. R. *Anal. Biochem.* **1994,** *219,* 207–217.
31. Medzihradszky, K. F.; Burlingame, A. L. *Methods: A Companion to Methods in Enzymology* **1994,** *6,* 284.