Protein Science (1995), *4*:791-803. Cambridge University Press. Printed in the USA.
Copyright © 1995 The Protein Society

Site-specific detection and structural characterization of the glycosylation of human plasma proteins lecithin:cholesterol acyltransferase and apolipoprotein D using HPLC/electrospray mass spectrometry and sequential glycosidase digestion

PATRICK A. SCHINDLER,[1,5] CHRISTINE A. SETTINERI,[1] XAVIER COLLET,[2] CHRISTOPHER J. FIELDING,[3] AND ALMA L. BURLINGAME[1,4]

[1] Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94143
[2] INSERM U326, Hospital Puspan, Toulouse, France
[3] Cardiovascular Research Institute and Department of Physiology, University of California, San Francisco, California 94143
[4] Liver Center, University of California, San Francisco, California 94143

(RECEIVED December 20, 1994; ACCEPTED February 1, 1995)

Abstract

Site-specific structural characterization of the glycosylation of human lecithin:cholesterol acyltransferase (LCAT) was carried out using microbore reversed-phase high performance liquid chromatography coupled with electrospray ionization mass spectrometry (HPLC/ESIMS). A recently described mass spectrometric technique involving monitoring of carbohydrate-specific fragment ions during HPLC/ESIMS was employed to locate eight different groups of glycopeptides in a digest of a human LCAT protein preparation. In addition to the four expected *N*-linked glycopeptides of LCAT, a di-*O*-linked glycopeptide was detected, as well as three additional glycopeptides. Structural information on the oligosaccharides from all eight glycopeptides was obtained by sequential glycosidase digestion of the glycopeptides followed by HPLC/ESIMS. All four potential *N*-linked glycosylation sites (Asn^{20}, Asn^{84}, Asn^{272}, and Asn^{384}) of LCAT were determined to contain sialylated triantennary and/or biantennary complex structures. Two unanticipated *O*-linked glycosylation sites were identified at Thr^{407} and Ser^{409} of the LCAT *O*-linked glycopeptide, each of which contain sialylated galactose$\beta 1 \rightarrow 3$*N*-acetylgalactosamine structures. The three additional glycopeptides were determined to be from a copurifying protein, apolipoprotein D, which contains potential *N*-linked glycosylation sites at Asn^{45} and Asn^{78}. These glycopeptides were determined to bear sialylated triantennary oligosaccharides or fucosylated sialylated biantennary oligosaccharides. Previous studies of LCAT indicated that removal of the glycosylation site at Asn^{272} converts this protein to a phospholipase (Francone OL, Evangelista L, Fielding CJ, 1993, *Biochim Biophys Acta 1166*:301-304). Our results indicate that the carbohydrate structures themselves are not the source of this functional discrimination; rather, it must be mediated by the structural environment around Asn^{272}.

Keywords: apolipoprotein D; exoglycosidase sequencing; glycopeptides; lecithin:cholesterol acyl transferase; oligosaccharides; selected ion monitoring

Reprint requests to: Alma L. Burlingame, Department of Pharmaceutical Chemistry, 513 Parnassus Avenue, Room S-1124, University of California, San Francisco, California 94143-0446; e-mail: alb@itsa.ucsf.edu.

[5] Present address: Ciba-Geigy Ltd., Basel CH-4002, Switzerland.

Abbreviations: LCAT, human lecithin:cholesterol acyl transferase; Apo D, apolipoprotein D; HPLC/ESIMS, high performance liquid chromatography coupled to electrospray ionization mass spectrometry; CID, collision-induced dissociation; HPLC/ESI/CID/MS, collision-induced dissociation during HPLC/ESIMS; SIM, selected ion monitoring; UV, ultraviolet; MS, mass spectrometry; FAB-MS, fast atom bombardment mass spectrometry; MALDI-MS, matrix-assisted laser desorption ionization mass spectrometry; Q2, quadrupole number two in a triple quadrupole mass analyzer; *O*-linked, serine/threonine-linked; *N*-linked, asparagine-linked; DTT, dithiothreitol; TFA, trifluoroacetic acid; Gal, galactose; GalNAc, *N*-acetylgalactosamine; GlcNAc, *N*-acetylglucosamine; NeuAc, *N*-acetyl neuraminic acid or sialic acid; Hex, hexose; HexNAc, *N*-acetylhexosamine; mU, milliunit; pGlu, pyroglutamic acid.

Lecithin:cholesterol acyltransferase catalyzes the transfer of an acyl group from the 2-position of phosphatidylcholine to the 3-hydroxyl group of cholesterol in blood plasma (Aron et al., 1978). In the absense of cholesterol, long-chain alcohols or water itself can act as acceptors for the acyl group, generating fatty esters or free acids, respectively (Subbaiah et al., 1980). The primary structure of human LCAT has been determined by amino acid (Yang et al., 1987) and cDNA (McLean et al., 1986) sequencing, and the sequence is shown in Figure 1. These studies of LCAT employing Edman degradation data suggested that all four potential Asn-linked (*N*-linked) glycosylation sites (as indicated by the consensus sequon Asn-Xxx-Ser/Thr, where Xxx is any amino acid except proline [Bause & Hettkamp, 1979]) at Asn^{20}, Asn^{84}, Asn^{272}, and Asn^{384} (shown in Fig. 1) were glycosylated (Yang et al., 1987). The carbohydrate composition was previously determined to contain 13% hexose, 6.2% hexosamine, and 5.4% sialic acid residues (Chong et al., 1983). In addition, further characterization of the nature of the glycosylation has been reported involving the use of glycosidases and specific inhibitors of glycoprotein assembly with cultured cells secreting LCAT activity. These studies concluded that LCAT contained high mannose and complex type *N*-linked oligosaccharides (Collet & Fielding, 1991). Although the function of the LCAT carbohydrates is not well understood, it has been shown that their presence is necessary for full activity of LCAT but not for synthesis or secretion of the LCAT protein containing no oligosaccharides (Collet & Fielding, 1991). It has been established by site-directed mutagenesis that removal of the glycosylation site at Asn^{272} converts LCAT into a phospholipase, generating fatty acids rather than cholesteryl esters from phosphatidylcholine (Francone et al., 1993). Therefore, this study was undertaken to determine the structural nature and heterogeneity of the glycosylation at the four known *N*-linked sites on this protein in order to ascertain whether qualitative structural differences were present only at Asn^{272}.

In the last decade, mass spectrometry has been used extensively to characterize posttranslational modifications of proteins (Biemann & Martin, 1987; Carr et al., 1991; Burlingame et al., 1992). Fast atom bombardment MS (Carr & Roberts, 1986; Carr et al., 1989; Poulter et al., 1991) and, more recently, electrospray ionization MS (Fenn et al., 1989; Loo et al., 1989; Hemling et al., 1990; Smith et al., 1990) and matrix-assisted laser desorption ionization MS (Karas et al., 1989; Siegel et al., 1991; Chait & Kent, 1992; Huberty et al., 1993) have been used successfully to identify and structurally characterize glycopeptides and carbohydrates. Although of recent origin, one of the most commonly used techniques today for mapping proteolytic fragments of proteins and glycoproteins is reversed phase high performance liquid chromatography coupled with electrospray ionization mass spectrometry (Ling et al., 1991; Duffin et al., 1992; Carr et al., 1993; Medzihradszky et al., 1994). In addition, ESI mass spectrometers can be used as selective detectors for glycopeptides, by monitoring for specific diagnostic sugar oxonium ions (selected ion monitoring) produced by increasing the collision excitation potential in the electrospray ionization source (collision-induced dissociation) during the HPLC/ESIMS experiment (Duffin et al., 1992; Carr et al., 1993; Huddleston et al., 1993). The dissociation of the molecular ions can be performed either in the electrospray source, by increasing the potential difference between the sampling orifice and the skimmer, or by achieving collisions in the second quadrupole of a triple quadrupole mass spectrometer (Carr et al., 1993). The former is the more sensitive method whereas the latter is more specific, because only precursor ions that decompose to yield the fragments of interest are detected. Several groups have published their own versions of the SIM technique (Duffin et al., 1992; Carr et al., 1993; Huddleston et al., 1993; Medzihradszky et al., 1994) and have demonstrated its sensitivity and selectivity to detect glycopeptides during liquid chromatographic separation of glycoprotein digests.

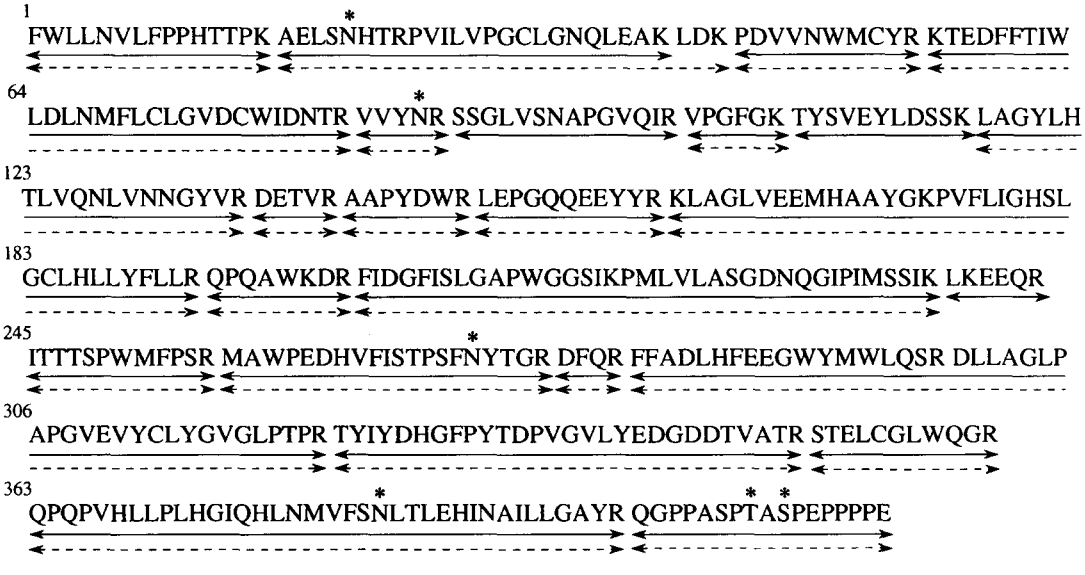

Fig. 1. Primary sequence of human LCAT. Glycosylation sites identified at Asn^{20}, Asn^{84}, Asn^{272}, and Asn^{384} (*N*-linked) as well as Thr^{407} and Ser^{409} (*O*-linked) are designated by asterisks. Underlining with the solid arrows indicates peptides and glycopeptides identified by HPLC/ESIMS, whereas the dashed arrows indicate those identified by Edman sequencing in our laboratory.

In the present study, a combination of trypsin digestion, HPLC/ESIMS, SIM, Edman degradation, and exo- and endoglycosidase digestions was used to obtain a global picture of the glycosylation on LCAT, including characterization of the structures and their heterogeneity at each glycosylation site. Using these methods, two previously unidentified *O*-linked glycosylation sites were identified and characterized. In addition, the power and sensitivity of these methods are further emphasized by the detection and characterization of three *N*-linked glycopeptides from an unsuspected low-level component in the LCAT preparation, apolipoprotein D.

Results

Localization of glycopeptides by HPLC/ESIMS with selected ion monitoring

Because the goal of this study was not only to characterize the carbohydrates from LCAT but also to assess the heterogeneity at each site, it was necessary to first detect and isolate the corresponding glycopeptides and then further analyze the carbohydrate structures attached to each peptide. First, approximately 1 nmol of reduced, carboxymethylated LCAT was digested with trypsin. One 300-pmol aliquot of the tryptic digest was analyzed by microbore reversed phase HPLC/ESIMS and another 300-pmol aliquot was analyzed by HPLC/ESI/CID/MS with SIM. The HPLC/ESIMS analysis of the tryptic digest was performed in the conventional operating mode, in which a mass chromatogram representing the total ion abundance, resulting from scanning the entire mass range (*m/z* 330–2,000), and a UV absorbance chromatogram were obtained. Only 5% of the material injected flowed into the mass spectrometer, while 95% was manually recollected after HPLC separation for further analysis. Because the primary sequence of LCAT was known, the sequences of the 21 tryptic peptides could be confirmed by identification of the molecular masses of the predicted tryptic peptides, as indicated by the solid arrows in Figure 1. As illustrated in Figure 2, for each UV peak there exists a corresponding MS (base peak intensity) peak (see top and middle panels, Fig. 2); however, their relative intensities are not necessarily the same.

Using a second 300-pmol aliquot of the tryptic digest, the HPLC/ESI/CID/MS with SIM experiment was then performed in order to identify which of the UV and MS peaks contain glycopeptides. Again, 95% of the injected material was manually collected into the vials containing the fractions that had the same retention time in the first experiment. The remaining 5% was directed into the ESI mass spectrometer, which was scanned in the SIM mode. For this experiment, the voltage of the sampling cone in the source of the mass spectrometer was adjusted to 180 V (from 60 V). This higher voltage increased collisions in

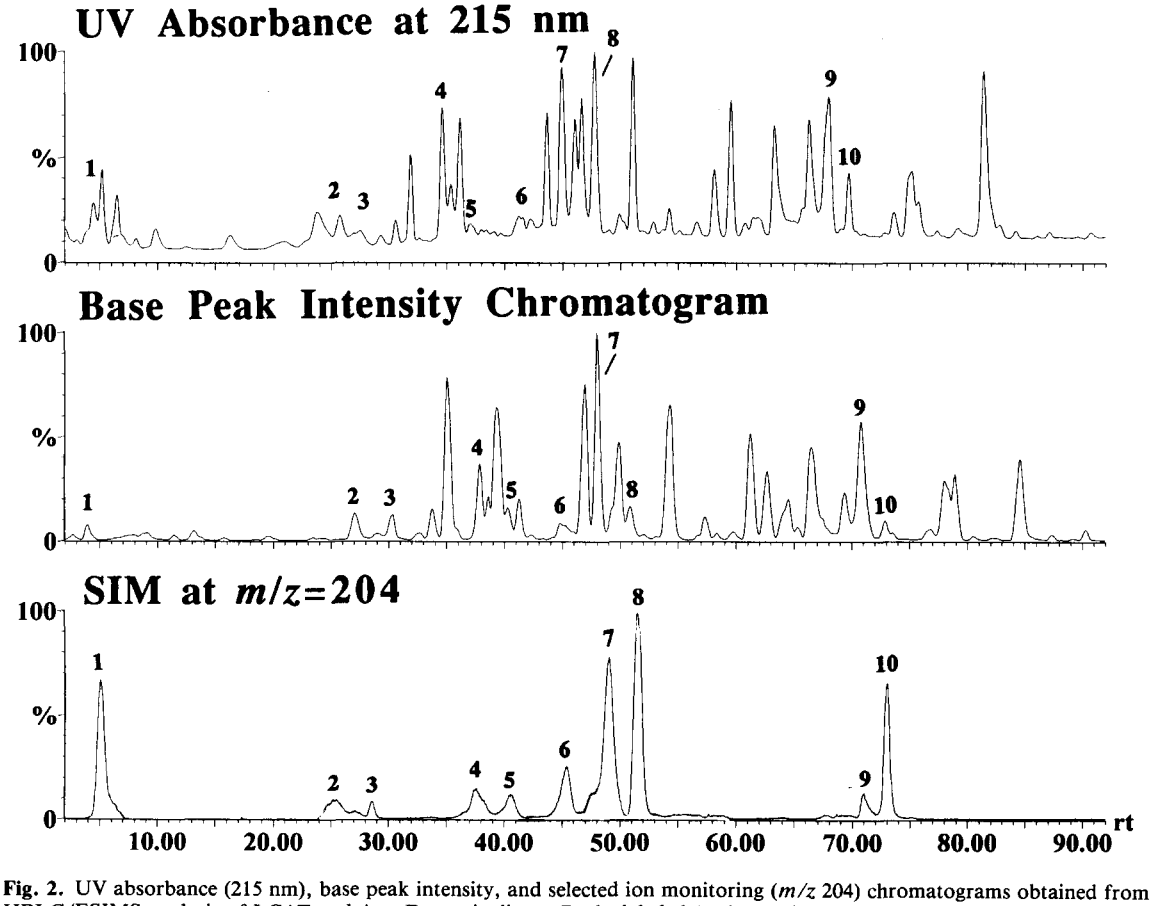

Fig. 2. UV absorbance (215 nm), base peak intensity, and selected ion monitoring (*m/z* 204) chromatograms obtained from HPLC/ESIMS analysis of LCAT and Apo D tryptic digest. Peaks labeled 1–10 were identified as glycopeptides and are discussed in the text.

the ESI source, which resulted in a high fragmentation yield. It has been demonstrated by several groups (Duffin et al., 1992; Carr et al., 1993; Huddleston et al., 1993) that under these conditions, carbohydrate moieties produce very characteristic fragments that can be used as diagnostic ions for the presence of glycopeptides. The mass spectrometer monitored the following selected mass values and the indicated carbohydrate moieties: m/z 204 (*N*-acetylhexosamine, $[HexNAc]^+$), m/z 292 (*N*-acetylneuraminic acid or sialic acid, $[NeuAc]^+$), m/z 366 (hexosyl *N*-acetylhexosamine, $[Hex\text{-}HexNAc]^+$), and m/z 657 (sialyl hexosyl *N*-acetylhexosamine, $[NeuAc\text{-}Hex\text{-}HexNAc]^+$). The chromatogram representing SIM at m/z 204 is illustrated in Figure 2 (lower panel). All four SIM chromatograms (shown in Fig. 3) could be virtually superimposed, indicating that the results obtained are definitely due to the presence of sialylated complex oligosaccharides (Kornfeld & Kornfeld, 1976) and not fortuitously formed peptide fragments. Ten different glycopeptide-containing fractions were identfied, as shown in Figure 2, where the corresponding UV, MS, and SIM peaks are labeled 1 through 10. Once the glycopeptide-containing peaks were identified, the electrospray mass spectra of the glycopeptides from the first analysis could be identified by direct comparison of the UV chromatograms from the two experiments and the retention times of these SIM peaks with the molecular masses of the intact glycopeptides determined.

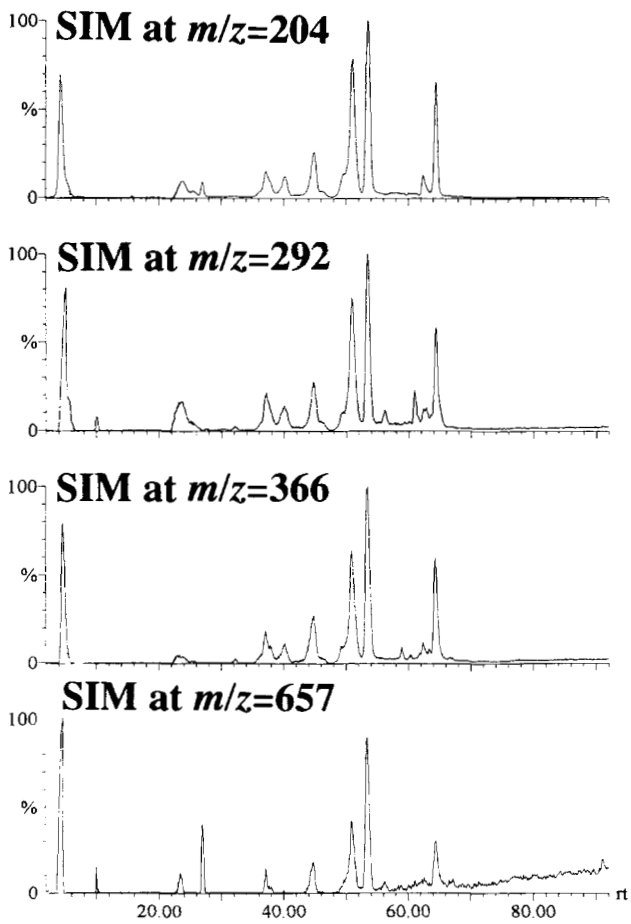

Fig. 3. Chromatograms obtained from selected ion monitoring at m/z 204 ($HexNAc^+$), m/z 292 ($NeuAc^+$), m/z 366 ($Hex\text{-}HexNAc^+$), and m/z 657 ($NeuAc\text{-}Hex\text{-}HexNAc^+$) from a 300-pmol injection of the LCAT + Apo D digest that was split 1:20 so that only 5% of the sample entered the mass spectrometer. Voltage of the sampling cone was adjusted from 60 V to 180 V during the SIM experiments in order to induce dissociation.

Because of the many ways in which monosaccharides can be linked together in an oligosaccharide, there is a very large number of possible carbohydrate structures that could represent a given composition. However, the types of *N*-linked structures typical of mammalian proteins (represented in three classes termed high mannose, hybrid, and complex [Kornfeld & Kornfeld, 1976]) contain a defined number of sequence and branching combinations, and the molecular weight of the carbohydrate portion of a glycopeptide enables one to predict the oligosaccharide composition as well as the structural class of the structure present on that glycopeptide. By subtracting the calculated molecular weight of the amino acids from the measured mass of the glycopeptides, it was possible to deduce the likely composition as well as the class and general structure (e.g., complex, trisialo triantennary) of the carbohydrate moieties. Based on the molecular weights and the SIM data, the peaks labeled 1 and 7–10 in Figure 2 contained the four predicted LCAT tryptic *N*-linked glycopeptides (where peaks 9 and 10 contained the same glycopeptide, except the peptide in fraction 10 contained an N-terminal pyroglutamic acid residue instead of a Gln based on a molecular weight difference of 17 Da and the expected occurrence of an N-terminal Gln in this peptide). The proposed oligosaccharide composition for each glycoform is presented in Table 1. Unexpectedly, the carbohydrate-containing components labeled 2 and 3 in Figure 2 revealed molecular weights that corresponded to the sum of the expected C-terminal tryptic peptide, residues 400–416, together with attached *O*-linked hexa- to undecasaccharides. This peptide contains no Asn but one threonine and two serine residues; the proposed oligosaccharide composition for each glycoform is listed in Table 2. Again, a similar glycopeptide molecular weight distribution was observed in fraction 3 but 17 Da less than that in fraction 2, and thus this peptide was assumed to be a pyroglutamic acid-containing glycopeptide because it also contained an N-terminal Gln. Finally, based on the identified molecular weights observed for fractions 4, 5, and 6 (Fig. 2), it appeared that these components were not related to LCAT *N*-linked or *O*-linked glycopeptides. Further characterization would be required to determine their structural nature and origin.

Characterization of glycopeptides by Edman degradation

Because several unexpected glycopeptide components were detected in the tryptic digest, all the collected fractions from the HPLC/ESIMS experiments were subsequently subjected to automated Edman degradation. Those complete peptide sequences determined by our Edman analyses are indicated in Figure 1 by the dashed arrows. For the glycopeptides underlined by dashed arrows, blank cycles were obtained at the Asn residues that were glycosylated (indicated by asterisks in Fig. 1). By confirming the peptide sequence of the glycopeptides, the Edman data, combined with molecular weight and SIM data obtained from the corresponding mass spectra, established the nature of the glycosylation for all of the expected LCAT *N*-linked glycopeptides and confirmed the presence of the unexpected *O*-linked glycopeptides, as listed in Tables 1 and 2, respectively. For example,

Table 1. *HPLC/electrospray mass spectrometry data obtained from sequential glycosidase digestion of LCAT N-linked glycopeptides*[a]

Fraction	Glycosylation site and sequence range	Intact glycopeptide MW (Da)	Carbohydrate composition	MW after neuraminidase digestion (Da)	MW after β-galactosidase digestion	MW after hexosaminidase digestion	Confirmed carbohydrate structure
1	Asn^{84} (81-85)	3,221.8	$Hex_6HexNAc_5NeuAc_2$	2,639.0	2,152.0	1,543.6	Disialo triantennary
		3,511.7	**$Hex_6HexNAc_5NeuAc_3$**	**2,639.0**	**2,152.0**	**1,543.6**	**Trisialo triantennary**
		3,656.8	$Hex_6HexNAc_5NeuAc_3Fuc_1$	2,784.7	2,297.6	1,891.6	Fucose trisialo triantennary
7	Asn^{20} (16-39)	4,821.1	$Hex_5HexNAc_4NeuAc_2$	4,240.2	3,914.6	3,510.8	Disialo biantennary
		5,186.7	$Hex_6HexNAc_5NeuAc_2$	4,605.0	4,118.5	3,510.8	Disialo triantennary
		5,478.1	**$Hex_6HexNAc_5NeuAc_3$**	**4,605.0**	**4,118.5**	**3,510.8**	**Trisialo triantennary**
		5,623.9	$Hex_6HexNAc_5NeuAc_3Fuc_1$	4,753.0	4,426.3	4,021.0	Fucose trisialo triantennary
		5,841.4	$Hex_7HexNAc_6NeuAc_3$	4,969.9	4,318.5	3,510.8	Trisialo tetraantennary
8	Asn^{272} (257-276)	4,269.2	$Hex_5HexNAc_4NeuAc_1$	3,980.0	3,655.2	3,247.5	Monosialo biantennary
		4,560.3	$Hex_5HexNAc_4NeuAc_2$	3,980.0	3,655.2	3,247.5	Disialo biantennary
		4,925.3	$Hex_6HexNAc_5NeuAc_2$	4,345.5	3,856.9	3,247.5	Disialo triantennary
		5,216.9	**$Hex_6HexNAc_5NeuAc_3$**	**4,345.5**	**3,856.9**	**3,247.5**	**Trisialo triantennary**
		5,364.4	$Hex_6HexNAc_5NeuAc_3Fuc_1$	4,491.3	4,164.7	3,757.2	Fucose trisialo triantennary
9	Asn^{384} (363-399)	6,111.8	$Hex_5HexNAc_4NeuAc_1$	5,822.0	2,715.9 (379-390)	2,310.5 (379-390)	Monosialo biantennary
		6,404.3	**$Hex_5HexNAc_4NeuAc_2$**	**5,822.0**	**2,715.9 (379-390)**	**2,310.5 (379-390)**	**Disialo biantennary**
10	pE (363-399)	6,095.5	$Hex_5HexNAc_4NeuAc_1$	5,805.0	2,715.9 (379-390)	2,310.5 (379-390)	Monosialo biantennary
		6,386.7	$Hex_5HexNAc_4NeuAc_2$	5,805.0	2,715.9 (379-390)	2,310.5 (379-390)	Disialo biantennary

[a] Bold type indicates the major species identified based on the relative abundance of the molecular ions in the mass spectra. Fucose is assumed to be the deoxyhexose because this is the most commonly reported deoxyhexose present on human oligosaccharide structures.

Edman degradation confirmed that the peak labeled 1 in Figure 2 corresponds to the peptide containing residues ^{81}VVYNR^{85} which contains a consensus sequon for *N*-linked glycosylation at Asn^{84}.

Similarly, Edman degradation confirmed that peak 7 in Figure 2 represents the peptide containing LCAT residues ^{16}AELSNHTRPVILVPGCLGNQLEAK^{39} and the consensus sequon for *N*-linked glycosylation at Asn^{20}. Edman degradation of peak 8 in Figure 2 yielded a sequence corresponding to residues ^{257}MAWPEDHVFISTPSFNYTGR^{276}, which contains a consensus sequon for *N*-linked glycosylation at Asn^{272}. Similarly, Edman degradation confirmed that the peak labeled 9 in Figure 2 represented the peptide ^{363}QPQPVHLLPLHGIQHL NMVFSNLTLEHINAILLGAYR^{399}, containing the consensus sequon for *N*-linked glycosylation at Asn^{384}. No Edman degradation data were obtained from peak 10.

Edman degradation also established that peak 2 in Figure 2 represented the C-terminal peptide of LCAT, ^{400}QGPPASPT ASPEPPPPE^{416}, but this peptide does not contain a consensus sequon for *N*-linked glycosylation and was not previously found to be glycosylated. However, as shown in Figure 1, this peptide contains two serine residues at positions 405 and 409 and one threonine residue at position 407, which could be glycosylated with *O*-linked oligosaccharides. The sequence obtained from Edman degradation of peak 2 yielded two blank cycles, which correspond to Thr^{407} and Ser^{409}, indicating the likely presence of *O*-linked glycosylation at these sites. Just as with peak 10, peak 3 in Figure 2 did not yield a sequence by Edman degradation.

Table 2. *HPLC/electrospray mass spectrometry data obtained from glycosidase digestion of LCAT O-linked glycopeptides*[a]

Fraction	Glycosylation site and glycopeptide sequence range	Intact glycopeptide MW (Da)	Carbohydrate composition	MW after neuraminidase + *O*-glycanase digestion	Confirmed carbohydrate structures
2	Thr^{407} and Ser^{409} (400-416)	**2,968.5**	**$Hex_2HexNAc_2NeuAc_2$**	**1,656.9**	**2[NeuAc-Galβ1 → 3GalNAc]**
		3,259.6	$Hex_2HexNAc_2NeuAc_3$	1,656.9	1[NeuAc-Galβ1 → 3GalNAc] + 1[$NeuAc_2$-Galβ1 → 3GalNAc]
		3,551.6	$Hex_2HexNAc_2NeuAc_4$	1,656.9	2[$NeuAc_2$-Galβ1 → 3GalNAc]
		3,624.9	$Hex_3HexNAc_3NeuAc_3$	ND[b]	
		3,914.9	$Hex_3HexNAc_3NeuAc_4$	ND	
		4,208.4	$Hex_3HexNAc_3NeuAc_5$	ND	
3	pE (400-416)	2,950.9	$Hex_2HexNAc_2NeuAc_2$	1,638.9	2[NeuAc-Galβ1 → 3GalNAc]

[a] Bold type indicates the major species identified based on the relative abundance of the molecular ions in the mass spectra.
[b] ND, not determined.

Edman degradation analysis of peaks 4, 5, and 6 confirmed that these fractions did not contain LCAT glycopeptides, but instead contained the following peptide sequences: CIQA-blank-YSLMENGK, CIQA-blank-YSLMENGKIK, and ADGTVNQ IEGEATPV-blank-LTEPAKKLEVK. From a search of the Protein Identification Resource (PIR) database, the sequences obtained by Edman were identified as peptides 41-53, 41-55, and 63-88 from a second glycoprotein, Apo D, which apparently copurified with the LCAT glycoprotein. This protein, which contains two potential sites of *N*-linked glycosylation at Asn^{45} and Asn^{78}, has been previously shown to associate with LCAT (Albers et al., 1976; Utermann et al., 1980). However, no data have previously been reported on its oligosaccharide structures.

Proposed oligosaccharide compositions based on Edman and molecular weight data

Based on the sequence ^{81}VVYNR^{85} identified by Edman for the glycopeptides in fraction 1 in Figure 2, the carbohydrate composition of the major component (based on signal intensity in the ESI mass spectrum, which was deduced from the difference between the expected [MH^+ = 650.8 Da] and the measured [MH^+ = 3,511.7 Da] molecular weight of this glycopeptide), represents a triantennary complex structure, as listed in Table 1. The observed electrospray mass values also indicated the presence of two minor glycoforms, one containing an additional deoxyhexose (fucose) residue and another containing only two sialic acid residues (see Table 1). (Fucose is assumed to be the deoxyhexose because this is the most commonly reported deoxyhexose present on human oligosaccharide structures. Earlier glycopeptide studies carried out in this laboratory have indicated that, when comparing glycopeptides with the same peptide sequence and different oligosaccharide components, the relative ion intensities in the ESIMS data correlate closely with the relative quantities of each glycoform [C.A. Settineri, unpubl. data].)

The mass spectrum for the glycopeptide identified by Edman as ^{16}AELSNHTRPVILVPGCLGNQLEAK^{39} from fraction 7 is shown in Figure 4. The molecular weights obtained require the major species to be a trisialo triantennary structure. In addition, the presence of other mass values indicates the presence of small amounts of biantennary, tetraantennary, and fucosylated triantennary structures (see Table 1). The mass spectrum obtained for the glycopeptide ^{257}MAWPEDHVFISTPSFNYTGR^{276} from fraction 8 contains five glycoforms, as shown in Figure 5A. The major component in the spectrum corresponds to a trisialo triantennary structure, as listed in Table 1, with small ammounts of sialylated biantennary and fucosylated trisialo triantennary as well.

The molecular masses obtained for the glycopeptide ^{363}QP QPVHLLPLHGIQHLNMVFSNLTLEHINAILLGAYR^{399} from fraction 9 indicated the presence of a disialo biantennary carbohydrate structure as the major species, and some monosialo biantennary as well (shown in Table 1). Although no Edman data were obtained for fraction 10, the ESIMS data (shown in Table 1) indicated that this fraction had a molecular weight that was 17 Da less than the major glycopeptide species identified in peak 9. Therefore, it very likely contained the same glycoforms as in peak 9, but with a pyroglutamic acid as the N-terminal amino acid. Our inability to obtain an Edman sequence further supports this modification of the N-terminal amino acid.

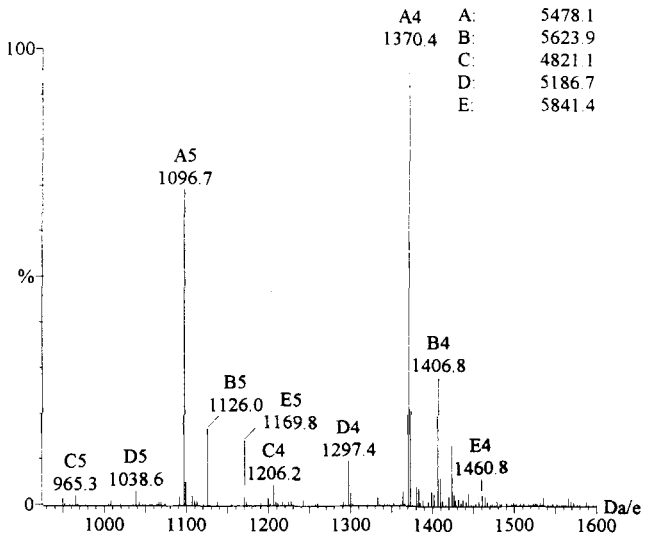

Fig. 4. Electrospray mass spectrum (fraction 7 in Fig. 2) of the LCAT tryptic peptide containing residues 16-39 (AELSNHTRPVILVPGCLG NQLEAK, MW_{avg} = 2,618.0 Da), which contains a consensus sequon for *N*-linked glycosylation at Asn^{20}. Molecular masses of the five glycopeptide components identified in this spectrum (4,821.1, 5,186.7, 5,478.1, 5,623.9, and 5,841.4 Da) represent this peptide with disialo biantennary, disialo triantennary, trisialo triantennary, trisialo fucosylated triantennary, and trisialo tetraantennary complex oligosaccharides attached at Asn^{20}, respectively.

Several molecular weights were identified in the HPLC/ESIMS data for fraction 2 from Figure 2, representing several different carbohydrate compositions, as listed in Table 2. The major *O*-linked species in fraction 2 contained a composition corresponding to two hexoses, two *N*-acetylhexosamines, and two sialic acids. The mass spectrometric data for fraction 3 contained a molecular weight that was 17 Da less than the major glycopeptide species identified in peak 2. This was assumed to be the same major glycopeptide component as in peak 2 (400-416) but with a pGlu as the N-terminal amino acid rather than the expected Gln; our inability to obtain an Edman sequence further supports this conclusion.

Molecular weights of the Apo D glycopeptides indicated that glycopeptides CIQANYSLMENGK and CIQANYSLMENG KIK from peaks 4 and 5 contain a composition corresponding to trisialo triantennary oligosaccharides, and glycopeptide ADGT VNQIEGEATPVNLTEPAKKLEVK from peak 6 contains a composition corresponding to fucosylated disialo biantennary oligosaccharides.

Characterization of N-linked carbohydrate structures of LCAT

In order to establish the structures possible from knowledge of the carbohydrate compositions listed in Table 1, in separate experiments, each of the four isolated glycopeptides was treated sequentially with specific exoglycosidases to remove the exposed nonreducing terminal monosaccharides, with analysis by HPLC/ESIMS after each step. Each exoglycosidase used was specific for one type of monosaccharide residue, such as *N*-acetyl glucos-

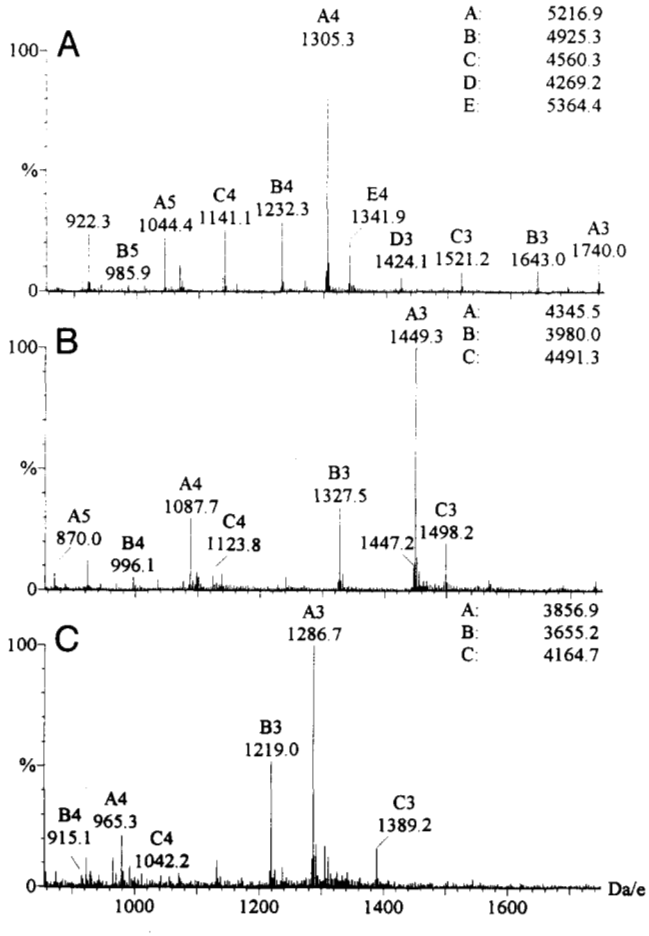

Fig. 5. A: Electrospray mass spectrum (fraction 8 in Fig. 2) of the tryptic glycopeptides containing LCAT residues 257–276 (MAWPEDHV FISTPSFNYTGR, MW_{avg} = 2,355.6 Da), which contains a consensus sequon for *N*-linked glycosylation at Asn^{272}. Molecular masses of the five glycopeptide components identified in this spectrum (4,269.2, 4,560.3, 4,925.3, 5,216.9, and 5,364.4 Da) represent this peptide with monosialo biantennary, disialo biantennary, disialo triantennary, trisialo triantennary, and trisialo fucosylated triantennary complex oligosaccharides attached at Asn^{272}, respectively, as illustrated by the structures. **B:** Electrospray mass spectrum of the glycopeptide mixture from A after digestion with *A. ureafaciens* neuraminidase. **C:** Electrospray mass spectrum of the glycopeptide mixture from B after digestion with *S. pneumoniae* β-galactosidase.

amine (GlcNAc) or galactose (Gal). The mass difference after each single exoglycosidase digestion and the monosaccharide specificity of each exoglycosidase were used to determine the sequence of the nonreducing terminal portion of the structures. For example, a change in mass of a glycopeptide after digestion with neuraminidase indicates first that sialic acid residues were present at the nonreducing terminus of the oligosaccharide, and the value of the mass difference indicates the number of sialic acid residues removed from the oligosaccharide. By using the enzymes sequentially, one can determine the sequence and branching of the oligosaccharides attached to the glycopeptides, and thereby establish the structures of the oligosaccharides at each site. To confirm the published specificity of the enzymes to be used, we first sequentially digested glycopeptides of known sequence from bovine fetuin (Green et al., 1988; Cumming et al., 1989) and analyzed them by HPLC/ESIMS (data not shown).

After establishing the specificities of the enzymes, each of the eight different glycopeptide mixtures (fractions 1 and 4–10) was first digested with *Arthrobacter ureafaciens* neuraminidase (which removes terminal sialic acid residues linked α2→3, 6, or 8 to Gal, NeuAc, GalNAc, or GlcNAc [Uchida et al., 1979]) and analyzed by HPLC/ESIMS. The change in the molecular masses was then used to determine the number of sialic acid residues removed from each glycopeptide. As with the earlier HPLC/ESIMS experiments, only 5% of the sample injected actually entered the mass spectrometer and 95% of the sample was recollected for digestion with the next exoglycosidase, β-galactosidase from *Streptococcus pneumoniae*, which removes terminal Gal residues linked β1→4 to GlcNAc or GalNAc (Distler & Jourdian, 1973; Kobata, 1979). After analysis by HPLC/ESIMS, the change in mass from before the β-galactosidase digestion indicated whether the oligosaccharides were biantennary, triantennary, or tetraantennary. For example, a biantennary structure will lose two Gal residues (a mass difference of 324 Da) upon β-galactosidase digestion, whereas a triantennary oligosaccharide will lose three Gal residues (a mass difference of 486 Da). Finally, the glycopeptides were digested with β-*N*-acetylhexosaminidase from chicken liver, which removes terminal GlcNAc residues linked β1→2, 3, or 4 to any sugar residue (Kobata, 1979). These data are summarized in Table 1.

The results of the HPLC/ESIMS analyses before and after digestion of the *N*-linked glycopeptide 257–276 with neuraminidase (shown in Fig. 5A and B, respectively) provided clear evidence that the major species present contained a trisialylated glycan, whereas the minor species contained mono- and disialylated glycans. The mass difference of 875 Da (5,217.3 Da minus 4,342.3 Da) for the major species corresponds to the mass of three sialic acid residues removed from this glycopeptide. After digestion with β-galactosidase, the HPLC/ESIMS data (shown in Fig. 5C) yielded a mass difference corresponding to three Gal residues (486 Da) for the major glycopeptide, and a mass difference corresponding to two or three Gal residues (324 or 486 Da) for the minor components. After digestion with β-*N*-acetylhexosaminidase, the HPLC/ESIMS data yielded a mass difference corresponding to three GlcNAc residues (609 Da) for the major glycopeptide, and a mass difference corresponding to two GlcNAc residues (406 Da) for the minor components, consistent with the original compositions suggested in Table 1. These data indicate that the original compositions from Table 1 are consistent with the presence of triantennary complex carbohydrate structures containing two and three sialic acid residues as well as zero and one fucose residues, and biantennary complex structures containing one and two sialic acid residues attached to Asn^{272}. However, only two Gal and two GlcNAc residues could be removed from the fucosylated triantennary structure. This was also found to be the case for the same glycan attached to Asn^{20}, whereas for the same glycan attached to Asn^{84}, the β-*N*-acetylhexosaminidase could only remove one GlcNAc. It is possible that the fucose on these glycans is attached to a Gal or GlcNAc on one of the three branches of these structures, because β-galactosidase is known not to cleave Gal residues from branches that are substituted with fucose on the terminal Gal or an attached GlcNAc (Kobata, 1979). Similar exoglycosidase experiments performed on the other three glycopeptides established the proposed structures listed in Table 1. Glycopeptide (81–85) contains primarily a trisialo triantennary structure attached to Asn^{84}, with minor fucosylated trisialo tri-

antennary and disialo triantennary species as well. A similar group of structures (trisialo triantennary [major], trisialo tetraantennary, fucosylated trisialo triantennary, disialo triantennary, and disialo biantennary) were identified on glycopeptide 16-39. The fourth and largest *N*-linked glycopeptide (363-399) was the only one of the four that was found to contain only biantennary structures. The major component was a disialo biantennary oligosaccharide, and the only other structure identified was a monosialo biantennary oligosaccharide. As shown in Table 1, we also experienced proteolytic cleavage of this peptide from digestion with the β-galactosidase. Other experiments with this enzyme have indicated that the β-galactosidase preparation contains some contaminating protease(s) that hydrolyze glycopeptides containing more than 25 amino acids (C.A. Settineri, unpubl. data). In addition, the exoglycosidase experiments enabled us to determine that the LCAT tryptic digest also contained *N*-linked glycopeptide components from a second glycoprotein, Apo D, which was copurified with the LCAT glycoprotein at a significantly lower quantity, clearly indicating the power of this technique for identifying and sequencing oligosaccharides on glycopeptides. These structures were established to be primarily trisialo triantennary oligosaccharides at Asn^{45} and fucosylated disialo biantennary oligosaccharides at Asn^{78}.

Characterization of O-linked carbohydrate structures from LCAT

The carbohydrate composition of the major *O*-linked glycopeptide (two hexoses, two *N*-acetylhexosamines, and two sialic acids) deduced from the Edman and molecular weight data was further investigated in order to determine the exact structures of the carbohydrates at Thr^{407} and Ser^{409}. Based on the previous structures of *O*-linked glycopeptides identified on human proteins (Conradt et al., 1989; Clogston et al., 1993; Nimtz et al., 1993), combined with the proposed composition described above, one would predict that both sites would contain sialyl hexosyl *N*-acetylhexosamine moieties. In order to first determine whether each site contained a single hexosyl *N*-acetylhexosamine disaccharide, the *O*-linked glycopeptide shown in Figure 6A was digested with neuraminidase (*A. ureafaciens*) to remove terminal sialic acid residues from the glycopeptide (Uchida et al., 1979), followed by endo-α-*N*-acetylgalactosaminidase (*Diplococcus pneumoniae*, *O*-glycanase™). The latter enzyme specifically cleaves galactoseβ1→3*N*-acetylgalactosamine (Galβ1→3GalNAc)-Ser/Thr structures between Ser/Thr and *N*-acetylgalactosamine (GalNAc) residues (Umemoto et al., 1977), resulting in the removal of these very specific *O*-linked glycans from glycopeptides or glycoproteins. The resulting digest was analyzed by HPLC/ESIMS, yielding the spectrum shown in Figure 6B. The molecular weight of the major glycopeptide component shifted from 2,950.9 Da in Figure 6A to 1,638.9 Da in Figure 6B. These data indicate that the *O*-linked glycosylation was completely removed from the glycopeptide, leading to the conclusion that Thr^{407} and Ser^{409} were each glycosylated with a Galβ1→3GalNAc structure. Based on the molecular weight data summarized in Table 2, the major *O*-linked glycopeptide species very likely contains primarily one sialic acid residue attached to each Galβ1→3GalNAc structure, and the other glycoforms—such as the component with a molecular weight of 3,551.6 and a composition of two hexoses, two *N*-acetylhexasamines, and four *N*-acetylneuraminic acids ($Hex_2HexNAc_2NeuAc_4$)—must represent two

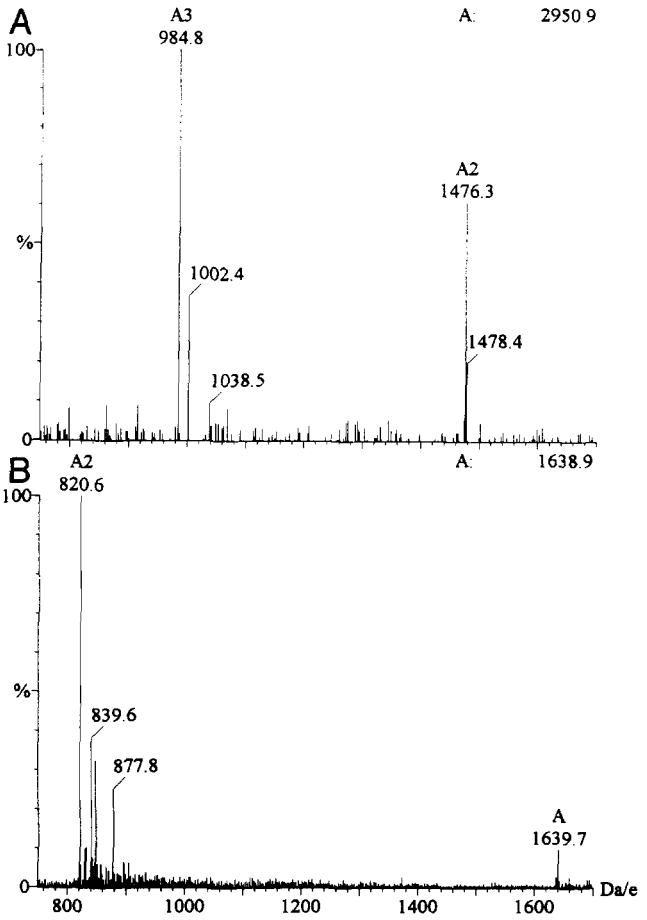

Fig. 6. A: Electrospray mass spectrum of LCAT C-terminal tryptic glycopeptide (fraction 3 from Fig. 2) containing primarily residues 400-416 (pEGPPASPTASPEPPPPE, MW_{avg} = 1,638.8 Da), where the N-terminal amino acid, Gln, has been cyclized to a pyroglutamic acid (pE, a decrease in mass of 17 Da). The molecular mass of the glycopeptide identified in this spectrum (component A, 2,950.9 Da) corresponds to this peptide with an oligosaccharide composition of $HexNAc_2Hex_2NeuAc_2$. **B:** Electrospray mass spectrum of the glycopeptide mixture in A after sequential digestion with *A. ureafaciens* neuraminidase followed by *D. pneumoniae* endo-α-*N*-acetylgalactosaminidase.

Galβ1→3GalNAc structures each bearing two terminal *N*-acetylneuraminic acid (sialic acid) residues. In addition, the presence of one very small peak at 25-30 min in the SIM chromatogram at *m/z* 366, combined with the presence of two larger peaks at 25-30 min in the SIM chromatogram at *m/z* 657 shown in Figure 3, led us to conclude that the major glycoform contains one sialic acid residue on each Galβ1→3GalNAc structure, rather than two sialic acid residues on one and zero on another. We were unable to determine whether the minor *O*-linked glycopeptides containing compositions of $Hex_3HexNAc_3NeuAc_{3-5}$ were glycosylated at all three potential sites or only at Thr^{407} and Ser^{409} (with larger structures), due to limited quantities of sample.

Characterization of N-linked carbohydrate structures from Apo D

Exoglycosidase sequencing combined with HPLC/ESIMS of the Apo D glycopeptides from peaks 4, 5, and 6 in Figure 2 indi-

cated that Asn[45] contains primarily trisialo triantennary oligosaccharides and Asn[78] contains primarily fucosylated disialo biantennary oligosaccharides. These data are summarized in Table 3.

Finally, the calculated average molecular weight of LCAT (59,187 Da) based on the combined major carbohydrate structures (12,105 Da) and the amino acid sequence (47,082 Da) is in good agreement with that previously estimated by sedimentation equilibrium data to be approximately 59,000 (Chung et al., 1979) to 60,000 Da (Chong et al., 1983).

Discussion

In this study, HPLC/ESIMS, Edman degradation, and glycosidase digestion were used to characterize the class of glycosylation structures at each *N*-glycosylation site of human LCAT, to assess the glycoforms present at each site, and to partially elucidate the structures of the oligosaccharides present at each site. In addition, this powerful methodology revealed the presence of unanticipated di-*O*-glycosylation on the C-terminal tryptic peptide as well as the presence of another glycoprotein, Apo D, at a significantly lower sample quantity. The results presented here are the first reported application of this method to a natural human glycoprotein bearing unknown glycosylation. The key factor that facilitated obtaining the complete characterization of the glycosylation on this protein was the use of a mass spectrometric-based strategy (HPLC/ESI/CID/MS) for selective identification of glycopeptides in a complex mixture of peptides and glycopeptides. An ESI mass spectrometer was used as a selective detector of glycopeptides by monitoring for diagnostic carbohydrate fragment ions generated by CID in the electrospray ion source. Our version of this technique (LC/ESI/CID/MS) differs slightly from those of other investigators (Duffin et al., 1992; Carr et al., 1993; Huddleston et al., 1993). For example, in the technique used by Carr et al., the orifice voltage was ramped from 120 V at *m/z* 150 to 65 V at *m/z* 500, and then it was held constant at 65 V from *m/z* 500 to 2,000. Our version, where two separate experiments are performed with two different but constant orifice voltages, has several advantages. The first advantage is the very specific and sensitive detection of glycopeptides because only carbohydrate marker ions are monitored in the SIM experiments. The SIM analysis also yields a high signal-to-noise ratio, which enables very low (subpicomolar) detection limits. However, the primary disadvantage of our technique is that a second HPLC/ESIMS analysis of the proteolytic digest, where the entire mass range (e.g., *m/z* 350–2,000) is scanned, is necessary in order to obtain the molecular weight information on the glycopeptides identified in the SIM analysis. This requires high chromatographic reproducibility, because it is necessary to compare the two UV chromatograms to match the peaks from the SIM trace with the peaks from the HPLC/ESIMS trace where the entire mass range was scanned.

The ability to use an ESI mass spectrometer as a selective detector to monitor for diagnostic fragment ions that are generated by the dissociation of molecular parent ions in the electrospray source has already been expanded to probe the post-translational modifications of parent ions other than glycopeptides, such as the selective detection of phosphopeptides (Huddleston et al., 1993), *S*-β-4-pyridylethyl-cysteine-containing peptides, and xenobiotic modified peptides (A. Ding, P.A.

Table 3. *HPLC/electrospray mass spectrometry data obtained from sequential glycosidase digestion of Apo D N-linked glycopeptides*[a]

Fraction	Glycosylation site and sequence range	Intact glycopeptide MW (Da)	Proposed carbohydrate composition	MW after neuraminidase digestion (Da)	MW after β-galactosidase digestion	Confirmed carbohydrate structure
4	Asn^{45} (41-53)	3,734.9	$Hex_5HexNAc_4NeuAc_2$	3,167.9	2,843.9	Disialo biantennary
		4,100.3	$Hex_6HexNAc_5NeuAc_2$	3,533.1	3,047.1	Disialo triantennary
		4,390.0	**$Hex_6HexNAc_5NeuAc_3$**	**3,533.1**	**3,047.1**	**Trisialo triantennary**
		4,682.3	$Hex_6HexNAc_5NeuAc_4$	3,533.1	3,047.1	Tetrasialo triantennary
		4,756.0	$Hex_7HexNAc_6NeuAc_3$	3,899.2	3,413.1	Trisialo triantennary Lac_1
		5,044.4	$Hex_7HexNAc_6NeuAc_4$	3,899.2	3,413.1	Tetrasialo triantennary Lac_1
5	Asn^{45} (41-55)	3,978.6	$Hex_5HexNAc_4NeuAc_2$	3,409.8	3,086.4	Disialo biantennary
		4,268.6	$Hex_5HexNAc_4NeuAc_3$	3,409.8	3,086.4	Disialo triantennary
		4,632.5	**$Hex_6HexNAc_5NeuAc_3$**	**3,774.9**	**3,287.9**	**Trisialo triantennary**
		4,924.8	$Hex_6HexNAc_5NeuAc_4$	3,774.9	3,287.9	Tetrasialo triantennary
				4,140.4	3,655.7	Triantennary Lac_1
				4,509.5	4,025.1	Triantennary Lac_2
6	Asn^{78} (63-88)	4,783.3	$Hex_5HexNAc_4NeuAc_1Fuc_1$	4,491.9	4,170.1	Fucose monosialo biantennary
		4,927.8	$Hex_5HexNAc_4NeuAc_2$	4,346.9	4,022.9	Disialo biantennary
		5,075.0	**$Hex_5HexNAc_4NeuAc_2Fuc_1$**	**4,491.9**	**4,170.1**	**Fucose disialo biantennary**
		5,443.2	$Hex_6HexNAc_5NeuAc_2Fuc_1$	4,858.2	4,372.2	Fucose disialo triantennary
		5,731.0	$Hex_6HexNAc_5NeuAc_3Fuc_1$	4,858.2	4,372.2	Fucose trisialo triantennary
		6,099.5	$Hex_7HexNAc_6NeuAc_3Fuc_1$	5,222.6	4,736.6	Fucose trisialo triantennary Lac_1
				5,589.6	4,936.3	Fucose tetraantennary Lac_1

[a] Bold type indicates the major species identified based on the relative abundance of the molecular ions in the mass spectra. Lac = Gal-GlcNac, indicating an additional lactosamine group extended on one of the branches.

Schindler, K.F. Medzihradszky, & A.L. Burlingame, unpubl. data). This method should rapidly become a widespread tool for the characterization of posttranslational and xenobiotic modifications, because of its speed, sensitivity, and selectivity compared to classical methods (Wold & Moldave, 1984; Biemann & Martin, 1987; Carr et al., 1991; Burlingame et al., 1992).

To obtain detailed structural information on the LCAT oligosaccharides, we used a method developed in our laboratory that involved specific degradation of the *N*-linked oligosaccharides by sequentially removing the nonreducing terminal monosaccharides from each branch of the structures using exoglycosidases, with analysis by HPLC/ESIMS after each digestion step (Settineri & Burlingame, 1993, 1994). The use of exoglycosidases for the structural analysis of oligosaccharides is not new (Li & Li, 1973; Kobata, 1984), however this method had previously been performed on free oligosaccharides and not glycopeptides, and on a larger scale (nanomolar versus picomolar). Recently, by using matrix-assisted laser desorption time-of-flight mass spectrometry to monitor such sequential digestions on *N*-linked glycopeptides, analysis at the low picomole range was attained (Sutton et al., 1993, 1994). In the latter method, all digestions were performed at the same time in one tube in 1-2-μL volumes, and different combinations of enzymes were used in separate digestions to determine the oligosaccharide sequences. Then, the whole mixture was analyzed for each separate digestion by MALDI-TOF MS. Therefore, sample losses were minimized; however, it was necessary to use experimental conditions that would suit all the different exoglycosidases. In addition, contaminants from the enzyme preparations as well as any other peptides that often coelute with the glycopeptides could interfere with the direct analysis of the mixtures by this method. The method described in this paper also requires only picomole quantities of sample and little sample handling, and does not require experimental conditions that suit all the different exoglycosidases. Using the specificity of the exoglycosidases, combined with the change in mass of the glycopeptides before and after each digestion, we could obtain sequence and branching information on the oligosaccharide portion of the glycopeptides. Of course, the amount of detailed structural information obtained is dependent on the specificity of the glycosidases used. The manufacturers' published specificity of the exoglycosidases is generally determined using free oligosaccharides. We verified that the enzymes retained the same cleavage specificity on oligosaccharides attached to peptides using glycopeptides of known sequence from bovine fetuin (Green et al., 1988; Cumming et al., 1989). We found that at certain (high) concentrations, the enzymes often did not exhibit their expected specificities, and some, such as the β-galactosidase from *S. pneumoniae*, exhibited peptidase activity on some peptides as well. However, it is still possible that a previously unknown structure may be cleaved by one of the enzymes, so the structures proposed can only be based on the fact that the enzymes will cleave only the monosaccharides that they have been shown to cleave. Nevertheless, this method of sequencing glycopeptides using exoglycosidases followed by MALDI-TOF MS or ESIMS is much more sensitive than the classical strategy of removing oligosaccharides from the glycopeptides or glycoprotein, followed by derivatization such as permethylation or peracetylation and analysis by FAB-MS (Carr et al., 1989; Dell, 1990).

A very specific endoglycosidase was used to characterize the two *O*-linked oligosaccharides identified on the proline-rich C-terminal peptide of LCAT, each of which contained sialylated Galβ1→3GalNAc structures. Their existence was not suspected before this work because they are only responsible for a small mass increment compared to that for the *N*-linked carbohydrates, and previous Edman sequencing of this protein did not identify these sites (Yang et al., 1987). The two sites are localized in the proline-rich (8 prolines of 17 residues in the tryptic peptide) C-terminus of LCAT. Although there is no consensus sequence for *O*-linked glycosylation, and therefore no concrete method for predicting sites of *O*-linked glycosylation, a statistical study of the amino acid sequences around all known *O*-linked glycosylation sites (Wilson et al., 1991) indicated that the most prominent feature in the vicinity of these sites is the increased frequency of proline residues. Furthermore, Wilson and coworkers found that in areas of multiple glycosylation sites, prolines are found most frequently in positions −1 and +3 amino acid residues relative to the glycosylation sites. Interestingly, these findings fit well with the sites of *O*-glycosylation found on LCAT (see Fig. 1). Edman degradation showed that only two of the three potential glycosylation sites (Thr^{407} and Ser^{409}, but not Ser^{405}) in the C-terminal tryptic peptide were found to bear carbohydrates. No proline residues are located in position −1 or +3 relative to Ser^{405}, whereas Ser^{409} has a proline located in position +3 and Thr^{407} contains prolines at position −1 as well as +3 to it. The possible biological significance of an *O*-glycosylated C-terminus for LCAT is unclear.

We established that three of the four *N*-linked glycosylation sites (Asn^{20}, Asn^{84}, and Asn^{272}) are glycosylated with the same group of major (trisialo triantennary) and minor (fucosylated trisialo triantennary, disialo triantennary, and disialo biantennary) carbohydrate structures (Asn^{20} contained a very small percentage of trisialo tetraantennary as well). The fourth site, Asn^{384}, contained only biantennary (mono- and disialo) structures. A previous study involving glycosidases and specific inhibitors of glycoprotein assembly with plasma LCAT and cultured cells secreting LCAT activity concluded that LCAT contained both high mannose and complex type *N*-linked oligosaccharides (Collet & Fielding, 1991). Treatment of plasma LCAT with endoglycosidase F (which hydrolyzes the the chitobiose core of high mannose and biantennary complex oligosaccharides [Elder & Alexander, 1982]) or endoglycosidase H (which hydrolyzes the the chitobiose core of high mannose and some hybrid oligosaccharides [Tarentino et al., 1974]) yielded two species on an SDS-polyacrylamide gel. Therefore, the presence of high mannose as well as complex oligosaccharides was concluded. However, we found no evidence of high mannose oligosaccharides. The Endo F would have removed the biantennary complex structures we identified, but the Endo H should not have cleaved any complex oligosaccharides. It is very likely that the Endo H preparation contained some contaminating Endo F activity and therefore cleaved the biantennary structures as well. In fact, it was published subsequent to this study that the source for Endo H also contained three different Endo F isozymes, only one of which contained similar activity to Endo H, whereas the others cleaved all types of *N*-linked oligosaccharides (Trimble & Tarentino, 1991). This clearly shows that one must be cautious when using enzymes isolated from natural sources and low-resolution analytical techniques such as SDS-PAGE, and again demonstrates the power of the mass spectrometric techniques for establishing these structures.

The exact function of the LCAT oligosaccharides is not known. However, site-directed mutagenesis has indicated that removal of the glycosylation site at Asn^{272} converts LCAT to a phospholipase, generating fatty acids rather than cholesteryl esters (Francone et al., 1993). Our results suggest that this effect is mediated by the overall structure of the protein in the area around residue 272 when a carbohydrate structure is present at that site rather than a specific oligosaccharide structure. Therefore, the enzyme substrate specificity may not depend on the presence of a specific carbohydrate structure at Asn^{272}, but may depend on the existence of any hydrophilic moiety at this special site on the protein and formation of the appropriate substrate binding pocket. An earlier study from some of the authors (Collet & Fielding, 1991) supports this hypothesis, because the data indicated that inhibitors of the processing of *N*-linked oligosaccharides during glycoprotein assembly in cultured cells secreting LCAT activity had no effect on the catalytic activity of LCAT. This also supports the hypothesis formulated by Francone and coworkers (1993), which states that the oligosaccharides at position 272 play an indirect role, possibly by aligning the sterol 3-hydroxyl group as the lecithin acyl group is transferred from Ser^{181} in LCAT. It would be interesting to verify what length or size is necessary for the oligosaccharides at position 272 to maintain full enzyme activity and the appropriate specificity. This could be achieved by trimming down this carbohydrate moiety by sequentially digesting LCAT with specific exoglycosidases and recording the enzymatic activity after each step.

Materials and methods

Isolation and purification of human LCAT

LCAT was purified from normal human plasma as described earlier (Collet & Fielding, 1991).

Reduction and alkylation of LCAT

LCAT in 0.1 mM NH_4HCO_3, 6 M guanidine-HCl, pH 8.0, was reduced by a 50-fold excess of DTT at 50 °C for 1.5 h under argon. A 20-fold excess of iodoacetic acid was used for alkylation in the dark at room temperature for 1.5 h. A $CHCl_3/CH_3OH$ (1:1, v/v) extraction was then performed to eliminate lipids from the sample, followed by dialysis overnight against 50 mM NH_4HCO_3, pH 8.0.

Tryptic digestion of LCAT

Reduced and alkylated LCAT was digested with sequencing grade trypsin (Boehringer Mannheim) at an enzyme:substrate ratio of 1:50 (w/w) for 8 h at 37 °C.

Fractionation of peptides and glycopeptides

Microbore reversed-phase HPLC separation was performed using a Carlo Erba Phoenix 20 dual syringe pump system. Separations were accomplished on an Aquapore 300 C-18 column, 1.0 mm i.d. × 250 mm (Applied Biosystems). Eluant A was 0.1% TFA in water, and eluant B was 0.08% TFA in acetonitrile. Water and acetonitrile were purchased from Fisher, and TFA was obtained from Pierce. Column effluent at 50 μL/min flowed directly into a variable-wavelength UV detector (Linear, model 204) equipped with a high-sensitivity U-Z View capillary flow cell (LC Packings). Positioned directly after the detector was an Isco μLC-500 microflow syringe pump, which allowed for postcolumn addition of a 2-methoxyethanol/2-propanol (Aldrich) (1:1, v/v) mixture to the column effluent at flow rates of 35–40 μL/min. Effluent was split approximately 1:19 after the mixing tee so that only 4–5 μL/min went into the mass spectrometer and the remainder was collected manually. The gradient program used to separate the LCAT tryptic digest was a linear gradient of 2–70% B over 120 min followed by 70–98% B over 10 min. The gradient programs used to analyze the glycopeptides after exo- and endoglycosidase digestions were either 2–40% B over 30 min or 10–50% B over 40 min.

Electrospray ionization mass spectrometry

The microbore HPLC system was interfaced to a VG BIO-Q (Fisons/VG) triple quadrupole mass spectrometer equipped with an atmospheric pressure electrostatic spray ion source. The mass spectrometer was scanned in non-continuum mode from *m/z* 330 to *m/z* 2,000 in 5 s. For the SIM experiment, the voltage of the sampling cone was adjusted from 60 to 180 V. Four different ions (*m/z* 204, 292, 366, and 657) were scanned individually over a window of 0.5 Da above and below each of these masses, with a scan time of 0.23 s and an interscan delay of 0.02 s. The accuracy of molecular weight measurements for the glycopeptides was generally 0.03% (e.g., ±0.9 Da for an average molecular mass of 3,000 Da).

Exo- and endoglycosidase digestions

The vacuum-centrifuged fractions containing *N*-linked glycopeptides (200–500 pmol) were solubilized in 10 μL of 50 mM NaOAc, pH 5.0, and 12.5 mU (2.5 μL) of neuraminidase (*A. ureafaciens*; Boehringer Mannheim) was added for a final enzyme concentration of 1.0 mU/μL. The digestion was performed at 37 °C for 5–7 h. After microbore HPLC/ESIMS analysis, the fractions were collected and vacuum dried. These fractions were then solubilized in 10 μL of 50 mM NaOAc, pH 5.0, and digested with 0.8 mU (5 μL) of β-galactosidase (*S. pneumoniae*; Oxford Glycosystems) (0.053 mU/μL final enzyme concentration) at 37 °C for 5–7 h; the samples were then reanalyzed by HPLC/ESIMS. The fractions of interest were again collected and vacuum dried, followed by solubilization in 10 μL of 50 mM NaOAc, pH 5.0, digestion with 12 mU (1.2 μL) of *N*-acetylhexosaminidase (chicken liver; Oxford Glycosystems) (1.1 mU/μL final enzyme concentration) at 37 °C in 50 mM NaOAc, pH 5.0, for another 5–7 h, and reanalysis. The dried *O*-linked glycopeptide fraction (50 pmol) was solubilized in 50 μL of 0.1 M Na_2HPO_4, pH 7.2, and digested with 10 mU (1 μL) neuraminidase (0.2 mU/μL final concentration) for 17 h, followed by 0.5 mU (1 μL) *O*-glycanase (*D. pneumoniae* Endo-α-*N*-acetylgalactosaminidase; Boehringer Mannheim) (0.01 mU/μL final enzyme concentration) for an additional 24 h at 37 °C.

Edman degradation

Microsequencing was performed on a Applied Biosystems 470A gas phase protein sequencer, with an Applied Biosystems 120A phenylthiohydantoin analyzer (Hunkapiller et al., 1983).

Acknowledgments

We thank Frank Masiarz of Chiron Corporation for performing the Edman sequence analysis on the di-*O*-linked glycopeptide and Zhonghua Yu and Corey Schwartz of UCSF for performing all other Edman sequence analysis. This work was supported by grants from the National Institutes of Health, National Center for Research Resources (grant RR01614), the National Institute of Environmental Health Sciences (grant ES04704) (to A.L.B.), and the National Heart, Lung and Blood Institute (Arteriosclerosis SCOR grant HI14237 to C.J.F.).

References

Albers JJ, Cabana VG, Dee Barden Stahl Y. 1976. Purification and characterization of human plasma lecithin: Cholesterol acyltransferase. *Biochemistry 15*:1084–1087.

Aron L, Jones S, Fielding CJ. 1978. Human plasma lecithin-cholesterol acyltransferase. Characterization of cofactor-dependent phospholipase activity. *J Biol Chem 253*:7220–7226.

Bause E, Hettkamp H. 1979. Primary structural requirements for *N*-glycosylation of peptides in rat liver. *FEBS Lett 108*:341–344.

Biemann K, Martin SA. 1987. Mass spectrometric determination of the amino acid sequences of peptides and proteins. *Mass Spectrom Rev 6*:1–76.

Burlingame AL, Baillie TA, Russell DH. 1992. Mass spectrometry. *Anal Chem 64*:467R–502R.

Carr SA, Hemling ME, Bean MF, Roberts GD. 1991. Integration of mass spectrometry in analytical biotechnology. *Anal Chem 63*:2802–2824.

Carr SA, Hemling ME, Folena WG, Sweet RW, Anumula K, Barr JR, Huddleston MJ, Taylor P. 1989. Protein and carbohydrate structural analysis of a recombinant soluble CD4 receptor by mass spectrometry. *J Biol Chem 264*:21286–21295.

Carr SA, Huddleston MJ, Bean MF. 1993. Selective identification and differentiation of *N*- and *O*-linked oligosaccharides in glycoproteins by liquid chromatography–mass spectrometry. *Protein Sci 2*:183–196.

Carr SA, Roberts GD. 1986. Carbohydrate mapping by mass spectrometry: A novel method for identifying attachment sites of Asn-linked sugars in glycoproteins. *Anal Biochem 157*:396–406.

Chait BT, Kent SBH. 1992. Weighing naked proteins: Practical, high-accuracy mass measurement of peptides and proteins. *Science 257*:1885–1894.

Chong KS, Jahani M, Hara S, Lacko AG. 1983. Characterization of lecithin-cholesterol acyltransferase from human plasma. 3. Chemical properties of the enzyme. *Can J Biochem Cell Biol 61*:875–881.

Chung J, Abano DA, Fless GM, Scanu AM. 1979. Isolation, properties, and mechanism of in vitro action of lecithin:cholesterol acyltransferase from human plasma. *J Biol Chem 254*:7456–7464.

Clogston CL, Hu S, Boone TC, Lu HS. 1993. Glycosidase digestion, electrophoresis and chromatographic analysis of recombinant human granulocyte colony-stimulating factor glycoforms produced in Chinese hamster ovary cells. *J Chromatogr 637*:55–62.

Collet X, Fielding CJ. 1991. Effects of inhibitors of *N*-linked oligosaccharide processing on the secretion, stability, and activity of lecithin:cholesterol acyltransferase. *Biochemistry 30*:3228–3234.

Conradt HS, Nimtz M, Dittmar KE, Lindenmaier W, Hoppe J, Hauser H. 1989. Expression of human interleukin-2 in recombinant baby hamster kidney, Ltk-, and Chinese hamster ovary cells. Structure of *O*-linked carbohydrate chains and their location within the polypeptide. *J Biol Chem 264*:17368–17373.

Cumming DA, Hellerqvist CG, Harris-Brandts M, Michnick SW, Carver JP, Bendiak B. 1989. Structures of asparagine-linked oligosaccharides of the glycoprotein fetuin having sialic acid linked to *N*-acetylglucosamine. *Biochemistry 28*:6500–6512.

Dell A. 1990. Preparation and desorption mass spectrometry of permethyl and peracetyl derivatives of oligosaccharides. In: McCloskey JA, ed. *Mass spectrometry*. San Diego: Academic Press. pp 647–660.

Distler JJ, Jourdian GW. 1973. The purification and properties of β-galactosidase from bovine testes. *J Biol Chem 248*:6772–6780.

Duffin KL, Welply JK, Huang E, Henion JD. 1992. Characterization of *N*-linked oligosaccharides by electrospray and tandem mass spectrometry. *Anal Chem 64*:1440–1448.

Elder JH, Alexander S. 1982. Endo-β-*N*-acetylglucosaminidase F: Endoglycosidase from *Flavobacterium meningosepticum* that cleaves both high-mannose and complex glycoproteins. *Proc Natl Acad Sci USA 79*: 4540–4544.

Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM. 1989. Electrospray ionization for mass spectrometry of large biomolecules. *Science 246*:64–71.

Francone OL, Evangelista L, Fielding CJ. 1993. Lecithin-cholesterol acyltransferase: Effects of mutagenesis at *N*-linked oligosaccharide attachment sites on acyl acceptor specificity. *Biochim Biophys Acta 1166*:301–304.

Green ED, Gabriela A, Baenziger J, Wilson S, Van Halbeek H. 1988. The aspargine-linked oligosaccharides on bovine fetuin. *J Biol Chem 263*: 18253–18268.

Hemling ME, Roberts GD, Johnson W, Carr SA, Covey TR. 1990. Analysis of proteins and glycoproteins at the picomole level by on-line coupling of microbore high-performance liquid chromatography with flow fast atom bombardment and electrospray mass spectrometry: A comparative evaluation. *Biomed Environ Mass Spectrom 19*:677–691.

Huberty MC, Vath JE, Yu W, Martin SA. 1993. Site-specific carbohydrate identification in recombinant proteins using MALD-TOF MS. *Anal Chem 65*:2791–2800.

Huddleston MJ, Bean MF, Carr SA. 1993. Collisional fragmentation of glycopeptides by electrospray ionization LC/MS and LC/MS/MS: Methods for selective detection of glycopeptides in protein digests. *Anal Chem 65*:877–884.

Hunkapiller MW, Hewick RM, Dreyer WJ, Hood LE. 1983. High-sensitivity sequencing with a gas-phase sequenator. *Methods Enzymol 91*:399–413.

Karas M, Ingendoh A, Bahr U, Hillenkamp F. 1989. Ultraviolet laser desorption ionization mass spectrometry of femtomolar amounts of large proteins. *Biomed Environ Mass Spectrom 18*:841–843.

Kobata A. 1979. Use of endo- and exoglycosidases for structural studies of glycoconjugates. *Anal Biochem 100*:1–14.

Kobata A. 1984. The carbohydrates of glycoproteins. In: Ginsburg V, Dobbins PW, ed. *Biology of carbohydrates*. New York: Wiley. pp 87–162.

Kornfeld R, Kornfeld S. 1976. Comparative aspects of glycoprotein structure. *Annu Rev Biochem 45*:217–237.

Li YT, Li SC. 1973. α-Mannosidase, β-*N*-acetylhexosaminidase, and β-galactosidase from jack bean meal. *Methods Enzymol 28*:699–713.

Ling V, Guzzetta AW, Canova-Davis E, Stults JT, Hancock WS, Covey T, Shushan B. 1991. Characterization of the tryptic map of recombinant DNA derived tissue plasminogen activator by high-performance liquid chromatography-electrospray ionization mass spectrometry. *Anal Chem 63*:2909–2915.

Loo JA, Udseth HR, Smith RD. 1989. Peptide and protein analysis by electrospray ionization-mass spectrometry and capillary electrophoresis-mass spectrometry. *Anal Biochem 179*:404–412.

McLean J, Fielding C, Drayna D, Dieplinger H, Baer B, Kohr W, Henzel W, Lawn R. 1986. Cloning and expression of human lecithin-cholesterol acyltransferase cDNA. *Proc Natl Acad Sci USA 83*:2335–2339.

Medzihradszky KF, Maltby DA, Hall SC, Settineri CA, Burlingame AL. 1994. Characterization of protein *N*-glycosylation by reversed-phase microbore liquid chromatography/electrospray mass spectrometry, complementary mobile phases and sequential exoglycosidase digestion. *J Am Soc Mass Spectrom 5*:350–358.

Nimtz M, Martin W, Wray V, Kloppel KD, Augustin J, Conradt HS. 1993. Structures of sialylated oligosaccharides of human erythropoietin expressed in recombinant BHK-21 cells. *Eur J Biochem 213*:39–56.

Poulter L, Karrer R, Burlingame AL. 1991. *n*-Alkyl *p*-aminobenzoates as derivatizing agents in the isolation, separation, and characterization of submicrogram quantities of oligosaccharides by liquid secondary ion mass spectrometry. *Anal Biochem 195*:1–13.

Settineri CA, Burlingame AL. 1993. Sequencing of glycopeptides using sequential digestion followed by HPLC/electrospray mass spectrometry. In: *Proceedings of the 41st ASMS Conference on Mass Spectrometry and Allied Topics*. San Francisco: ASMS. pp 94a–94b.

Settineri CA, Burlingame AL. 1994. Strategies for the characterization of carbohydrates from glycoproteins by mass spectrometry. In: Crabb JW, ed. *Techniques in protein chemistry V*. San Diego: Academic Press. pp 97–104.

Siegel MM, Hollander IJ, Hamann PR, James JP, Hinman L, Smith BJ, Farnsworth PH, Phipps A, King DJ, Karas M, Indengoh A, Hillenkamp F. 1991. Matrix-assisted UV-laser desorption/ionization mass spectrometric analysis of monoclonal antibodies for the determination of carbohydrate, conjugated chelator, and conjugated drug content. *Anal Chem 63*:2470–2481.

Smith RD, Loo JA, Edmonds CG, Barinaga CJ, Udseth HR. 1990. New developments in biochemical mass spectrometry: Electrospray ionization. *Anal Chem 62*:882–899.

Subbaiah PV, Albers JJ, Chen CH, Bagdade JD. 1980. Low density lipoprotein-activated lysolecithin acylation by human plasma lecithin-cholesterol acyltransferase. Identity of lysolecithin acyltransferase and lecithin-cholesterol acyltransferase. *J Biol Chem 255*:9275–9280.

Sutton CW, O'Neill JA, Cottrell JS. 1994. Site-specific characterization of glycoprotein carbohydrates by exoglycosidase digestion and laser desorption mass spectrometry. *Anal Biochem 218*:34–46.

Sutton CW, Poole AC, Cottrell JS. 1993. Carbohydrate characterization of a glycoprotein by matrix assisted laser desorption mass spectrometry. In: Angeletti RH, ed. *Techniques in protein chemistry IV*. San Diego: Academic Press. pp 109–116.

Tarentino AL, Plummer TH, Maley F. 1974. The release of intact oligosaccharides from specific glycoproteins by Endo-β-*N*-acetylglucosaminidase H. *J Biol Chem 249*:818–824.

Trimble RB, Tarentino AL. 1991. Identification of distinct endoglycosidase (endo) activities in *Flavobacterium meningosepticum*: Endo F1, Endo F2, and Endo F3. Endo F1 and Endo H hydrolyze only high mannose and hybrid glycans. *J Biol Chem 266*:1646–1651.

Uchida Y, Tsukada Y, Sugimori T. 1979. Enzymatic properties of neuraminidases from *Arthrobacter ureafaciens*. *J Biochem 86*:1573–1585.

Umemoto J, Bhavanandan VP, Davidson EA. 1977. Purification and properties of an endo-α-*N*-acetyl-D-galactosaminidase from *Diplococcus pneumoniae*. *J Biol Chem 252*:8609–8614.

Utermann G, Menzel HJ, Adler G, Dieker P, Weber W. 1980. Substitution in vitro of lecithin-cholesterol acyltransferase. Analysis of changes in plasma lipoproteins. *Eur J Biochem 107*:225–241.

Wilson IBH, Gavel Y, Heijne GV. 1991. Amino acid distributions around *O*-linked glycosylation sites. *Biochem J 275*:529–534.

Wold F, Moldave K, eds. 1984. *Post-translational modifications. Methods in enzymology*. New York: Academic Press.

Yang CY, Manoogian D, Pao Q, Lee FS, Knapp RD, Gotto AM, Pownall HJ. 1987. Lecithin:cholesterol acyltransferase. Functional regions and a structural model of the enzyme. *J Biol Chem 262*:3086–3091.