Identification and characterization of glycosylation sites in human serum clusterin. Protein Sci 6:2120–2123

ARTICLE *in* PROTEIN SCIENCE · OCTOBER 2008

Impact Factor: 2.85 · DOI: 10.1002/pro.5560061007 · Source: PubMed

7 AUTHORS, INCLUDING:

Martin Tenniswood
University at Albany, The State University of …
95 PUBLICATIONS **3,989** CITATIONS
SEE PROFILE

John W Crabb
Cleveland Clinic
183 PUBLICATIONS **8,746** CITATIONS
SEE PROFILE

Protein Science (1997), 6:2120–2133. Cambridge University Press. Printed in the USA.

Identification and characterization of glycosylation sites in human serum clusterin*

JAMES T. KAPRON,[1,3] GEORGE M. HILLIARD,[2] JOHNATHON N. LAKINS,[1]
MARTIN P.R. TENNISWOOD,[1] KAREN A. WEST,[1] STEVEN A. CARR,[2] AND JOHN W. CRABB[1]

[1]W. Alton Jones Cell Science Center, Lake Placid, New York 12946
[2]SmithKline Beecham Pharmaceuticals, King of Prussia, Pennsylvania 19406

(RECEIVED December 26, 1996; ACCEPTED June 12, 1997)

Abstract

Clusterin is a ubiquitous, heterodimeric glycoprotein with multiple possible functions that are likely influenced by glycosylation. Identification of oligosaccharide attachment sites and structural characterization of oligosaccharides in human serum clusterin has been performed by mass spectrometry and Edman degradation. Matrix-assisted laser desorption ionization mass spectrometry revealed two molecular weight species of holoclusterin (58,505 ± 250 and 63,507 ± 200). Mass spectrometry also revealed molecular heterogeneity associated with both the α and β subunits of clusterin, consistent with the presence of multiple glycoforms. The data indicate that clusterin contains 17–27% carbohydrate by weight, the α subunit contains 0–30% carbohydrate and the β subunit contains 27–30% carbohydrate. Liquid chromatography electrospray mass spectrometry with stepped collision energy scanning was used to selectively identify and preparatively fractionate tryptic glycopeptides. Edman sequence analysis was then used to confirm the identities of the glycopeptides and to define the attachment sites within each peptide. A total of six N-linked glycosylation sites were identified, three in the α subunit (α^{64}N, α^{81}N, α^{123}N) and three in the β subunit (β^{64}N, β^{127}N, and β^{147}N). Seven different possible types of oligosaccharide structures were identified by mass including: a monosialobiantennary structure, bisialobiantennary structures without or with one fucose, trisialotriantennary structures without or with one fucose, and possibly a trisialotriantennary structure with two fucose and/or a tetrasialotriantennary structure. Site β^{64}N exhibited the least glycosylation diversity, with two detected types of oligosaccharides, and site β^{147}N exhibited the greatest diversity, with five or six detected types of oligosaccharides. Overall, the most abundant glycoforms detected were bisialobiantennary without fucose and the least abundant were monosialobiantennary, trisialotriantennary with two fucose and/or tetrasialotriantennary. Clusterin peptides accounting for 99% of the primary structure were identified from analysis of the isolated α and β subunits, including all Ser- and Thr-containing peptides. No evidence was found for the presence of O-linked or sulfated oligosaccharides. The results provide a molecular basis for developing a better understanding of clusterin structure–function relationships and the role clusterin glycosylation plays in physiological function.

Keywords: amino acid sequencing; carbohydrate; clusterin; glycoprotein; glycosylation; high performance liquid chromatography; mass spectrometry; N-linked oligosaccharides

Clusterin is a secreted glycoprotein first described based upon the capacity to elicit the clustering of cells (Blaschuk et al., 1983) and is known by several acronyms, including TRPM-2, apolipoprotein J, serum protein-40,40, sulfated glycoprotein 2, and complement-lysis inhibitor. The protein has been implicated in an array of physiological roles, including lipid transport (de Silva et al., 1990b), inhibition of complement attack (Jenne & Tschopp, 1989), sperm maturation (Sylvester et al., 1984), and membrane re-modeling during apoptosis (Wilson et al., 1995). A commonality among the multiple possible functions may be a role in cholesterol uptake and recycling by membranes during times of cellular stress and change (Rosenberg & Silkensen, 1995; Wilson et al., 1995). The protein exhibits ubiquitous expression in normal vertebrate tissues and elevated expression in a variety of pathological conditions (Wong

Reprint requests to: John W. Crabb, W. Alton Jones Cell Science Center, Lake Placid, New York 12946; e-mail: jcrabb@northnet.org.

*A preliminary report of this work was presented at the Ninth Symposium of the Protein Society, Boston, Massachusetts, July 1995: Hilliard GM, Kapron JT, West K, Lakins J, Tenniswood M, Carr SA, Crabb JW. 1995. Characterization of carbohydrate on human clusterin. *Protein Sci* 4:183 (Suppl. 2, Abstract 687T).

[3]Present address: Advanced Bioanalytical Services, Inc., 15 Catherwood Road, Ithaca, New York 14850.

Abbreviations: CAM, carboxyamidomethyl; CHO, carbohydrate oligosaccharide; CNBR, cyanogen bromide; EDTA, ethylenediamine tetraacetic acid; LC-ESMS, liquid chromatography electrospray mass spectrometry; MALDI-TOF MS, matrix-assisted laser desorption ionization time of flight mass spectrometry; PBS, phosphate buffered saline; PBSE, PBS containing 5 mM EDTA; PMSF, phenylmethylsulfonyl fluoride; PTH, phenylthiohydantoin; RP-HPLC, reverse phase-HPLC; rt, retention time.

et al., 1994). Glycosylation likely plays an important role in the diverse apparent physiological roles of clusterin, yet this structural property remains poorly defined (Wagner et al., 1995). As part of on-going studies to decipher clusterin structure–function relationships, we have identified the carbohydrate attachment sites and characterized the oligosaccharide structures in human serum clusterin.

Previous studies of human serum clusterin have defined the protein and gene sequences and revealed several posttranslational processing events (Jenne & Tschopp, 1989; Kirszbaum et al., 1989, 1992; de Silva et al., 1990a, 1990b; Choi-Miura et al., 1992; Wong et al., 1994). Species homologues display remarkable conservation of amino acid sequence and genomic organization, however, no significant homology has yet been detected with other known structures (Wong et al., 1993). The protein is synthesized as a single polypeptide precursor that undergoes limited proteolysis to remove a leader peptide and to produce the α and β subunits, which are linked by four or five disulfide bonds (Choi-Miura et al., 1992). The protein mass calculated from amino acid residues alone is 50.1 kDa, yet by SDS-PAGE, holoclusterin exhibits an apparent molecular weight of 70–80 kDa and the α and β subunits appear to be 35–40 kDa. Clusterin glycosylation has long been known to contribute to the observed mass variations (Blaschuk & Fritz, 1984), and previous studies using lectin binding, susceptibility to glycosidases, and chemical analysis methods suggest that the protein contains sugar of the complex form, constituting 20–30% of the molecular weight (Blaschuk et al., 1983; Griswold et al., 1986; Burkey et al., 1991). Based on sequence motifs, various clusterin species may contain both N-linked sugar as well as O-linked sugar; for example, human clusterin contains six N-linked sequons and four O-linked sites predicted by the NetOglyc computer program (Hansen et al., 1995). Previous Edman sequence data suggest that human serum clusterin contains six sites of N-linked carbohydrate and no O-linked sugar (Kirszbaum et al., 1992). To clarify ambiguities regarding clusterin glycosylation and to establish a foundation for relating oligosaccharide content with clusterin function, we have used mass spectrometry and Edman degradation to define the carbohydrate covalently bound to human serum clusterin. This study represents the most rigorous molecular weight determination and detailed glycosylation site analysis of this intriguing glycoprotein to date.

Results

Molecular weight determination of holoclusterin and the α and β subunits

Molecular weight determination, amino acid analysis, and NH_2-terminal sequence analysis of human serum clusterin and the intact clusterin subunits were performed with RP-HPLC purified protein preparations. The RP-HPLC profiles in Figure 1 provide an indication of the purity of the immunoaffinity-purified holoclusterin and the chromatography yielded desalted protein suitable for MALDI-TOF mass spectrometric analysis. RP-HPLC of clusterin after reduction and S-carboxyamidomethylation allowed isolation of the intact, CAM-modified α and β subunits (Fig. 1B). Automated Edman sequence analysis of the purified holoprotein (about 10 pmol analyzed) revealed two sequences (DQTVS... and SLMPF...) consistent with the known human clusterin subunit sequences (Wong et al., 1994). Edman analysis of the purified

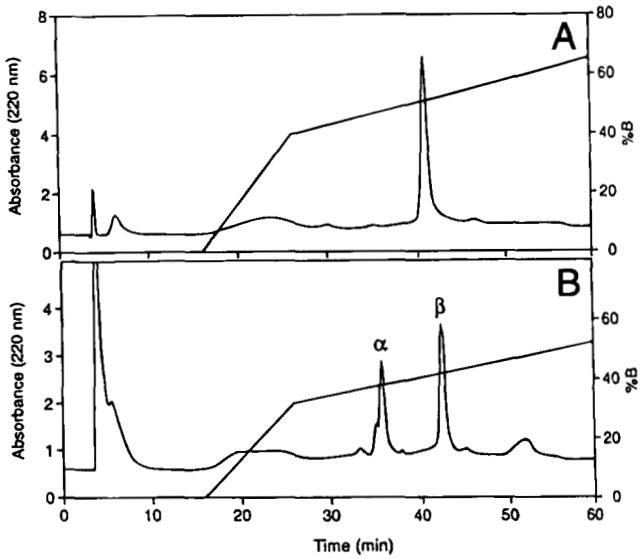


Fig. 1. RP-HPLC of holoclusterin (**A**) before and (**B**) after reduction and S-carboxyamidomethylation. Approximately 15 μg of holoclusterin per chromatography was injected onto a 5-μ Vydac C18 column (1 × 250 mm) at 0% B. Chromatography was performed with an Applied Biosystems model 120 HPLC system at a flow rate of 50 μL/min using the gradients indicated; 1-min fractions were collected automatically and absorbance monitored at 220 nm. Solvent A was 0.1% trifluoroacetic acid in H_2O and solvent B was 84% acetonitrile containing about 0.07% trifluoroacetic acid. The identity of the separated α and β clusterin subunits is indicated in panel B.

CAM subunits (5–38 pmol analyzed) revealed single sequences for each isolated protein and identified the earlier eluting component (Fig. 1B) as the α subunit (NH_2-terminal sequence DQTVSDNELQ...) and the later eluting component as the β subunit (NH_2-terminal sequence SLMPFSPYEP...). Amino acid analysis of holoclusterin and the purified subunits (Table 1) yielded compositional data in good agreement with the known compositions. The results from RP-HPLC, amino acid, and sequence analysis provide evidence of the quality of protein used for subsequent molecular weight determinations.

Previous SDS-PAGE-measured mass values (70–80 kDa for holoclusterin and 35–40 kDa for the subunits) have lead to tenuous estimates of the amount of carbohydrate associated with clusterin. The molecular weights of human serum clusterin and the isolated CAM subunits were determined by MALDI-TOF MS as shown in Figure 2 (mass values are the average of five measurements). For the protein preparation analyzed (holoclusterin preparation 2), MALDI-TOF MS revealed two molecular species of holoclusterin (58,505 ± 250 and 63,507 ± 200), three species of the CAM α subunit (24,196 ± 50; 26,464 ± 10; and 31,351 ± 25), and one species of the CAM β subunit (32,823 ± 20). The lower m/z species evident in Figure 2 represent multiply charged species. Weak ionization of the CAM β subunit was observed in LC-ESMS, nevertheless, the electrospray data suggest the existence of several molecular species, including but not limited to 32,787 ± 4, 33,157 ± 7, 33,443 ± 9, and 33,596 ± 4. Typical for glycoproteins, no useful molecular weight information was obtained by electrospray mass spectrometry for either holoclusterin or the CAM α subunit. Calculated from the gene sequence, the molecular weight of human holoclusterin is 50,071, the α subunit is 24,198, and the β subunit is 25,883. Assuming that the differences between the

Table 1. *Amino acid analysis of clusterin*[a]

	Residues/molecule					
	Holoclusterin		α Subunit		β Subunit	
Amino acid	Known[b]	Experimental	Known[b]	Experimental	Known[b]	Experimental
Asp/Asn (D/N)	46	47.7	23	21.0	23	22.7
Glu/Gln (E/Q)	68	63.8	34	31.4	34	31.4
Ser (S)	33	35.9	15	15.6	18	17.9
Gly (G)	11	20.7	7	10.0	4	13.5
His (H)	13	11.6	5	5.6	8	8.0
Arg (R)	29	25.2	14	12.4	15	14.4
Thr (T)	25	26.2	11	11.8	14	14.3
Ala (A)	16	19.6	6	7.7	10	12.6
Pro (P)	20	25.3	8	9.3	12	13.6
Tyr (Y)	11	12.1	5	4.7	6	5.8
Val (V)	25	24.2	9	9.3	16	17.2
Met (M)	14	15.0	7	6.3	7	3.3
Ile (I)	13	10.3	7	5.4	6	5.2
Leu (L)	39	42.1	19	22.0	20	23.6
Phe (F)	22	20.0	12	12.2	10	9.7
Lys (K)	28	27.8	16	15.8	12	13.3
Amount analyzed (μg)		0.8		0.8		0.2
Average compositional% error		10.3		11.5		11.4

[a]Phenylthiocarbamyl amino acid analysis with automatic vapor phase hydrolysis was performed on HPLC purified clusterin samples according to West and Crabb (1992). Average compositional % error was calculated using all residues except Gly, Trp, and Cys (West & Crabb, 1992).
[b]Wong et al. (1994).

calculated and observed mass values are due to glycosylation, the data indicate (for this protein preparation) that the α subunit contains 0–30% carbohydrate, the β subunit 27–30% carbohydrate, and holoclusterin 17–27% carbohydrate by weight. The observed mass variation is due to variable oligosaccharide content as shown below.

Identification of clusterin glycosylation sites

To identify glycosylation sites, human serum clusterin was reduced, S-carboxyamidomethylated, and tryptic digests of the intact CAM protein or the isolated CAM subunits analyzed by LC-ESMS. LC-ESMS data are presented in Figure 3 from tryptic digests of the individual α and β subunits. Additional LC-ESMS data from tryptic digests of two different holoclusterin preparations (Suppl. Figs. 1 and 2, respectively) are available in the Electronic Appendix. The similarity between RP-HPLC tryptic peptide maps recorded by ultraviolet absorbance and total ion current is shown in Supplementary Figure 1. Stepped collision energy scanning was used during all the on-line LC-ESMS analyses of clusterin to produce carbohydrate-specific marker ions as described in Materials and methods (Carr et al., 1993; Huddleston et al., 1993; Bean et al., 1995). The locations in the chromatograms where these ions maximize (usually two or more in concert) indicate a glycopeptide-containing fraction (see Fig. 3). Inspection of the underlying full-scan mass spectra provided the molecular mass of the glycopeptide and of the attached carbohydrate.

Select LC-ESMS fractions from Figure 3 and Supplementary Figure 1 were subjected to Edman degradation to determine NH_2-terminal peptide sequences; sequence data supporting the position of each glycosylation site is presented in Table 2. A total of six N-linked glycosylation sites were identified in clusterin from analysis of the isolated subunits (Fig. 3). Three N-linked sites were identified in the α subunit, namely ^{64}N, ^{81}N, and ^{123}N (Fig. 3B) and three sites were identified in the β subunit, namely, ^{64}N, ^{127}N, and ^{147}N (Fig. 3D). Peptides identified from analysis of the individual α and β subunits (Fig. 3) are listed in Tables 3 and 4, respectively. An overview of the identified clusterin glycosylation sites is presented within the context of the complete α and β subunit primary structures (Fig. 4). Unlabeled minor peaks in the "carbohydrate-specific" ion current traces include glycopeptides formed by partial tryptic and chymotryptic proteolysis fragments. Glycosylated peptides containing α subunit site ^{64}N were not found in holoprotein preparation one (Suppl. Fig. 1) and glycosylated peptides containing α subunit sites ^{64}N and ^{123}N were not found in holoprotein preparation two (Suppl. Fig. 2). Peptides identified from analysis of the two holoclusterin preparations are listed in Supplementary Tables 1 and 2 in the Electronic Appendix. The amount of protein sequence identified from each LC-ESMS analysis ranged from 82 to 100% for the α subunit and 82 to 98% for the β subunit. Overall, 99% of the clusterin primary structure was identified from the LC-ESMS analyses of the purified subunits (Fig. 3; Tables 3, 4); only β subunit residues 60–62 (EIR) and residues 78–79 (CR) were not found (Table 4) and these peptides do not contain potential glycosylation sites. The complete clusterin sequence was identified from the combined four LC-ESMS analyses. All Ser and Thr residues can be accounted for in the identified peptides and no O-linked sugar was detected in any of the analyses.

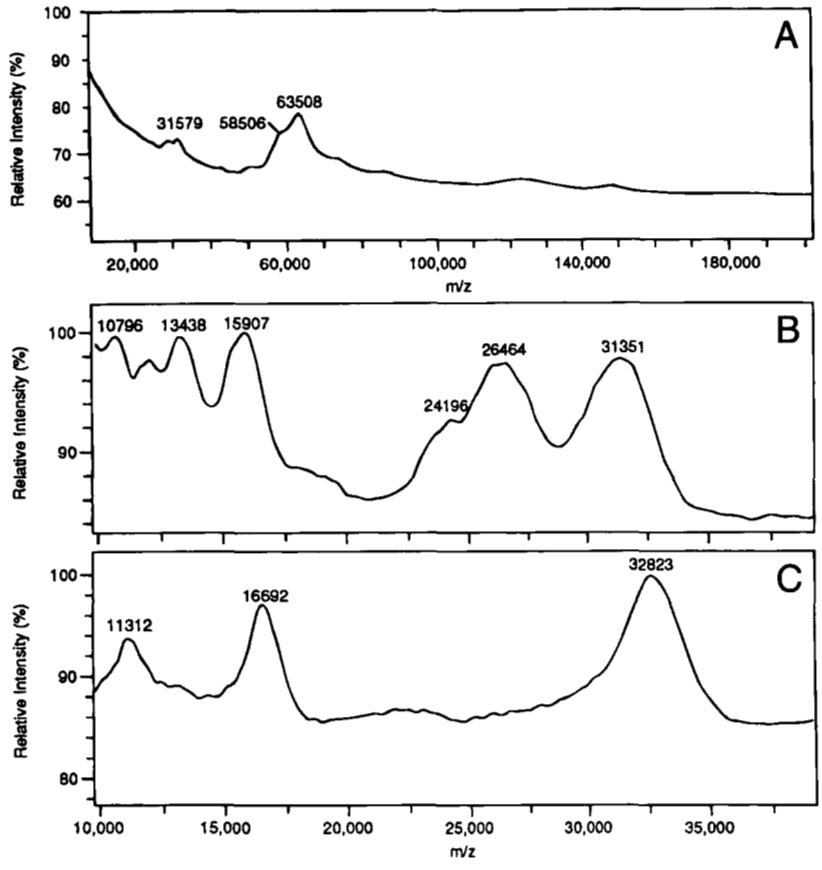


Fig. 2. MALDI-TOF mass spectra of holoclusterin and the α and β subunits. Holoclusterin and the CAM-modified α and β subunits were purified by RP-HPLC as shown in Figure 1 and about 10 pmol of each of the intact proteins analyzed using a VG TofSpec SE mass spectrometer. **A:** Holoclusterin masses of about 58.5 and 63.5 kDa are evident along with doubly charged species. **B:** CAM α-subunit, masses of about 24.2, 26.4, and 31.3 kDa are evident along with doubly and triply charged species. **C:** CAM β-subunit, a mass of about 32.8 kDa, is evident with doubly and triply charged species.

Table 2. *Edman sequence analysis of clusterin glycopeptides*[a]

LC-ESMS fraction (retention time)	Sequence determined	Amount analyzed (pmol)	Subunit (residues)	Glycosylation site
33 min (Fig. 3A,B)	KkEDALXEtr	19	α (58–67)	α^{64}N
	EDALXETR	4	α (60–67)	—
	YVNK	5	α (19–22)	—
	TLLsN...	4	α (47–51)	—
66 min (Suppl. Fig. 1)	LKELPgVcXe...	12	α (73–82)	α^{81}N
	ELPG...	2	α (75–78)	—
	LFDSD...	12	β (182–186)	—
59 min (Suppl. Fig. 1)	QLEeFLXQ...	4	α (117–124)	α^{123}N
19 min (Suppl. Fig. 1)	HXsTGXLR	7	β (63–70)	β^{64}N
85 min (Suppl. Fig. 1)	MLXTSSLLEQL...	42	β (125–135)	β^{127}N
49 min (Suppl. Fig. 1)	LAXLTQGEDQ...	55	β (145–154)	β^{147}N

[a]Indicated glycopeptide LC-ESMS fractions were analyzed by automated Edman degradation as described in Materials and methods. Determined sequences are listed along with the amount analyzed (based on PTH-amino acid yield), subunit residues, and glycosylation site. X denotes no residue assignment was made and lower-case letters indicate a tentative assignment. Sequence data interpretation was aided by knowledge of the protein sequence (Wong et al., 1994) and the mass spectra for the indicated fractions. Supplementary Figure 1 can be found in the Electronic Appendix. Supplementary Table 1 (a summary of peptides identified in Suppl. Fig. 1) and Spectra supporting Supplementary Table 1, can also be found in the Electronic Appendix.

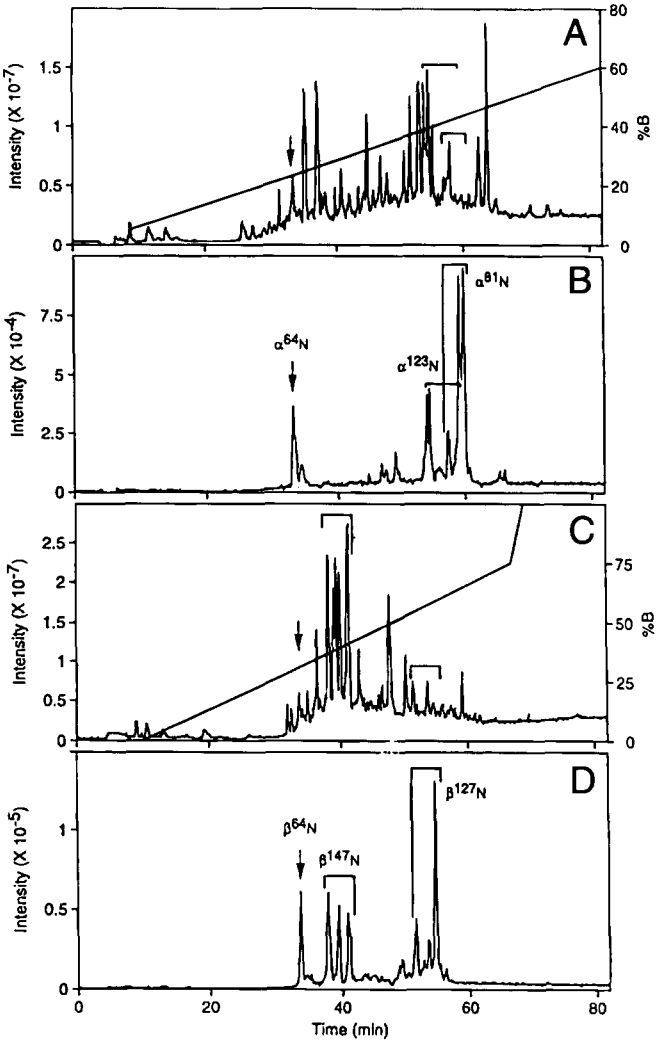


Fig. 3. LC-ESMS of tryptic digests of the clusterin subunits. CAM-modified clusterin subunits were purified as in Figure 1, digested with trypsin, and LC-ESMS performed with a PE Sciex API 300 triple quadrupole mass spectrometer using stepped collision energy scanning as described in Materials and methods. **A:** LC-ESMS of the α subunit tryptic digest (~15 μg). Total ion current (TIC) for the full scan m/z 400–2,400 and the HPLC gradient are shown. **B:** Selective ion monitoring for carbohydrate marker ions m/z 204, 292, and 366 for panel A. Arrows and brackets denote α subunit glycopeptide fractions and N-linked attachment sites. **C:** LC-ESMS of the β subunit tryptic digest (~15 μg). TIC for the full-scan m/z 400–2,400 and the HPLC gradient are indicated. **D:** Extracted ion current traces (summed) for carbohydrate-specific marker ions m/z 204, 292, and 366 for panel C. Arrows and brackets denote β subunit glycopeptide fractions and N-linked attachment sites.

Characterization of clusterin oligosaccharides

The structural class of oligosaccharide at each of the six clusterin glycosylation sites was determined by ESMS as described in Materials and methods. Glycopeptide electrospray mass spectra obtained from LC-ESMS analyses of the purified CAM α and β subunits are presented in Figures 5 and 6, respectively. Supporting glycopeptide mass spectra for two or three additional oligosaccharides identified from the LC-ESMS analyses of holoclusterin are presented in Figure 7. In Figures 5, 6, and 7, the measured masses of the glycopeptides, the calculated masses of the unmodified peptides, the oligosaccharide masses obtained by the difference between the unmodified peptide and glycopeptide masses, and the structural class of the oligosaccharide (arbitrarily categorized types 1–7 by ascending mass) are tabulated. No evidence of sulfated oligosaccharide was found using electrospray mass spectrometric extracted negative ion monitoring for SO_3^- (Bean et al., 1995).

The process used to characterize clusterin oligosaccharides will be illustrated with the two β subunit glycopeptides shown in Figure 6A with determined masses of 3,149.9 and 2,858.6. First, note that the difference in mass between these ions is that of a NeuAc (291.3 observed; 291.3 calculated), suggesting the presence of complex-type carbohydrates terminated in sialic acid. Subtracting the masses of the three β subunit tryptic peptides containing N-linked sequons (i.e., 944.0, 2,409.7, and 1,683.8) from the measured glycopeptide mass yields candidate masses for the attached oligosaccharide. For example, subtracting the above tryptic peptide masses from the glycopeptide of mass 3,149.9 yields candidate oligosaccharide masses of 2,205.9, 740.2, and 1,466.1, respectively. Only the candidate mass of 2,205.9 corresponds closely to a typical mammalian oligosaccharide structure, namely a bisialobiantennary structure (in-chain mass 2,206.0). This was the method of choice for interpretation of mass spectra when the identity of the tryptic peptide was known from sequence analysis (i.e., 944.0 for the above example). However, interpretation of mass spectra often preceded sequence analysis and also involved the following alternative approach. Subtracting the in-chain masses of typical bi-, tri-, and tetraantennary N-linked carbohydrates with and without NeuAc from the determined masses of the glycopeptides yields candidate masses for the peptide portion. For example, by subtracting 2,206.0 [for the bisialobiantennary glycoform (NeuAc-Hex-HexNAc)$_2$Hex$_3$HexNAc$_2$] and 2,862.6 [for the trisialotriantennary glycoform (NeuAc-Hex-HexNAc)$_3$Hex$_3$HexNAc$_2$] from the glycopeptide of mass 3,149.9, candidate masses of 943.9 and 287.3, respectively, are obtained for the peptide portion of the glycopeptide. Only the peptide mass of 943.9 corresponds closely to one of the expected N-linked tryptic glycopeptides, and Edman degradation subsequently confirmed that this peptide encompasses β subunit residues 63–70 (Tables 2, 4). In this example, carbohydrate must be attached at Asn 64 based on (1) the presence of a single N-linked consensus-site; (2) the absence of a PTH amino acid in cycle 2 of Edman degradation; (3) the presence of PTH-Ser and PTH-Thr in Edman cycles 3 and 4 (eliminating O-glycosylation); and (4) despite no identifiable signal in Edman cycle 6, carboxyamidomethyl Cys was present based on the mass of the peptide and the gene predicted sequence. The carbohydrate portion of the glycopeptide with observed mass 3,149.9 has a molecular mass consistent with that of a biantennary oligosaccharide with two NeuAc (i.e., type 2, in-chain mass 2,206.0); the glycopeptide with observed mass 2,858.6 (Fig. 6A) exhibits a mass consistent with a biantennary oligosaccharide with only one NeuAc (i.e., type 1, in-chain mass 1,914.8). Based on ESMS ion intensities, the relative amounts of these two oligosaccharides at site β^{64}N (in this particular clusterin preparation) are about 15% type 1 and 85% type 2, calculated as described in Materials and methods.

Seven different possible complex oligosaccharide structures were identified by mass and the relative amounts at each clusterin glycosylation site estimated by ESMS ion intensity. These results are summarized in Table 5. All attachment sites except β^{64}N were found to exhibit at least three oligosaccharides, namely bisialobiantennary structures without fucose (type 2, mass 2,206.0) and trisialotriantennary structures without fucose (type 4, mass 2,862.6)

Table 3. *Peptides identified by LC-ESMS from the clusterin α subunit*[a]

Residues	Retention time (min)	Calculated mass	Observed mass	Error (Da)	Sequence
1–18	44.7	2,008.9	2,008.4	0.5	DQTVSDNELQEMSNQGSK⟩Y
19–22	25.5	522.3	522.4	0.1	* K⟨YVNK⟩E
23–35	55.9	1,439.8	1,439.5	0.3	KLEIQNAVNGVKQIK⟩T
36–40	37.0	602.4	602.4	0	K⟨TLIEK⟩T
41–45	11.5	647.3	647.4	0.1	* K⟨TNEER⟩K
41–46	13.0	775.4	775.6	0.2	* K⟨TNEERK⟩T
41–56	45.6	1,874.0	1,873.8	0.2	* K⟨TNEERKTLLSNLEEAK⟩K
46–56	52.9	1,244.7	1,244.3	0.4	* R⟨KTLLSNLEEAK⟩K
47–57	53.5	1,244.7	1,244.9	0.2	* K⟨TLLSNLEEAKK⟩K
46–57	51.5	1,372.8	1,372.8	0	* R⟨KTLLSNLEEAKK⟩K
47–56	55.0	1,116.6	1,116.8	0.2	K⟨TLLSNLEEAK⟩K
57–67	33.1	1,331.5av	3,537.8	G	K⟨KKKEDALNETR⟩E
57–67	33.1	1,331.5av	4,193.9	G	K⟨KKKEDALNETR⟩E
57–67	33.1	1,331.5av	4,340.1	G	K⟨KKKEDALNETR⟩E
57–72	34.1	1,906.1av	4,768.5	G	K⟨KKKEDALNETRESETK⟩L
57–72	34.1	1,906.1av	4,914.6	G	K⟨KKKEDALNETRESETK⟩L
58–67	33.1	1,203.3av	3,409.0	G	K⟨KKEDALNETR⟩E
58–67	33.1	1,203.3av	4,065.4	G	K⟨KKEDALNETR⟩E
58–67	33.1	1.203.3av	4,211.8	G	K⟨KKEDALNETR⟩E
58–72	34.1	1,777.9av	3,984.0	G	* K⟨KKEDALNETRESETK⟩L
58–72	34.1	1,777.9av	4,640.1	G	* K⟨KKEDALNETRESETK⟩L
58–72	34.1	1,777.9av	4,786.5	G	* K⟨KKEDALNETRESETK⟩L
60–67	33.1	947.0av	3,956.0	G	K⟨EDALNETR⟩E
68–72	8.7	592.3	592.4	0.1	R⟨ESETK⟩L
73–96	57.3	2,937.5av	5,143.5	G	K⟨LKELPGVCNETMMALWEECKPCLK⟩Q
3–96	57.3	2,937.5av	5,800.0	G	K⟨LKELPGVCNETMMALWEECKPCLK⟩Q
73–96	57.3	2,937.5av	5,946.6	G	K⟨LKELPGVCNETMMALWEECKPCLK⟩Q
97–101	27.0	666.3	666.4	0.1	K⟨QTCMK⟩F
97–105	37.0	1,203.6	1,203.8	0.2	K⟨QTCMKFYAR⟩V
102–105	35.0	555.3	555.4	0.1	* K⟨FYAR⟩V
106–108	14.0	433.2	433.1	0.1	* R⟨VCR⟩S
109–116	35.0	731.4	731.6	0.2	* R⟨SGSGLVGR⟩Q
117–128	54.0	1,438.6av	3,644.8	G	* R⟨QLEEFLNQSSPF⟩Y
117–128	54.0	1,438.6av	4,300.8	G	* R⟨QLEEFLNQSSPF⟩Y
117–128	54.0	1,438.6av	4,446.9	G	* R⟨QLEEFLNQSSPF⟩Y
129–136	54.0	1,087.5	1,087.6	0.1	* F⟨YFWMNGDR⟩I
137–160	57.3	2,929.2av	2,929.6	0.4	R⟨IDSLLENDRQQTHMLDVMQDHFSR⟩A
146–160	52.9	1,871.8	1,871.7	0.1	R⟨QQTHMLDVMQDHFSR⟩A
161–172	63.8	1,392.7	1,393.0	0.3	R⟨ASSIIDELFQDR⟩F
173–176	40.6	569.3	569.6	0.3	R⟨FFTR⟩E
177–192	65.2	1,999.0	1,999.2	0.2	R⟨EPQDTYHYLPFSLPHR⟩R
193–205	73.6	1,686.0	1,686.0	0	R⟨RPHFFFPKSRIVR

[a]Clusterin peptides identified by LS ESMS in Figure 3A and B are listed by subunit, residues, total ion current retention time, calculated and observed masses, and amino acid sequences. Calculated masses are monoisotopic values unless indicated as chemical average masses (av). Error refers to the difference between the observed and calculated masses. G denotes a glycopeptide, an asterisk (*) indicates that the NH_2-terminal assignment is supported by Edman sequence analysis and the symbols (⟨,⟩) denote the proteolytic cleavage sites. Mass spectra for all listed peptides are available in Figure 5 or the Electronic Appendix.

and with fucose (type 5, mass 3,008.8). At site β^{64}N, only a monosialobiantennary structure (type 1, mass 1,914.8) and a bisialobiantennary structure (type 2) were detected. A type 1 structure was also detected at site β^{127}N. Five or six oligosaccharide structures were found at site β^{147}N, including the above three common types, a bisialobiantennary structure with one fucose (type 3, mass 2,352.2) and possibly a trisialotriantennary with two fucose (type 6, mass 3,154.9) that could have a sialyl Lewis type structure for one of the arms of the carbohydrate and/or possibly a tetrasialotriantennary structure (type 7, mass 3,153.9). The current analyses do not allow distinction between type 6 and type 7 oligosaccharides. Overall, the type 2 oligosaccharide structure was found to be most abundant and the type 1, 6, and/or 7 structures the least abundant. Relative amounts of glycoforms in Table 5 are approximations without adjustment for partial glycosylation site occupancy and should be regarded as semiquantitative estimates only, because strict quan-

Table 4. *Peptides identified by LC-ESMS from the clusterin β subunit*[a]

Residues	Retention time (min)	Calculated mass	Observed mass	Error (Da)	Sequence
1–13	53.5	1,540.7	1,541.0	0.3	R⟨SLMPFSPYEPLNF⟩H
14–49	59.0	4,317.9av	4,318.2	0.3	* F⟨HAMFQPFLEMIHEAQQAMDIHFHSPAFQHPPTEFIR⟩E
50–55	10.6	705.3	705.6	0.3	* R⟨EGDDDR⟩T
56–59	19.2	534.3	534.4	0.1	R⟨TVCR⟩E
63–70	33.4	944.0av	2,858.6	G	* R⟨HNSTGCLR⟩M
63–70	33.4	944.av	3,149.9	G	* R⟨HNSTGCLR⟩M
71–77	31.6	923.4	923.6	0.2	* R⟨MKDQCDK⟩C
80–95	38.9	1,761.8	1,761.9	0.1	R⟨EILSVDCSTNNPSQAK⟩L
96–98	32.3	443.3	443.4	0.1	* K⟨LRR⟩E
98–109	57.4	1,443.7	1,443.8	0.1	R⟨RELDESLQVAER⟩L
99–109	39.8	1,287.6	1,288.0	0.4	R⟨ELDESLQVAER⟩L
110–113	38.9	516.3	516.4	0.1	R⟨LTRK⟩Y
114–119	36.2	778.4	778.6	0.2	K⟨YNELLK⟩S
120–124	38.1	710.3	710.4	0.1	* K⟨SYQWK⟩M
125–139	51.0	1,767.0av	3,972.8	G	* K⟨MLNTSSLLEQLNEQF⟩N
125–144	54.4	2,409.7av	4,324.1	G	* K⟨MLNTSSLLEQLNEQFNWVSR⟩L
125–144	54.4	2,409.7av	4,615.5	G	* K⟨MLNTSSLLEQLNEQFNWVSR⟩L
125–144	54.4	2,409.7av	5,271.9	G	* K⟨MLNTSSLLEQLNEQFNWVSR⟩L
145–155	38.0	1,251.3av	34,56.8	G	* R⟨LANLTQGEDQY⟩Y
145–155	38.0	1,251.3av	4,259.8	G	* R⟨LANLTQGEDQY⟩Y
145–156	39.3	1,414.5av	3,620.2	G	* R⟨LANLTQGEDQYY⟩L
145–158	41.1	1,683.8av	3,889.5	G	* R⟨LANLTQGEDQYYLR⟩V
145–158	41.1	1,683.8av	4,035.5	G	* R⟨LANLTQGEDQYYLR⟩V
145–181	57.9	3,980.3av	3,980.4	0.1	R⟨LANLTQGEDQYYLRVTTVASHTSDSDVPSGVTEVVVK⟩L
159–181	41.1	2,314.5av	2,314.2	0.3	R⟨VTTVASHTSDSDVPSGVTEVVVK⟩L
182–198	47.6	1,873.0	1,873.2	0.2	K⟨LFDSDPITVTVPVEVSR⟩K
199–210	38.1	1,420.7	1,421.0	0.3	R⟨KNPKFMETVAEK⟩A
203–210	38.0	953.5	953.6	0.1	* K⟨FMETVAEK⟩A
211–216	39.5	778.4	778.6	0.2	* K⟨ALQEYR⟩A
211–217	39.3	906.6	906.8	0.2	* K⟩⟨ALQEYRK⟩K
218–222	15.9	697.4	697.6	0.2	K⟨KHREE
219–222	13.0	569.3	569.6	0.3	* K⟨HREE

[a]Clusterin peptides identified by LC-ESMS in Figure 3C and D are listed by subunit, residues, total ion current retention time, calculated and observed masses, and amino acid sequences. Calculated masses are monoisotopic values unless indicated as chemical average masses (av). Error refers to the difference between the observed and calculated masses. G denotes a glycopeptide, an asterisk (*) indicates that the NH_2-terminal assignment is supported by Edman sequence analysis and the symbols (⟨,⟩) denote the proteolytic cleavage sites. Mass spectra for all listed peptides are available in Figure 6 or the Electronic Appendix.

titation by MS without internal standards is not possible due to suppression effects, surface effects, charge state effects, and differences in ionization/desorption efficiencies among the various glycoforms present.

Discussion

Proof of six N-linked glycosylation sites in human serum clusterin and the identity of complex-type oligosaccharide structures at each attachment site was derived from Edman sequence and LC-ESMS analyses of clusterin tryptic peptides. Direct mass spectral analyses of glycopeptides provided the identity of three α subunit glycosylation sites (^{64}N, ^{81}N, and ^{123}N) and three β subunit glycosylation sites (^{64}N, ^{127}N, and ^{147}N). An earlier report inferred that these sites were glycosylated based upon sequence motif and the absence of phenylthiohydantoin asparagine in amino acid sequence analyses (Kirszbaum et al., 1992). The present study analyzed three different preparations of clusterin (including two sets of tryptic peptides from holoclusterin and one set from each of the isolated subunits) and found no evidence for any other attachment sites than the six N-linked glycosylation sites. The molecular weights of holoclusterin and the α and β subunits were determined by mass spectrometry, and the combined MALDI-TOF MS and ESMS analyses demonstrate mass heterogeneity due to variable amounts of glycosylation. These results highlight the sensitivity and efficacy of mass spectrometry in the microcharacterization of posttranslationally modified proteins. However, structural information of clusterin was limited by the relatively poor capability of mass spectrometry to resolve large molecules differing slightly in mass. Neither the extent of glycosylation heterogeneity nor its origins could be ascertained until clusterin was cleaved to peptides of smaller size (≤7,000 Da), where the accuracy and resolution of ESMS permitted single-dalton differences to be readily determined. On average, the deviation between calculated and observed masses in these ESMS measurements is less than 0.3 Da. Although very small peptides (di- and some tri-peptides) were below the full-

Clusterin α Subunit

```
D Q T V S D N E L Q  E M S N Q G S K Y V  N K E I Q N A V N G  V K Q I K T L I E K    40
T N E E R K T L L S  N L E E A K K K K E  D A L N E T R E S E  T K L K E L P G V C    80
N E T M M A L W E E  C K P C L K Q T C M  K F Y A R V C R S G  S G L V G R Q L E E   120
F L N Q S S P F Y F  W M N G D R I D S L  L E N D R Q Q T H M  L D V M Q D H F S R   160
A S S I I D E L F Q  D R F F T R E P Q D  T Y H Y L P F S L P  H R R P H F F F P K   200
S R I V R                                                                          205
```

Clusterin β Subunit

```
S L M P F S P Y E P  L N F H A M F Q P F  L E M I H E A Q Q A  M D I H F H S P A F    40
Q H P P T E F I R E  G D D D R T V C R E  I R H N S T G C L R  M K D Q C D K C R E    80
I L S V D C S T N N  P S Q A K L R R E L  D E S L Q V A E R L  T R K Y N E L L K S   120
Y Q W K M L N T S S  L L E Q L N E Q F N  W V S R L A N L T Q  G E D Q Y Y L R V T   160
T V A S H T S D S D  V P S G V T E V V V  K L F D S D P I T V  T V P V E V S R K N   200
P K F M E T V A E K  A L Q E Y R K K H R  E E                                      222
```

Fig. 4. Primary structure of the human clusterin α and β subunits. Amino acid sequences of the human clusterin subunits (Wong et al., 1994) are shown with identified asparagine carbohydrate attachment sites indicated in bold. More than 99% of the sequence was identified by peptide mass analysis, as summarized in Tables 3 and 4.

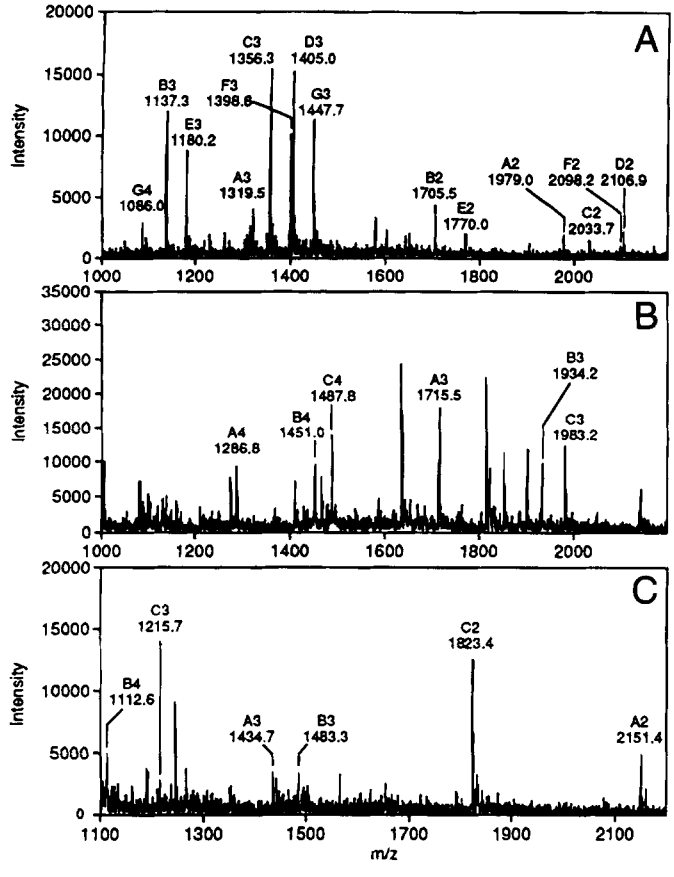
scan mass range used in the study, Edman degradation provided identification of many peptides. By using the combined MS and Edman approaches, essentially complete coverage of the 427-residue protein was obtained.

Carbohydrate heterogeneity

The total mass of carbohydrate attached to human serum holoclusterin was estimated to be 8.4–13.4 kDa or 17–27% by weight based upon MALDI-TOF MS measurements (Fig. 2A). Some nonglycosylated α subunit (24.2 kDa) was detected (Fig. 2B), but α-subunit species with about 2.2 and 7.1 kDa carbohydrate (9–30% by weight) appear more abundant. Completely nonglycosylated β-subunit was not found (Fig. 2C) and LC-ESMS demonstrated multiple β subunit species exist with about 6.9–7.7 kDa carbohydrate (27–30% by weight). The total carbohydrate mass estimated from the isolated subunits (8.6–14.8 kDa) is in reasonable agreement with that determined for the holoprotein (8.4–13.4 kDa). Although different oligosaccharide structures can occur at each site, if all six glycosylation sites exhibited the potentially common bisialobiantennary structure (type 2, mass 2,206), the total holoprotein mass due to carbohydrate would be 13.2 kDa, a value in agreement with the above-determined range.

Some of the N-linked attachment sites in human serum clusterin are only partially occupied by oligosaccharide. The lower molecular weight values observed by MALDI-TOF MS for holoclusterin and the subunits (Fig. 2) and the lack of detection of every glycosylation site in each LC-ESMS analysis support this conclusion. Partial glycosylation appears to be more prevalent on the α subunit than the β subunit and perhaps 20% of the α subunit population may at times have no attached carbohydrate, as suggested by the MALDI-TOF results from holoclusterin preparation 2 (Fig. 2). Nonglycosylated peptides were found for each of the α-subunit glycosylation sites in this clusterin preparation (Suppl. Table 2). The relative abundancy of nonglycosylated species in the other clusterin preparations used for LC-ESMS in Figure 3 and Supplementary Figure 1 was not estimated by MALDI-TOF MS. Notably, the only evidence found for partial glycosylation of the β subunit (peptide residues 145–181, Table 4) was at site β^{147}N, which also exhibits the greatest diversity in attached glycoforms. The observed oligosaccharide heterogeneity is associated in part with the fact that different clusterin preparations were utilized in the study. Although serum from a single individual was used throughout the study, blood was drawn at three different times over the course of 18 months for clusterin purification and characterization.

Notably, no O-linked glycosylation was detected, supporting an earlier report based on glycosidase experiments (Burkey et al., 1991). The NetOglyc computer program (Hansen et al., 1995) predicts that β chain residues ^{160}T, ^{161}T, ^{166}T, and ^{173}S are likely to be glycosylated, however, we consistently detected a tryptic peptide containing all of these residues as an unmodified peptide (β subunit residues 159–181 in Table 4 and Suppl. Tables 1 and 2). Furthermore, each LC-ESMS data set containing the β subunit was computer extracted for ions that would be generated if β peptide 159–181 was glycosylated on Ser and/or Thr [for example, with $(NeuAc)_n(HexHexNAc)_m$, where n = 0, 1, 2, 3, or 4 and m = 0, 1, 2, 4, 8, or 12]; only the nonglycosylated peptide was identified.



Peptide	Molecular Weight Glycopeptide	Peptide	CHO	CHO structure	Peptide residues	Attachment site
A	3956.0	947.0$_{av}$	3009.0	5	60-67	α^{64}N
B	3409.0	1203.3$_{av}$	2205.7	2	58-67	"
C	4065.4	"	2862.1	4	"	"
D	4211.8	"	3008.5	5	"	"
E	3537.8	1331.5$_{av}$	2206.3	2	57-67	"
F	4193.9	"	2862.4	4	"	"
G	4340.1	"	3008.6	5	"	"

Peptide	Molecular Weight Glycopeptide	Peptide	CHO	CHO structure	Peptide residues	Attachment site
A	5143.5	2937.5$_{av}$	2206.0	2	73-96	α^{81}N
B	5800.0	"	2862.5	4	"	"
C	5946.6	"	3009.1	5	"	"

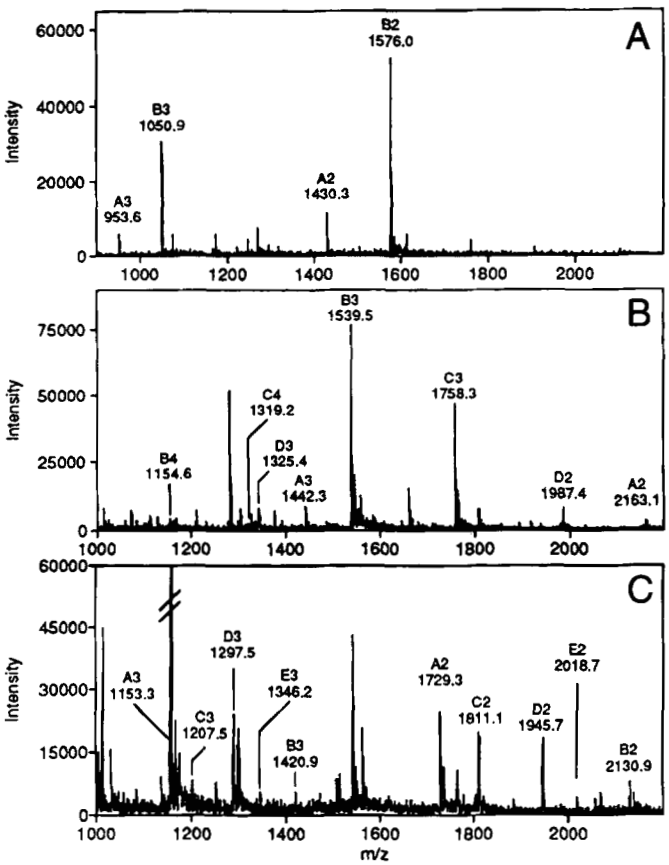
Peptide	Molecular Weight Glycopeptide	Peptide	CHO	CHO structure	Peptide residues	Attachment site
A	4300.8	1438.6$_{av}$	2862.2	4	117-128	α^{123}N
B	4446.9	"	3008.3	5	"	"
C	3644.8	"	2206.2	2	"	"

Fig. 5. Electrospray mass spectra of glycopeptides from clusterin α subunit. Glycopeptide mass spectra from the α subunit are presented by N-linked attachment site and retention time (rt) (from LC-ESMS, Fig. 3B). Glycopeptide *m/z* charge series are labeled by letter and positive charge (e.g., A2, A3, B2, B3, etc.). Unlabeled peaks correspond to co-eluting peptides. Measured glycopeptide masses as well as calculated peptide and oligosaccharide masses (CHO) are indicated in tabular form next to the relevant spectra. Based on the calculated oligosaccharide mass, the determined CHO structure is indicated by arbitrary type number where type 2 = bisialobiantennary, type 4 = trisialotriantennary, and type 5 = trisialotriantennary with one fucose. In-chain masses and structures of clusterin oligosaccharides are defined by type and summarized in Table 5. **A:** α^{64}N, rt = 33 min. **B:** α^{81}N, rt = 57–59 min. **C:** α^{123}N, rt = 54 min.

Oligosaccharide structures of human serum clusterin

Seven different possible complex types of oligosaccharide structures were identified and the relative amounts at each glycosylation site were estimated (Table 5). These structures are based on compositions derived from the determined molecular masses of the oligosaccharides and analogy to what is known about N-linked carbohydrates produced in mammalian cells. Three types of oligosaccharide structures were common to glycosylation sites α^{64}N, α^{81}N, α^{123}N, β^{127}N, and β^{127}N, namely bisialobiantennary structures without fucose and trisialotriantennary structures with one fucose or without fucose. Approximately equal amounts of types 2, 4, and 5 oligosaccharide structures were found at sites α^{64}N and α^{81}N, whereas the distribution of oligosaccharide structures at the other four glycosylation sites appear to exhibit more variability. Glycosylation sites β^{64}N and β^{127}N were the only sites where a monosialobiantennary structure was detected and β^{64}N exhibits the least apparent oligosaccharide diversity. Glycosylation site β^{147}N was the only site where a bisialobiantennary structure with one fucose (type 3), a possible sialyl Lewis type oligosaccharide (type 6), and/or a possible tetrasialotriantennary structure (type 7) were identified. Site β^{147}N exhibits the greatest glycosylation diversity, with five or six different oligosaccharides detected, and may play a special role in modulating clusterin biological activity.

The seven possible glycoforms listed in Table 5 appear to be the predominant oligosaccharide structures covalently associated with human serum clusterin, with type 2 structures the most abundant and types 1, 6, and/or 7 the least abundant. However, the carbohydrate content and attachment site occupancy also appears to fluctuate with protein preparation and other clusterin-bound oligosaccharides may be found in other preparations. The relative amounts of glycoforms listed in Table 5 reflect the approximate attachment site-specific distribution of oligosaccharides among those detected without adjustment for partial occupancy. It is important to emphasize that MS can only be used to provide an estimate of the relative amounts of glycoforms, and considerable caution must be exercised in deriving and using these values (see Roberts et al., 1995). Because serum clusterin represents the reservoir of clusterin synthesized in a number of tissues (including but not limited to liver, kidney, lung, and heart), it is possible that specific oligosaccharide structures may be found associated with clusterin depending upon tissue source, physiological status, disease state, and species.

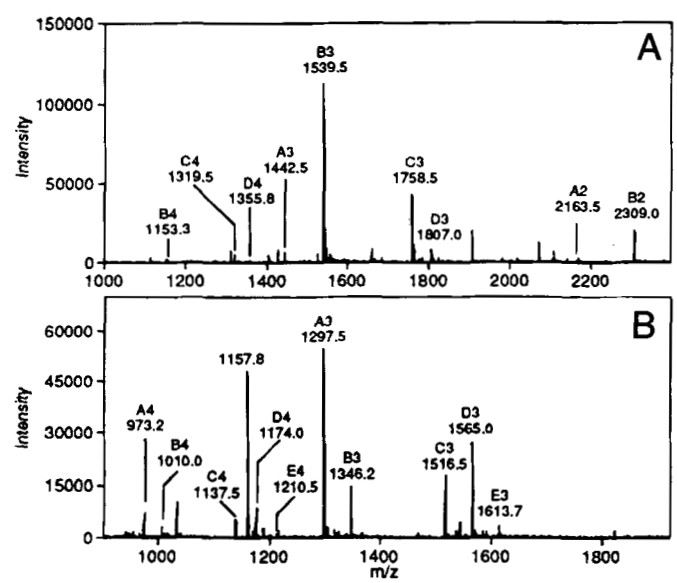


Peptide	Glycopeptide	Molecular Weight Peptide	CHO	CHO structure	Peptide residues	Attachment site
A	A: 2858.6	944.0$_{av}$	1914.6	1	63-70	β^{64}N
B	B: 3149.9	"	2205.9	2	"	"

Peptide	Glycopeptide	Molecular Weight Peptide	CHO	CHO structure	Peptide residues	Attachment site
A	4324.1	2409.7$_{av}$	1914.4	1	125-144	β^{127}N
B	4615.5	"	2205.8	2	125-144	"
C	5271.9	"	2862.2	4	"	"
D	3972.8	1767.0$_{av}$	2205.8	2	125-139	"

Peptide	Glycopeptide	Molecular Weight Peptide	CHO	CHO structure	Peptide residues	Attachment site
A	3456.8	1251.3$_{av}$	2205.5	2	145-155	β^{147}N
B	4259.8	"	3008.5	5	"	"
C	3620.2	1414.5$_{av}$	2205.7	2	145-156	"
D	3889.5	1683.8$_{av}$	2205.7	2	145-158	"
E	4035.5	"	2351.7	3	"	"

Fig. 6. Electrospray mass spectra of glycopeptides from clusterin β subunit. Glycopeptide mass spectra from the β subunit are presented by N-linked attachment site and retention time (rt) (from LC-ESMS in Fig. 3C,D). **A:** β^{64}N, rt = 33 min. **B:** β^{127}N, rt = 51–55 min. **C:** β^{147}N, rt = 38–41 min. Abbreviation and labels are as in Figure 5. The in-chain masses and structures of clusterin oligosaccharides are defined by type and summarized in Table 5.



Peptide	Glycopeptide	Molecular Weight Peptide	CHO	CHO structure	Peptide residues	Attachment site
A	4325.0	2409.7$_{av}$	1915.3	1	125-144	β^{127}N
B	4616.0	"	2206.3	2	"	"
C	5272.5	"	2862.8	4	"	"
D	5418.6	"	3008.9	5	"	"

Peptide	Glycopeptide	Molecular Weight Peptide	CHO	CHO structure	Peptide residues	Attachment site
A	3889.2	1683.8$_{av}$	2205.4	2	145-158	β^{147}N
B	4035.6	"	2351.8	3	"	"
C	4546.3	"	2862.5	4	"	"
D	4692.0	"	3008.2	5	"	"
E	4838.1	"	3154.3	6 or 7	"	"

Fig. 7. Electrospray mass spectra of glycopeptides from holoclusterin. Glycopeptide mass spectra from two holoclusterin LC-ESMS analyses are presented by N-linked attachment site and retention time (rt). **A:** β^{127}N, rt = 85 min (Suppl. Fig. 1). **B:** β^{147}N, rt = 27 min (Suppl. Fig. 2). Abbreviations and labels are as in Figure 5. The in-chain masses and structures of clusterin oligosaccharides are defined by type and summarized in Table 5.

Table 5. *Summary of clusterin oligosaccharide structures*[a]

Clusterin subunit	Attachment site	Type	Relative amount (%)	Oligosaccharide structure
α Subunit	^{64}N	2	30	Bisialobiantennary
		4	30	Trisialotriantennary
		5	40	Trisialotriantennary with one fucose
	^{81}N	2	33	Bisialobiantennary
		4	29	Trisialotriantennary
		5	38	Trisialotriantennary with one fucose
	^{123}N	2	65	Bisialobiantennary
		4	16	Trisialotriantennary
		5	19	Trisialotriantennary with one fucose
β Subunit	^{64}N	1	15	Monosialobiantennary
		2	85	Bisialobiantennary
	^{127}N	1	4	Monosialobiantennary
		2	68	Bisialobiantennary
		4	24	Trisialotriantennary
		5	4	Trisialotriantennary with one fucose
	^{147}N	2	44	Bisialobiantennary
		3	12	Bisialobiantennary with one fucose
		4	17	Trisialotriantennary
		5	24	Trisialotriantennary with one fucose
		6	3	Trisialotriantennary with two fucoses or
		7		Tetrasialotriantennary

[a]Oligosaccharide structures identified at each N-linked attachment site are listed. Approximate relative amount of glycoforms per attachment site was estimated as described in Materials and methods. All α subunit oligosaccharide amounts were estimated from Figure 5 (α^{64}N quantification was from residues 57 to 67); relative amounts for β subunit oligosaccharides were estimated from Figure 6A (β^{64}N), Figure 7A (β^{127}N), and Figure 7B (β^{147}N). Oligosaccharide in-chain chemical average masses and structures: Type 1, 1,914.8 [NeuAc(Hex-HexNAc)$_2$Hex$_3$HexNAc$_2$]; Type 2, 2,206.0 [(NeuAc-Hex-HexNAc)$_2$Hex$_3$HexNAc$_2$]; Type 3, 2,352.2 [(NeuAc-Hex-HexNAc)$_2$Hex$_3$HexNAc$_2$(dHex)]; Type 4, 2,862.6 [(NeuAc-Hex-HexNAc)$_3$Hex$_3$HexNAc$_2$]; Type 5, 3,008.8 [(NeuAc-Hex-HexNAc)$_3$Hex$_3$HexNAc$_2$(dHex)]; and Type 6, 3,154.9 [(NeuAc-Hex-HexNAc)$_3$Hex$_3$HexNAc$_2$(dHex$_2$)] or Type 7, 3,153.9 [(NeuAc)$_4$(Hex-HexNAc)$_3$Hex$_3$HexNAc$_2$].

Prior to this study, the oligosaccharides associated with rat Sertoli cell clusterin were the best characterized. Using methanolysis and gas chromatography, Griswold et al. (1986) reported the protein exhibits about 24% carbohydrate by weight, consisting of 1% fucose, 3.5% mannose, 4.1% galactose, 7.1% N-acetylglucosamine, and 8.0% N-acetylneuraminic acid, and proposed triantennary structures similar to fetuin. In addition, sulfate was found associated with the oligosaccharide, hence the acronym sulfated glycoprotein 2. We probed human serum clusterin samples by LC-ESMS for the presence of sulfate and found no evidence of sulfated oligosaccharide. If sulfate is present on human serum clusterin, it is present in trace amounts below the detection limits of the present analysis. Sulfation processes are prevalent in testicular tissue (particularly Sertoli cells) and sulfation of clusterin may be restricted to this tissue. Characterization of clusterin from testis rather than serum could enhance the likelihood of detecting sulfated oligosaccharide.

Clusterin has been associated with diverse physiological processes and is up regulated in response to various cellular stress and pathological conditions. Notably, the clusterin transcript and protein have been detected at low levels in normal brain, and at markedly elevated levels in patients with neurodegenerative diseases such as Alzheimer's, epilepsy, and retinitis pigmentosa (Wong et al., 1994). It is also expressed in a number of glandular tissues and tumors that undergo apoptosis in response to hormone ablation (Wong et al., 1994) and response to acute inflammatory stimuli (Hardardottir et al., 1994). Given these responses to cellular stress, the glycosylation state of clusterin likely plays a significant role in its biological activity. For example, the presence or absence of specific oligosaccharides (e.g., a sialyl Lewis type structure) at a given attachment site may influence intracellular or extracellular targeting of clusterin in different physiological conditions. Evaluating the role of glycosylation in clusterin biological function is particularly challenging because the function of the protein remains poorly defined. As an approach to determining the physiological role of clusterin glycosylation and whether variations in clusterin glycosylation are tissue specific, we have constructed a series of expression vectors containing human clusterin mutants lacking from one to five of the six glycosylation sites. These vectors will be used in transfections to compare mutant and wild-type clusterin expression in select cell lines (e.g., kidney, liver, prostate, mammary, and neural). Subsequent analysis of oligosaccharide structures by mass spectrometry will provide insights into how each cell type modifies clusterin in a specific fashion. In addition, recombinant mutant and wild-type clusterin will be produced in yeast and tested for binding to various natural and synthetic membrane fractions and cell types. Correlation of recombinant protein oligosaccharide structures with membrane-binding properties will provide insight into how clusterin elicits the clustering of cells and proposed functions of the protein. The present

study provides a firm molecular basis for these investigations and for relating specific clusterin glycosylation states with physiological function.

Materials and methods

Immunoaffinity purification of human serum clusterin

Clusterin was purified from human plasma by immunoaffinity chromatography using CNBr-activated Sepharose 4B with attached mouse monoclonal antibody G7 to human serum clusterin (Wilson & Easterbrook-Smith, 1992). The affinity matrix was a generous gift of Dr. Mark Wilson (Department of Biological Sciences, University of Wollongong, New South Wales, Australia). Plasma used for protein purification was obtained from one individual, a healthy 32 year-old male (blood type O^+) and co-author (J.L.). Blood was drawn at different times during the study; however, protein was purified and characterized from plasma recovered from individual bleeds. Briefly, blood (~140 mL) was drawn in syringes containing 0.5 mL of 0.2 M EDTA, pH 7.4 (5 mM EDTA final concentration), and centrifuged at 2,500 rpm/10 min at room temperature in a table-top IEC Centra 8R centrifuge. Protease inhibitors (benzamidine and PMSF; Sigma) were added (1 mM each) to the recovered plasma (~70 mL) and the preparation re-centrifuged in 30-mL Corex tubes coated with Gel Slick™ (AT Biochem) at 10,000 rpm for 15 min at 4 °C in a JA-17 rotor in a Beckman J6B centrifuge. The supernatant was diluted with one volume of cold PBS containing 0.5 M NaCl, 5 mM EDTA, pH 7.4 (high salt PBSE), and benzamidine, and PMSF added to 1 mM. Diluted plasma was applied to a column of the affinity matrix (2-mL bed volume) equilibrated with high-salt PBSE at 4 °C. The column was washed with high-salt PBSE at a flow rate of 0.5 mL/min until the 280-nm absorbance stabilized and was then washed successively with low-salt PBSE (PBS containing 5 mM EDTA), low-salt PBSE containing 0.5% Triton X-100, and low-salt PBSE to remove detergent. Clusterin was eluted with 0.2 M glycine-HCl, pH 2.7, and fractions (2 mL) collected in tubes containing 150 μL of 2 M Tris-HCl, pH 8.0 (at 22 °C), and mixed immediately to neutralize the pH. Clusterin-containing fractions were identified by 280-nm absorbance, pooled, concentrated, and dialyzed 3× against 10-fold dilutions of PBS in a CM-30 ultrafiltration unit (Amicon) at 4 °C. Following concentration and dialysis, about 2 mg of clusterin was typically recovered in about 250 μL as determined by amino acid analysis.

Alkylation, fragmentation, and RP-HPLC purification

Immunoaffinity-purified human serum clusterin was reduced with dithiothreitol and alkylated with iodoacetamide in the presence of 8 M urea, 200 mM ammonium bicarbonate, pH 8 (Stone et al., 1990), then digested with trypsin in 2 M urea or fractionated by RP-HPLC to isolate the intact α and β subunits. Alternatively, immunoaffinity purified clusterin was subjected to RP-HPLC purification without reduction and alkylation to obtain a desalted holoprotein sample appropriate for MALDI-TOF MS. Tryptic digestion of the isolated α and β subunits was performed as for the holoprotein (5% trypsin by weight, overnight at 37 °C). Tryptic digests and intact protein samples (≤ ~500 pmol or ~20 μg protein) were fractionated by RP-HPLC (Crabb et al., 1988) at a flow rate of 50 μL/min on 5-μ Vydac C4 or C18 microbore columns (1 × 250 mm) using aqueous acetonitrile/trifluoroacetic acid solvents and either an Applied Biosystems model 120 HPLC system equipped with a 75-μL dynamic mixer or a Beckman Gold HPLC system.

MALDI-TOF MS

About 5 pmol clusterin samples were analyzed by MALDI-TOF MS using a VG TofSpec laser desorption mass spectrometer. RP-HPLC-purified holoclusterin and the isolated α and β subunit samples in aqueous TFA/acetonitrile solvent were prepared for MALDI-TOF MS analysis by mixing 1 μL of approximately 10 pmol/mL glycoprotein with 1 μL of matrix (3,5-dihydroxybenzoic acid, 10 mg/mL in 50% ethanol or sinapinic acid, 10 mg/mL, in 25% ethanol, 25% acetonitrile) directly on the stainless steel target. Laser desorption/ionization mass spectra were acquired with a laser energy setting of 50% and an acceleration voltage of 25 kV; each spectrum was the sum of at least 50 laser shots. Bovine serum albumin (M_r 66,430 kDa) and bovine carbonic anhydrase (M_r 29,022 kDa) were used to calibrate the MALDI-TOF MS instrument.

ESMS

ESMS and LC-ESMS were performed with either a PE Sciex API 300 or a PE Sciex API III triple quadrupole mass spectrometer fitted with articulated ionspray plenums and atmospheric pressure ionization sources. Glycopeptides were detected selectively by LC-ESMS based on diagnostic sugar oxonium ions $HexNAc^+$ (produced by GalNAc or GlcNAc, *m/z* 204), $HexNAc^+$-$(H_2O)_2$ (*m/z* 168), NeuAc (N-acetylneuraminic acid, *m/z* 292), and Hex-$HexNAc^+$ (produced by Gal-GlcNAc, Man-GlcNAc, Gal-GalNAc, etc. *m/z* 366). The abundance of these low-mass marker ions in electrospray mass spectra was enhanced by collision-induced decomposition of the parent ions. This was accomplished by stepping the potential, which controls the extent of collision-induced decomposition of source-produced ions from a high voltage, to maximize fragment-ion production during acquisition of low *m/z* ions, to a lower voltage to yield intact ionized molecules during the remainder of the scan (Carr et al., 1993; Huddleston et al., 1993). In this way, both intact parent ions and abundant marker ions were observed in the same *m/z* scan. Carbohydrate marker ions at *m/z* 168 and/or *m/z* 204, 292, and 366 (dwell time, 100 ms each) were monitored in positive ion mode, at high orifice potential (90 V for the API 300 and 140 V for the API III). Full scans at *m/z* 475–1,925 or 400–2,400 (0.25 amu steps, scan time 6 s) were then acquired at a lower orifice potential (35 V with the API 300 and 65V with the API III). A bovine fetuin tryptic digest was infused into the mass spectrometer at 5 μL/min for optimizing the sensitivity of the carbohydrate marker ions. For LC-ESMS, the HPLC eluant was split, with 30–45% going to the mass spectrometer and the remainder collected in 1-min fractions. Portions of select LC-ESMS glycopeptide fractions were dried, resuspended in 0.2–5% formic acid, 50% methanol, and re-analyzed by infusion at 5 μL/min to confirm observed masses. LC-ESMS for sulfate detection (540 pmol tryptic digest of holoclusterin) monitored SO_3^- marker ions (*m/z* 80, dwell time 3 s) in negative ion mode at high orifice potential (160 V) and acquired full scans in positive ion mode at *m/z* 500–1,900 (0.25 amu steps, scan time 2 s) at lower orifice potential (35 V) in the same analysis using the API 300 instrument (Bean et al., 1995).

Attachment sites and the structural class of oligosaccharide at clusterin glycosylation sites were determined by a combination of Edman degradation and ESMS. The N-terminal sequence of glycopeptides and the position of the carbohydrate attachment sites were determined by Edman degradation and masses for the various peptides were calculated based on the known protein sequence and expected C-terminal proteolytic cleavage site. The peptide mass

was then subtracted from the measured mass of the glycopeptide to determine the mass of the oligosaccharide. Prior to or in the absence of sequence analysis, candidate peptide masses predicted from the known protein sequence were subtracted from the measured glycopeptide masses to determine the oligosaccharide mass; alternatively, typical mammalian oligosaccharide masses were subtracted from the glycopeptide mass to identify the peptide. The structural class of oligosaccharide was determined by comparison of the oligosaccharide mass with the established carbohydrate structures typically found in mammalian glycoproteins. A range of Hex (in-chain chemical average mass, $i = 162.14$), HexNAc ($i = 203.20$), dHex ($i = 146.14$), and NeuAc ($i = 291.26$) residues was used to interpret candidate carbohydrate masses in order to identify unusual structures, if present. Tryptic and chymotryptic proteolysis provided the major and minor C-terminal cleavage sites, respectively, but other possible cleavage sites were always considered during data interpretation. Prior to Edman analysis, initial oligosaccharide assignments were proposed based on candidate peptides containing the sequence motif for N-linked glycosylation: -Asn-X-Ser/Thr-, where X is any amino acid except Pro (Carr et al., 1993; Huddleston et al., 1993).

The relative amount of oligosaccharide structures at each glycosylation site was estimated based on electrospray positive ion intensity of the glycopeptides (Roberts et al., 1995). First, the observed ion intensities for each charge state of a glycoform were summed, then the total intensity for all the glycoforms was calculated. The relative amount or distribution of each glycoform at the site was estimated as percent of the total intensity. For example, in Figure 6A, charge-state ion intensity sums are A2 + A3 ~15,000 and B2 + B3 ~82,000; total ion intensity is A2 + A3 + B2 + B3 ~97,000; and relative amounts are type 1 (A2 + A3) ~15% and type 2 (B2 + B3) ~85%. Only a single glycopeptide species was used for quantifying each glycoform, even if multiple peptides were detected due to partial proteolysis. This quantification method assumes that the peptide picks up charge at low pH more readily than does the carbohydrate portion. By limiting the quantification to a single peptide species, the peptide portion of the glycopeptide serves as an internal standard for estimating the amount of the oligosaccharide species.

Peptide sequencing and amino acid analysis

Phenylthiocarbamyl amino acid analysis was performed as described previously using an automated analysis system (Applied Biosystems model 420H/130/920) (West & Crabb, 1992). Microsequence analysis by Edman degradation was performed using an Applied Biosystems gas phase sequencer (model 470/120/900) as described previously (Crabb et al., 1988).

Supplementary material in Electronic Appendix

The Electronic Appendix contains three primary folders entitled Supplementary Figures, Supplementary Tables, and Supplementary Mass Spectra. The Supplementary Figures folder contains two files, Supplementary Figure 1 (LC-ESMS of a Tryptic Digest of Holoclusterin Preparation One) and Supplementary Figure 2 (LC-ESMS of a Tryptic Digest of Holoclusterin Preparation Two). The Supplementary Tables folder contains two files, Supplementary Table 1 (Peptides Identified by LC-ESMS from Holoclusterin Preparation One) and Supplementary Table 2 (Peptides Identified by LC-ESMS from Holoclusterin Preparation Two). The Supplementary Mass Spectra folder contains four secondary folders entitled Spectra Supporting Table 3, Spectra Supporting Table 4, Spectra-Supplementary Table 1, and Spectra-Supplementary Table 2. Files within each secondary folder in Supplementary Mass Spectra are titled by retention time as listed in the relevant tables. Each supplementary mass spectra is labeled with the relevant table number and peptide residue numbers.

Acknowledgments

We thank Jeffery Hulmes, Scott Dodson, Michael Huddleston, and Gerry Roberts for expert technical assistance, and Drs. Roland Annan, Mark Bean, and Mark Hemling for valuable discussions. Special thanks are extended to Dr. Robert Trimble for reviewing the manuscript prior to publication. This work was supported in part by USPHS grants EY06603 and DK38639.

References

Bean MF, Annan RS, Hemling ME, Mentzer M, Huddleston MJ, Carr SA. 1995. LC-MS methods for selective detection of posttranslational modifications in proteins: Glycosylation, phosphorylation, sulfation and acylation. In: Crabb JW, ed. *Techniques in protein chemistry VI*. San Diego: Academic Press. pp 107–116.

Blaschuk O, Burdzy K, Fritz I. 1983. Purification and characterization of a cell-aggregating factor (clusterin), the major glycoprotein in ram rete testis fluid. *J Biol Chem 12*:7714–7720.

Blaschuk O, Fritz I. 1984. Isoelectric forms of clusterin isolated from ram rete testis fluid and from secretions of primary cultures of ram and rat Sertoli-cell-enriched preparations. *Can J Biochem Cell Biol 62*:456–461.

Burkey B, de Silva H, Harmony J. 1991. Intracellular processing of apolipoprotein J precursor to the mature heterodimer. *J Lipid Res 32*:1039–1048.

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

Choi-Miura NH, Takahashi Y, Nakano Y, Tobe T, Tomita M. 1992. Identification of the disulfide bonds in human plasma protein SP-40,40 (apolipoprotein-J). *J Biochem 112*:557–561.

Crabb JW, Johnson CM, Carr SA, Armes LG, Saari JC. 1988. The complete primary structure of the cellular retinaldehyde-binding protein from bovine retina. *J Biol Chem 263*:18678–18687.

de Silva HV, Harmony JAK, Stuart WD, Gil CM, Robbins J. 1990a. Apolipoprotein J: Structure and tissue distribution. *Biochemistry 29*:5380–5389.

de Silva HV, Stuart WD, Park YB, Mao SJT, Gil CM, Wetterau JR, Busch SJ, Harmony JAK. 1990b. Purification and characterization of apolipoprotein J. *J Biol Chem 265*:14292–14297.

Griswold M, Roberts K, Bishop P. 1986. Purification and characterization of a sulfated glycoprotein secreted by Sertoli cells. *Biochemistry 25*:7265–7270.

Hansen JE, Lund O, Engelbrecht J, Bohr H, Nielsen JO, Hansen JES, Brunak S. 1995. Prediction of O-glycosylation of mammalian proteins: Specificity patterns of UDP-galnac:polypeptide N-acetylgalactosaminyltransferase. *Biochem J 308*:801–813.

Hardardottir I, Kunitake S, Moser A, Doerrler W, Rapp J, Grunfeld C, Feingold K. 1994. Endotoxin and cytokines increase hepatic messenger RNA levels and serum concentrations of apolipoprotein J (clusterin) in Syrian hamsters. *J Clin Invest 94*:1304–1309.

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.

Jenne DE, Tschopp J. 1989. Molecular structure and functional characterization of a human complement cytolysis inhibitor found in blood and seminal plasma: Identity to sulfated glycoprotein 2, a constituent of rat testis fluid. *Proc Natl Acad Sci USA 86*:7123–7127.

Kirszbaum L, Bozas SE, Walker ID. 1992. SP-40,40, a protein involved in the control of the complement pathway, possesses a unique array of disulphide bridges. *FEBS Lett 297*:70–76.

Kirszbaum L, Sharpe JA, Murphy B, d'Apice AJF, Classon B, Hudson P, Walker ID. 1989. Molecular cloning and characterization of the novel, human complement-associated protein, SP-40,40: A link between the complement and reproductive systems. *EMBO J 8*:711–718.

Roberts GD, Johnson WP, Burman S, Anumula KR, Carr SA. 1995. An integrated strategy for structural characterization of monoclonal antibodies: Application to anti-respiratory syncytial virus Mab. *Anal Chem 67*:3613–3625.

Rosenberg M, Silkensen J. 1995. Clusterin: Physiologic and pathophysiologic considerations. *Int J Biochem Cell Biol 27*:633–645.

Stone KL, Elliott JI, Peterson G, McMurray W, Williams KR. 1990. Reversed-phase high performance liquid chromatography for fractionation of enzymatic digests and chemical cleavage products of proteins. *Methods Enzymol 193*:389–412.

Sylvester S, Skinner M, Griswold M. 1984. A sulfated glycoprotein synthesized by Sertoli cells and by epididymal cells is a component of the sperm membrane. *Biol Reprod 31*:1087–1101.

Wagner M, Morgans C, Koch-Brandt C. 1995. The oligosaccharides have an essential but indirect role in sorting GP80 (clusterin, TRPM-2) to the apical surface of MDCK cells. *Eur J Cell Biol 67*:84–88.

West KA, Crabb JW. 1992. Applications of automatic PTC amino acid analysis. In: Angeletti R, ed. *Techniques in protein chemistry III*. San Diego: Academic Press. pp 233–242.

Wilson M, Easterbrook-Smith S. 1992. Clusterin binds by a multivalent mechanism to the Fc and Fab regions of IgG. *Biochim Biophys Acta 1159*:319–326.

Wilson M, Easterbrook-Smith S, Taillefer D, Lakins J, Tenniswood M. 1995. The role of clusterin at sites of active cell death. In: Harmony J, ed. *Clusterin: Role in vertebrate development function and adaption*. Material Molecular Biology Intelligence Unit. Austin, Texas: RG Landes Co. pp 95–112.

Wong P, Pineault J, Lakins J, Taillefer D, Léger J, Wang C, Tenniswood M. 1993. Genomic organization and expression of the rat TRPM-2 (clusterin) gene, a gene implicated in apoptosis. *J Biol Chem 268*:5021–5031.

Wong P, Taillefer D, Lakins J, Pineault J, Chader G, Tenniswood M. 1994. Molecular characterization of human TRPM-2/clusterin, a gene associated with sperm maturation, apoptosis and neurodegeneration. *Eur J Biochem 221*:917–925.