

Identification of *N*-Glycosites in Chicken Egg White Proteins Using an Omics Strategy

Fang Geng,[†] Jinqiu Wang,[†] Dayu Liu,^{*,†} Yongguo Jin,[‡] and Meihu Ma^{*,‡}

[†]Key Laboratory of Coarse Cereal Processing, Ministry of Agriculture, College of Pharmacy and Biological Engineering, Chengdu University, No. 1 Upper Section of Shiling Street, Chengdu 610106, P. R. China

[‡]National R&D Center for Egg Processing, College of Food Science and Technology, Huazhong Agricultural University, No. 1 Shizishan Street, Wuhan 430070, P. R. China

S Supporting Information

ABSTRACT: Chicken egg white (CEW) is a perfect source of natural proteins that possesses outstanding functional properties and various bioactivities. The glycosylation structure of CEW proteins plays important roles in their functions, bioactivities, and allergies. The present work attempted to identify *N*-glycosites of CEW proteins using an omics strategy. CEW proteins were digested with trypsin and chymotrypsin; glycopeptides were enriched and deglycosylated using PNGase F in H₂¹⁸O water, followed by analysis using high-performance liquid chromatography/tandem mass spectrometry (HPLC-MS/MS). A total of 71 *N*-glycosites in 26 CEW glycoproteins were identified. Web-Logo analysis showed that most of the *N*-glycosites were at N-X-T (55%) and N-X-S (32%). Furthermore, two-dimensional electrophoresis of CEW clusterin demonstrated a series of spots horizontally distributed at 35–37 kDa with an extremely wide isoelectric point range of 4.54–6.68, indicating the heterogeneity of glycosylation of CEW clusterin. These results provided important information for the understanding of the structures, functions, and bioactivities of CEW glycoproteins.

KEYWORDS: chicken egg white, glycoproteins, *N*-glycosylation site, mass spectrometry, clusterin

INTRODUCTION

The functions of proteins depend on their structures, including their post-translational modifications and especially their glycosylations.^{1,2} Research in the field of egg science has shown that the glycosylation of chicken egg white (CEW) proteins plays important roles in egg white thinning, defense against microbes, and egg white allergies.^{3–6} The highly glycosylated ovomucin is considered the main provider of the viscosity of egg whites; the degradation of glycans on ovomucin will result in the thinning of egg white.⁷ Meanwhile, CEW glycoproteins and glycopeptides play important roles in resisting microbial invasion. For example, ovomucin and its derived glycopeptides demonstrated antiagglutinating activity toward Newcastle disease virus and *E. coli*.^{3,8} In addition, the mannosylated CEW protein and ovalbumin could attenuate an orally induced egg allergy in mice, indicating that glycan modification of CEW proteins would influence the egg-induced allergies.^{9,10}

Therefore, a number of studies have focused on the glycosylation sites and glycan structures of CEW proteins. The main CEW proteins, including ovalbumin, ovomucoid, ovomucin, and ovotransferrin, are all glycoproteins, and their glycosylation sites have been clarified.^{11,12} However, other CEW glycoproteins suffer from a lack of investigation. Several CEW proteins, such as ovoinhibitor, clusterin, and ovoglycoprotein, are known as glycoproteins, but details about their glycosylation are lacking. In addition, the glycosylation of hundreds of low-abundance CEW proteins, which were found in proteomic studies, has also not yet been explored.^{13,14}

The present work aims to identify the *N*-glycosites in CEW proteins using an omics strategy. CEW proteins were digested

with trypsin and chymotrypsin, and the glycopeptides were collected using lectins, followed by deglycosylation using PNGase F in H₂¹⁸O water. Deglycopeptides were identified using high-performance liquid chromatography/tandem mass spectrometry (HPLC-MS/MS), and *N*-glycosites were obtained from the MS/MS data via the MASCOT program. The results will provide important structural information about CEW glycoproteins and will be beneficial for clarifying the roles of glycosylation in the functional and biological activities of CEW proteins. These results will also aid in the understanding of the underlying mechanism of quality changes/deterioration during egg storage and processing.

MATERIALS AND METHODS

Egg White Sampling. Fresh chicken eggs laid within 24 h from white leghorns were collected in the morning from the Poultry Research Center farm of Huazhong Agricultural University and were used in the study.

Digestion, Enrichment, and Deglycosylation of CEW Glycopeptides. CEW proteins were processed based on the “filter-aided sample preparation” (FASP) protocol.^{15,16} Briefly, CEW lyophilized powder (400 μg) was dissolved with 200 μL of UA buffer (8 mol/L urea, 150 mmol/L Tris-HCl, pH 8.0) in a 10-kDa filtration tube (Amicon Ultra, Millipore, Bedford, MA). Then, 100 μL of iodoacetamide (60 mmol/L in UA buffer) was added, and the solution was treated for 30 min at room temperature. After elution with 25 mmol/L NH₄HCO₃, proteins were digested using trypsin or

Received: April 12, 2017

Revised: May 31, 2017

Accepted: June 6, 2017

Published: June 6, 2017

chymotrypsin (8 μ g in 40 μ L of NH_4HCO_3 , 100 mmol/L) at 37 °C for 12 h. After centrifugation, the filtrate containing digested peptides was transferred to a new 10-kDa filtration tube, and a lectin mixture (150 μ L) containing concanavalin A (Sigma, L7647), wheat germ agglutinin (Sigma, L9640), and RCA120 (Sigma, L7886, Louis, MO) in a 2:1 mass proportion to peptides was added to the top of the tube. After 1 h of incubation, unbound peptides were washed, and captured peptides were eluted twice with 25 mmol/L NH_4HCO_3 in H_2^{18}O . Deglycosylation was performed in H_2^{18}O at 37 °C, 4 μ g of PNGase F (Roche, 11365185001, Mannheim, Germany) was added, and incubation occurred at 37 °C for 3 h to release asparagine-linked oligosaccharides. The final solution was washed with 25 mmol/L NH_4HCO_3 and concentrated to 25 μ L.

Mass Spectrometric Analysis. Mass spectrometric measurements were performed via HPLC-ESI-MS/MS using a Q-Exactive mass spectrometer equipped with an Easy-nLC (Thermo-Fisher Science, Bremen, Germany). Ten microliters of deglycosylated peptides was loaded onto an Easy column (Thermo Scientific, 75 $\mu\text{m} \times 100$ mm, 3 μm), and separation was performed at a flow rate of 250 nL/min using a gradient constructed from solution A (2% acetonitrile, 0.1% formic acid) and solution B (84% acetonitrile, 0.1% formic acid): 4–10% B for 2 min; 10–20% B for 120 min; 20–45% B for 120 min; 45–100% B for 10 min; and 100% B for 18 min. The mass range for the MS scan was set to 300–1800 m/z , the MS resolution was 70 000 at m/z of 200, the AGC target was 3×10^6 , the maximum IT was 20 ms, the number of scan ranges was 1, and the dynamic exclusion was 25.0 s. For MS/MS analysis, peptides were subjected to fragmentation by high-energy collision-induced decomposition (HCD). The isolation window was 2 m/z , the resolution of MS/MS was 17 500 at m/z of 200, the maximum IT was 60 ms, the normalized collision energy was 27 eV, and the underfill ratio was 0.1%.^{17,18}

MS/MS Data Analysis. Analysis of the data was performed using MaxQuant software version 1.3.0.5 and comparing against the databases (NCBI, *Gallus*, v20160311). The search used was a MS/MS ion search, with trypsin or chymotrypsin as the enzyme, and up to two missed cleavages were allowed. Carbamidomethylation (C) was set as a fixed modification; deamidation with ^{18}O (N) and oxidation (M) were specified as variable modifications. Peptide ions were identified with a mass tolerance of 5 ppm, and fragment ions were identified with a mass tolerance of 0.1 Da. For identification, the false discovery rate (FDR) was specified to 1% at both the peptide and site levels.^{19,20} The complete operational processes are shown in Figure 1.

Two-Dimensional Electrophoresis and Clusterin Identification. CEW was diluted with distilled water (1:1, V/V), followed by a two-step polyethylene glycol (PEG) precipitation (PEG-8000, 5–9%, w/w) to obtain clusterin-containing fractions (P_{5-9}).²¹ Then, the P_{5-9} fractions were ultrafiltered using a cup ultrafilter (model SCM300, Shanghai Institute of Applied Physics, Shanghai, China) equipped with a 50 kDa molecular weight cutoff (MWCO) ultrafiltration membrane (SINAP Membrane Science and Technology Co. Ltd., Shanghai, China). The retained fraction (clusterin-rich fraction) was collected and used for the following analysis.

Two-dimensional electrophoresis (2-DE) analysis was performed as in the previous study.^{19,22} Briefly, the clusterin-rich fraction (40 μ g of total protein) was loaded onto DryStrip IPG strips (24 cm; pH 4–7) and isoelectrically focused. After equilibration to resolubilize proteins and to reduce disulfide bonds, the second-dimension electrophoresis was performed using a 10% SDS-PAGE. Protein spots were visualized by Coomassie Brilliant Blue staining, and data analysis was performed using the Image Master V 7.0 program (GE Healthcare). For identification, the candidate protein spots were excised from the gel and digested with trypsin (Promega, Madison, WI). Then, the digested peptides were identified using a matrix-assisted laser desorption/ionization time-of-flight MS/MS (MALDI-TOF MS/MS, Bruker, Karlsruhe, Germany). Data were searched against the protein database (NCBI, *Gallus gallus*) via the MASCOT program (<http://www.matrixscience.com>).²²

Bioinformatics Analysis. All of the potential glycosylation sites of the identified CEW glycoproteins were predicted by the NetNGlyc 1.0

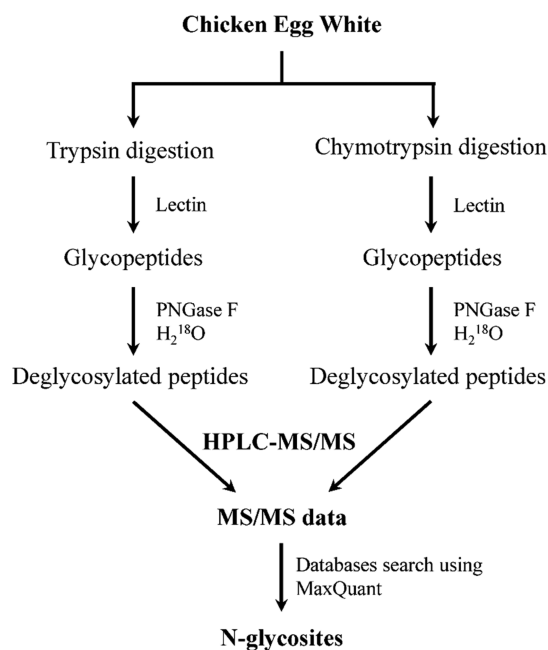


Figure 1. Flowchart of N-glycosite identification in chicken egg white (CEW) proteins.

Server (<http://www.cbs.dtu.dk/services/NetNGlyc/>) based on the sequences obtained from the UniProt database (<http://www.uniprot.org/>).

RESULTS AND DISCUSSION

Identified N-Glycosites in CEW. Glycopeptides derived from CEW proteins were enriched using mixed lectins and deglycosylated by PNGase F in H_2^{18}O , resulting in a transfer of the originally glycosylated asparagine to aspartic acid (occupied by ^{-18}OH) with a 2.99-Da increase in molecular weight, which can be exploited to identify N-glycosites.^{15,16} For obtaining higher coverage, two different proteases, trypsin and chymotrypsin, were used for the digestion of CEW proteins.

A total of 88 unique glycopeptides were identified (Table 1); these peptides contain 71 N-glycosites and belong to 26 egg white proteins. All of the identified glycopeptides were identified with high precision, and the mass tolerance of peptide ions was less than 5 ppm (Figure 2A). During the identification of the 88 glycopeptides, 53 of them were identified in the chymotrypsin digestion process, 35 of them were identified in the trypsin digestion process, and 17 of them were identified in the both digestion processes (Figure 2B).

The canonical N-glycosylation motif is N-X-[S/T] (where X is not proline); N-X-C has also been considered as one of the conserved sequences. Here, Web-Logo was employed to visualize the sequence motifs around the N-glycosites (9 amino acids on each side).²³ As shown in Figure 2C, threonine (T) and serine (S) were significantly overrepresented at the +2 position, meaning that most of the identified egg white N-glycosites were typical sites. A preference for serine (S), aspartic (D), glycine (G), and leucine (L) at the +1 position was also obvious. In the present study, N-glycosites that matched with N-X-T (55%) occurred more frequently than those that matched with N-X-S (32%). Additionally, there were 3 N-X-C motif N-glycosites (4%) in the identified result, and the other six identified N-glycosites (9%) are noncanonical sequons (Figure 2D). The representative MS/MS spectra of the N-glycopeptides are presented in Figure S1.

Table 1. List of N-Glycopeptides Resulting from a Database Search of HPLC-MS/MS Data

no. in UniProt	protein name	position of glycosite	peptide (digestion enzyme)	m/z charge	exp. mass (Da) ^a	calc. mass (Da) ^b	score ^c
Q8JIG5	α_1 -acid glycoprotein (ovoglycoprotein)	N ³⁶	VTADMDN*ATVDRLL (CT) ^d	768.88 ²⁺	1535.75	1532.75	157.86
		N ⁹⁰	VPADMDN*ATVDRLL (CT)	766.88 ²⁺	1531.76	1528.75	85.91
		N ¹⁰⁷	LN*ETCVVK (T)	483.25 ²⁺	964.49	961.47	107.66
			HN*STLTTHEDGQVSVMAELIHSKD (T)	638.55 ⁴⁺	2550.22	2547.20	267.09
O57579	aminopeptidase Ey		RHN*STLTTHEDGQVSVMAEL (CT)	709.67 ³⁺	2126.02	2123.01	180.08
		N ²⁰⁶	SEYEGN*VTK (T)	565.76 ²⁺	1129.52	1126.51	79.69
		N ⁶⁵	GSQVQAN*YSLK (T) ^e	615.29 ²⁺	1228.58	1168.55	65.40
		N ¹⁴¹	TNDLGSN*MTIGAVNSRGEF (CT)	993.46 ²⁺	1984.92	1981.92	85.81
		N ⁹⁹	QEVQCN*ETMLSLWECKPCLK (T)	853.05 ³⁺	2556.15	2382.07	92.68
			ALEKEKQLAEKQEVQCN*ETML (CT)	1197.59 ²⁺	2393.19	2333.16	168.20
		N ¹⁴¹	LN*RSSPF (CT)	412.21 ²⁺	822.43	819.42	70.27
		N ²⁷⁵	LGGFESESRN*F (CT)	623.29 ²⁺	1244.57	1241.57	90.78
		N ³⁵²	QAEMLN*TSLL (CT)	605.30 ²⁺	1208.60	1205.60	90.11
		N ³⁷²	RLGN*LTQGTDF (CT)	641.32 ²⁺	1280.64	1277.64	124.42
U6C3W5	EW135	N ¹²³	GEHNCNHGEDASVVCSGNN*KTVQL (CT)	877.05 ⁺	2628.14	2511.09	72.32
		N ³⁴⁸	AELLPVRLVN*GSNF (CT)	766.42 ²⁺	1530.84	1527.84	58.32
		N ²¹⁰	HGN*TSWGNAGDR (T)	425.52 ³⁺	1273.55	1270.54	212.59
		N ¹⁴⁵	TWFDSN*NSSVSGMDVIPKVISGPPY (CT)	1350.64 ²⁺	2699.28	2696.28	37.90
H9L385	hemopexin		TCRVTHN*GTSITKTL (CT)	564.63 ³⁺	1690.88	1630.85	54.40
P01875	Ig mu chain C region	N ²¹⁶	EN*ISDPTSLR (T)	616.31 ²⁺	1230.61	1227.61	100.55
P20763	Ig lambda chain C region	N ⁷³	ESNFN*QTATNRNTDGGSTDY (CT)	1069.44 ²⁺	2136.89	2133.88	94.39
E1BY93	Ig J chain	N ⁵⁷	FESNFNTQATN*R (T)	716.32 ²⁺	1430.65	1427.64	72.09
P00698	lysozyme	N ⁶²	NTQATN*RNTDGGSTDY (CT)	830.85 ²⁺	1659.70	1656.70	159.18
Q98UJ9	Mucin-5B	N ⁶⁹	VFASHQCN*DSYQDF (CT)	796.81 ²⁺	1591.63	1531.60	163.38
		N ¹⁸¹	VYSSGGTYSTPQCN*CTCK (T)	1036.92 ²⁺	2071.83	1897.76	63.92
			STPQCN*CTCKGGHW (CT)	565.89 ³⁺	1694.67	1520.60	92.86
		N ⁵²⁸	SITVDHSYQN*RTSGL (CT)	560.94 ³⁺	1679.82	1676.81	242.26
		N ⁶⁰	QCN*QSCRSLEDPDL (CT)	847.36 ²⁺	1692.71	1575.67	102.52
		N ⁷⁷²	YFN*QSSAGPGAIGSECKQSKTQDMHCY (CT)	1082.77 ³⁺	3245.31	3014.21	104.30
		N ⁸⁵⁵	QWN*CTDNPCK (T)	663.26 ³⁺	1324.52	1207.47	137.95
			N*CTDNPCKGTCTVY (CT)	846.83 ²⁺	1691.66	1517.59	94.29
		N ¹⁰³⁴	ITSTCSN*IN*MTDLQADQPFK (T)	1161.51 ²⁺	2321.01	2200.98	81.72
		N ¹⁰³⁶	ITSTCSN*IN*MTDLQADQPFK (T)	1161.51 ²⁺	2321.01	2200.98	175.21
O57579	aminopeptidase Ey		KITSTCSNIN*MTDL (CT)	800.88 ²⁺	1599.75	1539.72	140.08
		N ¹²¹⁹	TYPLN*ETIYSQTEGK (T)	924.44 ²⁺	1846.89	1843.88	149.30
		N ¹³⁰⁸	SIGN*ASQQL (CT)	649.29 ²⁺	1296.57	1179.53	65.50
		N ¹³⁷¹	FN*ESWDFGNQCIATCLGEENNIK (T)	1375.09 ²⁺	2748.18	2631.14	162.65
			FN*ESWDFGN*QIATCLGEENNIK (T)	918.06 ³⁺	2751.18	2631.14	76.19
		N ¹³⁷⁸	FN*ESWDFGN*QIATCLGEENNIK (T)	918.06 ³⁺	2751.18	2631.14	76.19
		N ¹⁴⁵²	EN*CTYVLVELIQPSSEK (T)	671.33 ³⁺	2010.99	1950.96	117.65
			HFKN*CTY (CT)	1101.46 ¹⁺	1100.46	1040.44	101.25
		N ¹⁷³²	LSRN*NTPVFEGCY (CT)	829.89 ²⁺	1657.78	1597.76	61.38
		N ¹⁹⁶⁴	KAPYDN*CTQY (CT)	631.77 ²⁺	1261.53	1201.51	69.98

Table 1. continued

no. in UniProt	protein name	position of glycosite	peptide (digestion enzyme, CT: chymotrypsin; T: trypsin)	m/z charge	exp. mass (Da) ^a	calc. mass (Da) ^b	score ^c
FINBL0	Mucin-6	N ¹¹⁰⁸	SYGSSVN*CTW (CT)	1163.46 ¹⁺	1162.46	1102.44	112.30
		N ¹¹³³	VNIEGCYN*CSHDEY (CT)	881.83 ²⁺	1761.66	1644.62	120.14
P01012	ovalbumin	N ³¹²	SSAN*LSGISSAEL (CT)	1412.66 ¹⁺	1411.66	1408.67	222.20
R9TNA6	ovalbumin-related protein X	N ¹¹¹	LSGITASKAN*Y (CT)	593.30 ²⁺	1184.60	1181.59	177.58
		N ³²⁶	IPSAN*LTGISSAEL (CT)	731.88 ²⁺	1461.76	1458.76	97.20
P01014	ovalbumin-related protein Y	N ⁹⁵	PN*ATYSLEADK (T)	662.83 ²⁺	1323.66	1320.66	342.42
			SEITRPN*ATY (CT)	577.78 ²⁺	1153.56	1150.56	279.86
		N ²¹⁵	EESKPVQMMCMN*NSFNVAITLPAEK (T)	925.42 ³⁺	2773.25	2697.23	95.14
		N ³¹²	SAN*LTGISSVDNLMISDAVHGVEVNEEGTEATGSTGAIGNIK (T)	1480.38 ³⁺	4438.13	4435.12	117.29
P10184	ovoinhibitor	N ⁶²	SRSAN*LTGISSVDNLMISDAVHGVE (CT)	865.10 ³⁺	2592.30	2589.29	210.31
		N ¹⁴¹	NLKPVCCTDGTSTYSN*EGGILYNR (T)	927.74 ³⁺	2780.23	2606.16	32.71
		N ³⁴	NAEHHTN*ISKL (CT)	633.82 ²⁺	1265.64	1262.64	202.00
P01005	ovomucoid		FPN*ATDKEGKDVLCNK (T)	646.66 ³⁺	1936.97	1876.94	138.22
			GAEVDCSRFPN*ATDKEGKDVLCNK (CT)	771.03 ³⁺	2310.09	2250.06	206.09
		N ⁶³	DLRPIQGTGDGVITYN*DCLLCAYSIEFGTN*ISK (T)	1220.56 ³⁺	3658.68	3481.62	56.66
		N ⁷⁷	DLRPIQGTGDGVITYN*DCLLCAYSIEFGTN*ISK (T)	1220.56 ³⁺	3658.68	3481.62	56.66
			SIEFGTN*ISKEHDGECKETVPMN*CSSY (CT)	1042.44 ³⁺	3124.33	3004.29	261.03
		N ⁹³	SIEFGTN*ISKEHDGECKETVPMN*CSSY (CT)	1042.44 ³⁺	3124.33	3004.29	261.03
		N ⁹⁹	ETVPMN*CSSYAN*TTSEDGK (T)	1048.92 ²⁺	2095.85	2032.84	179.09
			ETVPMN*CSSYAN*TTSEDGK (T)	1048.92 ²⁺	2095.85	2032.84	185.33
			AN*TTSEDGKVMVL (CT)	684.33 ²⁺	1366.67	1363.67	106.32
		N ¹¹⁶	AFN*PVCGTGDGVITYDN*ECLLCALCAHK (T)	882.05 ³⁺	2643.15	2469.08	46.58
		N ¹²⁸	AFN*PVCGTGDGVITYDN*ECLLCALCAHK (T)	882.05 ³⁺	2643.15	2469.08	35.75
		N ¹⁸²	CGSDN*KTY (CT)	474.19 ²⁺	946.37	886.35	148.14
		N ¹⁹⁹	CFNFCNAVESN*GTLTLSHFGK (T)	786.70 ³⁺	2357.09	2240.04	100.67
P20740	ovostatin		CNAVESN*GTLTLSHF (CT)	876.41 ²⁺	1750.82	1690.80	246.59
		N ⁴⁰³	VNNKNTHN*F (CT)	364.18 ³⁺	1089.53	1086.52	107.34
		N ³⁸⁸	TTSN*VSLVIEAANSF (CT)	813.91 ²⁺	1625.82	1622.82	123.76
		N ⁷⁵⁷	IWDIILN*STGKASVS (CT)	942.01 ²⁺	1882.01	1879.01	83.11
		N ¹¹⁴¹	ETASEKN*ITDIY (CT)	693.83 ²⁺	1385.66	1382.66	74.46
		N ¹³⁴⁷	SVQTSN*ASCPRDQPKF (CT)	941.44 ²⁺	1880.87	1820.85	116.82
Q4ADJ7	ovotransferrin	N ³⁸⁵	WSVVSN*GDVECTVVDTEK (T)	1013.96 ²⁺	2025.92	1965.90	99.02
		N ⁴⁹²	IHN*RTGTCTNFDEY (CT)	815.35 ²⁺	1628.69	1568.67	162.31
		N ⁶³⁷	GVN*GSEKSKF (CT)	528.27 ²⁺	1054.53	1051.53	85.36
		N ²⁶⁵	N*MTIPSM (CT)	909.43 ¹⁺	908.43	905.44	107.43
O42273	protein TENP	N ⁵³	AN*FTEQLAHSPKIVNSY (CT)	1061.54 ²⁺	2121.08	2118.07	156.85
P02752	riboflavin-binding protein	N ¹⁶⁴	CVPYSEMYAN*GTDMLCQSMWGESFK (T)	1441.06 ²⁺	2880.12	2763.08	158.37
			SEMYAN*GTDMLCQSMW (CT)	907.32 ²⁺	1812.65	1752.63	77.53
Q8JGM4	sulphydryl oxidase 1	N ²⁸⁸	LN*VTGSAIN*ETR (T)	640.83 ²⁺	1279.65	1273.66	142.83
		N ²⁹⁵	LN*VTGSAIN*ETR (T)	640.83 ²⁺	1279.65	1273.66	142.83
		N ³⁷¹	LRN*WTEPELPRLSAL (CT)	842.95 ²⁺	1683.90	1680.89	134.24
		N ⁴⁰¹	KEAVKNKEDASPAVLPTN*VTW (CT)	791.08 ³⁺	2370.25	2367.24	98.13
F1NCY6	TNFRSF6B	N ¹⁰⁵	QVEVQQCN*ATHNR (T)	793.86 ²⁺	1585.73	1525.71	90.73

Table 1. continued

no. in UniProt	protein name	position of glycosite	peptide (digestion enzyme, CT: chymotrypsin; T: trypsin)	m/z charge	exp. mass (Da) ^a	calc. mass (Da) ^b	score ^c
FINYJ8	BPI fold-containing family C protein	N ²⁴⁴	DLKGTVPVVGNG*HTDPPFPVAPF (CT)	1186.10 2+	2370.19	2367.19	48.97
R4GG81	death-associated protein kinase 3	N ²⁹⁷	SHSMPPN*NTYVN*FER (T)	943.41 2+	1884.82	1878.83	38.03

^aexp. mass, molecular weight of peptide based on the experimental data. ^bcalc. mass, molecular weight of peptide calculated theoretically, assuming there is no modification. ^cscore, N-linked glycosylation localization score calculated using MaxQuant. ^dN¹⁸O, conversion of Asn to (18O)Asp resulting from ¹⁸O/PNGase F treatment; ^eC, carbamidomethylated cysteines.

Although trypsin is the most commonly used enzyme tool in proteomic studies, chymotrypsin had a higher efficiency for the identification of *N*-glycosites in the present work, as many more *N*-glycosites were identified with the chymotrypsin digestion process (53) than with the trypsin digestion process (35). We had the same result in our previous ovomacroglobulin work.¹⁹ This phenomenon suggests that chymotrypsin maybe a more effective enzyme tool for the identification of *N*-glycosites.

Chicken Egg White Glycoproteins. Of the 26 identified egg white *N*-glycoproteins, approximately 46% (11) carried a single N-linked sugar chain (Table 1). Ovomucin (including Mucin-5B and Mucin-6) was the most heavily *N*-glycosylated protein in egg white, with 16 *N*-glycosites, followed by ovomucoid, with 9 *N*-glycosites (Figure 3).

Ovomucin. As the most heavily glycosylated protein in CEW, ovomucin is composed of two subunits, α -ovomucin (Mucin-5B) and β -ovomucin (Mucin-6).²⁴ Mucin-5B is mainly *N*-glycosylated and contains 24 predicted *N*-glycosites, 18 of which have been identified in previous research.¹⁷ Here, we identified 15 *N*-glycosites from Mucin-5B; 11 of them are included in Offengenden et al.'s work, while 4 of them are newly identified. Among the 15 sites, N⁶⁹ (NDS) was found unoccupied in Offengenden et al.'s work, but here it was identified as *N*-glycosylated with a Mascot Score of 163, indicating that this site has both glycosylated and non-glycosylated forms. Another two sites, N¹⁰³⁴ (NIN) and N¹³⁷⁸ (NCQ), are localized at the noncanonical sequence with Mascot Scores of 82 and 76, respectively, suggesting that they are nonconserved sites.

Conversely, Mucin-6 contains approximately 60% glycans, which are mostly O-glycosylated.²⁴ Even so, Mucin-6 has 11 predicted *N*-glycosites because of its long sequence (1185 AA). In Offengenden et al.'s work, N²²³ and N⁹³⁰ were occupied with *N*-glycan.¹⁷ Here, another two sites, N¹¹⁰⁸ and N¹¹³³, were identified as being *N*-glycosylated.

Glycans occupy the *N*-glycosites of ovomucin and are involved in its biological activities. For example, the carbohydrate moiety of ovomucin can interact with Mg²⁺ ions and result in a strengthening of its antiviral activity.⁸ Recently, Sun et al.'s work demonstrated that ovomucin-derived glycopeptides could disturb the adhesion of *E. coli* to porcine erythrocytes and thus have the potential to resist infectious diseases.³ All of these results implied that the glycans of ovomucin play key roles in the protein's bioactivities and indicated that the glycosylation structure of ovomucin needs to be further investigated. Therefore, the newly discovered glycosylation sites will be helpful in clarifying the glycosylation structure of ovomucin.

Ovomucoid. Another highly glycosylated egg white protein, ovomucoid, has been mostly investigated. According to sequence analysis, ovomucoid contains 6 potential *N*-glycosites.²⁵ A total of 9 asparagine residues were covalently attached to glycan groups in the present work, including 5 canonical sites that were confirmed previously and another canonical site (N¹⁸²) and 3 noncanonical sites (N⁶³, N¹¹⁶, and N¹²⁸) that were identified for the first time. Ovomucoid is the major allergen of egg white, and the glycans play important roles in the protein's structure and allergenicity,²⁶ suggesting that the location of glycosites may exert certain effects. Therefore, the newly identified sites and the linked *N*-glycans of ovomucoid should be investigated and characterized.

Lysozyme. Although CEW lysozyme does not contain the consensus sequence motif NX(S/T) in its sequence, there is

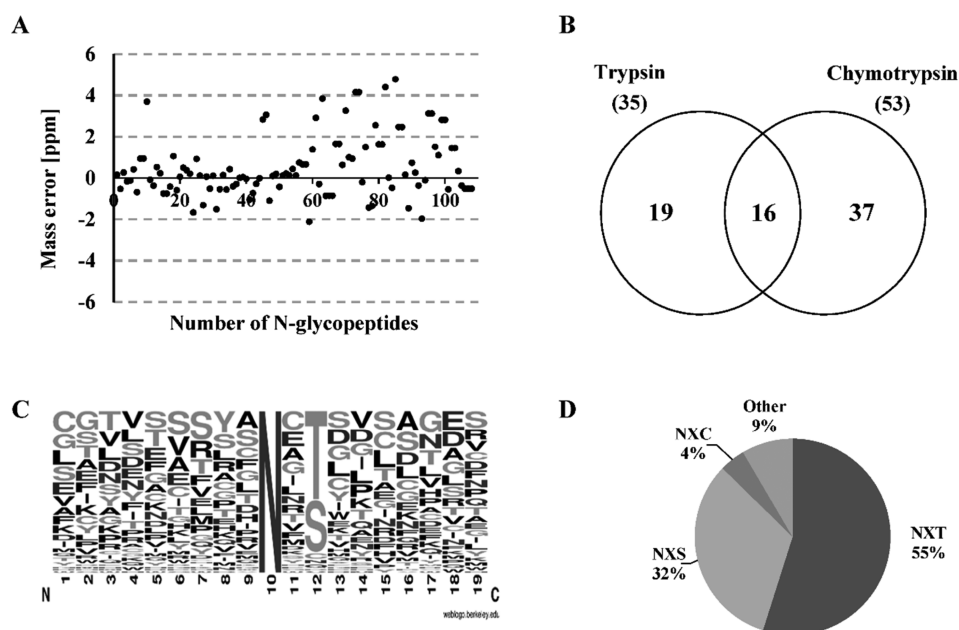


Figure 2. Common characteristics of the identified *N*-glycosites. A, mass error distribution of identified *N*-glycopeptides; B, the numbers of *N*-glycosites identified using trypsin and chymotrypsin; C, *N*-glycosylation sequences as derived using Web-Logo; D, distribution of recognized sequence motifs.

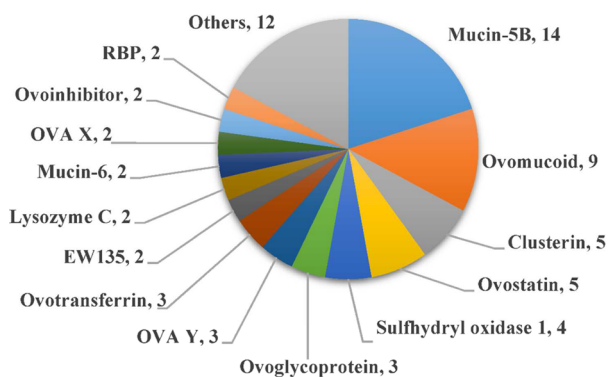


Figure 3. Distribution of identified *N*-glycosylation sites in CEW proteins.

undeniable evidence to demonstrate the existence of an *N*-glycosylated form of lysozyme at a low abundance level. Recently, Asperger et al. identified the *N*-glycosites of lysozyme for the first time using an optimized analytical approach and a novel data interpretation workflow.²⁷ The two newly discovered glycosylation sites, N³⁹ and N⁴⁴, are both localized at a nonconsensus sequon (NXN, NXQ). Surprisingly, these two low-abundance and unconventional *N*-glycosites of lysozyme also formed three different *N*-glycopeptides in the present work. In addition to the trypsin-derived *N*-glycopeptide FESNFNTQATN*R, which was identified in Asperger et al.'s work, another two new *N*-glycopeptides, ESNFN*T-QATNRNTDGSTDY and NTQATN*RNTDGSTDY, were identified in our mass spectrometry data. All three of the identified glycopeptides included both N³⁹ and N⁴⁴, but either N³⁹ or N⁴⁴ was *N*-glycosylated, suggesting that it is not common for glycosylation to occur at both adjacent sites.

Clusterin. Clusterin is a widely expressed secretory glycoprotein that is found in mammals and other vertebrates.²⁸ CEW clusterin is a disulfide-linked heterodimer of an alpha subunit (sequence from residues 232–448) and a beta subunit

(sequence from residues 21–231).²⁹ This protein is considered a glycoprotein based on the sequence analysis but has not yet been verified. In the present work, 5 potential *N*-glycosites were experimentally confirmed to be occupied by oligosaccharide chains, indicating for the first time that CEW clusterin is a glycoprotein.

Mammalian clusterin is a multifunctional glycoprotein that is distributed in ubiquitous tissues and is involved in many important physiological processes ranging from organ development to Alzheimer's disease.³⁰ The biological functions of clusterin should be due to its interaction with a wide range of molecules, such as lipids, immunoglobulins, and amyloid proteins.²⁸ The *N*-glycans of clusterin, as the most variable part of the molecule, may have important roles in the interactions with its target molecules. CEW clusterin is thought to interact with other CEW proteins and to stabilize them during the development of the embryo or storage of fresh eggs.^{29,31} Therefore, the glycosylation structure of CEW clusterin needs to be further characterized.

Ovalbumin-Related Proteins X and Y. Ovalbumin-related proteins X (OVA-X) and Y (OVA-Y) are glycoproteins, but details about their glycosylation are lacking.³² The alignment of OVA-X and OVA-Y revealed that the sequence identities of these two proteins were as high as 71% and that both of them contained 4 potential *N*-glycosites. Furthermore, these predicted *N*-glycosites were all located at identical positions and with the same sequences, except that the N¹¹² (NYS) of OVA-X differed from the N⁹⁵ (NYT) of OVA-Y. Here, 2 *N*-glycosites of OVA-X and 3 *N*-glycosites of OVA-Y were founded in the MS/MS analysis.

Other CEW Glycoproteins. Other well-known CEW glycoproteins, such as ovalbumin, riboflavin-binding protein (RBP, also known as ovoflavoprotein), and avidin, were also predictably identified as being *N*-glycosylated, and their acknowledged glycosylation sites were all reproduced in the present HPLC-MS/MS results (Table 1). Ovoinhibitor and α -acid glycoprotein (ovoglycoprotein) are also known as

glycoproteins, but their glycosites have not yet been reported. In this study, 2 *N*-glycosites on ovalbumin and 3 *N*-glycosites on ovoglobulin were experimentally verified for the first time. Ovomacroglobulin, as the second largest protein in egg white, contains up to 13 potential *N*-glycosites according to sequence analysis, and 12 of them had been confirmed experimentally in our previous work. Here, 5 sites were again confirmed to be *N*-glycosylated.

Low-abundance CEW proteins are not easily detected in MS/MS analysis because of their extremely low content and the interference from high-abundance CEW proteins.³³ Nevertheless, a number of low-abundance CEW proteins were experimentally confirmed as glycoproteins for the first time in the present study. These proteins included some enzymes in CEW, such as sulfhydryl oxidase 1 and aminopeptidase Ey; several proteins that are transferred from chicken serum, such as apolipoprotein D, hemopexin, EW135, and TENP; and some subunits of immunoglobulin, such as the Ig mu chain C region, lambda chain C region and J chain.

Heterogeneity of CEW Clusterin. Glycoproteins usually exist in multiple glycoforms, and the heterogeneity of glycosylation likely affects the properties of glycoproteins.³⁴ Here, the heterogeneity of CEW glycoproteins was demonstrated using clusterin as a representative.

After expansion in a 2-DE gel, the clusterin-rich CEW fraction showed two series of spots near the molecular weights of 35 and 37 kDa (Figure 4). These spots were further

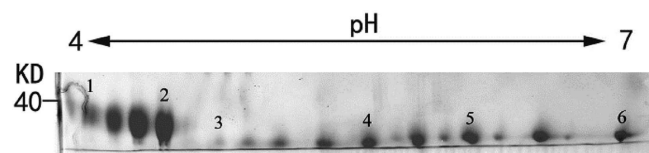


Figure 4. Visual illustration of the heterogeneity of clusterin glycosylation using 2-DE. Clusterin was separated by 2-DE using 24 cm (pH 4–7) IEF strips and 10% SDS-PAGE gels, and the labeled spots were identified using MALDI-TOF MS/MS.

identified using MALDI-TOF MS/MS analysis, and the results confirmed that spots 1 and 2 were the alpha subunit of clusterin, while spots 3–6 were the beta subunit (Table 2). Although the theoretical MWs of these two subunits are similar, the experimental MWs were different by 2–3 kDa. Coincidentally, the average molecular weight of an *N*-linked glycan is approximately 2.5 kDa,³⁵ implying that there may be one more *N*-glycan on the alpha subunit than on the beta subunit of CEW clusterin. The *pI* values of the alpha subunit, spots 1 and 2, ranged from 4.54 to 4.85, while the *pI* values of the beta subunit, spots 3 to 6, ranged from 5.06 to 6.68. The wide range of *pI* values indicated that the oligosaccharide chains

of clusterin are varied. It was speculated that the more acidic isoforms of the alpha subunit (*pI* = 4.54–4.85) would contain *N*-glycans with higher sialic acid contents. From the 2-DE results, it can be inferred that CEW clusterin has a high degree of heterogeneity in its *N*-glycan profiles.

In summary, this study is the first attempt of the analysis of the CEW glycoproteome, with the goal of trying to identify as many *N*-glycosites as possible. Many glycoproteins and *N*-glycosites were experimentally identified and confirmed for the first time. The *N*-glycosylated modification information is extremely important for understanding the structures, functions, bioactivities, and allergenicities of CEW glycoproteins. Nevertheless, *N*-glycosites are only one aspect of the glycosylation structure, and other glycosylation information, such as the composition and structure of *N*-glycans, heterogeneity of glycosylation forms, and integrated glycopeptide structure, need to be further investigated.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jafc.7b01706.

MS/MS spectra of peptides containing deglycosylated asparagine (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

*Fax: +86 28 84616063. E-mail: liudy1014@163.com.

*Fax: +86-27-87283177. E-mail: mameihuhn@163.com.

ORCID

Fang Geng: 0000-0002-8275-6770

Funding

This research was supported by the National Natural Science Foundation of China (Nos. 31601490 and 31230058).

Notes

The authors declare no competing financial interest.

■ ABBREVIATIONS USED

2-DE, two-dimensional electrophoresis; ACN, acetonitrile; CEW, chicken egg white; FDR, false discovery rate; HID, high-energy collision-induced decomposition; IEF, isoelectric focusing; MALDI-TOF, matrix-assisted laser desorption ionization-time-of-flight; MW, molecular weight; MWCO, molecular weight cutoff; PEG, polyethylene glycol; OVA-X, ovalbumin-related protein X; OVA-Y, ovalbumin-related protein Y; *pI*, isoelectric point; RBP, riboflavin-binding protein

■ REFERENCES

(1) Raman, R.; Raguram, S.; Venkataraman, G.; Paulson, J. C.; Sasisekharan, R. Glycomics: an integrated systems approach to

Table 2. Identification of Protein Spots in 2-DE by MALDI-TOF MS/MS

spot	gi no.	protein	score ^a	matched peptides/sequence coverage (%)	exptl <i>pI</i> /MW (kDa)	theor <i>pI</i> /MW (kDa)
1	gil45382467	clusterin alpha subunit	223	6/16	4.54/37.9	4.90/24.9
2	gil45382467	clusterin alpha subunit	216	5/15	4.85/36.5	4.90/24.9
3	gil45382467	clusterin beta subunit	254	11/22	5.06/34.5	7.08/24.5
4	gil45382467	clusterin beta subunit	127	8/19	5.68/34.6	7.08/24.5
5	gil45382467	clusterin beta subunit	114	8/13	6.10/34.9	7.08/24.5
6	gil45382467	clusterin beta subunit	136	10/22	6.68/35.2	7.08/24.5

^aProtein score is $-10 \log(P)$, where P is the probability that the observed match is a random event.

structure-function relationships of glycans. *Nat. Methods* **2005**, *2*, 817–824.

(2) Arnold, J. N.; Wormald, M. R.; Sim, R. B.; Rudd, P. M.; Dwek, R. A. The impact of glycosylation on the biological function and structure of human immunoglobulins. *Annu. Rev. Immunol.* **2007**, *25*, 21–50.

(3) Sun, X.; Gänzle, M. G.; Wu, J. Identification and Characterization of Glycopeptides from Egg Protein Ovomucin with Anti-Agglutinating Activity against Porcine K88 Enterotoxigenic *Escherichia coli* Strains. *J. Agric. Food Chem.* **2017**, *65*, 777.

(4) Offengenden, M.; Wu, J. Egg white ovomucin gels: structured fluids with weak polyelectrolyte properties. *RSC Adv.* **2013**, *3*, 910–917.

(5) Mine, Y.; Yang, M. Recent advances in the understanding of egg allergens: basic, industrial, and clinical perspectives. *J. Agric. Food Chem.* **2008**, *56*, 4874–4900.

(6) Benede, S.; Lopez-Exposito, I.; Molina, E.; Lopez-Fandino, R. Egg proteins as allergens and the effects of the food matrix and processing. *Food Funct.* **2015**, *6*, 694–713.

(7) KATO, A.; OGINO, K.; KURAMOTO, Y.; KOBAYASHI, K. DEGRADATION OF THE O-GLYCOSIDICALLY LINKED CARBOHYDRATE UNITS OF OVOMUCIN DURING EGG WHITE THINNING. *J. Food Sci.* **1979**, *44*, 1341–1344.

(8) Shan, Y.; Xu, Q.; Ma, M. Mg²⁺ binding affects the structure and activity of ovomucin. *Int. J. Biol. Macromol.* **2014**, *70*, 230–235.

(9) Rupa, P.; Nakamura, S.; Katayama, S.; Mine, Y. Effects of ovalbumin glycoconjugates on alleviation of orally induced egg allergy in mice via dendritic-cell maturation and T-cell activation. *Mol. Nutr. Food Res.* **2014**, *58*, 405–417.

(10) Rupa, P.; Nakamura, S.; Katayama, S.; Mine, Y. Attenuation of Allergic Immune Response Phenotype by Mannosylated Egg White in Orally Induced Allergy in Balb/c Mice. *J. Agric. Food Chem.* **2014**, *62*, 9479–9487.

(11) Mine, Y.; D'Silva, I. Bioactive components in egg white. In *Egg bioscience and biotechnology*; Mine, Y., Ed.; John Wiley & Sons, Inc.: Hoboken, NJ, 2008; pp 141–184.

(12) Geng, F.; Jin, Y. Chicken Egg White Glycoproteins and Its Allergenic Mechanism: A Review. *China Poultry* **2016**, *38*, 1–8.

(13) Qiu, N.; Ma, M.; Cai, Z.; Jin, Y.; Huang, X.; Huang, Q.; Sun, S. Proteomic analysis of egg white proteins during the early phase of embryonic development. *J. Proteomics* **2012**, *75*, 1895–1905.

(14) Qiu, N.; Ma, M.; Zhao, L.; Liu, W.; Li, Y.; Mine, Y. Comparative Proteomic Analysis of Egg White Proteins under Various Storage Temperatures. *J. Agric. Food Chem.* **2012**, *60*, 7746–7753.

(15) Zielinska, D. F.; Gnad, F.; Wiśniewski, J. R.; Mann, M. Precision mapping of an in vivo N-glycoproteome reveals rigid topological and sequence constraints. *Cell* **2010**, *141*, 897–907.

(16) Zielinska, D. F.; Gnad, F.; Schropp, K.; Wiśniewski, J. R.; Mann, M. Mapping N-glycosylation sites across seven evolutionarily distant species reveals a divergent substrate proteome despite a common core machinery. *Mol. Cell* **2012**, *46*, 542–548.

(17) Offengenden, M.; Fentabil, M. A.; Wu, J. N-glycosylation of ovomucin from hen egg white. *Glycoconjugate J.* **2011**, *28*, 113–123.

(18) Jiang, K.; Wang, C. J.; Sun, Y. J.; Liu, Y.; Zhang, Y.; Huang, L. J.; Wang, Z. F. Comparison of Chicken and Pheasant Ovotransferrin N-Glycoforms via Electrospray Ionization Mass Spectrometry and Liquid Chromatography Coupled with Mass Spectrometry. *J. Agric. Food Chem.* **2014**, *62*, 7245–7254.

(19) Geng, F.; Huang, X.; Majumder, K.; Zhu, Z.; Cai, Z.; Ma, M. Mass spectrometry and two-dimensional electrophoresis to characterize the glycosylation of hen egg white ovomacroglobulin. *J. Agric. Food Chem.* **2015**, *63*, 8209–8215.

(20) Li, S.; Geng, F.; Wang, P.; Lu, J.; Ma, M. Proteome analysis of the almond kernel (*Prunus dulcis*). *J. Sci. Food Agric.* **2016**, *96*, 3351.

(21) Geng, F.; Huang, X.; Yan, N.; Jia, L.; Ma, M. Purification of hen egg white ovomacroglobulin using one-step chromatography. *J. Sep. Sci.* **2013**, *36*, 3717–3722.

(22) Hu, S.; Qiu, N.; Liu, Y.; Zhao, H.; Gao, D.; Song, R.; Ma, M. Identification and comparative proteomic study of quail and duck egg white protein using 2-dimensional gel electrophoresis and matrix-

assisted laser desorption/ionization time-of-flight tandem mass spectrometry analysis. *Poult. Sci.* **2016**, *95*, 1137.

(23) Crooks, G. E.; Hon, G.; Chandonia, J.-M.; Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **2004**, *14*, 1188–1190.

(24) Robinson, D.; Monsey, J. Studies on the composition of egg-white ovomucin. *Biochem. J.* **1971**, *121*, 537–547.

(25) Zhu, F.; Trinidad, J. C.; Clemmer, D. E. Glycopeptide site heterogeneity and structural diversity determined by combined lectin affinity chromatography/IMS/CID/MS techniques. *J. Am. Soc. Mass Spectrom.* **2015**, *26*, 1092–1102.

(26) Benedé, S.; López-Fandiño, R.; Reche, M.; Molina, E.; López-Expósito, I. Influence of the carbohydrate moieties on the immunoreactivity and digestibility of the egg allergen ovomucoid. *PLoS One* **2013**, *8*, e80810.

(27) Asperger, A.; Marx, K.; Albers, C.; Molin, L.; Pinato, O. Low abundant N-linked glycosylation in hen egg white lysozyme is localized at nonconsensus sites. *J. Proteome Res.* **2015**, *14*, 2633–2641.

(28) Jones, S. E.; Jomary, C. Clusterin. *Int. J. Biochem. Cell Biol.* **2002**, *34*, 427–431.

(29) Mann, K.; Gautron, J.; Nys, Y.; McKee, M. D.; Bajari, T.; Schneider, W. J.; Hincke, M. T. Disulfide-linked heterodimeric clusterin is a component of the chicken eggshell matrix and egg white. *Matrix Biol.* **2003**, *22*, 397–407.

(30) Araki, S.; Israel, S.; Leskov, K.; Criswell, T.; Beman, M.; Klovov, D.; Sampath, L.; Reinicke, K.; Cataldo, E.; Mayo, L. Clusterin proteins: stress-inducible polypeptides with proposed functions in multiple organ dysfunction. *Br. J. Radiol.* **2005**, *27*, 106.

(31) Makkar, S.; Liyanage, R.; Kannan, L.; Packialakshmi, B.; Lay, J. O., Jr.; Rath, N. C. Chicken egg shell membrane associated proteins and peptides. *J. Agric. Food Chem.* **2015**, *63*, 9888–9898.

(32) Réhault-Godbert, S.; Labas, V.; Helloin, E.; Hervé-Grépinet, V.; Slugocki, C.; Berges, M.; Bourin, M.-C.; Brionne, A.; Poirier, J.-C.; Gautron, J. Ovalbumin-related protein X is a heparin-binding ov-serpin exhibiting antimicrobial activities. *J. Biol. Chem.* **2013**, *288*, 17285–17295.

(33) Liu, Y.; Qiu, N.; Ma, M. Comparative proteomic analysis of egg white proteins during the rapid embryonic growth period by combinatorial peptide ligand libraries. *Poult. Sci.* **2015**, *94*, 2495–2505.

(34) Bousfield, G. R.; Butnev, V. Y.; Rueda-Santos, M. A.; Brown, A.; Hall, A. S.; Harvey, D. J. Macro- and micro-heterogeneity in pituitary and urinary follicle-stimulating hormone glycosylation. *J. Glycomics Lipidomics* **2014**, *4*, 10.4172/2153-0637.1000125

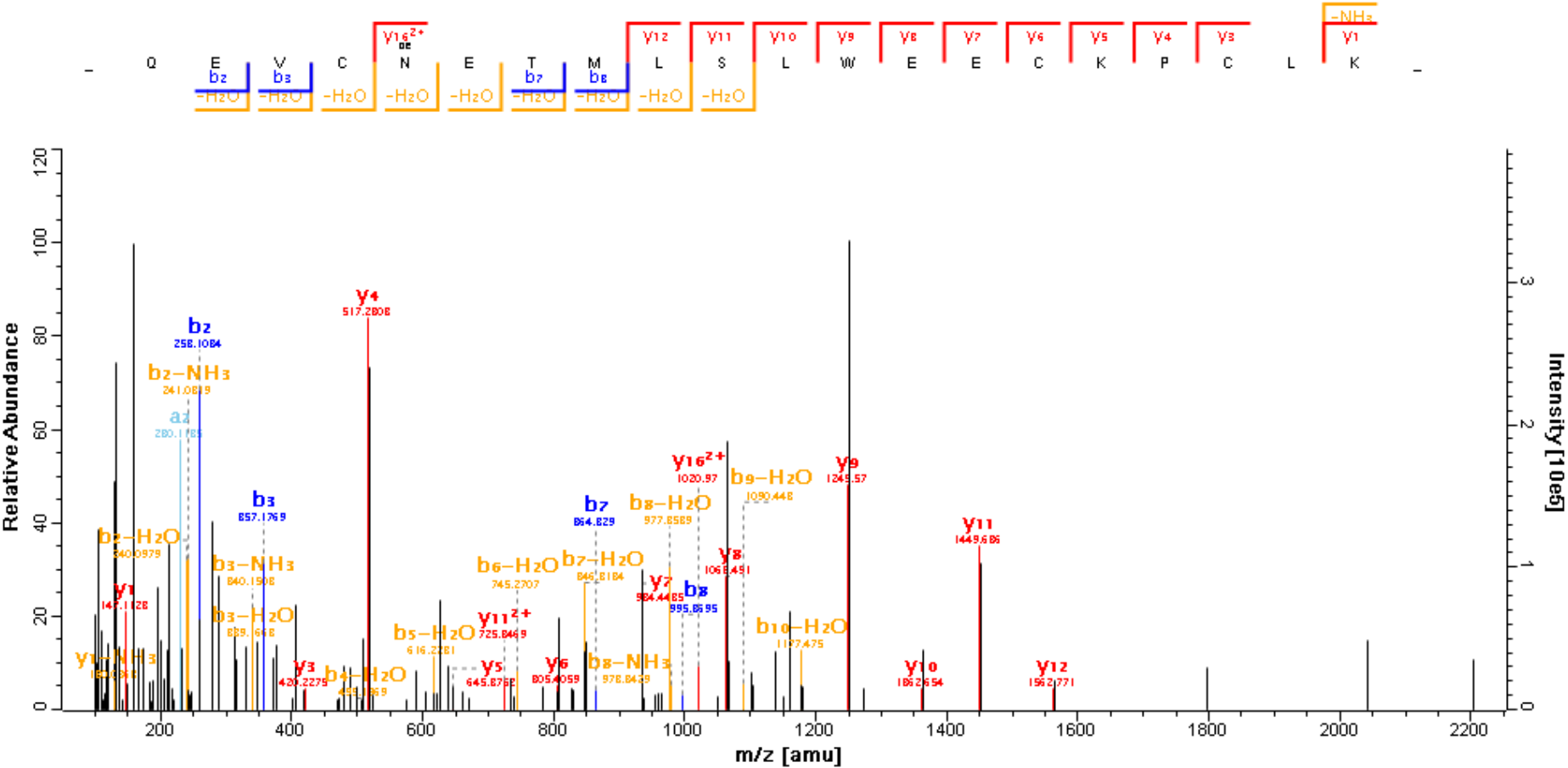
(35) Sunryd, J. C.; Cheon, B.; Graham, J. B.; Giorda, K. M.; Fissore, R. A.; Hebert, D. N. TMTC1 and TMTC2 are novel endoplasmic reticulum tetratricopeptide repeat-containing adapter proteins involved in calcium homeostasis. *J. Biol. Chem.* **2014**, *289*, 16085–16099.

Clusterin

N⁹⁹

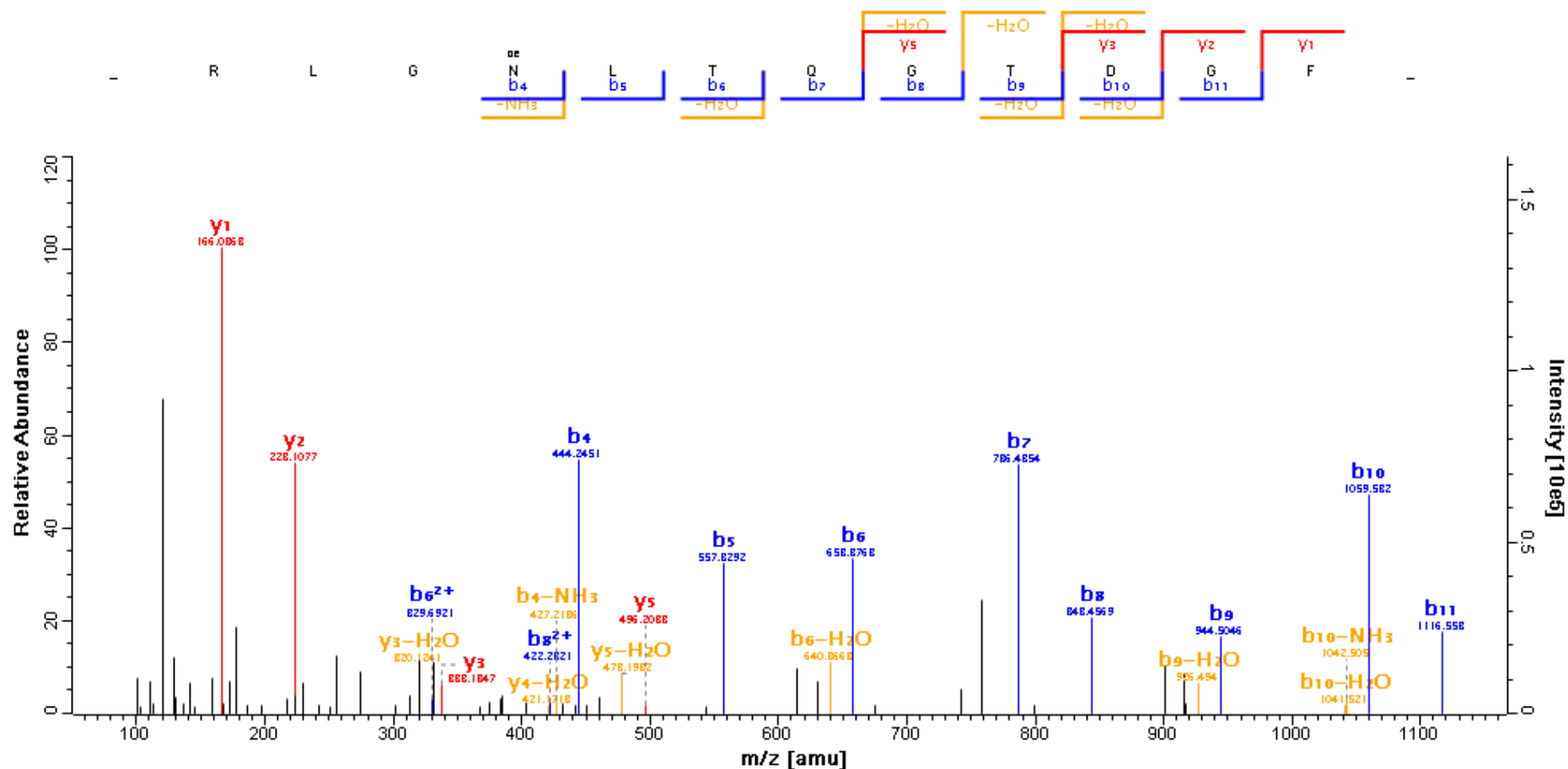
QEVCN*ETMLSLWEECKPCLK

Scan number 33066 Raw file P16086
Method FTMS; HCD Pepti... 92.68



RLGN*LTQGT DGF

Raw file	P16086
Peptide	124.42

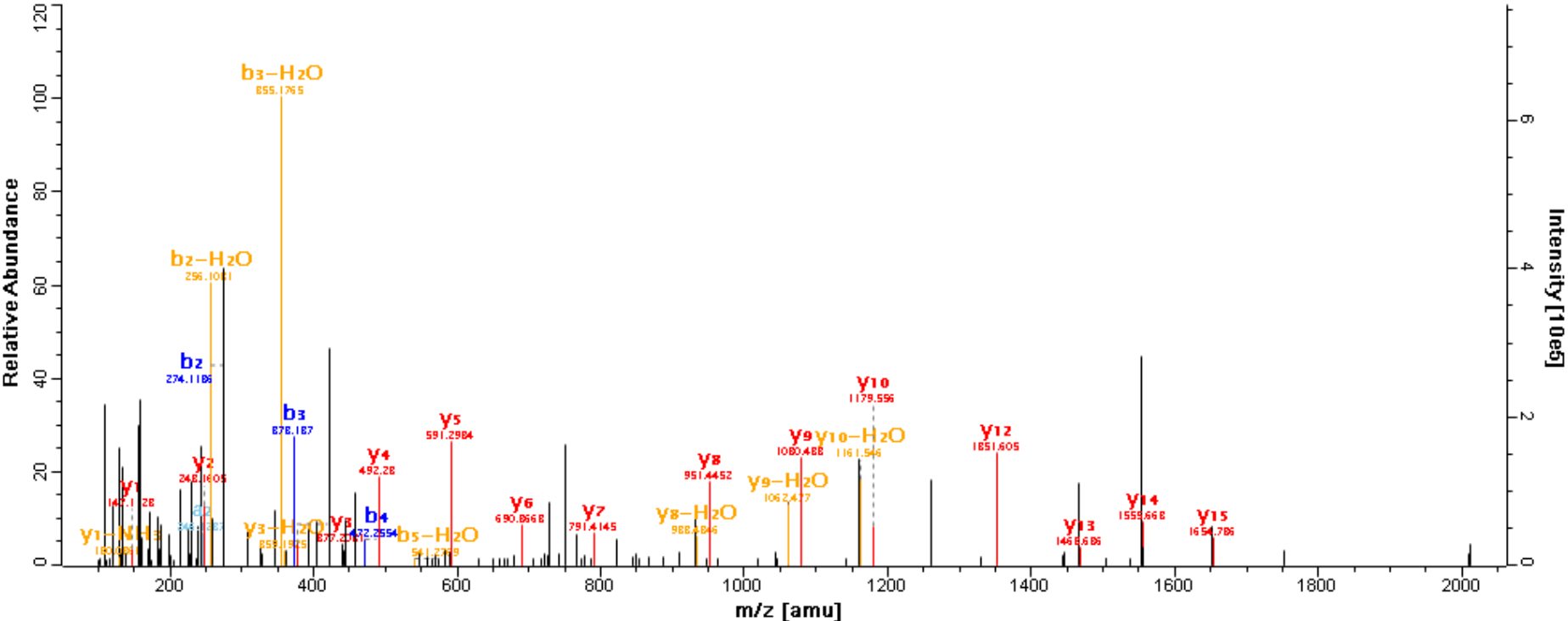
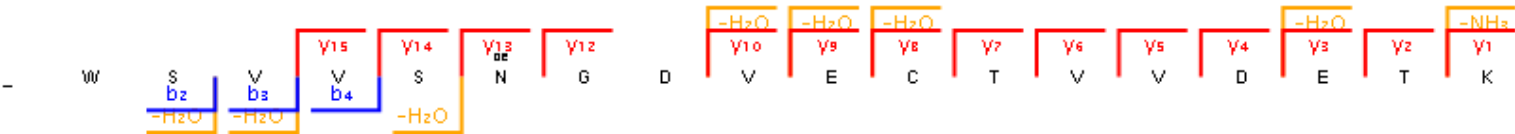


Ovotransferrin

N³⁸⁵

WSVVSN*GDVECTVVDETK

Scan number 27430 Raw file P16086
Method FTMS; HCD Pepti... 99.02

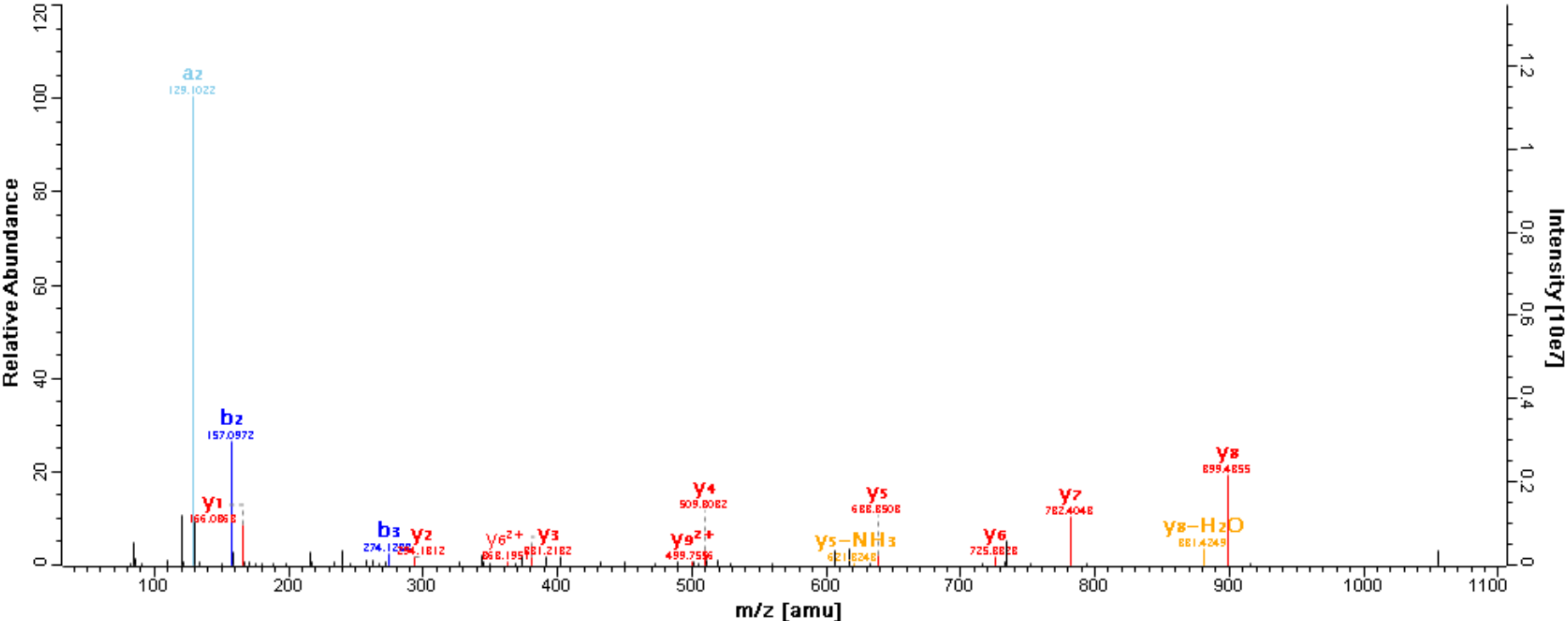
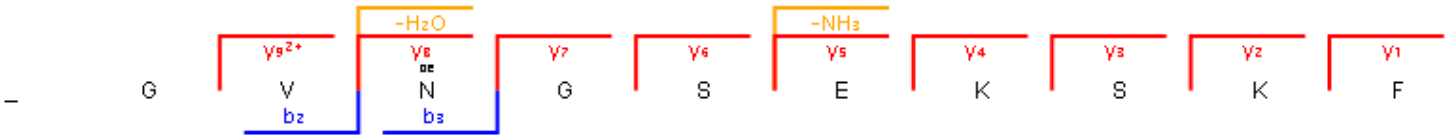


Ovotransferrin

N⁶³⁷

GVN*GSEKSKF

Scan number 3730 Raw file P16086
Method FTMS; HCD Pepti... 85.35

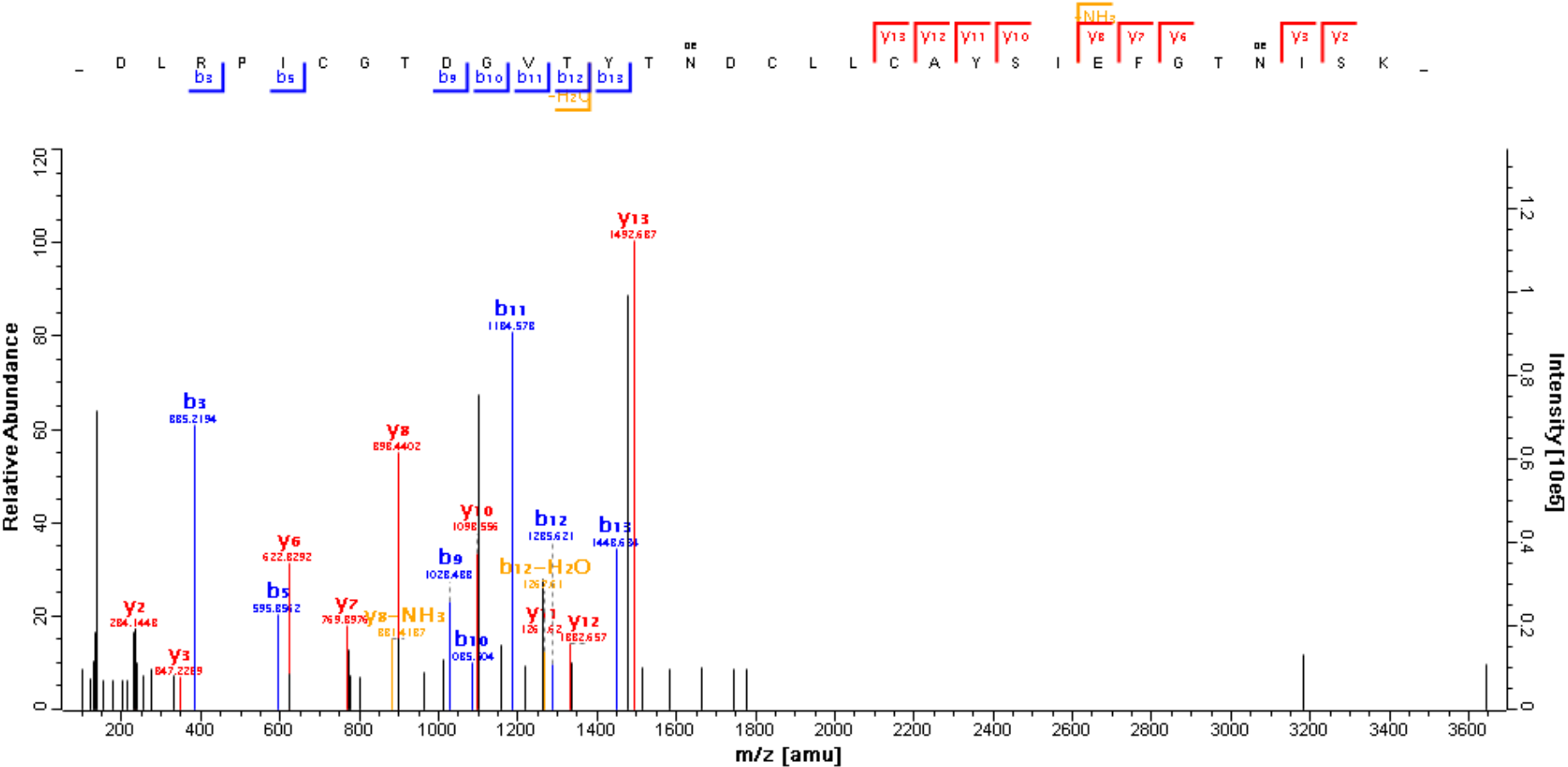


Ovomucoid

N⁶³ and N⁷⁷

DLRPICGTDGVTYTN*DCLLCAYSIEFGTN*ISK

Scan number 39569 Raw file P16086
Method FTMS; HCD Pepti... 56.66

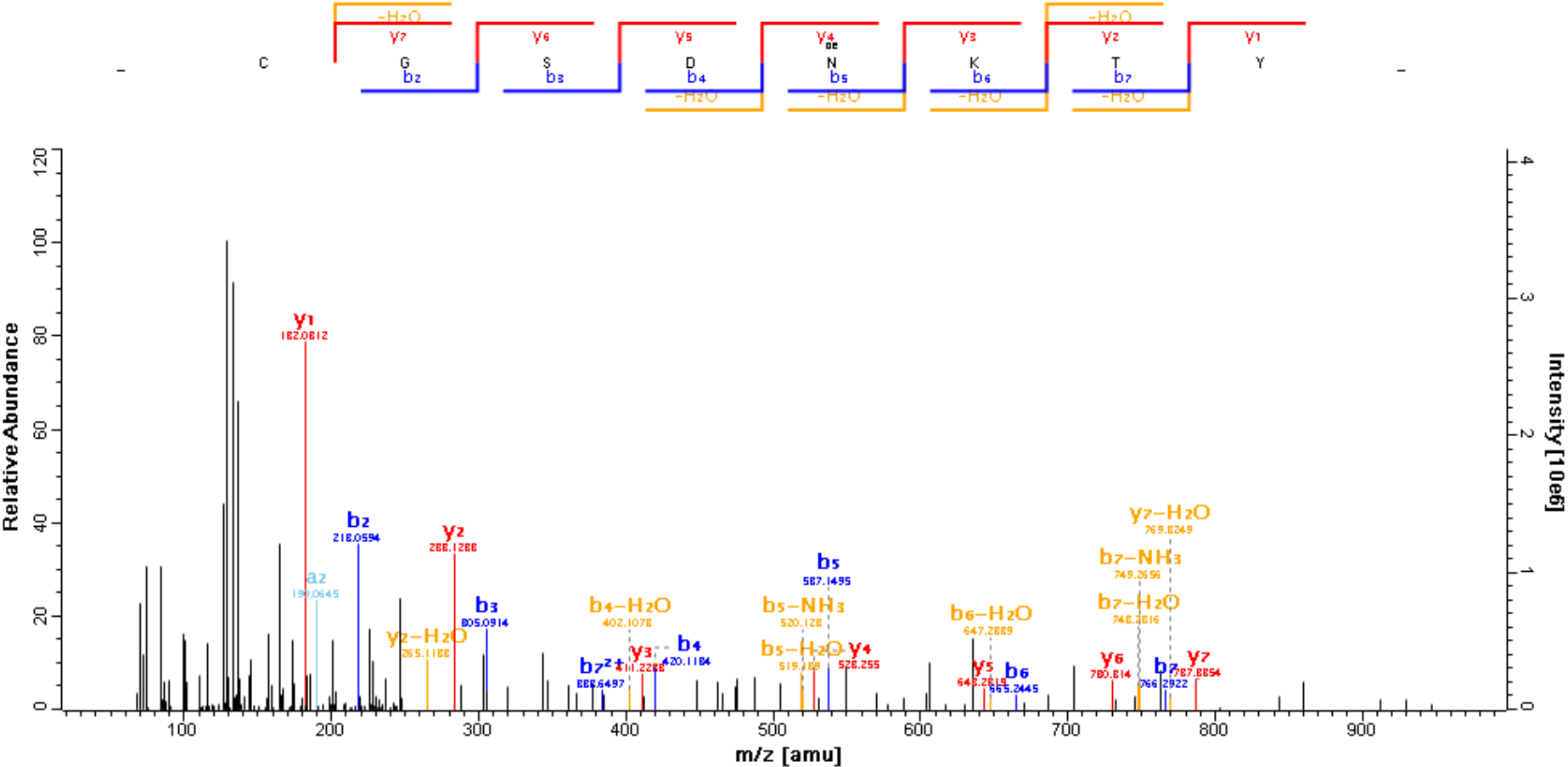


Ovomucoid

N¹⁸²

CGSDN*KTY

Scan number 2572 Raw file P16086
Method FTMS; HCD Pepti... 148.14

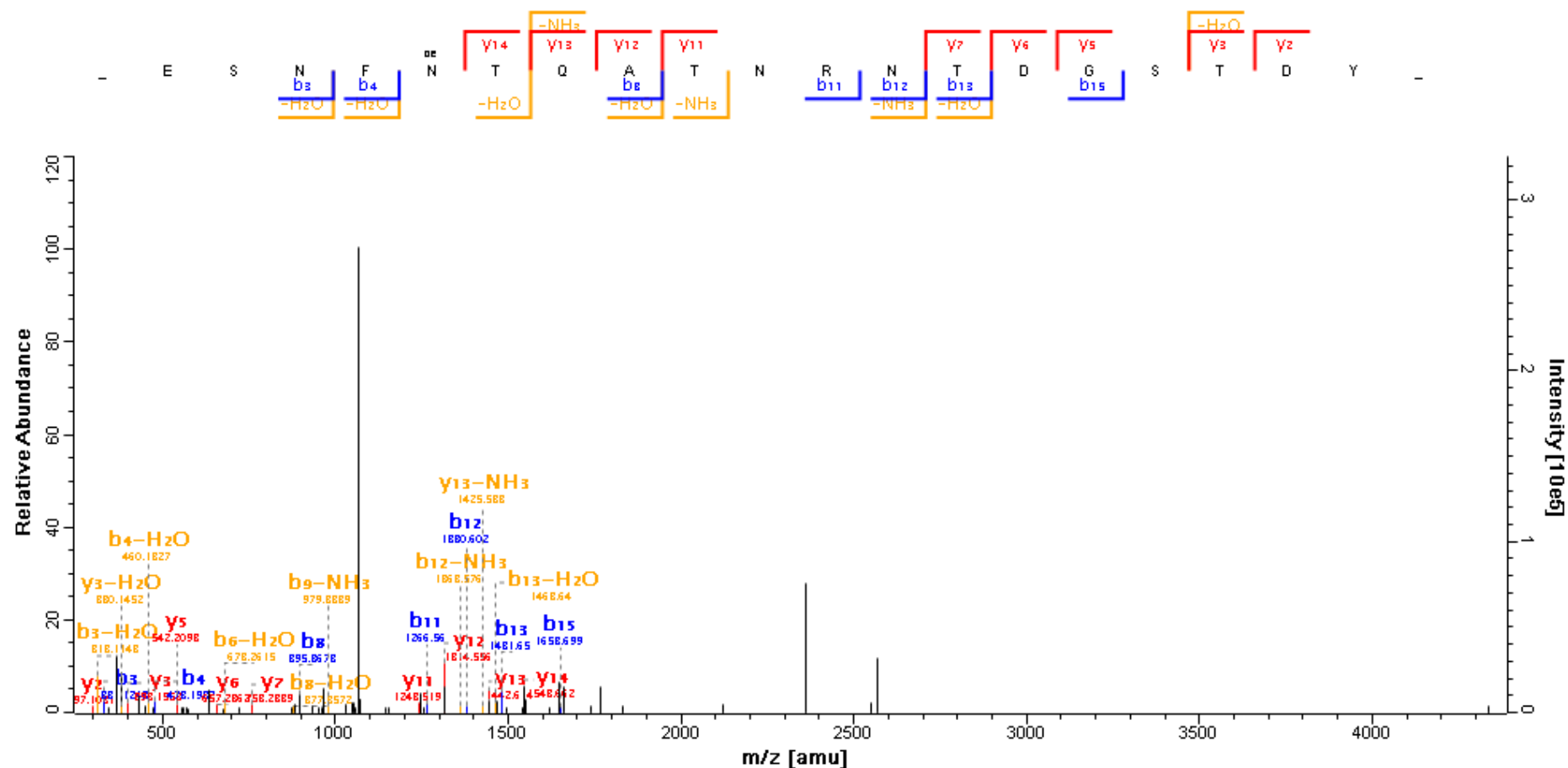


Lysozyme

N⁵⁷

ESNFN*TQATNRNTDGSTDY

Scan number 9187 Raw file P16086
Method FTMS; HCD Peptide... 94.39

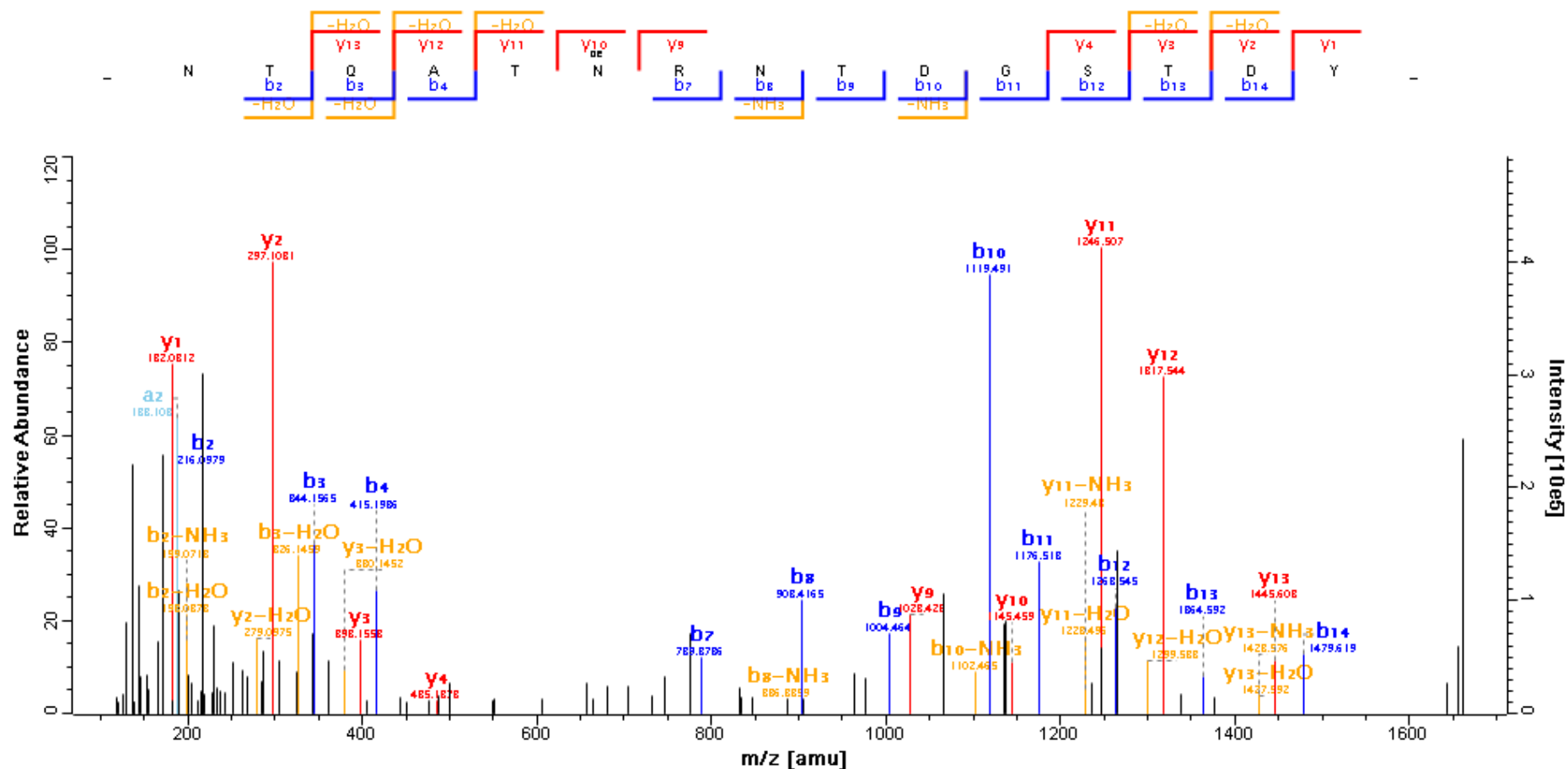


Lysozyme

N⁶²

NTQATN*RNTDGSTDY

Scan number 3769 Raw file P16086
Method FTMS; HCD Pepti... 159.18

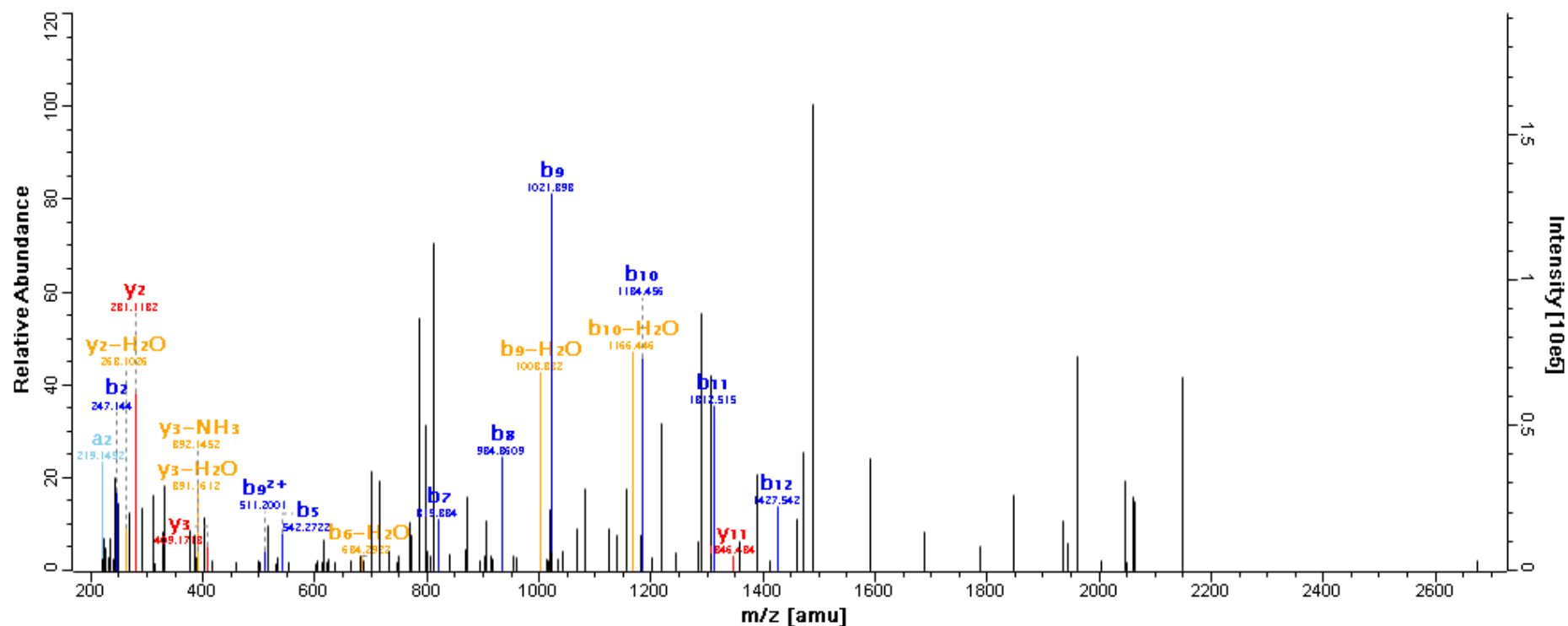
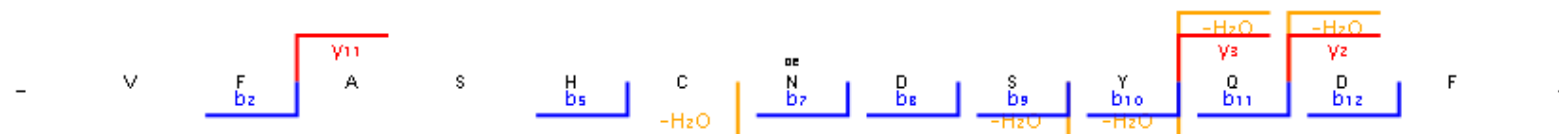


Mucin-5B

N⁶⁹

VFASHCN*DSYQDF

Scan number 15421 Raw file P16086
Method FTMS; HCD Pepti... 80.31



Mucin-5B

N¹³⁰⁸

SICN*ASCQIEL

Scan number 19844 Raw file P16086
Method FTMS; HCD Pepti... 65.5

