Anal. Chem. **2003,** *75,* 5628−5637

Determination of N-Glycosylation Sites and Site Heterogeneity in Glycoproteins

Hyun Joo An,[†] Thomas R. Peavy,[‡] Jerry L. Hedrick,[‡] and Carlito B. Lebrilla*[,†]

Department of Chemistry and Section of Molecular and Cell Biology, University of California, Davis, California 95616

An approach for the characterization of glycosylation sites and oligosaccharide heterogeneity in glycoproteins based on a combination of nonspecific proteolysis, deglycosylation, and matrix-assisted laser desorption/ionization Fourier transform mass spectrometry (MALDI-FT MS) is described. Glycoproteins were digested with Pronase yielding primarily glycopeptides and amino acids. Nonglycosylated peptide fragments were susceptible to complete Pronase digestion to their constituent amino acids. Steric hindrance prohibited the digestion of the peptide moiety attached to the glycan. Glycopeptides were desalted and concentrated using solid-phase extraction and analyzed by MALDI MS. The oligosaccharides were also analyzed by MALDI MS after releasing the glycans from glycoproteins using PNGase F. The peptide moiety of the glycopeptides was identified by subtracting the masses of the glycans derived from PNGase F treatment from the masses of the glycopeptides. The experimental strategy was validated using glycoproteins with known oligosaccharide structures, ribonuclease B and chicken ovalbumin. This procedure was then used to determine the N-glycosylation sites and site heterogeneity of a glycoprotein whose glycosylation pattern was unknown, namely, the *Xenopus laevis* egg cortical granule lectin. This procedure is useful for determining protein site heterogeneity and structural heterogeneities of the oligosaccharide moiety of glycoproteins.

Chart 1

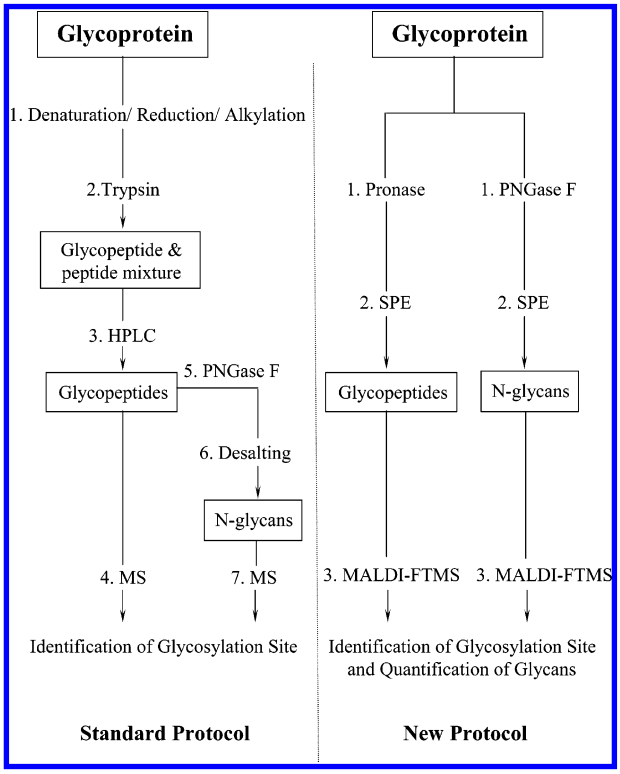
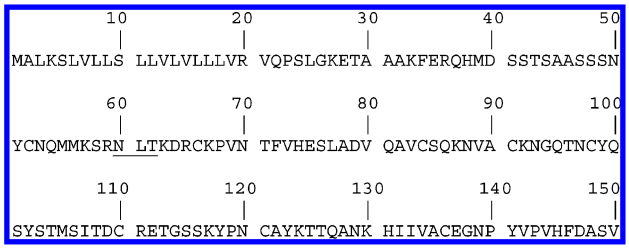


Chart 2



Glycosylation is one of the most common forms of posttranslational modification in eukaryotic proteins.[1] Glycans play key roles in protein folding, cell−cell recognition, cancer metastasis, and the immune system.[2,3] The number of oligosaccharide structures associated with a specific glycoprotein may be highly variable. For example, human erythrocyte CD59 has over 100 different oligosaccharide structures on a single glycosylation site (site heterogeneity).[4,5] Furthermore, there may be several potential N-linked glycosylation sites in the polypeptide that may or may not be glycosylated. Potential sites of N-glycosylation are readily identified with the consensus sequence AsnXSer or AsnXThr, where X can be any amino acid except proline.

Structural heterogeneity is an important characteristic of oligosaccharides and significantly complicates the structural analysis of glycoproteins. In 2-D gel electrophoresis, the structural heterogeneity of the oligosaccharides in glycoproteins results in

* To whom correspondence should be addressed. E-mail: cblebrilla@ucdavis.edu.

[†] Department of Chemistry.

[‡] Section of Molecular and Cell Biology.

(1) Wang, Y.; Tan, J.; Sutton-Smith, M.; Ditto, D.; Panico, M.; Campbell, R. M.; Varki, N. M.; Long, J. M.; Jaeken, J.; Levinson, S. R.; Wynshaw-Boris, A.; Morris, H. R.; Le, D.; Dell, A.; Schachter, H.; Marth, J. D. *Glycobiology* **2001**, *11*, 874.

(2) Helenius, A.; Aebi, M. *Science* **2001**, *291*, 2364−2369.

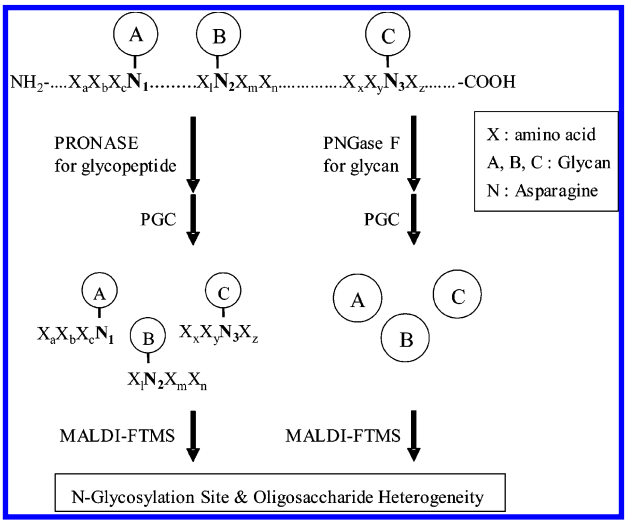
(3) Lowe, J. B. *Cell* **2001**, *104*, 809−812.

(4) Rudd, P. M.; Colominas, C.; Royle, L.; Murphy, N.; Hart, E.; Merry, A. H.; Hebestreit, H. F.; Dwek, R. A. *Proteomics* **2001**, *1*, 285−294.

(5) Rudd, P. M.; Elliott, T.; Cresswell, P.; Wilson, I. A.; Dwek, R. A. *Science* **2001**, *291*, 2370−2376.

10.1021/ac034414x CCC: $25.00
Published on Web 09/11/2003

Scheme 1



several spots, a smear, or a combination of both. Identification of specific glycosylation sites requires a lengthy process often involving trypsin digestion to produce glycopeptides, separation of the glycopeptides by high-performance liquid chromatography, and characterization of each fraction by mass spectrometry to determine the oligosaccharide and peptide structures of each fraction (left path, Chart 1). As the mass spectra often do not always yield unequivocal structural determination, tandem mass spectrometry (MS) is often necessary to identify unique fragments that correspond to the oligosaccharide moieties. Tandem MS (MS^n) or specifically collision-induced dissociation provides fragments that yield the peptide sequence, and hence the glycosylation site, as well as fragmentation of the oligosaccharides to determine their connectivity.

The traditional procedure in Chart 1 has several limitations. (1) Each oligosaccharide component gives the glycopeptide a different chromatographic property so that glycopeptides originating from the same glycosylation site may be distributed over the chromatographic profile. This behavior makes it difficult to identify all glycans from a given site and to quantify their relative abundances. (2) In the mass spectrum of an HPLC fraction, it is difficult to distinguish a glycopeptide from a large peptide based on mass alone. (3) Many glycoproteins are not susceptible to trypsin digestion, and therefore, preparation of glycopeptides is difficult.[6,7]

There is little or no alternative to the current paradigm. A newly emerging method of electron capture dissociation has been used to determine sites of glycosylation in whole intact glycoproteins.[8] While promising, the method is not widely available and may not necessarily make the characterization of heterogeneity significantly simpler.

In this report, we describe a relatively rapid procedure for the characterization of N-linked glycosylation sites based on a combination of general proteolysis coupled with solid-phase extraction of glycopeptides and high mass accuracy mass spectrometry (Chart 1). The enhancement in speed is due to the elimination of the HPLC separation and the associated MS analyses of each fraction from existing methods. The new protocol requires only two MS analyses. The glycoprotein was extensively digested with Pronase to produce glycopeptides with small peptide moieties (2−8 residues) except near the sites of glycosylation (Scheme 1). The oligosaccharide sterically blocks the digestion of the peptide segment directly connected to the oligosaccharide. These observations were also reported previously by Juhasz and Martin.[9] After solid-phase extraction with a porous graphitic carbon (PGC) cartridge, which concentrated the glycopeptide and removed salts, the mass spectrum of the glycopeptides was obtained. Because large nonglycosylated peptides were no longer present, glycopeptide masses were readily identified as they were essentially the only molecular species with signals above m/z 1000. For additional structural information, a portion of the initial glycoprotein sample was treated with PNGase F to release oligosaccharides. After porous graphitized carbon (PGC) purification, the mass spectrum was obtained and compared to the glycopeptide spectrum to obtain the peptide masses (and amino acid sequence), and hence identify the site of glycosylation.

EXPERIMENTAL SECTION

Materials and Methods. Ribonuclease B, chicken ovalbumin, and Pronase E were purchased from Sigma-Aldrich (St. Louis, MO) and used without further purification. Peptide-*N*-glycosidase F (PNGase F) was obtained from Calbiochem (La Jolla, CA), and modified trypsin was purchased from Promega Corp. (Madison, WI). Solvents were of HPLC grade. PGC cartridges for desalting were purchased from Alltech Associates, Inc. (Deerfield, IL). Evaporation of small amounts of solvent was done on a Centrivap Concentrator (Labconco Corp., Kansas City, MO) at 45 °C.

Preparation of the *Xenopus laevis* Jelly Ligand Column. Ovulated eggs were collected from hormonally induced *X. laevis* females and were dejellied using 45 mM mercaptoethanol in DeBoers solution adjusted to pH 9.5.[10] Solubilized jelly was centrifuged at 20000*g* at 4 °C and concentrated using ultrafitration. The jelly ligands for cortical granule lectin were separated from concentrated total jelly by preparative SDS−PAGE using a 4% stacking and 12% separating gel. After electroelution of the stacking gels, the ligand solution was dialyzed against water, concentrated, and conjugated to an activated-agarose gel (Affi-Gel; Bio-Rad Laboratories, Hercules, CA).

Purification of *X. laevis* Cortical Granule Lectin. Procedures based on Nishihara et al.[11] were used to purify the egg cortical granule lectin. After homogenizing dejellied eggs in column buffer (10 mM Tris, 150 mM NaCl, 1 mM $CaCl_2$, pH 7.8), yolk was separated from the lysate by adding an equal volume of 1,1,2-trichlorotrifluoroethane, mixing, centrifuging at 20000*g* for 10 min, and removing the supernatant solution.[12] The solution was added to the jelly ligand column and incubated overnight at 4 °C with gentle rocking. The column was washed with greater than 50 volumes of column buffer, and the lectin was eluted by addition

(6) Dezutterdambuyant, C.; Schmitt, D. A.; Dusserre, N.; Hanau, D.; Kolbe, H. V. J.; Kieny, M. P.; Gazzolo, L.; Mace, K.; Pasquali, J. L.; Olivier, R.; Schmitt, D. *Res. Virol.* **1991**, *142*, 129−138.
(7) Bezouska, K.; Sklenar, J.; Novak, P.; Halada, P.; Havlicek, V.; Kraus, M.; Ticha, M.; Jonakova, V. *Protein Sci.* **1999**, *8*, 1551−1556.
(8) Hakansson, K.; Cooper, H. J.; Emmett, M. R.; Costello, C. E.; Marshall, A. G.; Nilsson, C. L. *Anal. Chem.* **2001**, *73*, 4530−4536.
(9) Juhasz, P.; Martin, S. A. *Int. J. Mass Spectrom. Ion Processes* **1997**, *169*, 217−230.
(10) Hedrick, J. L.; Hardy, D. M. *Methods Cell Biol.* **1991**, *36*, 231−247.
(11) Nishihara, T.; Wyrick, R. E.; Working, P. K.; Chen, Y. H.; Hedrick, J. L. *Biochemistry* **1986**, *25*, 6013−6020.
(12) Gurdon, J. B.; Wickens, M. P. *Methods Enzymol.* **1983**, *101*, 370−386.

of 0.5 M galactose in calcium-free column buffer. The eluant was dialyzed against 10 mM ammonium bicarbonate and then dried by lyophilization.

Digestion of Glycoproteins with Pronase. Glycoprotein (0.1−1.0 mg) was dissolved in 0.1 M Tris-HCl buffer (pH 7.5), and Pronase (10 units of enzyme) was added to a solution and incubated at 37 °C for 36−48 h. The reaction mixture was boiled for 5 min to deactivate the enzyme. The digested glycoprotein was desalted and collected on a PGC cartridge.

Digestion of Glycoproteins with Trypsin. Glycoprotein (0.1−1.0 mg) was dissolved in 60 μL of 8 M urea (0.1 M Tris-HCl buffer pH 7.8) and incubated with 3 μL of 450 mM dithiothreitol for 45 min at 55 °C. The reduced sample was allowed to cool to room temperature, 6 μL of 500 mM iodoacetamide was added, and the mixture was incubated at room temperature for 20 min in the dark. The reaction was terminated by adding water to give a final concentration of 2 M urea. The reduced and alkylated glycoprotein was digested with 5 μg of modified trypsin at 37 °C overnight. The digest was put in −20 °C freezer to terminate the reaction.

HPLC Fractions of Glycoprotein from Tryptic Digest. The tryptic glycopeptides were purified and fractioned using a C_{18} column (5 μm, 120 × 200 mm). Glycopeptides were eluted from 0.1% TFA in water as solvent A and 0.08% TFA in acetonitrile as solvent B with a gradient that changed from 2 to 50% B within 70 min using a flow rate of 0.3 mL/min. The UV-active fractions were collected and concentrated in vacuo prior to MS analysis.

Deglycosylation of Tryptic Glycopeptides by PNGase F. Selected tryptic glycopeptides were incubated with PNGase F (1 U) in 100 mM NH_4HCO_3 (pH 7.5) for 12 h at 37 °C.

Deglycosylation of Glycoprotein by PNGase F. Denaturation of the glycoprotein was achieved by heating at 95 °C for 10 min with addition of the sodium dodecyl sulfate (total 2%). After cooling, PNGase F (1 unit) and 200 mM NH_4HCO_3 buffer were added to the solution (finally reaction mixture concentration was 100 mM NH_4HCO_3) and incubated overnight at 37 °C. An aliquot of 100% ethanol was added to the reaction mixture to make an 80% ethanol solution, and the resultant mixture was chilled to −20 °C for 30 min to precipitate all the proteins including the PNGase F. The reaction mixture was centrifuged, and the supernatant was removed. The supernatant was completely dried to remove the ethanol. The resulting oligosaccharides were dissolved in Nanopure water. The oligosaccharide solutions were applied to the PGC cartridge for desalting.

Glycopeptide and Oligosaccharide Purification Using a PGC-Solid-Phase Extraction (SPE). Digested glycopeptides with nonspecific protease and oligosaccharides released by PNGase F were purified by solid-phase extraction using a PGC cartridge. A PGC cartridge was washed with H_2O followed by 0.05% (v/v) TFA in 80% ACN/H_2O (v/v). The solution of digested glycoprotein or oligosaccharide was applied to the PGC cartridge. Subsequently, the cartridge was washed with Nanopure water at a flow rate of ~1 mL/min to remove salts and buffer. Glycopeptides and glycans were eluted with 10% ACN in H_2O, 20% ACN in H_2O, and 40% ACN in 0.05% TFA in H_2O. Each fraction was collected and concentrated in vacuo prior to MALDI analysis.

Mass Spectrometric Analysis. Mass spectra were recorded on an external source MALDI-FTICR-MS (HiResMALDI, IonSpec Corp., Irvine, CA) equipped with a 4.7-T magnet. MALDI-TOF mass spectra were recorded in linear mode using a Proflex III (Bruker Daltonics). The FT-ICR-MS was equipped with a pulsed YAG laser (266 nm). 2, 5-Dihydroxybenzoic acid was used as a matrix (5 mg/100 μL in ethanol). A saturated solution of NaCl (or in some cases KCl) in methanol was used as a dopant for the identification of the quasimolecular ion for either glycopeptides

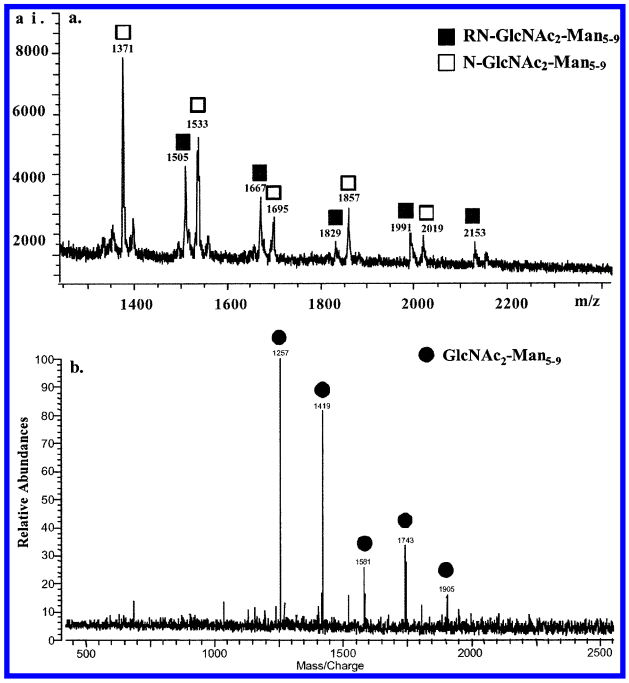


Figure 1. (a) MALDI-TOF MS of glycopeptides from Pronase digestion of ribonuclease B. Two series are observed corresponding to the peptide RN and the amino acid N. (b) MALDI-FT MS of oligosaccharides released from ribonuclease B by PNGase F.

Table 1 [a]

Section a

experimental mass data (Da)		peptide mass (Da)		
glycopeptide (■)	glycan	possible peptide sequence: RN		
$[M + H]^+$	$[M + Na]^+$	calcd	theor	mass error Δm (Da)
■ 1505.456	1257.420	288.119	288.155	0.036
■ 1667.478	1419.446	288.115	288.155	0.040
■ 1829.492	1518.462	288.113	288.155	0.042
■ 1991.614	1743.586	288.111	288.155	0.044
■ 2153.811	1905.779	288.115	288.155	0.040

Section b

experimental mass data (Da)		peptide mass (Da)		
glycopeptide (□)	glycan	possible peptide sequence: N		
$[M + Na]^+$	$[M + Na]^+$	calcd	theor	mass error Δm (Da)
□ 1371.386	1257.420	132.068	132.053	0.015
□ 1533.410	1419.446	132.066	132.053	0.013
□ 1695.449	1581.462	132.089	132.053	0.036
□ 1857.505	1743.586	132.021	132.053	0.032
□ 2019.763	1905.779	132.086	132.053	0.033

[a] All molecular ion mass and theoretical mass are monoisotopic mass.

Table 2

experimental mass data (Da)					
glycopeptide	glycan				
$[M + H]^+$	$[M + Na]^+$	composition	calcd peptide mass	mass error Δm (Da)	glycosylated peptide
		a. Hybride-Type Oligosaccharides			
○ 1442.544	1136.388	$(Hex)_3$-$(HexNAc)_3$	346.149	0.036	NLT
○ 1645.634	1339.498	$(Hex)_3$-$(HexNAc)_4$	346.131	0.055	NLT
○ 1848.712	1542.576	$(Hex)_3$-$(HexNAc)_5$	346.128	0.057	NLT
○ 2051.810	1745.658	$(Hex)_3$-$(HexNAc)_6$	346.144	0.041	NLT
○ 2254.860	1948.763	$(Hex)_3$-$(HexNAc)_7$	346.089	0.096	NLT
○ 2458.001	2151.864	$(Hex)_3$-$(HexNAc)_8$	346.129	0.056	NLT
$[M + Na]^+$	$[M + Na]^+$				
● 1484.567	1257.22	$(Hex)_5$-$(HexNAc)_2$	245.156	0.018	NL
● 1687.651	1460.502	$(Hex)_5$-$(HexNAc)_3$	245.159	0.022	NL
● 1890.708	1663.577	$(Hex)_5$-$(HexNAc)_4$	245.142	0.005	NL
△ 2279.927	1745.635	$(Hex)_5$-$(HexNAc)_6$	552.303	0.048	EKYN
△ 2441.971	1907.685	$(Hex)_5$-$(HexNAc)_6$	552.297	0.043	EKYN
▲ 2281.924	1866.650	$(Hex)_5$-$(HexNAc)_5$	433.284	0.067	NLTS
▲ 2485.013	2069.740	$(Hex)_5$-$(HexNAc)_6$	433.283	0.066	NLTS
◇ 2336.990	1501.529	$(Hex)_5$-$(HexNAc)_4$	853.472	0.054	EKYNLTS
◇ 2540.070	1704.599	$(Hex)_5$-$(HexNAc)_5$	853.482	0.063	EKYNLTS
		b. High-Mannose-Type Oligosaccharides			
$[M + Na]^+$	$[M + Na]^+$				
■ 1727.679	933.317	$(Hex)_3$-$(HexNAc)_2$	812.372	0.034	MEEKYN
■ 1889.753	1095.370	$(Hex)_4$-$(HexNAc)_2$	812.394	0.056	MEEKYN
■ 2051.816	1257.434	$(Hex)_5$-$(HexNAc)_2$	812.393	0.055	MEEKYN
■ 2213.887	1419.415	$(Hex)_6$-$(HexNAc)_2$	812.482	0.145	MEEKYN
□ 2159.863	1581.519	$(Hex)_7$-$(HexNAc)_2$	596.352	0.072	YNLTS
□ 2321.951	1743.671	$(Hex)_8$-$(HexNAc)_2$	596.289	0.009	YNLTS
□ 2484.036	1905.712	$(Hex)_9$-$(HexNAc)_2$	596.332	0.052	YNLTS

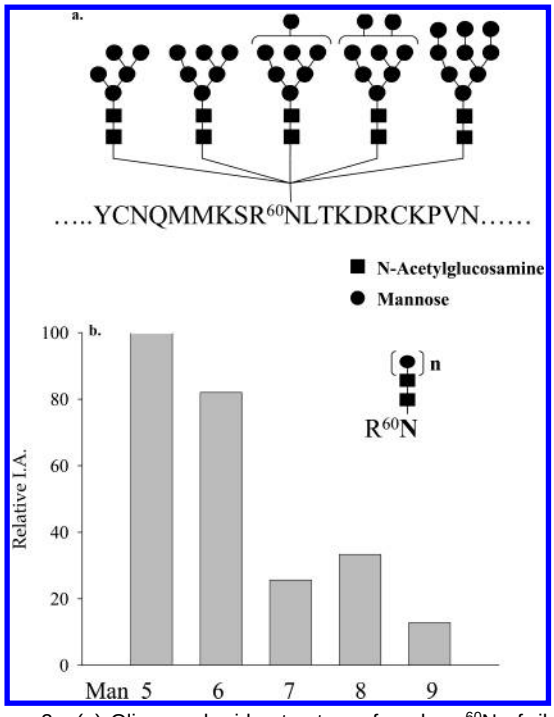


Figure 2. (a) Oligosaccharide structures found on ^{60}N of ribonuclease B. (b) Relative abundances of each oligosaccharide.

Chart 3

```
        10         20         30         40         50
         |          |          |          |          |
GSIGAASMEF CFDVFKELKV HHANENIFYC PIAIMSALAM VYLGAKDSTR

        60         70         80         90        100
         |          |          |          |          |
TQINKVVRFD KLPGFGDSIE AQCGTSVNVH SSLRDILNQI TKPNDVYSFS

       110        120        130        140        150
         |          |          |          |          |
LASRLYAEER YPILPEYLQC VKELYRGGLE PINFQTAADQ ARELINSWVE

       160        170        180        190        200
         |          |          |          |          |
SQTNGIIRNV LQPSSVDSQT AMVLVNAIVF KGLWEKAFKD EDTQAMPFRV

       210        220        230        240        250
         |          |          |          |          |
TEQESKPVQM MYQIGLFRVA SMASEKMKIL ELPFASGTMS MLVLLPDEVS

       260        270        280        290        300
         |          |          |          |          |
GLEQLESIIN FEKLTEWTSS NVMEERKIKV YLPRMKMEEK YNLTSVLMAM

       310        320        330        340        350
         |          |          |          |          |
GITDVFSSSA NLSGISSAES LKISQAVHAA HAEINEAGRE VVGSAEAGVD

       360        370        380
         |          |          |
AASVSEEFRA DHPFLFCIKH IATNAVLFFG RCVSP
```

or oligosaccharides. The glycopeptide/oligosaccharide solution (1 μL) was applied to the MALDI probe followed by matrix solution (1 μL). The sample was dried under a stream of air prior to mass spectrometric analysis.

Peptide Sequence Database. Glycoprotein amino acid sequences were obtained at Swiss-Prot and Trembl databases (accession numbers: Ribonuclease B, P00656; ovalbumin, P01012; CGL.1, Q91719; CGL.2, P79957). Potential N-glycosylation site, peptide was identified with the consensus sequence NXS/T, where X is any amino acid except P.

Table 3

high-mannose type					
glycopeptide	glycan				
$[M + Na]^+$	$[M + Na]^+$	composition	calcd peptide mass	mass error Δm (Da)	glycosylated peptide
		a. Experimental Mass Data (Da)			
● 2033.798	1419.465	$(Hex)_6$-$(HexNAc)_2$	632.344	0.031	GSANLTA
● 2195.856	1581.544	$(Hex)_7$-$(HexNAc)_2$	632.322	0.009	GSANLTA
● 2075.808	1419.465	$(Hex)_6$-$(HexNAc)_2$	674.354	0.006	LGSANLT
● 2237.872	1581.544	$(Hex)_7$-$(HexNAc)_2$	674.339	0.021	LGSANLT
● 2399.926	1743.586	$(Hex)_8$-$(HexNAc)_2$	674.351	0.008	LGSANLT
● 2251.879	1581.544	$(Hex)_7$-$(HexNAc)_2$	688.345	0.006	DLGSANL
● 2413.942	1743.586	$(Hex)_8$-$(HexNAc)_2$	688.367	0.028	DLGSANL
● 2377.906	1581.544	$(Hex)_7$-$(HexNAc)_2$	814.372	0.014	NLTASFY
● 2540.004	1743.586	$(Hex)_8$-$(HexNAc)_2$	814.429	0.143	NLYSAFY
■ 2117.821	1257.405	$(Hex_5$-$(HexNAc)_2$	878.427	0.050	VWHVPNK
■ 2279.874	1419.465	$(Hex)_6$-$(HexNAc)_2$	878.419	0.057	VWHVPNK
■ 2441.948	1581.544	$(Hex)_7$-$(HexNAc)_2$	878.414	0.062	VWHVPNK
■ 2343.937	1905.650	$(Hex)_9$-$(HexNAc)_2$	456.298	0.028	VPNK
■ 2505.969	2067.705	$(Hex)_{10}$-$(HexNAc)_2$	456.274	0.004	VPNK
■ 2159.831	1257.405	$(Hex)_5$-$(HexNAc)_2$	920.437	0.050	LGVWHVPN
■ 2321.883	1419.465	$(Hex)_6$-$(HexNAc)_2$	920.428	0.059	LGVWHVPN
		b. Hybrid/Complex Type			
○ 1857.724	974.330	$(Hex)_2$-$(HexNAc)_3$	901.405	0.013	NLTASFYS
○ 2019.785	1136.390	$(Hex)_3$-$(HexNAc)_3$	901.406	0.012	NLTASFYS
○ 2181.808	1298.439	$(Hex)_4$-$(HexNAc)_3$	901.379	0.039	NLTASFYS
○ 2343.937	1460.491	$(Hex)_5$-$(HexNAc)_3$	901.457	0.038	NLTASFYS
○ 1941.734	1542.555	$(Hex)_3$-$(HexNAc)_5$	417.189	0.034	NLTA (ANLT)
○ 2103.798	1704.608	$(Hex)_4$-$(HexNAc)_5$	417.200	0.022	NLTA (ANLT)
○ 2265.854	1866.661	$(Hex)_5$-$(HexNAc)_5$	417.204	0.018	NLTA (ANLT)
○ 2427.940	2028.714	$(Hex)_6$-$(HexNAc)_5$	417.236	0.014	NLTA (ANLT)
○ 2238.869	2053.745	$(Hex)_4$-$(HexNAc)_6$−Fuc	203.135	0.045	AN
○ 2505.969	2320.809	$(Hex)_6$-$(HexNAc)_5$-$(Fuc)_2$	203.171	0.080	AN
○ 2335.906	1298.439	$(Hex)_4$-$(HexNAc)_3$	1055.477	0.015	NLTASFYSPG
○ 2497.970	1460.491	$(Hex)_5$-$(HexNAc)_3$	1055.490	0.003	NLTASFYSPG
○ 2511.989	1460.491	$(Hex)_5$-$(HexNAc)_3$	1069.508	0.000	ANLTASFYSP
○ 2674.086	1622.552	$(Hex)_6$-$(HexNAc)_3$	1069.545	0.037	ANLTASFYSP
□ 1899.740	1688.623	$(Hex)_3$-$(HexNAc)_5$−Fuc	229.128	0.022	PN
□ 2061.793	1850.675	$(Hex)_4$-$(HexNAc)_5$−Fuc	229.128	0.022	PN
□ 2223.846	2012.708	$(Hex)_5$-$(HexNAc)_5$−Fuc	229.148	0.042	PN
□ 2385.941	2174.769	$(Hex)_6$-$(HexNAc)_5$−Fuc	229.182	0.075	PN

RESULTS AND DISCUSSION

Characterization of Glycosylation on Model Glycoproteins. (A) Ribonuclease B. To validate the experimental strategy, a model glycoprotein with a single glycosylation site and known oligosaccharide structures was characterized. Ribonuclease B is a well-studied glycoprotein containing high-mannose oligosaccharides.[13] The peptide chain contains 124 amino acids with a nominal mass of 13 678 Da. The sequence is shown with the glycosylation site on ^{60}Asn (Chart 2).

A positive ion mass spectrum of the glycopeptides isolated after Pronase digestion of ribonuclease B by PGC is shown in Figure 1a. This spectrum was obtained after minimal sample cleanup without HPLC (see Experimental Section). The only peaks observed above m/z 1000 corresponded to predicted glycopeptides. Large nonglycosylated peptide fragments were not present as they were digested to their amino acid constituents. Ionic species with m/z greater than 1000 were glycopeptides as the oligosaccharide moieties corresponded to at least 800 mass units ($GlcNAc_2Man_3$). There were two glycopeptide series present in the spectrum, based on peaks that differed by 162 Da or one hexose residue. Series 1 (open square) included the masses m/z 1371.386, 1533.410, 1695.449, 1857.505, and 2019.763. Series 2 (filled square) included the masses m/z 1505.456, 1667.478, 1829.492, 1991.614, and 2153.811.

Doping the sample with NaCl (see Experimental Section) provided additional information. The dopant enhanced the relative intensities of the five peaks in series 1, indicating that the ionic species were sodium-coordinated (Figure 1a). The other series was found to be protonated species whose intensities diminished upon the addition of the NaCl dopant (not shown).

To identify the peptide moieties in Figure 1a, the glycans were released from the glycoprotein with PNGase F and analyzed by MALDI MS (Figure 1b). Five glycan species were observed, reflecting the five glycopeptides observed in Figure 1a. The signals corresponded to sodium-adducted species of high-mannose oligosaccharides ranging in composition from $GlcNAc_2Man_5$ to $GlcNAc_2Man_9$ with m/z 1257.420, 1419.446, 1581.462, 1743.586, and 1905.779, respectively. The mass difference between adjacent peaks was 162 Da, corresponding to a mannose residue. The relative signal intensities of the five glycans (Figure 1b) were congruent with the five glycopeptides (Figure 1a).

The peptide moieties were identified by subtracting the masses of the observed glycans from the masses of the glycopeptides.

(13) Fu, D. T.; Chen, L.; Oneill, R. A. *Carbohydr. Res.* **1994**, *261*, 173−186.

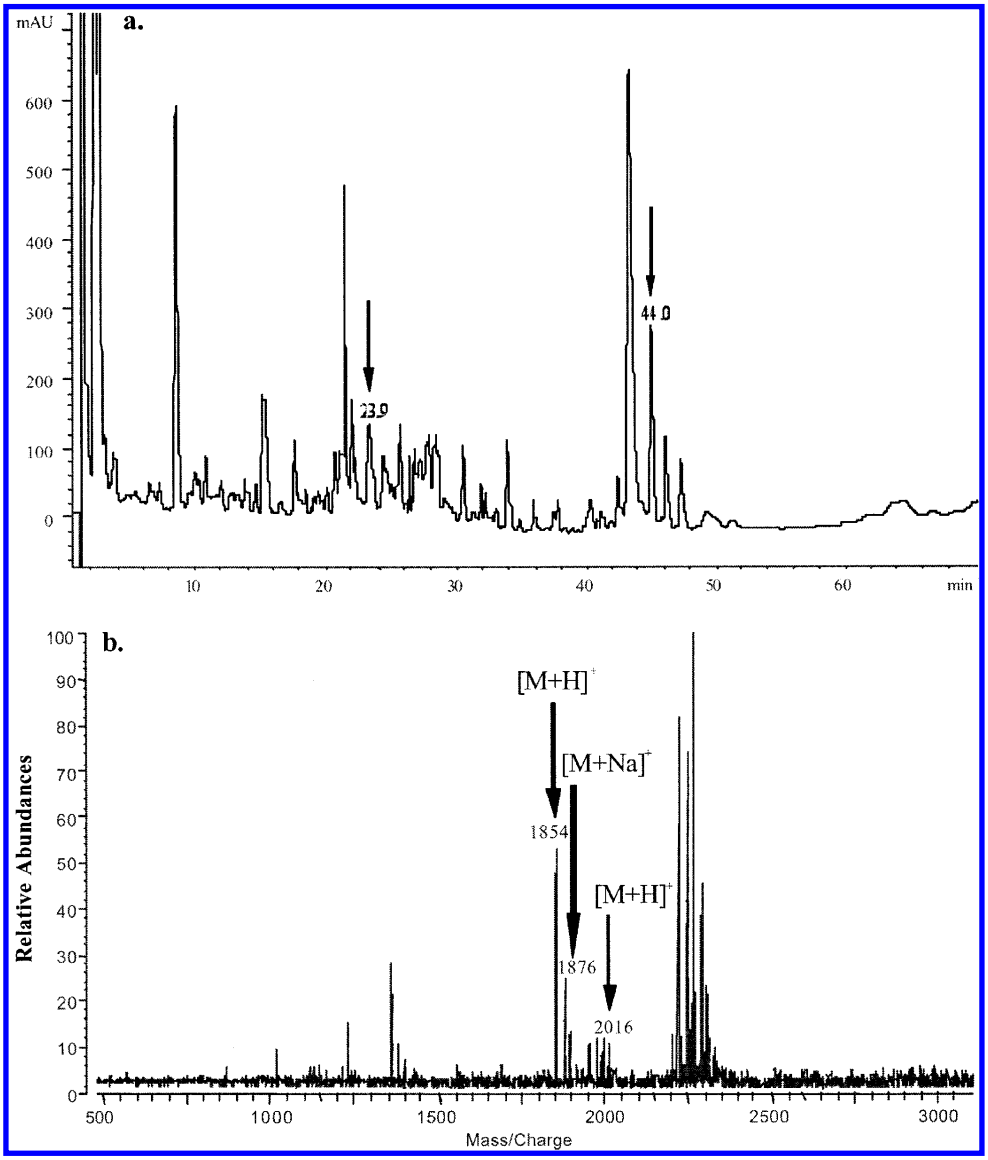


Figure 3. (a) HPLC with diode array detector (214 nm) of glycopeptides from trypsin digestion of ribonuclease B. The arrow indicates two HPLC fractions with glycopeptides. (b) MALDI-FT MS of HPLC fraction at 44.0 min. The arrows point to the quasimolecular ions of glycopeptides.

For example, the first peak of series 2 (m/z 1505.456, Figure 1a) corresponded to the glycopeptide with the glycan moiety $GlcNAc_2$-Man_5 (m/z 1257.420). The peptide mass was obtained by subtracting the observed glycan mass 1216.42 Da ($[M + Na]^+$ minus Na^+ and H_2O) from the observed glycopeptide mass 1504.45 Da ($[M + H]^+$ minus H^+). The peptide mass (288.12 Da) corresponded to the dipeptide Arg-Asn (theoretical mass 288.15 Da), with the glycosylation site at position ^{60}Asn (Table 1). The theoretical mass routinely varied by less the 20 ppm from the experimental mass. The glycopeptides of series 2 therefore consisted of the dipeptide Arg-Asn with high-mannose oligosaccharides ranging in size from $GlcNAc_2Man_5$ to $GlcNAc_2Man_9$. The lowest mass of series 1 (m/z 1371.39) corresponded to the glycopeptide with the oligosaccharide component $GlcNAc_2$-Man_5 (m/z 1257.42). The peptide mass of 132.07 Da corresponded to Asn (theoretical mass 132.05 Da). Glycopeptides of series 1 were, therefore, composed of a single Asn residue.

Juhasz and Martin[9] similarly examined the reaction of Pronase with ribonuclease B. They found that reacting the protein for an extended period cleaves down to the glycodipeptides. We find that it can even have a single amino acid. However, ribonuclease B is unique. For the reaction of glycoproteins, it is now typical that tri up to hepto residues can be left intact even after 48 h of reaction. For ribonuclease B, only a single Arg-Asn site exists and it overlaps the N-linked site as shown in Figure 2a. To illustrate that quantification is possible based on the relative ion abundances, the relative abundance of each glycan species was determined (Figure 2b). As these glycans have similar structures and sizes, their ionization efficiencies are similar and permit calculation of the relative oligosaccharides abundances in the glycoprotein.[14]

Characterization of Glycosylation Site in Ribonuclese B Using Trypsin Digestion. For comparison, ribonuclease B was digested with trypsin to illustrate the standard method for determining glycosylation. In this procedure the protein was digested by trypsin followed by separation and fractionation of

(14) Cancilla, M. T.; Wang, A. W.; Voss, L. R.; Lebrilla, C. B. *Anal. Chem.* **1999**, *71*, 3206−3218.

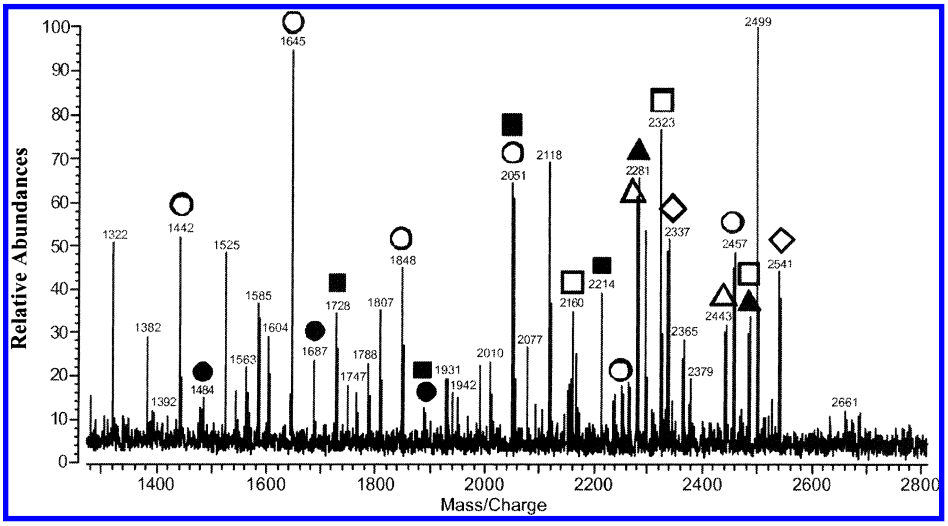


Figure 4. MALDI-FT MS of glycopeptides from the Pronase digestion of chicken ovalbumin. Each symbol marks an ion series that corresponds to an oligosaccharide series with a specific peptide residue.

the glycopeptides by HPLC (see Experimental Section).

Figure 3a shows the HPLC chromatogram of trypsin-digested ribonuclease B. To determine the fraction that contained glycopeptides, each fraction was collected and subsequently analyzed by MALDI MS. However, as fractions that contained glycopeptides often also contained peptides, identification of the glycopeptides was challenging. For example, fractions that eluted at 23.9 and 44.0 min contained glycopeptides. The positive mode mass spectrum of the fraction collected at 44.0 min is shown in Figure 3b. The peaks at m/z 1854.7 and 2016.8 correspond to the protonated adduct—a glycosylated peptide NLTK with $GlcNAc_2Man_6$ and $GlcNAc_2Man_7$, respectively. To identify the presence of glycopeptides in the fraction, the glycans were released using PNGase F followed by MALDI MS. No other glycopeptides were found in the 44.0-min fraction; the other large masses in the spectrum corresponded to nonglycosylated peptides. Other glycopeptides were found in other HPLC fractions after a repetitive MS analysis of each HPLC fraction.

(B) Characterization of Chicken Ovalbumin Glycosylation Sites. Chicken ovalbumin is a well-studied glycoprotein used to establish our proposed procedure. The protein moiety is composed of 385 amino acids with a molecular mass of 43 kDa (Chart 3). Ovalbumin has a single glycosylation site at ^{292}Asn with a second potential glycosylation at ^{311}Asn.[15−17] Figure 4 shows the mass spectrum of glycopeptides obtained after Pronase digestion and PGC separation. Doping the sample with NaCl enhanced the intensities of the quasimolecular ion ($[M + Na]^+$) while the addition of KCl shifted the signals accordingly. Five glycopeptide series were observed with signal separation of 203 Da, corresponding to an *N*-acetylhexosamine residue. The following series of peaks were identified: series 1 (open circles) with m/z 1442.544, 1645.634, 1848.712, 2051.810, 2254.860, and 2458.001; series 2 (filled circles) with m/z 1484.567, 1687.651, and 1890.708; series 3 (open triangles) with m/z 2279.927 and 2441.971; series 4 (filled triangles) with m/z 2281.924 and 2485.013; series 5 (open diamonds) with m/z 2336.990 and 2540.070. Two additional glycopeptide series were observed with signal separation of 162 Da, corresponding to hexose residue: series 6 (filled squares) with m/z 1727.679, 1889.753, 2051.816, and 2213.887 and series 7 (open squares) with m/z 2159.863, 2321.951, and 2484.036. Of the glycopeptides, only series 1 (open circles) corresponded to protonated parent ions.

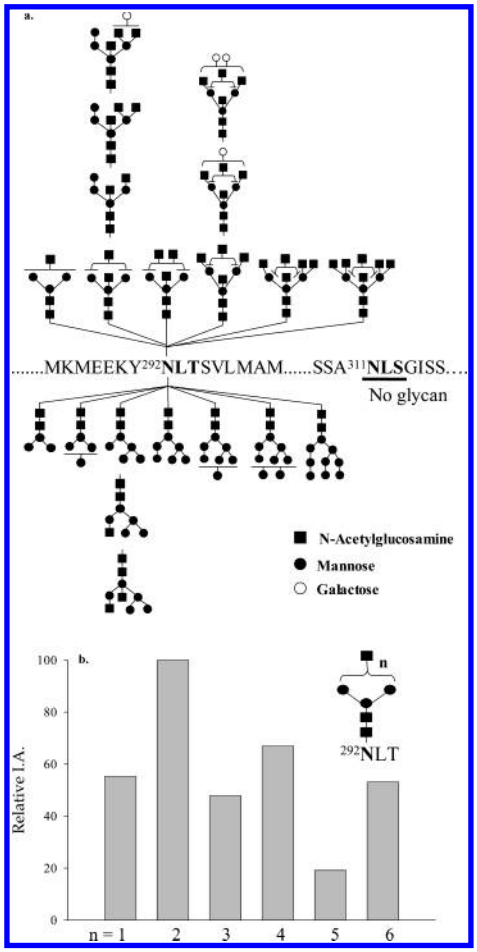
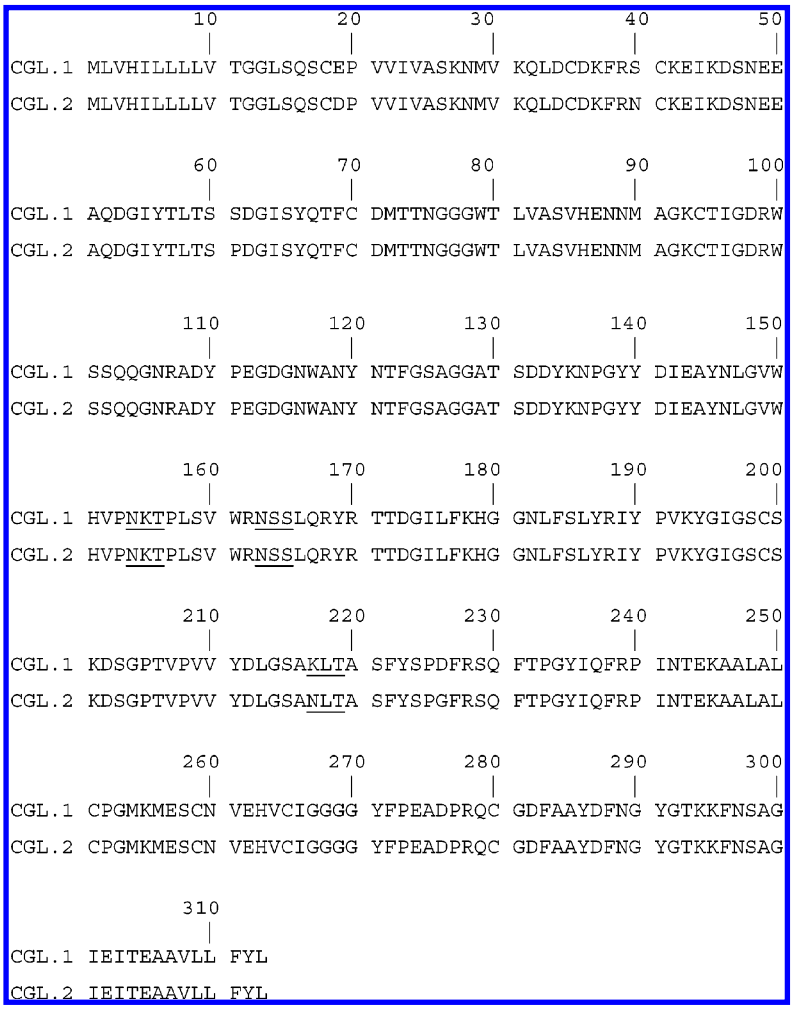


Figure 5. (a) Structures of ovalbumin glycopeptides based on data from Figure 4. (b) Relative abundances of each oligosaccharide on ^{292}NLT.

(15) Suzuki, T.; Kitajima, K.; Emori, Y.; Inoue, Y.; Inoue, S. *Proc, Natl, Acad, Sci. U.S.A.* **1997**, *94*, 6244−6249.
(16) Wilm, M.; Mann, M. *Anal. Chem.* **1996**, *68*, 1−8.
(17) Sheares, B. T. *J. Biol. Chem.* **1988**, *263*, 12778−12782.

Chart 4



Oligosaccharides were released from ovalbumin by PNGase F digestion, desalted by PGC and analyzed by MALDI. Based on the masses, the N-linked glycans of ovalbumin consisted of hybrid-type oligosaccharides $HexNAc_nHex_m$ ($n = 2-8$, $m = 3-5$) and high-mannose-type $GlcNAc_2Man_n$ ($n = 3-9$) (data not shown), in agreement with published reports.[18] The most abundant ovalbumin oligosaccharides belonged to a series of hybrid-type oligosaccharides with the composition $HexNAc_nHex_3$ ($n = 3-8$) (open circles, Figure 4). The fixed number of hexoses and varying number of HexNAc residues identify this series as belonging to hybrid-type oligosaccharides. To identify the peptide moieties and the glycosylation sites, the peptide masses were calculated by subtracting glycan masses from glycopeptide masses, both obtained experimentally (Table 2), as described previously. For example, the ion at m/z 1442.544 in Figure 4 was consistent with a glycopeptide having a glycan $HexNAc_3Hex_3$ (1136.388 m/z $[M + Na]^+$, from MALDI of PNGase F-treated protein). Subtraction of the two masses yielded a peptide mass of 346.149 Da. A similar analysis of the other glycopeptides in this series yielded the same peptide moiety to the sequence Asn-Leu-Thr (Table 2). In addition, the relative abundances of each hybrid-type glycan species with the peptide NLT could be determined from the relative ion abundances (Figure 5b).

Other glycopeptides in the hybrid oligosaccharide family were found and labeled by filled circles (NL), filled triangles (NLTS), and open diamonds (EKYNLTS). The oligosaccharide structures at NL and NLTS sites had the general formula $HexNAc_nHex_5$ ($n = 2-6$), while those at the EKYNLTS site had the general formula $HexNAc_nHex_4$ ($n = 4-5$).

A series of glycopeptides with oligosaccharides consistent with the $Hex_nHexNAc_6$ ($n = 3-4$) were associated with the peptides EKYN (Table 2). The series differed in mass by one hexose, usually an indication of high-mannose oligosaccharides, but were judged to be hybrid-type oligosaccharides because of the small number of hexoses in the general formula and the lack of other glycopeptides with more hexoses in the series (see below).

In general, high-mannose oligosaccharides may be distinguished by mass series in which the ions are separated by 162 Da or one hexose (mannose) residue. Two series of glycopeptides exhibited this pattern (Table 2b). The large number of mannose residues in the general formula $HexNAc_2Hex_n$ ($n = 3-6$) and $HexNAc_2Hex_n$ ($n = 7-9$), respectively, were consistent with high-mannose-type oligosaccharides. The corresponding peptide moieties are listed in Table 2b. Oligosaccharides $Hex_nGlcNAc_2$ ($n = 3-6$) were associated with the peptide MEEKYN and $Hex_nGlcNAc_2$ ($n = 7-9$) with the peptide YNLTS.

There were no glycopeptides found with the peptide moiety corresponding to the other potential glycosylation site (^{311}Asn).

(18) Küster, B.; Wheeler, S. F.; Hunter, A. P.; Dwek, R. A.; Harvey, D. J. *Anal. Biochem.* **1997**, *250*, 882−101.

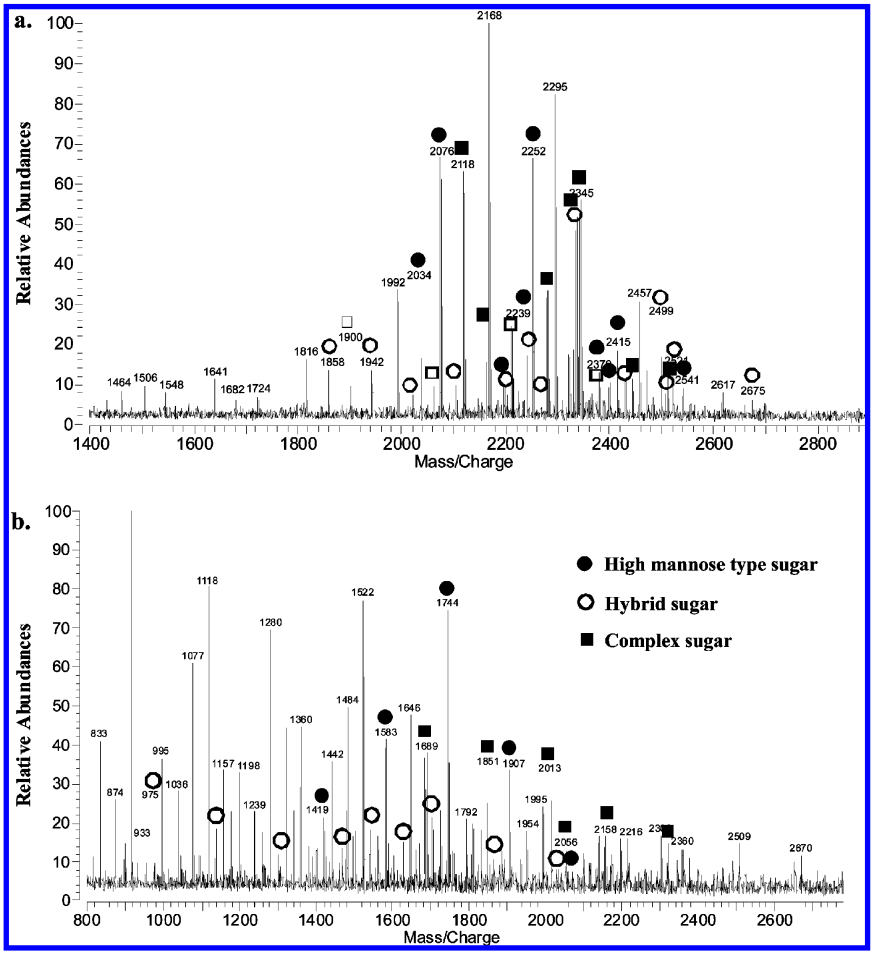


Figure 6. (a) MALDI-FT MS spectrum from Pronase digestion of *XL* CGL2. The most abundant peak, *m*/*z* 2168, is a self-proteolysis product of Pronase. High-mannose-type glycans were labeled by ■ and ● and hybrid/complex type glycans were labeled by □ and ○ at ^{154}Asn and ^{217}Asn, respectively. (b) MALDI-FT MS of oligosaccharides released from CGL.2 by PNGase F.

The results suggested that only ^{292}Asn of ovalbumin is a glycosylated site with the heterogeneity described in Figure 5a. Two types of oligosaccharides were observed corresponding to hybrid and high-mannose types. Glycan structures were not determined, but their arrangement in Figure 5a was consistent with those published in earlier reports.[19,20]

(C) Characterization of N-Glycosylation Sites in the Cortical Granule Lectin. A glycoprotein with an unknown glycosylation pattern, the *X. laevis* cortical granule lectin (CGL[21]), was investigated using this procedure. CGL derived from *X. laevis* is a glycoprotein released from the cortical granule after fertilization.[22,23] It has been postulated that CGL binds to oligosaccharide ligands on the egg jelly layer J1 to form a substrate that blocks polyspermy.[24] The processed glycoprotein is composed of 288 amino acids with a peptide mass of 32 kDa (accession number X82626).[11] Although the identity of CGL has been known for some time,[11] a new polymorphic protein has recently been discovered that is structurally different from that originally reported at amino acid 217. In the lectin discovered earlier (CGL.1), this position is occupied by Lys[11], while in CGL.2, this position is occupied by Asn[25] (Chart 4). CGL.1 has two glycosylation sites at ^{154}Asn and ^{163}Asn. The amino acid substitution in CGL.2 generates a third potential N-linked glycosylation site at ^{217}Asn. We attempted to determine the glycosylation pattern for purified CGL preparations by using the conventional trypsin method, but were unable to identify glycopeptides due to the resistance of CGL to tryptic digestion as evidenced by gel electrophoresis, HPLC and MS analysis (data not shown).

Tryptic peptides were generated from CGL preparations when deglycosylated using PNGase F, denatured using urea, and then trypsin digested. Amino acid sequence determination of tryptic peptides by MS−MS analysis revealed a CGL preparation that was CGL.2. This CGL.2 was digested with Pronase, and the glycopeptides were isolated and characterized as before. The mass spectrum of glycopeptides in the positive mode is shown in Figure 6a. The oligosaccharides from CGL.2 were released by PNGase F and analyzed by MALDI (Figure 6b). The oligosaccharides from CGL.2 consisted of high-mannose-type ($Man_nGlcNAc_2$) ($n = 5-9$), hybrid-, and complex-type oligosaccharides. Table 3 lists the

(19) Yamashita, K.; Tachibana, Y.; Kobata, A. *J. Biol. Chem.* **1978**, *253*, 3862−3869.
(20) Harvey, D. J.; Wing, D. R.; Küster, B.; Wilson, I. B. H. *J. Am. Soc. Mass Spectrom.* **2000**, *11*, 564−571.
(21) Wyrick, R. E.; Nishihara, T.; Hedrick, J. L. *Proc. Natl. Acad. Sci. U.S.A.* **1974**, *71*, 2067−2071.
(22) Quill, T. A.; Hedrick, J. L. *Dev. Growth Differ.* **1994**, *36*, 615−620.
(23) Quill, T. A.; Hedrick, J. L. *Arch. Biochem. Biophys.* **1996**, *333*, 326−332.
(24) Mozingo, N. M.; Hedrick, J. L. *Dev. Growth Differ.* **1996**, *38*, 647−652.
(25) Lee, J. K.; Buckhaults, P.; Wilkes, C.; Teilhet, M.; King, M. L.; Moremen, K. W.; Pierce, M. *Glycobiology* **1997**, *7*, 367−372.

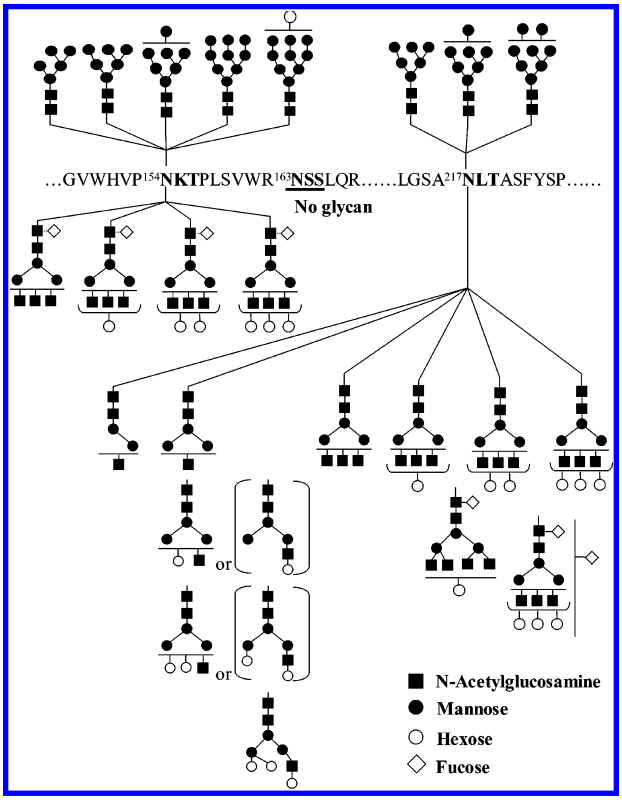


Figure 7. Distribution of the glycans on the three potential glycosylation sites. No glycan was found on ^{163}N.

glycopeptides found along with the peptide components after subtraction of the oligosaccharide moieties. The small differences in the mass error confirmed the assignments. We found only two glycosylation sites, ^{154}Asn and ^{217}Asn. The oligosaccharide structures found at these glycosylation sites are shown in Figure 7. No glycosylation was found at ^{163}Asn. The proposed glycan structures in Figure 7 are based solely on glycan masses, and proposed structures were not verified by additional analysis. However, these assignments are consistent with several structures as determined by NMR.[26] This analysis, as well as that of CGL.1 glycosylation sites, will be the subject of a future publication.

CONCLUSIONS

The procedure described in this report based on the combination of a nonspecific proteolysis to generate glycopeptides, solid-phase extraction of glycopeptides, PNGase F deglycosylation, and MALDI-FT MS provides a rapid and sensitive method for the identification of glycosylation sites and oligosaccharide heterogeneity. Each released glycopeptide contains a specific peptide moiety at the N-linked glycosylation site making it relatively simple to identify its origin. A substantially improved and efficient analysis is possible by elimination of HPLC separation of glycopeptides and repetitive MS analysis of separated glycopeptides. The new procedure is aided by the high mass accuracy routinely obtained by FT-MS. However, contemporary time-of-flight instruments achieve sufficient mass accuracy to perform similar analysis. Signal complexity due to large nonglycosylated peptides is eliminated by glycoprotein digestion using a nonspecific protease.

Quantitation of glycans at each glycosylation site is possible. For glycoproteins containing a single glycosylation site with little heterogeneity, such as ribonuclease B, the relative abundances of the glycopeptides in the mass spectrum mirror the relative abundances of the glycans (Figure 2). Neutral oligosaccharides of similar sizes tend to have very similar ionization efficiencies[14] thereby making quantitation possible. In situations where glycosylation is more diverse, such as in ovalbumin and CGL.2, glycans attached to identical peptide moieties can be directly compared. For example, the relative abundances of the $HexNAc_nHex_3$ (n = 3−8) in the mass spectrum of the glycopeptides derived from ovalbumin (Table 2a) also mirror the relative abundances of these glycan components (Figure 5b).

The new procedure presented in this paper will substantially improve the ease and efficiency for the determination and quantitation of protein site heterogeneity and structural heterogeneities of the oligosaccharide moiety of glycoproteins.

ACKNOWLEDGMENT

We gratefully acknowledge the financial support provided to H.J.A. and C.B.L. by an NIH R01 grant (GM049077-10) and an NSF grant (CHE-9982166), to J.L.H. by an NIH R01 grant and NSF, and to T.R.P. by an NIH NRSA fellowship (1F32HD083790A1) and an NIGMS MORE Institutional Research and Academic Career Development Award (http://prof.ucdavis.edu, grant K12GM00679). We also thank Drs. Robert Yen and Bruce Macher for the mass spectrometry analysis and amino acid determination of the *Xenopus* cortical granule lectin.

Received for review April 21, 2003. Accepted August 1, 2003.

AC034414X

(26) Song, Y.; Wardrip, N. J.; Shimoda, Y.; Hedrick, J. L. Unpublished results.