

Multiple Enzyme Approach for the Characterization of Glycan Modifications on the C-Terminus of the Intestinal MUC2Mucin

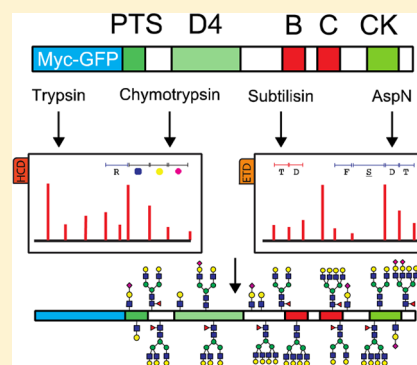
Sjoerd van der Post, Kristina A. Thomsson, and Gunnar C. Hansson*

Department of Medical Biochemistry, University of Gothenburg, Box 440, 405 30 Gothenburg, Sweden

Supporting Information

ABSTRACT: The polymeric mucin MUC2 constitutes the main structural component of the mucus that covers the colon epithelium. The protein's central mucin domain is highly O-glycosylated and binds water to provide lubrication and prevent dehydration, binds bacteria, and separates the bacteria from the epithelial cells. Glycosylation outside the mucin domain is suggested to be important for proper protein folding and protection against intestinal proteases. However, glycosylation of these regions of the MUC2 has not been extensively studied. A purified 250 kDa recombinant protein containing the last 981 amino acids of human MUC2 was produced in CHO-K1 cells. The protein was analyzed before and after PNGase F treatment, followed by in-gel digestion with trypsin, chymotrypsin, subtilisin, or Asp-N. Peptides were analyzed by nLC/MS/MS using a combination of CID, ETD, and HCD fragmentation. The multiple enzyme approach increased peptide coverage from 36% when only using trypsin, to 86%. Seventeen of the 18 N-glycan consensus sites were identified as glycosylated. Fifty-six N-glycopeptides covering 10 N-glycan sites, and 14 O-glycopeptides were sequenced and characterized. The presented method of protein digestion can be used to gain better insights into the density and complexity of glycosylation of complex glycoproteins such as mucins.

KEYWORDS: mucus, O-glycosylation, N-glycosylation, MUC2, chymotrypsin, subtilisin, ETD



INTRODUCTION

It was for a long time assumed that mucus merely served as a simple lubricator and a physical barrier against the surroundings. Today it is clear that the proteins building mucus have very specific properties, with an important role for the O-glycans in harboring the commensal bacterial flora.^{1,2} Mucus and its main component, MUC2 mucin, is secreted from the goblet cells that are scattered in between the intestinal enterocytes. The MUC2 protein (~550 kDa) constitutes the major protein component of mucus and is characterized by two PTS-domains, rich in proline, serine, and threonine, that become densely O-glycosylated to form the extended mucin domains.³ The attached glycans are estimated to increase the size of the molecule five times (~2.5 MDa) and to provide a nutritional source for the gut flora as well as making the mucin resistant against proteolytic degradation by endogenous and bacterial proteases.^{4,5} The N- and C-terminal parts of the MUC2 molecule are involved in di- and trimerization, enabling the formation of an enormous disulfide cross-linked net that is further polymerized to form the mucus gel upon secretion.^{6–8}

We have discovered that the inner stratified mucus layer in the colon is firmly attached to the epithelial cells and devoid of bacteria, as it acts as a physical barrier separating the bacteria from the underlying epithelial cells.¹ On the other hand, the commensal microbiota in the colon resides in the outer mucus layer, which has a more open structure that allows bacteria to enter. Results from mouse models have suggested that this

process is controlled by proteases.¹ Both layers exist in germ-free mice, indicating that the transition is independent of bacterial proteolytic activities and must be due to endogenous proteases.¹ Although we have identified additional proteins that are secreted in conjunction with MUC2,⁹ proteomics studies have so far not revealed any other proteins than MUC2 to be essential for the transition from the inner to the outer colonic mucus layer. It has therefore been proposed that the transition from the inner firm to the outer loose mucus layer involves proteolytic cleavages in the MUC2 molecule. The nature of the disulfide bonded covalent network formed by MUC2 and our earlier studies clearly suggest that the transition from the outer to the inner mucus layer is accompanied by cleavages in the MUC2 C-terminal part.¹ Studies of such proteolytic processing events require detailed biochemical knowledge of the MUC2 protein backbone and its posttranslational modifications. We and others have characterized the O-glycosylation of the central PTS-domains of MUC2 from the various parts of the human gastrointestinal tract.^{10–12} However, little is known about the glycosylation of the N- and C-terminal parts outside the central mucin domains of the human MUC2.

Due to its size and complexity, MUC2 has proven to be an extremely challenging molecule to study. Therefore, we have generated recombinant proteins encompassing selected regions

Received: August 20, 2014

Published: November 18, 2014

of the MUC2 molecule and have used these proteins to study the function of the individual parts of the MUC2 molecule.^{6,13,14} The aim of the current project was to obtain as complete protein backbone sequence coverage as possible of a recombinant MUC2 C-terminal protein expressed in CHO-K1 cells. This region of the protein contains the last 981 amino acids of the human MUC2 with ~10% cysteine residues, 18 N-glycosylation (N-X-S/T) consensus sites, and many potential O-glycosylation sites.

Recent developments in the field of mass spectrometry, and applied methodologies for selective purification and analysis of glycosylated peptides, have resulted in an increasing number of studies where glycan modifications are both characterized and localized.^{15–17} The adaptation of fragmentation techniques such as high-energy collisional dissociation (HCD) and electron-transfer dissociation (ETD) together with the introduction of high-resolution mass analyzers such as the Orbitrap hybrid mass spectrometer have facilitated elucidation of modified peptides.^{18,19} The analysis of fragment ions with high mass accuracy permits detection of diagnostic glycan oxonium ions that are used for rapid identification of MS/MS spectra containing glycopeptides.²⁰ ETD has the advantage of fragmenting the peptide backbone while still retaining the modifications, and it has successfully been applied to the analysis of glycosylated peptides.²¹ The analysis of glycopeptides from a complex protein sample requires selective enrichment steps or the release of the glycan moiety to facilitate analysis.^{19,22} Alternative studies have used less specific proteases to digest glycoproteins, and by this they leave only short peptides containing glycans.^{23,24} Here we aim to obtain as high protein backbone peptide coverage as possible of a purified protein in order to identify as many modified amino acids as possible. The main obstacle in this type of analysis is the limited number of tryptic cleavage sites in the protein backbone of the MUC2, resulting in poor sequence coverage. To overcome this problem, we explored the possibility to use multiple proteolytic enzymes in combination with ETD and HCD fragmentation. In this way we obtained a near full sequence coverage of the MUC2 C-terminus (MUC2-C, amino acids 4,198–5,179 of the human MUC2) and were able to characterize its posttranslational modification.

■ EXPERIMENTAL PROCEDURES

Materials

HPLC grade acetonitrile and analytical grade formic acid and trifluoroacetic acid were obtained from Sigma-Aldrich. Trypsin of proteomics grade (Promega), α -chymotrypsin (bovine pancreas), and subtilisin (protease from *Bacillus licheniformis* type VIII) were obtained from Sigma-Aldrich. Sequencing grade endoproteinase Asp-N and N-glycosidase F (PNGase F) were purchased from Roche Applied Science.

Sample preparations

Expression and purification of the recombinant C-terminus of MUC2 were performed as previously described.¹³ Briefly, CHO-K1 cells were transfected with a plasmid encoding for the C-terminal region (bases 12622–15708) of the human MUC2 with the addition of a Myc-tag and GFP, downstream of the immunoglobulin κ -chain signal sequence. Expressing clones were expanded and cultured in Iscove's modified Dulbecco's medium (Gibco), and media was harvested every 3–4 days. Collected supernatant was filtered through <100 kDa cutoff filters (YM-100; Millipore), to remove smaller components and

for buffer exchange to 50 mM Tris/HCl buffer (pH 8.0). MUC2-C was purified from the supernatant using ion exchange chromatography (MonoQ HR, 5/5, GE Healthcare), followed by anion exchange (Superose 6 HR 10/30, GE Healthcare) of the pooled MUC2-C containing fractions. Aliquots of the purified protein were reduced in 15 mM DTT for 15 min at 95 °C and then for 2 h at 37 °C, followed by alkylation in 40 mM iodoacetamide in the dark for 30 min. Protein aliquots used for de-N-glycosylation were dialyzed on 10,000 MW cutoff Slide-a-Lyzer cups (Millipore) against a 0.1 M Tris-HCl buffer pH 7.5, followed by addition of PNGase F, and left at 37 °C overnight. The samples were concentrated under vacuum using a SpeedVac, and 1 μ g of protein per condition was analyzed by SDS-PAGE on 8% gels.²⁵ The gels were developed using Imperial protein stain (Pierce), and the major band migrating at 250 kDa corresponding to the construct was excised and digested with the following enzymes: (I) endoproteinase Asp-N (1:50) in 25 mM Tris/HCl pH7.8, overnight at 37 °C, (II) trypsin (1:100) in 50 mM (NH)₄HCO₃, overnight at 37 °C, (III) chymotrypsin (1:100) in 50 mM (NH)₄HCO₃, overnight at 30 °C, (IV) subtilisin in 25 mM (NH)₄HCO₃, 2 h at 20 °C. The digestion reactions were quenched by the addition of 25 μ L of 0.1% formic acid, and peptides were extracted from the gel pieces by the addition of 30 μ L of 50% acetonitrile in water (twice). Extracted peptides were dried under vacuum and resolved in 15 μ L of 0.1% trifluoroacetic acid prior to analysis.

Mass spectrometry analysis

The protein digests were analyzed by online nLC-MS/MS using an Orbitrap XL mass spectrometer (Thermo Scientific) with the same configuration as described previously.¹⁴ A C18 stationary phase Reprosil-Pur 120 C18-AQ, 3 μ m (Dr Maisch gmbH, Germany) was used for both the pre- and analytical columns (2 cm \times 100 μ m i.d. and 20 cm \times 50 μ m i.d., respectively). The gradient was run at 5–50% B (A: 0.2% formic acid, B: acetonitrile) over 100 min at a flow rate of ~250 nl per minute. Data were acquired in a data-dependent mode automatically switching between full scan (m/z 350–1600) and MS/MS acquisition. Full scans were collected at a resolution of 60,000 at 400 m/z with the lock mass option enabled for real time calibration (m/z 371.101). The HCD scans for the glycoproteomics analyses were performed on three of the five most abundant ions from the full scan, a resolution of 7500, maximum 250 ms injection time, AGC (automatic gain control) of 100,000, an isolation width of 4 Da, and a normalized collision energy at 30%. For ETD, AGC target values were set to 30,000 and 200,000 for the fluoranthene cations ions and the precursors, respectively. The reaction time was set to 100 ms, and supplemental activation was enabled. Raw files from analyses of the different enzyme digests were converted to peaklists using extract_msn.exe (version 2, Thermo Scientific), and database searches were performed using Mascot (version 2.3, Matrix Science). Fixed modifications were set for: carbamidomethyl (C), variable modifications: oxidation (M) and deamidation (NQ) for all samples. Peptide tolerance: \pm 5 ppm and \pm 100 ppm for fragment ions; peptide charge: 2, 3, and 4+, and Mascot peptide ion score cutoff was set to 20. For the identification of copurified proteins the data were searched against all rodent entries in Swiss-Prot (v 2013_7, entries 26,115) combined with a list of known contaminants. At least two unique peptides were required for positive identification. Enzyme settings for Asp-N (cleavage N-terminal of D and E), chymotrypsin semispecific (F, L, Y, and

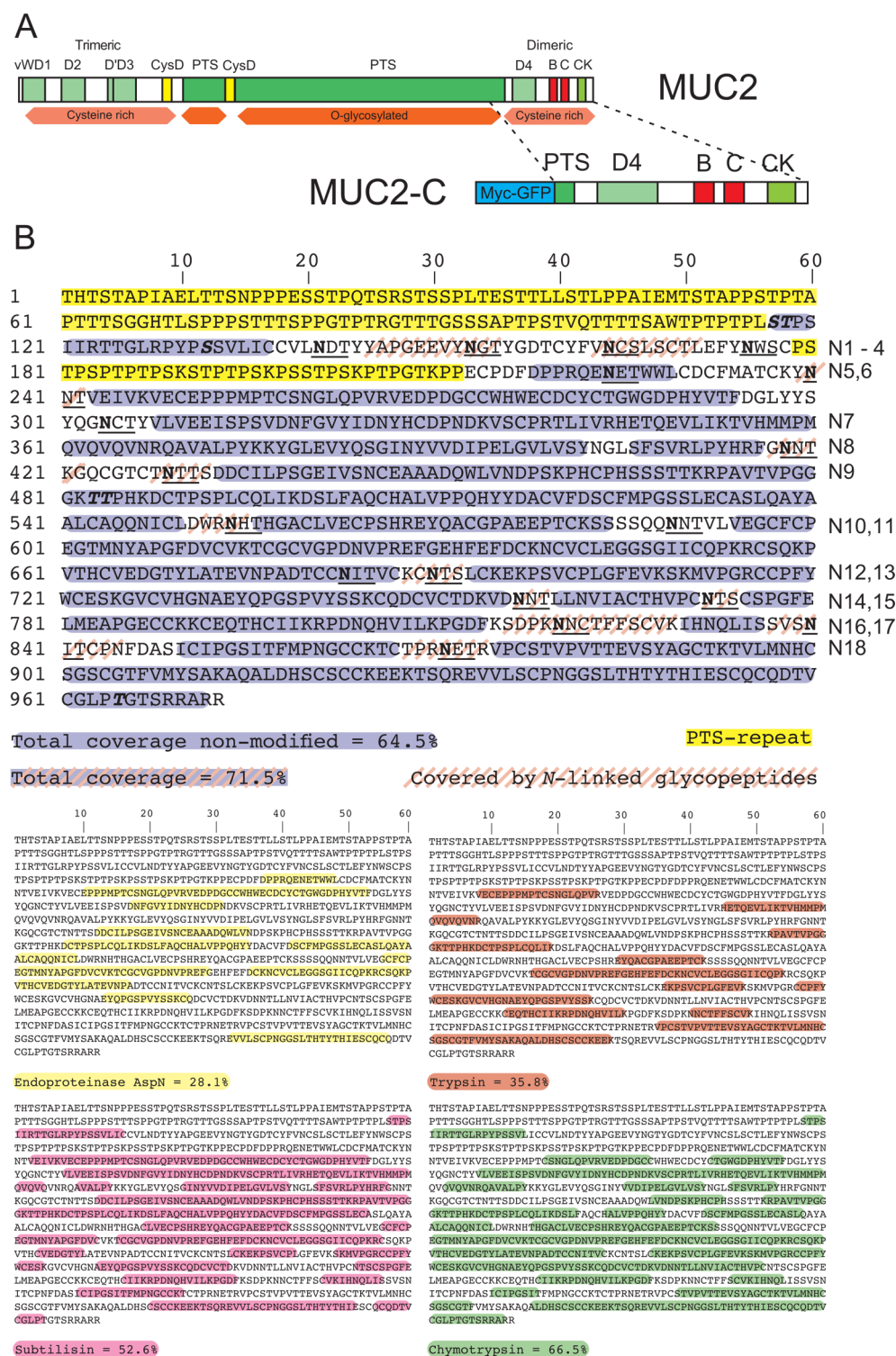


Figure 1. Overview of the MUC2-C sequence and the obtained sequence coverage. (A) The full MUC2 protein with the C-terminal highlighted. The recombinant protein used in this study is composed of the last 981 amino acids with the addition of a Myc-tag and GFP protein at the N-terminal. (B) Total peptide coverage (shaded in gray) from the combined MS analyses of tryptic, chymotryptic, subtilisin, or Asp-N digests. The GFP and Myc-tags are excluded from the figure. N-Glycosylation consensus sites (N-X-S/T) are underlined and in bold and named N1 to N18. Identified O-glycosylation sites (S/T) are highlighted in bold. Regions covered by peptides identified in the individual enzymatic digests are depicted below and shaded in color.

W) and subtilisin set to "none" to obtain no restrictions with respect to cleavage sites. The tryptic and Asp-N digests were searched against the database, allowing for one missed cleavage, and the chymotryptic and subtilisin digests were searched against the recombinant MUC2 C-terminal construct sequence

concatenated to an in-house database containing compiled mucins sequences (<http://www.medkem.gu.se/mucinbiology>). Manual selection and interpretation of HCD spectra of N- and O-glycopeptides were based on the detection of diagnostic oxonium ions (m/z 204.08 (HexNAc), 292.00 (Neu5Ac),

Table 1. N-Glycopeptides Identified in the Different Enzymatic Digests of MUC2-C Analyzed by LC/MS/MS

Start	End	Sequence	Site	Enzyme	MW calcd ^a		Glycopep. Obsd ^b		Glycan composition ^c					Rel. Amt ^d
					glycan	peptide	[M + 2H] ²⁺	[M + 3H] ³⁺	Δppm	NeuAc	Fuc	Hex	HexNAc	
239	247	YNNITVEIVK	²⁴⁰ N6	TRY	2352.85	1078.56		1144.81	1.9	0	0	7	6	++
				TRY	2643.94	1078.56		1241.84	0.8	1	0	7	6	++
				TRY	2790	1078.56		1290.53	0.7	1	1	7	6	+
				TRY	2935.04	1078.56		1338.87	0.6	2	0	7	6	++
				TRY	3226.13	1078.56		1435.91	0.4	3	0	7	6	++
416	421	FGNNITK	⁴¹⁸ N8	TRY	1768.64	679.33	1224.99	938.71	0.3	0	1	5	4	+
				TRY	2133.77	679.33	1408.06		-1.7	0	1	6	5	+
				TRY	2424.87	679.33		1035.74	0.2	1	1	6	5	++
				TRY	2498.9	679.33		1060.42	-0.8	0	1	7	6	+
				TRY	2715.96	679.33		1132.77	0	2	1	6	5	++
				TRY	2790	679.33		1157.45	-0.9	1	1	7	6	++
428	442	TNTTSDDCILPSGEI	⁴²⁹ N9	TRY	3081.09	679.33		1254.48	-0.2	2	1	7	6	++
				SUB	1622.58	1621.71		1082.44	-2.7	0	0	5	4	+
				SUB	1913.68	1621.71		1179.47	-1	1	0	5	4	++
				SUB	2204.77	1621.71		1276.51	-2.6	2	0	5	4	++
679	685	DTCCNIT	⁶⁸³ N12	SUB	1216.42	882.32	1050.38		-0.9	0	0	5	2	+
				SUB	1257.45	882.32	1070.89		-0.6	0	0	4	3	+
				SUB	1403.51	882.32	1143.92		-1.5	0	1	4	3	++
				SUB	1606.59	882.32	1245.46	830.65	0	0	1	4	4	+++
689	695	CNTSLCK	⁶⁹⁰ N13	SUB	1768.64	882.32	1326.49	884.66	-2	0	1	5	4	+++
				TRY	1257.45	881.37	1070.42		-0.1	0	0	4	3	+
				TRY	1622.58	881.37	1252.98	835.66	1.5	0	0	5	4	+
				TRY	1768.64	881.37	1326.01	884.35	1.4	0	1	5	4	++
				TRY	1913.68	881.37		932.69	-0.2	1	0	5	4	+
688	693	KCNTSL OR NTSICK	⁶⁹⁰ N13	TRY	2059.73	881.37		981.38	-0.2	1	1	5	4	+
				CHY	1216.42	721.34	969.89		-1	0	0	5	2	++
				CHY	1622.58	721.34	1172.97		-1.3	0	0	5	4	++
				CHY	1768.64	721.34	1246		-0.5	0	1	5	4	++
754	761	KVDNNITLL	⁷⁵⁷ N14	SUB	1768.64	915.50	1343.08	895.72	-0.7	0	1	5	4	++
				SUB	2059.73	915.50	1489	992.75	0.2	1	1	5	4	++
				SUB	2424.87	915.50		1114.47	-1	1	1	6	5	++
				SUB	2790	915.50		1236.18	-1.1	1	1	7	6	+
743	761	SSKQDCVCTDKVDNNITLL	⁷⁵⁷ N14	CHY	1768.64	2255.98		1342.55	-0.2	0	1	5	4	++
				CHY	2059.73	2255.98		1439.58	0.3	1	1	5	4	++
				CHY	2350.83	2255.98		1536.61	0.2	2	1	5	4	++
767	774	THVPCNTS	⁷⁷² N15	SUB	1403.51	914.39	1159.96		-0.2	0	1	4	3	+
				SUB	1768.64	914.39	1342.52	895.35	-1.1	0	1	5	4	++
				SUB	2059.73	914.39	1488.07	992.38	-1.2	1	1	5	4	+++
				SUB	2350.83	914.39		1089.42	-0.3	2	1	5	4	++
820	829	NNCTFFSCVK	⁸²¹ N16	SUB	2424.87	914.39		1114.09	-0.4	1	1	6	5	+
				TRY	1768.64	1275.54		1015.73	-1	0	1	5	4	++

Table 1. continued

Start	End	Sequence	Site	Enzyme	MW calcd ^a		Glycopep. Obsd ^b		Glycan composition ^c					Rel. Amt ^d
					glycan	peptide	[M + 2H] ²⁺	[M + 3H] ³⁺	Δppm	NeuAc	Fuc	Hex	HexNAc	
816	829	SDPKNNCTFFSCVK	821N16	TRY	2059.73	1275.54		1112.77	-1.1	1	1	5	4	++
				TRY	2350.83	1275.54		1209.8	-0.3	2	1	5	4	+
				TRY	1768.64	1702.74		1158.14	-2.3	0	1	5	4	++
813	825	DFKSDPKNNCTFF	821N16	TRY	2059.73	1702.74		1255.17	0.1	1	1	5	4	++
				SUB	2059.73	1618.71		1227.16	-0.7	1	1	5	4	+++
814	824	FKSDPKNNCTFF or KSDPKNNCTFF	821N16	SUB	2350.83	1618.71		1324.19	-2.3	2	1	5	4	+
				CHY	1768.64	1356.61		1042.76	-1.9	0	1	5	4	++
				2059.73	1357.62	1356.61	1139.79	-1	1	1	5	4	+++	
838	845	VSNITCPN	840N17	CHY	2350.83	1356.61		1236.82	-1.1	2	1	5	4	+
				SUB	1216.42	903.41	1060.93		-1.4	0	0	5	2	
837	845	SVSNIITCPN ^e	840N17	SUB	1216.42	990.44	1104.44		-1.3	0	0	5	2	
866	878	TCTPRNETRVPCS or CTPRNETRVPCST	871N18	SUB	2059.73	1576.71		1213.16	0.1	1	1	5	4	++
				2350.83	1577.72	1576.71	1310.19	-0.8	2	1	5	4	++	
				SUB	2424.87	1576.71	1334.87	-1.6		1	1	6	5	+

^aTheoretical peptide and glycan mass. ^bGlycopeptide mass measured in the orbitrap (details are described in Methods section). ^cNeuAc = N-acetylneuraminic acid; Hex = (Man and Gal); HexNAc = N-acetylhexosamine (GlcNAc); Fuc = Fucose. ^dRelative amounts are estimated by comparing peak intensities within the group of closely eluting glycoforms of the same peptide. ^eCoeluting precursor ions with composition GlcNAc₂, GlcNAc₂Man₁₋₃ were also detected, but may be caused by in-source fragmentation, since they are not biologically relevant.

366.11 (HexHexNAc), and 657.24 (HexHexNAcNeu5Ac)) using the Xcalibur browser (Thermo Scientific). The FindPept (for nonspecifically cleaved peptides) and GlycoMod tool (<http://www.expasy.org>) were used for identification of potential peptide candidates in the MUC2-C sequence, and all interpreted glycopeptides were identified within 5 ppm accuracy.

RESULTS

Sequence coverage of MUC2-C after de-*N*-glycosylation

The MUC2-C recombinant protein is made up of 1,300 amino acids containing a signal sequence followed by a MycTag, a GFP protein, and the last 981 amino acids of the human MUC2 protein¹³ (Figure 1A). Purified, reduced, and alkylated protein was analyzed by gel electrophoresis, and the fully glycosylated protein migrated at approximately 250 kDa. The expected mass of the protein is 187 kDa, indicating a large number of posttranslational modifications. The band was subjected to in-gel digestion with trypsin and analyzed by nLC-MS/MS. Database searches confirmed that the MUC2-C was the major component of the excised band. Twelve additional proteins were also detected and are compiled in Supporting Information Table S1. The tryptic peptide coverage of the MUC2-C (excluding the Myc and GFP tags) was 36%. None of the *N*-glycan motifs was found deamidated prior to PNGase F digestion, suggesting that all were modified by *N*-glycans; two asparagine containing peptides that were not part of a motif were found deamidated although only a minor fraction compared to the unmodified peptide, and likely introduced during sample preparation.

The sample was then deglycosylated with PNGase F and reanalyzed. Deglycosylation increased the peptide coverage of MUC2-C to 39%, and the obtained tryptic peptides from the de-*N*-glycosylated MUC2-C are compiled in Supporting Information Table S2. By this approach, all obtained peptides in the size range between 7 and 16 amino acids expected from a tryptic digest of MUC2-C were observed, with the exception of three peptides (Figure 1B).

To increase the peptide coverage, we evaluated other enzymes for protein digestion (one μ g protein). Asp-N, subtilisin, and chymotryptic digestions of the deglycosylated recombinant protein were analyzed by mass spectrometry, and the results are shown in Figure 1B and compiled in Supporting Information Tables S3, S4, and S5, respectively. The majority of the peptides from the chymotryptic were formed by nonspecific cleavages other than the C-terminal of the chymotrypsin typical amino acids F, L, Y, and W. The peptide coverage was therefore considerably improved when searches were set to semispecific digestion for the peptide identification. The chymotryptic and subtilisin digest were searched against a limited database compiled in-house containing only mucin sequences including the recombinant MUC2-C protein. 101 and 67 unique peptides were detected in the subtilisin and chymotryptic digest, respectively, covering 67% and 53% of the MUC2-C sequence. As expected, subtilisin was found to cleave nonspecifically after large uncharged amino acids, and many of these subtilisin derived peptides were overlapping on limited stretches in the protein backbone.

Most of the chymotryptic cleavages site were found C-terminally of large hydrophobic residues, as previously reported in the literature;²⁶ however, we also observed that cleavages with both subtilisin and chymotrypsin frequently occurred C-

terminally of iodoacetamide alkylated cysteines. Chymotrypsin and subtilisin were found to provide both overlapping peptides and complementary regions. The enzyme Asp-N cleaves N-terminally of the amino acids aspartic acid (D), and at a limited rate glutamic acid (E). Only 23 peptides were detected in the Asp-N digest, mostly in areas already covered by the other enzymes. The poor sequence coverage can be explained by the low number and poor distribution of aspartic acids in the MUC2-C sequence, resulting in a limited number of peptides suitable for mass spectrometry detection. When combining the results from the four enzymes, the total peptide coverage of the MUC2-C was increased to 86% (Figure 1B). The remaining noncovered regions were in the N-terminal part of MUC2-C (amino acids 1–58 and 180–210, Figure 1A). These two regions are rich in the amino acids proline, serine, and threonine and as such are expected to be highly *O*-glycosylated, and potential proteolytic cleavage sites are likely shielded by glycan moieties.

Site-specific *N*-glycosylation characterization from HCD fragmentation spectra

Seventeen of the 18 potential *N*-glycosylation sites in MUC2-C were detected as deamidated peptides after PNGase F treatment in the tryptic, chymotryptic, and subtilisin digests, and as such assumed to be modified. All the 17 sites appeared to be glycosylated, and the only site with a partial occupancy was found at N5 (Figure 1B). For this site a nonmodified peptide was detected in the Asp-N digest and a deamidated peptide in the chymotryptic digest (Supporting Information Tables S3 and S5).

To confirm the presence of *N*-glycans on the MUC2-C, protein digests were analyzed with LC-MS/MS using HCD fragmentation with orbitrap detection. HCD spectra corresponding to glycopeptides were initially selected for manual interpretation by screening for diagnostic ions in the lower mass range at m/z 204 (HexNAc), 366 (HexHexNAc), and m/z 657 (HexHexNAcNeuAc). In total, 56 *N*-glycopeptides were identified from the tryptic, subtilisin, and chymotryptic digests of the MUC2-C, including 38 glycoforms at 10 of the 18 potential *N*-glycosylation consensus sites. No *N*-glycopeptides were detected in the Asp-N digest. Results from the site-specific *N*-glycosylation survey are shown in Table 1.

The identified *N*-glycopeptides were modified by a broad variety of high mannose-, hybrid-, and complex type structures with up to seven different glycoforms per site. This is in line with previous analyses of wild type CHO cell *N*-glycosylation showing a broad heterogeneity.²⁷ Four *N*-glycosylation sites were covered by tryptic glycopeptides, three with chymotrypsin, and seven with subtilisin. In cases when glycopeptides covering the same *N*-glycosylation site were found in more than one enzymatic digest, the same *N*-glycan composition was generally found, together with other less abundant structures (Table 1, e.g. ⁶⁹⁰N13, ⁷⁵⁷N14, and ⁸²¹N16). De-*N*-glycosylation by PNGase F allowed subtilisin to cleave on the *N*-terminal side of the deamidated asparagine at ²²⁴N5, ⁴²⁹N9, and ⁷⁷²N15.

The majority of the glycopeptides eluted early on the reverse phase column, prior to the respective nonmodified peptides. However, glycopeptides from the subtilisin and chymotryptic digests covering site ⁷⁵⁷N14 eluted after the majority of the nonmodified peptides, and the glycopeptide covering site ⁸²¹N16 identified in all three enzymatic digests eluted even later. These long retention times reflect the hydrophobic nature of the peptides. Interestingly, the deglycosylated peptides

eluted shortly before the glycosylated versions, suggesting that the surface accessibility was affected by removal of the glycans. Examination of the full scans in areas where glycopeptides eluted revealed that all major glycopeptides were selected for fragmentation despite the narrow elution time windows.

Approximately 20% of the HCD spectra containing glycopeptide oxonium ions could not be interpreted due to weak fragment ions. This is likely due to the low abundance of the precursor and the fact that different glycopeptides require different collision energies to generate a sufficient number of fragment ions, as has been reported to be essential for HCD.²⁰ The observed precursor ions were doubly or triply charged $[M + 2H]^{2+}$ or $[M + 3H]^{3+}$ ions; however, glycopeptides above approximately 2500 Da were largely triply charged independently of the size of the peptide. The peptide size varied from 6 up to 19 amino acids. The HCD spectra were similar whether generated from tryptic or from nontryptic glycopeptides, but occasionally chymotryptic and subtilisin digests revealed fragments from the peptide backbone.

HCD fragmentation of a core fucosylated N-glycopeptide ¹⁵KSDPKNNCTFF from the chymotryptic digest of MUC2-C

The HCD spectrum of an N-glycopeptide at m/z 1042.76 $[M + 3H]^{3+}$ from a chymotryptic digest is shown in Figure 2. The

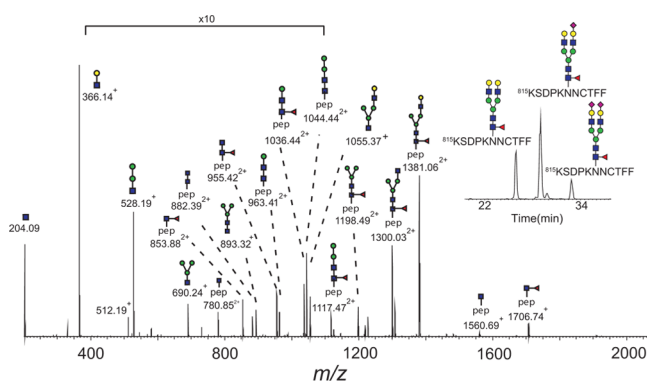


Figure 2. HCD fragmentation spectrum of the N-glycopeptide from a chymotryptic digest of MUC2-C detected at m/z 1042.76 $[M + 3H]^{3+}$, interpreted to have the sequence ⁸¹⁵KSDPKNNCTFF (site ⁸²¹N16, Figure 1B) and the N-glycan with the composition (Hex₅HexNAc₄Fuc). The mass chromatograms from LC/MS of the three N-glycopeptides with the same sequence are inserted.

lower region of the HCD spectra contains the characteristic diagnostic glycan oxonium b-ions at m/z 274/292, 366, 528, 657, and 893, which confirms the presence of a glycopeptide. However, in order to interpret the HCD spectrum, the main challenge is to identify the series of monoprotonated or doubly protonated y1 ions, which are abundant during HCD fragmentation and contain the peptide and the first GlcNAc residue.²⁰ In this spectrum, the presence of fragment ions from the sequential loss of residues from both the fucosylated core and the nonfucosylated glycopeptides made it difficult to identify this ion. However, the smallest doubly charged ions in the spectrum at m/z 780.85 $[M + 2H]^{2+}$ and m/z 853.88 $[M + 2H]^{2+}$ differed by 146, which corresponds to a Fuc residue. These ions correspond to the y1 ions of a core fucosylated N-glycan [peptide + GlcNAc + 2H⁺] and [peptide + GlcNAcFuc + 2H⁺], respectively. Subtracting the proposed N-glycan mass [Hex₅HexNAc₄Fuc] = 1768.639 from the precursor ion mass measured in the full scan resulted in two candidates within 10

ppm, corresponding to the peptide sequences ⁸¹⁵KSDPKNNCTFF at Δ ppm = 0.3.

HCD fragmentation of glycopeptides ⁷⁶⁷THVPCNTS and ⁶⁷⁹DTCCNIT from the subtilisin digest of MUC2-C

The pooled mass chromatograms of five glycoforms which showed similar HCD spectra are found in Figure 3A. The base peak chromatograms as well as the full scans between 14.5 and 15.0 min show that these closely eluting peaks belong to different glycoforms of the same peptide (Figure 4B). The HCD spectra of the doubly charged precursor ion of the nonsialylated N-glycan at m/z 1342.52 $[M + 2H]^{2+}$ and the triply charged precursor ion of the monosialylated N-glycan at m/z 992.38 $[M + 3H]^{3+}$ are shown in Figure 3C and D, respectively. As in the previous example, a mixture of fragment ions originating from sequential loss of monosaccharide residues from glycopeptides with and without core fucosylation was observed. The fragment ions at m/z 1118.5 and 1264.5 correspond to the peptide with the core GlcNAc and the corresponding fucosylated variant, allowing identification of the peptide as ⁷⁶⁷THVPCNTS (site ⁷⁷²N15, Figure 1A). In the HCD spectra of the nonsialylated glycopeptide, additional fragment ions were found at m/z 781.29 and m/z 915.39, which were interpreted as the nonglycosylated peptide and the peptide + GlcNAc that had lost the N-terminal amino acids Thr-His-Val (Figure 3C). Fragment ions from cleavages in the peptide backbone were consistently found in all the spectra of doubly charged precursor ions of these glycoforms but not in the triply charged ones.

We detected five glycoforms of a peptide eluting at 16.2–16.7 min that was interpreted as ⁶⁷⁹DTCCNIT (site ⁶⁸³N12, Figure 1A), of which two peptides were of a less abundant high mannose/hybrid type and three of a complex type (Table 1). The HCD spectrum of the high mannose type (GlcNAc₂Man₂) glycoform detected at m/z 1050.38 $[M + 2H]^{2+}$ is shown in Figure 4A. The four fragment ions at m/z 764.23, 883.32, 967.34, and 1086.40 were present in all HCD spectra of the respective glycoforms. The ions at m/z 883.32 and 1086.40 were assigned to the protonated peptide and the y1 ion (peptide + GlcNAc), and the m/z 764.23 and 967.34 ions were assigned to the same fragments after loss of the C-terminal threonine. Additional fragment ions at m/z 866.30 and 747.24 were interpreted as secondary and tertiary fragment ions formed after loss of the glycan, ammonia (17.02), and threonine from the peptide. A fragment ion at m/z 651.18 was interpreted as being formed from loss of the N-glycan and Ile-Thr from the C-terminal end, supporting the interpretation of the peptide sequence. A second glycoform on the same peptide was found at m/z 884.66 $[M + 3H]^{3+}$ with the glycan composition [Hex₅HexNAc₄Fuc] as shown in Figure 4B.

Site-specific O-glycosylation of MUC2-C digested with Asp-N, trypsin, chymotrypsin, or subtilisin

Spectra of 15 O-glycopeptides from four different sites were elucidated from HCD experiments of enzymatic digests of the MUC2-C, and the results are compiled in Supporting Information Table 6. Four additional O-glycopeptides from the GFP/Myc-tag regions were detected, but these are not discussed further. Based on the sequence, we would expect to identify more O-glycosylated sites. However, the first 120 amino acids of the recombinant MUC2-C terminal are almost solely composed of serine and threonine residues with the potential to become heavily O-glycosylated; hence, no peptides were observed due to the lack of proteolytic cleavage sites and

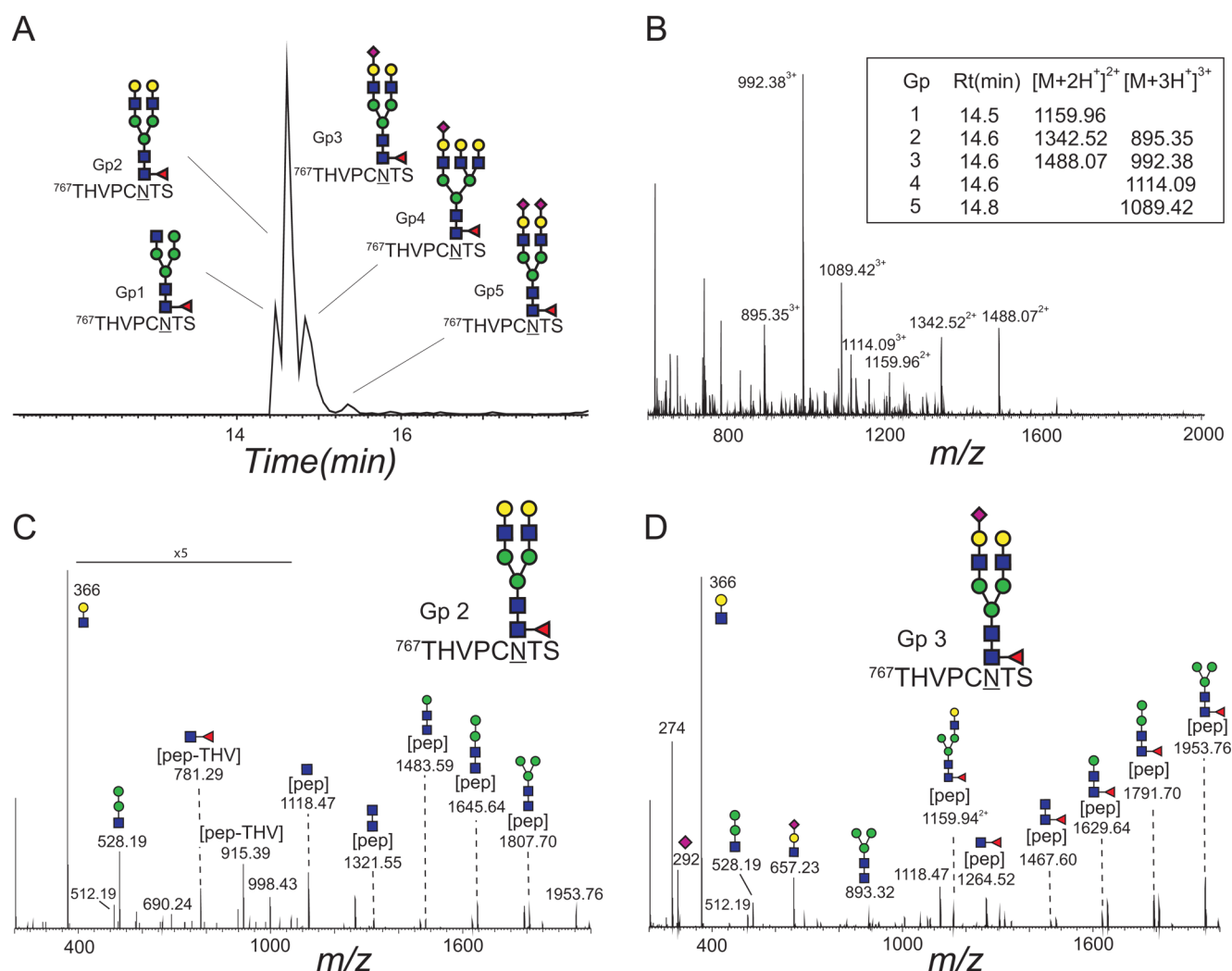


Figure 3. Pooled mass chromatograms of five *N*-glycopeptides labeled Gp 1–5 interpreted as glycoforms of the peptide ⁷⁶⁷THVPCNTS (site ⁷⁷²N15, Figure 1) from a subtilisin digest of MUC2-C (A) and pooled full scans collected between 14.5 and 15 min (B). Detected precursor ions and retention times of Gp1–5 are inserted in B. HCD fragmentation spectra of Gp 2 detected at *m/z* 1342.52 [*M* + 2*H*]²⁺ and Gp 3 detected at *m/z* 992.38 [*M* + 3*H*]³⁺ are shown in C and D, respectively.

potential glycan density. *O*-Glycopeptides were distinguished from *N*-glycopeptides by a more simple fragmentation pattern, and all the observed *O*-glycans were core 1 type *O*-glycans with 1–4 sugar residues, as previously reported for CHO-K1 cells.^{27,28} *O*-Glycan sites were only identified in three areas in the peptide backbone: between amino acids Ser118 and Ser132, on Thr483 or Thr484, and between Thr965 and Ser968 (Figure 1B). The HCD spectrum of the highly abundant ion at *m/z* 872.92 [*M* + 2*H*]²⁺ corresponded to an *O*-glycopeptide trisaccharide from the chymotryptic digest (Figure 5). This structure was dominated by three large fragment ions corresponding to sequential loss of the monosaccharide residues NeuAc, NeuAc + Hex, and NeuAc + Hex + HexNAc from the peptide, and the diagnostic ions at *m/z* 274, 292, and 366. The proposed peptide mass observed at *m/z* 1088.60 [*M* + *H*]⁺ matched the sequence ¹²⁶GLRPYPSSVL in MUC2-C Δppm = 1, and in addition fragment ions at *m/z* 771.41, 858.44, and 957.51 confirmed the proposed peptide sequence, and correspond to the peptide after loss of the amino acids S, V, and L from the C-terminal end. The ion at *m/z* 587.33 corresponds to the N-terminal amino acids ¹²⁶GLRPY. As the assigned peptide contained two potential glycosylated Ser

residues, we reanalyzed the chymotryptic material using ETD fragmentation, which made it possible to assign Ser132 (¹²⁶GLRPYPSSVL) as the glycosylation site (Supporting Information Table S6).

DISCUSSION

In the present study we demonstrate that substantial information on site-specific glycosylation of a large purified glycoprotein can be obtained when combining standard proteomics methods without including additional preparative steps selecting for glycopeptides. This is an attractive approach to use when it is possible to obtain relatively small amounts (low μg) of highly purified material. Most methodologies described for glycopeptide enrichment will have the drawback of selecting for specific types of *N*-glycans or *O*-glycans, such as lectin based approaches or by chemically modifying the glycans.^{29,30} Here, the use of multiple digestive enzymes and HCD fragmentation enabled the identification of 56 *N*- and 15 *O*-glycopeptides covering 10 *N*- and 4 *O*-glycosylation sites of a purified recombinant protein expressed in CHO-K1 cells containing the last 981 amino acids of the human MUC2 mucin.

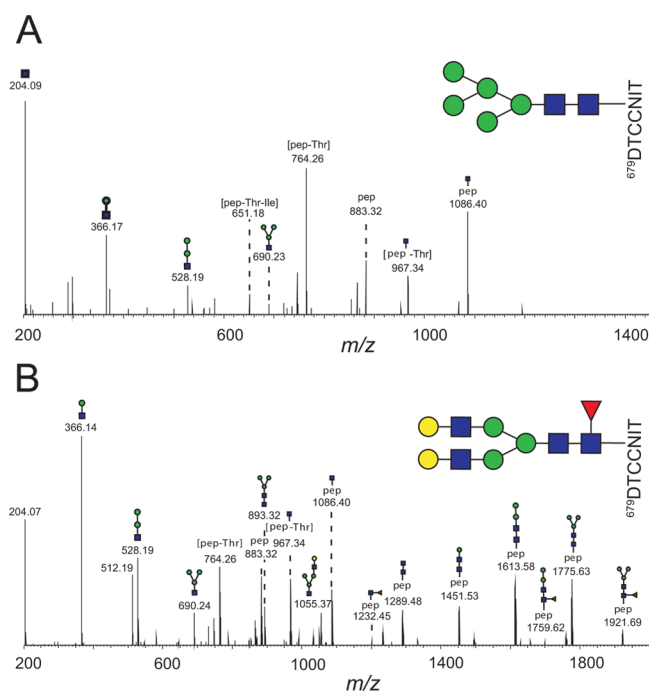


Figure 4. HCD fragmentation spectra of *N*-glycopeptides from a subtilisin digest of MUC2-C interpreted to have the sequence $^{679}\text{DTCCNIT}$ (site $^{683}\text{N12}$, Figure 1B) and the monosaccharide compositions $[\text{Hex}_5\text{HexNAc}_2]$ detected at m/z 1050.38 $[\text{M} + 2\text{H}]^{2+}$ (A) and $(\text{Hex}_5\text{HexNAc}_4\text{Fuc})$ at m/z 884.66 $[\text{M} + 3\text{H}]^{3+}$ (B).

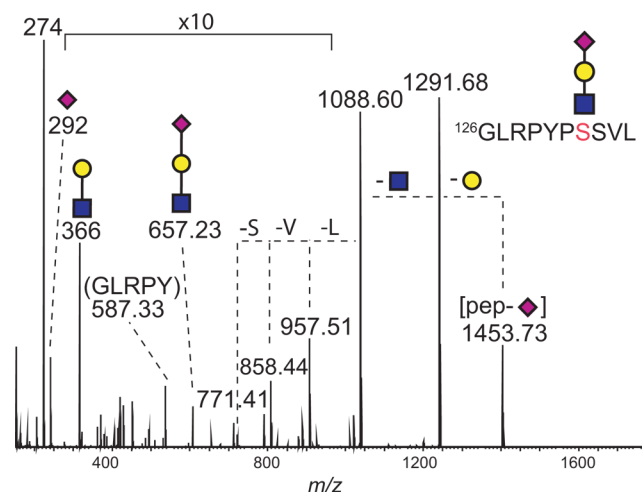


Figure 5. HCD fragmentation spectrum of an *O*-glycopeptide at m/z 872.92 $[\text{M} + 2\text{H}]^{2+}$ from a chymotryptic digest of MUC2-C interpreted to have the sequence $^{126}\text{GLRPYPSSVL}$ (site Ser132, Figure 1B) and the monosaccharide composition (HexHexNAc-NeuAc) .

The majority of the MUC2-C derived glycopeptides eluted early, as expected distributed over several minutes, allowing time for fragmentation of multiple glycoforms. The MUC2-C protein does not contain sufficient numbers of tryptic cleavage sites to generate reasonable sized peptides required for complete sequence coverage. To overcome this problem, we used additional enzymes for protein digestion, such as Asp-N and the less specific enzymes chymotrypsin and subtilisin. Chymotrypsin and subtilisin have been unattractive to use compared to trypsin, as these enzymes are less specific

compared to trypsin and generally generate smaller peptides with no fixed basic amino acids. The small size and reduced information obtained via fragmentation result in reduced peptide identifications and overall lower peptide confidence scores.^{31,32} However, when working with a purified protein and high-resolution mass spectrometry, these drawbacks can be overcome. In our analyses, digestion with chymotrypsin generated the highest sequence coverage. One reason for this was that overnight digestion generated many for these enzymes nontypical peptides. In total, we identified 64 (of 101) chymotryptic peptides formed by cleavages after residues other than L, W, F, and Y. We also found that chymotrypsin frequently cleaved C-terminally of cysteines alkylated with iodoacetamide (13 peptides), something that was advantageous in our case, as the MUC2 mucin is rich in cysteines.

Details on the characteristics of HCD spectra from tryptic *N*-glycopeptides have been published by others.^{18,19,33} We here show that nontryptic glycopeptides generate equally informative fragmentation spectra when analyzed using high-resolution HCD fragmentation. However, not all spectra were straightforward to interpret due to the presence of secondary fragment ions from Fuc on the core GlcNAc, and the presence of fragment ions originating from cleavages in the amino acid sequence. The diagnostic γ -fragment ion series containing the peptide and one or two core monosaccharide residues were detected in all 56 spectra but were not always among the more abundant ions. In addition to the nonspecific enzymes subtilisin and chymotrypsin, we also tried proteinase K, which has recently been used by others for glycoproteomics.^{23,24} However, digestion using proteinase K generated many overlapping glycopeptides of the same site which made spectral data interpretation time-consuming.

MUC2-C produced in CHO cells was found to be *N*-glycosylated on 17 out of 18 *N*-glycosylation consensus sites (N-X-S/T) and only one consensus site that was shown unmodified (N5, Figure 1B). The *N*-glycans of the individual sites were found to differ, with abundant core fucosylation on some sites and little or none on others. Terminal NeuAc was absent on N12, and larger *N*-glycans were found on sites N6 and N8 than on the other sites. We did not detect any NeuGc, something that might be because the CHO cell clone used or because the cells were grown serum free. A single *N*-glycan of high mannose type was detected at site N17, suggesting that protein folding may modify *N*-glycan trimming on this particular site when the protein passes through Golgi. As the protein was produced in CHO-K1 cells, the *N*-glycans composition might not reflect the structures found on the native MUC2 protein, although the same subset of glycosyltransferases involved in *N*-glycosylation is expressed in both species, which indicates that similar structures can be synthesized.³⁴ The domain structure of the human MUC2 mucin C-terminal is identical when compared to that of rodents, and the majority of the consensus sequences for *N*-glycosylation is conserved.³⁵

A total of seven *N*-glycan motifs were only covered after de-*N*-glycosylation, and their composition remains unknown. The first four motifs (*i.e.* $^{141}\text{N1}$, $^{153}\text{N2}$, $^{164}\text{N3}$, and $^{175}\text{N4}$) were not identified and are found in close proximity to each other with a limited number of potential enzymatic cleavage sites, which is a likely reason why no glycopeptides were identified in this region. In addition there might be *N*-glycopeptides derived from other sites with unfavorable properties for the analysis by reverse phase chromatography, or resulting in fragmentation

spectra with low ion intensities or limited sequence information. The identification of these missing sites might require N-glycopeptide enrichment using hydrophilic interaction chromatography (HILIC) prior to MS analyses to reduce the complexity of the samples by removing the majority of the unmodified peptides.³⁶

The human MUC2 C-terminus shows extensive sequence homology with the two other secreted human mucins MUC5B and MUC5AC and the von Willebrand factor (vWF) especially when it comes to the distribution of Cys residues. Interestingly, the vWF differs from the mucins, which are more similar, on the localization of the N-glycan sites. Only one single N-glycan site is conserved between vWF and human MUC2, and this is not conserved in two other secreted mucins (MUC5AC and MUC5B). This common Asn is N-glycosylated in vWF and has been shown to be essential for the folding/function of vWF, as mutations at this site give the severe von Willebrand type III disease.³⁷ N-Glycans are added cotranslationally and before folding and disulfide bond formation, suggesting that N-glycans are utilized to generate differently folded proteins and maybe different disulfide bonds.

There are to date 20 different GalNAc transferases described to be involved in O-glycan synthesis in the Golgi, adding the first monosaccharide residue to serine or threonine initiating O-glycosylation. These transferases have differing substrate specificity and have also been shown to be expressed in a cell and developmental stage specific manner (reviewed by Bennett et al.³⁸). Consequently the site specific O-glycosylation found on a recombinant protein produced in CHO-K1 cells will of course depend on the repertoire of GalNAc-transferases present in these cells and not fully reflect the MUC2 present in the human intestine.^{5,39} However, the identification of O-glycans on MUC2-C produced in a cell line still reveals important information. These O-glycosylated areas prove that these regions of the protein backbone must be situated on the exterior of the folded molecule. In this context, our O-glycan data fits well with our previously published work on certain pathogenic bacterial and parasite proteases which can disrupt the mucus gel.^{5,40} These proteases were found to target the same specific site in the MUC2 C-terminal, and the addition of a single GalNAc in proximity of the cleavage site abolished the proteolytic activity (⁴³⁷LSTPSIIR↓T↓TGLRPYPSSVL; arrows show cleavage sites and underlines show the observed glycosylation site in the present study).

Proteolytic processing of mucins by endogenous or bacterial proteases is important for maintaining a functional colonic mucus layer. A prerequisite for understanding these complex processes is knowledge of protein modifications, especially glycosylation. As we identified the majority of the modified sites on the C-terminal of MUC2, our approach can be used as a reference when studying site-specific glycosylation on mucus collected from clinical samples. The protein quantities used for this method are possible to obtain from these sources, although one has to consider the increase in complexity by introducing peptides derived from the N-terminal of MUC2 and other mucus associated proteins. We have now shown that by using multiple enzymes for protein digestion in combination with high-resolution mass spectrometry and different types of ion fragmentation one can obtain detailed insight into complex and large glycoproteins such as the mucins.

■ ASSOCIATED CONTENT

§ Supporting Information

Table S1. Proteins identified in the tryptic digest of de-N-glycosylated MUC2-C purified from CHO-K1 cells and analyzed with SDS-PAGE and LC/MS/MS. Table S2. Peptides identified in the tryptic digest of de-N-glycosylated MUC2-C. Table S3. Peptides identified in the chymotryptic digest of de-N-glycosylated MUC2-C. Table S4. Peptides identified in the subtilisin digest of de-N-glycosylated MUC2-C. Table S5. Peptides identified in the Asp-N digest of de-N-glycosylated MUC2-C. Table S6. O-glycopeptides from enzymatic digests of MUC2-C analyzed with LC/MS/MS. This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Author

* E-mail: gunnar.hansson@medkem.gu.se. Telephone: +46 31 7863488. Fax: +46 31 7862150.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the Swedish Research Council (no. 7461, 21027), The Swedish Cancer Foundation, The Knut and Alice Wallenberg Foundation, the IngaBritt and Arne Lundberg Foundation, Sahlgrén's University Hospital (LUA-ALF), Wilhelm and Martina Lundgren's Foundation, Torsten och Ragnar Söderbergs Stiftelser, The Sahlgrén's Academy, the National Institute Of Allergy And Infectious Diseases (U01AI095473, the content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH), and The Swedish Foundation for Strategic Research—The Mucus-Bacteria-Colitis Center (MBC) of the Innate Immunity Program.

■ REFERENCES

- (1) Johansson, M. E. V.; Phillipson, M.; Petersson, J.; Velcich, A.; Holm, L.; Hansson, G. C. The inner of the two Muc2 mucin-dependent mucus layers in colon is devoid of bacteria. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 15064–15069.
- (2) Staubach, F.; Künzel, S.; Baines, A. C.; Yee, A.; McGee, B. M.; Bäckhed, F.; Baines, J. F.; Johnsen, J. M. Expression of the blood-group-related glycosyltransferase B4galnt2 influences the intestinal microbiota in mice. *ISME J.* **2012**, *6*, 1345–1355.
- (3) Johansson, M. E. V.; Sjövall, H.; Hansson, G. C. The gastrointestinal mucus system in health and disease. *Nat. Rev. Gastroenterol. Hepatol.* **2013**, *10*, 352–361.
- (4) Hooper, L. V.; Gordon, J. I. Glycans as legislators of host–microbial interactions: spanning the spectrum from symbiosis to pathogenicity. *Glycobiology* **2001**, *11*, 1R–10R.
- (5) van der Post, S.; Subramani, D. B.; Bäckström, M.; Johansson, M. E. V.; Vester-Christensen, M. B.; Mandel, U.; Bennett, E. P.; Clausen, H.; Dahlén, G.; Sroka, A.; et al. Site-specific O-glycosylation on the MUC2 mucin protein inhibits cleavage by the *Porphyromonas gingivalis* secreted cysteine protease (RgpB). *J. Biol. Chem.* **2013**, *288*, 14636–14646.
- (6) Godt, K.; Johansson, M. E. V.; Lidell, M. E.; Mörgelin, M.; Karlsson, H.; Olson, F. J.; Gum, J. R.; Kim, Y. S.; Hansson, G. C. The N terminus of the MUC2 mucin forms trimers that are held together within a trypsin-resistant core fragment. *J. Biol. Chem.* **2002**, *277*, 47248–47256.
- (7) Asker, N.; Bäckström, D.; Axelsson, M. A.; Carlstedt, I.; Hansson, G. C. The human MUC2 mucin apoprotein appears to dimerize before O-glycosylation and shares epitopes with the

“insoluble” mucin of rat small intestine. *Biochem. J.* **1995**, 308 (Pt 3), 873–880.

(8) Ambort, D.; Johansson, M. E.; Gustafsson, J. K.; Nilsson, H. E.; Ermund, A.; Johansson, B. R.; Koeck, P. J.; Hebert, H.; Hansson, G. C. Calcium and pH-dependent packing and release of the gel-forming MUC2 mucin. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, 109, 5645–5650.

(9) Johansson, M. E. V.; Thomsson, K. A.; Hansson, G. C. Proteomic Analyses of the Two Mucus Layers of the Colon Barrier Reveal That Their Main Component, the Muc2Mucin, Is Strongly Bound to the Fcgbp Protein. *J. Proteome Res.* **2009**, 8, 3549–3557.

(10) Capon, C.; Maes, E.; Michalski, J. C.; Leffler, H.; Kim, Y. S. Sd(a)-antigen-like structures carried on core 3 are prominent features of glycans from the mucin of normal human descending colon. *Biochem. J.* **2001**, 358, 657–664.

(11) Robbe, C.; Capon, C.; Coddeville, B.; Michalski, J. Structural diversity and specific distribution of O-glycans in normal human mucins along the intestinal tract. *Biochem. J.* **2004**, 384, 307–316.

(12) Holmen Larsson, J. M.; Karlsson, H.; Sjövall, H.; Hansson, G. C. A complex, but uniform O-glycosylation of the human MUC2 mucin from colonic biopsies analyzed by nanoLC/MSn. *Glycobiology* **2009**, 19, 756–766.

(13) Lidell, M. E.; Johansson, M. E. V.; Mörgelin, M.; Asker, N.; Gum, J. R.; Kim, Y. S.; Hansson, G. C. The recombinant C-terminus of the human MUC2 mucin forms dimers in Chinese-hamster ovary cells and heterodimers with full-length MUC2 in LS 174T cells. *Biochem. J.* **2003**, 372, 335–345.

(14) Ambort, D.; van der Post, S.; Johansson, M. E. V.; Mackenzie, J.; Thomsson, E.; Krengel, U.; Hansson, G. C. Function of the CysD domain of the gel-forming MUC2 mucin. *Biochem. J.* **2011**, 436, 61–70.

(15) Wührer, M.; Catalina, M. I.; Deelder, A. M.; Hokke, C. H. Glycoproteomics based on tandem mass spectrometry of glycopeptides. *J. Chromatogr. B: Analyt. Technol. Biomed. Life Sci.* **2007**, 849, 115–128.

(16) Frost, D. C.; Li, L. *Recent advances in mass spectrometry-based glycoproteomics*, 1st ed.; Elsevier Inc.: 2014; Vol. 95, pp 71–123.

(17) Thaysen-Andersen, M.; Packer, N. H. Advances in LC–MS/MS-based glycoproteomics: Getting closer to system-wide site-specific mapping of the N- and O-glycoproteome. *Biochim. Biophys. Acta (BBA)—Proteins Proteomics* **2014**, 1844, 1437–1452.

(18) Scott, N. E.; Parker, B. L.; Connolly, A. M.; Paulech, J.; Edwards, A. V.; Crossett, B.; Falconer, L.; Kolarich, D.; Djordjevic, S. P.; Højrup, P. Simultaneous glycan-peptide characterization using hydrophilic interaction chromatography and parallel fragmentation by CID, higher energy collisional dissociation, and electron transfer dissociation MS applied to the N-linked glycoproteome of *Campylobacter jejuni*. *Mol. Cell Proteomics* **2011**, 10, M000031–MCP201.

(19) Darula, Z.; Medzihradsky, K. F. Affinity enrichment and characterization of mucin core-1 type glycopeptides from bovine serum. *Mol. Cell Proteomics* **2009**, 8, 2515–2526.

(20) Segu, Z. M.; Mechref, Y. Characterizing protein glycosylation sites through higher-energy C-trap dissociation. *Rapid Commun. Mass Spectrom.* **2010**, 24, 1217–1225.

(21) Wiesner, J.; Premisler, T.; Sickmann, A. Application of electron transfer dissociation (ETD) for the analysis of posttranslational modifications. *Proteomics* **2008**, 8, 4466–4483.

(22) Jensen, P. H.; Kolarich, D.; Packer, N. H. Mucin-type O-glycosylation—putting the pieces together. *FEBS J.* **2010**, 277, 81–94.

(23) Clowers, B. H.; Dodds, E. D.; Seipert, R. R.; Lebrilla, C. B. Site Determination of Protein Glycosylation Based on Digestion with Immobilized Nonspecific Proteases and Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *J. Proteome Res.* **2007**, 6, 4032–4040.

(24) Neue, K.; Mormann, M.; Peter-Katalinić, J.; Pohlentz, G. Elucidation of glycoprotein structures by unspecific proteolysis and direct nanoESI mass spectrometric analysis of ZIC-HILIC-enriched glycopeptides. *J. Proteome Res.* **2011**, 10, 2248–2260.

(25) Laemmli, U. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **1970**, 227, 680–685.

(26) Schilling, O.; Huesgen, P. F.; Barr Eacute, O.; Keller, U. A. D.; Overall, C. M. Characterization of the prime and non-prime active site specificities of proteases by proteome-derived peptide libraries and tandem mass spectrometry. *Nat. Protoc.* **2011**, 6, 111–120.

(27) North, S. J.; Huang, H.-H.; Sundaram, S.; Jang-Lee, J.; Etienne, A. T.; Trollope, A.; Chalabi, S.; Dell, A.; Stanley, P.; Haslam, S. M. Glycomics profiling of Chinese hamster ovary cell glycosylation mutants reveals N-glycans of a novel size and complexity. *J. Biol. Chem.* **2010**, 285, 5759–5775.

(28) Bäckström, M.; Link, T.; Olson, F.; Karlsson, H.; Graham, R.; Picco, G.; Burchell, J.; Taylor-Papadimitriou, J.; Noll, T.; Hansson, G. Recombinant MUC1 mucin with a breast cancer-like O-glycosylation produced in large amounts in Chinese-hamster ovary cells. *Biochem. J.* **2003**, 376, 677–686.

(29) Zielinska, D. F.; Gnäd, 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.

(30) Nilsson, J.; Rüetschi, U.; Halim, A.; Hesse, C.; Carlsohn, E.; Brinkmalm, G.; Larson, G. Enrichment of glycopeptides for glycan structure and attachment site identification. *Nat. Methods* **2009**, 6, 809–811.

(31) Pirmoradian, M.; Budamgunta, H.; Chingin, K.; Zhang, B.; Astorga-Wells, J.; Zubarev, R. A. Rapid and deep human proteome analysis by single-dimension shotgun proteomics. *Mol. Cell Proteomics* **2013**, 12, 3330–3338.

(32) Käll, L.; Canterbury, J. D.; Weston, J.; Noble, W. S.; MacCoss, M. J. Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nat. Methods* **2007**, 4, 923–925.

(33) Catalina, M. I.; Koeleman, C. A. M.; Deelder, A. M.; Wührer, M. Electron transfer dissociation of N-glycopeptides: loss of the entire N-glycosylated asparagine side chain. *Rapid Commun. Mass Spectrom.* **2007**, 21, 1053–1061.

(34) Xu, X.; Nagarajan, H.; Lewis, N. E.; Pan, S.; Cai, Z.; Liu, X.; Chen, W.; Xie, M.; Wang, W.; Hammond, S.; et al. The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line. *Nat. Biotechnol.* **2011**, 29, 735–741.

(35) Lang, T.; Hansson, G. C.; Samuelsson, T. Gel-forming mucins appeared early in metazoan evolution. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, 104, 16209–16214.

(36) Wührer, M.; de Boer, A. R.; Deelder, A. M. Structural glycomics using hydrophilic interaction chromatography (HILIC) with mass spectrometry. *Mass Spectrom. Rev.* **2009**, 28, 192–206.

(37) Eikenboom, J. C.; Ploos van Amstel, H. K.; Reitsma, P. H.; Briët, E. Mutations in severe, type III von Willebrand's disease in the Dutch population: candidate missense and nonsense mutations associated with reduced levels of von Willebrand factor messenger RNA. *Thromb. Haemost.* **1992**, 68, 448–454.

(38) Bennett, E. P.; Mandel, U.; Clausen, H.; Gerken, T. A.; Fritz, T. A.; Tabak, L. A. Control of mucin-type O-glycosylation: A classification of the polypeptide GalNAc-transferase gene family. *Glycobiology* **2012**, 22, 736–756.

(39) Yang, Z.; Halim, A.; Narimatsu, Y.; Joshi, H. J.; Steentoft, C.; Schjoldager, K. T.-B. G.; Schulz, M. A.; Sealover, N. R.; Kayser, K. J.; Bennett, E. P.; et al. The GalNAc-type O-glycoproteome of CHO cells characterized by the SimpleCell strategy. *Mol. Cell. Proteomics* **2014**, DOI: 10.1074/mcp.M114.041541.

(40) Lidell, M. E.; Moncada, D. M.; Chadee, K.; Hansson, G. C. Entamoeba histolytica cysteine proteases cleave the MUC2 mucin in its C-terminal domain and dissolve the protective colonic mucus gel. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, 103, 9298–9303.