Determination of the Site-specific O-Glycosylation Pattern of the Porcine Submaxillary Mucin Tandem Repeat Glycopeptide

MODEL PROPOSED FOR THE POLYPEPTIDE:GalNAc TRANSFERASE PEPTIDE BINDING SITE*

(Received for publication, December 19, 1996, and in revised form, January 30, 1997)

Thomas A. Gerken‡§, Cheryl L. Owens¶, and Murali Pasumarty‡
From the W. A. Bernbaum Center for Cystic Fibrosis Research, Departments of Pediatrics, Biochemistry, and Molecular and Microbiology, Case Western Reserve University, Cleveland, Ohio 44106-4948

The heterogeneously glycosylated 81-residue tryptic tandem repeat glycopeptide from porcine submaxillary mucin (PSM) has been isolated and its glycosylation pattern determined by amino acid sequencing. Key to these studies is the ability to trim the structurally heterogeneous PSM oligosaccharide side chains to homogeneous GalNAc monosaccharide side chains by mild trifluoromethanesulfonic acid treatment. Trypsin treatment of trifluoromethanesulfonic acid-treated PSM releases the 81-residue tandem repeat as an ensemble of 81-residue glycopeptides with different glycosylation patterns. Automated amino acid sequencing using Edman degradative chemistry of the repeat was used to determine the extent of glycosylation of nearly every Ser and Thr residue. The Thr residues are all highly glycosylated within the range of 73–90%, giving an average Thr glycosylation of 83%. In contrast, the Ser residues display a wide range of glycosylations, ranging between 33 and 95%, giving an average Ser glycosylation of 74%. These data are consistent with the known elevated glycosylation of Thr peptides over Ser peptides for the porcine UDP-N-acetylgalactosamine:polypeptide N-acetylgalactosaminyltransferase. It is also observed that the extent of glycosylation of the repeat correlates poorly with published predictive methods. An examination of the sequences surrounding the glycosylation sites reveals that nearly all of the highly glycosylated sites have a penultimate Gly residue, whereas those that are less highly glycosylated have medium to large side chain penultimate residues. As observed by others, glycosylation also appears to be modulated by the presence of Pro residues. On the basis of these findings we suggest that the acceptor peptide binds the transferase in a β-like conformation and that penultimate residue side chain steric interactions may play a role in determining extent that a given Ser or Thr is glycosylated. A model for the GalNAc transferase peptide binding site is proposed.

Mucin glycoproteins are heavily O-glycosylated glycoproteins secreted by higher organisms that serve vital roles, protecting and lubricating epithelial cell surfaces from biological, chemical, and mechanical insult. Mucins and mucin-like molecules attached to membrane and cell surfaces play additional important roles by modulating for example immune response, inflammation, and tumorigenesis (1–5). The O-glycosylated domains of mucins and mucin-like glycoproteins typically contain 50–80% carbohydrate and possess expanded conformations. These regions typically contain high Ser and Thr contents, commonly composed of polypeptide tandem repeats containing clusters of Ser and Thr residues. It has been demonstrated, in the case of mucins, that the O-linked oligosaccharide side chains, attached via α-N-acetylgalactosamine (GalNAc)1 to Ser and Thr, are solely responsible for their 3-fold expanded peptide chain dimensions (6). Chemical and NMR studies indicate that on average 75% or more of the Ser and Thr residues in mucins are glycosylated (7, 8). Littler, however, is known of the actual distribution of the carbohydrate in these clusters along the mucin polypeptide core. In addition, it is unknown whether the distribution of oligosaccharides along the peptide core is random or whether specific Ser/Thr residues are preferentially glycosylated over others. Obtaining this information by peptide mapping approaches would be a significant analytical undertaking due to the vast array of glycopeptides with different glycosylation patterns and oligosaccharide structures that would be needed to be quantitatively isolated and characterized.

Several predictive methods, based on the analysis of reported O-glycosylation sites compiled from protein data bases, are available for estimating the relative propensity for a given Ser or Thr to be glycosylated (9–11). Application of these predictive methods to the highly glycosylated mucins may not be fully valid because mucins and mucin-like molecules were not a significant component of the “training” data sets, which were dominated by globular glycoproteins, and where presented, the data on mucins and mucin-like glycoproteins glycosylation are commonly incomplete or in error. Furthermore, the propensity for O-glycosylation of Ser and Thr residues at the surface of globular glycoproteins may vary considerably from those found in mucin-like glycoproteins, which presumably serve different structural functions.

In vitro O-glycosylation studies using synthetic peptide acceptors have also been utilized to determine the propensity for a given Ser and Thr to be O-glycosylated (12–19). Unfortunately, the isolated peptide α-N-acetylgalactosaminylntrans-

The abbreviations used are: α-GalNAc, α-N-acetylgalactosamine; -OG, glycosylated Ser or Thr residue; TFMSA, trifluoromethanesulfonic acid; PTH, phenylthiohydantoin; PSM, porcine submaxillary gland mucin; R-PSM, reduced and carboxymethylated PSM; TR-PSM, trypsin digested R-PSM; TTR-PSM, mild TFMSA treated TR-PSM; TTTR-PSM, trypsin digested TTR-PSM; GTTTR-PSM, Glu-C digested TTR-PSM; HPLC, high performance liquid chromatography.
ferases give different extents of Ser/Thr glycosylation on peptide substrates compared with that observed in vivo (14, 20). This may be the result of altered enzyme specificity as a result of inappropriate solution conditions, absence of cofactors, the presence of more than one transferase (21–23), or the result of the artifacts resulting from the use of relatively small peptides containing charged N and C termini. Only by characterizing and quantifying the specific in vivo glycosylation of native and/or expressed recombinant proteins and peptides (20) is it likely that valid and useful data will be obtained for determining the true in vivo transferase specificities.

To begin to address the structural effects of O-glycosylation and to determine the extent that mucin O-glycosylation is modulated in vivo in a site-specific manner, we have undertaken the isolation and characterization of the porcine submaxillary gland mucin (PSM)-glycosylated tandem repeat. We have determined the glycosylation pattern of the isolated 81-residue tandem repeat and have isolated and characterized several smaller PSM tandem repeat-derived glycopeptides (24). This work provides the first detailed analysis of the glycosylation pattern of a mucin and suggests that mucin glycosylation is modulated by peptide sequence but not entirely as expected by the existing O-glycosylation prediction algorithms. Based on the observed glycosylation pattern a model for the GaINAc transferase peptide binding site has been proposed.

**EXPERIMENTAL PROCEDURES**

**Materials**

α-GalNAc-Thr and α-GalNAc-Ser were a kind gift of R. Koganti, Biomerica Inc. Edmonton, Alberta, Canada. Except where noted, all chemicals and enzyme reagents were obtained from Sigma or Fisher.

**Isolation of PSM**—Porcine submaxillary gland mucin was obtained from frozen porcine submaxillary glands, in gram quantities, as described by Shogren et al. (25). Reduction, Carboxymethylation, and Trypsinolysis of PSM—PSM was reduced and carboxymethylated by the methods of Gupta and Jenoff (32) giving R-PSM. R-PSM (1 g/50 ml) was digested with 1-tosylalanyl-2-phenylthioethyl chloromethyl ketone (15 mg/g) R-PSM ( Worthington) overnight at 37 °C in 50 mm ammonium bicarbonate, pH 8.3. A second aliquot of trypsin was added to ensure complete digestion and incubated 5–8 h. Tolueno was added to both incubation solutions to prevent microbial growth. After exhaustive dialysis and the removal of insoluble debris by centrifugation, trypsinized R-PSM (TR-PSM) was lyophilized.

**Partial Deglycosylation by Trifluoromethanesulfonic Acid (TFMSA)**—The TTTR-PSM-T3 tandem repeat and Glu-C-digested C-terminal tandem repeat peptides on reverse phase HPLC. The mixture was fractionated by S200 chromatography. The major glycopeptide peak (GTTTR-PSM) was isolated and characterized. The N- and C-terminal tandem repeat peptides on reverse phase HPLC.

**HPLC Purification of PSM Tandem Repeat and Glu-C-digested Tandem Repeat**—The TTTR-PSM-T3 tandem repeat and Glu-C-digested repeat (GTTTR-PSM) were further purified by reverse phase HPLC chromatography on a 0.46 × 15 cm C18 ODS II column (Alltech Associates Inc., Dearfield IL) using 0.05% trifluoroacetic acid, water/faceto- nitric gradient as described in the figure legend. All isolations were performed on a Varian 5000 HPLC system (Varian Associates, Walnut Creek CA) equipped with a Schamitsu UV/VIS detector.

**Amino Acid Sequencing**—Pulsed liquid phase Edman degradation amino acid sequencing of the isolated PSM tandem repeats and Glu-C-derived glycopeptide was performed on either an Applied Biosystems 477A or Applied Biosystems Premise 494 protein sequencer (Perkin-Elmer) typically using standard manufacturer recommended pulsed-liquid cycles (24). Samples of 2000–5000 pmol were dried on trifluoroacetic acid washed glass fiber filters (ABI number 401111) spotted with 1.5 mg of BioBrene Plus (ABI 400385). Amino acid phenylthiohydantoin (PTH) derivatives were chromatographed on standard ABI 5-μm C18 PTH columns using the Fast Normal 1 gradient program and were monitored by the absorbance at 289 nm. The PTH-Thr/Ser-O-GalNAc derivatives were found to elute as two diastereotopic peaks at unique positions in the chromatogram (24). Due to increased peak broadening and changes in elution position as the PTH columns age, some variability and overlap of the glycosylated PTH derivatives with Ser, Thr, Gly, and Asp PTH derivatives were commonly observed. Based on the ratio of the areas of the diastereotopic PTH-Ser/Thr-O-GalNAc derivatives was found to be relatively constant (typically 45/55 for Thr), the extent of glycosylation usually could be determined by use of simple algegra. Modifications in the HPLC gradient were found to produce only marginal improvement in the separation of the PTH-Thr/Ser-O-GalNAc derivatives from the elution positions of neighboring PTH-derivatives. Prior to data analysis, long term cycle prior and lag for each PTH-derivative was eliminated by a base-line subtraction approach. This was performed by subtracting the 5-cycle minimum value running average across the entire sequencing run for each peak. Due to the length of the peptides and high content of similar amino acids (i.e. Gly, Ser, Thr, and Ala), no attempts were made to include quantifications for adjacent residue cycle lag or preview. Response factors for the PTH-Thr/Ser-O-GalNAc derivatives were obtained from shorter glycopeptide sequencing experiments (24) and were found to be similar to those of Ser and Thr. Thr/Ser and their glycosylated PTH-derivatives were further adjusted for each sequencing run to obtain consistent picomole yields relative to the non-Ser/Thr residues.

**Prediction of O-Glycosylation Sites and Secondary Structure**—The PSM tandem repeat sequence was analyzed for potential O-glycosylation sites using software kindly provided by Dr. A. Elhammer (9) and by using the E-mail NetOglyc server of Hansen et al. (10). The sequence coupled vector projection predictions were kindly performed by Dr. K. Chou (11). Peptide secondary structure predictions were performed by the SOPM internet server (29). Predictions were performed for at least 10 residues beyond the tandem repeat N and C-terminal boundaries to eliminate end effects.

**Peptide Modeling**—Peptides were modeled using the Biopolymer module of InsightII (MSI Inc. formally Biosym Technologies, San Diego CA).

**RESULTS**

**Isolation of PSM Tandem Repeat Glycopeptides**—The structure of porcine submaxillary mucin (PSM) polypeptide based on the nucleotide sequencing of Timpte et al. (30) and Eckhardt et al. (31) is shown in Fig. 1A. The mucin’s structure is dominated by the presence of highly O-glycosylated, multiple repeating 81-residue tandem repeats. These repeats make up the vast majority of the mucin’s amino acid sequence and represent the major glycosylation sites in PSM.

**Trypsic Tandem Repeat Glycopeptide**—There are two potential trypsin cleavage sites, Arg-Ile and Arg-Pro, in each PSM tandem repeat. The Arg-Pro site, however, is expected to be inactive; therefore, trypsin treatment is expected to yield single

---

**PSM Tandem Repeat Glycosylation Pattern**—The TTTR-PSM-T3 tandem repeat (30 mg) in 5 ml of 25 mM ammonium bicarbonate, pH 7.8, was digested with 1 mg of protease Glu-C (Boehringer Mannheim) for 20 h at 25 °C. After a second addition of Glu-C and further digestion, the mixture was fractionated by S200 chromatography. The major glycopeptide peak (GTTTR-PSM) was isolated and characterized.

**Partial Deglycosylation by Trifluoromethanesulfonic Acid (TFMSA)**—The TTTR-PSM-T3 tandem repeat and Glu-C-digested repeat (GTTTR-PSM) were further purified by reverse phase HPLC chromatography on a 0.46 × 15 cm C18 ODS II column (Alltech Associates Inc., Dearfield IL) using 0.05% trifluoroacetic acid, water/faceto- nitric gradient as described in the figure legend. All isolations were performed on a Varian 5000 HPLC system (Varian Associates, Walnut Creek CA) equipped with a Schamitsu UV/VIS detector.

**Amino Acid Sequencing**—Pulsed liquid phase Edman degradation amino acid sequencing of the isolated PSM tandem repeats and Glu-C-derived glycopeptide was performed on either an Applied Biosystems 477A or Applied Biosystems Premise 494 protein sequencer (Perkin-Elmer) typically using standard manufacturer recommended pulsed-liquid cycles (24). Samples of 2000–5000 pmol were dried on trifluoroacetic acid washed glass fiber filters (ABI number 401111) spotted with 1.5 mg of BioBrene Plus (ABI 400385). Amino acid phenylthiohydantoin (PTH) derivatives were chromatographed on standard ABI 5-μm C18 PTH columns using the Fast Normal 1 gradient program and were monitored by the absorbance at 289 nm. The PTH-Thr/Ser-O-GalNAc derivatives were found to elute as two diastereotopic peaks at unique positions in the chromatogram (24). Due to increased peak broadening and changes in elution position as the PTH columns age, some variability and overlap of the glycosylated PTH derivatives with Ser, Thr, Gly, and Asp PTH derivatives were commonly observed. Based on the ratio of the areas of the diastereotopic PTH-Ser/Thr-O-GalNAc derivatives was found to be relatively constant (typically 45/55 for Thr), the extent of glycosylation usually could be determined by use of simple algebra. Modifications in the HPLC gradient were found to produce only marginal improvement in the separation of the PTH-Thr/Ser-O-GalNAc derivatives from the elution positions of neighboring PTH-derivatives. Prior to data analysis, long term cycle prior and lag for each PTH-derivative was eliminated by a base-line subtraction approach. This was performed by subtracting the 5-cycle minimum value running average across the entire sequencing run for each peak. Due to the length of the peptides and high content of similar amino acids (i.e. Gly, Ser, Thr, and Ala), no attempts were made to include quantifications for adjacent residue cycle lag or preview. Response factors for the PTH-Thr/Ser-O-GalNAc derivatives were obtained from shorter glycopeptide sequencing experiments (24) and were found to be similar to those of Ser and Thr. Thr/Ser and their glycosylated PTH-derivatives were further adjusted for each sequencing run to obtain consistent picomole yields relative to the non-Ser/Thr residues.

**Prediction of O-Glycosylation Sites and Secondary Structure**—The PSM tandem repeat sequence was analyzed for potential O-glycosylation sites using software kindly provided by Dr. A. Elhammer (9) and by using the E-mail NetOglyc server of Hansen et al. (10). The sequence coupled vector projection predictions were kindly performed by Dr. K. Chou (11). Peptide secondary structure predictions were performed by the SOPM internet server (29). Predictions were performed for at least 10 residues beyond the tandem repeat N and C-terminal boundaries to eliminate end effects.

**Peptide Modeling**—Peptides were modeled using the Biopolymer module of InsightII (MSI Inc. formally Biosym Technologies, San Diego CA).
copies of the 81-residue tandem repeat with the sequence given in Fig. 1B. Indeed, Timpte and co-workers (30) have shown that after full deglycosylation and digestion of the fully deglycosylated (apo) mucin with trypsin, this trypsin peptide is obtained. In contrast, Gupta and Jentoft (32) have shown that trypsin treatment of fully glycosylated, reduced and carboxymethylated mucin fails to yield single copies of the repeat and instead yields relatively high molecular weight glycopeptides composed of undegraded tandem repeats. Apparently, the longer oligosaccharide side chains inhibit the digestion of the PSM polypeptide by trypsin, preventing the isolation of individual tandem repeats. The Sephacryl S200 gel filtration chromatogram of this species, TR-PSM, is given in Fig. 2A.

Only after quantitatively trimming the oligosaccharide side chains to the peptide-linked α-GalNAc residue by mild TFMSA treatment (26–27), Fig. 2B, does the PSM glycopeptide core become susceptible to cleavage by trypsin, Fig. 2C, presumably releasing monomeric glycosylated trypsin tandem repeats (peak T3). Mild TFMSA treatment does not appreciably degrade the mucin, since on gel filtration untreated and treated mucin (TR-PSM and TTR-PSM respectively) contain similar high molecular weight glycosylated peaks, Fig. 2A and B. Integrations of the α-carbon resonances of the 13C NMR spectra of native and TTR-PSM confirm that 96 to 97% of the glycosylated Ser and Thr residues of the native mucin retain intact unsubstituted α-GalNAc residues after mild TFMSA treatment (data not shown, see Ref. 26).

The suspected monomeric glycosylated trypsin tandem repeat, pooled as indicated in Fig. 2C, gave a single sharp peak after rechromatography on S200 as shown in Fig. 2D (■). On reverse phase HPLC, Fig. 3A, this peak gives a single somewhat broadened peak, representing the ensemble of heterogeneously glycosylated tandem repeats. This peak was pooled for amino acid sequencing and for subsequent digestion by Glu C (discussed below). Amino acid sequencing of this peak confirmed the isolation of the PSM tandem repeat.

Since each step in the above procedure typically yields a single major glycopeptide species that is pooled for the subsequent step, the obtained tandem repeat glycosylation pattern (see below) is expected to represent the majority of the PSM tandem repeats of the native mucin. After correcting for the different carbohydrate contents of native and TFMSA-modified PSM (26), it is calculated that between 40 and 50% (uncorrected for the nonspecific losses of material at each step) of the initial TR-PSM peptide is isolated in the pooled TTR-PSM-T3 fraction. The possibility exists, however, that a subpopulation of differently glycosylated species may have been excluded, since the lower molecular weight regions of the glycopeptide peaks in Fig. 2, B and C, were not pooled. These lower molecular weight species are thought to represent nonspecific (protease and TFMSA) degradation products of the tandem repeat and other non-tandem repeat glycopeptides arising from the N- and C-terminal domains of PSM (see Fig. 1A). Evidence for

Fig. 1. Polypeptide structure and trypsin tandem repeat sequence of the porcine submaxillary gland mucin (PSM). A, a model of the PSM polypeptide based on the nucleotide sequencing of Timpte et al. (30) and Eckhardt et al. (31) and the biophysical studies of Gupta and Jentoft (32), Shogren et al. (6), and Perez-Vilar et al. (33). The PSM polypeptide consists of a very long highly extended O-glycosylated domain containing multiple copies, n, of an 81-residue tandem repeat which is followed by a relatively small, poorly glycosylated C-terminal domain. The glycosylated domain consists of several thousand residues comprising 25 or more tandem repeats (31). The C-terminal (and perhaps the N-terminal) domains have been shown to dimerize, thus accounting for the very large size of native mucin (33).

B, the sequence of the trypsin tandem repeat.

Fig. 2. Sephacryl S200 gel filtration chromatography showing the isolation of the PSM trypsinic tandem repeat and its digestion by Glu C. For A–C protein content is monitored by the absorbance at 229 nm (■) and carbohydrate content by periodic acid-Shiff, absorbance at 555 nm (○). A, S200 chromatogram of trypsinized, reduced, and carboxymethylated PSM (TR-PSM), B, chromogram of TR-PSM after partial deglycosylation by TFMSA at 0 °C giving TTR-PSM (see “Experimental Procedures”), C, S200 chromatogram of the indicated TTR-PSM fraction in B after digestion with trypsin, yielding TTTT-PSM. The third peak, T3, represents the trypsinicized glycosylated PSM tandem repeat. D, S200 chromatogram of the glycosylated trypsin tandem repeat, T3, from C (■), absorbance at 229 nm) and the chromatogram of Glu-C digested T3 (isolated on reverse phase HPLC, Fig. 3A) giving GTTTR-PSM (○, absorbance at 229 nm).

The appearance of peaks T1 and T2 in Fig. 2C, after the second trypsinolysis, varies from preparation to preparation, ranging from their absence to what is observed in Fig. 2C. Carbon-13 NMR spectroscopy of these fractions reveals the presence of longer oligosaccharide side chains; therefore, these peaks represent incompletely deglycosylated mucin that is resistant to complete trypsin digestion. The higher carbohydrate content of the trypsin-resistant fractions is also shown by their higher carbohydrate to protein ratios (absorbance 555/229 nm) compared with the trypsinic tandem repeat, T3.
nonspecific protease degradation arises from the observation that the use of non-L-1-tosylamido-2-phenylethyl chloromethyl ketone-treated trypsin gives a T3 peak that is further broadened to lower molecular weight. By eliminating these lower molecular weight species we are able to reduce the background sequencing “noise,” thereby permitting the sequencing of longer segments of the tandem repeat glycopeptide (discussed below). We believe, however, that these degradation products would not have significantly different glycosylation patterns compared with the intact tandem repeat.

Glu-C Glycopeptides—The PSM tandem repeat contains three potential cleavage sites for endoproteinase Glu-C from Staphylococcus aureus V8 (cleaving at C terminus of Glu). Digestion of the tryptic PSM tandem repeat with Glu-C will therefore produce three glycopeptides of 40, 38, and 3 residues each, proceeding from the N to C terminus, respectively. We chose to isolate the 38-residue glycopeptide (residues 39–78) since its analysis would further help confirm the glycosylation pattern of the C-terminal half of the tandem repeat.

The HPLC-purified tryptic tandem repeat (Fig. 3A) was digested with Glu-C and fractionated on S200 chromatography, Fig. 2D. As shown in the figure the Glu-C digest (○) migrates at a lower molecular weight than undigested tryptic repeat (□). The 40- and 38-residue glycopeptides are not expected to be resolved on S200; therefore, the production of a single sharp peak after Glu-C digestion indicates the repeat has been fully cleaved by the enzyme. The remaining 3-residue glycopeptide was not specifically isolated or identified but presumably appears near the included volume of the column approximately fractions 120–140.

On reverse phase HPLC the pooled product of the Glu-C digest (pooled as indicated in Fig. 2D, ○) reveals a complex pattern comprised of two major broad peaks, labeled (GTTTR-PSM)-GII and (GTTTR-PSM)-GIII, as shown in Fig. 3B. Peak GII elutes at a lower acetonitrile content than the intact tryptic tandem repeat, whereas peak GIII elutes at a similar acetonitrile content as the intact tryptic tandem repeat. On the basis of the expected differences in hydrophobicities, the least hydrophobic fraction, GII, was tentatively assigned to the least hydrophobic glycopeptide 39–78 (containing 11 hydrophobic residues: Ala, Pro, Val, and Ile), and the more hydrophobic fraction, GIII, was tentatively assigned to glycopeptide 1–38 (containing 15 hydrophobic residues). Proton NMR spectroscopy, at 600 MHz, of each pooled fraction confirmed their identities based on their different amino acid compositions (data not shown). Fraction GTTTR-PSM-GII was unambiguously identified as residues 39–78 on the basis of its amino acid sequence, presented below.

Amino Acid Sequencing of PSM Tandem Repeat Glycopeptides—Amino acid sequencing was performed on the reverse phase HPLC-pooled product of the Glu-C digest (pooled as indicated in Fig. 2D, ○) reveals a complex pattern comprised of two major broad peaks, labeled (GTTTR-PSM)-GII and (GTTTR-PSM)-GIII, as shown in Fig. 3B. Peak GII elutes at a lower acetonitrile content than the intact tryptic tandem repeat, whereas peak GIII elutes at a similar acetonitrile content as the intact tryptic tandem repeat. On the basis of the expected differences in hydrophobics, the least hydrophobic fraction, GII, was tentatively assigned to the least hydrophobic glycopeptide 39–78 (containing 11 hydrophobic residues: Ala, Pro, Val, and Ile), and the more hydrophobic fraction, GIII, was tentatively assigned to glycopeptide 1–38 (containing 15 hydrophobic residues). Proton NMR spectroscopy, at 600 MHz, of each pooled fraction confirmed their identities based on their different amino acid compositions (data not shown). Fraction GTTTR-PSM-GII was unambiguously identified as residues 39–78 on the basis of its amino acid sequence, presented below.
the amino acid-PTH-derivative produces diastereomers with different retention times. The \(\alpha\)-GalNAc-Ser-PTH-derivatives elute as an unresolved doublet, labeled \(S^*\) and \(S^{**}\), early in the gradient near the position of PTH-Asp, and the \(\alpha\)-GalNAc-Thr-PTH diastereomers elute later in the gradient as resolved peaks \(T^*\) and \(T^{**}\), near the positions of PTH-Ser and PTH-Thr, respectively. These peaks are readily identified in the sequencing chromatograms of glycopeptide TTTR-PSM-T3 for residues Ser2, Ser6, and Thr22 as shown in Fig. 4, B–D. On the basis of the relative sizes of the glycosylated and nonglycosylated PTH-Ser and PTH-Thr derivatives in the TTTR-PSM-T3, it appears that Ser6 and Thr22 are more highly glycosylated than Ser2, which seems to be poorly glycosylated.

Sequencing chromatograms were quantified to obtain the residue-specific extent of glycosylation as illustrated in Figs. 5 and 6 for sequencing run 2 of TTTR-PSM-T3 tandem repeat glycopeptide. Fig. 5A displays the uncorrected area data for the Gly-PTH peak plotted as a function of sequence cycle. The figure shows a pronounced base-line curvature that is also observed for the other amino acid residues and glycosylated Ser/Thr (data not shown). Since the extent of base-line curvature correlated with the residue’s percent mole fraction, we conclude that the curvature is due to the cumulative effects of cycle preview and lag and perhaps due to heterogeneous cleavage of the tandem repeat peptide. Since non-zero base lines will interfere with the accurate determination of the extent of glycosylation and with the sequence determination at high cycle numbers, we eliminated the curvature by the base-line subtraction approach described under “Experimental Procedures.” The effectiveness of the base-line correction approach is shown in the corrected data for Gly, Ile, Val, and Ala of Fig. 5, C–F, and for glycosylated and nonglycosylated Ser and Thr of Fig. 6, A–D. Note that after base-line correction the sequence can be read well beyond residue 60 as demonstrated by the expanded plots of Fig. 5, D and F, and Fig. 6, B and D. A plot of the single residue picomoles recovered versus cycle number, Fig. 5B, indicates that we have achieved reasonable sequential residue quantification and recovery. An average apparent repetitive yield of 99% is obtained from the data in Fig. 5B.

Table I lists the calculated extent of Ser/Thr glycosylation obtained from the multiple sequencing of the PSM tryptic tandem repeat (TTTR-PSM-T3) and its C-terminal Glu-C glycopeptide (GTTTR-PSM-GII). Sequence determinations 1 and 2 represent data from the same sample sequenced on different instruments. Note the excellent agreement between the two sequencing experiments. Sequence determination 3 represents
PSM Tandem Repeat Glycosylation Pattern

The abbreviations used are: TTTR-PSM-T3, oligosaccharide trimmed PSM tryptic tandem repeat, residues 1–81, HPLC purified in Fig. 3A; GTTTR-PSM-GII, protease Glu-C-cleaved TTTR-PSM-T3, residues 39–78, purified as shown in Fig. 3B. For further explanation see text.

Table I. Sequence-specific glycosylation patterns of the PSM tandem repeat

| Residue | TTTR-PSM-T3 OG | GTTTR-PSM-GII OG | NetGlyc valuea | h valuec | Δ valuedd |
|---------|----------------|------------------|----------------|---------|-----------|
| S77     | 33             | 2               | 0.29           | 0.11    | 0.19      |
| T70     | 10             | 7               | 0.17           | 0.05    | 0.35      |
| S66     | 86             | 13              | 0.58           | 0.50    | 0.45      |
| T60     | 50             | 46              | 0.69           | 0.75    | 0.43      |
| S64     | 45             | 19              | 0.12           | 0.52    | 0.16      |
| T64     | 88             | 10              | 0.48           | 0.25    | 0.21      |
| S63     | 91             | 45              | 0.75           | 0.17    | 0.08      |
| T62     | 96             | 63              | 0.84           | 0.32    | 0.55      |
| S61     | 44             | 34              | 0.32           | 0.78    | 0.08      |
| T56     | 79             | 80              | 0.88           | 0.83    | 0.94      |
| S59     | 76             | 62              | 0.60           | 0.47    | 0.14      |
| T55     | 91             | 88              | 0.24           | 0.59    | 0.65      |
| S54     | 91             | 91              | 0.85           | 0.95    | 0.60      |
| T50     | 96             | 96              | 0.45           | 0.56    | 0.35      |
| S51     | 92             | 92              | 0.52           | 0.17    | 0.89      |
| T47     | 67             | 62              | 0.27           | 0.45    | 0.02      |
| S48     | 81             | 93              | 0.20           | 0.59    | 0.18      |
| T45     | 80             | 90              | 0.74           | 0.66    | 0.60      |
| S47     | 93             | 89              | 0.34           | 0.38    | 0.08      |
| T40     | 53             | 47              | 0.42           | 0.46    | 0.12      |
| S45     | 92             | 92              | 0.52           | 0.61    | 0.20      |
| T35     | 79             | 83              | 0.34           | 0.59    | 0.51      |
| S39     | 88             | 92              | 0.82           | 0.85    | 0.21      |
| T30     | 77             | 77              | 0.32           | 0.80    | 0.49      |
| S29     | 62             | 48              | 0.79           | 0.87    | 0.50      |
| T26     | 56             | 55              | 0.55           | 0.91    | 0.17      |
| S25     | 71             | 75              | 0.40           | 0.82    | 0.87      |
| T21     | 78             | 77              | 0.56           | 0.48    | 0.28      |
| S20     | 96             | 107             | 0.45           | 0.30    | 0.14      |
| T19     | 86             | 87              | 0.24           | 0.27    | 0.02      |
| S19     | 73             | 73              | 0.26           | 0.27    | 0.61      |

Average 78

The tandem repeat obtained from a different PSM preparation. Again, the results of determination 3 are nearly indistinguishable from the results of sequence determinations 1 and 2. The glycosylation patterns of the C-terminal Glu-C tryptic tandem repeat glycopeptide, determinations 4 and 5, are in good agreement with the data from the full tryptic tandem repeat.

An examination of sequence determinations 1–5, Table I, reveals that Ser2, Ser14, Ser32, Ser47, Ser54, Ser62, and Ser63 are consistently the least glycosylated. As a group, 74% of the Ser residues are glycosylated compared with 83% for the Thr residues, values consistent with the 13C NMR data analysis (data not shown, see Ref. 26). Combined, 78% of the Ser and Thr residues are glycosylated (Table I).

The average sequence-specific glycosylation obtained from determinations 1–5 along with the O-glycosylation predictions of Hansen et al. (10) (NetGlyc activity values), Elhammer et al. (9) (h values) and Chou (11) (Δ values) are tabulated in Table I. For the predictions, plus symbols to the right of the value indicate that the residue is predicted to be glycosylated based on the original published cutoff criteria, and minus symbols indicate the residue would not be glycosylated. As evident from the table and visually when the observed versus predicted glycosylations are plotted together (Fig. 7), the predictions do not completely agree with each other nor do they correlate well with the observed glycosylation.

DISCUSSION

O-Linked Glycoprotein Sequencing—These studies demonstrate that the partial deglycosylation of mucin-type O-linked glycoproteins by mild TFMSA yields glycoprotein derivatives with monosaccharide α-GalNAc side chains that are both susceptible to protease digestion and suitable for standard amino acid sequencing. Using this approach the heterogeneous glycosylated 81-residue PSM tryptic tandem repeat has been isolated and quantitatively sequenced, revealing its glycosylation pattern.

There are numerous reports of the use of automated Edman sequencing for the semiqualitative sequencing of O-linked glycoproteins and for the determination of the in vitro glycosylation patterns of the UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase. Sites of glycosylation have typically been estimated by the presence of “blank” cycles (see for example Refs. 23, 35–36). Other studies of the GalNAc transferase have relied on the incorporation of radiolabeled UDP-GalNAc into substrate peptide and scintillation counting of the released products after Edman sequencing (see Refs. 9, 18–19, 34, 37).

Few workers have attempted to chromatographically characterize or quantitatively analyze the resultant glycosylated Ser and Thr PTH derivatives obtained from standard Edman sequencing. Abernethy and co-workers (34) have described and partially characterized the α-GalNAc-O-Thr-PTH-derivative.
derived from the sequencing of a series of glycopeptide acceptors of the GalNAc transferase. Although they failed to demonstrate its use, these workers had suggested that a TFMSA-Edman sequencing approach would be useful for characterizing α-GalNAc-Thr containing glycopeptides. In the laboratory of Gooley and co-workers (38–41), an Edman sequencing protocol has been developed, using the more hydrophilic solvent trifluoroacetic acid, for the sequence analysis of O-linked glycoproteins containing intact oligosaccharides. Unfortunately, for most glycopeptides with intact oligosaccharide side chains, the presence of heterogeneous oligosaccharide structures complicates the sequence analysis. Thus, quantitation of the extent of glycosylation is still a difficult task. Another drawback of this approach is that the presence of full-length oligosaccharide side chains will interfere with protease digestions making the isolation of reasonably sized glycopeptides with homogeneous peptide sequences difficult. Therefore, for several reasons, the use of mild TFMSA to trim heterogeneous O-linked oligosaccharide side chains to homogeneous peptide-linked GalNAc residues followed by standard Edman sequencing may be the most effective approach for determining the primary glycosylation pattern of heavily O-glycosylated glycoprotein domains.

PSM Tandem Repeat Glycosylation Pattern—The glycosylation pattern obtained for the PSM tandem repeat reveals that all potential glycosylation sites can be glycosylated, although to different extents. Interestingly, the Ser residues display a wider range of observed glycosylations, ranging from about 30% to nearly 100%, whereas the Thr residues show a much narrower range of glycosylation, ranging between 70 and 90%. The extent of Ser O-glycosylation of the PSM tandem repeat is higher than expected based on the in vitro glycosylation studies of the isolated porcine UDP-GalNAc-polypeptide N-acetylgalactosamintransferase (14–15). With this enzyme and other GalNAc transferases isolated to date (see Refs. 9, 17–20, 35, 42–45), Ser residues are typically very poor substrates in in vitro glycosylation studies. In contrast, in vivo glycosylation studies reveal higher extents of O-glycosylation at both Ser and Thr (20). As expected the observed PSM glycosylation pattern is consistent with these observations.

As shown in Fig. 7, none of the three most recent peptide O-glycosylation predictive approaches were capable of successfully predicting the PSM tandem repeat glycosylation pattern. A comparison of the figure suggests that Ekhammer’s and Chou’s (9, 11) predictions may be more useful for mucins as they correctly predicted the largest number of glycosylated residues. Unfortunately, none of the approaches performed well in predicting the poorly glycosylated residues, although all three approaches predicted the least glycosylated residue, Ser.

Having available such a large data base in the PSM tandem repeat (i.e. a total of 31 different glycosylated Ser and Thr residues), an attempt was made to determine whether any specific patterns could be associated with a given sequence’s degree of glycosylation. In Tables II and III we have listed, in order of increasing extent of glycosylation, the heptad peptide sequences for each Ser and Thr residue in the tandem repeat. In addition, due to the large number of Ser/Thr dyad sequences (9 pairs) and the presence of a single Ser triad, we have also listed their peptide sequences in Table IV for comparison.

An examination of the Ser peptide data (Table II) suggests several possible trends. Sequences with Gly at positions +2 or −2 from the potential site of glycosylation appear to be associated with a higher degree of glycosylation; Gly at positions +3 or −3 may be associated with reduced glycosylation, and Gly at positions +1 or −1 apparently are glycosylation neutral as shown by their average glycosylation values at the bottom of Tables II–IV. Sequences containing Pro are usually more heavily glycosylated, with the Pro typically located at the +3 or −3 positions, and sequences with high Ser/Thr contents appear to be more poorly glycosylated. No obvious correlation with the extent of glycosylation beyond the uniform presence of coil and extended secondary structures is found. The presence of large hydrophobic residues, Ile and Val, does not apparently directly correlate with the extent of glycosylation (however, their positions may be important as discussed below). These trends also appear for Thr and the dyad and triad peptide sequences (Tables III and IV), although since their ranges in glycosylation are much smaller the trends are less apparent. It is noteworthy

3 This approach was not further exploited by these workers, apparently because of their belief that the diastereotopic peaks observed for α-GalNAc-O-Thr-PTH represented breakdown products and therefore was not suitable for automated sequencing.

4 Compared with the standard chlorobutane-based solvent systems, trifluoroacetic acid is more efficient at extracting the more polar oligosaccharide Ser/Thr PTH derivatives.

5 The extent of glycosylation of Thr and Ser was determined from a single sequencing run of the full tandem repeat and therefore may contain significant errors. We believe, however, that the reported values are reasonable since in our earlier glycopeptide isolations (24) the major collagenase glycopeptide that is isolated is the fully glycosylated 14-residue C-terminal glycopeptide Gly<sup>40</sup>-Arg<sup>41</sup> which contains glycosylated Thr<sup>70</sup>, Ser<sup>75</sup>, Thr<sup>79</sup>, and Ser<sup>80</sup>.
sequences are identical to those observed for the sequences aligned for sequence; thus, as one would expect, the trends observed for the dyad sylation could be rationalized in terms of specific peptide con-

1 residue at dyad positions 

that 7 of the 9 dyad sequences, Table IV, have at least 1 Gly residue at dyad positions +1 or −1 while 4 of the sequences have Gly residues at both +1 and −1 dyad positions.

To examine the possibility that the observed extent of glycosylation could be rationalized in terms of specific peptide conformation and structure, we built models for the peptide sequences in Tables II–IV. An extended β-conformation was found to best account for the nearest neighbor and penultimate positional effects of Gly. In a β-conformation, nearest neighbor residues will have their side chains directed away from each other on opposite sides of the peptide backbone, whereas penultinate residues (at positions +2 or −2) will have their side

that 7 of the 9 dyad sequences, Table IV, have at least 1 Gly residue at dyad positions +1 or −1 while 4 of the sequences have Gly residues at both +1 and −1 dyad positions.

To examine the possibility that the observed extent of glycosylation could be rationalized in terms of specific peptide conformation and structure, we built models for the peptide sequences in Tables II–IV. An extended β-conformation was found to best account for the nearest neighbor and penultimate positional effects of Gly. In a β-conformation, nearest neighbor residues will have their side chains directed away from each other on opposite sides of the peptide backbone, whereas penultinate residues (at positions +2 or −2) will have their side

Note that in a dyad sequence, positions +1 and −1 also represent positions +2 and −2 for the penultimate Ser/Thr residue in the dyad sequence; thus, as one would expect, the trends observed for the dyad sequences are identical to those observed for the sequences aligned for single Ser and Thr residues in Tables II and III.

6 Note that in a dyad sequence, positions +1 and −1 also represent positions +2 and −2 for the penultimate Ser/Thr residue in the dyad sequence; thus, as one would expect, the trends observed for the dyad sequences are identical to those observed for the sequences aligned for single Ser and Thr residues in Tables II and III.

7 Since deglycosylated mucin polypeptides typically behave as solvent-exposed random-coil structures (6, 26, 30, 44) and because the ensemble of random coil conformations will contain extended β-like structures, the modeling of these short PSM sequences in β-conformations is not inconsistent with their known solution conformations.
chains adjacent to each other on the same side of the peptide backbone. Thus, the observation that penultimate Gly residues may favor glycosylation suggests that penultimate residues with larger side chains may interfere with glycosylation. Furthermore, since the side chains of residues at positions +1 and −1 would not be expected to sterically interfere with the Ser or Thr side chain, the presence or absence of residues containing bulky side chains at these positions would be expected to have little effect on glycosylation. Thus, Ser2, Ser54, and Ser47 which have N- and C-terminal penultimate residues with side chains are poorly glycosylated, whereas Ser43, Ser57, and Ser73 which have a single penultimate neighbor with side chains are more highly glycosylated. Ser73 having 2 penultimate Gly residues might also be expected to be highly glycosylated; however, it is only moderately glycosylated. The reduced glycosylation of Ser73 is consistent with the results of in vitro glycosylation studies that suggest the added flexibility of multiple Gly residues may reduce glycosylation (see Refs. 15 and 45). All of the single residue Thr sequences in Table III are highly O-glycosylated, and the ranking of Thr56, Thr57, and Thr25 would follow the order of decreasing penultimate residue side chain size. Ser43 and Thr39 appear to be exceptions to the “rule” in that they contain 2 penultimate residues with side chains and are nevertheless very highly glycosylated. We suggest that other factors, such as the presence of Pro residues in their sequences, may be responsible for enhancing their glycosylation, as discussed below.

The unexpectedly high glycosylation of Ser43 may be rationalized in terms of the conformational effects of −1 Pro. Model building reveals that the Pro residue preceding Ser43 can alter the peptide backbone conformation so that the −2 side chain would no longer be adjacent to the Ser side chain. The proposed conformational effects of −1 Pro may explain the elevated incidence for observing Pro at this position at known O-glycosylation sites (9–10, 16, 46). Experimentally, the introduction of a −1 Pro has been shown to enhance the in vitro glycosylation of a human von Willebrand factor dodecapeptide (17) and the in vitro glycosylation of recombinant erythropoietin (47).

The glycosylation of Thr39 is apparently enhanced by the presence of a +3 Pro. Increased occurrences of Pro at positions −3 and +3 are consistently observed at known O-glycosylation sites (9–10, 16, 46), with prolines at +3 having the greatest predictive power over any other residue or position (9, 10). This is supported by the work of O’Connell and co-workers (16–17), where the removal of a +3 Pro was found to reduce in vitro glycosylation. Peptides with high Pro content, such as AcTPPP, furthermore, are relatively good substrates for the transferase (12–13, 48). Isolated Pro residues appear to be less prevalent at the −2 and +2 positions in PSM and in the glycosylation surveys of others (9–10, 16, 45), again suggesting the importance of having small side chains at the +2 and −2 positions. For example, the peptide sequence APDTRPA, in the human mucin Muc1 tandem repeat, which contains +2 and −2 Pro residues, cannot be glycosylated in vitro (18–19, 34). Likewise, the introduction of Pro at the −2 position relative to Ser39 in recombinant erythropoietin inhibits its in vivo glycosylation (47).

The Ser/Thr residues in the hydroxyamino acid dyad sequences are typically highly and uniformly glycosylated giving a combined average glycosylation of ~85% (Table IV). From the sequencing data, however, we can only estimate a range of values in which both residues of a dyad would be glycosylated; this value ranges from ~60 to ~95% depending on whether the initial glycosylation is inhibitory or stimulatory toward the second (Table IV). Consistent with the presence or absence of glycosylation enhancing neighbors Thr79, Ser52, and Ser14 are less highly glycosylated than their dyad neighbor. The relatively high extent of Ser dyad glycosylation further suggests that the initial glycosylation of a residue in a dyad may enhance the glycosylation propensity of the second. Similar conclusions were drawn from the in vitro glycosylation studies of Wang et al. (14–15), although the glycopeptide studies of Brockhausen and co-workers (45) suggest the reverse.

The PSM tandem repeat contains a triad sequence of serine residues, Ser94-Ser64 (Table IV). Ser62 which lacks penultimate Gly or glycosylation-enhancing Pro residues is relatively poorly glycosylated, and Ser63 is only moderately glycosylated, perhaps due to its 2 penultimate Gly residues. Interestingly Ser64 is only moderately glycosylated despite having an enhancing +3 Pro.

Several studies show that charged residues appear to be relatively absent from O-glycosylation sites (9–10, 16, 46).
O’Connell and co-workers (17) and Nehrke and co-workers (20) have shown that the in vitro and in vivo transferase activity is highly sensitive to the specific placement of pairs of charged residues. In contrast, Thr and Thr in the PSM tandem repeat are highly glycosylated despite having a pair of charged residues (Arg and Glu) in their sequences (Tables III and IV). The high glycosylation of these sequences may stem from the presence of P or +3 Pro residues or the transferase may tolerate the specific charge distribution of these peptides.

Model of the GalNAc Transferase Peptide Binding Site—From the above steric and structural arguments a rough model of the transferase peptide binding site has been proposed to help explain several features of the transferase and lead to testable hypotheses (Fig. 8). As proposed, substrate peptide would bind in an extended β-like conformation. The UDP-GalNAc binding site would be in the vicinity of the Ser/Thr residue at position P0. This model is similar to that proposed by Elhammer and co-workers (9) who suggested that interactions between the substrate and transferase would occur at both the reactive Ser or Thr residue and through the binding of an extended β-like substrate peptide. These workers, however, found no sequence specific side chain specificity, beyond the presence of +3 Pro for residues surrounding the acceptor Ser or Thr. Peptides with extended β- or β-turn conformations have also been suggested by others to be favorable substrate conformations for the transferase (10, 16, 46, 49).

In the model, side chains at penultimate positions relative to the central Ser or Thr are proposed to sterically interfere in some manner with glycosylation. Since all of the PSM tandem repeat glycosylation site are at least partially glycosylated in vivo, the interactions of the penultimate side chains must not completely inhibit the transfer of α-GalNAc, but rather reduce the glycosylation efficiency by affecting substrate binding processes and/or GalNAc transfer. Pro (and perhaps the presence of specific patterns of charged residues) appears to further modulate transferase activity, with the prevalence of +3 Pro suggesting the possibility of a specific transferase binding site for Pro at this position. These latter effects would most likely operate at the level of substrate binding rather than at the GalNAc transfer step.

Because of the alternating nature of β-structured peptides, substrate would be expected to bind the active site in two possible orientations, each differing by the direction in which its side chains point. The ability to bind opposite sides of the peptide would readily allow for the glycosylation of dyads. To permit the glycosylation of the second residue and to maintain flexibility regarding the order of glycosylation, both the P and P+1 sites on the transferase (see Fig. 8) would be expected to be large enough to accommodate a peptide-linked α-GalNAc residue. To further favor dyad glycosylation we suggest that these sites may weakly bind peptide-linked α-GalNAc residues.

Our model of the GalNAc transferase binding site is consistent with the kinetic analysis of Wragg and co-workers (48) that supports a 2-site model for the random, non-competitive binding of peptide substrate and UDP-GalNAc. These workers suggest that peptide substrate selection may be a 2-step process, involving the initial low specificity binding of peptide followed by a second process involving the specific interaction of the hydroxyamino acid with the transferase active site. Enhancements in the binding or lifetimes at either of these steps could result in apparent increases in transferase activity. We propose that this second binding/transfer step may explain the differences in Thr and Ser acceptor activities and the higher sensitivity of Ser to peptide sequence. Thr residues, having a methyl group, may possess enhanced binding and/or optimal substrate geometry for GalNAc transfer compared with Ser residues that lack the methyl group. Therefore, for peptides with similar primary binding affinities, Thr acceptor peptides would be expected to be more efficiently glycosylated than Ser peptides.8

8 It is worth noting that for N-glycosylation the Asn-Xaa-Thr sequon is more efficiently glycosylated compared with the Asn-Xaa-Ser sequon (50) perhaps for the same reasons presented here.
Acknowledgments—We thank Drs. K. L. Gerken and W. C. Merrick for reading the manuscript and for their helpful suggestions. Mathew Gombrich is also acknowledged for his contributions.

REFERENCES

1. Varki, A. (1993) Glycobiology 3, 97–130
2. Shimizu Y., and Shaw, S. (1993) Nature 366, 639–631
3. Springer, T. (1994) Cell 76, 391–314
4. Hilken, J. (1988) Cancer Rev. 11–12, 25–54
5. Difilipo, N., Yamakami, K., Washington, S., and Bhavanandan, V. (1994) Glycosylation & Disease 1, 21–30
6. Shogren, R. L., Gerken, T. A., and Jentoft, N. (1989) Biochemistry 28, 5525–5536
7. Gerken, T. A., and Dearborn, D. G. (1984) Biochemistry 23, 1485–1497
8. Gerken, T. A., and Jentoft, N. (1987) Biochemistry 26, 4689–4699
9. Elhammer, A. P., Poorman, R. A., Brown, E., Maggiora, L. L., Hoogerheide, J. G., and Kezdy, F. J. (1993) J. Biol. Chem. 268, 10028–10038
10. Hansen, J. E., Lund, O., Engelbrecht, J., Bohr, H., Nielsen, J. O., Hansen, J. E. S., and Brunak, S. (1995) Biochem. J. 308, 801–813
11. Chou, K. C. (1995) Protein Sci. 4, 1365–1383
12. Young, J. D., Tsuchiya, D., Sandlin, D. E., and Holroyde, M. J. (1979) Biochemistry 18, 4444–4448
13. Briand, J. P., Andrews, S. P., Jr., Cahill, E., Conway, N. A., and Young, J. D. (1981) J. Biol. Chem. 256, 12205–12207
14. Wang, Y., Abernethy, J. L., Eckhardt, A. E., and Hill, R. L. (1992) J. Biol. Chem. 267, 12709–12716
15. Wang, Y., Agrwal, N., Eckhardt, A. E., Stevens, R. D., and Hill, R. L. (1993) J. Biol. Chem. 268, 22979–22983
16. O’Connell, B. C., Tabak, L. A., and Ramasubbu, N. (1991) Biochem. Biophys. Res. Commun. 180, 1024–1030
17. O’Connell, B. C., Hagen, F. K., and Tabak, L. A. (1992) J. Biol. Chem. 267, 25010–25018
18. Nishimori, I., Johnson, N. R., Sanderson, S. D., Perini, F., Mountjoy, K., Cerny, R. L., Gross, M. L., and Hollingsworth, M. A. (1994a) J. Biol. Chem. 269, 16123–16130
19. Nishimori, I., Perini, F., Mountjoy, K. P., Sanderson, S. D., Johnson, N., Cerny, R. L., Gross, M. L., Fontenot, D., and Hollingsworth, M. A. (1994c) Cancer Res. 54, 3738–3744
20. Neves, K., Hagen, F. K., and Tabak, L. A. (1994) J. Biol. Chem. 271, 7061–7065
21. Clausen, H., and Bennett, E. (1996) Glycobiology 6, 635–646
22. Marth, J. (1996) Glycobiology 6, 701–705
23. Sorensen, T., White, T., Wandal, H. H., Kristensen, A. K., Roepstorff, P., and Clausen, H. (1995) J. Biol. Chem. 270, 24166–24173
24. Gerken, T. A., Owens, C. L., and Pasunarrarry, M. (1997) in Techniques in Glycobiology (Townsend, R. R., ed) pp. 247–270, Marcel Dekker, Inc., New York
25. Shogren, R. L., Jamieson, A. M., Blackwell, J., and Jentoft, N. (1986) Biochemistry 25, 1505–1517
26. Gerken, T. A., Gupta, R., and Jentoft, N. (1992) Biochemistry 31, 639–648
27. Edge, A. S. B., Faltynek, C. R., Hof, L., Reichert, L. E., and Weber, P. (1981) Anal. Biochem. 118, 131–137
28. Mantle, M., and Allen, A. (1978) Biochem. Soc. Trans. 6, 607–609
29. Georjun, C., and Deleage, G. (1994) Protein Eng. 7, 157–164
30. Timpte, C. S., Eckhardt, A. E., Abernethy, J. L., and Hill, R. L. (1988) J. Biol. Chem. 263, 1081–1088
31. Eckhardt, A. E., Timpte, C. S., Abernethy, J. L., Zhao, Y., and Hill, R. L. (1991) J. Biol. Chem. 266, 9678–9686
32. Gupta, R., and Jentoft, N. (1989) Biochemistry 28, 6114–6121
33. Perez-Vilar, J., Eckhardt, A. E., and Hill, R. L. (1996) J. Biol. Chem. 271, 9845–9850
34. Abernethy, J. L., Wang, Y., Eckhardt, A. E., and Hill, R. L. (1992) in Techniques in Protein Chemistry III (Angeletti, R. H., ed) pp. 277–286, Academic Press, New York
35. Stadie, T. R. E., Chai, W., Lawson, A. M., Byfield, P. G. H., and Hanisch, F. G. (1995) Eur. J. Biochem. 229, 140–147
36. Rohrer, J. S., Cooper, G. A., and Townsend, R. R. (1993) Anal. Biochem. 212, 7–16
37. Lorenz, C., Strahl-Bolsinger, S., and Ernst, J. F. (1992) Eur. J. Biochem. 205, 1163–1167
38. Gooley, A. A., Clasen, B. J., Marschalek, R., and Williams, K. L. (1991) Biochem. Biophys. Res. Commun. 178, 1194–1201
39. Pisano, A., Redmond, J. W., Williams, K. L., and Gooley, A. A. (1993) Glycobiology 3, 429–435
40. Pisano, A., Packer, N. H., Redmond, J. W., Williams, K. L., and Gooley, A. A. (1994) Glycobiology 4, 837–844
41. Gooley, A. A., and Williams, K. L. (1994) Glycobiology 4, 413–417
42. Hagen, F. K., VanWuyckhuyse, B., and Tabak, L. A. (1993) J. Biol. Chem. 268, 12960–12965
43. O’Connell, B. C., and Tabak, L. A. (1993) J. Dent. Res. 72, 1554–1558
44. Gerken, T. A., Butenhof, K., and Shogren, R. (1989) Biochemistry 28, 6538–6543
45. Brockhausen, I., Toki, D., Brockhausen, J., Peters, S., Bielfelt, T., Kleen, A., Paulsen, H., Meldal, M., Hagen, F., and Tabak, L. A. (1996) Glycocon. J. 13, 649–656
46. Wilson, I. B. H., Gavel, V., and von Heijne, G. (1991) J. Biochem. (Tokyo) 275, 529–534
47. Elliott, S., Bartley, T., Delorme, E., Derby, P., Hunt, R., Lorenzini, T., Parker, V., Rohde, M., and Stoney, K. (1994) Biochemistry 33, 11237–11245
48. Wragg, S., Hagen, F. K., and Tabak, L. A. (1995) J. Biol. Chem. 270, 16947–16954
49. Aubert, J., Biserte, G., and Loucheux-Lefebvre, M. (1976) Arch. Biochem. Biophys. 175, 410–418
50. Kasturi, L., Eshleman, J., Wunner, W., and Shakin-Eshleman, S. (1995) J. Biol. Chem. 270, 14756–14761

PSM Tandem Repeat Glycosylation Pattern

9719
Determination of the Site-specific O-Glycosylation Pattern of the Porcine Submaxillary Mucin Tandem Repeat Glycopeptide: MODEL PROPOSED FOR THE POLYPEPTIDE: GalNAc TRANSFERASE PEPTIDE BINDING SITE

Thomas A. Gerken, Cheryl L. Owens and Murali Pasumarthy

J. Biol. Chem. 1997, 272:9709-9719.
doi: 10.1074/jbc.272.15.9709

Access the most updated version of this article at http://www.jbc.org/content/272/15/9709

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 48 references, 16 of which can be accessed free at http://www.jbc.org/content/272/15/9709.full.html#ref-list-1