Determination of the site-specific oligosaccharide distribution of the O-glycans attached to the porcine submaxillary mucin tandem repeat:

Further evidence for the modulation of O-glycan side chain structures by peptide sequence.

*Thomas A. Gerken¹, Marc Gilmore & Jiexin Zhang
Departments of Pediatrics and Biochemistry
W.A. Bernbaum Center for Cystic Fibrosis Research
School of Medicine
Case Western Reserve University
Cleveland OH 44106

*This work was supported by grants from the National Cancer Institute, National Institutes of Health, RO1-CA-78834 and from the Cystic Fibrosis Foundation.

+To whom correspondence should be addressed:
Case Western Reserve University
Department of Pediatrics, School of Medicine, BRB,
2109 Adelbert Rd
Cleveland OH 44106-4948.
Telephone 216-368-4556, fax 216-368-4223
e-mail: txg2@po.cwru.edu
Running title: Modulation of Mucin Side Chain Structure
Abstract

Little is known of the degree that polypeptide sequence and local environment modulate the structures of O-linked glycans. Towards this understanding, the site-specific mono-(GalNAc-O-), di- (β-Gal (1-3) α-GalNAc-O-) and tri- (α-Fuc(1-2) β-Gal (1-3) α−GalNAc-O-) saccharide distribution has been determined for 29 of the 31 O-glycosylated Ser/Thr residues in the tandem repeat domains of blood group A-negative porcine submaxillary gland mucin. The glycosylation patterns obtained from 3 individual animals are in agreement with earlier incomplete determinations on a pooled mucin (J. Biol. Chem. 272, 9709 (1997) & J. Biol Chem. 273, 26580 (1998)), confirming that the addition of the peptide linked GalNAc and its substitution by β(1-3) Gal are sensitive to local peptide sequence in a highly reproducible manner in vivo. The present data further support earlier suggestions of an inverse correlation of the density of hydroxyaminoacid residues (and by inference the density of peptide GalNAc) with the extent of substitution of the peptide linked GalNAc by β (1-3) Gal. This effect is highly correlated for Ser linked glycans but not for Thr linked glycans. A similar correlation is observed with respect to the in vivo peptide GalNAc glycosylation pattern. In contrast, the addition of α-1-2 Fuc to β-Gal shows no apparent correlation with hydroxyaminoacid density, although a marked elevation in the fucosylation of Ser linked glycans compared to Thr linked glycans is observed. The above effects may represent both steric and conformational factors acting to alter the relative accessability and activity of the glycosyltransferases towards substrate. These results demonstrate that the porcine submaxillary gland core-1 β3-galactosyltransferase and α2-fucosyltransferase exhibit unique peptide/glycopeptide sensitivities which may provide mechanisms for the modulation of O-linked side chain structures.
Introduction

The O-glycosylation of serine and threonine residues with glycans linked through N-acetylgalactosamine (GalNAc) is a common post-translational modification of a wide range of secreted and membrane associated proteins. One class of highly O-glycosylated proteins are the mucus glycoproteins commonly called mucins(1). These glycoproteins are typically 20-30% Ser and Thr, are between 50 and 80% carbohydrate by weight and commonly contain tandemly repeated peptide sequences. Mucins and glycoproteins containing mucin-like domains play a major role protecting epithelial cell surfaces and are thought to be involved in modulating many biological processes including the immune response, adhesion, inflammation, tumorigenesis and perhaps development (2-9).

The first step in O-glycan synthesis is the transfer of GalNAc to Ser/Thr residues, by UDP-GalNAc:polypeptide α-GalNAc transferase (ppGalNAc transferase). To date, ten ppGalNAc transferase isoforms have been described (9-16). Although each transferase has not been fully characterized, their peptide substrate specificities vary within the family(11, 17-21); many show sensitivity to prior glycosylation and others require prior addition of GalNAc for activity(9, 14, 20, 22-24). The expression of ppGalNAc transferase isoforms with different peptide and/or glycopeptide specificities therefore represents the first step in the regulation of O-glycan structure by peptide sequence in vivo. Subsequent elongation of O-linked glycans proceeds by the step-wise addition of single sugar residues via a series of substrate specific Golgi resident transferases (25). It is well accepted that Golgi localization, nucleotide sugar concentration and competition among transferases are major dictates of O-glycan structure and elongation (25-31). Depending on the initial and subsequent substitutions on the GalNAc residue, a wide range of O-linked core structures is possible (25). Little, however, is known of the effects of local peptide sequence on O-glycan elongation (23, 32-37), although a number of examples of site-specific elongation of N-linked glycans are known (30, 38-46). Factors affecting site-specific N-glycan structures include specific peptide sequence motifs, protein surface structures and the folded state of the protein. The paucity of our understanding of the influence of the acceptor peptide structure on O-glycan elongation in mucin-like domains arises from the difficulty in determining the site-specific glycosylation pattern of mucins and other heavily O-glycosylated glycoproteins.
With the goal of understanding the influence of peptide sequence and structure on mucin O-glycosylation, our laboratory has previously reported the site-specific GalNAc-O-Ser/Thr and core-1 ([R1-,R2-]β-Gal (1-3) α-GalNAc-O-Ser/Thr) glycosylation patterns of 29 of the 31 glycosylated residues in the 81 residue tandem repeat of the porcine submaxillary mucin (PSM) pooled from several different animals (34, 37) (see Figure 1 for the glycan structures and tandem repeat sequence of PSM(47-52)). We found that the extent of GalNAc core glycosylation was site-specific and that the degree of core-1 glycosylation appeared to inversely correlate to the density of Ser and Thr residues in the polypeptide sequence. From the latter finding, we suggested that the oligosaccharide side chain length may also correlate with the density of glycosylated Ser and Thr residues. These results suggested the reasonable possibility of steric interactions of neighboring glycosylated Ser and Thr residues affecting glycan elongation.

In this manuscript we have extended these studies to the characterization of the site-specific glycosylation pattern of the PSM tandem repeat isolated from a group of A blood group-negative animals, obtaining the mono- (GalNAc-O-Ser/Thr) di- (β-Gal (1-3) α-GalNAc-O-Ser/Thr) and tri- (α-Fuc (1-2) β-Gal(1-3) α-GalNAc-O-Ser/Thr) saccharide distributions at each individual glycosylation site. The analysis of the glycan distributions from PSM obtained from 3 different animals with similar oligosaccharide compositions gave highly reproducible site-specific glycosylation patterns with GalNAc and core-1 glycosylation patterns nearly identical to those previously described for pooled mucin(34, 37). The analysis of the complete glycosylation pattern of these mucins provides additional support for the notion that oligosaccharide side chain structures and side chain lengths may be modulated to a significant extent by the density of neighboring (presumably partially glycosylated) hydroxyaminoacid residues in the polypeptide sequence and that this effect is primarily observed on Ser-linked glycans at the initial step of β-Gal addition by the core-1 β3-galactosyltransferase, forming the core-1 structures. In contrast, our results indicate that the porcine α2- fucosyltransferase in these animals prefers to add Fuc to Ser linked glycans compared to Thr linked glycans by a factor of ~1.6, while exhibiting no correlation to hydroxyaminoacid density for either linkage residue. These results unambiguously show that the elongation of O-linked mucin type glycans is significantly and reproducibly affected, whether directly or
indirectly, by local peptide sequence. This work represents the most complete characterization to date of the *in vivo* O-linked glycosylation pattern of any highly O-glycosylated glycoprotein.

**Methods**

*Mucin isolation and characterization.* Paired frozen porcine submaxillary glands from individual animals were obtained from Pel-Freez (Rogers, AK) and screened for the presence of A blood group determinants by the ability of gland extracts (47) to inhibit the agglutination of human A-positive erythrocytes with *Phaseolus* *Limensis* lectin (Sigma Chemical Co., St. Louis, MO). Mucin used in this work was purified from paired blood group A-negative glands. Oligomeric PSM tandem repeat glycosylated domains, called TR-PSM, were obtained after trypsinization and gel filtration chromatography of the reduced and carboxymethylated mucin as previously described (37). Tandem repeat domains were further purified from contaminating globular proteins by passage through SP-Sephadex (Pharmacia) equilibrated with 50 mM formic acid pH 4 followed by passage through Octyl Sepharose (Pharmacia) equilibrated with 1 M (NH₄)₂SO₄, 100 mM NaH₂PO₄ pH 5. The oligosaccharide composition of the purified mucin was obtained by integration of the anomeric carbon resonances of the carbon-13 NMR spectrum (50), see below.

*Mucin glycan modifications.* Mild, 0°C, trifluoromethanesulfonic acid (TFMSA)/anisole treatments (37, 53) were used to trim O-linked glycans from intact or modified TR-PSM domains, leaving intact peptide linked GalNAc residues. This treatment on intact TR-PSM gives TTR-PSM containing GalNAc residues at all positions originally containing glycans, irrespective of glycan length or structure. We shall refer to glycopeptides derived by this approach as “Native-GalNAc”-PSM tandem repeats.

Selective removal of C3 unsubstituted GalNAc residues on TR-PSM by periodate oxidation (100 mM IO₄⁻, 166 mM NaCl, 100 mM NaHOAc, pH 4.5, 0°C) followed by NaBH₃ reduction (~0.67 M, in 0.5 M Na₂HPO₄, pH 8.0, 0°C) and subsequent mild 0°C TFMSA/anisole treatment was performed as previously described (34). This procedure gives TROTR-PSM, which contains GalNAc residues only at those sites originally containing core-1 glycans (i.e. di and tri saccharides in A-negative blood type mucin). These derivatives will be referred to as “Core-1” PSM tandem repeats. Several changes were made to further optimize these procedures by: 1) using a 10-20 fold molar excess periodate over
oxidizable carbohydrate and an oxidation time of 5-6 hr, 2) destroying periodate by NaI/NaHCO$_3$/NaS$_2$O$_3$ (200, 200 & 800mM) (34), 3) using a maximum NaBH$_3$ reduction time of 15 minutes with a pH maximum of ~8.5, and 4) treating with 0°C TFMSA/anisole for no more than 4 hr. Under these conditions the degradation of the peptide core was kept to a minimum while completely removing oxidized GalNAc. These preparations were further purified by gel filtration chromatography on Sephacryl S200 as previously described (34, 37).

The trisaccharide site-specific glycosylation pattern was determined by the enzymatic conversion of the disaccharide side chains to monosaccharide side chains followed by periodate oxidation and mild TFMSA. Unsubstituted β-(1-3) Gal residues of the disaccharide were selectively removed from 20-30mg of TR-PSM in 2-4 ml of buffer by placing in dialysis tubing with 5-8 U of bovine testes β-galactosidase (Sigma, or Glyco Inc, Novato CA) and dialyzing against 250mL 100 mM sodium citrate pH 4.5, overnight at 33°C. A prior treatment with 1 U of Clostridium perfringens neuraminidase (Sigma) dialyzed against 250mL 100 mM sodium acetate pH 5.0, for 4-8 hr at 33°C was performed to remove sialic acid which might interfere with the removal of β-Gal. Glycosidase treatments were repeated until complete removal of β-Gal and sialic acid was demonstrated by the complete loss of the β-Gal and sialic acid anomer carbon-13 NMR resonances at 105 and 101 ppm respectively(50). These preparations after oxidation/reduction and TFMSA treatments (as described above) yield mucin containing GalNAc residues only at sites that originally contained fucosylated trisaccharide. These TROGNTR-PSM preparations (designated “Trisaccharide” PSM tandem repeat domains), were further purified by Sephacryl S200 chromatography (data not shown) giving a major high molecular weight excluded volume peak identical to those previously reported (34, 37).

To confirm the specificity of the chemical periodate oxidation/reduction approach for removing unsubstituted GalNAc residues, an enzymatic approach was also performed using the α-N-acetylgalactosaminidase from chicken liver (Sigma). Briefly 5-10 mg of β-galactosidase- neuraminidase treated mucin, was incubated with 5-10 U of α-N-acetylgalactosaminidase and dialyzed against ~250 mL of 100mM sodium citrate/phosphate buffer overnight at 42°C. Glycosidase treatments were repeated until the complete removal of unsubstituted GalNAc was confirmed by carbon-13 NMR, by monitoring
the loss of the unique monosaccharide GalNAc C5, C4 and C3 carbon resonances at 72.57, 69.92 and 69.03 ppm (50).

Fully deglycosylated apo-PSM was obtained from “Native GalNAc” PSM oligomeric tandem repeats (containing only peptide linked GalNAc residues) as described earlier by oxidation and alkaline elimination (53).

**PSM tandem repeat isolation.** The 81 residue tandem repeat glycopeptides were obtained from the above “Native GalNAc”, “Core-1” and “Trisaccharide” PSM preparations after trypsinolysis with 1% modified sequence grade trypsin (TRSEQZ, Worthington Biochemicals) in 50mM (NH₄)₂CO₃ pH 8.5 for 7 hr at 37°C (37). Trypsin was inhibited by the addition of 1 mM phenylmethylsulfonylfluoride. Tandem repeat glycopeptides were pooled after chromatography on Sephacryl S200. The GluC - C-terminal portion of the tandem repeat glycopeptide (residues 39-79) was obtained after N-terminal biotinylation and digested with protease Glu C as described (34). The large C-terminal glycopeptide beginning at residue 39 of the tandem repeat was isolated from the N-terminal glycopeptide after passage through a 1 ml immobilized Avidin column (Pierce) and subsequent Sephacryl S200 gel filtration.

**Amino acid sequencing.** Pulsed liquid phase Edman degradation amino acid sequencing was performed on an Applied Biosystems Procise 494 protein sequencer (Perkin-Elmer, Applied Biosystems Div., Foster City CA) using manufacturer recommended pulse liquid cycles as described previously (34, 37). Amino acid phenylthiohydantoin (PTH) derivatives were chromatographed on an ABI 5µm C18 PTH column using the fast normal I gradient program. The PTH-Ser/Thr-O-GalNAc elute as two diastereotopic peaks in the chromatogram at relatively unique positions (37). Pico-moles of PTH derivatives were obtained after eliminating long range cycle preview and lag for each PTH derivative by a simple base line subtraction approach as described (37). Corrections for the overlap of the second eluting PTH-Thr-O-GalNAc diastereomer with PTH-Thr were made using the previously obtained area ratios for the two PTH-Thr-O-GalNAc diastereomers obtained from fully glycosylated glycopeptides (34, 37). The extent of glycosylation was determined by comparing the relative picomoles of nonglycosylated and glycosylated PTH derivatives after taking into account the relative recovery of the individual PTH species. The recovery values used in this work were adjusted to give constant summed Ser and Thr
amino acid compositions (obtained by summing the non-baseline corrected picomole content over all cycles of the sequence determination, typically greater than 40 cycles) regardless of the extent of GalNAc glycosylation of the tandem repeat. With these values, all preparations, from apo- to highly glycosylated “Native-GalNAc” PSM, gave identical amino acid compositions regardless of their extent of glycosylation, with standard deviations of 1 mole percent or less for all residues including Ser and Thr (utilizing all 17 full length tandem repeat determinations performed in this study as well as 3 apo tandem repeat determinations). Using the N-terminal sequence of the tandem repeat, the summed Edman derived amino acid composition was found to be consistent with a 97.5% repetitive yield. The same set of response factors were utilized for the processing of all Edman sequencing data described in this work. Typically 2-8 nmoles of glycopeptide were sequenced to ensure sequencing to at least 40 cycles. The reproducibility between sequencing runs is relatively good, with standard deviations typically in the range of less than 5 percentage values.

**NMR analysis** Carbon-13 NMR spectra were obtained using Varian Inova 300 or Inova 600 spectrometers using D$_2$O as solvent. Integrals of the oligosaccharide anomic carbons unique to each oligosaccharide side chain were used to obtain the oligosaccharide side chain distribution(50). The absence of the tetrasaccharide was indicated by the complete absence of the terminal GalNAc anomic carbon resonance at 92 ppm. The integrals of the GalNAc C1, C5 and N-acetyl carbon resonances and the $\alpha$-carbon resonances of Ser, Thr and Gly after the TFMSA treatments were used to approximate the relative extent of glycosylation (50, 53). Comparisons of the integrals between fully relaxed NOE suppressed spectra (5.27 sec recycle time) and partially relaxed spectra with full NOE typically revealed no systematic changes in resonance areas utilized for the analysis.

Proton NMR spectra were obtained at 600 MHz in 99.99 % D$_2$O. Spectra were obtained at 45°C in order to eliminate overlap of the HDO resonance with the GalNAc anomic protons (54). HDO resonance precaturation was not typically utilized to avoid alterations in the anomic proton resonance areas. The areas of the GalNAc anomic proton resonances at 4.85 ppm and the GalNAc H2 protons which show sensitivity to Ser and Thr linkages (4.13 and 4.06 ppm, respectively, (54)) were used to determine the extent of glycosylation. Areas were normalized to the methyl resonances of Ile and Val at
0.9 ppm and the extent of glycosylation was obtained by utilizing the expected amino acid composition of the 81 residue PSM tandem repeat.

**Amino acid analysis.** Amino acid analyses were performed by the Protein/Peptide Core Facility of the Massachusetts General Hospital on a Beckman 6300 amino acid analyzer. To determine the extent of destruction of Ser and Thr, hydrolysis times in 6N HCl at 110°C of 6, 12, 18 and 24 hr were obtained. No significant losses of Ser or Thr were detected as a function of hydrolysis time. Nevertheless, the 18 and 24 hr time points were used to extrapolate a 0 time nanomole content for each amino acid residue. Quantitation of O-glycosylation was estimated by amino acid analysis of β-eliminated (0.1M NaOH, 0.6 mM NaBH₄, 7hr at 45°C) and reduced (0.66 M NaBH₄ + 0.016 M PdCl₂ in 0.8M HCl) samples as described by Downs et al (55). Extent of glycosylation was calculated from the loss of (glycosylated) Ser & Thr relative to the non reactive amino acid residues and by the gain in Ala in the case of Ser.

**Sequence 'density' determinations and data analysis.** Sequence weighted average Ser/Thr density values were obtained by performing a 7 residue center weighted running average along the tandem repeat, using arbitrary weights of 0.5, 0.75, 1.0, 1.0, 0.75, 0.5 at each position and a denominator of 5.5. Ser and Thr residues were valued as 1 and all other residues valued as 0 (34)Footnote 2. The coefficients were arbitrarily chosen to more heavily weight the central residues and to provide smoothing to the function. In vivo GalNAc density (OG) values were obtained using an identical weighting scheme while using the actual in vivo site-specific Ser or Thr percent glycosylation values. Statistical analyses were performed using Pearson product moment correlation procedure in the Sigma Stat statistical software package (version 2.0) (SPSS Inc., Chicago IL). Correlations were deemed significant for p values less than 0.05.

**Results**

**PSM oligosaccharide characterization.** The possible O-linked glycan side chain structures previously described for PSM, Figure 1, range from the monosaccharide Tn antigen, GalNAc-O-Ser/Thr, to the tetrasaccharide, blood group A structure, α-GalNAc(1-3) [α-Fuc(1-2)]-β-Gal(1-3) α-GalNAc-O-Ser/Thr (47, 51, 52). Each mono through tetra saccharide structure may also have a sialic acid (in the form of N-glycoylneuraminic acid, NeuNGl) 2-6 linked to the peptide GalNAc residue, thus up to 8
different oligosaccharide structures are possible. Earlier carbon-13 NMR studies from our laboratory (50) have revealed that the relative proportion of these glycans varies among pigs, ranging from animals having over 50% the A blood group tetrasaccharide to animals having nearly equal amounts of the di- and tri-saccharide and no tetrasaccharide. In the present study we chose to determine the site-specific glycosylation pattern of the latter class of animals having nearly equal amounts of the mono, di and tri saccharide and no tetrasaccharide (see Table I). It was anticipated that from the analysis of these mucins we would be able to determine the extent that the activity of the \( \alpha_2 \)-fucosyltransferase may be modulated by local peptide sequence.

**Selective Deglycosylation of PSM.** The determination of a glycoprotein’s site-specific glycosylation pattern relies on the ability of mild TFMSA/anisole treatment to trim O-linked glycans to the peptide linked GalNAc residue, the ability of periodate to selectively oxidize peptide-linked GalNAc residues which contain no C3 substitutents and the ability to distinguish GalNAc-O-Ser/Thr residues by automated Edman amino acid sequencing(37, 53). By using a combination of periodate oxidation followed by mild TFMSA/anisole treatment, giving the “Core-1” PSM tandem repeat, those sites containing core-1 (or longer) side chains (i.e those attached to GalNAc via C3) can be identified and quantified by Edman sequencing(34). This approach can be extended to the characterization of longer core-1 glycan side chains by sequential glycosidase treatments prior to the periodate cleavage step. To reveal the positions of the PSM fucosylated trisaccharide (\( \alpha \)-Fuc(1-2)\( \beta \)-Gal(1-3)\( \alpha \)-GalNAc-O-Ser/Thr), bovine testis \( \beta \)-galactosidase was used to remove the \( \beta \)-Gal residues from the disaccharide (\( \beta \)-Gal(1-3)\( \alpha \)-GalNAc-O-Ser/Thr), leaving mucin containing only monosaccharide and trisaccharide glycans. Subsequent periodate oxidation and mild TFMSA/anisole treatment removed the monosaccharide GalNAc residues while trimming the trisaccharide to GalNAc. The determination and quantification of the sites of trisaccharide attachment was obtained by the Edman sequencing of the so derived “Trisaccharide” PSM tandem repeat obtained after trypsinolysis. Thus, by sequencing a series of sequentially deglycosylated “Native-GalNAc”, “Core-1” & “Trisaccharide” PSM tandem repeat preparations, the site-specific mono, di and tri saccharide distributions can be obtained.

Figure 2 shows the carbon-13 NMR spectra for several of the deglycosylated preparations PSM B10
isolated from a single animals. As discussed in the figure legend, the spectra clearly demonstrate the
selective removal of carbohydrate by neuraminidase & β-galactosidase, Figure 2B, and the increasing
removal of GalNAc between the “Native GalNAc”, “Core 1” and “Trisaccharide” PSM tryptic tandem
repeats, Figures 2D-F. Also included in Figure 2 are the NMR spectra of fully deglycosylated apo
mucin, Figure 2G, and of a preparation of β-galactosidase PSM B10 that has had its monosaccharide
GalNAc residues removed by α-N-acetylgalactosaminidase, Figure 2C. As discussed in the figure
legend and below, the latter spectrum shows that α-N-acetylgalactosaminidase removes nearly the same
amount of unsubstituted GalNAc as does the periodate oxidation-reduction-TFMSA procedure, Figure
2F.

The approximate extent of glycosylation determined from the NMR resonance areas of selected
GalNAc and peptide carbons and protons (data not shown) is given in Table II. (Note that the average
extent of glycosylation from the NMR data may not fully agree with the individual Ser and Thr
glycosylations given since different sets of resonances were used for each determination; see Table II
legend). For the most part the NMR derived values agree reasonably well with the expected extent of
glycosylation, although as discussed below the net “Trisaccharide” glycosylation is somewhat higher than
expected. Both the proton and carbon-13 NMR analysis of the “Trisaccharide” derivatives indicate that
Ser residues are more highly glycosylated than Thr residues (Table II). This is readily apparent by the
relative areas of the glycosylated (SOG & TOG) and nonglycosylated (SOH & TOH) α-carbon
resonances of Thr and Ser in the 54-60 ppm range in spectrum F of Figure 2 and by the relative areas of
the GalNAc H2 proton resonances of similar preparations (data not shown). Nearly identical differential
glycosylation of Ser over Thr is also observed in the “Trisaccharide”-like preparation where α-N-
acetylgalactosaminidase was used to remove monosaccharide GalNAc residues rather than periodate and
TFMSA, Figure 2C, and Table II. Therefore, the differences in Ser and Thr glycosylation are not likely
artifacts of the chemical deglycosylation approach.

Amino acid sequencing and glycosylation pattern determination. Edman amino acid sequencing
was performed on the differentially deglycosylated TFMSA treated, 81 residue “Native GalNAc”, “Core
1” and “Trisaccharide” tryptic tandem repeat glycopeptides obtained from each of the three PSM
preparations described above. The 40 residue C-terminal glycopeptides resulting from Glu-C digestion were also sequenced to complete the analysis of the C-terminal portion of the tandem repeat. Representative Ser and Thr sequence profiles are given in Figure 3 and the average residue-specific glycosylation for each PSM preparation (and the average of all three) are summarized in Table III. These data are further summarized in Table II, and are in good agreement with the proton and carbon-13 NMR derived glycosylation values. As with the NMR analysis, the Edman sequence analysis reveals a higher extent of glycosylation of Ser residues relative to Thr in the “Trisaccharide” PSM derivatives. These differences in Ser/Thr glycosylation are also detected when the extent of Ser and Thr O-glycosylation is estimated by the amino acid analysis of the β-eliminated and reduced samples, Table II. Therefore, with three different PSM preparations using two different NMR approaches (13C and 1H NMR), two different analytical chemical approaches (Edman sequencing and amino acid analysis after β-elimination and reduction) and two different methods for the removal of monosaccharide GalNAc residues (periodate oxidation and α-N-acetylgalactosaminidase) it is consistently found that Ser residues are more highly glycosylated by the trisaccharide than Thr residues. Combining all the analytical data in Table II, the average trisaccharide enhancement of Ser residues over Thr residues is approximately 1.6 fold.

Edman sequencing derived values for the oligosaccharide distribution and percent GalNAc glycosylation for each of the three different PSM preparation are given in parentheses in Table I for comparison to the expected values. The Edman sequencing derived monosaccharide and summed di+ tri saccharide percentages and the “Core 1” percent GalNAc glycosylation values (columns 1, 4 & 5 respectively) are in excellent agreement with expected, even maintaining the relative differences between each preparation. On the other hand, values for the percent glycosylation of the “Trisaccharide” (far right column) are about 10 percentage values higher than expected (29-37% vs an expected 20-29%), representing a 25-45 % elevation in glycosylation over the expected values. Similar elevations in “Trisaccharide” percent glycosylation are observed by NMR and amino acid analysis (Table II). The elevated glycosylation of the “Trisaccharide” derivatives leads to a 30 to 40% increase in the calculated trisaccharide distribution at the expense of the disaccharide distribution over expected (columns 2 & 3,
Nevertheless, the general differences among PSM preparations are preserved, with PSM B10 containing more disaccharide and less tetrasaccharide than PSM 12 or B7. After this discrepancy was noted, a series of studies were performed to determine the origins of the elevated glycosylation of the “Trisaccharide” derivatives.

First, it was necessary to determine whether the observed enhanced glycosylation of Ser could be the source of the increased total glycosylation due to specific destruction of nonglycosylated Ser as a result of the periodate oxidation or TFMSA treatments. This does not appear to be the case, since using α-N-acetylgalactosaminidase to remove mono saccharide GalNAc residues from β-galactosidase-treated mucin also reveals the enhanced glycosylation of Ser over Thr (13C-NMR data Figure 2C & F & Table II). Furthermore, standard amino acid analysis of the B7 and B10 derivatives show no significant loss of Ser or Thr among the differently deglycosylated derivatives (i.e. 22.5, 23.6 & 25.6% Ser and 13.8, 13.8 & 13.0 % Thr for the “Native GalNAc”, “Core 1” and “Trisaccharide” derivatives of B10 respectively; for comparison the expected values of Ser and Thr are 24.7% and 13.6% respectively). In addition, a control oxidation and reduction procedure performed on fully deglycosylated apo-PSM also failed to alter its amino acid composition (data not shown).

That the amino acid compositions are constant regardless of extent of glycosylation suggests the possibility that one or more steps in the procedure may be selectively degrading less highly glycosylated mucin tandem repeats or even intact mucin molecules. Several lines of evidence support this possibility. First, immobilized lectin chromatography of native PSM isolated from a single animal with intact oligosaccharide side chains can be fractionated into mucin species with different oligosaccharide distributions (unpublished results). These results suggest that isolated mucin is made up of a distribution of differently glycosylated molecules whose glycosylation presumably reflects the biosynthetic range of oligosaccharide structures, reflecting their different ER/Golgi localizations at the time of harvesting. It is also possible that heterogeneity may arise at the cellular level, with different cells producing differently glycosylated mucins. Regardless, the glycosylation patterns of individual mucin molecules would range from nascent-like poorly glycosylated peptide cores containing mostly short oligosaccharides, to more heavily glycosylated mature mucin peptide cores containing more fully elongated oligosaccharides. It is
expected that those mucin molecules and tandem repeats with fewer and shorter oligosaccharide side chains would be more susceptible to degradation and loss. The large material losses (50-30% relative to theory) observed after a second TFMSA treatment of apo-PSM and mucin containing only oxidized/reduced GalNAc side chains (i.e. mucin previously treated with TFMSA & subsequently oxidized and reduced) compared to the 100% recovery of mucin containing intact GalNAc side chains (i.e. mucin previously treated with TFMSA only) after the same TFMSA treatment, further suggest that deglycosylated apo mucin and mucin containing only oxidized and reduced GalNAc residues are less stable towards TFMSA than mucins containing intact GalNAc residues. Intact GalNAc residues and longer oligosaccharide side chains, therefore, appear to protect the peptide core from TFMSA degradation whereas oxidized-reduced GalNAc residues and regions of deglycosylated peptide do not. Other studies in our laboratory reveal an extreme sensitivity of apo mucin to proteolytic degradation (unpublished results). Therefore, the higher than expected extent of glycosylation observed in the “Trisaccharide” PSM derivatives reflects the nonspecific destruction by TFMSA of poorly glycosylated tandem repeat regions of the mucin molecule and the loss of a population of immature mucin molecules containing short glycan side chains, both of which become more susceptible to TFMSA after β-galactosidase and periodate treatment.footnote 3 We estimate that a 20-30% loss of poorly glycosylated mucin could account for the observed results. This results in an increase in the trisaccharide at the expense of the disaccharide over what would be expected (Table I) but no change in the core 1 or monosaccharide compositions. Therefore, the obtained site-specific glycosylation patterns are weighted towards the mucin fractions containing the more mature and more heavily glycosylated mucin molecules. These glycosylation patterns are expected to reasonably reflect the overall trends in site-specific glycosylation of the intact mucin.

**PSM tandem repeat site-specific glycosylation pattern.** Using the data in Table III we have calculated the apparent site-specific mono, di and trisaccharide content for nearly every Ser and Thr in the tandem repeat. The results are given graphically in Figure 4 for each of the three PSM preparations where the black, white and grey bars represent respectively the mono- di- and trisaccharide content. The oligosaccharide distribution is given sequentially by residue number in the left panels of Figure 4 (A-D)
and grouped separately by Ser and Thr residue type in the right panels (E-H). The average site-specific distribution for all three preparations is shown at the bottom of the figure, panels D and H. (Note that the data in Figure 4 are displayed so that the sum of the mono, di and tri saccharides represents each Ser or Thr residue’s extent of GalNAc glycosylation, further indicating the degree of Ser or Thr glycosylation at each site). Also plotted in each panel is the derived S/T density function (grey lines), which is a measure of the relative density of Ser and Thr residues in the tandem repeat sequence (34). As discussed below, the S/T density appears to correlate with various aspects of the oligosaccharide distribution and with the in vivo peptide GalNAc (OG) density. Plots evaluating the correlation of the glycosylation patterns shown in Figure 4 with S/T density are given in Figure 5.

A detailed examination of Figure 4 reveals that the glycosylation patterns of the three PSM derivatives are remarkably similar, although as discussed above, the trisaccharide content is greater than expected due to losses of poorly glycosylated tandem repeat. The distributions otherwise reflect each preparation’s original relative oligosaccharide distribution given in Table I; PSM 12 and B7 have very similar distributions, while PSM B10 has a distribution elevated in the disaccharide and lower in the mono and trisaccharide. In general, the monosaccharide values agree well with the values given in Table I, while the trisaccharide values are increased at the expense of the disaccharide. The appropriate monosaccharide values and consistent trends observed in the di and trisaccharide values between the different mucin preparations are taken as evidence that the Edman sequencing-derived values are indeed representative of the site-specific oligosaccharide distribution of the mucins under study.

An analysis of the left panels in Figure 4 reveal several trends. First there appears to be 3 regions where the monosaccharide content appears elevated in all three PSM preparations, in the range of residues 29-33, 49-50 and 57-64. These data confirm what we have shown previously on pooled mucin (34), that the monosaccharide (black bars, Figure 4) and core-1 substitution patterns (sum of white & grey bars, Figure 4) are not uniformly distributed along the tandem repeat and that their distribution appears to be directly and indirectly associated, respectively, to the Ser/Thr density plot (grey line, plots showing these correlations are shown in Figure 5 and are discussed below). Overall glycosylation and oligosaccharide distribution also appear to be Ser and Thr specific as shown by the right panels, E-H, in Figure 4. As
shown earlier (34), Ser residues are more variably glycosylated by GalNAc than Thr (total height of bars),
while the monosaccharide content of Thr residues is somewhat elevated relative to Ser residues (black bars).
As discussed above, the plots also show more trisaccharide side chains attached to Ser residues compared to Thr residues (grey bars).

The differences between the glycosylation of Ser and Thr are also revealed by the S/T density plots in Figure 5.
Plots of the percent monosaccharide versus S/T density show a statistical correlation with S/T density for Ser residues (Figure 5A, \( p<0.0001, r^2=0.28 \)) but not for Thr (Figure 5F, \( p=0.35, r^2=0.08 \)).
Inverse correlations with the core 1, di- and tri- saccharide contents with S/T density are also shown for Ser (Figure 5B, \( p<0.001, r^2=0.51 \); Figure 5C \( p=0.002, r^2=0.15 \); Figure 5D \( p<0.0001, r^2=0.47 \) respectively) but not for Thr (Figure 5G, \( p=0.026, r^2=0.15 \); Figure 5, \( p=0.079, r^2=0.11 \); Figure 5I, \( p=0.11, r^2=0.034 \) respectively).
Even the plots of percent GalNAc on Ser/Thr are strongly inversely correlated with S/T density for Ser residues (Figure 5E, \( p<0.0001, r^2=0.33 \)) and only barely inversely correlated for Thr residues (Figure 5J, \( p=0.022, r^2=0.097 \)).
The lack of a strong correlation for Thr may be the result of the relatively high and more uniform glycosylation of Thr residues by GalNAc relative to Ser. This is consistent with most in vitro studies showing that Thr residues are typically an order of magnitude more readily glycosylated compared to Ser (11, 17-21).
Note that in Figure 5, data from Ser 2 was excluded from the statistical analysis (open diamonds).
This was justified because of the very low level of “Native” glycosylation observed for Ser 2, \(~10\%\) , which leads to very large errors in determining its glycosylation pattern by difference. In addition, Ser 2 is clearly an outlier on many of the plots.

Unusual local sequence effects, beyond S/T density, presumably dominate the glycosylation of Ser 2 (34, 37).

Several trends are predicted by the S/T density plots in Figure 5. At high S/T densities (i.e. \(~1\)) the plots suggest that Ser residues will only be \(~30\%\) glycosylated by the monosaccharide and will contain no di or trisaccharide. In contrast at very low S/T densities (i.e. \(~0\)), Ser residues are predicted to be nearly fully elongated. They would be expected to contain \(~20\%\) and \(~80\%\) the di and trisaccharide side chains respectively, while containing few monosaccharide side chains. For Thr, there are few statistically significant trends with respect to S/T density, although, it would be expected that Thr residues
would be relatively uniformly glycosylated at ~80% regardless of S/T density and would have more di and fewer trisaccharide side chains.

The normalized core 1 substitution pattern (i.e. %Gal on GalNAc) is given in Figure 6 A & E (where the white and grey bars represent the di and trisaccharide proportions respectively). This plot provides the relative site-specific propensity of the core 1, UDP-galactose: α-N-acetylgalactosamine β-1,3-galactosyltransferase (core-1 β3-Gal transferase) to add β-Gal to GalNAc irrespective of extent of Ser or Thr glycosylation by GalNAc. Note that for ease of presentation, averaged values obtained from the three PSM preparations have been plotted in Figure 6. As we have shown previously on pooled mucin (34), these data clearly demonstrate that the propensity to add Gal varies uniquely along the peptide sequence, ranging from ~20 to nearly 100%. In addition, the plot suggests that the core 1 propensities appear to be indirectly associated with Ser/Thr density (grey line). Plots showing these correlations are given in Figure 7A and F for Ser and Thr respectively. Again, the propensity for core 1 glycosylation is highly correlated with S/T density for Ser (p<0.0001, r²=0.49) but not for Thr (p=0.19, r²=0.09). Note that individual data points from each PSM preparation are plotted in Figure 7 and data for Ser 2 have been omitted in the correlation calculations for the reasons stated above.

Figure 6B also shows the normalized site-specific fucosylation pattern (i.e % Fuc on Gal) which reports on the propensity of the porcine GDP-fucose: galactose β1,3 α1,2-fucosyltransferase (α2-Fuc transferase) to transfer Fuc to β-Gal. The plots suggest that the fucosyltransferase adds sugar to the β-Gal residue in a nonuniform, site-specific manner, ranging from 35% to nearly 100%. However, when data for Ser and Thr linked glycans are plotted separately, Figure 6F, it is evident that within each residue type, the normalized extent of Fuc substitution is relatively uniform: ~80% for Ser and ~45% for Thr. The site-specific variations in fucosylation in Figure 6B seem to be primarily due to Ser/Thr linkage type rather than any other unique local property of the peptide sequence. As confirmation, plots of normalized percent fucosylation vs S/T density, Figure 7B & G, do not reveal significant correlations with Ser/Thr density (p=0.06, r²=0.04 & p=0.59, r²=.003, for Ser & Thr respectively). These results suggest the porcine salivary gland α2-Fuc transferase recognizes differences in glycan Ser/Thr linkages but is otherwise relatively insensitive to polypeptide sequence or S/T density.
Note that in Figure 5D the trisaccharide content on Ser residues appears inversely proportional to the residue’s S/T density, while in Figure 7B the normalized propensity for adding Fuc to β-Gal is not correlated with S/T density. This apparent contradiction arises at the previous step in the biosynthesis of the trisaccharide by the core-1 β3- Gal transferase which produces the acceptor substrate for the α2-Fuc transferase. It is the relative synthesis of the disaccharide by the core 1 β3- Gal transferase, shown to be sensitive to S/T density (for Ser linked glycans, Figures 5B and 7A), that is responsible for the inverse correlation between trisaccharide content and S/T density.

From the obtained site-specific oligosaccharide distribution, the site-specific glycan side chain lengths can be obtained. Side chain length can be calculated in two different way: one that includes the non-glycosylated proportion of Ser or Thr residues in the calculation, normalizing to Ser/Thr, and the other that excludes the non-glycosylated proportion of Ser and Thr, normalizing to peptide GalNAc. The former value provides an indication of the full extent of glycosylation of an individual Ser or Thr residue, while the latter provides information on the average length of the oligosaccharide side chains that are actually attached to a given residue regardless of the extent that the Ser or Thr is O-glycosylated. These different side chain length distributions, averaged over the three PSM preparations, for the ease of presentation, are shown in Figure 6 C, D, G & H. These plots also show that the side chain lengths are not uniformly distributed along the peptide core (ranging from ~0.2 to ~2.2 and ~1.5 to ~2.8 residues respectively) and that there also appears to be an inverse association of side chain length with Ser/Thr density (grey line). This is borne out for Ser-linked glycans (Figure 7C & D, p<0.0001, r²=0.48 & p<0.0001, r²=0.46) and less so for Thr-linked glycans (Figure 7H & I, p=0.0053, r²=0.17 & p=0.32, r²=0.06). No significant difference between the average side chain length of Ser linked and Thr linked glycan side chains is observed when normalized to Ser or Thr (1.4 and 1.5 residues respectively, Figure 6G), although Ser residues show greater variability ranging from ~0.2 to ~2.2 residues versus ~1.2 to ~1.8 residues for Thr. Only after normalizing to peptide GalNAc do differences in the average side chain lengths become evident, with Ser residues on average having slightly longer side chains than Thr (2.4 versus 2.0 residues respectively, Figure 6H). The variability between Ser and Thr linked glycans are similar, with Ser ranging from ~1.4 to ~2.7 residues and Thr ranging from ~1.4 to ~2.4 residues. Least
square fits of the Ser side chain length data in Figures 7C & D suggests that at very low Ser/Thr densities, the side chain lengths would be in the range of 2.5 to 3 residues. At high S/T densities, lengths in the range of ~0 and ~1 residues would be expected. Plots for Thr linked glycans (Figures 7H & I) are not statistically correlated with S/T density.

Figures 7E and 7J show the Ser- and Thr-specific plots of S/T density versus overall peptide GalNAc (OG) density derived from the “Native” glycosylation patterns given in Table III. From these plots, it is clear that the S/T density values are highly correlated for both Ser and Thr residues, with the in vivo density of GalNAc (OG) glycosylation (Figure 7E, p<0.0001, r²=0.69, Figure 7J, p<0.0001, r²=0.86). These high correlations indicate that S/T density may be taken as a reasonable representation of a site’s local density of peptide GalNAc glycosylation. On this basis we propose that the trends observed in Figures 5 and 7 with respect to S/T density and particularly with Ser linked glycans, can be explained by inhibitory effects of neighboring GalNAc glycosylation due to steric interactions of neighboring GalNAc residue with transferase. Thus at high S/T densities, and by inference high neighboring GalNAc (OG) densities, the extent of site-specific Ser/Thr glycosylation by GalNAc is decreased due to the reduced access of the ppGalNAc transferase to the peptide core (Figure 5E)². Likewise, the accessibility of the core 1 β3-Gal transferase to a given GalNAc may also be reduced at high S/T density (i.e. high OG density) due to steric interactions. At high S/T (OG) densities it is found that the extent of core 1 substitution is decreased (Figure 5B, C & D, & Figure 7A) while the monosaccharide content is increased (Figure 5A). These trends cumulatively result in the observed inverse behavior of side chain length with Ser/Thr (OG) density observed in Figures 7C & D. It is intuitively very reasonable to expect that ppGalNAc transfease and core 1 β3-Gal transferase would be sensitive to local GalNAc glycosylation density, while α2-Fuc transferase, whose acceptor substrate is 2 residues removed from the peptide core would not. Attempts to relate the observed site-specific glycosylation patterns to other properties of the PSM mucin sequence such as predicted secondary structure have thus far been uninformative.

**Discussion**

The goal of this work was to further determine the extent that O-glycosylation patterns may be
affected by local peptide sequence by extending our studies to the characterization of the site-specific
distribution of the blood group H trisaccharide, $\alpha$-Fuc(1-2) $\beta$-Gal(1-3) $\alpha$-GalNAc-O-Ser/Thr, on the PSM
tandem repeat. This work represents the most complete determination of the in vivo site-specific
oligosaccharide distribution of any heavily O-glycosylated mucin type glycoprotein Footnote 5.

The present work was undertaken using mucin isolated from 3 individual blood group A-negative
animals, with nearly identical overall oligosaccharide side chain distributions (Table I). This was done to
eliminate potential difficulties arising from characterizing mucin pooled from individuals with widely
varying oligosaccharide side chain distributions. The obtained glycosylation patterns were found to be
very similar among the different animals (Figure 4), and in keeping with the peptide-linked GalNAc and
core 1 glycosylation patterns reported earlier on a pooled mucin preparation composed of a mixture of
blood group A-positive and negative mucins (34, 37). These findings suggest that in the porcine salivary
gland, the activities of the initiating ppGalNAc transferases and core-1 $\beta$-3-Gal transferase that synthesize
the core and core 1 glycans remain relatively constant between animals regardless of individual blood
group activities. Individual to individual variability in the PSM oligosaccharide structures, therefore,
arises predominantly at the stage of tri- and tetra- saccharide biosynthesis.

As discussed in the Results section, the procedure for obtaining the trisaccharide distribution leads to
the loss of a fraction (20-30%) of poorly glycosylated mucin, either by the specific loss of nascent poorly
glycosylated mucin molecules or by nonspecific loss of poorly glycosylated tandem repeats. As a result,
the obtained glycosylation patterns are representative of a more mature more highly glycosylated fraction
of mucin molecules.

This work has confirmed our earlier findings (34) that the porcine core-1 $\beta$3-Gal transferase is
sensitive to peptide sequence and that a significant extent of its activity, particularly for Ser linked
glycans, appears to be inversely correlated to local hydroxyamino acid density (Figures 5B & 7A). This
effect was proposed to be due to inhibition of the transferase by local steric effects of neighboring
glycosylated Ser and Thr residues (34). The highly correlated relationship of the S/T density function
with the similarly derived in vivo peptide GalNAc (OG) density function, shown in Figures 7E & J,
supports this interpretation. As it is observed in vitro, it is likely that in vivo Thr residues are more
rapidly glycosylated by GalNAc than Ser residues; therefore, the initially glycosylated Thr residues may interfere with the subsequent glycosylation and elongation of neighboring Ser residues. In contrast, the more rapidly glycosylated Thr residues are likely to initially have poorly glycosylated neighbors, hence their initial rates of core 1 substitution would be less likely to affected by neighboring site glycosylation. We may be observing the differential in Ser and Thr glycosylation kinetics, which may explain why the glycosylation of Ser residues is more highly inversely correlated to S/T density than the glycosylation of Thr residues. The above findings are also consistent with earlier observations from in vitro glycosylation studies on model glycopeptides where the specificities of both the ppGalNAc and core-1 transferases were proposed to be affected by steric effects of neighboring glycosylated residues (22, 23, 32, 33, 58, 59).

Other presently unknown sequence or conformational factors associated with S/T density may be the origin of the observed correlations. The termination of monosaccharide elongation by α2-6 sialylation of GalNAc is not a likely alternative explanation, since previous studies of the PSM oligosaccharide alditols (51) indicate that only about ~25% of the monosaccharide GalNAc residues are substituted by NeuNGl while ~50% of the core 1 substituted glycans were NeuNGl substituted. One would expect higher levels of monosaccharide sialylation if sialylation were playing a significant role in terminating GalNAc elongation. Work is in progress determining the site-specific sialylation patterns of the PSM monosaccharide side chains to confirm this conclusion. Peptide substrate specificities for the sialyltransferases that specifically add α2-6 sialic acid to unsubstituted peptide linked GalNAc (ST6GalNAc I & II) have yet to be characterized (56, 57).

In this work we have also shown that the extent of site-specific glycosylation of Ser residues by peptide-linked GalNAc appears to be inversely correlated with its S/T density (Figure 5 E), suggesting that in vivo one or more of the porcine salivary gland ppGalNAc transferase(s) may be affected by steric interactions of neighboring glycosylated residues. Indeed, our in vitro glycosylation studies on apo-PSM with ppGalNAc T1 tend to support this conclusion (unpublished data).

For the 3 PSM preparations studied, the tandem repeat trisaccharide distribution is shown to vary in a relatively reproducible sequence-dependent manner (Figures 4 and 6B). However, after grouping the
Ser and Thr residues separately it was shown that these differences were predominantly due to the glycan’s Ser/Thr glycosidic linkage (Figure 6F) rather than to differences in neighboring peptide sequence or to Ser/Thr density (Figure 7B & G). These results suggest that the porcine salivary gland α2-Fuc transferase, responsible for forming the trisaccharide, is particularly sensitive to Ser/Thr glycosidic linkage and less so to peptide sequence or Ser/Thr density. Possible origins for these differences are discussed below. The total trisaccharide content at a given site appears to be inversely correlated to Ser/Thr density on Ser linked glycans (Figure 5D). As discussed in the Results, this is due to the apparent sensitivity of the core-1 β3-Gal transferase, that synthesizes the precursor disaccharide, to S/T density (Figures 5B & 7A). Ser-linked glycan side chain lengths are also correlated with Ser/Thr density (Figure 7C & D) for similar reasons.

The porcine α2-Fuc transferase responsible for forming the trisaccharide, in contrast to the core 1 β3-Gal transferase, shows no apparent sensitivity to adjacent peptide sequence (Figure 6F) or to local Ser/Thr density (Figure 7B & G). This suggests that the terminal β-galactose acceptor may project far enough away from the peptide core that neighboring peptide and carbohydrate side chains fail to significantly influence transferase activity and/or accessibility. However, a pronounced difference in α2-Fuc transferase activity against glycans attached to Ser and Thr is observed, with Ser linked glycans containing ~1.6 fold more Fuc than Thr linked glycans (Figure 6F). This unusual finding suggests that there may be differences in the conformation or dynamics of the acceptor disaccharide. These differences may influence the binding or accessibility of the acceptor to the transferase or the transferase substrate binding site may specifically recognize the peptide glycosidic linkage. Threonine O-glycosidic linkages are more restricted and less flexible than Ser linkages and may have different conformations (50, 60-63). Differential α2-6 sialylation by STGalNAc I or II of Ser or Thr glycans may also play a role in directing the core 1 glycan fucosylation towards Ser-linked glycans. The Ser/Thr peptide substrate specificities of the STGalNAc transferases active against the peptide linked disaccharide T antigen remain unexplored.

Our $^{13}$C-NMR analysis of the oligosaccharide structures on PSM isolated from a series of A blood group-negative glands reveals what appears to be two general classes of mucins with different fucosylation patterns. One class of mucins, whose site-specific patterns were studied here, contains
typically 1/3 each of the mono, di and tri saccharides while the second class contains typically ~1/3 the monosaccharide and 1/2 to 2/3 the trisaccharide, and very little disaccharide. It is possible that differences in the expression or activity of the porcine FUT2 (Se enzyme) α2-fucosyltransferase may be responsible for these differences among A blood group-negative animals (64, 65). Since mucin substrates containing β-Gal(1-3) α-GalNAc-O-Ser/Thr can be fully fucosylated by excess porcine salivary gland α2-Fuc transferase (presumably the Se enzyme) in vitro (66, 67), a simple difference in copy number of the enzyme could readily account for the observed different glycosylation patterns. This explanation is in keeping with the generally observed lower activity (lower \( V_{\text{max}}/K_{\text{M}} \) values) of the Se enzyme compared to the H blood group α2 fucosyltransferase, FUC1 (68-71). On this basis, the observed preference for the fucosylation of Ser linked glycans over Thr residues, in the less highly fucosylated mucins could arise as a result of a low copy number of the Se enzyme combined with a moderate kinetic preference for Ser linked glycans over Thr linked glycans for the transferase. Recently it has been shown that a polymorphism in the porcine FUT1 gene leads to protection against edema and post-weaning diarrhea in swine infected with toxigenic *E. coli* (64). The expression of FUT1 in the porcine salivary gland is unknown but it is conceivable that potential differences in the expression of FUT1 and its alleles may further account for the variable fucosylation observed among individual swine and between Ser and Thr linked glycans.

It is interesting to compare the glycosylation patterns of individual residues. For example the 3 least glycosylated residues of Figure 4D, Ser2 Ser54 and Ser62 (9 to 32% glycosylated), all appear to have relatively high core-1 substitutions (54-99%, Figure 6A) and high levels of fucosylation (84-95%, Figure 6B), and therefore relatively long average side chain lengths (2.1-2.9 sugar residues per GalNAc, Figure 6D). In contrast, Ser32 and Ser63 which are 45% and 39% glycosylated on average by GalNAc, respectively, are poorly core-1 substituted (21%, Figure 6A) but these are highly fucosylated (100 & 90%, Figure 6B). These residues therefore have somewhat shorter average side chain lengths (~1.4 sugars per GalNAc, Figure 6D). The glycans on Thr29, Thr30, Thr49 and Thr50 also have shorter average side chain lengths (1.4-1.8 sugars per GalNAc, Figure 6D) but this is due to both lower core-1 glycosylation (30-55%, Figure 6A) and lower fucosylation (33-45%, Figure 6B). Thr37, Thr39 and Thr52 have
somewhat longer average side chain lengths (2.2-2.4 sugars per GalNAc, Figure 6D) compared to the other Thr residues, due to both higher core-1 substitution (78-93%, Figure 6A) and higher fucosylation (48-72%, Figure 6B). Those residues with shorter side chains, Ser32, S63, Thr29, Thr30, Thr49 & Thr50, appear near peaks in the Ser/Thr density function (Figure 6D). Also note that the residues with the lowest and highest Ser/Thr densities, Ser43 and Ser63 have nearly the longest and shortest average side chain lengths, 2.7 and 1.4 residues per GalNAc respectively (Figure 6D). Placing the longer side chains in less densely glycosylated regions of the peptide may help render the peptide core more uniformly protected against proteolytic degradation.

These studies clearly show that a mucin’s polypeptide sequence, either directly or indirectly, affects the distribution of O-glycan structures in vivo, in a reproducible, site-specific manner. Apparent correlations of the density of hydroxyamino acids with the distribution of Ser linked glycan structures have lead us to propose that steric interactions of neighboring glycosylated residues may play a significant role in the initial glycosylation and core 1 elongation of the porcine mucin, where presumably most Ser and Thr residues would be in sequences and conformations optimized for relatively efficient O-glycosylation. We further suggest that the observed differences in sensitivity of Ser and Thr to S/T density, may arise because of their different rates of glycosylation by GalNAc. Since there are reproducible deviations in the correlations with S/T density, additional factors such as local peptide sequence or conformation must also contribute to the modulation of the initial O-glycosylation of the mucin. It is also likely that these latter factors may play more significant roles for substrates that may not be as optimized for O-glycosylation or that are less densely glycosylated as the mucins. In contrast, the distribution of the trisaccharide suggests that the addition of the Fuc residue by the porcine α2-fucosyltransferase is relatively insensitive to local peptide sequence but instead appears to be sensitive to a presently unknown property that differs between the Ser and Thr peptide glycosidic linkages. These findings indicate that the initial elongation of O-glycans is not an entirely passive or random process and that considerable work needs to be done before we have an understanding of the molecular basis of its regulation.

Acknowledgments
The authors would like to acknowledge the technical assistance of Cheryl Owens of the Molecular Biology Core Laboratory of the CWRU Cancer Center, CWRU Center for AIDS Research, and Skin Diseases Research Center (funded in part by NIH awards P30-CA43703, P30-AI36219 and P30-AR39750). We would also like to thank Ashok Kahatri of the Protein/Peptide Core Facility Massachusetts General Hospital for the amino acid analysis and for helpful discussions. The assistance of Jessica Levine is also acknowledged.
References


Footnotes

Footnote 1:
Abbreviations: α2-Fuc transferase, GDP-fucose: galactose β1,3 α1,2-fucosyltransferase; core-1 β3-Gal transferase, UDP-galactose: α-N-acetylgalactosamine β-1,3-galactosyltransferase; Fuc, fucose; Gal, galactose; GalNAc, N-acetylgalactosamine; NeuNGl, N-glycoloylneuraminic acid or sialic acid; ppGalNAc transferase, UDP-GalNAc:polypeptide α-GalNAc transferase; PSM, porcine submaxillary mucin; PTH, phenylthiohydantoin; “Native GalNAc” or (T)TTTR-PSM, “Core 1” or (T)TROTR-PSM, “Trisaccharide” or (T)TROGNTR-PSM; a series of differentially deglycosylated preparations of the PSM multiple tandem repeat domain described in the methods and results, where (T) represents the 81 residue tandem repeat domain obtained after trypsinolysis; TFMSA, trifluoromethanesulfonic acid.

Footnote 2:
Note that the weighting coefficients reported in this manuscript differ from the incorrectly given coefficients in reference (34). The numerical values listed for the Ser-Thr densities of the PSM tandem repeat in reference (34), however, are correct and are actually based on the coefficients given in this work.

Footnote 3.
The differences in overall glycosylation of the α-Nacetylgalactosaminidase derived “Trisaccharide”-PSM B10 compared to the derivative obtained via oxidation and reduction after TFMSA treatment (~44 vs ~30%, Table II) suggest that the presence of oxidized and reduced GalNAc residues afford some protection from degradation by TFMSA. Therefore, the oxidation-reduction-TFMSA approach for selectively eliminating unsubstituted GalNAc residues is at present the best available approach for these studies.

Footnote 4
It should be noted that it is not inconsistent to have S/T density inversely correlated with the site-specific glycosylation of Ser residues by GalNAc while also having S/T density directly correlated with GalNAc (OG) density. In the former comparison a specific site is compared to S/T density while in the latter the net properties of several neighboring residues are compared to S/T density.

Footnote 5
All that remains in order to complete the characterization of the PSM tandem repeat glycosylation pattern is the determination of its’ site-specific sialylation pattern, which is currently in progress.
### Table I

$^{13}$C NMR derived glycan compositions of individual A' PSM preparations and estimated “Core 1” PSM and “Trisaccharide” PSM percent GalNAc-O-Ser/Thr compositions

<table>
<thead>
<tr>
<th>Preparation</th>
<th>mono</th>
<th>di</th>
<th>tri</th>
<th>di+tri</th>
<th>“Core 1” PSM (^c)</th>
<th>“Trisaccharide” PSM (^d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSM 12</td>
<td>36(31)(^e)</td>
<td>34(16)</td>
<td>32(53)</td>
<td>66(69)</td>
<td>50(42)</td>
<td>24(32)</td>
</tr>
<tr>
<td>PSM B7</td>
<td>29(27)</td>
<td>33(19)</td>
<td>38(54)</td>
<td>71(73)</td>
<td>53(50)</td>
<td>29(37)</td>
</tr>
<tr>
<td>PSM B10</td>
<td>24(22)</td>
<td>49(35)</td>
<td>27(43)</td>
<td>76(78)</td>
<td>57(53)</td>
<td>20(29)</td>
</tr>
<tr>
<td>Avg</td>
<td>30(29)</td>
<td>38(23)</td>
<td>32(50)</td>
<td>70(73)</td>
<td>53(48)</td>
<td>24(33)</td>
</tr>
</tbody>
</table>

\(^a\) as defined in Figure 1.

\(^b\) percent peptide linked GalNAc present on Ser/Thr calculated from the given oligosaccharide distribution and using a 75 % initial native GalNAc-O-Ser/Thr glycosylation.

\(^c\) Extent of glycosylation expected for “Core 1” PSM preparations, (T)TROTR-PSM.

\(^d\) Extent of glycosylation expected for “Trisaccharide” PSM preparations, (T)TROGNTR-PSM.

\(^e\) Values in parenthesis derived from the Edman sequencing data given in Table III.
Table II.

*GalNAc glycosylation of partially deglycosylated PSM tandem repeat domains.*

<table>
<thead>
<tr>
<th>PSM Preparation</th>
<th>“Native GalNAc” PSM&lt;sup&gt;a&lt;/sup&gt;</th>
<th>“Core 1” PSM&lt;sup&gt;b&lt;/sup&gt;</th>
<th>“Trisaccharide” PSM&lt;sup&gt;c&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>%Thr-OG %SerOG %CHO</td>
<td>%Thr-OG %SerOG %CHO</td>
<td>%Thr-OG %SerOG %CHO</td>
</tr>
<tr>
<td><strong>Expected&lt;sup&gt;d&lt;/sup&gt;</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Avg</td>
<td>- - 75</td>
<td>- - 53</td>
<td>- - 24</td>
</tr>
<tr>
<td>range</td>
<td>- - 75</td>
<td>- - 46-57</td>
<td>- - 20-29</td>
</tr>
<tr>
<td><strong>13C NMR&lt;sup&gt;e&lt;/sup&gt;</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>79 59 64</td>
<td>46 47 44</td>
<td>32 32 29</td>
</tr>
<tr>
<td>B7</td>
<td>71 75 76</td>
<td>47 53 50</td>
<td>21 40 31</td>
</tr>
<tr>
<td>B10</td>
<td>74 71 71</td>
<td>58 59 58</td>
<td>26 40 29</td>
</tr>
<tr>
<td>B10(TR)&lt;sup&gt;f&lt;/sup&gt;</td>
<td>71 65 62</td>
<td>56 52 51</td>
<td>32 34 32</td>
</tr>
<tr>
<td><strong>13C NMR Avg</strong></td>
<td>74 68 68</td>
<td>52 53 51</td>
<td>28 36 30</td>
</tr>
<tr>
<td>B10 Enz&lt;sup&gt;g&lt;/sup&gt;</td>
<td>- - -</td>
<td>- - -</td>
<td>15 38 31</td>
</tr>
<tr>
<td><strong>1H NMR&lt;sup&gt;h&lt;/sup&gt;</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B7(TR)&lt;sup&gt;f&lt;/sup&gt;</td>
<td>93 61 74</td>
<td>46 50 50</td>
<td>36 58 43</td>
</tr>
<tr>
<td>B10(TR)</td>
<td>82 58 69</td>
<td>68 53 60</td>
<td>28 42 28</td>
</tr>
<tr>
<td><strong>1H NMR Avg</strong></td>
<td>87 59 71</td>
<td>57 51 55</td>
<td>33 50 36</td>
</tr>
<tr>
<td><strong>Edman sequencing&lt;sup&gt;i&lt;/sup&gt;</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12(TR)&lt;sup&gt;f&lt;/sup&gt;</td>
<td>72 54 61</td>
<td>38 42 42</td>
<td>21 37 32</td>
</tr>
<tr>
<td>B7(TR)</td>
<td>70 63 68</td>
<td>48 49 49</td>
<td>23 43 37</td>
</tr>
<tr>
<td>B10(TR)</td>
<td>71 63 68</td>
<td>55 50 53</td>
<td>19 33 29</td>
</tr>
<tr>
<td>Edman Avg</td>
<td>71 60 65</td>
<td>46 46 47</td>
<td>21 37 32</td>
</tr>
<tr>
<td>B10 Enz&lt;sup&gt;j&lt;/sup&gt; (TR)</td>
<td>- - -</td>
<td>- - -</td>
<td>31 54 44</td>
</tr>
<tr>
<td><strong>Amino acid analysis after β-elimination&lt;sup&gt;k&lt;/sup&gt;</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B7(TR)&lt;sup&gt;f&lt;/sup&gt;</td>
<td>77 75 76</td>
<td>nd nd nd</td>
<td>33 55 48</td>
</tr>
<tr>
<td>B10(TR)</td>
<td>nd nd nd</td>
<td>nd nd nd</td>
<td>34 43 40</td>
</tr>
<tr>
<td>AAA Avg</td>
<td>77 75 76</td>
<td>nd nd nd</td>
<td>33 49 44</td>
</tr>
</tbody>
</table>

<sup>a</sup> GalNAc glycosylation of partially deglycosylated PSM tandem repeat domains.  
<sup>b</sup> Core 1 PSM tandem repeat domain.  
<sup>c</sup> Trisaccharide PSM tandem repeat domain.  
<sup>d</sup> Expected values.  
<sup>e</sup> 13C NMR values.  
<sup>f</sup> B10(TR) PSM tandem repeat domain.  
<sup>g</sup> B10 Enz (TR) PSM tandem repeat domain.  
<sup>h</sup> 1H NMR values.  
<sup>i</sup> Edman sequencing values.  
<sup>j</sup> B10 Enz (TR) PSM tandem repeat domain.  
<sup>k</sup> Amino acid analysis after β-elimination.
Legend to Table II

a) “Native GalNAc” PSM tandem repeat domains having glycans trimmed to GalNAc by TFMSA, giving ‘native’ glycosylation pattern.
b) “Core 1” PSM tandem repeat domains after removal of mono saccharide GalNAc by periodate and trimmed by TFMSA, giving core-1 glycosylation pattern.
c) “Trisaccharide” PSM tandem repeat domains obtained by β-galactosidase and periodate treated and subsequently trimmed by TFMSA, giving trisaccharide glycosylation pattern. This column also includes a B10 sample whose monosaccharide GalNAc residues were removed by α-N-acetylgalactosaminidase, see notes g & j below.
d) Obtained using a native glycosylation of 75% and oligosaccharide composition values for PSM 12, B7 & B10 given in Table I.
e) %Thr-OG and %Ser-OG represent the extent of glycosylation of the Ser and Thr residues obtained from the GalNAc C1 anomeric carbon and α-Ser and α-Thr carbon resonances. %CHO values are based on the GalNAc C1, C5, C2, & N-acetyl methyl carbons in addition to the α-Ser and α-Thr resonances. Because additional resonances were utilized to obtain the %CHO values these values are not necessarily consistent with the average of the listed %Ser-OG and %Thr-OG values. All values were normalized to the α-Gly resonance area and percent glycosylation determined utilizing the area ratios based on the theoretical amino acid composition of the PSM 81 residue tandem repeat.
f) (TR) corresponds to data from the 81 residue tryptic tandem repeat.
g) Values for PSM B10 containing only intact trisaccharide side chains after having the mono and disaccharide side chains removed enzymatically by neuraminidase, β-galactosidase and α-N-acetylgalactosaminidase treatments. The sample was not treated with TFMSA nor trypsin.
h) Proton NMR spectra at 45°C of the tryptic 81 residue tandem repeat glycopeptides that have been sequenced. Percent glycosylation are based on the areas of GalNAc anomic protons relative to the Ile & Val methyl resonances utilizing the theoretical amino acid composition. Percent Ser and Thr glycosylation are based on the Ser/Thr GalNAc H2 proton resonances after correcting for non GalNAc overlap.
i) Values derived from Table III.
j) Values for PSM B10 tryptic tandem repeat glycopeptide after the mono and disaccharides were enzymatically removed by neuraminidase, β-galactosidase and α-N-acetylgalactosaminidase, and the trisaccharide trimmed to GalNAc by mild TFMSA/anisole treatment.
k) Values obtained by amino acid analysis after β-elimination and NaBH₄/PdCl₂ reduction (55). nd, not determined.
**Table III.**

Sequence specific glycosylation of selectively deglycosylated PSM tryptic tandem repeat glycopeptides

<table>
<thead>
<tr>
<th>Residue</th>
<th>12 avg</th>
<th>12 sd</th>
<th>B7 avg</th>
<th>B7 sd</th>
<th>B10 avg</th>
<th>B10 sd</th>
<th>All avg</th>
<th>All sd</th>
<th>PSM Preparatione</th>
<th>PSM Preparatione</th>
<th>PSM Preparatione</th>
</tr>
</thead>
<tbody>
<tr>
<td>S2</td>
<td>9</td>
<td>2</td>
<td>12</td>
<td>8</td>
<td>1</td>
<td>9</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>S6</td>
<td>75</td>
<td>4</td>
<td>88</td>
<td>86</td>
<td>1</td>
<td>82</td>
<td>6</td>
<td>69</td>
<td>3</td>
<td>88</td>
<td>80</td>
</tr>
<tr>
<td>S7</td>
<td>82</td>
<td>2</td>
<td>90</td>
<td>87</td>
<td>1</td>
<td>85</td>
<td>3</td>
<td>75</td>
<td>7</td>
<td>90</td>
<td>82</td>
</tr>
<tr>
<td>S13</td>
<td>79</td>
<td>1</td>
<td>80</td>
<td>86</td>
<td>0</td>
<td>82</td>
<td>4</td>
<td>68</td>
<td>3</td>
<td>74</td>
<td>79</td>
</tr>
<tr>
<td>S14</td>
<td>47</td>
<td>3</td>
<td>60</td>
<td>52</td>
<td>4</td>
<td>51</td>
<td>6</td>
<td>41</td>
<td>0</td>
<td>55</td>
<td>50</td>
</tr>
<tr>
<td>S17</td>
<td>67</td>
<td>2</td>
<td>76</td>
<td>76</td>
<td>0</td>
<td>72</td>
<td>5</td>
<td>52</td>
<td>2</td>
<td>64</td>
<td>66</td>
</tr>
<tr>
<td>T22</td>
<td>80</td>
<td>2</td>
<td>81</td>
<td>86</td>
<td>3</td>
<td>83</td>
<td>3</td>
<td>43</td>
<td>0</td>
<td>50</td>
<td>64</td>
</tr>
<tr>
<td>S23</td>
<td>76</td>
<td>1</td>
<td>77</td>
<td>91</td>
<td>7</td>
<td>82</td>
<td>8</td>
<td>70</td>
<td>4</td>
<td>76</td>
<td>77</td>
</tr>
<tr>
<td>T29</td>
<td>77</td>
<td>1</td>
<td>79</td>
<td>83</td>
<td>9</td>
<td>80</td>
<td>6</td>
<td>17</td>
<td>2</td>
<td>30</td>
<td>27</td>
</tr>
<tr>
<td>T30</td>
<td>77</td>
<td>3</td>
<td>76</td>
<td>83</td>
<td>6</td>
<td>79</td>
<td>5</td>
<td>27</td>
<td>1</td>
<td>34</td>
<td>47</td>
</tr>
<tr>
<td>S32</td>
<td>43</td>
<td>3</td>
<td>46</td>
<td>46</td>
<td>3</td>
<td>45</td>
<td>2</td>
<td>8</td>
<td>3</td>
<td>11</td>
<td>10</td>
</tr>
<tr>
<td>S33</td>
<td>66</td>
<td>3</td>
<td>70</td>
<td>71</td>
<td>5</td>
<td>69</td>
<td>4</td>
<td>39</td>
<td>0</td>
<td>48</td>
<td>42</td>
</tr>
<tr>
<td>T37</td>
<td>79</td>
<td>0</td>
<td>77</td>
<td>78</td>
<td>8</td>
<td>78</td>
<td>5</td>
<td>47</td>
<td>1</td>
<td>59</td>
<td>77</td>
</tr>
<tr>
<td>T39</td>
<td>72</td>
<td>2</td>
<td>80</td>
<td>82</td>
<td>9</td>
<td>78</td>
<td>8</td>
<td>72</td>
<td>10</td>
<td>80</td>
<td>70</td>
</tr>
<tr>
<td>S43</td>
<td>75</td>
<td>2</td>
<td>86</td>
<td>79</td>
<td>1</td>
<td>80</td>
<td>5</td>
<td>71</td>
<td>5</td>
<td>75</td>
<td>79</td>
</tr>
<tr>
<td>S47</td>
<td>35</td>
<td>4</td>
<td>55</td>
<td>50</td>
<td>1</td>
<td>47</td>
<td>9</td>
<td>29</td>
<td>3</td>
<td>39</td>
<td>44</td>
</tr>
<tr>
<td>T49</td>
<td>77</td>
<td>0</td>
<td>82</td>
<td>75</td>
<td>11</td>
<td>78</td>
<td>7</td>
<td>32</td>
<td>3</td>
<td>34</td>
<td>60</td>
</tr>
<tr>
<td>T50</td>
<td>73</td>
<td>1</td>
<td>77</td>
<td>78</td>
<td>2</td>
<td>76</td>
<td>3</td>
<td>34</td>
<td>2</td>
<td>50</td>
<td>48</td>
</tr>
<tr>
<td>T52</td>
<td>72</td>
<td>3</td>
<td>69</td>
<td>68</td>
<td>4</td>
<td>70</td>
<td>4</td>
<td>50</td>
<td>5</td>
<td>63</td>
<td>65</td>
</tr>
<tr>
<td>S54</td>
<td>15</td>
<td>3</td>
<td>26</td>
<td>24</td>
<td>3</td>
<td>22</td>
<td>5</td>
<td>22</td>
<td>6</td>
<td>21</td>
<td>20</td>
</tr>
<tr>
<td>S57</td>
<td>58</td>
<td>2</td>
<td>82</td>
<td>76</td>
<td>9</td>
<td>72</td>
<td>12</td>
<td>48</td>
<td>3</td>
<td>61</td>
<td>61</td>
</tr>
<tr>
<td>S59</td>
<td>64</td>
<td>4</td>
<td>72</td>
<td>72</td>
<td>3</td>
<td>71</td>
<td>5</td>
<td>35</td>
<td>6</td>
<td>38</td>
<td>38</td>
</tr>
<tr>
<td>T60</td>
<td>69</td>
<td>1</td>
<td>73</td>
<td>72</td>
<td>1</td>
<td>71</td>
<td>5</td>
<td>47</td>
<td>6</td>
<td>58</td>
<td>71</td>
</tr>
<tr>
<td>S62</td>
<td>31</td>
<td>1</td>
<td>33</td>
<td>31</td>
<td>6</td>
<td>32</td>
<td>4</td>
<td>14</td>
<td>0</td>
<td>18</td>
<td>21</td>
</tr>
<tr>
<td>S63</td>
<td>34</td>
<td>-</td>
<td>40</td>
<td>40</td>
<td>0</td>
<td>39</td>
<td>3</td>
<td>7</td>
<td>1</td>
<td>8</td>
<td>11</td>
</tr>
<tr>
<td>S64</td>
<td>60</td>
<td>-</td>
<td>59</td>
<td>63</td>
<td>3</td>
<td>61</td>
<td>5</td>
<td>45</td>
<td>0</td>
<td>46</td>
<td>55</td>
</tr>
<tr>
<td>S66</td>
<td>56</td>
<td>-</td>
<td>78</td>
<td>75</td>
<td>6</td>
<td>73</td>
<td>9</td>
<td>49</td>
<td>1</td>
<td>53</td>
<td>64</td>
</tr>
<tr>
<td>T70</td>
<td>68</td>
<td>-</td>
<td>76</td>
<td>81</td>
<td>-</td>
<td>75</td>
<td>6</td>
<td>50</td>
<td>0</td>
<td>70</td>
<td>72</td>
</tr>
<tr>
<td>S73</td>
<td>51</td>
<td>-</td>
<td>79</td>
<td>77</td>
<td>-</td>
<td>69</td>
<td>13</td>
<td>44</td>
<td>0</td>
<td>55</td>
<td>60</td>
</tr>
</tbody>
</table>

Avg 61 68 68 66 6 42 50 53 47 7 32 37 29 32 6
SerOG 54 63 63 60 6 42 49 50 46 5 37 43 33 37 5
Legend to Table III

a) Extent of glycosylation obtained from Edman sequencing of the N terminal (residues 1- >40) and C terminal, Glu C (residues 39--75), glycopeptides as described in the methods.

b) “Native GalNAc” PSM (TTTR-PSM), PSM tryptic tandem repeat glycopeptide with glycans trimmed to GalNAc by TFMSA, giving total ‘native’ (full) glycosylation pattern.

c) “Core-1” PSM (TTROTR-PSM), PSM tryptic tandem repeat glycopeptide with mono saccharide GalNAc glycans removed by periodate and trimmed by TFMSA, giving core-1 glycosylation pattern.

d) “Trisaccharide” PSM (TTROGNTR-PSM), PSM tryptic tandem repeat glycopeptide β-galactosidase and periodate treated, subsequently trimmed by TFMSA, giving trisaccharide glycosylation pattern.

e) Residue specific percent GalNAc glycosylation values for PSM preparations 12, B7, B10 and their combined average. Standard deviations are given under ‘sd’ labeled column where appropriate.

Number of determinations: PSM 12: “Native GalNAc” PSM, 2 N-terminal sequence determinations; “Core-1” PSM, 2 N-terminal and 2 C-terminal sequence determinations; “Trisaccharide” PSM, 4 N-terminal and 3 C-terminal sequence determinations. PSM B7: 1 N-terminal and 1 C-terminal determination for each differently deglycosylated preparation. PSM B10: 2 N-terminal and 1 C-terminal determination for each differently deglycosylated preparation. Combined average values under ‘All’ columns represent the average of all determinations in the group therefore averages for “Native GalNAc”, “Core-1” and “Trisaccharide” represent a total of 7, 9 and 12 combined N-terminal and C-terminal sequence determinations.
Figure Legends

Figure 1.
Porcine mucin O-glycan structures and tandem repeat amino acid sequence. A, possible O-linked glycan structures ranging from the monosaccharide, Tn, determinant to the tetrasaccharide, A blood group determinant (47, 51, 52). For each oligosaccharide a NeuNGl, sialic acid, residue may be 2,6-linked to the peptide core GalNAc residue. Typically 40-50% of the oligosaccharides contain NeuNGl(50). B, sequence of the PSM 81 residue tryptic tandem repeat (48). The repeat contains 31 hydroxyamino acids, 20 Ser and 11 Thr, dispersed along the sequence. Native mucin contains about 100 contiguous repeats of this sequence each containing one trypsin susceptible site (49).

Figure 2.
Carbon 13 NMR spectra of native and selectively deglycosylated porcine mucin B10. Spectra A through C are for native and enzymatically deglycosylated mucin while spectra D-F are for the series of “Native”, “Core-1”, and “Trisaccharide” mucin tryptic tandem repeats containing only peptide linked GalNAc residues. The glycosylation patterns of the latter were determined by Edman amino acid sequencing (see Figure 3 and Table III). A) TR-PSM containing intact oligosaccharide side chains. B) TR-PSM containing only the mono and trisaccharide side chains, obtained after treatment with neuraminidase and β-galactosidase. C) TR-PSM containing only trisaccharide side chains, obtained after the treatments in B followed by α-N-acetylgalactosaminidase. Selective oligosaccharide side chains removal is shown by the anomeric carbon assignments in the 90-110 ppm region. Resonances from NeuNGl are labeled, N2-11, in A. Glycosylated (SOG & TOG) and nonglycosylated (SOH & TOH) Ser and Thr α-carbons resonances, respectively, are labeled in spectra B and C in the 52-60 ppm region. Note the change in relative areas of the glycosylated and nonglycosylated α-Ser and α-Thr resonances after removal of GalNAc in spectrum C. D) “Native GalNAc” tryptic tandem repeat glycopeptide obtained after trimming all glycans to GalNAc by mild TFMSA. Peptide GalNAc carbons are labeled C1-6, & NAc. Note the resonance intensities for the glycosylated and nonglycosylated Ser and Thr α-carbons in D are identical to those in A and B indicating no losses of peptide linked GalNAc after TFMSA treatment. E) “Core 1” tryptic tandem repeat glycopeptide obtained after removal of non C3 substituted GalNAc residues by periodate oxidation followed by mild TFMSA. This preparation contains peptide
GalNAc residues only at positions originally substituted by core 1 (di and tri) oligosaccharides. Note the decrease in the anomeric carbon resonances of Ser and Thr linked GalNAc (~98 ppm) and the changes in the Ser and Thr α-carbon resonances both indicating the partial deglycosylation of the mucin (i.e. decrease in TOG & SOG resonances and increase in TOH & SOH resonances).  F) “Trisaccharide” tryptic tandem repeat glycopeptide obtained after neuraminidase and β-galactosidase treatment (spectrum B) followed by periodate and mild TFMSA. This preparation contains peptide GalNAc residue only at positions originally substituted by the trisaccharide. Note the further loss of carbohydrate with respect to spectrum E and the similarity of the Ser and Thr α-carbon resonance patterns to the enzymatically deglycosylated derivative containing intact trisaccharide, side chains, spectrum C. G) Fully deglycosylated apo-PSM tandem repeat domain showing the complete loss of carbohydrate. The β-carbons of nonglycosylated Ser and Thr are labeled as are the α-carbons of Ala and Gly. All spectra were plotted with a nearly constant height for the Gly α-carbon resonance, 43 ppm. Sample sizes ranged from ~25 to ~3 mg.

Figure 3

Amino acid sequencing profiles of selectively deglycosylated tandem repeats from PSM preparation B10. A and D represent Ser and Thr profiles for “Native” mucin tandem repeats, B and E represent Ser and Thr profiles for “Core-1” tandem repeats. C and F represent Ser and Thr profiles for “Trisaccharide” tandem repeats. GalNAc glycosylated (black lines, open square) and nonglycosylated (grey line, closed diamond) Ser and Thr residues are indicated. Data were baseline corrected as previously described (34, 37).

Figure 4

Sequence specific mono, di and trisaccharide glycosylation patterns of the PSM tryptic tandem repeats from three A negative blood group animals and their average. Each residue’s mono, di, & trisaccharide substitution is indicated by the black, white and grey bars, respectively. The sum of the bars represent the total glycosylation of each residue by GalNAc. Panels A through D display the data in sequential order along the tandem repeat while panels E through H group the data separately by Ser and Thr. Panels A & E represent PSM 12, panels B & F, PSM B7, panels C & G, PSM B10, and panels D & H the average of all three. The Ser/Thr density function is plotted as the open squares connected by the
grey line (34). Data were derived from Table III. Where necessary values from Table III were adjusted to eliminate inconsistent (negative) values by increasing the “Native GalNAc” values or by lowering the “Trisaccharide” values given in Table III as necessary. Out of the 116 glycosylation values listed in Table III only 7 were adjusted by 1-3 percentage values while 4 were adjusted by 4 to 7 percentage values. These changes are well within the estimated errors of the determinations. No other values in Table III were modified.

Figure 5

PSM tandem repeat site-specific glycosylation patterns versus tandem repeat Ser/Thr density (34). Left panels, A-E, represent data for Ser linked glycans while the right panels, F-J, represent data for Thr linked glycans. The lines in each plot represent the linear least square fit to the data, omitting data for Ser2 (open symbols) for the reasons discussed in the text. Note that individual data points representing each of the three PSM preparations are plotted in each panel and that for only the Ser linked glycans are the trends with respect Ser/Thr density statistically significant (p<0.05). A & F, percent monosaccharide vs Ser/Thr density (Ser: \( r^2=0.28, p<0.0001 \), Thr: \( r^2=0.08 \) p=0.35). B & G; percent core 1 substitution (sum of di and trisaccharide) vs Ser/Thr density (Ser: \( r^2=0.51, p<0.0001 \), Thr: \( r^2=0.15 \) p=0.026). C & H, percent disaccharide vs Ser/Thr density (Ser: \( r^2=0.15, p=0.002 \), Thr: \( r^2=0.11 \) p=0.08). D & I, percent trisaccharide vs Ser/Thr density (Ser: \( r^2=0.47, p<0.0001 \), Thr: \( r^2=0.03 \) p=0.11). E & J; percent peptide linked GalNAc on Ser/Thr vs Ser/Thr density (Ser: \( r^2=0.32, p<0.0001 \), Thr: \( r^2=0.10 \) p=0.02). Note that \( r^2 \) and p were obtained omitting data for Ser2. Data points were derived from Table III (9 values out of 87 were adjusted as described in Figure 4 to eliminate inconsistencies).

Figure 6

Average normalized sequence specific glycosylation patterns of the PSM tryptic tandem repeat from a negative blood group animals. Panels A through D display the data in sequential order along the tandem repeat while panels E through H group the data separately by Ser and Thr. Panels A & E percent Gal on peptide GalNAc (i.e. percent core 1), percent disaccharide white bars, percent trisaccharide grey bars. Panels B & F percent Fuc on Gal. Panels C & G average side chain length normalized to Ser or Thr. Panels D & H average side chain length normalized to peptide linked GalNAc. The Ser/Thr density function is plotted as the open squares connected by the grey line (34). Data were obtained from the
average glycosylation values given in Table III (1 value out of 29 was adjusted as described in Figure 4 to eliminate inconsistencies).

**Figure 7**

PSM tandem repeat normalized site-specific glycosylation patterns versus tandem repeat Ser/Thr density (34). Left panels, A-E, represent data for Ser linked glycans while the right panels, F-J, represent data for Thr linked glycans. The lines in each plot represent the linear least square fit to the data, omitting data for Ser2 (open symbols) for the reasons discussed in the text. Individual data points representing each of the three PSM preparations are plotted in each panel. A & F, percent Gal (i.e. core 1 glycans) on peptide GalNAc vs Ser/Thr density (Ser: r²=0.49, p<0.0001, Thr: r²=0.09 p=0.19). B & G; percent Fuc on Gal vs Ser/Thr density (Ser: r²=0.04, p=0.055, Thr: r²=.003, p=0.59). C & H, hydroxyaminoacid normalized side chain length vs Ser/Thr density (Ser: r²=0.47, p<0.0001, Thr: r²=0.17, p=0.005). D & I, peptide linked GalNAc normalized side chain length vs Ser/Thr density (Ser: r²=0.46, p<0.0001, Thr: r²=0.06 p=0.31). E & J, ; in vivo glycosylation (OG) density vs Ser/Thr density (Ser: r²=0.69, p<0.0001, Thr: r²=0.86 p<0.0001). Excluding the percent Fuc on Gal vs Ser/Thr density plot (panel B) the trends with respect Ser/Thr density are statistically significant (p<0.05) for Ser linked glycans only. Note that r² and p were obtained omitting data for Ser2. Data points were derived from Table III (9 values out of 87 were adjusted as described in Figure 4 to eliminate inconsistencies).
A

\[
\begin{align*}
\alpha\text{-GalNAc-O-Ser/Thr} & \quad \text{Mono (Tn)} \\
\beta\text{-Gal (1-3)} & \\
\alpha\text{-GalNAc-O-Ser/Thr} & \quad \text{Di (T)} \\
\alpha\text{-NeuGlc (2-6)} & \\
\alpha\text{-Fuc (1-2) } \beta\text{-Gal (1-3)} & \\
\alpha\text{-GalNAc-O-Ser/Thr} & \quad \text{Tri (H)} \\
\alpha\text{-Fuc (1-2) } \beta\text{-Gal (1-3)} & \\
\alpha\text{-GalNAc (1-3)} & \quad \text{Tetra (A)}
\end{align*}
\]

B

ISVAGSSGP_{10}AVSSGASQAA_{20}GTSGAGPTT_{30}ASSVGVTETA_{40}
RPSVAGSGTT_{50}GTVSGASGST_{60}GSSSGSGA_{70}GASIGQPETS_{80}R

Fig 1
Fig 2
Fig 3

A

B

C

D

E

F

Ser

Thr
Fig 4
FIG 5

A

B

C

D

E

F

G

H

I

J

Ser

Thr

% MONO

% Core 1

% DI

% TRI

% GaINAc

S/T DENSITY

S/T DENSITY

S/T DENSITY

S/T DENSITY

S/T DENSITY

S/T DENSITY

S/T DENSITY
Fig 7
Determination of the site-specific oligosaccharide distribution of the O-glycans attached to the porcine submaxillary mucin tandem repeat: Further evidence for the modulation of O-glycan side chain structures by peptide sequence

Thomas A. Gerken, Marc Gilmore and Jiexin Zhang

J. Biol. Chem. published online January 2, 2002

Access the most updated version of this article at doi: 10.1074/jbc.M111690200

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 0 references, 0 of which can be accessed free at http://www.jbc.org/content/early/2002/01/02/jbc.M111690200.citation.full.html#ref-list-1