Role of Peptide Sequence and Neighboring Residue Glycosylation on the Substrate Specificity of the Uridine 5'-Diphosphate−α-*N*-acetylgalactosamine:Polypeptide *N*-acetylgalactosaminyl Transferases T1 and T2: Kinetic Modeling of the Porcine and Canine Submaxillary Gland Mucin Tandem Repeats[†]

Thomas A. Gerken,* Chhavy Tep, and Jason Rarick

*W. A. Bernbaum Center for Cystic Fibrosis Research, Departments of Pediatrics and Biochemistry,
Case Western Reserve University School of Medicine, Cleveland, Ohio 44106*

Received April 23, 2004; Revised Manuscript Received June 1, 2004

ABSTRACT: A large family of uridine 5′-diphosphate (UDP)−α-*N*-acetylgalactosamine (GalNAc): polypeptide *N*-acetylgalactosaminyl transferases (ppGalNAc Ts) initiates mucin-type *O*-glycan biosynthesis at serine and threonine. The peptide substrate specificities of individual family members are not well characterized or understood, leaving an inability to rationally predict or comprehend sites of O-glycosylation. Recently, a kinetic modeling approach demonstrated neighboring residue glycosylation as a major factor modulating the O-glycosylation of the porcine submaxillary gland mucin 81 residue tandem repeat by ppGalNAc T1 and T2 [Gerken et al. (2002) *J. Biol. Chem. 277*, 49850−49862]. To confirm the general applicability of this model and its parameters, the ppGalNAc T1 and T2 glycosylation kinetics of the 80+ residue tandem repeat from the canine submaxillary gland mucin was obtained and characterized. To reproduce the glycosylation patterns of both mucins (comprising 50+ serine/threonine residues), specific effects of neighboring peptide sequence, in addition to the previously described effects of neighboring residue glycosylation, were required of the model. Differences in specificity of the two transferases were defined by their sensitivities to neighboring proline and nonglycosylated hydroxyamino acid residues, from which a ppGalNAc T2 motif was identified. Importantly, the model can approximate the previously reported ppGalNAc T2 glycosylation kinetics of the IgA1 hinge domain peptide [Iwasaki, et al. (2003) *J. Biol. Chem. 278*, 5613−5621], further validating both the approach and the ppGalNAc T2 positional weighting parameters. The characterization of ppGalNAc transferase specificity by this approach may prove useful for the search for isoform-specific substrates, the creation of isoform-specific inhibitors, and the prediction of mucin-type O-glycosylation sites.

O-Glycosylated mucin-like domains serve important structural and biological roles in many secreted and membrane-associated glycoproteins. For example, biological process such as the protection of epithelial cell surfaces, cellular adhesion, cellular protein targeting, the immune and inflammatory responses, and the immune evasion of tumor cells are modulated by glycoproteins containing mucin-like domains, which are required for their biological properties (for example, refs *1*−*7*). These domains typically contain 20−30% hydroxyamino acids, serine and threonine, and are commonly composed of heavily O-glycosylated polypeptide tandem repeats. The mucin-type O-linked glycans, which typically range from one to more than 10 carbohydrate residues in length, are attached to serine or threonine via α-*N*-acetylgalactosamine (GalNAc) and impart an extended polypeptide conformation (*8*−*10*). The unique placement of specific mucin-type *O*-glycan structures in the peptide sequence in some instances is required for full biological activity (*2*, *11*−*16*) and perhaps even for development (*17*, *18*). Recently we have demonstrated that the extent of substitution and distribution of *O*-glycan structures on the porcine submaxillary gland mucin (PSM)[1] 81 residue tandem repeat varies in a reproducible manner along the peptide sequence (*19*). Together, these findings suggest that the glycosyltransferases involved in the initial steps of mucin-type *O*-glycan biosynthesis are uniquely sensitive to features of the peptide sequence that are presently not fully understood.

In the Golgi, the transfer of GalNAc to the peptide core is performed by a family of uridine 5′-diphosphate (UDP)−GalNAc:polypeptide *N*-acetylgalactosaminyl transferases

[†] Supported by the National Institutes of Health, National Cancer Institute, Grant RO1-CA-78834, and by the Cystic Fibrosis Foundation.

* Corresponding Author. Mailing address: Department of Pediatrics, Case Western Reserve University, School of Medicine, BRB, Cleveland, OH 44106-4948. Telephone: 216-368-4556. Fax: 216-368-4223. E-mail: txg2@cwru.edu.

[1] Abbreviations: ppGalNAc T, UDP−GalNAc:polypeptide α-GalNAc transferase; CSM, canine submaxillary gland mucin; PSM, porcine submaxillary gland mucin; r^2, least-squares correlation coefficient; SD, standard deviation; W_{OH_n} and W_{OG_n}, positional weighting coefficients for the presence of a nonglycosylated or glycosylated serine/threonine residue at position *n* (see eqs 2 and 3 of ref *46*); *f*(OG+OH), glycosylation state rate constant multiplier function (see eq 4 of ref *46*); *g*(Pro,Glu,Arg), peptide sequence rate constant multiplier function defined in Experimental Procedures and eq 2; *F*(Xaa)$_n$, residue specific positional rate factor defined in Experimental Procedures and eq 3.

10.1021/bi049178e CCC: $27.50 © 2004 American Chemical Society
Published on Web 07/08/2004

(ppGalNAc Ts) (*20*, *21*). This step represents the first committed step of mucin-type *O*-glycan biosynthesis and serves to identify the glycan attachment site and may further serve to modulate subsequent glycosylation events. Presently there are 14 members described in the mammalian ppGalNAc transferase family, ppGalNAc T1−T14 (*17*, *18*, *22*−*32*). Family member homologues have also been found in *Drosophila* and *Caenorhabditis elegans* (*17*, *18*, *33*, *34*), suggesting that there may be evolutionarily conserved roles/substrates for several of these transferases. The peptide and glycopeptide substrate specificity of the ppGalNAc T family members are not well characterized or understood. The most characterized members, ppGalNAc T1−T4, reveal different, as well as overlapping, peptide substrate preferences (*17*, *23*, *35*−*39*) and poorly understood sensitivities to peptide substrate glycosylation (*6*, *40*−*44*). Moreover, ppGalNAc T7 and T10 have an apparent absolute requirement for prior GalNAc addition for activity (*26*, *29*, *45*).

Using a kinetic modeling approach, we have previously demonstrated that neighboring residue glycosylation is an important factor in modulating the site-specific glycosylation of nearly all of the 31 serine and threonine residues in the PSM tandem repeat by ppGalNAc T1 and T2 (*46*). In this relatively simple kinetic model, first-order serine and threonine glycosylation rate constants were proportionally decremented as a function of neighboring glycosylation state. Using the model, we could reasonably reproduce the experimental ppGalNAc T1 and T2 glycosylation patterns of the PSM tandem repeat, each with a unique set of positional weighting parameters (*46*). By a similar kinetic modeling approach, we have recently shown that the in vivo substitution of the GalNAc residue by *β*-galactose (*β*-Gal), forming the so-called Core 1 structure (*β*-Gal(1−3)-α-GalNAc-Ser/Thr), may also be modulated by neighboring glycosylation effects in PSM (*47*). Although the possibility was discussed (*46*), the effect of peptide sequence on the modeling of the ppGalNAc transferases was not originally incorporated in the model since neighboring glycosylation effects could adequately reproduce the experimental glycosylation time course of PSM. Since the ppGalNAc T1 and T2 model fitting parameters were based only on the glycosylation of the PSM tandem repeat, it is important to perform further in vitro glycosylation and modeling studies on additional mucin tandem repeats for confirmation of the overall assumptions of the model and for further refinement of the positional weighting parameters of each transferase.

In this work, a newly described 80+ residue tandem repeat from the canine submaxillary gland mucin (CSM) was isolated, and its peptide sequence and in vivo glycosylation pattern were determined by Edman amino acid sequencing. The fully deglycosylated apo-CSM tandem repeat domain was used as a substrate for ppGalNAc T1 and T2, and the site-specific glycosylation time course for each transferase was analyzed by the kinetic modeling approach.

EXPERIMENTAL PROCEDURES

ppGalNAc Transferases and Porcine Mucin Glycosylation. Soluble recombinant bovine ppGalNAc T1 and human ppGalNAc T2 (see ref *46*) were a kind gift of Dr. Ake Elhammer (Kalamazoo, MI). The glycosylation kinetics of the porcine salivary gland mucin (PSM) tandem repeat by ppGalNAc T1 and T2 has been previously reported (*46*).

Canine Mucin Isolation and Characterization. Frozen mixed-breed canine submaxillary glands were obtained from Pel-Freez (Rodgers, AR). Canine salivary mucin (CSM) was isolated from pooled glands and reduced and carboxymethylated by the procedures described for the isolation and carboxymethylation of the porcine salivary gland mucin (PSM) reported earlier (*48*). Unlike PSM, canine mucin with its intact glycans is highly susceptible to trypsin; therefore, no pretreatment of CSM with trypsin to remove the N- and C-terminal globular domains was performed. The O-linked carbohydrate side chains on CSM were quantitatively trimmed to the peptide-linked GalNAc residue by mild trifluoromethanesulfonic acid (TFMSA) treatment (*48*, *49*) giving so-called T(−t)R-CSM. This material tends to be partially insoluble in water; hence, only the soluble portion was characterized further.

CSM tandem repeats of ∼86+ residues were obtained after a 2 h digestion of soluble T(−t)R-CSM with 0.01% protease Glu-C (50 mM ammonium bicarbonate, pH 7.8) and isolated on Sephacryl S200 chromatography after the addition of protease inhibitor *N*-α-*p*-tosyl-L-lysine chloromethyl keytone (TLCK). The ∼40 residue C-terminal portion of tandem repeat (peptides B and C, see below) was obtained from the N-terminal biotinylated tandem repeat after overnight 1% protease Glu-C digestion as described previously for PSM (*19*, *46*, *50*). The avidin affinity column step was occasionally omitted because it was found that the biotinylation modification completely blocked the Edman sequencing of the modified peptide. Sephacryl S200 chromatography was performed in 50 mM acetic acid buffer, pH 4.0 (NH_4OH), to eliminate protease contamination (*46*). Fully deglycosylated CSM was obtained from T(−t)R-CSM after treatment with α-*N*-acetylgalactosaminidase (chicken liver enzyme, Sigma; sodium citrate/phosphate, pH 3.7, 43 °C) in the presence of protease inhibitors (Sigma P8304 and P8849) as previously described (*19*). We found for CSM that the oxidation−elimination approach for removing peptide-linked GalNAc (*49*) resulted in greater peptide core degradation than the enzymatic approach. Each deglycosylation (and reglycosylation) modification was assessed by 150 MHz carbon 13 NMR spectroscopy for the quality and completeness of the modification as previously described (*19*, *46*). Only completely deglycosylated mucins were used for transferase substrate.

Pulsed liquid Edman amino acid sequencing of apo-CSM and T(−t)R-CSM was performed on an Applied Biosystems Procise 494 Edman protein sequencer (Foster City, CA) as previously described (*19*, *46*, *48*, *50*). Improved separation of the closely eluting phenylthiohydantoin (PTH)−Ser-OH and PTH−Thr-O−GalNAc peaks was obtained by lowering the C18 PTH column temperature by 5−10 °C. Glycosylation site data analysis was performed as previously described (*19*, *48*, *50*). The site-specific glycosylation of overlapping residues in the mixed C terminal CSM peptides (peptides B and C, see below) was obtained by a mathematical deconvolution approach.

ppGalNAc T1 and T2 Reglycosylation of the apo-CSM Tandem Repeat. High molecular weight apo-CSM was reglycosylated by the procedures previously described for apo-PSM (*46*). Briefly, reactions were performed in 0.5−1.5 mL volumes containing 5 mg/mL apo-mucin, 10 mM $MnCl_2$, 0.1 mM EDTA, 100 mM HEPES, pH 7.5, 5 mM

UDP−GalNAc, and 22 μg/mL ppGalNAc transferase T1 or T2 in the presence of protease inhibitor cocktails (Sigma P8304 and P8849). Reaction mixtures were allowed to incubate for 4−24 h at 37 °C and subsequently transferred to dialysis membranes for dialysis at 4 °C against 2−3 changes of 500 mL of reaction buffer lacking UDP−GalNAc to remove free UDP. GalNAc transfer was reinitiated at 37 °C by re-adding both protease inhibitors and UDP−GalNAc (to 5 mM). This was repeated until the desired net reaction time was reached (up to ∼10 days for ppGalNAc T2). Fresh ppGalNAc T2 (∼11 μg/mL) was added at every 2−3 days (approximate half-live 2.5 days (*46*)) while for ppGalNAc T1, no additional additions were made (approximate half-life 5 days (*46*)). Each incubation time point was repeated 2−4 times. The net incubation times reported have been adjusted to roughly account for dilution effects and the loss of transferase activity with extended incubation time. After purification on S200 chromatography, the high molecular weight reglycosylated CSM was digested with protease Glu C to release the tandem repeat for Edman sequencing, as described above and in the Results section. Carbon-13 NMR spectra were obtained after each incubation to confirm the transfer of GalNAc to the mucin as is shown in Figure S1 in the Supporting Information for ppGalNAc T2.

Inclusion of Peptide Sequence Effects into Kinetic Model. A kinetic model incorporating the inhibitory effects of neighboring glycosylated and nonglycosylated hydroxyamino acid residues has been previously described (*46*). In this model, the first-order rate constant for serine and threonine glycosylation, k_{Ser} or k_{Thr}, is multiplied by the value of the $f(\mathrm{OG}+\mathrm{OH})_i$ function, which is a measure of local glycosylation status (plus and minus three residues of the site of glycosylation), incorporating positional weighting coefficients W_{OG_n} and W_{OH_n} the values of which range from 0 to 1 (*46*). To incorporate additional sequence-specific effects into the model, an additional function, $g(\mathrm{Pro,Glu,Arg})_i$, has been devised to incorporate both stimulatory and inhibitory effects of neighboring proline and charged residues (glutamic acid/aspartic acid and arginine/lysine/histidine). This gives the modified first-order rate equation

$$\mathrm{d[OG]}_i/\mathrm{d}t = k_{(\mathrm{Ser\ or\ Thr})} f(\mathrm{OG}+\mathrm{OH})_i\, g(\mathrm{Pro,Glu,Arg})_i [\mathrm{OH}]_i \quad (1)$$

where the $g(\mathrm{Pro,Glu,Arg})_i$ function is defined in eq 2 as

$$g(\mathrm{Pro,Glu,Arg})_i = (1 + (\mathrm{Pro})_i)(1 + (\mathrm{Glu})_i)(1 + (\mathrm{Arg})_i) \quad (2)$$

In eq 2, $(\mathrm{Xaa})_i$ (where Xaa = Pro, Glu, Arg) represents the sum of the individual residue-specific positional rate factors, $F(\mathrm{Xaa})_n$, over plus and minus three residues of the site of glycosylation as defined by eq 3.

$$(\mathrm{Xaa})_i = F(\mathrm{Xaa})_{(i-3)}[\mathrm{Xaa}]_{(i-3)} + F(\mathrm{Xaa})_{(i-2)}[\mathrm{Xaa}]_{(i-2)} + F(\mathrm{Xaa})_{(i-1)}[\mathrm{Xaa}]_{(i-1)} + F(\mathrm{Xaa})_{(i+1)}[\mathrm{Xaa}]_{(i+1)} + F(\mathrm{Xaa})_{(i+2)}[\mathrm{Xaa}]_{(i+2)} + F(\mathrm{Xaa})_{(i+3)}[\mathrm{Xaa}]_{(i+3)} \quad (3)$$

In eq 3, values of the positional rate factors, $F(\mathrm{Xaa})_n$, can range from −1 to any positive number, while $[\mathrm{Xaa}]_n$ values are either 0 or 1, representing the absence or presence of the Xaa residue at position *n*. In this formalism, values of $F(\mathrm{Xaa})_n$ of zero will result in no change in the rate constant, while negative values will decrease the rate, and positive values will increase the rate. Note, however, that in this formalism the sum of the positional F(Xaa) factors for any Xaa residue type may not exceed −1.

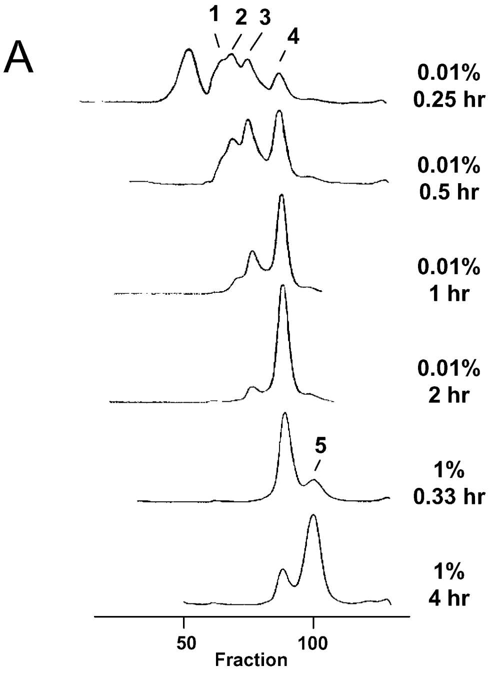
As a final modification of the model, we have decided to make the $f(\mathrm{OG}+\mathrm{OH})_i$ function for serine glycosylation identical to that for threonine glycosylation (eq 4 of ref *46*). In the previous simulation, the rate constant for serine residues was more heavily decremented than threonine residues with increasing glycosylation density by this function. With the present modification, serine residues will be more realistically treated, having their rate constants decremented proportional to increasing glycosylation density as shown by the plot of this function given in Supporting Figure S2.

Numerical simulations were performed as described previously (e.g., eq 4 of ref *46*) using Lotus 123 spread sheet software, release 9.7 (Lotus Development Co. Cambridge, MA). Simulations were performed utilizing 360 incremental time points and were optimized by manually incrementing the above positional parameters and serine/threonine rate constants until the best fit (by least-squares correlation coefficient, r^2, and standard deviation) to the experimental glycosylation values was obtained.

Extraction of Site-Specific Glycosylation from Published IgA1 Glycosylation Data. The pathway and time course of the glycosylation of a 20 residue peptide representing the IgA1 hinge domain by ppGalNAc T2 has recently been reported by Iwasaki and co-workers (*51*). Using their structural assignments (ref *51*, Figure 6) and the respective peak heights in the reverse-phase HPLC elution profiles (ref *51*, Figure 3A), we could construct the glycosylation time course of each serine and threonine in the peptide as shown in Figure S8 in the Supporting Information.

RESULTS

Isolation, Sequence Determination and in Vivo GalNAc Glycosylation Pattern of the CSM Tandem Repeat Domain. Mild TFMSA-treated CSM (containing only mono-GalNAc side chains) was incubated with a series of proteases to test for the presence of tandem repeats. Trypsin and protease Arg-C rapidly produced small peptide fragments with multiple N-terminal sequences, while protease Lys-C was inactive (data not shown). A time course performed with 0.01% and 1% protease Glu-C reveals the presence of five unique molecular weight species on Sephacryl S200 over the course of digestion (Figure 1A), peaks 4 and 5 being the major products of the 0.01% and 1% protease Glu-C digestions, respectively. Edman amino acid sequencing of peaks 1−4 revealed the identical N-terminal sequence, while peak 5 gave a mixture of three sequences, one of which was identical to the sequence obtained for peaks 1−4. Further sequencing of peak 4 revealed the sequence of an 80+ residue peptide (peptide A in Figure 1B), which also contained the sequence of the peptides found in peak 5 (peptides B and C, Figure 1B, representing Glu-C cleavage of peptide A at Glu39 and Glu48, respectively). The Edman sequencing of a second Glu-C digest of the N-terminal biotinylated (blocked) peak 4 (peptide A) gives only the sequence of peptides B and C (*19*, *46*, *50*). Upon exhaustive



B

Peptide A:

AGSVGRTAGG$_{10}$PGFTSPGRVA$_{20}$
GGTGSPTASA$_{30}$RGGTPGGSEG$_{40}$
FTAAPGSEST$_{50}$GGHSGAPGTT$_{60}$
LAGRAGTTLG$_{70}$PRSEPSGTGV$_{80}$
GGSPVATTL

Peptide B:

G$_{40}$FTAAPGSEST$_{50}$GGHSG....

Peptide C:

ST$_{50}$GGHSGAPGTT$_{60}$LAGR....

FIGURE 1: Isolation and peptide sequence of the canine submaxillary gland mucin (CSM) protease Glu-C tandem repeat glycopeptides. Panel A shows sequential digestion of partially deglycosylated (mild TFMSA-treated) CSM by protease Glu-C (0.01% or 0.1%) fractionated on Sephacryl S200 gel filtration chromatography and monitored by absorbance at 220 nm. The first eluting unlabeled peak in the top chromatograph represents undigested mucin, while the numbered peaks represent CSM tandem repeat glycopeptides. Panel B shows the amino acid sequences of the CSM tandem repeat glysopeptides. Peptide A represents the sequence obtained from peak 4, while peptides B and C represent the N-terminal sequences obtained from peak 5, which represent the further cleavage of peptide A at Glu39 and Glu48.

Glu-C digestion only peptide C is found, at the position of peak 5, on S200, while the nine residue peptide remaining from the cleavage of peptide B is found near the included volume of the column (data not shown). Peak 4, therefore, represents the monomeric CSM tandem repeat, while peaks 1−3 represent multiples of the tandem repeat. Identical protease Glu-C sensitivity, S200 chromatographic behavior, and peptide sequence were found for fully deglycosylated apo-CSM and its partially reglycosylated derivatives (data not shown).

An analysis of the Edman sequencing data indicates that the CSM tandem repeat is relatively nondegenerate, similar to that of PSM (*19*, *48*, *50*). Network protein sequence analysis (*52*) consensus secondary structure predictions reveal that the sequence should be entirely random coil except for three short (one to two residue) regions with predicted extended structures at Val18, Phe40, and Thr59. The amino acid composition of the repeat is typical of mucins with glycine, serine, threonine, alanine, and proline accounting for 80% of its residues (31% glycine, 14% threonine, 13% serine, 12% alanine, 10% proline, 6% arginine, 5% valine, 3% glutamic acid, 2% each phenylalanine and leucine, and 1% histidine). Ten percent of its residues are charged; the low arginine and glutamic acid content and lack of lysine readily account for the sensitivity of CSM to trypsin and Glu-C and the resistance of CSM to protease Lys-C digestion. The composition of the tandem repeat is very similar to the reported amino acid analysis of native CSM (*53*). The derived partial CSM tandem repeat sequence has been deposited to the Swiss-Prot protein database, accession number P83762. MULTALIN (*52*, *54*) and CLUSTALW (*52*, *55*) sequence alignments of the CSM tandem repeat with the PSM tandem repeat (see Supporting Table S1) reveal 25−30 identical residues in each alignment; however, the longest contiguous sequence of identical residues does not exceed three residues. Of the 25 hydroxyamino acid residues in CSM, 11−15 can be aligned with those in PSM, but there are no sequence homologies found adjacent to these residues. The CSM and PSM tandem repeats, although similar in length, share no significant sequence homologies.

The in vivo site-specific glycosylation of 23 of the serine and threonine residues in the CSM tandem repeat obtained by quantitative Edman sequencing is given in Table S2 in the Supporting Information. The in vivo glycosylation ranges from ∼10% to ∼80% with an average serine/threonine residue glycosylation of ∼60%. Similar to the in vivo glycosylation of PSM, serine residues are more poorly glycosylated on average (average of 52%) and display a wider range of glycosylation (∼10−80%) compared to threonine residues, which are more uniformly glycosylated (average of 67%, range of ∼60−75%). These data along with the ppGalNAc transferase data (see below) are plotted in sequential order in Figure 2 (light gray bars) and grouped by serine/threonine residue type in Figure S3 of the Supporting Information. Note that regions of low glycosylation do not overlap with the predicted short regions of extended secondary structure.

ppGalNAc T1 and T2 Reglycosylation of the apo-CSM Tandem Repeat Domain. The relative rates of site-specific glycosylation by ppGalNAc T1 and T2 were obtained from the quantitative Edman sequence analysis of the CSM Glu-C tandem repeat as described in the Experimental Procedures. Figures 2 and S3 show that ppGalNAc T1 is capable of glycosylating nearly all of the serine/threonine residues of the CSM repeat resulting in an average glycosylation of 51% at the longest incubation time (∼24 h). As previously found for PSM (*46*), serine residues are glycosylated less rapidly than threonine residues and display a greater variability in their extent of glycosylation. At the longest ppGalNAc T1 incubation time, threonine residues are 63% glycosylated on average (range 30−75%) while serine residues are 39% glycosylated on average (range 8−66%). The similarity of the native in vivo and the ppGalNAc T1 in vitro glycosylation patterns shown in Figures 2 and S3 suggests that ppGalNAc T1 is capable of partially reproducing the in vivo glycosylation of CSM.

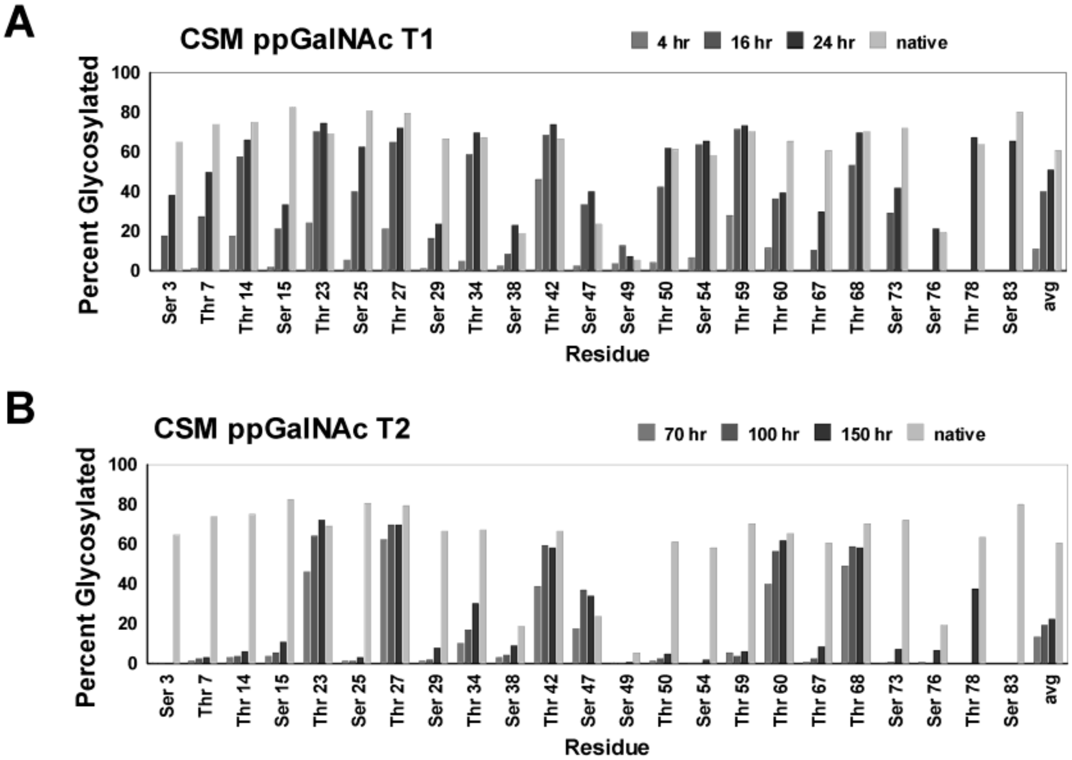


FIGURE 2: In vitro glycosylation of the apo-CSM tandem repeat by ppGalNAc T1 and T2: (A) ppGalNAc T1 glycosylation pattern; (B) ppGalNAc T2 glycosylation pattern. In each panel, the clustered dark gray bars, increasing from left to right, represent each residue's glycosylation for the approximate net incubation times of 4, 16, and 24 h for ppGalNAc T1 (A) and 70, 100, and 150 h for ppGalNAc T2 (B). The right most light gray bar in each cluster represents the observed in vivo glycosylation of the residue. Note for Thr68, Ser73, Thr78, Thr88, and Ser83 that glycosylation data may be missing at one or more time points. See Figure S3 in the Supporting Information for the identical plots grouped by serine and threonine. Numerical data are tabulated in Table S2 of the Supporting Information.

In contrast to ppGalNAc T1, ppGalNAc T2 glycosylates only about one-third of the serine/threonine residues of the CSM tandem repeat (Figures 2 and S3). Only seven residues, nearly all threonine, show substantial glycosylation, while five of the 12 threonine and nine of the 11 serine residues appear to be nearly refractory to glycosylation by ppGalNAc T2. The near-exclusive glycosylation of threonine residues by ppGalNAc T2 was confirmed by ^{13}C NMR (Figure S1). At the longest incubation time point (∼150 h), the average glycosylation of CSM by ppGalNAc T2 is 23%. At this time point, the average threonine residue glycosylation is 35% (range of 3−73%), while the average serine residue glycosylation is 8% (range 0−35%). Except for residues that are poorly glycosylated in vivo, the ppGalNAc T2 glycosylation pattern fails to reproduce the native in vivo glycosylation pattern (Figure 2 and S3). Interestingly, ppGalNAc T2 is capable of more highly glycosylating Thr60 compared to ppGalNAc T1. Overall, the glycosylation behavior of ppGalNAc T1 and T2 toward the CSM tandem repeat mirrors their activities previously reported for the PSM tandem repeat (*46*).

Kinetic Modeling of PSM and CSM by ppGalNAc T1. The glycosylation of the PSM tandem repeat by ppGalNAc T1 (and T2) has recently been simulated using a kinetic model that accounts for the inhibitory effects of neighboring residue glycosylation (*46*). Applying the positional weighting parameters (W_{OG_n} and W_{OH_n}) derived from the PSM work to the CSM ppGalNAc T1 data results in a moderately successful simulation of the data (see Figure S4 in the Supporting Information). However, the quality of the simulation, based on standard deviation and correlation coefficient, is not as significant as observed that for the PSM simulation (*46*).

Recognizing that the simulation does not include direct peptide sequence effects, an improved version of the model was developed (described in the Experimental Procedures) that includes the effects of neighboring proline and charged (i.e., glutamic acid/aspartic acid and arginine/lysine/histidine) residues. These residues were chosen since several studies have demonstrated that neighboring proline and charged residues can significantly alter transferase activity both in vivo and in vitro (for example, refs *56*−*60*). In this version of the model, the rate constant is multiplied by the additional g(Pro,Glu,Arg) function (described in the Experimental Procedures) the value of which is derived from the residue-specific positional weighing factors $F(\mathrm{Xaa})_n$ [where Xaa = Pro, Glu (and Asp), and Arg (and Lys,His) and where n represents the position plus and minus three residues of the site of glycosylation]. Values of $F(\mathrm{Xaa})_n$ can range from −1 to any positive number; where values of 0 represent no effect, negative values represent inhibitory effects, and positive values represent stimulatory effects. The distribution of serine and threonine, proline, glutamic acid, and arginine residues relative to serine and threonine in CSM, PSM, and the combined CSM/PSM tandem repeats are given in Supporting Table S3. In the combined CSM/PSM data set, the hydroxyamino acids, proline, and arginine, and to a lesser extent glutamic acid, appear to be randomly distributed relative to serine and threonine (Table S3).

After a series of manual optimizations on each tandem repeat and subsequently on both repeats together, a set of positional weighting parameters was obtained capable of

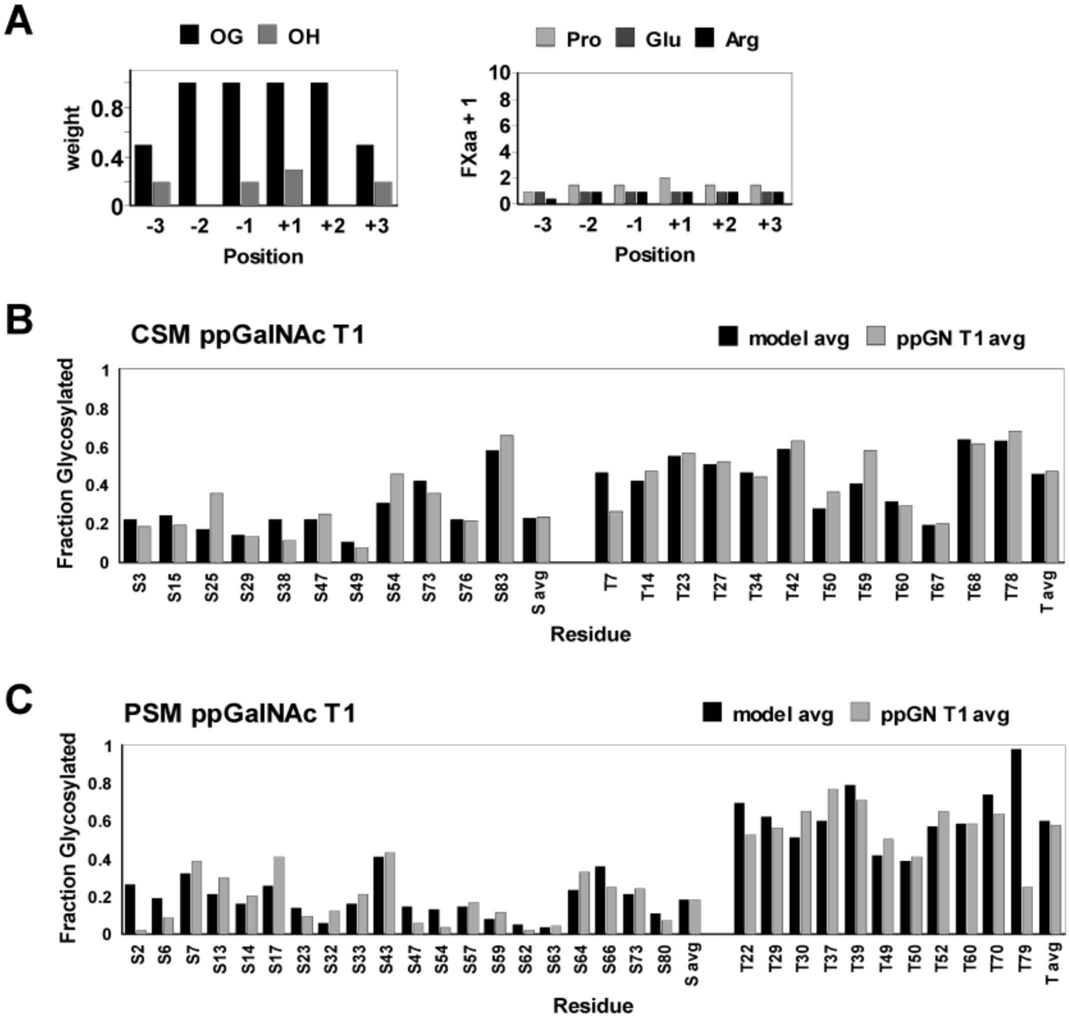


FIGURE 3: Model fitting of the in vitro glycosylation of the CSM and PSM tandem repeats by ppGalNAc T1. In part A, the left panel gives the values for the positional weighting parameters W_{OG_n} and W_{OH_n} (black and gray bars, respectively) for the presence and absence, respectively, of a glycosylated serine or threonine residue. The right panel gives the values for the residue specific positional weighting factors ($F(\mathrm{Xaa})_n$ +1), where $F(\mathrm{Xaa})_n = F(\mathrm{Pro})_n$, $F(\mathrm{Glu})_n$, $F(\mathrm{Arg})_n$ (light-gray, gray, and dark-gray bars respectively). Part B presents a comparison of the averaged calculated (black bars) and experimental (gray bars) ppGalNAc T1 glycosylation of the CSM tandem repeat ($k_{\mathrm{Thr}} = 0.048$, $k_{\mathrm{Ser}} = 0.018$ mole fraction/h). Part C presents a comparison of the averaged calculated (black bars) and experimental (gray bars) ppGalNAc T1 glycosylation of the PSM tandem repeat ($k_{\mathrm{Thr}} = 0.10$, $k_{\mathrm{Ser}} = 0.014$ mole fraction/h) (experimental data from ref *46*). Note that the averages do not include time points for which no experimental data were available. See also Figure S5 in the Supporting Information for a full display of the individual time point glycosylation data.

reproducing the overall ppGalNAc T1 glycosylation patterns of both CSM and PSM (see Figures 3 and 4). A significant improvement in the fit for CSM with little change in the fit for PSM was obtained (see legend to Figure 4). Additional plots further visualizing the ability of the model to fit the experimental data are given in Figure S5 found in the Supporting Information. To compensate for potential differences in transferase activities and reaction conditions between the PSM and CSM experiments, the serine and threonine intrinsic rate constants were optimized separately for each mucin.

The newly optimized CSM/PSM ppGalNAc T1 glycosylation state sensitive parameters, W_{OG_n} and W_{OH_n}, were found to be very similar to those obtained for the original PSM model (*46*) (compare Figure 3A and Supporting Figure S4A). Inhibitory effects of neighboring residue glycosylation ($W_{\mathrm{OG}n}$ values) for the combined PSM/CSM model remained unchanged, while inhibitory effects of neighboring nonglycosylated hydroxyamino acid residues (W_{OH_n} values) were slightly elevated compared to the original PSM work. In the new model, the presence of neighboring proline residues gave rate enhancements of 1.5−2-fold for all sites, except at the −3 position, ($F(\mathrm{Pro})_n$ +1 values). Charged residues, in contrast, appear to have little effect on glycosylation rates by ppGalNAc T1, except that an arginine at the −3 position may reduce the rate constant by half ($F(\mathrm{Glu})_n$ +1 and $F(\mathrm{Arg})_n$ +1 values). Since there are few charged residues in CSM and PSM, this conclusion may be revised with additional study. Analysis of the importance of the various parameters on the individual CSM and PSM models indicates that CSM is most sensitive to the rate enhancements of neighboring proline residues, while PSM is most sensitive to the inhibition effects of neighboring glycosylated residues. These differences in sensitivites reflect differences in the relative distributions of proline and hydroxyamino acid residues in the two mucin sequences; that is, CSM has a higher neighboring proline residue content, while PSM has a higher neighboring hydroxyamino acid residue content (see Table S3).

We conclude from the present modeling parameters for ppGalNAc T1 that rate reductions due to neighboring group glycosylation and rate enhancements due to the presence of

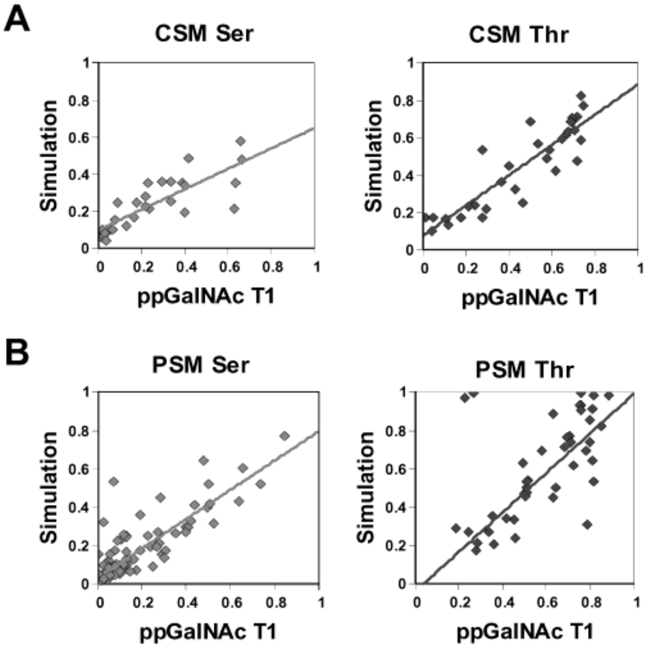


FIGURE 4: Plots of the simulated versus experimental ppGalNAc T1 site specific glycosylation for the CSM and PSM tandem repeats (individual time point data): (A, left) serine residues in CSM (SD = 0.10 mole fraction, $r^2 = 0.67$) and (right) threonine residues in CSM (SD = 0.11 mole fraction, $r^2 = 0.80$) using $k_{\mathrm{Thr}} = 0.048$ and $k_{\mathrm{Ser}} = 0.018$ mole fraction/h; (B, left) serine residues in PSM (SD = 0.09 mole fraction, $r^2 = 0.80$) and (right) threonine residues in PSM (SD = 0.14 mole fraction, $r^2 = 0.66$) using $k_{\mathrm{Thr}} = 0.10$ and $k_{\mathrm{Ser}} = 0.014$ mole fraction/h. Note that the statistical analysis for PSM omitted values for Ser2 and Thr79 as described previously (*46*).

proline appear to be the dominant factors modulating its substrate activity. It should, however, be noted that the ratio of the ppGalNAc T1 threonine and serine rate constants independently optimized for CSM and PSM were nearly 2.5-fold different for the two mucins (2.7 versus 7.1, respectively).[2] This difference may in part be due to differences in the current CSM and original PSM ppGalNAc T1 reglycosylation procedures. The original PSM ppGalNAc T1 work was performed using different dialysis procedures and prior to our routine use of the protease inhibitor cocktails (*46*). Since ppGalNAc T1 shows a relatively sharp pH maximum between pH 7.5 and 8 (*61*), it is possible that small changes in pH may have differentially altered the serine and threonine intrinsic rate constants.

Kinetic Modeling of PSM and CSM by ppGalNAc T2. The ppGalNAc T2 positional weighting parameters derived from the original PSM model (*46*) were unable to fully reproduce the observed ppGalANc T2 glycosylation pattern of CSM (see Supporting Figure S6). However, when both the PSM and CSM ppGalNAc T2 data were manually fit to the expanded model, described in the Experimental Procedures, a significantly improved fit was obtained, capable of reasonably reproducing the glycosylation patterns of both CSM and PSM (Figures 5 and 6 and Supporting Figure S7). Note that the statistically poor fit for the serine residues in CSM (Figure 6A) arises from the overall very low glycosylation of serine by ppGalNAc T2 and from the underprediction of the glycosylation of Ser47 and the overprediction of Ser76. These latter discrepancies may further reflect errors in the experimental data since both residues are at the C-terminus of the sequenced peptides.

The optimized CSM/PSM ppGalNAc T2 glycosylation state sensitive W_{OG_n} values (Figure 5A) are identical to the originally obtained PSM values (*46*)(see Supporting Figure S6A), while the nonglycosylated threonine and serine residue sensitive W_{OH_n} values are somewhat reduced, although the high inhibitory effect for a nonglycosylated serine or threonine at the +1 position is maintained. As discussed previously (*46*) the high +1 positional weighting, in effect, adds a peptide sequence dependence to the model and is required for the model to correctly reproduce the ppGalNAc T2 glycosylation patterns of both PSM and CSM. In contrast to ppGalNAc T1, ppGalNAc T2 shows larger and more specific proline residue rate enhancements at positions −3, −1, and +3 (Figure 5A, $F(\mathrm{Pro})_n$ +1 values of 2.5, 10, and 3.5, respectively). As discussed below, the very large proline enhancement for the −1 position is consistent with previous peptide substrate studies with ppGalNAc T2 and suggests that a preceding proline residue may represent a key ppGalNAc T2 motif. Similar to ppGalNAc T1, neighboring charged residues appear to have little effect on the rates of ppGalNAc T2 glycosylation, except for the presence of glutamic acid at the −1 position, which reduces the rate by 0.5 (Figure 5A). The ratios of the ppGalNAc T2 threonine and serine rate constants, independently optimized for CSM and PSM, are similar within experimental error (5.5 and 3.8, respectively).[3]

An analysis of the relative importance of the various ppGalNAc T2 weighting parameters reveals that the specificity of this transferase is dominated by the specific rate enhancements of neighboring proline residues and the specific inhibitory effects of an adjacent nonglycosylated hydroxyamino acid residue. As expected from their different amino acid distributions, the simulation on CSM shows a larger dependence on the proline parameters as compared to PSM (see Table S3). Interestingly, the inhibitory effects of neighboring glycosylated residues, although maximized in the model, nevertheless are found to play relatively minor roles in the CSM/PSM ppGalNAc T2 simulations compared to the more pronounced effects of neighboring proline and nonglycosylated hydroxyamino acid residues.

Kinetic Modeling of IgA1 Hinge Domain Peptide by ppGalNAc T2. The human IgA1 hinge domain contains two eight-residue mucin-like tandem repeats, which are partially O-glycosylated in vivo (*62*). Recently, Iwasaki and coworkers (*51*) reported the time course and pathway for the in vitro glycosylation by ppGalNAc T2 of a 20 residue peptide containing nine potential glycosylation sites representing the human IgA1 hinge domain (see Figure 7A). Using

[2] If the reglycosylation conditions (pH, buffers, etc.) were identical for each tandem repeat, the ratio of the serine and threonine rate constants would be expected to be nearly the same for both tandem repeats irrespective of moderate differences in substrate and transferase concentration.

[3] These values are essentially identical keeping in mind that the CSM serine rate constant could easily be increased with no significant worsening of the fit. Note also that the CSM and PSM ppGalNAc T2 experiments were performed using nearly identical procedures and reaction conditions, including the use of protease inhibitor cocktails. ppGalNAc T2 may also be less sensitive than ppGalNAc T1 to differences in reaction conditions as shown by the IgA1 analysis that follows.

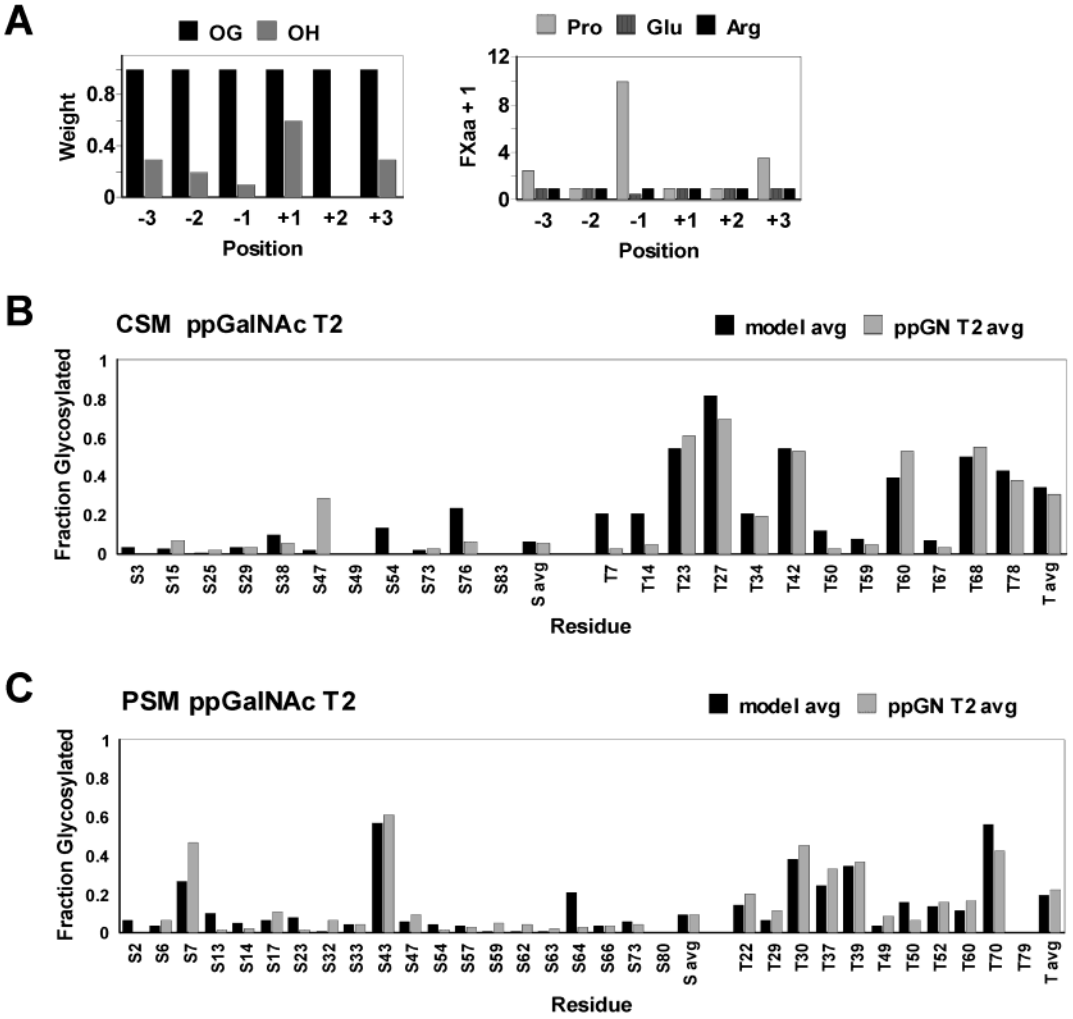


FIGURE 5: Model fitting of the in vitro glycosylation of the CSM and PSM tandem repeats by ppGalNAc T2. In panel A, the left panel gives values for the positional weighting parameters W_{OG_n} and W_{OH_n} (black and gray bars, respectively) for the presence and absence, respectively, of a glycosylated serine or threonine residue. The right panel gives values for the residue specific positional weighting factors ($F(\mathrm{Xaa})_n$ +1), where $F(\mathrm{Xaa})_n = F(\mathrm{Pro})_n$, $F(\mathrm{Glu})_n$, $F(\mathrm{Arg})_n$ (light-gray, gray and dark-gray bars, respectively). Panel B presents a comparison of the averaged calculated (black bars) and experimental (gray bars) ppGalNAc T2 glycosylation of the CSM tandem repeat ($k_{\mathrm{Thr}} = 0.0022$, $k_{\mathrm{Ser}} = 0.0004$ mole fraction/h). Panel C presents a comparison of the averaged calculated (black bars) and experimental (gray bars) ppGalNAc T2 glycosylation of the PSM tandem repeat ($k_{\mathrm{Thr}} = 0.0034$, $k_{\mathrm{Ser}} = 0.0009$ mole fraction/h) (experimental data from ref *46*). Note that the averages do not include time points for which no experimental data were available. See also Figure S7 in the Supporting Information for a full display of the individual time point glycosylation data.

the Iwasaki et al. data, we constructed the glycosylation time course for each serine and threonine (see Experimental Procedures and Supporting Figure S8) for comparison to the predictions of the ppGalNAc T2 kinetic model (Figure 7B,C and Supporting Figure S9). As shown in the figures, the ppGalNAc T2 positional weighting parameters derived from the CSM/PSM simulation can approximate most of the features of the experimental IgA1 glycosylation pattern: that is, the poor to very low glycosylation of Thr4_b, Thr12_b, and Ser17_d and the high glycosylation of Thr7_c, Thr15_c, Ser11_a, and Ser19.[4] Although the plots of simulated versus experimental glycosylation (Figure 7C**)** show no correlation for serine (principally due to Ser3_a, and Ser9_d), a very high correlation is found for threonine (see Figure S9). The value of 4.0 for the ratio of the optimized IgA1 threonine/serine rate constants, furthermore, is in agreement with the values obtained for CSM and PSM. We take this relatively successful simulation of the IgA1 domain as validation of both the general kinetic modeling approach and the CSM/PSM-derived ppGalNAc T2 weighting parameters.

DISCUSSION

In this work, an 80+ residue tandem repeat from the canine submaxillary gland mucin (CSM) containing over 25 hydroxyamino acid residues was isolated and sequenced, and its peptide-linked GalNAc glycosylation pattern was determined. This represents the third mucin tandem repeat to have its in vivo glycosylation pattern quantitatively determined (e.g., PSM and human MUC1 (*48*, *50*, *63*, *64*)). The properties of the CSM tandem repeat are summarized in the Results.

[4] In the model, Thr7_c and Thr15_c are most rapidly glycosylated since they have proline at the +3 position. Their glycosylation will reduce or inhibit the glycosylation of neighboring Thr4_b and Thr 12_b and Ser9_d and Ser17_d. Among these residues, Ser9_d and Ser17_d would be expected to be more rapidly glycosylated than Thr4_b and Thr12_b because of their preceding proline residues (see Figures 7 and S9). Because of the higher density of hydroxyamino acid residues in the IgA1 sequence (~50%) compared to CSM and PSM (~35%) the W_{OH_n} and W_{OG_n} values are found to play a more significant role in fitting IgA1 compared to CSM or PSM.

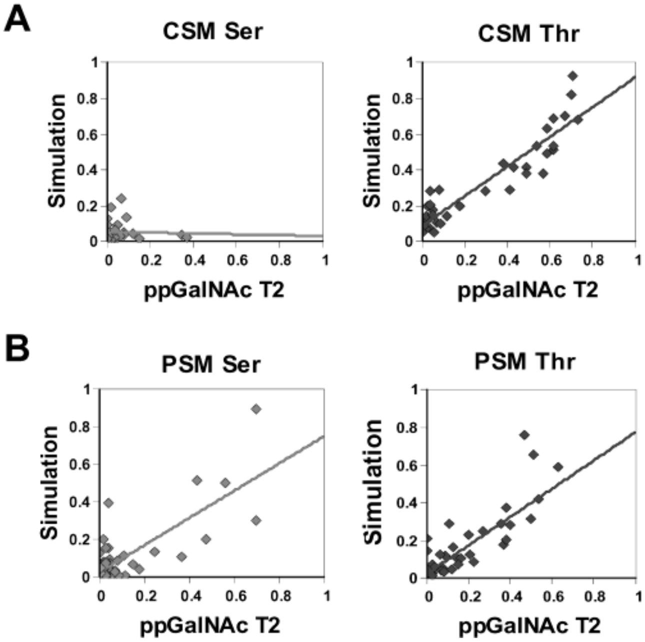


FIGURE 6: Plots of the simulated versus experimental ppGalNAc T2 site specific glycosylation for the CSM and PSM tandem repeats (individual time point data): (A, left) serine residues in CSM (SD = 0.12 mole fraction, $r^2 = 0$) and (right) threonine residues in CSM (SD = 0.09 mole fraction, $r^2 = 0.86$) using $k_{\mathrm{Thr}} = 0.0022$ and $k_{\mathrm{Ser}} = 0.0004$ mole fraction/h; (B, left) serine residues in PSM (SD = 0.11 mole fraction, $r^2 = 0.60$) and (right) threonine residues in PSM (SD = 0.11 mole fraction, $r^2 = 0.73$) using $k_{\mathrm{Thr}} = 0.0034$ and $k_{\mathrm{Ser}} = 0.0009$ mole fraction/h. Note that the statistical analysis for PSM omitted values for Ser2 and Thr79 as described previously (*46*).

With the isolation of the multimeric apo-CSM tandem repeat domain, a second large mucin substrate has become available for characterizing the substrate specificities of ppGalNAc transferases. By using such large apo-mucin tandem repeat domains containing multiple acceptor sites, the need to individually characterize large arrays of short peptides and the need to account for end-effects are effectively eliminated (i.e., ref *59*). Since all sites within the apo-mucin are exposed to the same conditions, experimental variability is reduced and differences between sites can be readily compared. As a final advantage, their glycosylation patterns can be modeled numerically to reveal the effects of local sequence and neighboring residue glycosylation (i.e., this work and refs *46* and *47*).

Recently, we reported a kinetic modeling approach capable of approximating the experimental ppGalNAc T1 and T2 glycosylation patterns on the PSM tandem repeat (*46*). In this model, serine and threonine glycosylation rate constants were reduced as a function of neighboring residue glycosylation status, imparting sensitivity to the presence of both neighboring glycosylated and nonglycosylated hydroxyamino acid residues. This model, however, only partially reproduces the presently obtained CSM glycosylation data for either transferase. Since the original model is clearly oversimplified, a more realistic model was developed that included neighboring proline and charged residue effects. Proline residues enhance O-glycosylation in vivo and in vitro and are predictors of O-glycosylation (see refs *56*−*60*). Neighboring charged residues typically reduce or inhibit O-glycosylation in vivo and in vitro (*57*−*59*), although neighboring glutamic acid residues occasionally enhance glycosylation (*56*, *60*). To maintain simplicity, no other residue types were included in the expanded model, although others could be easily included as additional experimental data warrant. Using this model, the ppGalNAc T1 and T2 glycosylation patterns of CSM and PSM are more reasonably reproduced. The ability of the model to approximate the reported ppGalNAc T2 glycosylation pattern of an IgA1 hinge domain peptide further validates both the kinetic modeling approach and the apo-mucin-derived ppGalNAc T2 parameters. Nevertheless, a small number of residues are poorly reproduced by the ppGalNAc T1 or T2 models. As suggested previously for PSM, local nonrandom peptide secondary structure may account for some of the deviations (i.e., Ser2 and Thr79 in PSM (*46*)). The inability to fully glycosylate a residue in the MUC1 tandem repeat has also been attributed to the formation of nonrandom secondary structure (*44*, *65*, *66*).

All known ppGalNAc transferases possess three ricin-like lectin domains at their C-terminus that are proposed to variably bind GalNAc-glycosylated peptide substrates (*20*, *26*, *29*, *38*, *44*, *45*, *67*−*69*). Tenno and co-workers (*68*, *69*) have shown that ppGalNAc T1 appears to utilize its ricin domains against an apo-bovine mucin substrate containing multiple glycosylation sites but not against small peptide substrates containing single glycosylation sites (*68*, *69*). Although, not examined by these workers, the glycosylation kinetics of apo-mucin with ppGalNAc T1 would be expected to be complex with an initial slow phase followed by a more rapid phase as the apo-mucin becomes increasingly glycosylated. We, however, do not observe such lag-like behavior in our ppGalNAc T1 or T2 studies on PSM or CSM. This may be due to our limited number of time points, due to the stepwise nature of our reglycosylation procedure, or both. However, the time course of the glycosylation of the IgA domain peptide by ppGalNAc T2 (*51*) reveals such effects on four residues (see Figure S8). The origins of these effects on IgA1 are unknown and may be either due to glycopeptide activity or due to conformational effects as a function of glycosylation. Highly detailed kinetic studies comparing ricin-domain mutant and wild-type transferases are needed to address the origins of these effects.

On the basis of the CSM/PSM-optimized positional weighting parameters, ppGalNAc T1 is highly inhibited by neighboring glycosylated residues, weakly inhibited by neighboring nonglycosylated hydroxyamino acid residues, and modestly (1.5−2-fold) enhanced by neighboring proline at nearly all positions. In contrast, Yoshida et al. (*59*) observe large rate enhancements (∼10- to ∼20-fold) for proline residues at positions +1 and +3 for a series of nine-residue peptide substrates. These differences may arise from peptide end effects or differences in the secondary structures of apo-mucin (highly extended random coil-like) and short peptides (the secondary structure/conformation of which may vary with amino acid substitution). Our finding for an apparent low sensitivity of ppGalNAc T1 to neighboring peptide residues suggests that this transferase may simply recognize a highly expanded random coil-like conformation for optimal activity. This is consistent with the recent NMR and molecular modeling studies of Kinarsky et al. (*66*) who suggest that ppGalNAc T1 may prefer residues in extended β-like conformations. This suggests that in vivo ppGalNAc T1 may serve as an early or initializing transferase designed to specifically glycosylate highly extended random coil-like

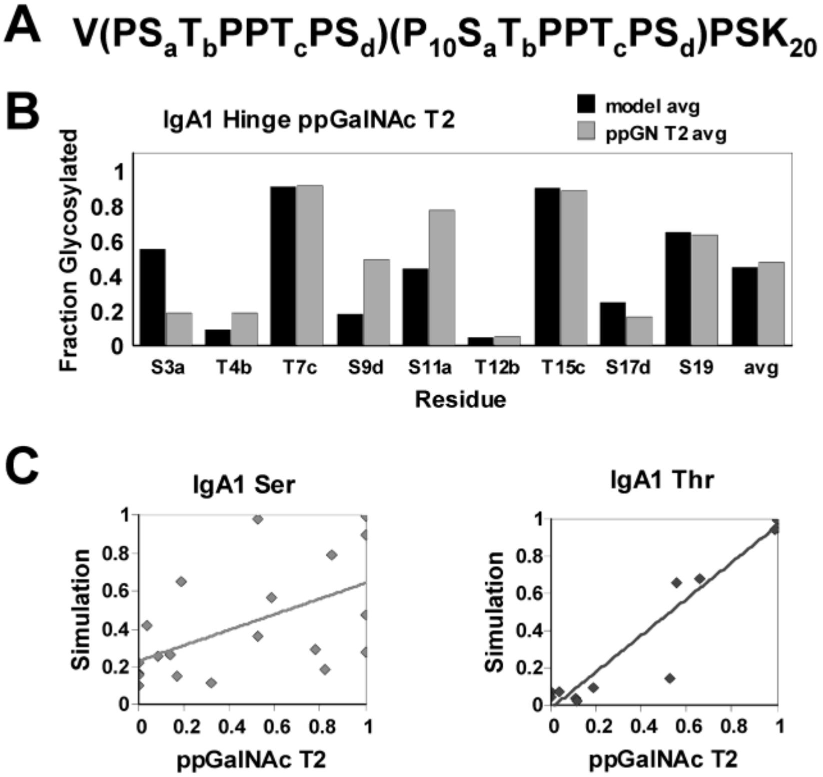


FIGURE 7: Modeling the in vitro glycosylation of the IgA1 hinge domain peptide by ppGalNAc T2: (A) peptide sequence of IgA1 domain showing two eight-residue tandem repeats (*51*); (B) comparison of the averaged calculated (black bars) and experimental (gray bars) ppGalNAc T2 glycosylation of the IgA1 peptide using the positional weighting parameters and factors given in Figure 6A ($k_{\mathrm{Thr}} = 0.028$, $k_{\mathrm{Ser}} = 0.007$ mole fraction/min). Experimental residue specific ppGalNAc T2 glycosylation data were derived from ref *51* as described in the Experimental Procedures. Panel C presents plots of simulated versus experimental ppGalNAc T2 site-specific glycosylation for the IgA1 peptide (individual time point data): (left) serine residues (SD = 0.30 mole fraction, $r^2 = 0.06$); (right) threonine residues (SD = 0.06 mole fraction, $r^2 = 0.94$). Note that the threonine residue statistics are better than they appear from the plots because of the clustering of several data points at 100%. See also Figures S8 and S9 in the Supporting Information for a full display of the individual time point glycosylation data. Note that the serine/threonine letter subscripts denote identical positions within tandem repeats.

(apo-mucin) peptides. As such ppGalNAc T1 could be expected to perform the bulk of the transfer of GalNAc to these molecules, thereby imprinting its sensitivity to neighboring glycosylation onto the mature mucin glycoprotein, indeed as we found for PSM (*19*). This is consistent with the generally high expression levels of ppGalNAc T1 (*21*) and accounts for our current observation that the in vivo glycosylation patterns of CSM and PSM tend to display characteristics of their in vitro ppGalNAc T1 glycosylation patterns.

The optimized modeling parameters for ppGalNAc T2 reveal maximal sensitivities to neighboring residue glycosylation and pronounced inhibitory sensitivities to neighboring hydroxyamino acid residues, particularly at the C-terminal +1 position relative to the site of glycosylation. This latter effect reduces the glycosylation of the N-terminal residue in hydroxyamino acid pairs (i.e., the CSM Thr59−60 and Thr67−68 and the PSM Ser6−7 and Thr29−30 dyad pairs). In contrast to ppGalNAc T1, proline residues exhibit a unique pattern of enhancements on ppGalNAc T2; proline at the N-terminal, −1, position gives a 10-fold enhancement, while proline at the −3 and +3 positions give 2.5- and 3-fold enhancements. The most rapidly glycosylating residues in CSM and PSM have N-terminal proline residues (i.e., Thr27 in CSM and Ser43 in PSM (*46*)).[5] Consistent with our detection of a large N-terminal proline rate enhancement, peptides shown to be good ppGalNAc T2 substrates typically contain Pro-Thr or Pro-Ser sequences (see Supporting Table S3 (*23*, *35*, *38*, *51*, *70*−*72*)).[6] Furthermore, ppGalNAc T2 is selectively absorbed (compared to ppGalNAc T1) on an affinity column of bound MUC2 peptide containing multiple Pro-Thr sequences (Table S3) (*70*). The association of high ppGalNAc T2 activity with peptides containing Pro-Thr and Pro-Ser sequences, coupled with our modeling, suggests that this sequence may represent the first described ppGalNAc T2 motif. The strong preference for an adjacent N-terminal proline residue is in keeping with the findings of Kinarsky et al. (*66*) who suggest that ppGalNAc T2 may prefer substrates with polyproline II-like conformations over extended β-strand conformations. Overall the strong neighboring sequence effects observed for ppGalNAc T2 tend to reduce the relative contribution of the inhibitory effect of neighboring glycosylation on the specificity of this transferase.

It has been proposed that members of the large family of evolutionarily conserved ppGalNAc transferases may have unique substrate specificities maintained for the efficient

[5] Ser76 in CSM, the only other hydroxyamino acid residue to have a preceding proline residue, is poorly glycosylated experimentally and slowly glycosylated in the ppGalNAc T2 simulation. An analysis of the model indicates that the glycosylation of this serine residue is slowed by the glycosylation of neighboring Thr78.

[6] Note that the MUC1 peptide, listed in Table S3, which lacks Pro-Ser or Pro-Thr sequences, is nevertheless a good substrate for ppGalNAc T2. This is presumably due to the 2.5−3.5-fold rate enhancements of the proline residues at the −3 or +3 positions or both relative to the sites of glycosylation.

glycosylation of specific proteins or peptide sequences (*17*, *20*, *21*). For example, a functional *Drosophila* ppGalNAc transferase [(*l(2)35Aa* or *pgant35A*, a homologue to the mammalian ppGalNAc T11] is essential for development in the fly, suggesting that no other transferase can perform its specific function (*17*, *18*). In keeping with this notion, Iwasaki and co-workers (*51*) propose that the initial O-glycosylation of the IgA1 hinge domain is performed in vivo primarily by ppGalNAc T2. ppGalNAc T2 is the only ppGalNAc transferase able to extensively glycosylate the IgA1 hinge domain peptide in vitro, and its glycosylation pattern is similar to the native in vivo pattern (see Figure S9, (*51*, *62*)). Based on our ppGalNAc T2 model, the IgA1 hinge domain is expected to be an excellent ppGalNAc T2 substrate (i.e., seven of nine hydroxyamino acid residues are preceded by proline and several have −3 or +3 proline residues), and the model largely parallels the sequential site-specific glycosylation observed experimentally. Therefore, on the basis of a relatively simple model, one can begin to account for the transferase specificity and glycosylation pattern of an important peptide substrate, which had not been previously understood (*51*). Our kinetic modeling of this substrate−transferase pair further demonstrates that transferases with apparently simple substrate motifs, nevertheless, may be capable of targeting the efficient glycosylation of a select subset of protein substrates.

In summary, an expandable kinetic modeling approach capable of uniquely characterizing ppGalNAc transferase substrate specificity has been demonstrated. The model is capable of reproducing the glycosylation patterns of two ppGalNAc transferase isoforms against two mucin substrates and is further validated by reproducing the features of the glycosylation of the IgA1 hinge peptide by ppGalNAc T2. These findings along with the recent demonstration of the ability to similarly model mucin Core 1 (*β*-Gal(1−3)-α-GalNAc-O-Ser/Thr) glycosylation patterns (*47*) suggests that in the future such modeling approaches may become useful for predicting the patterns of mucin O-glycosylation. In addition, the ppGalNAc transferase isoform-specific weighting parameters should become useful for the in vivo and in vitro search for transferase-specific protein substrates and for the development of specific transferase inhibitors.

ACKNOWLEDGMENT

We thank Dr. Ake Elhammer, Kalamazoo, MI, for the generous gifts of ppGalNAc T1 and T2. The technical assistance of Jessica Levine is also acknowledged.

SUPPORTING INFORMATION AVAILABLE

Supporting Tables S1−S4 and Supporting Figures S1−S9. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Van Klinken, B. J., Dekker, J., Buller, H. A., and Einerhand, A. W. (1995) Mucin gene structure and expression: Protection vs adheasion, *Am. J. Physiol. 269*, G613−G627.
2. Zheng, X., and Sadler, J. E. (2002) Mucin-like Domain of Enteropeptidase Directs Apical Targeting in Madin-Darby Canine Kidney Cells, *J. Biol. Chem. 277*, 6858−6863.
3. Rudd, P. M., Wormald, M. R., Stanfield, R. L., Huang, M., Mattsson, N., Speir, J. A., DiGennaro, J. A., Fetrow, J. S., Dwek, R. A., and Wilson, I. A. (1999) Roles for glycosylation of cell surface receptors involved in cellular immune recognition, *J. Mol. Biol. 293*, 351−366.
4. Fukuda, M., and Tsuboi, S. (1999) Mucin-type O-glycans and leukosialin, *Biochim. Biophys. Acta 1455*, 205−217.
5. Dennis, J. W., Granovsky, M., and Warren, C. E. (1999) Glycoprotein glycosylation and cancer progression, *Biochim. Biophys. Acta 1473*, 21−34.
6. Hanisch, F. G., and Muller, S. (2000) MUC1: the polymorphic appearance of a human mucin, *Glycobiology 10*, 439−449.
7. Tsuiji, H., Takasaki, S., Sakamoto, M., Irimura, T., and Hirohashi, S. (2003) Aberrant O-glycosylation inhibits stable expression of dysadherin, a carcinoma-associated antigen, and facilitates cell−cell adhesion, *Glycobiology 13*, 521.
8. Shogren, R., Gerken, T. A., and Jentoft, N. (1989) Role of glycosylation on the conformation and chain dimensions of O-linked glycoproteins: light-scattering studies of ovine submaxillary mucin, *Biochemistry 28*, 5525−5536.
9. Gerken, T. A., Butenhof, K. J., and Shogren, R. (1989) Effects of glycosylation on the conformation and dynamics of O-linked glycoproteins: Carbon-13 NMR studies of ovine submaxillary mucin, *Biochemistry 28*, 5536−5543.
10. Coltart, D. M., Royyuru, A. K., Williams, L. J., Glunz, P. W., Sames, D., Kuduk, S. D., Schwarz, J. B., Chen, X. T., Danishefsky, S. J., and Live, D. H. (2002) Principles of mucin architecture: structural studies on synthetic glycopeptides bearing clustered mono-, di-, tri-, and hexasaccharide glycodomains, *J. Am. Chem. Soc. 124*, 9833−9844.
11. Leppanen, A., White, S. P., Helin, J., McEver, R. P., and Cummings, R. D. (2000) Binding of Glycosulfopeptides to P-selectin Requires Stereospecific Contributions of Individual Tyrosine Sulfate and Sugar Residues, *J. Biol. Chem. 275*, 39569−39578.
12. Xu, Z., and Weiss, A. (2002) Negative regulation of CD45 by differential homodimerization of the alternatively spliced isoforms, *Nat. Immunol. 3*, 764−771.
13. Moody, A. M., Chui, D., Reche, P. A., Priatel, J. J., Marth, J. D., and Reinherz, E. L. (2001) Developmentally regulated glycosylation of the CD8α*β* coreceptor stalk modulates ligand binding, *Cell 107*, 501−512.
14. Merry, A. H., Gilbert, R. J. C., Shore, D. A., Royle, L., Miroshnychenko, O., Vuong, M., Wormald, M. R., Harvey, D. J., Dwek, R. A., Classon, B. J., Rudd, P. M., and Davis, S. J. (2003) O-Glycan Sialylation and the Structure of the Stalk-like Region of the T Cell Co-receptor CD8, *J. Biol. Chem. 278*, 27119−27128.
15. Alfalah, M., Jacob, R., Preuss, U., Zimmer, K. P., Naim, H., and Naim, H. Y. (1999) O-linked glycans mediate apical sorting of human intestinal sucrase-isomaltase through association with lipid rafts, *Curr. Biol. 9*, 593−596.
16. Breuza, L., Garcia, M., Delgrossi, M. H., and Le Bivic, A. (2002) Role of the membrane-proximal O-glycosylation site in sorting of the human receptor for neurotrophins to the apical membrane of MDCK cells, *Exp. Cell Res. 273*, 178−186.
17. Schwientek, T. J., Bennett, E. P., Flores, C., Thacker, J., Hollman, M., Reis, C. A., Behrens, J., Mandel, U., Keck, B., Schafer, M. A., Hazelmann, K., Zubarev, R., Roepstorff, P., Hollingsworth, M. A., and Clausen, H. (2002) Functional conservation of subfamilies of putative UDP-N-acetylgalactosamine: Polypeptide N-acetylgalactosaminyltransferases in drosophila, *C. elegans* and mammals: One subfamily comprised of l(2)35Aa is essential in drosophila, *J. Biol. Chem. 277*, 22623−22638.
18. Ten Hagen, K. G., and Tran, D. T. (2002) A UDP-GalNAc: polypeptide *N*-acetylgalactosaminyltransferase is essential for viability in *Drosophila melanogaster*, *J. Biol. Chem. 277*, 22616−22622.
19. Gerken, T. A., Gilmore, M., and Zhang, J. (2002) 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, *J. Biol. Chem. 277*, 7736−7751.
20. Ten Hagen, K. G., Fritz, T. A., and Tabak, L. A. (2003) “All in the Family” − The UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferases, *Glycobiology 13*, 1R−16R.
21. Young, W. W., Jr., Holcomb, D. R., Ten Hagen, K. G., and Tabak, L. A. (2003) Expression of UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase isoforms in murine tissues determined by real-time PCR: a new view of a large family, *Glycobiology 13*, 549−557.

22. Clausen, H., and Bennett, E. P. (1996) A family of UDP-GalNAc: polypeptide N-acetylgalactosaminyl-transferases control the initiation of mucin-type O-linked glycosylation, *Glycobiology 6*, 635−646.
23. Bennett, E. P., Hassan, H., Mandel, U., Mirgorodskaya, E., Roepstorff, P., Burchell, J., Taylor-Papadimitriou, J., Hollingsworth, M. A., Merkx, G., van Kessel, A. G., Eiberg, H., Steffensen, R., and Clausen, H. (1998) Cloning of a human UDP-N-acetyl-α-D-Galactosamine:polypeptide N- acetylgalactosaminyltransferase that complements other GalNAc- transferases in complete O-glycosylation of the MUC1 tandem repeat, *J. Biol. Chem. 273*, 30472−30481.
24. Ten Hagen, K. G., Hagen, F. K., Balys, M. M., Beres, T. M., Van Wuyckhuyse, B., and Tabak, L. A. (1998) Cloning and expression of a novel, tissue specifically expressed member of the UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase family, *J. Biol. Chem. 273*, 27749−27754.
25. Bennett, E. P., Hassan, H., Mandel, U., Hollingsworth, M. A., Akisawa, N., Ikematsu, Y., Merkx, G., van Kessel, A. G., Olofsson, S., and Clausen, H. (1999) Cloning and characterization of a close homologue of human UDP-N-acetyl- α-D-galactosamine:Polypeptide N-acetylgalactosaminyltransferase-T3, designated GalNAc-T6. Evidence for genetic but not functional redundancy, *J. Biol. Chem. 274*, 25362−25370.
26. Ten Hagen, K. G., Tetaert, D., Hagen, F. K., Richet, C., Beres, T. M., Gagnon, J., Balys, M. M., VanWuyckhuyse, B., Bedi, G. S., Degand, P., and Tabak, L. A. (1999) Characterization of a UDP-GalNAc:polypeptide N- acetylgalactosaminyltransferase that displays glycopeptide N- acetylgalactosaminyltransferase activity, *J. Biol. Chem. 274*, 27867−27874.
27. White, K. E., Lorenz, B., Evans, W. E., Meitinger, T., Strom, T. M., and Econs, M. J. (2000) Molecular cloning of a novel human UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase, GalNAc-T8, and analysis as a candidate autosomal dominant hypophosphatemic rickets (ADHR) gene, *Gene 246*, 347−356.
28. Toba, S., Tenno, M., Konishi, M., Mikami, T., Itoh, N., and Kurosaka, A. (2000) Brain-specific expression of a novel human UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase (GalNAc-T9), *Biochim. Biophys. Acta 1493*, 264−268.
29. Ten Hagen, K. G., Bedi, G. S., Tetaert, D., Kingsley, P. D., Hagen, F. K., Balys, M. M., Beres, T. M., Degand, P., and Tabak, L. A. (2001) Cloning and characterization of a ninth member of the UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase family, ppGaNTase-T9, *J. Biol. Chem. 276*, 17395−17404.
30. Guo, J. M., Zhang, Y., Cheng, L., Iwasaki, H., Wang, H., Kubota, T., Tachibana, K., and Narimatsu, H. (2002) Molecular cloning and characterization of a novel member of the UDP-GalNAc: polypeptide N-acetylgalactosaminyltransferase family, pp-GalNAc-T12, *FEBS Lett. 524*, 211−218.
31. Zhang, Y., Iwasaki, H., Wang, H., Kudo, T., Kalka, T. B., Hennet, T., Kubota, T., Cheng, L., Inaba, N., Gotoh, M., Togayachi, A., Guo, J., Hisatomi, H., Nakajima, K., Nishihara, S., Nakamura, M., Marth, J. D., and Narimatsu, H. (2003) Cloning and Characterization of a New Human UDP-N-Acetyl-α-D-galactosamine: Polypeptide N-Acetylgalactosaminyltransferase, Designated pp-GalNAc-T13, that is Specifically Expressed in Neurons and Synthesizes Tn Antigen, *J. Biol. Chem. 278*, 573−584.
32. Wang, H., Tachibana, K., Zhang, Y., Iwasaki, H., Kameyama, A., Cheng, L., Guo, J., Hiruma, T., Togayachi, A., Kudo, T., Kikuchi, N., and Narimatsu, H. (2003) Cloning and characterization of a novel UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase, pp-GalNAc-T14, *Biochem. Biophys. Res. Commun. 300*, 738−744.
33. Hagen, F. K., and Nehrke, K. (1998) cDNA Cloning and Expression of a Family of UDP-N-acetyl-D-galactosamine: Polypeptide N-Acetylgalactosaminyltransferase Sequence Homologues from *Caenorhabditis elegans*, *J. Biol. Chem. 273*, 8268−8277.
34. Ten Hagen, K. G., Tran, D. T., Gerken, T. A., Stein, D. S., and Zhang, Z. (2003) Functional Characterization and Expression Analysis of Members of the UDP-GalNAc:Polypeptide N-Acetylgalactosaminyltransferase Family from *Drosophila melanogaste*r, *J. Biol. Chem. 278*, 35039−35048.
35. Wandall, H. H., Hassan, H., Mirgorodskaya, E., Kristensen, A. K., Roepstorff, P., Bennett, E. P., Nielsen, P. A., Hollingsworth, M. A., Burchell, J., Taylor-Papadimitriou, J., and Clausen, H. (1997) Substrate specificities of three members of the human UDP-N-acetyl- α-D-galactosamine:Polypeptide N-acetylgalactosaminyltransferase family, GalNAc-T1, -T2, and -T3, *J. Biol. Chem. 272*, 23503−23514.
36. Kirnarsky, L., Nomoto, M., Ikematsu, Y., Hassan, H., Bennett, E. P., Cerny, R. L., Clausen, H., Hollingsworth, M. A., and Sherman, S. (1998) Structural analysis of peptide substrates for mucin-type O-glycosylation, *Biochemistry 37*, 12811−12817.
37. Iida, S., Takeuchi, H., Hassan, H., Clausen, H., and Irimura, T. (1999) Incorporation of N-acetylgalactosamine into consecutive threonine residues in MUC2 tandem repeat by recombinant human N-acetyl-D- galactosamine transferase-T1, T2 and T3, *FEBS Lett. 449*, 230−234.
38. Hassan, H., Reis, C. A., Bennett, E. P., Mirgorodskaya, E., Roepstorff, P., Hollingsworth, M. A., Burchell, J., Taylor-Papadimitriou, J., and Clausen, H. (2000) The lectin domain of UDP-N-acetyl-D-galactosamine: polypeptide N-acetylgalactosaminyltransferase-T4 directs its glycopeptide specificities, *J. Biol. Chem. 275*, 38197−38205.
39. Elhammer, A. P., Kezdy, F. J., and Kurosaka, A. (1999) The acceptor specificity of UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferases, *Glycoconjugate J. 16*, 171−180.
40. Brockhausen, I., Moller, G., Merz, G., Adermann, K., and Paulsen, H. (1990) Control of mucin synthesis: the peptide portion of synthetic O-glycopeptide substrates influences the activity of O-glycan core 1 UDP-galactose:N-acetyl-α-galactosaminyl-R β 3-galactosyltransferase, *Biochemistry 29*, 10206−10212.
41. Brockhausen, I., Toki, D., Brockhausen, J., Peters, S., Bielfeldt, T., Kleen, A., Paulsen, H., Meldal, M., Hagen, F., and Tabak, L. A. (1996) Specificity of O-glycosylation by bovine colostrum UDP-GalNAc: polypeptide α-N-acetylgalactosaminyltransferase using synthetic glycopeptide substrates, *Glycoconjugate J. 13*, 849−856.
42. Hanisch, F. G., Reis, C. A., Clausen, H., and Paulsen, H. (2001) Evidence for glycosylation-dependent activities of polypeptide N-acetylgalactosaminyltransferases rGalNAc-T2 and -T4 on mucin glycopeptides, *Glycobiology 11*, 731−740.
43. Kato, K., Takeuchi, H., Miyahara, N., Kanoh, A., Hassan, H., Clausen, H., and Irimura, T. (2001) Distinct Orders of GalNAc Incorporation into a Peptide with Consecutive Threonines, *Biochem. Biophys. Res. Commun. 287*, 110−115.
44. Hanisch, F. G., Muller, S., Hassan, H., Clausen, H., Zachara, N., Gooley, A. A., Paulsen, H., Alving, K., and Peter-Katalinic, J. (1999) Dynamic epigenetic regulation of initial O-glycosylation by UDP-N- Acetylgalactosamine:Peptide N-acetylgalactosaminyltransferases. site- specific glycosylation of MUC1 repeat peptide influences the substrate qualities at adjacent or distant Ser/Thr positions, *J. Biol. Chem. 274*, 9946−9954.
45. Bennett, E. P., Hassan, H., Hollingsworth, M. A., and Clausen, H. (1999) A novel human UDP-N-acetyl-D-galactosamine: polypeptide N- acetylgalactosaminyltransferase, GalNAc-T7, with specificity for partial GalNAc-glycosylated acceptor substrates, *FEBS Lett. 460*, 226−230.
46. Gerken, T. A., Zhang, J., Levine, J., and Elhammer, A. (2002) Mucin core O-glycosylation is modulated by neighboring residue glycosylation status: Kinetic modeling of the site-specific glycosylation of the apo-porcine sumbaxillary mucin tandem repeat by UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferases T1 and T2, *J. Biol. Chem. 277*, 49850−49862.
47. Gerken, T. A. (2004) Kinetic modeling confirms the biosynthesis of mucin Core 1 (β-Gal(1−3) α-GalNAc-*O*-Ser/Thr) *O*-glycan structures are modulated by neighboring glycosylation effects, *Biochemistry 43*, 4137−4142.
48. Gerken, T. A., Owens, C. L., and Pasumarthy, M. (1997) 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, *J. Biol. Chem. 272*, 9709−9719.
49. Gerken, T. A., Gupta, R., and Jentoft, N. (1992) A novel approach for chemically deglycosylating O-linked glycoproteins: The deglycosylation of submaxillary and respiratory mucins, *Biochemistry 31*, 639−648.
50. Gerken, T. A., Owens, C. L., and Pasumarthy, M. (1998) Site-specific core 1 O-glycosylation pattern of the porcine submaxillary gland mucin tandem repeat: Evidence for the modulation of glycan length by peptide sequence, *J. Biol. Chem. 273*, 26580−26588.
51. Iwasaki, H., Zhang, Y., Tachibana, K., Gotoh, M., Kikuchi, N., Kwon, Y. D., Togayachi, A., Kudo, T., Kubota, T., and Narimatsu, H. (2003) Initiation of O-glycan synthesis in IgA1 hinge region is determined by a single enzyme, UDP-N-Acetyl-α-D-galac-

tosamine: Polypeptide N-acetylgalactosaminyltransferase 2; pp-GalNAc-T2, *J. Biol. Chem. 278*, 5613−5621.

52. Combet, C., Blanchet, C., Geourjon, C., and Deleage, G. (2000) NIPS@:Network Protein Sequence Analysis, *Trends Biochem. Sci. 25*, 147−150.
53. Lombart, C., and Winzler, R. J. (1972) Isolation and characterization of canine submaxillary mucin, *Biochem. J. 128*, 975−977.
54. Corpet, F. (1988) Multiple sequence alignment with hierarchial clustering, *Nucleic Acids Res. 16*, 10881−10890.
55. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice, *Nucleic Acids Res. 22*, 4673−4680.
56. Elhammer, A. P., Poorman, R. A., Brown, E., Maggiora, L. L., Hoogerheide, J. G., and Kezdy, F. J. (1993) The specificity of UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase as inferred from a database of in vivo substrates and from the in vitro glycosylation of proteins and peptides, *J. Biol. Chem. 268*, 10029−10038.
57. Nehrke, K., Hagen, F. K., and Tabak, L. A. (1996) Charge distribution of flanking amino acids influences O-glycan acquisition in vivo, *J. Biol. Chem. 271*, 7061−7065.
58. Nehrke, K., Ten Hagen, K. G., Hagen, F. K., and Tabak, L. A. (1997) Charge distribution of flanking amino acids inhibits O-glycosylation of several single-site acceptors in vivo, *Glycobiology 7*, 1053−1060.
59. Yoshida, A., Suzuki, M., Ikenaga, H., and Takeuchi, M. (1997) Discovery of the Shortest Sequence Motif for High Level Mucin-type O-Glycosylation, *J. Biol. Chem. 272*, 16884−16888.
60. Hansen, J. E., Lund, O., Tolstrup, N., Gooley, A. A., Williams, K. L., and Brunak, S. (1998) NetOglyc: prediction of mucin type O-glycosylation sites based on sequence context and surface accessibility, *Glycoconjugate J. 15*, 115−130.
61. Elhammer, A., and Kornfeld, S. (1986) Purification and characterization of UDP-N-acetylgalactosamine: polypeptide N-acetylgalactosaminyltransferase from bovine colostrum and murine lymphoma BW5147 cells, *J. Biol. Chem. 261*, 5249−5255.
62. Mattu, T. S., Pleass, R. J., Willis, A. C., Kilian, M., Wormald, M. R., Lellouch, A. C., Rudd, P. M., Woof, J. M., and Dwek, R. A. (1998) The glycosylation and structure of human serum IgA1, Fab, and Fc regions and the role of N-glycosylation on Fc α receptor interactions, *J. Biol. Chem. 273*, 2260−2272.
63. Muller, S., Goletz, S., Packer, N., Gooley, A., Lawson, A. M., and Hanisch, F. G. (1997) Localization of O-glycosylation sites on glycopeptide fragments from lactation-associated MUC1. All putative sites within the tandem repeat are glycosylation targets in vivo, *J. Biol. Chem. 272*, 24780−24793.
64. Muller, S., Alving, K., Peter-Katalinic, J., Zachara, N., Gooley, A. A., and Hanisch, F. G. (1999) High-density O-glycosylation on tandem repeat peptide from secretory MUC1 of T47D breast cancer cells, *J. Biol. Chem. 274*, 18165−18172.
65. Kirnarsky, L., Prakash, O., Vogen, S. M., Nomoto, M., Hollingsworth, M. A., and Sherman, S. (2000) Structural effects of O-glycosylation on a 15-residue peptide from the mucin (MUC1) core protein, *Biochemistry 39*, 12076−12082.
66. Kinarsky, L., Suryanarayanan, G., Prakash, O., Paulsen, H., Clausen, H., Hanisch, F. G., Hollingsworth, M. A., and Sherman, S. (2003) Conformational studies on the MUC1 tandem repeat glycopeptides: implication for the enzymatic O-glycosylation of the mucin protein core, *Glycobiology 13*, 929−939.
67. Hagen, F. K., Hazes, B., Raffo, R., deSa, D., and Tabak, L. A. (1999) Structure−function analysis of the UDP-N-acetyl-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase. Essential residues lie in a predicted active site cleft resembling a lactose repressor fold, *J. Biol. Chem. 274*, 6797−6803.
68. Tenno, M., Saeki, A., Kezdy, F. J., Elhammer, A. P., and Kurosaka, A. (2002) The Lectin Domain of UDP-GalNAc:Polypeptide N-Acetylgalactosaminyltransferase 1 Is Involved in O-Glycosylation of a Polypeptide with Multiple Acceptor Sites, *J. Biol. Chem. 277*, 47088−47096.
69. Tenno, M., Kezdy F. J., Elhammer A. P., and Kurosaka A. (2002) Function of the lectin domain of polypeptide N-acetylgalactosaminyltransferase 1, *Biochem. Biophys. Res. Commun. 298*, 755−759.
70. Sorensen, T., White, T., Wandall, H. H., Kristensen, A. K., Roepstorff, P., and Clausen, H. (1995) UDP-N-acetyl-α-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase. Identification and separation of two distinct transferase activities, *J. Biol. Chem. 270*, 24166−24173.
71. Tetaert, D., Richet, C., Gagnon, J., Boersma, A., and Degand, P. (2001) Studies of acceptor site specificities for three members of UDP-GalNAc:N-acetylgalactosaminyltransferases by using a synthetic peptide mimicking the tandem repeat of MUC5AC, *Carbohydr. Res. 333*, 165−171.
72. Takeuchi, H., Kato, K., Hassan, H., Clausen, H., and Irimura, T. (2002) O-GalNAc incorporation into a cluster acceptor site of three consecutive threonines: Distinct specificity of GalNAc-transferase isoforms, *Eur. J. Biochem. 269*, 6173−6183.

BI049178E