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

(Received for publication, September 2, 1998, and in revised form, January 20, 1999)

Franz-Georg Hanisch‡‡, Stefan Müller‡, Helle Hassan‡, Henrik Clausen‡, Natasha Zachara‡, Andrew A. Gooley§, Hans Paulsen‡‡, Kim Alving‡‡, and Jasna Peter-Katalinic‡‡

From the ‡Institute of Biochemistry, University of Cologne, Joseph-Stelzmann-Strasse 52, 50931 Cologne, Germany, the §Department of Oral Diagnostics, School of Dentistry, University of Copenhagen, DK2200 Copenhagen, Denmark, the ¶Macquarie Center for Analytical Biotechnology, Macquarie University, Sydney NSW 2109, Australia, the **Institute of Organic Chemistry, University of Hamburg, 20146 Hamburg, Germany, and the ‡‡Institute of Medical Physics and Biophysics, University of Münster, 48149 Münster, Germany.

In search of possible epigenetic regulatory mechanisms ruling the initiation of O-glycosylation by polypeptide:N-acetylgalactosaminyltransferases, we studied the influences of mono- and disaccharide substituents of glycopeptide substrates on the site-specific in vitro addition of N-acetylgalactosamine (GalNAc) residues by recombinant GalNAc-Ts (rGalNAc-T1, -T2, and -T3). The substrates were 20-mers (HGV20) or 21-mers (AHG21) of the MUC1 tandem repeat peptide carrying GalNAcα or Galβ1–3GalNAcα at different positions. The enzymatic products were analyzed by MALDI mass spectrometry and Edman degradation for the number and sites of incorporated GalNAc. Disaccharide placed on the first position of the diad Ser-16-Thr-17 prevents glycosylation of the second, whereas disaccharide on the second position of Ser-16-Thr-17 and Thr-5-Ser-6 does not prevent GalNAc addition to the first. Multiple disaccharide substituents suppress any further glycosylation at the remaining sites. Glycosylation of Ser-16 is negatively affected by glycosylation at position −6 (Thr-10) or −10 (Ser-6) and is inhibited by disaccharide at position −11 (Thr-5), suggesting the occurrence of glycosylation-induced effects on distant acceptor sites. Kinetic studies revealed the accelerated addition of GalNAc to Ser-16 adjacent to GalNAc-substituted Thr-17, demonstrating positive regulatory effects induced by glycosylation on the monosaccharide level. These antagonistic effects of mono- and disaccharides could underlie a postulated regulatory mechanism.

Although no strict sequence dependence is known for the initiation of O-glycosylation by polypeptide:N-acetylgalactosaminyltransferases (ppGalNAc-Ts),§ functionally expressed recombinant enzymes display a distinct selectivity for peptide motifs in the vicinity of putative glycosylation sites (1–3). Until now, the prediction of O-glycosylation sites was based on two different approaches: the analysis of in vitro or of in vivo glycosylated peptides (4–6). Strikingly, these two methodological strategies revealed deviating results when the patterns of GalNAc addition to MUC1 tandem repeat peptide were analyzed (7, 8). All five putative sites were identified as glycosylation targets in vivo (8), whereas only Thr within the VTSA motif and/or Thr and Ser within the GSTA motif was glycosylated in vitro by the enzymes from tumor cells (7, 9) or milk (7) or by the recombinant GalNAc-Ts (T1–T3) (1–3). These differences could be explained by substrate specificities of the enzymes involved. On the other hand, the source of ppGalNAc-Ts used in the in vitro studies and that of the in vivo processed MUC1 were the same. However, distinct enzyme species may have been lost during preparation or may not be active under the conditions used for in vitro glycosylation. It could also be assumed that there is a need for these enzymes to act spatially or temporally in specific subcellular compartments that are not retained in the in vitro system. Finally, a further explanation should be considered that assumes that initial glycosylation of a peptide substrate influences the subsequent glycosylation events at vicinal or distant Ser/Thr positions. There are two observations that would favor this concept: 1) although ppGalNAc-Ts have largely been localized to the cis-Golgi (10), there are reports demonstrating a more diffuse distribution of these enzymes throughout the Golgi system (11); 2) the mucin core peptide can cycle between cis-Golgi and the endoplasmic reticulum (11) and GalNAc may be added successively to incompletely glycosylated substrates carrying short, core-type glycans. Positive or negative influences on the acceptor qualities of the remaining, still unglycosylated positions could be postulated to be exerted on the mono- or oligosaccharide level and could account for the partial unpredictability of the actual glycosylation sites. According to this concept, the site specificity of initial O-glycosylation would not merely be ruled by the peptide sequence around putative target sites. Evidence for negative effects on vicinal sites induced by mono- or disaccharide substituents has previously been reported for a series of glycopeptide substrates based on the MUC2 repeat peptide (12). Another in vitro glycosylation study indicated that positive regulatory effects on vicinal sites could also be responsible for the accelerated GalNAc transfer (7).

We followed this line of considerations by testing a panel of mono- and disaccharide substituted peptides corresponding to one MUC1 tandem repeat and carrying GalNAc or Galβ1–3GalNAc at single or multiple positions. The site-specific activ-
ities of three recombinant enzymes (rGalNAc-T1, -T2, and -T3) were assayed and compared with those of the soluble enzymes shed into human milk (milk GalNAc-Ts). The products were separated and quantitated by reversed phase HPLC, identified by MALDI mass spectrometry and the "terminally" glycosylated peptides with a maximum number of incorporated GalNAc were sequenced by Edman degradation to localize the sites of glycosylation. We were able to demonstrate that negative vicinal effects on glycosylation are exerted on the disaccharide level and are site-restricted. On the monosaccharide level, positional effects could be proven to occur in vitro, resulting in greatly accelerated GalNAc transfer to a vicinal position. Finally, also distant effects on initial O-glycosylation were observed that suggest that glycosylation-induced conformational changes of the peptide substrate may influence the accessibility of particular acceptor sites for ppGalNAc-Ts.

**EXPERIMENTAL PROCEDURES**

**Materials**

Glycopeptides and Peptides—The glycopeptides listed in Table I were synthesized as described previously (13) and were analyzed by 1H-NMR spectroscopy (400 MHz) (14) and by MALDI mass spectrometry (15). The peptides A1 to A8 correspond to a 21-mer of the MUC1 tandem repeat domain starting with the AHG motif and carrying one to four O-linked disaccharides Galβ1–3GalNAc at various positions. Glycopeptides A11 and A13 carrying GalNAc in defined positions of a 20-mer (HGV20) or 21-mer (AHG21) were synthesized similarly (13). Nonglycosylated peptide TAP25, corresponding to one repeat and five overlapping amino acids (starting with the TAP motif), was kindly provided by Dr. Taylor-Papadimitriou (Imperial Cancer Research Fund, London, United Kingdom).

Enzymes—Polypeptide GalNAc-transferases were obtained as follows: secreted, soluble recombinant forms of GalNAc-T1, -T2, and -T3 were expressed in insect cells as described previously (2, 3). Enzymes were partially purified from serum-free culture supernatant of transfectants of High-Five™ (Invitrogen) cells as described previously (3). The ppGalNAc-transferases from human milk were enriched by ultrafiltration through ultrafree MC membranes (Millipore, Eschborn, Germany) with a nominal cut-off of 10 kDa. Kinetic parameters were calculated by double reciprocal Lineweaver-Burk transformations at four substrate concentrations using assay conditions described by Brockhausen et al. (12).

**Measurement of In Vitro Glycosylation by Reversed Phase High Performance Liquid Chromatography**

Aliquots (50–100 μl) of the reaction mixtures were injected onto a PLRP-S column (250 × 4.6 mm, Polymer Laboratories, Shropshire, United Kingdom) or a narrow-bore ODS Ultrasphere column (150 × 2 mm, Beckman Instruments, Munich, Germany) and chromatographed on an HPLC system (System Gold Beckman Instruments, Munich, Germany) by gradient elution in a mixture of acetonitrile in water (0.1% trifluoroacetic acid) from 2% (solvent A) to 80% (solvent B) (duration, 80 min). Alternatively, gradient elution was performed starting from 0% solvent B to 6% solvent B (duration, 3 min), followed by a gradient from 6% solvent B to 16% solvent B (duration, 30 min). The glycopeptides were run at 1 ml/min and detected photometrically at 214 nm.

**Measurement of GalNAc Incorporation into Glycopeptide Substrates by Matrix-assisted Laser Desorption Ionization Mass Spectrometry**

The HPLC-purified glycopeptides were dried in a speed vac and dissolved in a mixture of water and methanol (1:1, v/v) to yield concentration of approximately 1 mg/ml 2,6-Dihydroxyacetophenone (concentration, 10 g/liter in 50% aqueous methanol) as a MALDI matrix was co-crystallized with the analyte in a dried droplet preparation. MALDI-time-of-flight experiments were performed on a VISION 2000 prototype mass spectrometer with a 2.3 m flight tube. Laser at 337 nm was used. Measurements were carried out in linear mode using appropriate delay time and potential to focus the ions of interest.

**TABLE I**

| Synthetic (glyco)peptides used in this study |
|---------------------------------------------|
| A1  | Ala-His-Gly-Val-Thr-Ser-Ala-Pro-Arg-Thr-Pro-Ala  |
| A2  | Ala-His-Gly-Val-Thr-Ser-Ala-Pro-Arg-Thr-Pro-Ala  |
| A3  | Ala-His-Gly-Val-Thr-Ser-Ala-Pro-Arg-Thr-Pro-Ala  |
| A4  | Ala-His-Gly-Val-Thr-Ser-Ala-Pro-Arg-Thr-Pro-Ala  |
| A5  | Ala-His-Gly-Val-Thr-Ser-Ala-Pro-Arg-Thr-Pro-Ala  |
| A6  | Ala-His-Gly-Val-Thr-Ser-Ala-Pro-Arg-Thr-Pro-Ala  |
| A7  | Ala-His-Gly-Val-Thr-Ser-Ala-Pro-Arg-Thr-Pro-Ala  |
| A8  | Ala-His-Gly-Val-Thr-Ser-Ala-Pro-Arg-Thr-Pro-Ala  |
| A9  | Ala-His-Gly-Val-Thr-Ser-Ala-Pro-Arg-Thr-Pro-Ala  |
| A10 | Ala-His-Gly-Val-Thr-Ser-Ala-Pro-Arg-Thr-Pro-Ala |
| A11 | Ala-His-Gly-Val-Thr-Ser-Ala-Pro-Arg-Thr-Pro-Ala |
| A12 | Ala-His-Gly-Val-Thr-Ser-Ala-Pro-Arg-Thr-Pro-Ala |
| A13 | Ala-His-Gly-Val-Thr-Ser-Ala-Pro-Arg-Thr-Pro-Ala |
| A14 | Ala-His-Gly-Val-Thr-Ser-Ala-Pro-Arg-Thr-Pro-Ala |
| A15 | Ala-His-Gly-Val-Thr-Ser-Ala-Pro-Arg-Thr-Pro-Ala |
| A16 | Ala-His-Gly-Val-Thr-Ser-Ala-Pro-Arg-Thr-Pro-Ala |
| A17 | Ala-His-Gly-Val-Thr-Ser-Ala-Pro-Arg-Thr-Pro-Ala |

GalNAc GalNAc-GalNAc

Glycopeptide A8 is contaminated with a <20% admixture of A6b containing an additional Thr at position 18.
Glycopeptides were sequenced on a Hewlett-Packard G1000A protein sequencer with 3.1 (solid) chemistry that uses methanol rather than ethyl acetate to transfer the ATZ-amino acid (16). The Sequelon AA™ membranes were from the Perceptive Biosystems division of Perkin-Elmer. Quantitation of the O-glycosidic serine/threonine substitution was calculated using a glycopeptide derived from Dictyostelium discoideum recombinant expressed glycoprotein PsA, which contains a 100% localization of glycosylation sites by Edman degradation of glycopeptides. The ultrafiltrated reaction mixtures were applied to FPLPS columns and run in water-acetonitrile gradients as described under "Experimental Procedures." A, the glycopeptide substrate A1 was incubated with the recombinant GalNAc transferases rGalNAc-T1 (A1G1), rGalNAc-T2 (A1G2), and rGalNAc-T3 (A1G3) and yielded the product indicated by A1-1. B, the glycopeptide A5 was incubated with rGalNAc-T1 (A5G1), rGalNAc-T2 (A5G2), and rGalNAc-T3 (A5G3) and yielded the products indicated by A5-1 and A5-2. C, the glycopeptide substrate A11 was incubated with a mixture of the rGalNAc-Ts (A11-G1-3) and yielded one product, indicated by A11-1.
Structural Analyses of Terminally Glycosylated Peptides—The glycopeptide substrates A1–A8, A11, and A13 (Table I) were incubated with the single or combined recombinant enzymes (or with the GalNAc-Ts from milk) for at least 72 h or up to 7 days in order to identify the sites and to measure the maximal number of GalNAc residues added to the Thr/Ser positions (Table III). With respect to positions Ser-6 and Thr-10, which are not target sites for rGalNAc-T1 to -T3 (HPLC and Edman) or with the GalNAc-Ts from human skin milk (HPLC and MALDI) separated by reversed phase HPLC and analysed by MALDI mass spectrometry and/or Edman sequencing for the number and sites of GalNAc addition.

| Peptide | Expected | HPLC | MALDI mass spectrometry | Edman |
|---------|----------|------|-------------------------|--------|
| A1      | 2        | 1    | 1                       | 1      |
| A2      | 3        | 3    | 3                       | 3      |
| A3      | 2        | 2    | 2                       | 2      |
| A4      | 3        | 3    | 3                       | 3      |
| A5*     | 2        | 1 (2) | 1 (2)                   | 1 (2)  |
| A6      | 1        | 0    | ND                      | ND     |
| A7      | 0        | 0    | ND                      | ND     |
| A8      | 1        | 1    | ND                      | ND     |
| A11     | 1        | 1    | 1 (2)                   | 1      |
| A12     | 3        | 3    | ND                      | ND     |
| TAP25   | 3        | 3    | 3                       | 3      |
| A13     | 3        | 3    | ND                      | ND     |

ND, not determined.
* Maximum numbers of GalNAc residues expected to be incorporated into MUC1 repeat peptide on the basis of previous in vitro glycosylation studies with TAP25 (7).

The intensity of the pseudomolecular ion M + H corresponding to A11-GalNAc, at m/z 2741.5 was below 4% of the major product A11-GalNAc after 168 h of reaction time.

substituted GlcNAc-Thr in position 4. The corrected yield for this cycle (repetitive yield for the glycopeptide ITATPAPT) was 95%. The retention time of GlcNAc-Thr was almost identical to the corresponding GalNAc derivative on reversed phase HPLC.

RESULTS

Dynamic Regulation of Initial O-Glycosylation

Vicinal and Distant Glycosylation—The glycopeptides (A1–A8, A11, and A13) and the nonglycosylated control peptide TAP25 were analyzed for the sites of glycosylation after in vitro transfer of the maximal number of GalNAc residues (Table III). On the basis of these qualitative data, several site-specific effects of substrate glycosylation on vicinal, proximal, and distant target sites were revealed.

Long Range Effects—The terminally glycosylated product of glycopeptide A1 did not carry GalNAc at Ser-16 (Table III). By contrast, the nonglycosylated control peptide TAP25 and the glycopeptide substrates, including glycopeptide A4, were glycosylated at this position. This finding indicates that a disaccharide substituted at Thr-5 can negatively affect a distant target site with respect to its acceptor qualities. The long range effect, which is exerted over a peptide stretch of 11 amino acids, is site-specific, because a disaccharide substitution at the adjacent Ser-6 (glycopeptide A4) had no influence on Ser-16 glycosylation (Table III). Exertion of the distance effect is also strikingly dependent on the glycan substituted at Thr-5, i.e., glycopeptide A11 with GalNAc at Thr-5 exhibited Ser-16 glycosylation.

Vicinal Effects—Results obtained for another group of glycopeptide substrates demonstrated that besides long range effects disaccharide substituents can also exert negative effects on vicinal target sites. In glycopeptide A5 (Table I), the disaccharide at Ser-16 is located adjacent to the acceptor position Thr-17 and may exert a negative influence, possibly mediated by steric hindrance, on the binding of ppGalNAc-transferases and hence prevents glycosylation of Thr-17 (Table III). A sequence variant of A5, glycopeptide A5b (Table I), with an additional Thr in position 18 was found to be glycosylated at this site (Table III). It can, therefore, be argued that the negative influence of the disaccharide in Ser-16 is solely exerted vicinal and not proximal. The relatively low amount of this minor glycopeptide (below 20%) renders unlikely substrate competition as an explanation for nonglycosylation of A5, the major glycopeptide in the substrate mixture. Interestingly, a similar negative effect with complete inhibition of vicinal glycosylation was not observed for GalNAc addition to Ser-16 in the substrates A3 (disaccharide at Thr-17) and A4 (disaccharide at Ser-6) (Table III). It may be suggested, accordingly, that disaccharides placed on the first position of an ST diad (here, Ser-16) could prevent the glycosylation of the second position, whereas disaccharide on the second position of ST or TS diads (here, Thr-17 or Ser-6) may not necessarily inhibit addition to the first.

Proximal Effects—Ser-16 glycosylation was not affected by disaccharide substitution at the proximal position –6 in glycopeptide substrate A2 (Table III). The same holds true for glycopeptide A13 exhibiting monosaccharide substitution of Thr-10 (Table III). Although the pattern of site-specific GalNAc addition on glycopeptides A2 and A13 agrees with that on the nonglycosylated control peptide TAP25, the rate of Ser-16 glycosylation in A2 and A13 was found to be affected negatively by glycans at Thr-10 (see kinetic analyses and Fig. 2).

Effects Induced by Multiple Glycosylation—Remarkably, substrate peptides substituted with more than one disaccharide (A6, A7, and A8) could not serve as acceptors for rGalNAc-Ts or milk GalNAc-Ts at any of the remaining possible positions (Table III). With respect to positions Ser-6 and Thr-10, which are not in vitro target sites for rGalNAc-T1 to -T3 or milk GalNAc-Ts, this finding excludes the possible existence of positive regulatory mechanisms mediated by core1-disaccharides. On the other hand, the negative effect on Ser-16 glycosylation in A6 induced by disaccharide substitution at Thr-5 and Thr-17 is solely exerted on the disaccharide substituted at Thr-5.
charide level, because the structural analogue of the glycopeptide, the monosaccharide substituted glycopeptide A11, was a substrate of ppGalNAc-Ts (Table III). Most likely, the prevention of Ser-16 glycosylation in A6 is caused by disaccharide substitution of Thr-5 and is exerted in the same way as the negative long range effect on Ser-16 glycosylation in glycopeptide A1. It should be recalled at this point that vicinal substitution with disaccharide at Thr-17 did not suppress Ser-16 glycosylation in A3 (Table III).

Kinetic Studies—The rates of GalNAc incorporation into the three actual in vitro glycosylation sites of MUC1 repeat peptide (Thr-5, Ser-16, and Thr-17) were different for each position and ppGalNAc-T used. Kinetic studies were performed by quantitative HPLC/Edman sequencing and by radiometric measurement of $[^{14}C]$GalNAc incorporation to determine the apparent kinetic constants. This approach allowed the identification and quantitation of each intermediate or final glycopeptide product. It was also possible to assign for some of the glycopeptide substrates apparent $K_m$ and $V_{max}$ values to distinct glycosylated sites.

Site Preferences of ppGalNAc-Ts—The ppGalNAc-Ts from milk glycosylated Thr-5 at higher rates than Thr-17 or Ser-16. This is evidenced from previous HPLC/Edman analysis of TAP25 products (7), but also from the apparent kinetic constants (Table III) obtained for glycopeptide substrates A1 (glycosylation of Thr-17; see also Fig. 2C), A3 (mainly glycosylation of Thr-5; see also Fig. 2, A and C), and A5 (glycosylation of Thr-5). The lowest relative reaction rates were measured for the Ser-16 position, which made glycosylation of this site the rate-limiting step in the glycosylation of peptides A2, A3, A4, A13, and TAP25 (Fig. 2, A and B). Accordingly, no Ser-16 glycosylation was generally found in HPLC fractions corresponding to intermediate products. The same relative rates of GalNAc incorporation into the three positions were found when using the recombinant enzymes in a mixture. The single rGalNAc-T species exhibited distinct preferences for the in vitro glycosylation sites within the VTSA and GSTA motifs, respectively (Tables IV and V). As demonstrated previously for rGalNAc-T1 and -T3 (1), the best substrates for these enzymes are characterized by a nonglycosylated Thr-5 (compare A3 to A1 in Table IV), whereas rGalNAc-T2 prefers the nonglycosylated GSTA motif (compare A1 to A3 and A5 in Table IV). rGalNAc-T2 is the enzyme species that adds GalNAc most efficiently to Ser-16 (glycosylation of A11). This is also evident

| Peptide | Sites of glycosylation in final products $^a$ | Kinetic data for milk-GalNAc-Ts $^b$ |
|---------|---------------------------------------------|--------------------------------------|
|         |                                             | $\text{pmol h}^{-1}$ $\text{mM}^{-1}$ | $K_{max}$ (nmol h$^{-1}$) |
| A1      | T S T S T S T T                              | 490                                  | 1.3 |
| A4      | T S T S T S T T                              | 1954                                 | 0.4 | 2.33 |
| A5      | T S T T S T T T T                            | 2690                                 | 0.3 | 3.51 |
| A5b     | T S T S T T T T                             | 3359                                 | 0.1 | 3.95 |
| A3      | T S T S T S T T                            | 0                                     | 0.6 | 0.38 |
| A2      | T S T S T S T                              | 237                                  | 0.6 | 0.38 |
| A11     | T S T S T S T                             | 0                                     | 0.6 | 0.38 |
| A7      | T S T S T S T                             | 0                                     | 0.6 | 0.38 |
| A8      | T S T S T S T                             | 0                                     | 0.6 | 0.38 |

$^a$The Ser positions carrying the disaccharide Gal1-3GalNAc are marked by $\bullet$, those substituted with GalNAc by $\triangle$. Positions, where GalNAc residues were introduced in vitro glycosylation, are marked by arrows. For each glycopeptide substrate the maximal number of incorporated GalNAc residues is shown. The glycopeptide substrates were incubated with a mixture of GalNAc-T1 to -T3 or with the milk-GalNAc-Ts. The final glycopeptide products obtained after 72 h incubation and HPLC separation were analyzed by MALDI-MS and Edman degradation for the number and sites of GalNAc addition.

$^b$The substrate mixtures were incubated with the ppGalNAc-Ts from human milk at four different peptide concentrations (0.1, 0.5, 1.0, and 2.0 mM) to determine the apparent kinetic constants. Since the milk-GalNAc-Ts are regarded as a mixture of enzymes with different site preferences, no attempt was made to quantitate the individual site constants. The specific activity of the enzyme preparation was measured at a substrate concentration of 1.0 mM and given as pmol h$^{-1}$. The assays were carried out as described in the “Methods” section. The estimated errors for the measurement of reaction rates were in the range of 5%.
from HPLC analysis of the final product of A3 glycosylation (Table V). Whereas the formation of A3(Thr-5, Ser-16)-GalNAc2 catalyzed by the rGalNAc-T1 and rGalNAc-T3 did not exceed 7% of the total glycopeptides (substrate and products) after reaction times of 72 h, the same product represented 28% of total glycopeptide using rGalNAc-T2 (Table V).

**Negative Glycosylation-induced Effects**—Glycosylation of Ser-16 is apparently the most sensitive indicator of effects induced by a series of glycosylation events at the proximal sites 2, 6, 2, 11 relative to the target Thr/Ser (Fig. 2, A and B). Relative to the nonglycosylated control peptide TAP25, all substrates with GalNAc substitution at Thr-10 (A13), Galβ1–3GalNAc substitution at Thr-10 (A2) and Thr-5 (A1), or Ser-6 (A4) exhibit reduced (A2, A4, and TAP25) or even no detectable glycosylation of Ser-16 (A1, not shown). Compared with the nonglycosylated control peptide TAP25, Thr-17, also, is affected negatively by the disaccharide in Thr-5 (A1), because even after 72 h, more than 60% of substrate was detected in the reaction mixture (Fig. 2, C and D).

**Positive Glycosylation-induced Effects**—Besides negative effects exerted by glycan substituents on vicinal, proximal and distant sites, the acceleration of reaction rates was also demonstrated to be induced by vicinal glycosylation. Position Ser-16, which except for rGalNAc-T2 is reportedly a poor substrate of the GalNAc-Ts (3), becomes a significantly better target site after substitution of the vicinal Thr-17 with GalNAc (glycopeptide A11, Fig. 2A). Accordingly, for TAP25 lacking a GalNAc substitution at Thr-17 the percentage of Ser-16 glycosylated products measurable after 24 h was below 2%. When the glycopeptide A11 was used as a substrate, around 17% of Ser-16 glycosylated products were measurable at the same time (Fig. 2A). Whereas with TAP25 a maximal incorporation of 15% was reached after 72 h, about 42% of Ser-16 glycosylated products were detected with A11. The positive effect described above is exerted only on the monosaccharide level (see Ser-16 glycosylation of A3 in Fig. 2A).

**DISCUSSION**

Previous *in vitro* studies of initial O-glycosylation of MUC1 repeat peptides revealed that only three out of five putative sites were targets for the ppGalNAc-Ts from different sources and recombinant forms of the first three cloned rGalNAc-Ts.
(1–3, 7, 9). This in vitro pattern, however, contrasted strikingly with the glycosylated sites identified on VNTR glycopeptides derived from purified milk MUC1, which exhibited at least partial glycosylation of five potential glycosylation positions (8). These observations prompted us to look for possible explanations that could abrogate this discrepancy: 1) undefined site-specific ppGalNAc-Ts may exist that could add the sugar residue to the positions Ser-6 within the VTSA motif or to Thr-10 within the DTRP motif, 2) the assay conditions might be not optimal for the respective enzymes used, or 3) glycosylation at one site could influence the substrate qualities at other potential O-glycosylation sites. The latter influence could be expected to be negative due to a glycosylation-induced decrease in the accessibility of the peptide to the enzymes. On the other hand, only positive glycosylation-induced effects on poor substrate positions would explain the higher degree of site occupation found in vivo.

The reported results clearly demonstrate that all of the above assumptions can be regarded as valid in the in vitro model of MUC1 glycosylation. Negative glycosylation effects on the GalNAC transfer into vicinal sites, but also into distant sites, were observed on the disaccharide level of glycopeptide substrates (see Table III). The stimulation exerted by GalNAc substitution of Thr-17 in GSTA on the vicinal Ser glycosylation was the only positive effect detected in this model (see Fig. 2, A and B). Accordingly, poor target sites Ser-6 and Thr-10 remained unglycosylated on all peptide and glycopeptide substrates, even when reaction time was extended to 7 days. Traces of GalNAc added to one of these sites by milk enzymes after such a long reaction time may indicate that the kinetics of the stochastic in vitro glycosylation deviates by orders of magnitude from that of in vivo, which occurs at a highly organized membraneous surface. These considerations may explain the lack of in vitro glycosylation of Ser-6 and Thr-10. On the other hand, results with novel rGalNAc-Ts show that among these enzymes, rGalNAc-T4 (17) exhibits specificity for these two sites preferentially, with peptide substrates substituted already with GalNAc at Thr-5, Ser-16 and Thr-17. The present study demonstrates that work with nonglycosylated peptide substrates could principally yield different glycosylation patterns compared with the use of glycopeptide substrates (refer to the glycosylation patterns of A1 and A5). Hence, the addition of GalNAc in vitro is not merely ruled by the sequence context of acceptor sites but can be affected by preceding glycosylation of other sites of the peptide substrate. This finding should be considered when analyzing the specificities of individual ppGalNAc-Ts in vitro. The use of glycosylated peptides as substrates is justified by the assumption that in vivo, also, the preferred target sites should be glycosylated with GalNAc (or even with complex, core-type glycans) before initiation at other potential, but poor substrate positions has been started. Accordingly, the effects observed in vitro could also rule the patterns of GalNAc addition to the mucin core peptide in the Golgi system. We have previously shown that although all five potential sites within the repeat peptide are glycosylation targets in vivo, on the average, only 2.6 sites per repeat carry glycan moieties (8). The degree of substitution at each site as revealed by quantitative Edman degradation, in conjunction with evi-

### Table IV

| Substrate (preferred target site) | Enzyme GalNAc-T | pmol h⁻¹ | K₉ | V₉₉ |
|----------------------------------|-----------------|-----------|----|------|
| A1 (Thr-17)                      | T1              | 64        | ND | ND   |
|                                 | T2              | 1696      | 1.4| 4.00 |
|                                 | T3              | 109       | ND | ND   |
| A3 (Thr-5)                       | T1              | 432       | ND | 0.4   |
|                                 | T2              | 203       | 2.5| 0.63 |
|                                 | T3              | 1317      | 0.3| 1.82 |
| A11 (Ser-16)                     | T1              | <50       | ND | ND   |
|                                 | T2              | 128       | 0.5| 0.19 |
|                                 | T3              | <50       | ND | ND   |
| A2 (Thr-17)                      | T2              | 1682      | 0.7| 3.03 |
| A4 (Thr-17)                      | T2              | 601       | 0.2| 0.85 |
| A5 (Thr-5)                       | T2              | 309       | 2.0| 1.00 |

### Table V

| Substrate | Enzyme          | Intermediate product | Final product |
|-----------|-----------------|----------------------|---------------|
|           |                 | Product Percentage   |                |
| A1        | GalNAc-T1       |                      | 33.0          |
| A1        | GalNAc-T2       |                      | 100.0         |
| A1        | GalNAc-T3       |                      | 54.2          |
| A3        | GalNAc-T1       | 87.5                 | 6.4           |
| A3        | GalNAc-T2       | 91.2                 | 4.5           |
| A3        | GalNAc-T3       |                      | 84.2          |
| A5        | GalNAc-T1       |                      | 59.5          |
| A5        | GalNAc-T2       |                      | 84.0          |
| A5        | GalNAc-T3       |                      |               |

Glycosylation-induced effects on the activities of rGalNAc-T1 to -T3 as revealed by the apparent kinetic constants for glycopeptide substrates

The reaction conditions were identical to those given in Table 3 except for the replacement of milk enzymes by the individual rGalNAc-Ts. In case of substrates A1, A5, and A11, where only one GalNAc is added to a defined peptide position, the apparent kinetic constants correspond to the individual site constants: Thr-17 (A1), Thr-5 (A5), and Ser-16 (A11). Considering the negligible rates of Ser-16 glycosylation, an estimation of the individual site constants for GalNAc addition to Thr-5 (A3) is possible. Because the transferase rGalNAc-T2 glycosylates preferentially sites within the GSTA motif and Ser-16 glycosylation occurs at very low rates compared to Thr-17 glycosylation, the apparent constants for A2 and A4 primarily reflect glycosylation of Thr-17. Errors of rate determination were estimated to be in the range of 5%. ND, not determined.

Glycosylation of the underlined peptide position indicates the rate limiting reaction during formation of the final product. Substrates and products were quantitated by rpHPLC after 72 h of reaction time.

| Substrate | Enzyme          | Intermediate product | Final product |
|-----------|-----------------|----------------------|---------------|
|           |                 | Product Percentage   |                |
| A1        | GalNAc-T1       |                      |               |
| A1        | GalNAc-T2       |                      |               |
| A1        | GalNAc-T3       |                      |               |
| A3        | GalNAc-T1       | 87.5                 |               |
| A3        | GalNAc-T2       | 91.2                 |               |
| A3        | GalNAc-T3       |                      |               |
| A5        | GalNAc-T1       |                      |               |
| A5        | GalNAc-T2       |                      |               |
| A5        | GalNAc-T3       |                      |               |
dence from mass spectrometry indicate a heterogeneous pattern of site-specific glycosylation within the repeat peptide. This heterogeneity could be explained in part by the findings discussed above.

What sorts of conformational or steric effects could underlie the positive or negative changes of site-specific substrate qualities? On the basis of NMR data and molecular modeling techniques, it has recently been suggested that ppGalNAc-Ts (T1 and T3) require at least three points of contact to stabilize the spatial orientation of the target hydroxyl group (18). Although one of these groups is the target site itself, the second and third sites are located in the vicinity (between +3/+6 and between −2/−6). It was further suggested that five or six residues adjacent to both sides of the target Thr/Ser should be in an extended β-like or an inverse γ-turn conformation, which was proven to exist in the VTSA motif (18). Already previous studies on model peptides had shown that in vitro O-glycosylation of potential target sites was dependent on a random-coil conformation of the peptide substrate (4). Native mucins like those from submaxillary glands tend to adopt such random-coil structure leading to extended conformations with peptide dimensions about 3-fold more expanded than found for deglycosylated mucins (19). However, apomucins from submaxillary glands, also, provably exist in a more extended, nonglobular conformation with high content of aperiodic structure and β-turns facilitating addition of GalNAc to the peptide scaffold (20). Nonglycosylated (21) and glycosylated VNTR peptides of MUC12,3 have been analyzed by measurement of circular dichroism2 and by NMR spectroscopy,2,3 and they turned out to form a left-handed poly-L-proline II helix stabilized by addition of GalNAc residues.2 This type of secondary structure is a characteristic feature of mucins, because also for other members of this subclass of glycoproteins, like the human salivary mucin MG2, the existence of a significant population of poly-L-proline II-type helices has been demonstrated in aqueous solution (22). Of particular significance in the context of this study are interactions between the sugar substituent and the peptide backbone with influences on the peptide conformation and substrate qualities of potential O-glycosylation sites. NOE connectivities between the N-acetyl NH proton of the GalNAc and the yH protons of the prolines at +3 indicate that the GalNAc residues are positioned along the peptide backbone of MUC1 repeats in a C-terminal direction (1). This finding is in partial agreement with work on model glycopeptides for antifreeze glycoprotein (23) and ovine submaxillary mucin (19). In both studies, the authors were able to demonstrate the existence of an intramolecular hydrogen bond between the amide proton of GalNAc and the carbonyl oxygen of the glycosylated threonine resulting in a C-terminal orientation of the sugar relative to the protein backbone. NMR analyses of the glycopeptides used in this study are still in progress, but the available data confirm a relatively extended conformation of the VNTR peptide disrupted by polypeptide helical elements and a turn-like self-stabilizing PDTR motif.3

The knowledge summarized above on peptide conformation of MUC1 at the glycosylated motifs and on glycan/peptide interaction, in conjunction with the suggested requirements for substrate binding by the ppGalNAc-Ts, may partly explain why core-type glycosylation at −6 relative to the Ser-16 target site (A2, A13) negatively influences GalNAc addition at this position (see Fig. 2B). Similar considerations should be relevant in the context of GalNAc addition to Thr-17 in A5 and Thr-17/Thr-18 in A5b, where the Thr at +1 remains unglycosylated, whereas that at +2 is a target for the respective ppGalNAc-T. On the basis of available data, no straightforward explanation can be given to the negative effect exerted by Galβ1–3GalNAc at Thr-5 on the glycosylation at Ser-16, in particular, because no such effect was found with GalNAc substituted at Thr-5. While long range interactions between the VTSA motif and the PDTR motif were not observed (18), information on interactions of distant motifs with the GSTA motif is lacking. With concern to the positive effect on Ser-16 glycosylation exerted by GalNAc at Thr-17, it may be allowed to speculate that the sugar interaction with the proline residue at +3 reduces the flexibility of the peptide chain and in this way may stabilize a particular conformation that is preferred by the transferase. Alternatively, it cannot be excluded that the binding site of the enzyme forms specific contacts with the hydroxyl groups of sugar residues at position +1.

The message of this paper is regarded to be found in a mechanic model that could explain the regulation of initial O-glycosylation. The postulated dynamic regulation of peptide

---

2 D. I. R. Spencer, S. Missailidis, C. DeMatteis, M. S. Searle, S. J. B. Tendler, and M. R. Price, presented at the Workshop on Mucin O-Glycosylation: Sites and Processing, Copenhagen, Denmark 1997.

3 J. Dejahn, C. Dietel, H. Paulsen, and B. Meyer, B., presented at the 5th International Workshop on Carcinoma-associated Mucins, Cambridge, 1998, and personal communication by Hans Paulsen.
glycosylation during endoplasmic reticulum/cis-Golgi cyclization of proteins may supplement the site-specific addition of GalNac by specialized ppGalNAc-Ts. It remains to be established whether glycosylation-induced effects on GalNac transfer to particular target sites on monorepeats of MUC1 can also be observed on multiple repeats, because artificial effects related to the sizes of the peptide substrates cannot be ruled out. No such effects had been observed by Nishimori et al. (9) using nonglycosylated oligomeric MUC1 repeats. In conclusion, the differential acceptor qualities of individual peptide positions in conjunction with enzyme repertoire and activity of ppGalNAc-Ts in Golgi compartments should result in an ordered sequence of glycosylation, which is partly regulated by competitive glycosyltransferases involved in the synthesis of core-type glycans. As presented in Fig. 3, the density of peptide glycosylation could accordingly be regulated by the competition between initiation by ppGalNAc-Ts and the core glycan synthesis by β-6-glucosaminyltransferase (core2 enzyme) or β-3-galactosyltransferase (core1 enzyme). Based on this postulated regulatory model, carcinoma cells with a deficient glycosylation machinery lacking one of these core enzymes or expressing these at low activity levels (Fig. 3B) should glycosylate peptides to a higher degree of substituted sites compared with normal cells (Fig. 3A). This assumption has recently been confirmed for the breast cancer cell line T47D, which does not express a functional core2 enzyme, by showing that on the average, 4.8 of the five putative sites within VNTR peptide are glycosylated. This finding is in striking contrast to that obtained for the lactation-associated glycoform, where only half of the putative sites per repeat had been found to be glycosylated on the average (8). A similar relationship between density of O-glycosylation and the chain lengths of glycans has recently been demonstrated by structural analysis of porcine submaxillary mucin (24).

Further studies should show on which level of glycosylation the initiation by ppGalNAc-Ts is inhibited most effectively and whether the postulated model has relevance also for the glycosylation of other mucin peptides. In particular, the occurrence of positive stimulatory effects on initial O-glycosylation that were shown to be induced by prior GalNac addition could explain the formation of clustered O-linked glycans on mucins, in general. Similar positive effects exerted by prior glycosylation on poor substrate positions as in the MUC1 model were not described yet for other mucin peptides. However, the present study calls the hypothesis into question that such effects could explain the discrepancies found between site-specific glycosylation of MUC1 VNTR peptide in vitro and in vivo. Results from in vitro glycosylation (25–27) and the derived rules on how sequences that flank Ser/Thr target sites influence the enzyme binding to substrate have to be critically re-evaluated. Site-specific glycosylation patterns predicted on the basis of in vivo data (28) agree better with the actual localization of O-linked glycans.

Acknowledgment—We thank Dr. F. Hillenkamp (University of Münster, Germany) for kind permission to use the VISION 2000 instrument for the performance of MALDI mass spectrometry.

REFERENCES
1. Sørensen, T., White, T., Wandall, H. H., Kristensen, A. K., Roepstorff, P., and Clausen, H. (1995) J. Biol. Chem. 270, 24166–24173
2. Bennett, E. P., Hassan, H., and Clausen, H. (1996) J. Biol. Chem. 271, 17006–17012
3. 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) J. Biol. Chem. 272, 23563–23571
4. O’Connell, B. C., Tabak, L. A., and Ramasubbu, N. (1993) Biochem. Biophys. Res. Commun. 180, 1024–1030
5. Nehrke, K., Hagen, F. K., and Tabak, L. A. (1996) J. Biol. Chem. 271, 7061–7065
6. Gooley, A. A., and Williams, K. L. (1994) Glycobiology 4, 413–417
7. Stadie, T. R., Chai, W., Lawson, A. M., Byfield, P. G., and Hanisch, F.-G. (1995) Eur. J. Biochem. 229, 140–147
8. Müller, S., Goletz, S., Packer, N., Gooley, A. A., Lawson, A. M., and Hanisch, F.-G. (1997) J. Biol. Chem. 272, 24780–24793
9. Nishimori, I., Johnson, N. R., Sanderson, S. D., Perini, F., Mountjoy, K., Cerny, R. L., Gross, M. L., and Hollingsworth, M. A. (1994) J. Biol. Chem. 269, 16123–16130
10. Roth, J., Wang, Y., Eckhardt, A. E., and Hill, R. L. (1994) Proc. Natl. Acad. Sci. U. S. A. 91, 8935–8939
11. Röttger, S., White, T., Wandall, H. H., Olivo, J. C., Stark, A., Bennett, E. P., Whitehouse, C., Berger, E. G., Clausen, H., and Nilsson, T. (1998) J. Cell Sci. 111, 45–60
12. Brochhausen, I., Taki, D., Brochhausen, J., Peters, S., Bielefeldt, T., Kleen, A., Paulsen, H., Meldal, M., Hagen, F., and Tabak, L. A. (1996) Glycoconj. J. 13, 849–856
13. Karsten, U., Diotel, C., Klig, G., Paulsen, H., Goletz, S., Müller, S., and Hanisch, F.-G. (1998) Cancer Res. 58, 2541–2549
14. Mathieux, N., Paulsen, H., Meldal, M., and Bock, K. (1997) J. Chem. Soc. Perkin Trans. 1, 2359–2368
15. Goletz, S., Thiede, B., Hanisch, F.-G., Schultz, M., Peter-Katalinic, J., Müller, S., Seitz, O., and Karsten, U. (1997) Glycoconjugate J. 7, 881–896
16. Zachara, N. E., and Gooley, A. A. (1998) in Mucin Methods and Protocols: Methods in Molecular Biology (Corfield, T., ed) Humana Press, Totowa, NJ, in press
17. 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) J. Biol. Chem. 273, 30472–30481
18. Kirnarsky, L., Nomoto, M., Ikematsu, Y., Hassan, H., Bennett, E. P., Cerny, R. L., Clausen, H., Hollingsworth, M. A., and Sherman, S. (1998) Biochemistry 37, 12811–12817
19. Butenholz, K. J., and Gerken, T. A. (1995) Biochemistry 34, 2650–2663
20. Eckhardt, A. E., Timpke, C. S., Abernethy, J. L., Touniadje, A., Johnson, N. R., and Hill, R. L. (1987) J. Biol. Chem. 262, 11339–11344
21. Fontenot, J. D., Tjandra, N., Bu, D., Ho, C., Montelaro, R. C., and Finn, O. J. (1993) J. Biol. Chem. 268, 5386–5394
22. Antonyraj, K. J., Karunanarakan, T., and Raj, P. A. (1998) Arch. Biochem. Biophys. 356, 197–206
23. Minura, Y., Yamamoto, Y., Inoue, Y., and Chuj, R. (1992) Int. J. Biol. Macromol. 14, 242–248
24. Gerken, T. A., Owens, C. L., and Pasumathy, M. (1998) J. Biol. Chem. 273, 26580–26588
25. Wilson, J. B. H., Gavel, Y., and von Heijne, G. (1991) Biochem. J. 275, 529–534
26. Wang, Y., Abernethy, J. L., Eckhardt, A. E., and Hill, R. L. (1992) J. Biol. Chem. 267, 12759–12766
27. Wang, Y., Agrwal, N., Eckhardt, A. E., Stevens, R. D., and Hill, R. L. (1993) J. Biol. Chem. 268, 22979–22983
28. Hansen, J. E., Lund, D., Engelbrecht, J., Bohr, H., Nielsen, J. O., Hansen, J.-E. S., and Brunak, S. (1993) Biochem. J. 308, 801–813

4 S. Müller, K. Alving, J. Peter-Katalinic, N. Zachara, A. A. Gooley, and F.-G. Hanisch, submitted for publication, and presented at the 5th International Workshop on Carcinoma-associated Mucins, Cambridge, 1998.