[CANCER RESEARCH 54, 3738–3744, July 15, 1994]

N-Acetylgalactosamine Glycosylation of MUC1 Tandem Repeat Peptides by Pancreatic Tumor Cell Extracts[1]

Isao Nishimori, Fulvio Perini, Kimberly P. Mountjoy, Sam D. Sanderson, Nicole Johnson, Ronald L. Cerny, Michael L. Gross, J. Darrell Fontenot, and Michael A. Hollingsworth[2]

Eppley Institute for Research in Cancer and Allied Diseases, University of Nebraska Medical Center, Omaha, Nebraska 68198-6805 [I. N., F. P., K. P. M., S. D. S., N. J., M. A. H.]; First Department of Internal Medicine, Kochi Medical School, Kochi 783, Japan [I. N.]; Department of Chemistry, University of Nebraska, Lincoln, Nebraska 68588-0304 [R. L. C., M. L. G.]; and Theoretical Biology and Biophysics, T10, MS-K710, Los Alamos National Laboratory, Los Alamos, New Mexico 87545 [D. F.]

ABSTRACT

Synthetic peptides corresponding to the human mucin MUC1 tandem repeat domain (20 residues) were glycosylated *in vitro* by using UDP-*N*-[^{3}H]acetyl-D-galactosamine (GalNAc) and lysates of pancreatic tumor cell lines. Results obtained with peptides of different lengths (from one to five repeats) suggest that increasing the number of tandem repeats has neither a positive nor a negative effect on the density of glycosylation along the MUC1 tandem repeat protein backbone. Purified glycopeptides were sequenced on a gas-phase sequencer, and glycosylated positions were determined by measuring the incorporated radioactivity in fractions collected following each round of Edman degradation. The results showed that two of three threonine residues on the MUC1 tandem repeat peptides were glycosylated by pancreatic tumor cell lysates at the following positions: GVTSAPDTRPAPGSTAPPAH (underlined T indicates position of GalNAc attachment). None of the serine residues were glycosylated. Determination of the mass of the glycopeptides by mass spectrometry confirmed that a maximum of two molecules of GalNAc were covalently linked to each 20-residue repeat unit in the peptides. The data presented here show that acceptor substrate specificity of the UDP-GalNAc:polypeptide *N*-acetylgalactosaminyltransferase detected in lysates of pancreatic and breast tumor cell lines is identical and is limited to some but not all threonines in the MUC1 tandem repeat peptide sequence. The influence of primary amino acid sequence on acceptor substrate activity was evaluated by using several peptides that contain single or double amino acid substitutions (relative to the native human MUC1 sequence). These included substitutions in the residues that were glycosylated and substitutions of the surrounding primary amino acid sequence. The results of these studies suggest that primary amino acid sequence, length, and relative position of the residue to be glycosylated dramatically affect the ability of peptides to serve as acceptor substrates for the UDP-GalNAc:polypeptide *N*-acetylgalactosaminyltransferase.

INTRODUCTION

It has been shown previously that mucins produced by human adenocarcinomas often contain tumor-associated epitopes (1). These include both carbohydrate structures and protein structures. Generally, these tumor-associated carbohydrate epitopes are formed by (*a*) incomplete or blocked synthesis, which causes an accumulation of precursor forms; (*b*) a shift in the pathway resulting in the expression of minor antigens; (*c*) neosynthesis of carbohydrate through the activation of certain glycosyltransferases; or (*d*) organizational changes in the cell membrane exposing carbohydrate antigens that may otherwise be obscured (2). In the case of at least one mucin core protein, MUC1, aberrant glycosylation by breast adenocarcinomas results in the exposure of epitopes detected by monoclonal antibodies on the core protein that are not commonly found on forms of the mucin produced by corresponding normal tissues (3). Similar antibody-defined epitopes were not found on mucin secreted by pancreatic tumor cell lines unless it was chemically deglycosylated; however, these epitopes can be detected in pancreatic tumor cells, presumably on incompletely processed mucin that is not secreted (4). For breast, pancreatic, and ovarian adenocarcinomas, MUC1 core protein epitopes recognized by T-cells have been described (5, 6).

One predominant epitope identified as a tumor-associated antigen on the MUC1 core protein includes the amino acid sequence PDTRP. One current hypothesis proposes that the addition of truncated oligosaccharides to tumor-derived mucin (compared to longer oligosaccharide structures on mucin produced by normal tissues) facilitates antibody binding to this epitope by unmasking this portion of the core protein; this epitope is inaccessible to some monoclonal antibodies on some forms of MUC1, possibly because of the presence of large, branched carbohydrate structures (3). Another possibility, which has not been ruled out, is that there are differences in the positions of attachment of the *O*-glycosyl-linked oligosaccharides to the MUC1 core protein. A combination of these explanations is also possible. The studies reported herein address the latter hypothesis. We recently used breast tumor cell line extracts to glycosylate, *in vitro*, synthetic peptides with sequences based on the MUC1 tandem repeat (7). The results of these studies showed that the threonine found in the sequence PDTRP was not glycosylated *in vitro* by a breast tumor cell line, whereas the two other threonines (but no serines) in the MUC1 tandem repeat were glycosylated. The results reported herein demonstrate that human pancreatic tumor cell lines glycosylate the same relative positions on the MUC1 tandem repeat. In addition, the levels of GalNAc transferase[3] activity were determined for a number of cell lines, including pancreas, breast, and colon cancer cell lines.

MATERIALS AND METHODS

Tumor Cell Lines. Human pancreatic adenocarcinoma cell lines HPAF, Capan-1 (well differentiated), T3M4, BxPC-3 (moderately differentiated), PANC-1, HGC-25 (poorly differentiated), and PANC-89 (undefined differentiation); human breast adenocarcinoma cell lines BT-20 (moderately differentiated), SK-BR-3 (poorly differentiated), and MCF7; human colon adenocarcinoma cell line HT 29; human melanoma cell line SK-MEL-28; and human neuroblastoma cell line SK-N-SH were grown in Eagle's minimum essential medium with 10% fetal calf serum, 2 mM L-glutamine, 100 unit/ml penicillin G, and 0.1 mg/ml streptomycin (unless otherwise indicated by the American Type Culture Collection catalogue).

Preparation of Cell Extracts. Tumor cells were harvested from culture flasks (~80% confluence) and washed three times with standard phosphate-buffered saline. Cells were incubated on ice with 1% Triton X-100 (Sigma

Received 3/21/94; accepted 5/13/94.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[1] This work was supported in part by NIH Grants CA57362, DK47662, and CA36727; American Cancer Society Grant SIG 16; and the Nebraska Cancer and Smoking Disease Research Program. The four-sector tandem mass spectrometer was purchased with funds awarded by former National Science Foundation Regional Instrumentation Grant CHE8620177 and the University of Nebraska-Lincoln. Ongoing support for this facility was provided by National Science Foundation Biology Division Grant DIR901762. N. J. was supported by NIH Training Grant T32 CA09476.

[2] To whom requests for reprints should be addressed, at Eppley Institute for Research in Cancer and Applied Diseases, University of Nebraska Medical Center, 600 South 42nd Street, Box 986805, Omaha, NE 68198-6805.

[3] The abbreviations used are: GalNAc transferase, UDP-*N*-acetyl-D-galactosamine: polypeptide *N*-acetylgalactosaminyltransferase (EC 2.4.1.41); GalNAc, *N*-acetyl-D-galactosamine; HPLC, high pressure liquid chromatography.

Chemical Co., St. Louis, MO) in phosphate-buffered saline for 30 min. After centrifugation (1000 × *g*, 5 min), the protein concentration of the supernatant was adjusted to 8 mg/ml (Bio-Rad Protein Assay; Bio-Rad, Richmond, CA) and the lysate was then stored at −70°C for later use.

Synthesis of Acceptor Substrate Peptides. Synthetic peptides used in this study included 8, 9, 20, 30, 40, 60, 80, and 105 residue peptides based on the native MUC1 tandem repeat sequence (GSTAPPAHGVTSAPDTRPAP). The sequence of the 20-residue peptide (corresponding to a single tandem repeat unit) was APGSTAPPAHGVTSAPDTRP, and the sequence of the 30-residue peptide corresponding to 1.5 of the tandem repeats was GVTSAPDTRPAPG-STAPPAHGVTSAPDTRP. The sequences of 40-, 60-, and 80-residue peptides were 2, 3, and 4 repeats of the sequence PDTRPAPGSTAPPAHGVTSA, respectively. The sequence of the 105-residue peptide was 5 repeats of the tandem repeat plus an additional sequence, GVTSA, at the amino terminus. Sixteen overlapping 9-residue peptides based on the native MUC1 tandem repeat sequence, and a number of nonnative (substituted) 9-residue peptides that contained substitutions of one or two amino acids into the native sequence were also used. Short peptides (30 residues or less) were synthesized according to the general principles of solid phase methodology by using an Applied Biosystems Model 430A peptide synthesizer as described previously (8). Long peptides (>30 residues) were synthesized as described previously (9). Peptides were purified by reverse phase HPLC as described previously (7). The concentration, purity, and fidelity of the peptides were confirmed by amino acid composition analysis (10) and mass spectrometry (7).

Assay of GalNAc Transferase Activity. GalNAc transferase activity was measured as described previously (7) with modifications as described below. The standard assay mixture contained 50 mM imidazole-HCl (pH 7.2), 1% Triton X-100, 20 mM $MnCl_2$, 64 μM peptide, 50 μM UDP-[^{3}H]GalNAc (DuPont New England Nuclear, Boston, MA; diluted with cold UDP-GalNAc to 260 dpm/pmol for the experiments reported here), and 0.8 mg protein/ml cell lysate in a total volume of 50 μl. Variations in this standard assay condition are reported in "Results" or in the figure legends. Reactions were performed at 37°C for 4 h and were terminated by adding 200 μl of water. Reaction mixtures were then applied to columns of ion-exchange resin, 1 ml of Dowex 1-X8-200 (Sigma) in a Pasteur pipet plugged with glass wool, and eluted with 2.5 ml water. Radioactivity in fractions from the Dowex column were quantified in scintillation fluid (ScintiVerse II; Fisher Chemical, Fair Lawn, NJ) by liquid scintillation spectrometry (LS3801; Beckman, Fullerton, CA). Net radioactivity incorporated into peptide substrates was determined after subtracting background values from assays that did not include peptide substrates. The reported results are the mean of three independent experiments; SE is also reported.

Purification of Glycosylated Peptide. Glycosylated peptides were purified and characterized as described previously (7). Briefly, fractions from the Dowex 1 chromatography step that contained peptide components were concentrated, applied to a P-10 gel filtration column (Bio-Gel P-10, fine; Bio-Rad), and eluted with water (14 × 460 mm; flow rate, 0.32 ml/min). The radioactivity in each fraction (0.96 ml) was quantified by liquid scintillation spectroscopy. Fractions from selected experiments were also tested for the presence of MUC1 peptide by using a dot blot test evaluating reactivity with monoclonal antibody HMFG-2 (kindly provided by Drs. Sandra Gendler and Joyce Taylor-Papadimitriou, Imperial Cancer Research Fund, London, England), which recognizes the amino acid sequence DTR in the MUC1 tandem repeat. Those fractions that contained radiolabeled components and that were reactive with monoclonal antibody HMFG-2 were concentrated to 1 ml in siliconized tubes on a vacuum centrifuge concentrator (Jouan, Winchester, VA).

Peptides (glycosylated and nonglycosylated) in the P-10 column fractions were purified to homogeneity by using reverse phase HPLC and the conditions described below. The instrumentation included an ISCO (Lincoln, NE) Model 2360 low pressure gradient programmer, an ISCO Model 2350 pump, and an ISCO Model V4 absorbance detector. A Waters (Milford, MA) Delta-Pak C_{18} column (3.9 × 150 mm) was used for separation of the 20- and 30-residue peptides. Solvent A was 0.1% trifluoroacetate (Advanced Chemtech, Louisville, KY) in water; solvent B was 0.095% trifluoroacetate in acetonitrile:water (80:20). Gradients were A plus 0–35% buffer B in 35 min at 1 ml/min. MUC1 peptides (20-residue; nonglycosylated) were eluted at 25 min and 30-residue MUC1 peptides (nonglycosylated) were eluted at 29 min. Absorbance was monitored at 214 nm.

Mass Spectrometry. The mass spectra were obtained with the first stage of a VG ZAB-T mass spectrometer operating at base resolution (R = 1000) to provide maximum sensitivity. Samples (10 pmol or more) were dissolved in 10 μl of acetonitrile:water; (1:1) a 1-μl aliquot was placed on the stainless steel probe tip along with 1 μl of the matrix (1:1 glycerol/*m*-nitrobenzyl alcohol). Ions were desorbed with a 30-keV Cs^+ ion beam. The instrument was operated at an accelerating potential of 8 keV (giving an upper m/z of 10,000) and scanned over a mass range from 700 to 4600 daltons at a rate of 15 s/decade of mass.

Identification of *O*-Glycosylation Sites with a Gas-Phase Sequencer. An aliquot of each sample was covalently attached to Sequelon AA membrane (Milligen, Burlington, MA) according to the manufacturer's instructions. The membrane disk was then placed directly on the lower cartridge insert of a Porton (Beckman) PI2090E Microsequencer. A Zitex membrane was placed over it and the reaction cartridge was sealed. The membrane was washed with 1 ml of extraction solvent prior to sequencing.

The extraction solvent was 70% aqueous methanol with 16 μg/ml dithiothreitol. The standard sequencing program was modified to eliminate the postcleavage drying step, the number of extractions was increased from two to three, and the reconstitution volume was increased from 200 to 300 μl. One-half of each reconstituted cycle was diverted to a fraction collector, and the fractions from each Edman degradation cycle were evaluated by scintillation counting to determine the radioactive content. The remainder was injected onto a Hewlett Packard (Kennett Square, PA) AminoQuant (2.1- × 200-mm) column for identification of the phenylthiohydantoins.

RESULTS

GalNAc Transferase Activity. The effects of varying assay conditions (time, 0 to 8 h; pH, 4.5–9.0; cell extract/enzyme concentration, 0–6 mg/ml; MUC1 peptide substrate concentration, 0–1024 μM; and UDP-GalNAc concentration, 0–200 μM) on total GalNAc transferase activity were evaluated, and appropriate conditions for different aspects of this study were selected. The standard assay conditions selected for the studies reported here are described under "Materials and Methods." Representative apparent K_m values for selected peptide acceptor substrates shown here were: for the 30-residue peptide, 122 μM for HPAF cell lysates; for the 20-residue peptide, 200 μM for HPAF cell lysates.

Effect of Peptide Length on Glycosylation *in Vitro*. We have determined previously the positions of attachment of GalNAc to MUC1 tandem repeat peptides that were glycosylated *in vitro* by using extracts from a breast tumor cell line (7). These assays used peptides up to 30 amino acid residues long, which contained 1.5 of the MUC1 tandem repeat units. It was reasoned that 1.5 repeats would contain a sufficient amount of overlapping sequence to present all possible primary amino acid sequence combinations as a substrate for evaluating the specificity of GalNAc transferase activity. Two lines of experimental analysis, direct sequencing and mass spectrometry, suggest that there are two predominant sites of glycosylation on each MUC1 tandem repeat unit (7). It has been postulated that longer stretches of MUC1 tandem repeat assume higher levels of ordered secondary structure (11). In theory, these might present different conformations that would alter their ability to act as substrates for the GalNAc transferase.

We tested this hypothesis by evaluating enzyme activity on peptides with different numbers of tandem repeats to determine whether longer peptides receive higher or lower densities of glycosylation per tandem repeat unit. GalNAc transferase assays were performed with several peptides that contained multiple MUC1 tandem repeats (from 1 to 5). The results of these experiments (Fig. 1) show that there is a linear correlation (R = 0.986) between the number of tandem repeat units and the relative enzyme rate, suggesting that the amount of GalNAc incorporated into each tandem repeat during *in vitro* assays is not affected by tandem repeat length. This result suggests that increas-

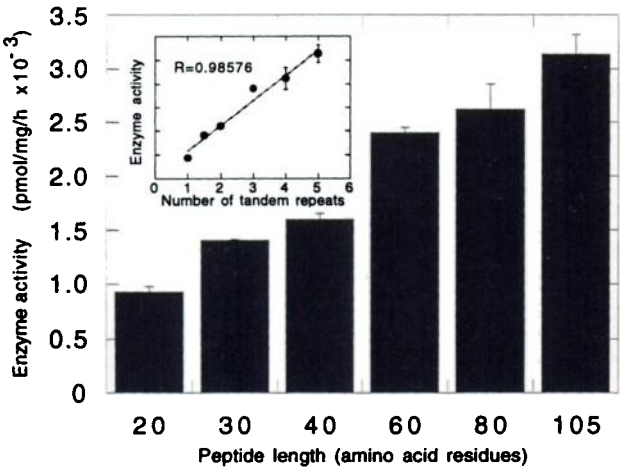

Fig. 1. Relative GalNAc transferase activity with peptides that contain from 1 to 5 MUC1 tandem repeat units. The relative enzyme activity of GalNAc transferase in BXPC-3 cell lysates (*ordinate*) was determined by using as acceptor substrates synthetic peptides that contained from 1 to 5 of the MUC1 tandem repeat units (*abscissa*). The sequence of the 20-residue peptide corresponding to a single tandem repeat unit was APGSTAPPAHGVTSAPDTRP, and the sequence of the 30-residue peptide corresponding to 1.5 of the tandem repeats was GVTSAPDTRPAPGSTAPPAHGVTSAPDTRP. The sequences of 40-, 60-, and 80-residue peptides were 2, 3, and 4 repeats of the sequence PDTRPAPGSTAPPAHGVTSA, respectively. The sequence of the 105-residue peptide was 5 repeats of the tandem repeat plus an additional sequence GVTSA at the amino terminus. Enzyme assays were performed using the standard conditions described under "Materials and Methods." *Bars*, SE.

ing the numbers of tandem repeats does not have either a positive or a negative effect on the density of glycosylation along the MUC1 tandem repeat protein backbone.

Sequence Analysis and Mass Spectrometry of Glycosylated Peptides. Peptides (20 residue and 30 residue) glycosylated *in vitro* by HPAF lysates were purified and sequenced by using an Edman degradation procedure in concert with scintillation counting (see "Materials and Methods") to determine the positions of glycosylation. The results of sequencing the 20-residue peptide are shown in Fig. 2. The same relative positions were glycosylated on the 30-residue peptide (data not shown). We detected the correct amino acid residues through 18 cycles under these experimental conditions. Two major peaks of radioactive GalNAc were seen at the cycles corresponding to the T positions in GSTAPP and GVTSAP for both the 20- and 30-residue peptides. The comparatively high levels of radioactivity seen in the cycles immediately following these two T positions were due to sequencing lag effects caused by cumulative inefficient Edman degradation reactions during previous cycles, as evidenced by analysis of the HPLC retention times of the phenylthiohydantoins. In addition, the results of analysis by fast atom bombardment mass spectrometry of the glycosylated 30-residue peptide showed components with masses corresponding to the peptide with 1, 2, and 3 GalNAc residues attached, but no more than 3 (data not shown). These results suggest that under the conditions of this assay, the 20-residue peptide is glycosylated at (underlined) APGSTAPPAHGVTSAPDTRP. The three sites of glycosylation of the 30-residue tandem repeat peptide are (underlined) GVTSAPDTRPAPGSTAPPAHGVTSAPDTRP. There are no detectable differences in the sites of glycosylation between the pancreatic tumor cell line HPAF reported here and the breast tumor cell line MCF-7 (7) under the conditions used for these assays.

GalNAc Transferase Activity Using Overlapping 9-Residue Peptides. We performed GalNAc transferase assays with 16 different 9-residue peptide substrates that represent a moving set through MUC1 tandem repeat sequence to evaluate the contribution of substrate primary amino acid sequence to the relative activity and specificity of the human GalNAc transferase expressed in the pancreatic tumor cell lines. It was important to perform these experiments with pancreatic tumor cell line extracts to determine whether there was any evidence for distinct GalNAc transferase activities (different substrate specificities) among those cell lines or between the pancreatic tumor cell lines and a breast tumor cell line that was evaluated previously (7). Fig. 3 presents data obtained when the overlapping set of 9-residue peptides were used as substrates in GalNAc transferase assays; lysate from the pancreatic tumor cell line BxPC3 was the enzyme source. Similar results were obtained when lysates from two other pancreatic tumor cell lines, HPAF and PANC-1 (data not shown), were evaluated. Standard assay conditions were used for evaluating GalNAc transferase activity with this set of peptides. The data presented in Fig. 3 show that there were two particularly effective peptide substrates with the sequences PGSTAPPAH and HGVTSAPDT. Relatively high enzyme rates were detected with peptides that had overlapping sequences with these. These results are consistent with the sequencing results described above that show the glycosylated positions on MUC1 peptides to be threonines (underlined) at the following positions: PGSTAP and GVTSAP. Other data supporting the hypothesis that the threonine is the predominant site of glycosylation on these peptides are presented in Fig. 5 and are discussed below. The data presented in Fig. 3 support the hypothesis that the threonine in the sequence position PDTRP is not glycosylated by pancreatic tumor, breast tumor, or colon carcinoma cell lines. These data closely resemble the data obtained when the same set of peptides was tested with a breast adenocarcinoma cell extract (7), providing support for the hypothesis that there are no differences in the substrate specificity of GalNAc transferase activity between these types of tumors. Moreover, the results presented in Fig. 3 confirm the previous observation (7) that GalNAc transferase activity in these cell extracts is affected by the length of the substrate primary amino acid sequence around the glycosylated residue.

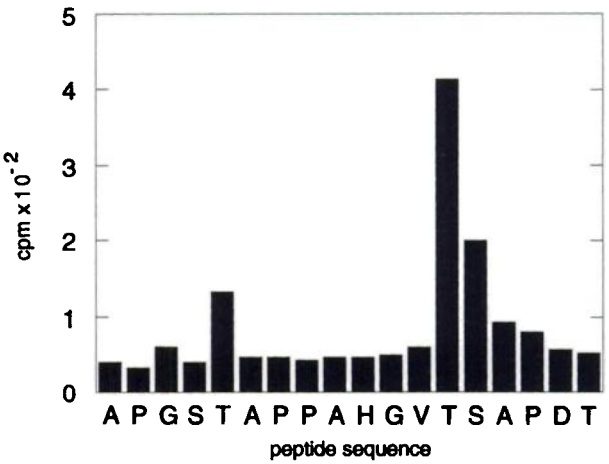

Fig. 2. Identification of *O*-glycosylation sites on a 20-residue MUC1 tandem repeat peptide. Representative results of sequencing the glycosylated 20-residue peptide (*abscissa*) and quantification of the radioactivity detected in each sequencing cycle (*ordinate*). Sequencing experiments with a 20-residue peptide substrate showed that two sites were glycosylated (**bold** with underline indicate glycosylation): APGSTAPPAHGVTSAPDT. Similar results were obtained with the 30-residue peptide. There was a reproducible preferential glycosylation of the GVTSA site over the GSTAP site (approximately 2-fold more radioactivity detected under these experimental conditions) on both 20- (see figure) and 30-residue (data not shown) peptides. The significant peaks of radioactivity detected in the later cycles of the 20-residue peptide (sequence, SAP) are most likely the result of a sequencing lag effect resulting from the accumulation of inefficiently cleaved residues from earlier Edman degradation cycles that are cleaved in subsequent cycles.

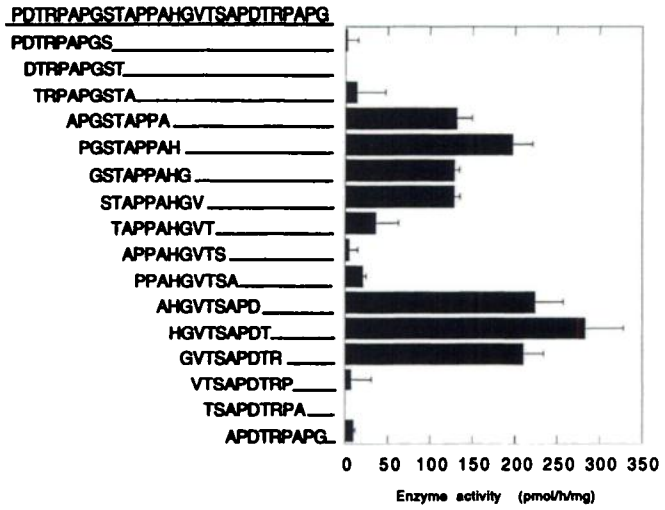

Fig. 3. Relative GalNAc transferase activity with a series of overlapping 9-residue MUC1 tandem repeat peptides as acceptor substrates. GalNAc transferase assays with a series of overlapping 9-residue peptides based on the MUC1 tandem repeat sequence (*ordinate*) as acceptor substrates and lysates of the human pancreatic tumor cell line BxPC-3 as the enzyme source were performed using standard assay conditions as described under "Materials and Methods." *Abscissa*, relative enzyme rates. Results are presented as mean of three independent experiments. *Bars*, SE.

GalNAc Transferase Activity with 9-Residue Peptides in a Panel of Cell Lines. Three peptide substrates (PDTRPAPGS, GSTAPPAHG, and GVTSAPDTR) were used to evaluate levels of enzyme activity in a number of tumor cell lysates (Fig. 4). There were two principal reasons for performing this series of experiments. The first was to evaluate the relative levels of enzyme activity among tumor cell lines derived from different organs and from tumors of different differentiation grades. The second was to confirm the observation that the PDTRPAPGS peptide was not glycosylated by a separate enzyme with distinct substrate specificity that may be expressed in some cell lines but not in others. No enzyme activity above background values was detected with peptide PDTRPAPGS in any of the tested cell lines, supporting the hypothesis that this site is not glycosylated. In all cell lines tested, peptide GVTSAPDTR showed relative levels of enzyme activity that were at least 2-fold higher than peptide GSTAPPAHG, consistent with the previous observation that the former is a better substrate (7). It is also noteworthy that the three breast tumor cell lines evaluated here showed higher relative levels of enzyme activity/mg of protein than any of the pancreatic cancer cell lines that were tested. The panel of pancreatic tumor cell lines tested here includes cell lines that were derived from tumors of distinct differentiation grades (12), including poorly differentiated (PANC-1, HGC 25), moderately differentiated (T3M4, BxPC-3, PANC-89), and well differentiated (HPAF, CAPAN-1). The two cell lines from poorly differentiated pancreatic tumors expressed lower levels of enzyme activity than cell lines from moderately differentiated or well differentiated tumors.

GalNAc Transferase Activity on Substituted 9-residue Peptides. Several 9-residue peptides, in which one or two amino acids were substituted into the sequence of the native 9-residue MUC1 peptide, were tested as substrates in the GalNAc transferase assay. These experiments were performed to confirm the positions of glycosylation on the 9-residue peptide substrates, and to evaluate the contribution of selected features of substrate primary amino acid sequence on GalNAc transferase activity. The results shown in Fig. 5 are for HPAF cell extract using standard assay conditions except for the acceptor substrate concentration, which was 1024 μM. Similar but proportionally smaller results were obtained with peptide concentrations of 64 μM (data not shown). Eleven peptides that contained substitutions relative to the native peptide sequence GVTSAPDTR were utilized for these assays (Fig. 5*A*), as were eight substituted peptides based on the sequence GSTAPPAHG (Fig. 5*B*), and two substituted peptides based on the sequence PDTRPAPGS (Fig. 5*C*).

Substitution of the third position threonine (relative to the amino terminus) with valine in the peptide GVVSAPDTR [the underlined amino acid indicates substitution(s)] (Fig. 5*A*) reduced enzyme activity to values that were indistinguishable from background (1.7% of the activity obtained with the native sequence peptide). Substitution of serine for this same threonine yielded a very low level of relative activity that was slightly above background (3.2% of activity obtained with the native sequence). Substitution of the eighth position threonine with valine or serine (GVTSAPDVR and GVTSAPDSR) resulted in a slightly enhanced relative enzyme activity (117.6 and 111.1%, respectively). Substitution of the first glycine with proline (PVTSAPDTR) showed no effect on relative enzyme activity. Interestingly, either substitution of the sixth proline (at the relative third position to the carboxyl side of threonine) with glycine (GVTSAGDTR) or deletion of the ninth arginine (the eight residue peptide, GVTSAPDT) significantly reduced relative activity (7.6% and 18.7%, respectively). A peptide with the same amino acid composition but a scrambled sequence (TARVPTGSD) showed no activity above background values (1.7%). Substitution of the third threonine with D-threonine (indicated as small "t"; GVtSAPDTR) also eliminated enzyme activity. An exchange of the third threonine with the second valine (GTVSAPDTR) yielded an ineffective substrate; however, an exchange of the third threonine with the fourth serine (GVSTAPDTR) yielded a substrate with low relative enzyme activity (4.1% of the activity obtained with the native peptide). Taken together, the loss of enzyme activity with peptide substrate GVVSAPDTR, which contains an isosteric substitution of valine for threonine at position 3, the loss of activity with the peptide containing the D-threonine substitution at position 3, and the high level of activity seen with peptide GVTSAPDVR, which contains an isosteric substitution of valine for threonine at

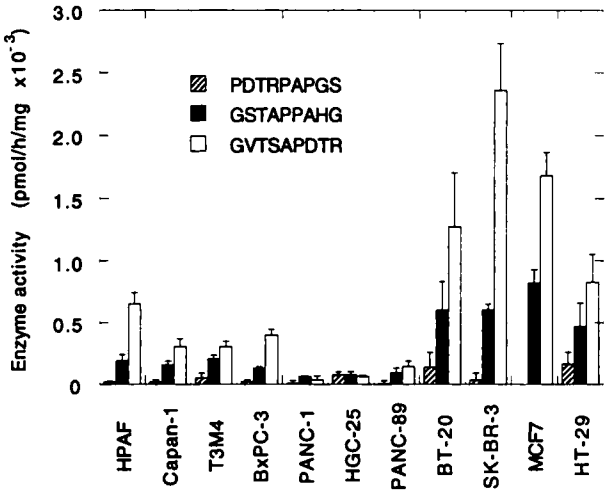

Fig. 4. Evaluation of GalNAc transferase activity in lysates of several pancreatic and breast tumor cell lines with selected 9-residue MUC1 tandem repeat peptides as acceptor substrates. GalNAc transferase assays with the 9-residue MUC1 tandem repeat peptides PDTRPAPGS, GSTAPPAHG, and GVTSAPDTR as acceptor substrates were performed (standard assay conditions described in "Materials and Methods") using lysates of several tumor cell lines as the enzyme source. The reported relative rates (*ordinate*) are the means (*bars*, SE) for three or more independent assays for each cell line. Pancreatic tumor cell lines, HPAF, Capan-1, T3M4, BxPc-3, PANC-1, HGC-25, and PANC-89; breast tumor cell lines, BT-20, SK-BR-3, and MCF7; colon tumor cell line, HT-29.

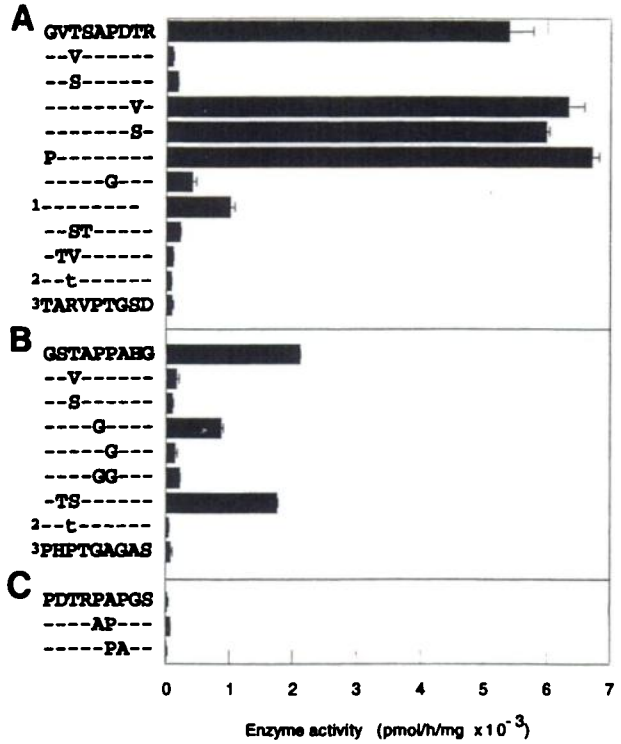

Fig. 5. Relative GalNAc transferase activity with 9-residue and 8-residue MUC1 tandem repeat peptides that contain sequence substitutions as acceptor substrates. Peptides evaluated as acceptor substrates included several 9-residue peptides and an 8-residue peptide that contained substitutions or a deletion relative to the sequence of the two native acceptor substrate peptides, GVTSAPDTR (*A*) and GSTAPPAHG (*B*). Synthetic peptides that contained substitutions relative to the native peptide PDTRPAPGS (*C*), which does not show detectable activity as a substrate, were also prepared and evaluated. Sequence positions were selected for substitution based on their relative position in relation to the position of glycosylation (see Fig. 2). Standard assay conditions were used as described in "Materials and Methods" with the following modifications: acceptor peptide concentrations, 1024 μM. *Dashes*, conservation of the native peptide sequence (*top* of each *panel*). The mean relative rate (*abscissa*) of three independent experiments is presented; *bars*, SE. *1*, 8-residue peptide, which lacks the carboxyl-terminal R found in the native sequence. In *2*, lower case t indicates D-threonine at that position. *3*, peptides with the same amino acid composition but a scrambled sequence as compared to the native sequence peptide.

position 8, confirm that this peptide is glycosylated at the position 3 threonine and not at the position 8 threonine or at the position 4 serine, in agreement with the results of sequencing the 20-residue glycosylated peptide (discussed above).

Relative enzyme activity for peptides with substitutions in the GSTAPPAHG sequence are shown in Fig. 5*B*. Substitution of the third position threonine with valine (GSVAPPAHG) or with serine (GSSAPPAHG) reduced enzyme activity with this substrate to values that were indistinguishable from background, supporting the assignment of glycosylation to threonine at this position by the sequencing studies with the 20-residue peptide (discussed above). Three peptides contained substitutions of glycine for the fifth and/or sixth proline residues (GSTAGPAHG, GSTAPGAHG, and GSTAGGAHG). Among these, only GSTAGPAHG showed significant enzyme activity (41.0% of that obtained with the native peptide). A peptide in which the second serine and the third threonine were exchanged (GTSAPPAHG) was an effective substrate (83.2% of the activity obtained with the native peptide). A peptide that contained a substitution of L-threonine with D-threonine (GStAPPAHG) and a peptide with the same amino acid composition as the native peptide but a scrambled sequence (PHPTGAGAS) were ineffective as substrates in these assays.

The peptides shown in Fig. 5*C* were designed to determine whether subtle changes in the sequence of the PDTRPAPGS peptide would convert it into an effective substrate. The substitutions were designed to evaluate the importance of the relative positions of proline in this peptide (proline is moved to the relative third position to the carboxyl side of threonine) and the potential negative influence of the bulky, charged residues aspartic acid and arginine at positions flanking the threonine. In the substituted peptides, the sixth alanine was exchanged with either the fifth or seventh proline. This produced substrates in which threonines, prolines, and alanines were in the same relative positions as the GSTAPPAHG or the GVTSAPDTR peptides described above. Neither the substituted peptides (PDTRPPAGS, PDTRAPPGS) nor the native peptide show activity above background values (Fig. 5*C*).

Effect of Acceptor Substrate Concentration on the Rate of the GalNAc Transferase. GalNAc transferase activity was evaluated with selected peptides at several substrate concentrations from 0.256 to 8.192 mM in assays that utilized HPAF cell lysate as the enzyme source (Fig. 6). The purpose of these experiments was to confirm enzyme activity in cases where the relative enzyme rates obtained with some of the substituted peptides were very low at low substrate concentrations. Other assay conditions were as described for the standard assay in "Materials and Methods." Enzyme rates were increased at increased substrate concentrations for the peptides GVTSAPDTR, GSTAPPAHG, and GTSAPPAHG, as expected. Increasing the concentrations of peptides GVSSAPDTR and GSSAPPAHG yielded slight increases in the reaction rates (Fig. 6), consistent with the observation of a low but reproducible relative rate that was obtained for these peptides at a concentration of 1024 μM (Fig. 5). In contrast, the use of peptide PDTRPAPGS as a substrate did not yield detectable enzyme rate data even at high substrate concentrations, confirming that this peptide is not an effective substrate (Fig. 6).

DISCUSSION

The studies described here were designed to determine the positions of *O*-linked glycosylation on peptides based on the MUC1 mucin core protein that were glycosylated *in vitro* by pancreatic adenocarcinomas. This question is of interest because certain epitopes on the tandem repeat portion of the MUC1 core protein, primarily those contained within and around the polypeptide sequence PDTRP, are

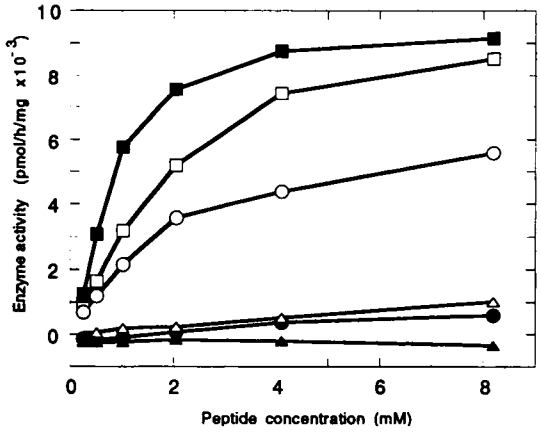

Fig. 6. GalNAc transferase activity in HPAF cell lysates with increasing concentrations of selected 9-residue peptides as acceptor substrates. The relative rate of GalNAc transferase activity in HPAF lysates was evaluated using as acceptor substrates different concentrations (4–1024 μM, *abscissa*) of the indicated 9-residue MUC1 tandem repeat peptides. Relative rates (*ordinate*) were calculated as described in "Materials and Methods." ■, GVTSAPDTR; □, GSTAPPAHG; ○, GTSAPPAHG; ●, GVSSAPDTR; △, GSSAPPAHG; ▲, PDTRPAPGS.

exposed on forms of this mucin produced by breast adenocarcinoma cells (3). These same epitopes are masked from antibody recognition when the mucin is expressed by corresponding normal breast tissues and by some adenocarcinomas derived from different organ sites, including the pancreas (3, 4). One hypothesis that would explain this differential antibody reactivity would be that breast and pancreatic tumor cells glycosylate the tandem repeat portion of the MUC1 protein at different positions. This hypothesis was tested in the work described in this article.

The results of testing peptide substrates with multiple tandem repeats (Fig. 1) support the use of short peptide sequences such as those described in this paper for evaluating the substrate specificity and relative rate of the human GalNAc transferase. One interpretation of these data is that higher levels of ordered secondary structure that may be created by increasing the number of tandem repeats (11) do not affect the relative rate of the GalNAc transferase for each unit of tandem repeat substrate. The linear correlation between tandem repeat number and relative enzyme rate supports the hypothesis that the glycosylation patterns of short tandem repeat peptides accurately reflect the glycosylation patterns found on molecules that contain longer tandem repeat components. Moreover, the hypothesis that the mucin primary amino acid sequence plays a significant role in determining the positions of *O*-linked glycosylation is also supported. These findings call into question previous proposals that the tandem repeat portions of mucins are glycosylated in a stochastic manner because they exist in a random coil conformation (13). Certainly, many structural features of mucins are conferred by the addition of carbohydrates to serine and threonine residues. It will be important to investigate further the relative contribution of both carbohydrate and protein moieties to mucin structure and function.

The results of sequencing peptides that were glycosylated *in vitro* by the HPAF pancreatic tumor cell line showed that these cell extracts glycosylated the peptides at the same positions (Fig. 2) as were reported previously for the MCF7 breast tumor cell line (7). This conclusion is supported by the results of the experiments shown in Figs. 3 and 5, which are essentially identical to results obtained with the breast tumor cell line (7). If these results accurately reflect the *in vivo* processes, they suggest that pancreatic and breast tumor cell lines express identical GalNAc transferase activities that recognize and glycosylate the same substrate sequences. Furthermore, the results described above and the results presented in Figs. 4 and 6 suggest that the polypeptide sequence PDTRP, which is the major epitope in this tumor-associated antigen, is not a site of glycosylation in human breast tumor cell lines, pancreatic tumor cell lines, or colon tumor cell lines.

Another interesting finding is the observation that extracts of breast tumor cell lines demonstrate higher relative levels of enzyme activity than extracts of pancreatic tumor cell lines (Fig. 4). Mucin produced by these breast adenocarcinoma cell lines is less heavily glycosylated (50% by weight) than mucin produced by the pancreatic tumor cell lines (80% by weight) (4, 14). Moreover, the breast tumor-derived mucin contains epitopes on the core protein that are not exposed on the pancreatic tumor-derived mucin until it is deglycosylated (4, 14). The fact that the breast tumor cell lines show higher levels of GalNAc transferase activity, together with the finding that there was no difference in the positions that were glycosylated using breast or pancreatic tumor cell extracts (discussed above), supports the hypothesis that the exposure of selected core protein epitopes on this mucin is not the result of the breast tumor cells glycosylating the MUC1 tandem repeat at fewer positions or at different positions. Instead, the differential accessibility of the core proteins to specific antibodies is probably explained by differences in the structure of oligosaccharides that are attached by different cell types. It is very important to point out, however, that the analyses of glycosylation presented here are necessarily limited to the enzyme activities detectable by *in vitro* assay. Thus, it is possible that the assays performed here will not detect distinct GalNAc transferase activities that are present in intact cells. Studies are under way to confirm these *in vitro* findings with *in vivo*-derived samples.

It is notable that two poorly differentiated pancreatic tumor cell lines (12) showed much lower levels of GalNAc transferase activity compared to most of the well differentiated or moderately differentiated cell lines; however, any general conclusions about relative levels of GalNAc transferase activity in secretory cells of distinct differentiation states remains to be established by future studies.

Our previous work showed that extracts from breast tumor cells glycosylate the two sites on the MUC1 tandem repeat at different relative rates, and that the primary amino acid sequence surrounding the glycosylated residue affects the specificity and relative rate of enzyme activity (7). The results in Figs. 3 and 5 confirm those observations in a pancreatic tumor cell line. The general significance of these findings as they relate to the activity of the GalNAc transferase has been discussed previously (7).

There are currently several laboratories evaluating the potential of using the MUC1 mucin core protein as a target for the development of improved diagnostic reagents and as a tumor vaccine reagent (15). Other laboratories and biotechnology companies (16) are interested in combining tumor-associated carbohydrate structures with specific mucin core proteins or peptides as vaccine reagents. Information about the positions of attachment of oligosaccharide may be critical to the design of these reagents. Future studies will address questions about differences in the posttranslational processing and structure of this mucin as it is expressed by different tumors and normal tissues and how these differences affect its functional properties.

ACKNOWLEDGMENTS

The authors thank Dr. Olivera Finn for providing some of the peptides used in this study and Dr. David Colcher for helpful discussions. The authors gratefully acknowledge the technical assistance of Thomas Caffrey, Jan Williamson, and Charles Troxel.

REFERENCES

1. Metzgar, R. S., Hollingsworth, M. A., and Kaufman, B. Pancreatic mucins. *In:* V. L. W. Go *et al.* (eds.), The Pancreas: Biology, Pathology, and Disease, pp. 351–367, Ed. 2. New York: Raven Press, 1993.
2. Singhal, A., and Hakomori, S. Molecular change in carbohydrate antigens associated with cancer. Bioessays, *12:* 223–230, 1990.
3. Gendler, S. J., Spicer, A. P., Lalani, E. N., Duhig, T., Peat, N., Burchell, J., Pemberton, L., Boshell, M., and Taylor-Papadimitriou, J. Structure and biology of a carcinoma-associated mucin, MUC1. Am. Rev. Respir. Dis., *144:* S42–S47, 1991.
4. Lan, M. S., Hollingsworth, M. A., and Metzgar, R. S. Polypeptide core of a human pancreatic tumor mucin antigen. Cancer Res., *50:* 2997–3001, 1990.
5. Barnd, D. L., Lan, M. S., Metzgar, R. S., and Finn, O. J. Specific, major histocompatibility complex-unrestricted recognition of tumor-associated mucins by human cytotoxic T-cells. Proc. Natl. Acad. Sci. USA, *86:* 7159–7163, 1989.
6. Jerome, K. R., Barnd, D. L., Bendt, K. M., Boyer, C. M., Taylor-Papadimitriou, J., McKenzie, I. F. C., Bast, R. C., and Finn, O. J. Cytotoxic T-lymphocytes derived from breast adenocarcinoma patients recognize an epitope present on the protein core of a mucin molecule preferentially expressed by malignant cells. Cancer Res., *51:* 2908–2916, 1991.
7. Nishimori, I., Johnson, N. R., Sanderson, S. D., Perini, F., Mountjoy, K., Cerny, R. L., Gross, M. L., and Hollingsworth, M. A. The influence of acceptor substrate primary amino acid sequence on the activity of human uridine diphosphate-*N*-acetylgalactosamine:polypeptide *N*-acetylgalactosaminyltransferase: studies with the MUC1 tandem repeat. J. Biol. Chem., *269:* 16123–16130, 1994.
8. Ember, J. A., Sanderson, S. D., Taylor, S. M., Kawahara, M., and Hugli, T. E. Biologic activity of synthetic analogues of C5a anaphylatoxin. J. Immunol., *148:* 3165–3173, 1992.
9. Fontenot, J. D., Finn, O. J., Dales, N., Andrews, P. C., and Montelaro, R. C. Synthesis

of large multideterminant peptide immunogens using a poly-proline b-turn helix motif. Pept. Res., *6:* 330–336, 1993.

10. Damme, E. M., Kaku, H., Perini, F., Goldstein, I. J., Peeters, B., Yagi, F., Decock, B., and Peumans, W. J. Biosynthesis, primary structure and molecular cloning of snowdrop (*Galanthus nivalis* L.) lectin. Eur. J. Biochem., *202:* 23–30, 1991.
11. Fontenot, J. D., Tjandra, N., Bu, D., Ho, C., Montelaro, R. C., and Finn, O. J. Biophysical characterization of one-, two-, and three-tandem repeats of human mucin (muc-1) protein core. Cancer Res., *53:* 5386–5394, 1993.
12. Hollingsworth, M. A., Strawhecker, J. M., Caffrey, T. C., and Mack, D. R. Expression of MUC1, MUC2, MUC3, and MUC4 mucin mRNAs in human pancreatic and intestinal tumor cell lines. Int. J. Cancer, *57:* 198–203, 1994.
13. Jentoft, N. Why are proteins o-glycosylated? Trends Biochem. Sci., *15:* 291–294, 1990.
14. Lan, M. S., Batra, S. K., Qi, W. N., Metzgar, R. S., and Hollingsworth, M. A. Cloning and sequencing of a human pancreatic tumor mucin cDNA. J. Biol. Chem., *265:* 15294–15299, 1990.
15. Hareuveni, M., Gautier, C., Kieny, M., Wreschner, D., Chambon, P., and Lathe, R. Vaccination against tumor cells expressing breast cancer epithelial tumor antigen. Proc. Natl. Acad. Sci. USA, *87:* 9498–9502, 1990.
16. Fung, P. Y. S., Madej, M., Koganty, R. R., and Longenecker, B. M. Active specific immunotherapy of a murine mammary adenocarcinoma using a synthetic tumor-associated glycoconjugate. Cancer Res., *50:* 4308–4314, 1990.