Recombinant MUC1 probe authentically reflects cell-specific O-glycosylation profiles of endogenous breast cancer mucin: High-density and prevalent core2-based glycosylation

Stefan Müller and Franz-Georg Hanisch

Institute of Biochemistry II, Medical Faculty of the University, Joseph-Stelzmann-Str. 52, D-50931 Köln, Germany

Key words: O-glycosylation, MUC1, breast cancer

Running title: MUC1 O-glycosylation in breast cancer

Corresponding address:

Dr. Stefan Müller and
Prof. Dr. Franz-Georg Hanisch
Institut für Biochemie II
der Medizinischen Fakultät
Joseph-Stelzmann-str. 52
50931 Köln
Germany
Tel: +49 221 478 4493
Fax: +49 221 478 6977
e-mail: stefan.mueller@uni-koeln.de
       franz.hanisch@uni-koeln.de
Abstract

Knowledge on the O-linked glycan chains of tumor-associated MUC1 is primarily based on enzymatic and immunochemical evidence. To obtain structural information and to overcome limitations by the scarcity of endogenous mucin, we expressed a recombinant glycosylation probe corresponding to six MUC1 tandem repeats in four breast cancer cell lines. Comparative analyses of the O-glycan profiles were performed after hydrazinolysis and normal-phase chromatography of 2-aminobenzamide labeled glycans. Except for a general reduction in the O-glycan chain lengths and a high density glycosylation, no common structural pattern was revealed. T47D fusion protein exhibits an almost complete shift from core2 to core1 expression with a preponderance of sialylated glycans. By contrast, MCF-7, MDA-MB231, and ZR75-1 cells glycosylate the MUC1 repeat peptide preferentially with core2-based glycans terminating mostly with α3-linked sialic acid (MDA-MB231, ZR75-1) or α2/3-linked fucose (MCF-7). Endogenous MUC1 from T47D and MCF-7 cell supernatants revealed almost identical O-glycosylation profiles compared to the respective recombinant probes indicating that the fusion proteins reflected the authentic O-glycan profiles of the cells. The structural patterns in the majority of cells under study are in conflict with biosynthetic models of MUC1 O-glycosylation in breast cancer, which claim that the truncation of normal core2-based polylactosamine structures to short sialylated core1-based glycans is due to the reduced activity of core2-forming β6-N-acetyl-glucosaminyltransferases and/or to overexpression of competitive α3-sialyltransferase.
The cell membrane-associated human mucin MUC1 is expressed by most epithelia (1) and some subsets of lymphocytes (2, 3). It is overexpressed by many carcinomas and an altered glycosylation pattern results in tumor-specific exposure of peptide epitopes (4), making MUC1 a promising tumor antigen with diagnostic, as well as therapeutic potential in the treatment of cancer (5-7).

Mature MUC1 consists of two subunits that are proteolytically derived from a common precursor peptide and form a stable heterodimeric complex. The smaller subunit contains a C-terminal cytoplasmic domain, the membrane spanning domain and a short extracellular sequence that is non-covalently linked to the larger, extracellular subunit which contains the extensively O-glycosylated mucin domain (8). After its first appearance on the cell membrane the complex is internalized and recycled several times for follow-up sialylation in the trans-Golgi (9). Mature glycoforms of the mucin can remain on the cell surface or become shed by still unknown mechanisms (9).

MUC1 contains five potential N-glycosylation sites, three of which are located in the membrane-associated subunit and two near the C-terminus of the extracellular subunit. However, the bulk of glycosylation, which can make up between 50 to 80 % of the total mass, is O-linked to numerous threonine and serine residues in the mucin domain. This domain comprises a variable number of tandemly repeated 20 amino acid sequences, each containing five potential O-glycosylation sites (10, 11). Although each of these sites is an O-glycosylation target \textit{in vivo}, the average density of glycans may vary considerably among MUC1 glycoforms. As an instructive example, tandem repeat glycopeptides that were prepared from MUC1 expressed in the lactating breast (12) or from the breast cancer cell line T47D (13), have been demonstrated to contain an average of 2.6 and 4.8 glycans, respectively.
The pool of O-glycan structures produced by a single cell is the product of a complex biosynthetic process, which is not template guided and requires the ordered action of multiple glycosyltransferases. Accordingly, O-glycan patterns are often cell and tissue specific and may differ substantially from one cell type to another. The O-glycan profiles of MUC1 glycoforms from breast milk, urine and two breast carcinoma cell lines have been investigated so far. The O-glycan profile of lactation-associated MUC1 is dominated by core2 based linear or branched polylactosamine chains, which are substituted with up to three fucose residues (14). Mono- and disialylated structures were also present but accounted for less than 25% of total glycans (15). Urinary MUC1, in contrast, exhibits an O-glycan profile with significantly shorter neutral and acidic glycans, which are based on core1 as well as core2 structures (16). According to three reports, tumor-associated glycoforms, which had been isolated from T47D (17) and BT 20 (18) breast carcinoma cell line supernatants or from cell lysates (19), were demonstrated to contain truncated precursor structures like core-GalNAc or the core1 disaccharide Gal(β1-3)GalNAc as well as its mono- and disialylated derivatives.

It has been proposed, that breast-cancer associated changes in MUC1 O-glycosylation reflect a general switch of core2 to core1 expression, which leads to a reduction of O-glycan chain length (18) and may be the result of decreased or lacking expression of core2 specific β6 N-acetylgalcosaminytransferases (20). In addition, an increased α2,3-sialylation of core1 structures has been claimed to represent a biosynthetic stop signal and to inhibit core2 formation (21). However, due to the limited availability of chemical data, these hypotheses rely mainly on indirect enzymatic and immunochemical evidence, including the analysis of glycosyltransferase expression patterns by in situ hybridisations (22), northern blot analysis or enzyme activity assays (20). Immunochemical analysis of O-glycosylation inhibited cells (23) revealed an increased binding of tumor-specific
monoclonal antibodies like SM3 or HMFG-1, supporting the idea that tumor-associated MUC1 is underglycosylated with regard to the average glycan chain lengths.

Increased levels of MUC1 in the sera of cancer patients (24) as well as tumor-associated truncation of O-glycans, render MUC1 tandem repeat peptide epitopes accessible to the immune system. Humoral and cellular responses have been demonstrated in cancer patients (25, 26), but also in pregnant women (27) and healthy individuals (24). Although these natural responses are usually insufficient to fight the progress of malignant diseases, MUC1 derived peptides or glycopeptides are used in clinical trials to trigger therapeutically and prophylactically efficient immune reactions in humans (28, 29).

To enable the design of efficient tumor vaccines the cancer-associated glycosylation profiles on MUC1 repeat peptides need to be analysed in more detail and with state-of-the-art methodologies. Chemical studies on MUC1 O-glycosylation in cancer that have been reported so far, suffered from severe limitations of sample amounts. To overcome the analytical problems that are associated with low sample amounts and to extend the data on MUC1 O-glycosylation in mammary carcinomas by chemical evidence, we have expressed a MUC1 fusion protein in the breast cancer cell lines ZR75-1, MDA-MB 231, MCF-7 and T47D. O-glycan pools were released from the purified fusion proteins by hydrazinolysis and the 2-aminobenzamide labeled glycans were profiled by normal-phase HPLC. Complementary information on the O-glycan profiles was revealed by an independent approach using non-reductive β-elimination and methylation of the oligosaccharides followed by their mass spectrometric analysis. To confirm that the recombinant probes reflect the authentic glycosylation profiles of the individual cell lines, we analysed in two cases also the endogenous mucin by applying mass spectrometric profiling.
Materials and Methods

Construction of the MUC1 fusion protein (MFP6)

A MUC1 tandem repeat sequence containing terminal restriction sites was constructed by annealing and ligating four 5’ phosphorylated 60mer oligonucleotides. 5’ overhangs were filled up with Klenow enzyme and the construct was cloned into the *BamH I* and *Hind III* sites of pBluescript SK (Stratagene, Heidelberg, Germany) resulting in pBS TR1. Since there is a single *Sma I* site in each tandem repeat sequence, limited *Sma I* digestion of a 3.4 kb *Hind III* / *Apa I* cleaved MUC1 cDNA resulted in a series of fragments of 2-40 tandem repeat sequences and migrating in the 0.2-3 kb range on a 1.6 % agarose gel. Fragments ranging from 0.8-1.8 kb were eluted from the gel and ligated into *Sma I* / *Hind III* cleaved pBS TR1. Out of several clones containing 10 – 30 tandem repeats and the 3´ sequence of the original construct, a single clone with approximately 25 tandem repeats was chosen. *BamH I* digestion followed by limited *Sma I* digestion removed the 5´sequence of the MUC1 cDNA and a random number of tandem repeats and resulted in several fragments, migrating between 3.5 and 5.5 kb on 1.2 % agarose gels. Fragments in the 4 kb range were eluted from the gel and ligated to a 300 bp *Sma I* / *Not I* fragment that was exised from pBS TR1 and represented the 5´sequence of pBSTR1. The procedure resulted in several clones that contained 2-16 tandem repeats flanked by the 3´and 5´sequences of the original pBS TR1. A clone containing 6 tandem repeats was choosen and subcloned into the pCEP-PU expression vector (30) using the 5´ *Nhe I* and the 3´ *Not I* sites. pCEP-PU already contained the signal peptide of the BM40 extracellular matrix protein, followed by a hexa-histidine sequence, a myc tag and the *Nhe I* and *Not I* restriction sites that were used for in-frame insertion of the MUC1 tandem repeat construct. Restriction enzymes and all other DNA modifying enzymes were obtained from New England Biolabs, Frankfurt am Main, Germany.
Cell culture

The breast cancer cell lines MDA-MB231, T47D, MCF-7 and ZR75-1 were obtained from the American Type Culture Collection and were cultured in 80 or 160 cm$^2$ flasks at 37 °C in the presence of 5 % CO$_2$. Minimal essential medium, supplemented with Glutamax I, 10 % fetal calf serum, 0.1 mM sodium pyruvate, 100 i.U. Penicillin and 100 µg/ml Streptomycin was used for MDA-MB231, MCF-7 and ZR75-1. T47D cells were grown in RPMI 1640, supplemented as described above and containing 10 µg/ml bovine insulin. After transfection with MFP6 0.1 µg/ml Puromycin were added to all media. Media were exchanged every 3-4 days and the cells were passaged when they reached 80 - 90 % confluency using 0.1 % trypsin and 0.2 mM EDTA. Puromycin and insulin were from Sigma, Taufkirchen, Germany, all other cell culture reagents were obtained from Life Technologies, Karlsruhe, Germany.

Expression of MFP6

pCEP-PU (30) is a mammalian episomal expression vector, containing a puromycin resistance gene and an origin of replication that allows episomal propagation in mammalian cells. Protein expression is driven by the cytomegalovirus promoter. All cell lines were transfected using the Superfect (Quiagen, Hilden, Germany) lipofection reagent according to the manufacturer’s instructions. Cells were allowed to recover in standard media and 0.1 µg/ml puromycin were added after 24 h. Non-transfected cells died within five days and the transfectants were expanded and cultured in the presence of 0.1 µg/ml puromycin to keep selective pressure on the episomal vector.
Purification of MFP6

Conditioned supernatants from confluent cell layers were collected, centrifuged at 1000 x g at 4 °C for 10 min and dialyzed against several changes of demineralized water. Dialyzed supernatants were adjusted to 50 mM sodium phosphate pH 8.0, 200 mM sodium chloride, 1 mM imidazol, 5 mM 2-mercaptoethanol and 10 % ethanol and centrifuged at 20,000 x g (4 °C) for 45 min. Up to 1 L of the adjusted supernatant was loaded onto a column of 2 ml Ni-NTA Superflow (Quiagen, Hilden, Germany). The column was washed with 30 ml 50 mM sodium phosphate pH 6.5, 500 mM sodium chloride, 10 mM imidazol, 10 mM 2-mercaptoethanol and 10 % ethanol. After equilibrating the column with 5 ml 20 mM sodium phosphate, pH 6.5, proteins were eluted with 8 x 1 ml 0.2 M trifluoroacetic acid (TFA) in 10 % acetonitril. Fractions 2 - 5 appeared to contain the fusion proteins and were subjected to HPLC purification on a reversed-phase column (Vydac 208TP1015, MZ Analysentechnik, Mainz, Germany). Samples were injected in 500 µl aliquots and the column was eluted with a gradient of 2 to 80 % acetonitril in 0.1 % TFA over 30 min. A flow-rate of 1 ml/min was used and the chromatogram was registered spectrophotometrically at 214 nm. Peaks were collected manually and the quality of the preparations was analyzed by sodium dodecyl sulfate polyacrylamide gel electrophoresis. The gels were stained with silver or blotted onto nitrocellulose membranes and the fusion proteins were detected with an anti-myc monoclonal antibody (Santa Cruz Biotechnology, Heidelberg, Germany). Purified MFP6 was quantified by its extinction at 280 nm.

Purification of endogenous mucin

Secretory MUC1 was isolated from the supernatants of T47D and MCF-7 cells by affinity chromatography on anti-MUC1 antibody columns (1 mg of a mixture of repeat peptide-specific monoclonal antibodies B27.29 and BW835 coupled covalently to NHS-activated...
HighTrap columns from Amersham Biosciences) as described previously (13). The MUC1 containing fractions which could be registered by enzyme immunoassay were collected, dialyzed against dilute ammonium hydrogencarbonate, pH 8.0 and dried by vacuum centrifugation prior to gel permeation chromatography. Final purification of the mucin (solubilized in 4 M guanidinium hydrochloride) was achieved by size exclusion on a Superdex 200 HR column (1x30 cm, Amersham Biosciences) equilibrated and run in 0.1 M ammonium hydrogencarbonate buffer, pH 8.0 using a flow rate of 0.5 ml/min. Purity of the mucin containing fractions monitored by enzyme immunoassay was finally checked after western blotting and staining of the nitrocellulose membrane with the Glycan Detection Kit (Roche, Mannheim, Germany).

**Hydrazinolysis**

20 to 50 µg of lyophilized MFP6 were subjected to hydrazinolysis (31), using the O-glycan release kit (Glyko, Novato, CA) according to the manufacturer’s instructions, with some minor modifications. Briefly, the treatment of extensively dried, salt free fusionproteins was performed for 5 h at 60 °C, followed by evaporation of the reagent under reduced pressure. After re-N-acetylation for 15 min at 4 °C using 2 mM acetic anhydride in ice cold saturated bicarbonate, the reaction mixture was desalted with the supplied ion exchange resin in a batch procedure and the glycans were dried by vacuum centrifugation.

**2-Aminobenzamide labeling of O-glycan pools**

2-Aminobenzamide (Fluka, Seelze, Germany) was 2 x recristallized from ethanol and stored under argon at -20°C. A detailed description of the labeling chemistry is given in (32). Briefly, dry glycans in 0.5 ml microcentrifuge tubes were mixed with 2 µl 1 M
2-aminobenzamide in acetic acid and 3 µl 2 M sodium cyanoborohydride in dimethyl sulfoxide. After a 2 h incubation at 60 °C the samples were spotted onto chromatography paper (Schleicher & Schüßl, Dassel, Germany) and excessive reagents were removed by ascending paper chromatography in n-butanol : ethanol : water (4:1:1). The application points were cut out and eluted with 500 µl of water in order to recover the labeled glycans which do not migrate under the chromatographic conditions described above. Eluted fractions were filtered through 0.2 µm membranes in centrifugal microfiltration devices and stored at -20°C.

Preparation of 2AB labeled standards

Lacto-N-tetraose, 3-sialyllactosamine and dextran hydrolysate were obtained from Glyko, Novato, CA. NeuAcα2,3Galβ1-3GalNAc and NeuAcα2,3Galβ1-3(NeuAcα2-6)GalNAc were prepared from human fetuin (Glyko, Novato, CA) and human glycophorin A (Sigma, Taufkirchen, Germany). Galβ1-4GlcNAcβ(Galβ1-3)GalNAc was prepared from MUC1 isolated from human milk as described (14). Hydrazinolysis of the glycoproteins and the mucin, as well as 2AB labeling was performed as described above.

Normal phase-HPLC of 2-AB labeled O-glycans

HPLC runs were performed on a System Gold HPLC work station (Beckman Instruments, Muenchen, Germany) equipped with a Shimadzu RF-10AXL fluorescence detector and using the Beckman Nouveau software for data acquisition. Wavelength settings on the detector were 330 nm for excitation and 420 nm for emission. Aliquots of 2-AB labeled glycans were dried in a centrifugal evaporator. Prior to analysis the samples were dissolved in 75 % acetonitril in water and 20 µl were injected onto a polymer-based
aminophase column (Astec NH₂ polymer, 5 µm, 4.6 x 250 mm, Alltech, Unterhaching, Germany). The column was eluted with a linear gradient of 80 % acetonitril / 20 % 250 mM ammonium formiate to 40 % acetonitril / 60 % 250 mM ammonium formiate over 80 min at a flow rate of 0.5 ml/min.

Anionexchange HPLC of 2-AB labeled O-glycans

Anionexchange HPLC was performed according to (33). The HPLC system described above was used with a polymer-based anion exchange column (Q HyperD 10, 10 µm 4.6 x 100 mM, Beckman Instruments, Muenchen, Germany) at a flow rate of 1 ml/min. Buffer A was water, buffer B was 500 mM ammonium formiate pH 9.0. A gradient of 0 % B for 1 min, 0-5 % B over 12 min, 5-21 % B over 13 min 21 % – 80 % over 25 min followed by 80 % to 100 % over 4 min was used. Samples were dissolved in water and 20 µl were injected.

Exoglycosidase digestion of 2AB labeled glycans

Neuraminidase (Clostridium perfringens), α2-3 sialidase (Salmonella typhimurium), α1-3/4 fucosidase (Xanthomonas manihotis) and α1-2 fucosidase (Xanthomonas manihotis) were obtained from New England Biolabs, Frankfurt am Main, Germany. Digestion of glycan pools or isolated glycans was performed in 50 mM sodium citrate pH 4.5 at 37 °C over night.

Preparation of partially deglycosylated tandem repeat peptides

10 µg of MFP6 was partially deglycosylated by sequential treatment with neuraminidase (Clostridium perfringens), α1-3/4 fucosidase (Xanthomonas manihotis), α1-2 fucosidase
(Xanthomonas manihotis), β-galactosidase (bovine testes, Glyko, Novato, CA) and β-hexosaminidase (jack bean, Glyko, Novato, CA) in 50 mM sodium citrate pH 4.5. Galactosidase and β-hexosaminidase treatment was repeated once. The samples were desalted by gel filtration on Sephadex G25 (NAP5, Amersham Pharmacia Biotech, Freiburg, Germany) in water and dried by vacuum centrifugation. Samples were solubilized in 100 µl 20 mM sodium phosphate, pH 7.5, 0.2 mM calcium acetate and 5 mM dithiotreitol. Clostripain (Sigma, Taufkirchen, Germany) was activated in 5 mM dithiotreitol, 1 mM calcium acetate for 3 h at room temperature and the samples were digested overnight at 37 °C with 2 µg of enzyme. The digests were subjected to reversed-phase chromatography on a C18 column (ODS Ultrasphere 5U, 2.0 x 150 mm, Beckman Instruments, Muenchen, Germany) as described previously (12) and aliquots of 0.3 ml fractions were analyzed by MALDI-TOF mass spectrometry. Tandem repeat glycopeptide containing fractions were combined and stored frozen.

**MALDI-TOF mass spectrometry of partially deglycosylated tandem repeat peptides**

Aliquots (10 µl) of HPLC fractions were mixed with an equal volume of a saturated solution of α-cyano 4-hydroxycinnamic acid (Sigma, Taufkirchen, Germany) in 2:1 acetonitril-water (0.1% TFA). 1 µl of each sample was applied to a polished stainless steel target and air dried. Reflectron MALDI-TOF mass spectrometry was performed on a Bruker Reflex III instrument (Bruker Daltonik, Bremen, Germany) in the positive ion mode using a pulsed laser beam (nitrogen laser, λ = 337 nm). The acceleration voltage was set to 20 kV , the reflector voltage was 22,5 kV.
Mass spectrometric analysis of methylated O-glycans

MPF6 (10 - 20 µg) or appr. 5 µg of endogenous MUC1 were dried by vacuum centrifugation and suspended under argon in dry, ultrapure dimethyl sulfoxide (Aldrich, Steinheim, Germany). After addition of 0.5 vol. methyl iodide (Fluka, Seelze, Germany), the samples were immediately mixed with one volume of a fine suspension of sodium hydroxide in DMSO (25 mg / ml) and incubated overnight at ambient temperature. The reaction mixture was extracted from chloroform with water (3 x 0.5 vol.) and the chloroform phase was dried in a stream of nitrogen. The methylated oligosaccharides were solubilized in 20 µl of methanol and 0.4 µl aliquots were mixed with 1.6 µl matrix solution (dihydroxy benzoic acid, 10 mg / ml of acetonitril-aqueous 0.1% TFA, 1:2, v/v). Reflectron MALDI-TOF mass spectrometry was performed on a Bruker Reflex III or IV instrument (Bruker Daltonik, Bremen, Germany) in the positive ion mode using a pulsed laser beam (nitrogen laser, λ = 337 nm). The acceleration voltage was set to 20 kV, the reflector voltage was 22,5 kV.
Results

Generation of a fusion protein containing six MUC1 tandem repeats (MFP6)

Fusion proteins containing the secretory signal peptide of the BM40 extracellular matrix protein and 2–16 tandem repeats of a MUC1 cDNA were constructed. A hexa-histidine and a myc tag were introduced to allow glycosylation independent affinity purification and immuno detection. MFP6 is a clone with six tandem repeats and was choosen as a glycosylation probe (Fig. 1A), since we noticed that fusion proteins with smaller repeat numbers were expressed poorly in the breast cancer cell lines used in this study (data not shown) and substantially larger ones were expected to be difficult to chromatograph in reversed-phase HPLC. DNA sequencing of the MFP6 construct confirmed the expected sequence given in figure 1A. Several tandem repeat sequences in this construct deviate from the ´conserved´ sequence in two positions, namely the PDTR (PESR) and PPAH (PAAH; PQAH) motive. These sequence variants originate from the MUC1 cDNA and represent a general sequence polymorphism that has been described previously (11,13,34).

Expression and purification of MFP6

The breast carcinoma cell lines T47D, MCF-7, MDA-MB231 and ZR75-1 were transfected with the episomal expression vector pCEP-PU containing the MFP6 construct. Westernblot analysis using an anti-myc monoclonal antibody revealed that MFP6 was expressed and secreted into the supernatant in the four cell lines (data not shown). Expression remained constant over several weeks and passages, as long as selective pressure was applied by culturing the cells in the presence of puromycin. Since serum reduction or depletion might have resulted in an artificial glycosylation pattern, 10 % fetal calf serum were always
included in the media. MFP6 was isolated from the conditioned supernatants by affinity chromatography on immobilized Ni\(^{2+}\) ions and subsequent reversed-phase chromatography on a C8-silica column. Prior to further analysis, the MFP6 preparations were rechromatographed on the same column in order to confirm the expected quantities and the homogeneity of the sample, which was checked by gel electrophoresis and silver staining combined with western blot analysis (Fig. 1B). According to photometric quantification of pure MFP6 the yields ranged from 0.5 to 2.0 mg protein per liter of conditioned supernatant.

**Purification of endogenous MUC1**

Endogenous MUC1 was eluted from the antibody affinity column at low pH and appeared in the first 3 – 6 fractions according to enzyme immunoassay detection with anti-MUC1 antibody B27.29. After dialysis and concentration by vacuum centrifugation the collected fractions were applied onto a gel permeation FPLC column with an exclusion limit of >1000 kDa and the mucin was effectively separated from low molecular weight proteins and peptides eluting at \(V_E > 2.5V_0\) according to the UV profiles at 280 nm. MUC1 positive fractions were identified by western blot analysis with B27.29 antibody (Fig. 1C), and the absence of contaminating glycoproteins was verified by using digoxigenin labeling of protein-bound glycans (Fig. 1C).

**HPLC analysis of 2AB labeled glycans reveals cell-specific O-glycan profiles**

The complete O-glycan pools were released from MFP6 by hydrazinolysis and the reducing ends of the glycans were labeled with the fluorescent dye 2-aminobenzamide
using reductive amination chemistry. The O-glycan pools were profiled by anion exchange HPLC and normal-phase HPLC on an aminophase column. Structures were identified by comparison to known standards and by sequential degradation with exoglycosidases. A comprehensive list of the assigned structures and their relative amounts is given in table 1.

According to column calibration with the 2AB derivatives of neutral (lacto-N-tetraose), monosialo (NeuAcα2-3Galβ1-3GlcNAc, NeuAcα2-3Galβ1-3GalNAc, NeuAcα2-6(Galβ1-3)GalNAc) and disialo (NeuAcα2-6(NeuAcα2-3Galβ1-3)GalNAc) glycans, the anion exchange chromatograms can be divided into roughly three segments comprising clusters of neutral, mono- and disialylated O-glycans. While neutral glycans eluted as a single peak with the break through fraction, seven mono- and disialylated species were well resolved. The corresponding structures and the relative amounts are listed in table 1. Glycans containing more than two sialic acids or inorganic modifications like sulfate or phosphate were expected to elute later in the gradient with salt concentrations higher than 300 mM (33) and were not detected on the fusion proteins expressed in any of the breast carcinoma cell lines.

Two alternative buffer systems were used in normal phase HPLC. A high salt buffer system (250 mM ammonium formiate in buffer B) suppresses ionic interaction of acidic glycans with the positively charged amino groups of the resin and was routinely used for the simultaneous separations of neutral, mono- and disialylated structures in a single run (Fig. 2A). Under these conditions the glycan pools were resolved into more than eleven different species. Digestion with neuraminidase (C. perfringens) which cleaves α2-3 as well as α2-6 linked sialic acids, identified the acidic glycans and revealed the underlying neutral structures, which were assigned as GalNAc, Galβ1-3GalNAc (peak 1) and Galβ1-4GlcNAcβ1-6(Galβ1-3)GalNAc (peak 2; refer to figures 2 A and 3 A) according to external
standards that were prepared from synthetic glycopeptides or commercially available glycoproteins. Peaks 3 and 5 appeared to be resistant to treatment with α2-3 sialidase (S. typhimurium), but digestion with neuraminidase (C. perfringens) resulted in the formation of GalNAc and Galβ1-3GalNAc, respectively. Accordingly, peak 3 was identified as NeuAcα2-6GalNAc and peak 5 as NeuAcα2-6(Galβ1-3)GalNAc, in accordance with the retention times of authentic, mass spectrometrically confirmed standard compounds. Similarly, peak 4 was assigned as NeuAcα2-3Galβ1-3GalNAc according to a standard trisaccharide and in agreement with this assignment the desialylation with α2,3 sialidase resulted in a peak isographic with Galβ1-3GalNAc. Limited α2-3 sialidase digestion of peak 7 yielded peaks 6 and 2, complete digestion resulted in peak 2 only and, hence, peak 7 was assigned as NeuAcα2-3Galβ1-4GlcNAcβ1-6(NeuAcα2,3Galβ1-3)GalNAc. Peak 8 was identified as NeuAcα2-6(NeuAcα2-3Galβ1-3)GalNAc, since it coeluted with an external standard that was prepared from human glycophorin A. This interpretation was supported by digestion with α2,3 sialidase (S. typhimurium) and neuraminidase (C. perfringens), which resulted in peaks isographic with NeuAcα2-6(Galβ1-3)GalNAc and Galβ1-3GalNAc, respectively. Peaks 6a and 6b were both shifted to peak 2 by digestion with α2,3 sialidase (S. typhimurium) and therefore should represent the two possible positional isomers resulting from α2,3 monosialylation of a core2 tetrasaccharide.

To analyze minor neutral constituents in the MCF-7 profile after enzymatic desialylation, we used an alternative low salt buffer system (50 mM ammonium formiate in buffer B). The use of this system allowed the analysis of fucosylated structures, since residual monosialylated glycans were shifted to higher retention times (refer to peaks 7 and 8 in figure 3B). Peak 10 was sensitive to α1,3/4 fucosidase (X. manihotis) and digestion yielded a peak isographic with the core2 tetrasaccharide implying that peak 10
corresponds to the core2-based Lewis$^X$ structure Galβ1-4(Fucα1-3)GlcNAcβ1-6(Galβ1-3)GalNAc. α1,2 fucosidase (X. manihotis) digestion of peak 9 revealed that this structure is also derived from the core2 tetrasaccharide, with a single α-fucose residue linked to position C2 of one of the two terminal galactose residues. Peak 11 represents a difucosylated core2 structure corresponding to Lewis$^Y$, since α1,2 fucosidase (X. manihotis) cleaved off one residue and generated the Lewis$^X$ pentasaccharide (peak 10). Several minor peaks eluted in the 45 to 55 min range and were also sensitive to fucosidase treatment. Concerted digestion with α1,2 and α1,3/4 fucosidase (X. manihotis) removed these minor signals and led to a corresponding increase of peak 12, which was assigend as core2-based structure with two lactosamine units on the β1-6 branch. The spiked shape of this peak implies a heterogeneity, which is probably due to the presence of both, type I and type II lactosamine units. Although it was not possible to identify the fucosylated derivatives of this structure in detail, it is evident, that highly fucosylated polylactosamine-type glycans are present and account for approximately 5% of the O-glycans that were derived from MFP6 expressed in MCF-7 cells. Note that fucosidase treatment did not affect any of the sialylated peaks, that elute at 60 and 70 min, indicating that sialylated Lewis structures are not present or below the limit of detection. Moreover, fucosidase treatment did not affect the profiles of any of the other fusion proteins, implying that MCF-7 exhibits an unusual O-glycosylation pattern among the breast carcinoma cell lines analyzed in this study.

Summarizing the structural data in table 1, which were derived from the corresponding profiles in figure 2A, the O-glycans on MUC1 fusion protein revealed cell-specific expression patterns. Core2-based structures accounted for less than 5% of the T47D derived glycans and the most prominent species of this cell line were represented by the α2,3 (to Gal) and α2,6 (to GalNAc) sialylated trisaccharides. Neutral glycans were of minor
abundance and no fucosylated species were detected. A completely opposite pattern was revealed for MDA-MB231 and ZR75-1 cells, which primarily expressed sialylated glycans based on core 2. The structures were generally larger and more complex and terminated preferentially with α3 (to Gal) linked sialic acid. While the profiles were qualitatively similar, the two cell lines exhibited quantitative differences with respect to the contribution of either core-type and the relative amount of sialylated species. It is noteworthy, that high proportions of core 2-based glycans are associated with a high degree of α2,3 sialylation in the ZR75-1 profile. A distinct and unique profile was registered for MCF-7 fusion protein, which carried primarily core2-based, more extended neutral glycans terminating in a considerable portion with α2/3-linked fucose.

In case of T47D-MFP6 the results were in agreement with our previous profiling studies on the O-linked glycans analysed as alditols after reductive cleavage from endogenous MUC1 isolated from cell supernatants (17). Attempts to liberate O-glycans from endogenous mucin by applying hydrazinolysis failed despite extensive desalting and drying of the samples. These findings are in accordance with experience from other laboratories (31), in particular when small amounts of mucin in the low microgram range are treated.

**MALDI-TOF analysis of permethylated O-glycan pools**

Independent information on the profiles of O-linked chains and confirmation of structural assignments made on the basis of chromatographic criteria (see table 1) came from mass spectrometric analysis of the permethylated glycans, which were liberated by non-reductive β-elimination (Tab. 2 and Fig. 4). Each glycan species was represented in the MALDI mass spectra by its pseudomolecular ion M+Na and mass increment calculation revealed the composition of the oligosaccharides in terms of monosaccharide constituents.
The patterns measured for the different cancer cell lines agreed mainly with those obtained by chromatographic analysis. Several oligosaccharides, however, could not be identified in the chromatograms in figure 2 and 3, probably due to interference with coeluting components or due to their scarcity. Vice versa the trace component NeuAca2-6GalNAc in the HPLC profiles (peak 3 in table. 1) was not detected as the corresponding pseudo-molecular ion (m/z 675) in the mass spectra. Two cell lines, T47D and MCF-7, were selected for the analysis of endogenous MUC1, because they represent the most extreme cases with respect to core1 and core2 expression on the recombinant probes. The results obtained for the endogenous mucin demonstrate that absence of core2 expression in T47D and prevalence of core2 expression in MCF-7 cells are authentically reflected in the profiles measured for the recombinant probes (Fig. 4, Tab. 2).

O-glycan density on partially deglycosylated tandem repeats

In addition to the structural O-glycan profiles, the density of O-linked chains in the peptide sequence is a second important parameter in the characterisation of mucin O-glycosylation. To address this issue we used an array of exoglycosidases to degrade the O-glycans to the level of the peptide bound N-acetylgalactosamine. Subsequent proteolytical degradation with clostripain, which cleaves once in each tandem repeat, resulted in a mixture of tandem repeat glycopeptides, containing two to five N-acetylgalactosaminyl residues. The relative proportion of each glycoform was revealed on a semi-quantitative basis by reflectron MALDI-TOF mass spectrometry (Fig. 5). Tandem repeat glycopeptides derived from MFP6 expressed in T47D, ZR75-1 or MDA-MB231 contained three to five GalNAc residues, with the five GalNAc peptide being the most intensive signal in the mass spectra. The glycopeptide mixture that was derived from MFP6 expressed in MCF-7 exhibited two to five GalNAc per tandem repeat peptide.
The strongest signal cluster in this spectrum corresponds to the peptide with three GalNAc residues, indicating a substantially lower glycosylation density with regard to the other cell lines. Note that none of the MFP6 preparations yielded tandem repeat peptides with less than two GalNAc residues.
Discussion

Structural evidence on the profiles and densities of O-linked glycans on MUC1 repeat peptide is presented using a truncated recombinant probe as reporter protein of endogenous mucin glycosylation. Utility of the approach using an artificial protein to probe *in vivo* O-glycosylation in tumor cell lines was validated by comparison with the endogenous mucin and the demonstration that the natural, cell-specific profiles of O-linked glycans are authentically reflected on the fusion protein. The results obtained for four cell lines reveal additional insight into the process of O-glycosylation in breast cancer cells.

1) No general pattern was found that could be regarded as breast cancer-associated. By contrast, each cell line expresses an individual, cell-specific O-glycosylation. 2) The profiles which demonstrate a preponderance of core2-based structures are in striking conflict with previous reports and disprove an earlier posited hypothesis that breast cancer-associated MUC1 is deficient in core2-type glycans. 3) Also in contrast to an accepted hypothesis, but in confirmation of a previous study on the endogenous mucin, we found that O-glycosylation density on recombinant MUC1 probes expressed in breast cancer cells is generally higher than on lactation-associated mucin. 4) In confirmation of previous investigations on this topic breast cancer-associated glycans appear to be generally reduced in their lengths compared to the polylactosamine-type structures found in milk.

In breast cancer information on MUC1 O-glycosylation was primarily based on enzymatic and immunochemical studies (20, 21, 22, 23), while structural evidence was limited (17, 18, 19). In summary, the latter studies had revealed that the tumor-associated profiles were characterized by reduced or deficient core2 expression. Studies on the enzymatic mechanisms underlying the aberrant O-glycosylation in breast cancer cells had revealed two important common features 1) low activities or deletion of the core2-specific β6-N-
acetylglucosaminyl-transferase (C2GnT), which converts core1-disaccharide (Galβ1-3GalNAc) into the corresponding core2-trisaccharide (20), and 2) over-expression of the core1-specific α3-sialyltransferase introducing NeuAc at Gal of core1 (21, 22). Since sialic acid represents a potent biosynthetic stop signal, the trisaccharide NeuAcα2-3Galβ1-3GalNAc can not be further glycosylated by C2GnT (35). Both changes on the enzymatic level are expected to reduce core2 formation in breast cancer cells, and consequently MUC1 O-glycosylation would be dominated by the sialylated derivatives of core1-disaccharide. Previous structural work was largely in agreement with the above outlined biosynthetic model. With one exception (T47D) our data on the O-glycan profiles on human breast cancer MUC1 are in disagreement with these earlier reports. In particular, the most prominent structures on MUC1 from ZR75-1 and MDA-MB231 have two characteristic features in common: they are based on core2-trisaccharide and terminate with α3-sialylated galactose. The latter structural element may reflect the over-expression of the respective sialyltransferase observed by Burchell et al. (22) in a variety of breast cancer cells. However, core2 formation was claimed to compete with termination of core1 by α3-sialylation catalyzed by the over-expressed enzyme (21). Hence, if the biosynthetic model was correct, α2-3 mono- and in particular α2-3 disialylated core2-based glycans (structures 6a, 6b, 7 in table 1) would not be expected to represent the major structures. A competition between C2GnT and α3ST would suggest an at least partial co-localization of the enzymes in one of the Golgi subcompartments. Expectedly, the core-specific enzyme should be found in the cis-Golgi and act on its substrate, before sialyltransferases in the trans-Golgi come into play. However, the actual subcellular distribution of the two enzymes has not yet been published.

In a previous contribution we could demonstrate that breast cancer cells exhibit an unexpected increase in the density of MUC1 O-glycosylation rather than a decrease (13).
This evidence was based on structural studies of endogenous mucin from T47D cells, and hence could represent an extreme feature of this particular cell line. The results of this study demonstrate that 1) MFP6 expressed in the same cell line exhibits a similar high density as the endogenous mucin; and 2) that also other breast cancer cell lines glycosylate the MUC1 tandem repeat at significantly higher densities than lactating epithelial cells.

A further common feature of O-linked chains on the cancer-associated MUC1-glycoforms is their reduced lengths. While the normal mucin was reported to carry polylactosamine-type chains with up to 16 monosaccharide units and average sizes of about 6-7 units (14, 15), the glycans from cancer-associated MUC1 did not exceed the lengths of a nonasaccharide with average sizes of 3-4 units. Chain truncation may result from reduced activity of the extension enzyme β3-GnT or from enhanced activity of α3ST, introducing a biosynthetic stop signal on the core2-tetrasaccharide before chain elongation and polylactosamine formation take place.

Independent of the specific MUC1-related topics the presented evidence potentially reveals also insight into aspects of the O-glycosylation process. 1) Overexpression of the recombinant probe compared to the endogenous mucin seems to have no significant effect on the O-glycosylation profiles. 2) The fusion protein passes the golgi compartments in a soluble form, while endogenous MUC1 is O-glycosylated as a membrane-integrated mucin. This difference is not reflected in the O-glycosylation patterns of secreted or shed glycoforms isolated from the cell supernatants. Both observations, however based on still rather limited evidence, may indicate that epigenetic parameters, like the rate of golgi passage and the topology of the protein substrate within the golgi compartment are less important with respect to the final glycosylation profiles than the cellular repertoire of glycosyltransferases.
Among the multifarious functions of O-linked glycans on proteins their impact on the antigenicity and immunogenicity of the underlying peptide core may be one of the most important. Differentiation-dependent changes in the cellular activity of glycosyltransferases could indirectly play a role in generating or masking of epitopes and could explain why tumor cells exhibit distinct glycosylation profiles compared to their normal counterparts. MUC1 O-glycosylation has been demonstrated to modulate antigenicity of the peptide core in different ways (23, 36), and influences of complex glycosylation on MUC1 up-take, processing and presentation by dendritic cells have also been shown (37). Recently it was demonstrated that natural anti-MUC1 antibodies from breast cancer patients exhibit a preference for the GalNAc substituted MUC1-peptides, indicating a direct contribution of the carbohydrate moiety to cancer-associated MUC1 epitopes (38). It is therefore of utmost importance to learn about the O-glycosylation profiles of this and possibly other mucins in cancer and it can be anticipated that the revealed individually fluctuating O-glycosylation profiles will have consequences for the design of efficient tumor vaccines in breast cancer.

Independent of the specific context in which this study was performed, the recombinant approach in combination with state-of-the-art methodology in the O-glycan profiling might be also of more general interest for investigators trying to unravel the complex regulation of O-glycosylation or functional aspects of O-glycosylation. Major advantages of the approach are that it provides specific recombinant probes in good yields, high purity and with a defined profile of O-linked glycans. There are many open questions which could be answered by using the respective recombinant constructs: 1) Are membrane-bound mucins differently glycosylated compared to their secretory counterparts? 2) Has the expression rate of a mucin construct influence on its cell-specific O-glycosylation? 3) How does O-glycosylation influence the trafficking of membrane-bound mucin constructs? 4) In
which way do O-linked glycans on mucin constructs affect endocytosis, processing and presentation of antigens by dendritic cells?
Literature

1. Zotter, S., Hageman, P.C., Lossnitzer, A., Mooi, W.J., Hilgers, J. (1988) Cancer Rev. 11, 55-101
2. Agrawal, B., Krantz, M.J., Parker, J., Longenecker, B.M. (1998) Cancer Res. 58, 4079-4081
3. Dent, G.A., Civalier, C.J., Brecher, M.E., Bentley, S.A. (1999) Am. J. Clin. Pathol. 111, 741-747
4. Girling, A., Bartkova, J., Burchell, J., Gendler, S., Gillett C., Taylor-Papadimitriou J. (1989) Int. J. Cancer 43, 1972-1976
5. Apastolopoulos, V., McKenzie, I.F.C. (1994) Crit. Rev. Immunol. 14, 293-309
6. Finn O.J., Jerome, K.R., Henderson, R.A., Pecher, G., Domenech, N., Magarian-Blander, J., Barrat-Boyes, S.M. (1995) Immunol. Rev. 145, 61-89
7. Miles, D.W., Taylor-Papadimitriou, J. (1999) Pharmacol. Ther. 82, 97-106
8. Ligtenberg, M.J.L., Kruishaar, L.; Buijs, F., van Meijer, M., Litvinov, S.V., Hilkens, J. (1992) J. Biol. Chem. 267, 6171-6177
9. Litvinov, S.V., Hilkens, J. (1993) J. Biol. Chem. 268, 21364-21371
10. Gendler, J.S., Spicer, A.P. (1995) Annu. Rev. Physiol. 57, 607-634
11. Hanisch, F-G., Müller, S. (2000) Glycobiology 9, 1181-1189
12. Müller, S., Goletz, S., Packer, N., Gooley, A., Lawson, A.M., Hanisch, F-G. (1997) J. Biol. Chem. 272, 24780-24793
13 Müller, S., Alving, K., Peter-Katalinic, J., Zachara, N., Gooley, A. A., Hanisch, F-G. (1999) *J. Biol. Chem.* **274**, 18165-18172

14 Hanisch, F-G., Uhlenbruck, G., Peter-Katalinic, J., Egge, H., Dabrowsky, J., Dabrowsky, U. (1989) *J. Biol. Chem.* **264**, 872-883

15 Hanisch, F-G., Uhlenbruck, G., Peter-Katalinic, J., Egge, H., Dabrowsky, U. (1990) *Glycoconj. J.* **7**, 525-543

16 Bhavanandan, B.H., Zhu, Q., Yamakami, K., Dilulio, N.A., Nair, S., Capon, C., Lemoine, J., Fournet, B. (1998) *Glycoconj. J.* **15**, 37-49

17 Hanisch, F-G., Stadie, T.R.E., Deutzzmann, F., Peter-Katalinic, J. (1996) *Eur. J. Biochem.* **236**, 318-327

18 Hull, S.R., Bright, A.S., Carraway, K.L., Abe, M., Hayes, D.F., Kufe, D.W. (1989) *Cancer Commun.* **1**, 261-267

19 Lloyd, K.L., Burchell, J., Kudryashov, V., Yin, B.W.T., Taylor-Papadimitriou, J. (1996) *J. Biol. Chem.* **271**, 33325-33334

20 Brockhausen, I., Yang, J-M., Burchell, J., Whitehouse, C., Taylor-Papadimitriou, J. (1995) *Eur. J. Biochem.* **233**, 607-617

21 Whitehouse, C., Burchell, J., Gscheissner, S., Brockhausen, I., Lloyd, K., Taylor-Papadimitriou, J. (1997) *J. Cell. Biol.* **137**, 1229-1241

22 Burchell, J., Poulson, R., Hanby, A., Whitehouse, C., Cooper, L., Clausen, H., Miles, D., Taylor-Papadimitriou, J. (1999) *Glycobiology* **9**, 1307-1311

23 Burchell, J., Taylor-Papadimitriou, J. (1993) *Epith. Cell Biol.* **2**, 155 - 162
24 Hilkens, J., Kroezen, V., Bonfrer, J.M., De Jong-Bakker, M., Bruning, P.F. (1986) 
*Cancer Res.* **46**, 2582-2587

25 Kotera, Y., Fontenot, D.J., Piecher, G., Metzgar, R.S., Finn, O.J. (1994) 
*Cancer Res.* **54**, 2856-60

26 Barnd. D.L., Lan, M.S., Metzgar, R.S., Finn, O.J. (1989) 
*Proc. Natl. Acad. Sci. USA* **86**, 7159-7163

27 Agrawal, B., Reddish, M.A., Krantz, M.J., Longenecker, M.M. (1995) 
*Cancer Res.* **55**, 2257-2261

28 Karanikas, V., Hwang, L.A., Pearson, J., Ong, C.S., Apostolopoulos, V., Vaughan, H., Xing, P.X., Jamieson, G., Pietersz, G., Tait, B., Broadbent, R., Thynne, G., McKenzie, I.F.C. (1997) *J. Clin. Invest.* **100**, 2783-2792

29 Goydos, J.S., Elder, E., Whiteside, T.L., Finn, O.J., Lotze, M.T. (1996) 
*J. Surgical Res.* **63**, 298-304

30 Kohfeldt, E., Maurer, P., Vannahme, C., Timpl, R. (1997) *FEBS Lett.* **414**, 557-561

31 Patel, P., Bruce, J., Merry, A., Bigge, C., Wormald, M., Jaques, A., Parekh, R. (1993) *Biochemistry* **32**, 679-693

32 Bigge, J.C., Patel, T.P., Bruce, J.A., Goulding, P.N., Charles, S.M., Parekh, R.B. (1995) *Anal. Biochem.* **230**, 229-238

33 Guile, G.R., Wong, S.Y.C., Dwek, R.A. (1994) *Anal. Biochem.* **222**, 231-235

34 Engelmann, K., Baldus, S.E., Hanisch, F.-G. (2001) *J. Biol. Chem.* **276**, 27764 - 27769

35 Brockhausen, I., (1999) *Biochim. Biophys. Acta* **1473**, 67-95
36 Karsten, U., Diotel, C., Klich, G., Paulsen, H., Goletz, S., Müller, S., Hanisch, F.-G. (1998) *Cancer Res.* **58**, 2541 - 2549

37 Hitbold, E.M., Alter, M.D., Ciborowski, P., Finn, O.J. (1999) *Cell. Immunol.* **194**, 143-149

38 von Mensdorff-Pouly, S., Petrakou, E., Kenemans, P., van Uffelen, K., Verstraeten, A.A., Snijdwint, F.G.M., van Kamp, G.J., Schol, D.J., Reis, C.A., Price, M.R., Livingston, P.O., Hilgers, J. (2000) *Int. J. Cancer* **86**, 702-712
Legends

Fig. 1  A  Amino acid sequence of MFP6

A recombinant fusion protein containing a secretory signal peptide (italics), a hexa-histidine (underlined) and a myc tag (underlined) followed by six MUC1 tandem repeat units (boxed) was constructed from oligonucleotides and a MUC1 cDNA. The sequences flanking the tandem repeats were designed to encode two tyrosine residues and several protease cleavage sites, to allow photometric quantification of the protein and proteolytic release of the entire tandem repeat unit. Deviations from the MUC1 tandem repeat consensus sequence are shaded and represent a sequence polymorphism in the MUC1 gene.

B  Reversed-phase HPLC of affinity-isolated MFP6 (MCF-7) on C8 column

MFP6 fusion protein from MCF-7 cells was re-chromatographed on a C8 reversed phase column (208TP10415, Vydac) in a gradient of acetonitril in 0.1 % aqueous TFA (2-80% over 30 min) at a flow rate of 1 ml/min. Eluting protein was registered spectrophotometrically at 214 nm and collected manually. Inserts: Westernblots with anti-myc antibody (lane 1) show the glycosylated MCF-7 fusion protein in HPLC fraction 16 and 17 with an apparent molecular mass of 40-60 kDa. Lane 2 shows a westernblot of fraction 16 stained with the DIG Glycan Detection Kit.

C  Westernblot analysis of affinity-purified endogenous MUC1 from MCF-7 cell supernatants

Excluded fraction from Superdex 200 HR size-exclusion chromatography stained with anti-MUC1 antibody B27.29 (lane 1); the same fraction stained with the DIG Glycan Detection Kit (Roche) (lane 2).
Fig. 2 Normal phase and anion exchange HPLC analysis of 2AB labeled O-glycans

Complete O-glycan pools were released from the fusion proteins by hydrazinolysis and the glycans were labelled with 2-AB.

(A) Samples from MFP6 fusion protein were dissolved in 75 % acetonitril in water and 20 µl were injected onto a polymer-based aminophase column \((\text{NH}_2\text{-polymer, Astec})\) equilibrated in 80 % acetonitril / 20 % 250 mM ammonium formiate pH 4.5. The column was eluted by decreasing the percentage of acetonitril to 20 % over 120 min at a flow rate of 0.5 ml/min. Excitation and emission wavelength were set to 330 and 420 nm respectively.

(B) Samples from MFP6 were dissolved in water and 20 µl were injected onto a polymer-based strong anion exchange column (Q HyperD 10, Beckman). The column was equilibrated in water and eluted with an increasing concentration of ammonium formiate pH 9 (dotted line). The flow rate was 1ml/min. Peak numbers in (A) and (B) refer to the assigned structures listed in table 1.

Fig. 3 Glycosidase degradation of 2-AB labeled MCF-7 derived O-glycan pools

The complete 2-AB labeled MCF-7 derived O-glycan pool was digested with the indicated exoglycosidases. Samples were dissolved in 75 % acetonitril in water and 20 µl were injected onto a polymer-based aminophase column. The column was eluted with a linear gradient of 80 % acetonitril / 20 % 50 mM ammonium formiate to 20 % acetonitril / 80 % 50 mM ammonium formiate over 120 min at a flow rate of 0.4 ml/min. Excitation and emission wavelength were set to 330 and 420 nm respectively. Peak numbers refer to figure 2 and to the assigned structures listed in
table 1. Downward and upward arrows indicate peaks that disappeared or increased after glycosidase treatment of the sample.

**Fig. 4** MALDI mass spectra of methylated O-glycans from recombinant and endogenous MUC1 expressed by MCF-7 cells

Solutions of the methylated glycans derived from approx. 10 µg of fusion protein (A) or mucin (B) were mixed with matrix and analysed on a Bruker Reflex IV MALDI-tof mass spectrometer for the patterns of pseudo-molecular ions M+Na. Identification of the ion masses is given in table 2.

**Fig. 5** Reflectron MALDI-TOF MS of partially deglycosylated tandem repeat peptides

Recombinantly expressed MFP6 from supernatants of breast cancer cells T47D (A), MCF-7 (B), MDA-MB231 (C), and ZR75-1 (D) was sequentially deglycosylated with the exoglycosidases neuraminidase, α2- and α3/4-fucosidases, β-galactosidase and β-hexosaminidase. Proteolytic cleavage products from clostripain digestion were analysed after rpHPLC by MALDI-TOF mass spectrometry in the positive ion mode. Triplett clusters of pseudomolecular ions MH correspond to the original PAP20 sequence (calculated average mass: 1888.0) and its sequence variants (Pro > Ala: -26 Da; Pro > Gln: +31 Da) substituted with 2 to 5 GalNAc residues (mass increment: 203.2). Signal clusters marked by 5* correspond to penta-GalNAc-substituted PAP20 peptides with one residual galactosyl moiety. Except for PAP20 glycopeptides from MCF-7 cells, all other cell lines revealed high density O-glycosylation of MUC1 tandem repeats with predominance of the pentasubstituted repeat peptide.
Acknowledgements

This study was supported by grants from the Deutsche Forschungsgemeinschaft Ha2092 / 4-4 and from the National Institute of Health 1RO1 CA84106. The expression vector pCEP-PU, including the BM40 signal sequence and tags was kindly provided by Dr. Patrik Maurer, Institute of Biochemistry, Med. Faculty, University of Cologne, Cologne, Germany. We thank Mrs. Kirsten Ottenberg for her skillful technical assistance during preparation of endogenous MUC1 samples.
**BM40-signal peptide**

MRAWIFFLCLAGRALA

A PLVHHHHHGPLVDVASNE

hexa-his-tag myc-tag

Q KLISEEDLASMTGGQQMGR

DIEGRGLAPYALKEMAPP AH

GVTSAPDTRPAPGSTAP A AH

GVTSAPESRPAPGSTAPP A AH

GVTSAPDTRPAPGSTAPPAH

GVTSAPDTRPAPGSTAPPAH

GVTSAPDTRPAPGSTAPPAH

GVTSAPDTRPAPGSTAPPAH

GVTSAPDTRPAPGSTAPPAH

GVTSAPDTRPAPGSTAPPAH

GVTSAPDTRPAPGSTAPPAH

GVTSAPDTRPAPGSTAPPAH

GVTSAPDTRPAPGSTAPPAH

GVTSAPDTRPAPGSTAPPAH

GVTSAPDTRPAPGSTAPPAH

GVTSAPDTRPAPGSTAPPAH

GVTSAPDTRPAPGSTAPPAH

**QVMKEALPYPRPLEAGKAGS**

**RHDKIH**
Fig. 1B

Retention time [min]

MW (kDa)

Fr. 16-17

Downloaded from https://www.jbc.org/ on March 24, 2020 by guest.
Fig. 1C
Fig. 2

(A) Retention time [min] vs. relative fluorescence for T47D, MCF-7, MDA 231, and ZR75-1.

(B) Retention time [min] vs. relative fluorescence for T47D, MCF-7, MDA 231, and ZR75-1.
Fig. 3

Retention time [min]

Relative fluorescence

Native

Neuraminidase

α1-2 fucosidase

α1-2 + α1-3/4 fucosidase

Peaks labeled with numbers correspond to specific retention times and relative fluorescence levels for each enzyme treatment.
Fig. 4
| Proposed structures | HPLC peak | T47D | MCF-7 | MDA 231 | ZR 75-1 |
|---------------------|-----------|------|-------|---------|---------|
| Gal (β1-3) GalNAc   | 1         | 6.9  | 12.9  | 5.3     | trace   |
| Gal (β1-4) GlcNAc, (β1-6) Gal (β1-3) GalNAc | 2 | - | 54.3 | 13.5 | 4.3 |
| NeuAc (α2-6) GalNAc | 3 | trace | - | trace | trace |
| NeuAc (α2-3) Gal (β1-3) GalNAc | 4 | 45.9 | 2.3 | 24.6 | 14.5 |
| NeuAc (α2-6) Gal (β1-3) GalNAc | 5 | 24.6 | 2.9 | 4.8 | 2.3 |
| NeuAc (α2-3) Gal (β1-4) GlcNAc, (β1-6) Gal (β1-3) GalNAc | 6 | trace | - | 21.6 | 24.7 |
| NeuAc (α2-3) Gal (β1-4) GlcNAc, (β1-6) NeuAc (α2-3) Gal (β1-3) GalNAc | 7 | trace | - | 10.9 | 32.1 |
| NeuAc (α2-6) NeuAc (α2-3) Gal (β1-3) GalNAc | 8 | 22.7 | - | 8.4 | 14.8 |
| Fuc (α1-3) Gal (β1-4) GlcNAc, (β1-6) Gal (β1-3) GalNAc | 9 | - | 7.9 | - | - |
| Fuc (α1-2) Gal (β1-4) GlcNAc, (β1-6) Gal (β1-3) GalNAc | 10 | - | 9.2 | - | - |
| Fuc (α1-3) GlcNAc (β1-6) Gal (β1-3) GalNAc | 11 | - | 2.2 | - | - |
| Fuc (α1-3) GlcNAc (β1-6) Fuc (α1-2) Gal (β1-3) GalNAc | 12 | - | 5.0 | - | trace |
| GalNAc / tandem repeat peptide [Mol/Mol] | | 3-5 | 2-5 | 3-5 | 3-5 |
| core2 [%] | < 5 | 83 | 45 | 75 |
| acidic glycans [%] | 93.2 | 5.2 | 70.3 | 88.4 |

O-glycan pools were released by hydrazinolysis and the 2AB-derivatives were analyzed by normal phase HPLC as described in the text. Integrated peak areas were used for quantification. Peak identification numbers refers to the chromatograms shown in Fig. 2 and Fig. 3.
Table 2: MALDI-TOF MS analysis of methylated O-glycans from recombinant MUC1 fusion protein MFP6 or endogenous MUC1

| Oligosaccharide composition | HPLC peak identification | Pseudomolecular ion M+Na (average mass) | Human breast cancer cells |
|-----------------------------|---------------------------|----------------------------------------|--------------------------|
|                            |                           |                                        | T47D         | MCF-7         | MDA-MB231     | ZR75-1       |
|                            |                           |                                        | R  | E  | R  | E  | R  | E  |
| Hex$_1$ HexNAC$_1$         | 1                         | 518                                    | +  | n.d. | +  | n.d. | (+) | -  |
| dHex$_1$ Hex$_1$ HexNAC$_1$| n.d.                      | 692                                    | +  | n.d. | ( +) | n.d. | -   | -  |
| Hex$_1$ HexNAC$_2$         | n.d.                      | 763                                    | -  | -   | +  | -   | (+) | -  |
| NeuAc$_1$ Hex$_1$ HexNAC$_1$| 4, 5                      | 879                                    | +  | +   | -  | -   | +   | +  |
| Hex$_2$ HexNAC$_2$         | 2                         | 967                                    | -  | -   | +  | -   | +   | -  |
| dHex$_1$ Hex$_2$ HexNAC$_2$| 9, 10                     | 1141                                   | -  | -   | +  | -   | +   | -  |
| NeuAc$_2$ Hex$_1$ HexNAC$_1$| 8                         | 1240                                   | +  | +   | +  | -   | +   | +  |
| dHex$_2$ Hex$_2$ HexNAC$_2$| 11                        | 1315                                   | -  | -   | ( +) | ( +) | trace | -  |
| NeuAc$_1$ Hex$_2$ HexNAC$_2$| 6                         | 1329                                   | -  | -   | trace | ( +) | +   | +  |
| Hex$_3$ HexNAC$_3$         | 12                        | 1416                                   | -  | -   | ( +) | +   | -   | -  |
| dHex$_1$ Hex$_3$ HexNAC$_3$| n.d.                      | 1590                                   | -  | -   | ( +) | ( +) | -   | -  |
| NeuAc$_2$ Hex$_2$ HexNAC$_2$| 7                         | 1690                                   | trace | -   | -   | +   | +   | +  |
| dHex$_2$ Hex$_3$ HexNAC$_3$| n.d.                      | 1764                                   | -  | -   | trace | trace | -   | -  |
| NeuAc$_1$ Hex$_3$ HexNAC$_3$| n.d.                      | 1777                                   | -  | -   | -   | trace | -   | -  |
| Hex$_4$ HexNAC$_4$         | n.d.                      | 1865                                   | -  | -   | -   | ( +) | -   | -  |
| dHex$_1$ Hex$_4$ HexNAC$_4$| n.d.                      | 2040                                   | -  | -   | -   | trace | -   | -  |

The O-linked glycans were liberated from the fusion protein by β-elimination / methylation, extracted into chloroform and analysed after evaporation and resolubilization in methanol in a mixture with matrix dihydroxy bezolic acid (1 : 4, v / v). The presence of major ion species is indicated by + , that of minor species by ( +). Numbers under „HPLC peak identification“ refer to the chromatographic profiles in figures 2 and 3, and to table 1. Endogenous MUC1 (E) was isolated from...
supernatants of MCF-7 cells and analysed in parallel to the respective fusion protein (R). n.d. refers to glycans, which were not detectable by HPLC, or to molecular ions, which could not be unambiguously identified due to clusters of low molecular mass compounds.
Recombinant MUC1 probe authentically reflects cell-specific O-glycosylation profiles of endogenous breast cancer mucin: High-density and prevalent core2-based glycosylation
Stefan Müller and Franz-Georg Hanisch

J. Biol. Chem. published online May 8, 2002

Access the most updated version of this article at doi: 10.1074/jbc.M202921200

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

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