The N Terminus of the MUC2 Mucin Forms Trimers That Are Held Together within a Trypsin-resistant Core Fragment*

Klaus Godl‡‡, Malin E. V. Johansson‡‡, Martin E. Lidell‡, Matthias Mörgelin‡, Hasse Karlsson‡, Fredrik J. Olson‡, James R. Gum, Jr.‡‡*, Young S. Kim‡‡*, and Gunnar C. Hansson‡‡ ‡‡

From the ‡Department of Medical Biochemistry, Göteborg University, Medicinaregatan 9A, 413 90 Gothenburg, Sweden, ¶Cell and Molecular Biology, Lund University, 221 84 Lund, Sweden, and the ¶Gastrointestinal Research Laboratory, Department of Veterans Affairs Medical Center, University of California, San Francisco, California 94121

The N terminus of the human MUC2 mucin (amino acids 1–1397) has been expressed as a recombinant tagged protein in Chinese hamster ovary cells. The intracellular form was found to be an endoglycosidase H-sensitive monomer, whereas the secreted form was an oligomer that gave monomers upon disulfide bond reduction. The secreted MUC2 N terminus contained a trypsin-resistant core fragment. Edman sequencing and mass spectrometry of the peptides obtained localized this core fragment to the C-terminal end of the recombinant protein. This core retained its oligomeric nature with an apparent mass of ~240 kDa. Upon reduction, peptides of ~85 kDa were found, suggesting that the N terminus forms trimers. This interpretation was also supported by gel electrophoresis and gel filtration of the intact MUC2 N terminus. Electron microscopy revealed three globular domains each linked via an extended and flexible region to a central part in a trefoil-like manner. Immunostaining with gold-labeled antibodies localized the N-terminal end to the three globular structures, and the antibodies directed against the Myc and green fluorescent protein tags attached at the C terminus localized these to the stalk side of the central trefoil. The N terminus of the MUC2 mucin is thus assembled into trimers that contain proteolytically stable parts, suggesting that MUC2 can only be partly degraded by intestinal proteases and thus is able to maintain a mucin network protecting the intestine.

Mucins are highly glycosylated proteins protecting the mucosal surfaces of the body. All mucins are characterized by their heavy O-glycosylation, which is clustered in mucin domains rich in the amino acids Ser, Thr, and Pro. Mucins can be structured in different ways, either as membrane-bound with a transmembrane domain or secreted as monomers or as polymers (1, 2). The latter mucins are also described as gel-forming, as these give the mucus its viscoelastic properties. Four human gel-forming mucins have been found, MUC2, MUC5AC, MUC5B, and MUC6, which are clustered on chromosome 11p15 (3). However, the best studied mucin is the porcine submaxillary mucin (PSM),1 which has become a model also for the human mucins of this type (3–6).

The MUC2 mucin is the main gel-forming mucin of the small and large intestines and is produced by the intestinal goblet cells (2, 7–10). It is a major structural component of the mucous barrier covering the epithelium, protecting the epithelial cells from microorganisms as well as digestive enzymes. In fact, one of the remarkable features of this mucin is that it can withstand the pancreatic digestive enzymes. That the highly glycosylated mucin domains are totally resistant to these enzymes is easy to understand. However, the globular, less glycosylated ends must also be relatively proteolytically resistant, as the polymers must remain intact to maintain the mucous gel. The MUC2 mucin is difficult to solubilize, and it is insoluble in 4 M guanidinium chloride (8, 10), a property that might be related to the appearance of nonreducible covalent linkages in addition to the disulfide bonds (10, 11). The MUC2 mucin has been reported to be aberrantly expressed in the airways upon Pseudomonas aeruginosa infection in cystic fibrosis, where its properties might contribute to the abnormal mucous viscosity in this disorder (12).

Both N and C termini of human MUC2 show sequence similarities to the respective termini of the von Willebrand factor (vWF) and PSM. The vWF is dimerized in the endoplasmic reticulum via intermolecular disulfide bond formation of its C terminus. Multimerization takes place in the Golgi via disulfide bonds between cysteines in the N terminus, where the vWF forms end-to-end oligomers (13, 14). This, thus, the vWF will form linear polymers (13, 14). This observation also suggests that the polymerized gel-forming mucins are linear, which is consistent with biophysical studies of mucins (15, 16). However, when the N terminus of PSM was recombinantly expressed, the results suggested that it formed trimers instead of the predicted dimers (5, 6). This suggestion has been disputed, and it has also been suggested that PSM may be oligomerized differently from other mucins.

To study how the N terminus takes part in the multimerization of the MUC2 mucin, the MUC2 N terminus was expressed in Chinese hamster ovary (CHO) cells. The secreted oligomerized mucin was purified and shown by electron microscopy to

1 The abbreviations used are: PSM, porcine submaxillary mucin; vWF, von Willebrand factor; CHO, Chinese hamster ovary; GFP, green fluorescent protein; mAb, monoclonal antibody; BiTris, 2-bis[2-hydroxyethyl]amino]–2-[2-hydroxyethyl]propane-1,3-diol; MALDI, matrix-assisted laser desorption ionization; MS, mass spectrometry; MS/MS, tandem mass spectrometry; ESI, electrospray ionization.
be held together in a central condense core surrounded by three large globular domains. The oligomer was also shown to have a large portion that was resistant to trypsin digestion. The trypsin-resistant “core fragment” was localized and shown to be a trimer.

**EXPERIMENTAL PROCEDURES**

**Tissue Culture**—CHO-K1 cells were obtained from American Type Culture Collection (CCL-61). Cells were cultured in Iscove’s modified Dulbecco’s medium (Invitrogen) containing 10% (v/v) fetal calf serum and 1% (v/v) penicillin/streptomycin and supplemented with sodium pyruvate (110 mg/liter), L-arginine (116 mg/liter), L-glutamine (290 mg/liter), L-asparagine (36 mg/liter), L-lysine (10 mg/liter), and β-mercaptoethanol (3.49 μl/liter). Transfected cells were kept in G418 (250 μg/ml).

**Construction of Expression Vector**—pSNMUC2-MG—DNA clones D8, SMUC343 (7), and SMUC313 * contain bp 1–617, 565–1792, and 1685–4208, respectively, of the published sequence (GenBankTM/EBI accession with enhanced GFP was removed from the pEGFP vector (Clontech) by digestion with XbaI and EcoRI site at bp 566, and SMUC313 was added to the end of SMUC343 by the natural SfiI site at bp 1678. RNA from the cell line LS174T was used to obtain a 241-bp reverse transcription-PCR product using primers 5'-ACGACGGTACGAGAAGATCTT and 5'-TTCATAGGTCGTATCACTTATCATGACGCC. The PCR product was ligated to the SfiI site to give a unique XhoI site at the 3'-end. The construct was made in pBluescript (Stratagene). The DNA was sequenced; compared with the published sequence; and found to be identical with one exception: at positions 4079 and 4080, AT was substituted with TA. This change gives rise to an amino acid substitution of His to Leu and has previously been reported. Because each of the positive clones was selected by adding 250 g/ml of 2-Mercaptoethanol, were subcloned in microtiter plates, and strongly fluorescent clones were pooled (no forms of intact SNMUC2-MG were excluded), dialyzed against 10 mM Tris-HCl (pH 7.4) and diluted with the same buffer to 50 μM, was applied to the column at 200 μl/min. The column was washed at 500 μl/min with 60 ml of 10 mM Tris-HCl (pH 7.4) and 500 mM NaCl. The antibodies were eluted with 30 ml of 100 mM glycine HCl (pH 2.5) into a tube with 3 ml of 10 mM Tris-HCl (pH 8.0). Additional 2 mM Tris base was added to bring the eluate to a pH of 7. The column was equilibrated with 25 ml of 10 mM Tris-HCl (pH 7.4), 150 mM NaCl, and 0.02% (v/v) NaN3 by ultrafiltration (3000 g × g × 4 °C; Vivaspin 2, R.C., 30,000 cutoff, Sartorius Corp.). The anti-MUC2N3 antiserum (19) was used unpurified, and the anti-Myc mAb (clone 9E10, American Type Culture Collection, CRL-1729) was used as hybridoma spent culture medium. The anti-GFP mAb was from BD Biosciences.

**Immunoblotting**—Proteins were boiled into nitrocellulose membranes (Trans-Blot transfer medium, Bio-Rad) in 25 mM Tris and 192 mM glycine using a semidry transfer cell (Trans-Blot SD, Bio-Rad) at 2 mA/cm2 for 2 h. The membrane was blocked for 3 h at room temperature in 10% milk powder in phosphate-buffered saline with 0.1% Tween 20 and incubated at 4 °C overnight with the primary antibody. The membrane was incubated for 1 h at room temperature in the secondary antibody (1:1000 dilution; anti-rabbit or mouse alkaline phosphatase-conjugated IgG, Dako Corp.) and developed with nitro blue tetrazolium/5-bromo-4-chloro-3-indolyl phosphate.

**Immunoprecipitation**—50 μl of Dynabeads M-450 (goat anti-mouse IgG, 4 × 108 beads/ml; Dynal, Inc.) were washed three times with 500 μl of phosphate-buffered saline, 0.1% bovine serum albumin, and 0.1% sodium azide and incubated on a shaker for 1 h at room temperature with 200 μl of anti-Myc mAb hybridoma supernatant. The beads were washed four times with 500 μl of 50 mM Tris-HCl (pH 7.9), 150 mM NaCl, and 1% Triton X-100 and resuspended in 50 μl of the same buffer. 2 ml of metabolically labeled culture medium and cell lysate, respectively, were added and incubated overnight at 4 °C. After washing 10 times with 500 μl of 50 mM Tris-HCl (pH 7.9), 150 mM NaCl, and 1% Triton X-100, the beads were eluted in 50 μl of 50 mM Tris-HCl (pH 6.8) and 1% SDS for 5 min at 95 °C.

**Purification of SNMUC2-MG**—100 ml of culture medium and cell lysates, respectively, containing SNMUC2-MG were filtered in an Amicon YM-10 stirred cell through an Amicon YM100 ultrafiltration membrane (Millipore Corp.). The volume was reduced to 25 ml, diluted to 300 ml with BT buffer (20 mM BisTris-HCl (pH 6.0), and reduced again to 25 ml. This procedure was repeated four times. Aliquots (5 ml) of the concentrate were loaded onto a Superose 6 prepartion grade column (1000 × 26 mm; Amersham Biosciences) in BT buffer at a flow rate of 1.5 ml/min. The eluting proteins were monitored at 280 nm. The SNMUC2-MG-containing fractions, as determined by SDS-PAGE, were pooled and directly loaded onto an ion-exchange column (Mono Q 5/5, Amersham Biosciences) in BT buffer. Medium-derived SNMUC2-MG was washed with 210 mM NaCl in BT buffer and eluted with 250 mM NaCl in BT buffer. Lysate-derived SNMUC2-MG was washed with 170 mM NaCl in BT buffer and eluted with 210 mM NaCl in BT buffer. The collected 1 ml fractions were individually analyzed by SDS-PAGE and silver-stained. After gel filtration as well as ion-exchange chromatography, the fractions were also subjected to SDS-PAGE and Western blotting using the anti-Myc mAb and anti-MUC2N3 antiserum. The fractions containing SNMUC2-MG were pooled (no forms of intact SNMUC2-MG were excluded), dialyzed

---

2 J. R. Gum, Jr., unpublished data.
against H2O, and lyophilized. The purification finally resulted in a yield of ~0.5 mg of SNMUC2-MG protein.

**Analytical Gel Filtration**—50 μg of nonreduced purified SNMUC2-MG were dissolved in 100 μl of 4% guanidinium chloride and 50 mM Tris-HCl (pH 8.0) and applied to a Superose 6 HR 10/30 column eluted at 0.25 ml/min using an OptoLab DSP refractometer (Wyatt Technology Corp.) as the detector. The column had a void volume of 7.7 ml and the standards β-amylose (200 kDa), apoferritin (445 kDa), and thyroglobulin (669 kDa) eluted at 14.61, 14.51, and 10.70 ml, respectively.

**Preparation of Tryptic Core Fragment**—Purified medium-derived SNMUC2-MG (100 μg) was incubated in 0.1% SDS for 5 min at 95 °C and subsequently diluted with 9 volumes of 100 mM ammonium acetate (pH 8.5). After addition of 5 μg of tosylphenylalanyl chloromethyl ketone (TPCK)-trypsin (Sigma), the samples were incubated overnight at 37 °C. For deglycosylation, the trypsin digest was stopped with 100 μg/ml aprotinin and 1 mM phenylmethylsulfonyl fluoride at room temperature for 4 h, and 4 units of N-glycosidase F (Roche Molecular Biochemicals) were added to the sample and incubated overnight at 37 °C. Purification was done on lyophilized material that was redissolved in 0.1 M ammonium acetate (pH 8.5) and separated by gel filtration on a Superose 6 PC 3.2/30 column (Amersham Biosciences) using a Smart HPLC system (Amersham Biosciences). Fractions containing the trypptic core fragment were lyophilized with and without prior reduction in 10 mM dithiothreitol.

**Deglycosylation with Hydrogen Fluoride**—Immunoprecipitated 35S-labeled SNMUC2-MG was lyophilized before addition of 100 μl of ammonium hydrogen fluoride and incubated for 30 min on ice. Hydrogen fluoride was evaporated under a stream of nitrogen and finally for 15 min in a SpeedVac concentrator. The material was redissolved in 20 μl of SDS sample buffer with 200 μM dithiothreitol and analyzed on a 5% SDS-polyacrylamide gel.

**Enzymatic Deglycosylation**—35S-Labeled SNMUC2-MG was immunoprecipitated, and the beads were resuspended in 50 μl of 50 mM Tris-HCl (pH 7.9), 150 mM NaCl, and 1% Triton X-100. Aliquots (10 μl) were washed twice with 100 μl of 50 mM sodium acetate (pH 6.0), 4 mM calcium chloride, and 100 μg/ml bovine serum albumin and resuspended in 50 μl of the same buffer before addition of 0.5 μl of neuraminidase (1 unit/ml), 0.5 μl of O-glycosidase (0.5 units/ml), or 0.5 μl of endoglycosidase H (5 units/ml) (Roche Molecular Biochemicals), respectively. Aliquots (10 μl) were washed twice with 100 μl of 50 mM sodium phosphate buffer (pH 7.0), 10 mM EDTA, and 100 μg/ml bovine serum albumin and resuspended in 50 μl of the same buffer before addition of 0.5 μl of Pepsin (1000 units/ml). All samples were incubated on a shaker overnight at 37 °C. After removal of the supernatant, the beads were eluted in 20 μl of SDS sample buffer containing 10% mercaptoethanol for 5 min at 95 °C and analyzed by SDS-PAGE and autoradiography.

**In-gel Digest**—The deglycosylated trypptic core fragment was separated by SDS-PAGE on 4–15% gradient Ready-Gel (Bio-Rad) and stained with Coomassie Blue. After destaining in 30% methanol and 7% acetic acid, the gel was soaked in H2O overnight. The excised band was cut into 1-mm pieces. After successive washing steps for 30 min at room temperature on a shaker in 0.2 M ammonium bicarbonate, 0.2 M ammonium bicarbonate and 25% acetonitrile, H2O and 25% acetonitrile, H2O and 50% acetonitrile, and 100% acetonitrile, the gel pieces were air-dried for 2 h. 1 volume of trypsin (sequencing grade, Roche Molecular Biochemicals; 20 μg/ml in 0.2 M ammonium bicarbonate) was added to the beads, and 10 min later, an additional volume of 0.2 M ammonium bicarbonate was added and incubated overnight at 37 °C. The peptides were eluted from the gel pieces in 4 volumes of 0.1% trifluoroacetic acid (for MALDI-MS) or 0.2% formic acid (for ESI-MS/MS) for 6 h at room temperature.

**MALDI-MS**—The reduced and alkylated trypptic core fragment without deglycosylation was in-gel trypsin-digested (see above), lyophilized, and analyzed with a Trimmer Tip C18 column (MForeground Corp.) according to the manufacturer's instructions. The MALDI mass spectra were collected on a Micromass Tof-Spec E time-of-flight mass spectrometer with delayed extraction in the reflectron mode. 0.5 μl of matrix solution (10 mg of α-cyanoc-4-hydroxycinnamic acid (Sigma) in 1 ml of 1:1 acetonitrile/water) were mixed on the target with 0.5 μl of sample and then left to dry. The monoisotopic peptide masses from the MALDI mass spectra were matched against theoretical trypptic peptides from the full-length SNMUC2-MG sequence for identification.

**ESI-MS/MS**—0.1 volume of the in-gel trypsin-digested core fragment was purified on a Zip-Tip C18 column according to the manufacturer's instructions. The eluted volume (50%, ~2 μl) was analyzed by nanoflow ESI-MS/MS on a Micromass Q-Tof 1 mass spectrometer. Argon was used as the collision gas with a collision energy of 40 eV.

**Edman Sequencing**—The lyophilized trypptic core fragment (1 pmol) was redissolved in SDS sample buffer containing 100 mM dithiothreitol and separated by SDS-PAGE on 5% Ready-Gel. After Western blotting onto polyvinylidene difluoride membrane (Immobilon-P, Millipore Corp.) in 50 mM sodium borate buffer (pH 8.9) and 10% methanol for 30 min at 2 mA/cm2, the membrane was stained with 0.1% Coomasie Blue R-250 (Bio-Rad). The stained band was cut out and N-terminally sequenced by Edman degradation on a Procise 492 Protein Sequencer (Applied Biosystems). A pulsed-liquid sequencing method for polyvinylidene difluoride-blotted protein was used according to the manufacturer.

**Electron Microscopy**—SNMUC2-MG was analyzed by negative staining electron microscopy as described previously (20). The usual sample concentration was ~20 μg/ml in 50 mM Tris-HCl (pH 7.4) and 0.15 M NaCl. Aliquots (5 μl) were adsorbed onto carbon-coated grids for 1 min, washed with 2 drops of water, and stained with 2 drops of 0.75% uranyl formate. The grids were rendered hydrophilic by glow discharge at low pressure in air. In some experiments, the N-terminal particles were identified with the anti-Myc mAb (9E10), the anti-GFP mAb (BD Biosciences), or the purified rabbit anti-MUC2N1 antisera, all labeled with colloidal thorocyanate gold (21). Specimens were observed in a Jeol JEM 1230 electron microscope operating at an accelerating voltage of 60 kV. Images were recorded with a Gatan Multiscan 791 CCD camera.

**RESULTS**

**Expression of the MUC2 N Terminus**—A plasmid expressing the MUC2 N terminus up to amino acid 1307 (7) followed by a Myc tag and GFP was constructed and called pSNMUC2-MG. The MUC2 sequence contains all of the MUC2 N terminus, including the last cysteine (position 1305), and stops at the beginning of the first and small mucin domain. The plasmid was transfected into CHO-K1 cells, and stable cell lines were selected from fluorescent colonies. These were further cloned and analyzed for secretion of the MUC2 N terminus. Several clones showed high-level and consistent secretion of proteins migrating identically upon gel electrophoresis when stained chemically or with antibodies reacting with SNMUC2-MG (anti-MUC2N3 antisera, anti-Myc mAb, and anti-GFP mAb). One of these clones was selected, expanded, and used for the purification of SNMUC2-MG.

To study the expression and sizes of the MUC2 N terminus expressed in the selected CHO cell lines, the cells were metabolically labeled, immunoprecipitated using the anti-Myc mAb, and analyzed by SDS-PAGE. Only one band was precipitated from the cell lysate with an apparent mass of ~220 kDa, and one band from the medium with an apparent mass of ~280 kDa, both analyzed in their reduced forms (using the sizes of the Precision protein standards) (Fig. 1A). The nonreduced secreted MUC2 N terminus gave only one large band with a size exceeding that of apoB-100 (512 kDa), which was used as a molecular mass marker. The lack of a marker larger than the N terminus made estimation of the size of the secreted oligomeric product difficult. However, the fact that it migrated significantly slower than the apoB-100 marker suggested that its oligomeric state was higher than a dimer.

To analyze the processing of the N terminus in CHO cells, pulse-chase studies were performed. The cells were metabolically labeled for 15 min, followed by a chase for up to 3 h, and analyzed by immunoprecipitation using the anti-Myc mAb and SDS-PAGE. The intracellular band was observed already after the pulse period, and this band vanished at the same time (90–135 min) that the secreted oligomer appeared in the medium (Fig. 1B). These results suggest that the N terminus of MUC2 is oligomerized relatively late in the secretory pathway and that the secreted form is present as a disulfide bond-stabilized oligomer. To further study the reason for the difference in size between the intracellular form and the secreted...
oligomer, both the intracellular and secreted forms were purified.

**Purification of SNMUC2-MG**—To purify SNMUC2-MG in its secreted oligomeric form as well as in its intracellular monomeric form, identical techniques were used. At first, both the culture medium and cell lysate were concentrated by four rounds of ultrafiltration through a 100-kDa cutoff filter with intermediate dilution with buffer. The concentrates thus lacked small components and were further purified by gel filtration on Sepharose 6. Fractions containing SNMUC2-MG were pooled and directly loaded onto a Mono Q ion-exchange column. The bound proteins were stepwise eluted with increasing amounts of NaCl. The secreted oligomeric form of SNMUC2-MG could be purified to homogeneity, as confirmed by SDS-PAGE, revealing the reduced monomers at an apparent molecular mass of 260 kDa (Fig. 2, lane 1). The purified intracellular monomers with an apparent molecular mass of 220 kDa were still contaminated by a weak band of ~170 kDa (Fig. 2, lane 2). On Western blots, the 260- and 220-kDa bands were detected by the anti-MUC2N3 antisera, the anti-GFP mAb, and the anti-Myc mAb. However, the 170-kDa band was stained only by the anti-MUC2N3 antisera, suggesting that it was a product of proteolytic cleavage, where the remaining protein lacked the Myc tag and GFP (data not shown).

**Glycosylation Status of the MUC2 N Terminus**—When the intracellular and secreted forms of SNMUC2-MG were analyzed in their reduced forms, a difference in size of ~40 kDa was revealed (Fig. 2). As glycosylation is one of the possible explanations for this, N-glycosylation and O-glycosylation were analyzed. CHO cells preferably make mono- and disialylated Galβ1-3GalNAc-(Ser/Thr), O-glycosylation can be studied by neuraminidase treatment, followed by O-glycosidase treatment. As shown in Fig. 3, when analyzed in their reduced forms, the intracellular form (lysate) was unaffected by these enzymes, whereas the secreted form was reduced in size by this treatment. The secreted form was still larger than the intracellular one, suggesting additional differences or incomplete O-deglycosylation.

When the N-glycans were studied by endoglycosidase H digestion, the intracellular N terminus was reduced in size, whereas the secreted one was unaffected (Fig. 3A). Both forms were affected by the N-glycosidase F treatment. That the intracellular form was endoglycosidase H-sensitive and lacked O-glycans suggested that this N terminus was trapped in the endoplasmic reticulum. This information, together with the pulse-chase studies, suggests a relatively slow folding process and an efficient endoplasmic reticulum exit control for the MUC2 N terminus, resulting in accumulation in this compartment.

The secreted form of SNMUC2-MG was still larger than the
intracellular form after removal of the O-glycans (Fig. 3A). Part of this is due to the N-glycans attached at the 10 potential sites, but all of the difference is not accounted for in this way. To further study this, the secreted and intracellular forms of the N terminus were subjected to anhydrous hydrogen fluoride treatment, known to effectively remove both N- and O-linked glycans, followed by PAGE (Fig. 3B). The sizes of both forms decreased, but secreted form of SNMUC2-MG was still slightly larger (10–20 kDa) than the intracellular form. The sizes of the endoglycosidase H- and hydrogen fluoride-treated intracellular forms of SNMUC2-MG were similar, suggesting that the hydrogen fluoride treatment effectively removed all glycans. This suggests that the secreted MUC2 N terminus has modifications in addition to glycosylation giving an apparent higher molecular mass. It is less likely that the intracellular form is smaller due to proteolytic cleavage, as this band was stained by all available antibodies against SNMUC2-MG, but this cannot be excluded.

Trypsin-resistant Core Fragment—To further study how the secreted oligomerized MUC2 N terminus was assembled, it was subjected to proteolytic digestion. The purified nonreduced and reduced/alkylated forms of SNMUC2-MG from the spent culture medium were treated with trypsin and analyzed by SDS-PAGE under reducing conditions. As expected, the reduced/alkylated form of SNMUC2-MG did not give any larger peptides analyzable by SDS-PAGE. However, the trypsin-digested nonreduced form of SNMUC2-MG gave a single distinct band migrating with an apparent molecular mass of 85 kDa when analyzed on a reducing gel (Fig. 4) compared with the non-trypsinized molecule (260 kDa). When trypsin-digested nonreduced SNMUC2-MG was analyzed under nonreducing conditions, a band with an apparent molecular mass of 240 kDa was observed (Fig. 4B). This suggests that the disulfide bond-stabilized N-terminal oligomer is held together within this fragment, here called a core. Kinetic studies revealed that the SNMUC2-MG protein was converted to the 85-kDa band within 1 h, and no further degradation was observed for up to 24 h (data not shown). It was possible to digest SNMUC2-MG to the core fragment also without the preincubation with 0.1% SDS, but SDS pretreatment gave more consistent results. A band with an apparent mass of ~85 kDa was also obtained when nonreduced SNMUC2-MG was treated (without SDS pretreatment) with the proteases chymotrypsin, Glu-C, and thermolysin. Proteolytic digestion with Pronase gave a single band of ~60 kDa (data not shown). Subtilisin was the only protease tested that degraded nonreduced SNMUC2-MG to smaller peptides.

To delineate the part of the MUC2 N terminus that was involved in the oligomerization corresponding to the core fragment, epitope mapping and protein sequencing studies were performed. Two antisera (anti-MUC2N1 and anti-MUC2N3) against the MUC2 N terminus as well as the anti-Myc mAb were used in Western blot experiments on trypsin-digested SNMUC2-MG separated by SDS-PAGE under reducing conditions (Fig. 4A). The anti-MUC2N1 antiserum did not react with the 85-kDa core fragment, whereas the anti-MUC2N3 antiserum did, revealing that the C-terminal part of MUC2 in SNMUC2-MG was part of the core. The Myc epitope was inserted directly after the MUC2 N terminus, replacing the first Arg1022 in the MUC2 sequence. The C terminus of the core was localized within the last part of the D3 domain of MUC2 (7). As the anti-Myc mAb did not react with the core fragment, the core was localized within the last part of the D3 domain of the MUC2 N terminus as outlined in Fig. 5.

The N-terminal end of the core fragment was determined by Edman sequencing of the 85-kDa core fragment after SDS-PAGE and blotting of trypsin-treated SNMUC2-MG. A single sequence was obtained, EAPTXDP, where X corresponds to a cysteine that could not be analyzed in this experiment (Fig. 5 and Table I). The trypsin-resistant core fragment thus had a homogeneous N-terminal end and was due to cleavage after Arg1022 in the MUC2 sequence. The C terminus of the core fragment has to extend at least over amino acids 1168–1181, as the anti-MUC2N3 antiserum showed reactivity (see Fig. 5).
To further determine the C-terminal end of the core fragment, it was purified and subjected to SDS-PAGE under reducing conditions. The 85-kDa band was in-gel trypsin-digested, and the peptides obtained were analyzed by MALDI-MS and ESI-MS using nanospray on a Q-ToF mass spectrometer. The trypsin cleavage sites are shown in Fig. 5, and the observed peptides in Table I. The molecular mass of the peptide corresponding to the N-terminal end as revealed by Edman sequencing was found at m/z 2534.1 in both the MALDI-MS and ESI-MS experiments. Several of the internal peptides were observed, and some were possible to sequence by ESI-MS/MS (underlined sequences in Fig. 5). The peptides containing amino acids 1147–1166 and 1228–1251 contained one and two potential N-glycosylation sites, respectively, and were detected only after N-glycanase treatment.

The most C-terminal peptide found was at m/z 1022.42 and 1012.97 (masses of 2043.8 and 2024.8 Da for the mercaptoethanol- and carboxamidomethyl-modified cysteine, respectively) and corresponded to amino acids 1367–1383. This peptide could be sequenced by the electrospray mass spectrometer as shown in Fig. 5. This sequence can be read from the C-terminal end as y ion series all the way up to y10 at m/z 1800.8 and from the N terminus as b ion series up to b10 at m/z 634.3. The masses revealed that the cysteine had been modified by mercaptoethanol and that Asn deaminated to Asp, preparatory artifact.

A summary of the Edman sequencing is presented in Table I. The sequence had been deaminated into Asp. To determine whether the core fragment was still held together as an oligomer, nonreduced SNMUC2-MG purified from the spent culture medium was treated with trypsin and analyzed before and after reduction by SDS-PAGE. The nonreduced core fragment gave a band at ~240 kDa (Fig. 4B). Upon reduction, this was reduced to 85 kDa. Both of these bands were now within range of the molecular mass markers, allowing accurate size estimations. The size discrepancy due to the glycosylation of the core fragment and the core fragment itself with or without deglycosylation is presented in Table I. The size difference between the nonreduced and reduced core fragment was ~90 kDa. As reduced SNMUC2-MG eluted at 10.70 ml, slightly before the largest molecular mass standard, thyroglobulin, with a mass of 669 kDa, which eluted at 10.70 ml (see “Experimental Procedures”). The size of the MUC2 N terminus could thus not be accurately determined, but was estimated to be 690–750 kDa. As reduced SNMUC2-MG had an estimated mass of 260 kDa, the MUC2 N terminus appears to be organized into a trimer.

Electron Microscopy of the Secreted MUC2 N-terminal Tri-mer—To further analyze the oligomeric status as well as macromolecular structure, the MUC2 N terminus was subjected to electron microscopy after negative staining. To be able to localize the domains, anti-Myc, anti-GFP, and immunopurified anti-MUC2N1 antibodies were gold-labeled, and SNMUC2-MG were immunostained. The overall picture of the MUC2 N terminus is a variable three-dimensional structure (Fig. 7). However, a general molecular structure with a central domain connected with single flexible threads to three identical globular domains could be observed. These flexible parts make the images variable, with many molecules less well spread and exposed on the grid. The central domain, holding the molecule together, was stained by both the anti-GFP (Fig. 7, C–E, and F) and anti-Myc (Fig. 7D) antibodies. Only one antibody dot could be observed per molecule despite the fact that each trimer should contain three sites. The reason for this it not known, but is most likely due to steric hindrance, an interpretation supported by the observation that it is impossible to catch and

Table I

| Peptide sequence | aa position | [M + H]+calc | [M + H]+obs | MALDI-MS | ESI-MS | Comments (Edman sequencing) |
|------------------|-------------|-------------|-------------|----------|--------|-----------------------------|
| EAPTYPD         | 1023–1029   | 2534.15     | 2534.13     | Cys(Cm)  |
| EAPTDPVISNEPCSLPHRR | 1023–1044 | 1024.47     | 1024.50     | Cys(Cm)  |
| EGACVFWR        | 1105–1112   | 2300.07     | 2301.11     | Cys(ME), N → D (1149e, 1154f) |
| TINGHHSNISVSYLEGCPY | 1147–1166 | 2301.05     | 2301.05     | Cys(ME), N → D (1149e, 1154f) |
| DRPIYEDLKK      | 1170–1179   | 1277.64     | 1277.65     | Cys(ME), N → D (1149e, 1154f) |
| ILNQTDGDACFYEICGPNNTVEK | 1228–1251 | 2840.19     | 2840.30     | Cys(ME), N → D (1230f, 1246f) |
| VQCDVSVGFIK     | 1355–1366   | 1405.75     | 1405.81     | Cys(ME), N → D (1237f) |
| NEDQFGNPPGLCYDYK | 1367–1383 | 2043.81     | 2043.81     | Cys(ME), N → D (1373f) |
| D (1373 )      | 2024.83     | 2024.81     | 2024.81     | Cys(ME), N → D (1373f) |
| D (1149 )      | 2043.81     | 2043.81     | 2043.81     | Cys(ME), N → D (1373f) |

a, amino acid; Cys(Cm), carboxamidomethylated Cys; Cys(ME), mercaptoethanol-modified Cys.

b, Asn deaminated to Asp, preparatory artifact.

c, Asn to Asp due to deglycosylation with N-glycosidase F.
detect this protein with the same antibody (anti-Myc or anti-
GFP) in sandwich enzyme-linked immunosorbent assay exper-
iments. When the target was a little more strongly negatively
stained, the three large globular domains were less pro-
nounced, whereas the central domain was more easily visual-
ized (Fig. 7, E and F). In some molecules, the central part had
an electron-dense trefoil-like structure (Fig. 7 F). This trefoil
could be part of the D3 domain where the MUC2 N terminus is
held together. The flexible thread is probably a single peptide
chain linking the central core with the large globular domain
made up of the D1 and D2 domains (see Fig. 5). This interpre-
tation is supported by the gold-labeled anti-MUC2N1 antibody,
generated against a peptide from the D1 domain (Fig. 5). This
antibody can stain one, two, or three of these globular domains
(Fig. 7, G–I). The gold particles are located farthest away from
the central core with the large globular domain
made up of the D1 and D2 domains (see Fig. 5). This interpre-
tation is supported by the gold-labeled anti-MUC2N1 antibody,
generated against a peptide from the D1 domain (Fig. 5). This
antibody can stain one, two, or three of these globular domains
(Fig. 7, G–I). The gold particles are located farthest away from
the central parts, suggesting that the N-terminal D1 domain is
located in the outer part of the molecule. Some images (Fig.
7G) suggest a substructure in the large globular parts, maybe
due to the D1 and D2 domains. This interpretation of the
electron microscopy images is schematically explained in Fig.
7K. The size of the large globular domain is estimated to be
~10 nm, and the length of one subunit is estimated to be 20–25
nm when maximally extended. The extended parts have not
been localized in detail due to a lack of functional antibodies.
It can thus be concluded that the MUC2 mucin N terminus is
oligomerized into a trimer and that this is held together with
disulfide bonds in a trypsin-resistant, relatively compact do-
main made up by amino acids 1023–1383 of MUC2.

DISCUSSION

CHO cells permanently expressing and secreting the N ter-
minus of the MUC2 mucin in high quantities were instrumen-
tal in the studies presented here. The selection of clones was
greatly facilitated by the presence of GFP at the C terminus, as
only green cell colonies were picked and screened for secretion.
The initial use of COS cells was abandoned, as these cells
secrete only small amounts of the recombinant protein, proba-
bly due to a less efficient endoplasmic reticulum folding ma-
chinery for this protein. However, CHO cells also seem to have
some difficulties in folding the MUC2 N terminal, as suggested
by the presence of relatively large amounts of endoglycosidase
H-sensitive MUC2 monomers in the cell lysates. Despite this,
the spent culture medium contained large quantities of the
recombinant protein, allowing an easy purification procedure
and subsequent biochemical studies.

Already at the time of cDNA sequencing of the MUC2 mucin
(7), it was suggested that the assembly of this mucin followed
the principles of the vWF (13, 14). The assembly process is
initiated in the endoplasmic reticulum, where dimers are
formed by the C terminus via its cysteine knot. This has been
shown to be the case for the full-length vWF (14), MUC2 (19,
22), and MUC5AC (23) as well as for the recombinantly ex-
pressed C terminus of PSM (4), rat Muc2 (24), and human
MUC2.3 That this process is localized to the endoplasmic retic-
ulum is suggested by the rapid formation of the dimers, the
disulfide bond-stabilized oligomer into the medium. The forma-

3 M. E. Lidell, M. E. V. Johansson, N. Asker, C. Weinberg, J. R. Gum,
Jr., Y. S. Kim, and G. C. Hansson, unpublished data.
Trimeric Trypsin-resistant Core of the MUC2 Mucin

Fig. 7. Electron microscopy of the purified SNMUC2-MG protein trimer stained with gold-labeled antibodies. The purified SNMUC2-MG protein was immunostained on the grid with gold-labeled anti-Myc (D), gold-labeled anti-GFP (A–C, E, and F), and gold-labeled purified anti-MUC2N1 (G–I) antibodies. The grid was negatively stained for a shorter time in B–D and for a longer in A, E, and F and analyzed by electron microscopy. A schematic of the SNMUC2-MG trimer and a potential interpretation of the pictures are given in K. The arrow in F shows the central trefoil-like structure holding the trimer together. The images shown are representative of the structures captured in an extended form; but due to the flexible parts, some structures were more curled up, and the structure less well presented. The bars represent 50 nm in A and 25 nm in B–I (shown in I).

The formation of N-terminal trimers in MUC2 and PSM suggests that at least two mucin species do not form linear polymers. The easiest interpretation is that these mucins will form branched networks. This may not be consistent with most of the biophysical information for soluble mucins, including PSM, which has predicted linear polymers. This interpretation was based on light scattering, ultracentrifugation, and molecular electron microscopy studies (15, 16, 25). One possibility is that MUC2 and PSM are different from the other gel-forming mucins. This is true for MUC2, as it is known to form insoluble complexes (8, 10), as is a minor portion of MUC5B (26); but there is no information suggesting that PSM can become insoluble. To understand this, we have to await further studies on the other gel-forming mucins to determine whether they follow the suggested assembly for MUC2 and PSM or that of vWF with linear polymers. We observed the secretion of only N-terminal trimers; and thus, it is less likely that forms of MUC2 other than N-terminal trimers will be secreted from normal cells. However, it cannot be excluded that the mucin domain that should follow the N terminus could interfere or modulate the formation of N-terminal trimers in vivo. During studies of the assembly of full-length MUC2 in LS174T cells, we observed that not only C-terminal dimers of MUC2 are transferred out of the endoplasmic reticulum and become glycosylated in the Golgi apparatus, but also O-glycosylated monomers were detected (11). The reason for this is not known, but could be due to some MUC2 escaping the dimerization process or a yet undefined proteolytic cleavage. As the dimerization in the C terminus is believed to be an endoplasmic reticulum-specific process, these monomers cannot be extended once they enter into the Golgi apparatus or other later parts of the secretory pathway. The N terminus of these MUC2 monomers can most likely be incorporated into N-terminal trimers and could as such block further elongation. There is currently no proof of such a hypothetical model of mucin assembly, but regulation of the level of mucin monomers passing out of the endoplasmic reticulum could be one out of several ways to regulate mucin structure and thus its properties (11).

The three identical peptides that made up the trypsin-resistant core fragment were localized to the most C-terminal end of the MUC2 N terminus. This shows that the disulfide bonds holding the N terminus together must be localized in this part of the molecule. This is not a surprise, as this has been sug-
gested for the vWF and PSM (4). Site-directed mutagenesis of recombinantly expressed PSM (6) suggests that Cys \textsuperscript{1130} of MUC2 in the conserved sequence EC\textsuperscript{E}WHY is one of the cysteines forming the intermolecular disulfide bonds. The localization of this cysteine in the core fragment is marked in Fig. 5. To allow for trimerization, there must be at least one additional intermolecular disulfide bond stabilizing the N-terminal trimer. The localization of this cysteine is not known today.

The molecular structure of the MUC2 N-terminal trimer as revealed by electron microscopy is consistent with the sizes of the domains and their potential function. The D1 and D2 domains are likely to harbor the disulfide bond isomerase activity catalyzing the trimerization. These domains are probably localized to the three large globular domains that were attached to the central core with a single flexible peptide. The precise localization of this peptide has not been determined, but it could be in the D’ domain or may be more likely a part of the D3 domain, as this contains a long sequence (72 amino acids) lacking Cys (positions 917–988). The flexibility of this link could allow the enzymes to move and come close to their catalytic targets in the central core. This core, probably largely made up by the D3 domain, could be localized with two antibodies directed toward the recombinantly attached Myc tag and GFP. This core was revealed as a very dense trefoil-like structure, consistent with the localization of the trypsin-resistant core to this region. The apparent localization of the GFP and Myc tag to the stalk of the trefoil (Fig. 7) suggests that the first small mucin domain of MUC2 is localized here, close to the central core. That the flexible link between the large globular domains and the central core is susceptible to trypsin cleavage is understandable. The D1 and D2 domains can probably also be cleaved off under physiological circumstances as shown for the vWF (4) and MUC5B from salivary glands (26). However, no cleavage of the MUC2 N terminus was observed in CHO cells, nor has this been found for PSM expressed in COS cells. This study suggests that despite the flexible nature of this part, there is a need for specific enzymes present only in certain cells.

The native MUC2 mucin produced in intestinal cells, both in cell culture and in cells in situ, is also held together by additional non-disulfide intermolecular bonds (10, 11). The nature of these bonds, as well as their location within the MUC2 mucin, is presently unknown. Reducing the SNMUC2-MG protein expressed in CHO cells gave only monomers; and thus, we have not been able to find any nonreducible linkage during these studies. The reason for this could be either that these bonds are formed in other parts of MUC2 or that necessary components for the formation of this type of linkage are missing in CHO cells.

The goblet cells of the small and large intestines are the normal site for MUC2 mucin biosynthesis. One of the main challenges of a digestive system is protection against digestive enzymes. The MUC2 mucin is the main intestinal mucin localized to the mucous gel covering the epithelial cells. This suggests that it has an important role in this protective system and that it must be able to withstand the action of pancreatic enzymes. How this works is easy to understand for the mucin domains, as the peptide core is protected by the dense O-glycosylation. However, it is even more important to protect the oligomeric nature of MUC2 as organized by the N and C termini. The present finding of a trypsin-resistant core fragment still held together as a trimer located in close proximity to the highly glycosylated domain and therefore able to maintain the oligomeric nature of the mucin must be physiologically very important. This proteolytic resistance is not maintained by a lack of amino acids where trypsin can cleave. The core fragment is resistant to cleavage by not only trypsin, but also by most proteolytic enzymes tested except subtilisin. Although no information is currently available on how this proteolytic resistance is maintained, the high number of intramolecular disulfide bonds must be crucial, as reduction makes the MUC2 N terminus just as degradable as other proteins. We have thus been able to express the MUC2 N terminus and to show that it forms trimers held together within a trypsin-resistant core fragment, suggesting that mucins will form branched networks and not simple linear polymers.

Acknowledgments—We are indebted to Karin Axelsson and Maria Baumgarten for excellent technical assistance.

REFERENCES

1. Gendler, S. J., and Spicer, A. P. (1995) Annu. Rev. Physiol. 57, 697–634
2. Farstner, J. F., and Farstner, G. G. (1994) in Physiology of the Gastrointestinal Tract (Johnson, L. R., ed) 3rd Ed., pp. 1255–1263, Raven Press, New York
3. Perez-Vilar, J., and Hill, R. L. (1999) J. Biol. Chem. 274, 31753–31754
4. Perez-Vilar, J., and Hill, R. L. (1998) J. Biol. Chem. 273, 6982–6988
5. Perez-Vilar, J., Eckhardt, A. E., de Luca, A., and Hill, R. L. (1998) J. Biol. Chem. 273, 14442–14449
6. Perez-Vilar, J., and Hill, R. L. (1998) J. Biol. Chem. 273, 34527–34534
7. Gum, J. R., Jr., Hicks, J. W., Toribara, N. W., Siddiki, B., and Kim, Y. S. (1994) J. Biol. Chem. 269, 2440–2446
8. Carlstedt, I., Herman, A., Karlsson, H., Sheehan, J., Fransson, L.-A., and Hansson, G. C. (1993) J. Biol. Chem. 268, 18771–18781
9. Hansson, G. C., Backstrom, D., Carlstedt, I., and Klinga-Levan, K. (1994) Biochem. Biophys. Res. Commun. 196, 181–190
10. Herrmann, A., Davies, J. R., Lindell, G., Märtensson, S., Packer, N. H., Swallow, D. M., and Carlstedt, I. (1999) J. Biol. Chem. 274, 15828–15836
11. Axelsson, M. A. B., Asker, N., and Hansson, G. C. (1998) J. Biol. Chem. 273, 18864–18870
12. Li, J.-D., Dohrman, A. F., Gallup, M., Miyata, S., Gum, J. R., Kim, Y. S., Nadel, J. A., Prince, A., and Basbaum, C. B. (1997) Proc. Natl. Acad. Sci. U. S. A. 94, 967–972
13. Wagner, D. D. (1990) Annu. Rev. Cell Biol. 6, 217–246
14. Sadler, J. E. (1996) Annu. Rev. Biochem. 67, 395–424
15. Carlstedt, I., Lindingen, H., and Sheehan, J. K. (1983) Biochem. J. 213, 427–435
16. Sheehan J. R., and Thornton, D. J. (2000) Methods Mol. Biol. 125, 87–96
17. Toribara, N. W., Gum, J. R., Jr., Culhane, P. J., Lague, R. E., Hicks, J. W., Petersen, G. M., and Kim, Y. S. (1991) J. Clin. Invest. 88, 1005–1013
18. Blum, H., Beier, H., and Gross, H. J. (1987) Electrophoresis 8, 93–99
19. Asker, N., Axelsson, M. A. B., Olofsson, S.-O., and Hansson, G. C. (1998) J. Biol. Chem. 273, 18864–18870
20. Engel, J., and Pumphrey, H. (1987) Methods Enzymol. 145, 3–78
21. Baecher, W., Lucoq, J. M., and Roth, J. (1985) Histochemistry 83, 409–411
22. Asker, N., Baechstrom, D., Axelsson, M. A. B., Carlstedt, I., and Hansson, G. C. (1995) Biochem. J. 308, 873–880
23. Asker, N., Axelsson, M. A. B., Olofsson, S.-O., and Hansson, G. C. (1998) Biochem. J. 335, 383–397
24. Bell, S. L., Khatri, I. A., Xu, G. Q., and Farstner, J. F. (1998) Eur. J. Biochem. 253, 123–131
25. Sheehan, J. K., Oates, K., and Carlstedt, I. (1986) Biochem. J. 239, 147–153
26. Wickstrom, C., and Carlstedt, I. (2001) J. Biol. Chem. 276, 47116–47121
The N Terminus of the MUC2 Mucin Forms Trimers That Are Held Together within a Trypsin-resistant Core Fragment
Klaus Godl, Malin E. V. Johansson, Martin E. Lidell, Matthias Mörgelin, Hasse Karlsson, Fredrik J. Olson, James R. Gum, Jr., Young S. Kim and Gunnar C. Hansson

J. Biol. Chem. 2002, 277:47248-47256.
doi: 10.1074/jbc.M208483200 originally published online October 8, 2002

Access the most updated version of this article at doi: 10.1074/jbc.M208483200

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

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

This article cites 25 references, 15 of which can be accessed free at http://www.jbc.org/content/277/49/47248.full.html#ref-list-1