Studies on the “Insoluble” Glycoprotein Complex from Human Colon

IDENTIFICATION OF REDUCTION-INSSENSITIVE MUC2 OLIGOMERS AND C-TERMINAL CLEAVAGE

(Received for publication, October 29, 1998, and in revised form, January 8, 1999)

Ankatrin Herrmann‡, Julia R. Davies‡, Gert Lindell‡, Stig Mårtensson‡, Nicole H. Packer†, Dallas M. Swallow‡‡, and Ingemar Carlstedt‡ ‡‡

From the ‡Mucosal Biology Group, Department of Cell and Molecular Biology, Section for Molecular Pathogenesis, Lund University, P. O. Box 94, 221 00 Lund, Sweden, the Departments of ‡Surgery and ‡Urology, Lund University Hospital, 221 85 Lund, Sweden, the ‡Macquarie University Centre for Analytical Biotechnology, Sydney 2019, Australia, and **Human Biochemical Genetics Unit, University College, London, NW1 2HE United Kingdom

The “insoluble” glycoprotein complex was isolated from human colonic tissue and mucin subunits were prepared following reduction. Antibodies raised against peptide sequences within MUC2 revealed that virtually all of this mucin occurs in the insoluble glycoprotein complex. In addition, reduction released a 120-kDa C-terminal MUC2 fragment, showing that proteolytic cleavage in this domain may occur and leave the fragment attached to the complex via disulfide bonds. The variable number tandem repeat region and the irregular repeat domain were isolated after trypsin digestion and shown to have molecular weights of 930,000 and 180,000, respectively, suggesting a molecular weight for the entire MUC2 monomer of approximately 1.5 million. Gel chromatography and agarose gel electrophoresis revealed several populations of MUC2 subunits, and analytical ultracentrifugation showed that these have molecular weights on the order of 2 million, 4 million, and 5 million, corresponding to monomers, dimers, and trimers, respectively. Agarose gel electrophoresis of subunits from individuals expressing both a “long” and a “short” MUC2 allele revealed a larger number of populations, consistent with the presence of short and long monomers and oligomers arising from permutations of the two types of monomers. In addition to disulfide bonds, MUC2 monomers are apparently joined by a “novel,” reduction-insensitive bond.

The mucosal surface of the gastrointestinal tract is protected by a visco-elastic mucus gel formed by high molecular mass (0.5–25 × 10^6) glycoproteins referred to as mucins. The protein backbones of mucins are heavily substituted with O-linked oligosaccharides attached to serine and/or threonine residues, and these amino acids are, together with proline, typically enriched within so-called mucin domains. Several mucin domains may occur in a single mucin subunit and these regions are flanked by less heavily glycosylated segments of the protein core. Large secreted mucins from the respiratory tract, stomach, and endocervix have been shown to be linear oligomers formed by subunits joined via disulfide bonds (1, 2). However, less is known about the macromolecular architecture of intestinal mucins.

The mucin genes MUC2, MUC3, MUC4, MUC5B, and MUC6 are expressed in normal human colon (3–7). MUC2 is believed to be the dominating mucus-forming species in this tissue and is, so far, the only “large” mucin of which the cDNA has been fully sequenced (8). The apoprotein contains two mucin domains, which differ in length and are separated by a cysteine-containing region (Fig. 1A). The longer domain, which is referred to as the variable number tandem repeat (VNTR) region is composed largely of tandemly repeated 23 amino acid peptide units, which vary in number between alleles and are rich in threonine and proline (9, 10). The shorter mucin domain comprises a 347-amino acid-long irregular repeat rich in threonine, serine, and proline. The regions flanking the mucin domains each show a high degree of similarity to the N- and C-terminal cysteine-containing D-domains of the prepro-von Willebrand factor (8), and it has therefore been suggested that MUC2 is processed and oligomerized in a fashion similar to this protein. In support of this hypothesis, it has been shown that dimerization occurs before O-glycosylation and the formation of larger structures (11, 12). The temporal relationship between further oligomerization and O-glycosylation is not clear, but data from Sheehan et al. (13) suggest that there is a slow assembly of what appears to be fully glycosylated units. Other investigators have identified biosynthetic intermediates that have been interpreted as being O-glycosylated monomers and dimers (14). However, the presence of an O-glycosylated monomer population is in apparent conflict with dimerization preceding O-glycosylation.

The rat Muc2 homologue has been partially cloned and sequenced (15) and shows significant homology to human MUC2, both in the N and C termini (16, 17). Thiol reduction of rat and human small intestinal mucins gives rise to a 118-kDa glycoprotein, previously referred to as the “link” protein (18, 19). However, cDNA analysis has shown that this structure is, in fact, part of the MUC2/Muc2 gene (17), and the finding is explained by proteolytic cleavage within the C-terminal domain of MUC2/Muc2. Proteolytic cleavage in the N-terminal region of rat Muc2 has recently been reported (20).

Mucins isolated from rat small intestine are assembled into a complex that cannot be solubilized by breaking noncovalent

* This work was supported by Swedish Medical Research Council Grants 7902, 9823, and 9711; the Medical Faculty of Lund; the Swedish Cancer Foundation; Crafoordska Stiftelsen; The European Union (contract number BMH4-CT96-2222); the Swedish Foundation for Health Care Sciences and Allergy Research; and the National Health and Medical Research Council of Australia (Grant 990475). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

†† To whom correspondence should be addressed. Tel.: 46-46-222-46-49; Fax: 46-46-15-77-56; E-mail: Ingemar.Carlstedt@medkem.lu.se.

‡‡ To whom correspondence should be addressed. Tel.: 46-46-15-77-56; E-mail: Ingemar.Carlstedt@medkem.lu.se.

1 The abbreviations used are: VNTR, variable number tandem repeat; PAS, periodic acid-Schiff; PBS, phosphate-buffered saline; Fmoc, N-(9-fluorenylmethoxycarbonyl).
bonds; however, reduction of disulfide bonds brings the complex into solution. Two high molecular mass glycopeptides (650 and 335 kDa, respectively) were isolated after reduction and trypsin digestion of the insoluble glycoprotein complex from rat small intestine (21). The larger of these glycopeptides apparently corresponds to a large mucin domain from rat Muc2 (22). In the human colon, most of the mucins also appear as a glycoprotein complex that is insoluble in guanidinium chloride (23).

Here, we have isolated the insoluble glycoprotein complex from human colon and shown that it contains virtually all the MUC2 mucin in this tissue. Reduction of the complex revealed oligomeric forms of the MUC2 monomer that are insensitive to reduction and, in some samples, proteolytic cleavage was found to occur in the C-terminal part of the apoprotein. Preliminary communications have already been published (24, 25).

**EXPERIMENTAL PROCEDURES**

**Materials**—Disopropylphosphofluoridate was from Fluka, dithiothreitol was from Merck, and N-ethylmaleimide and trypsin (EC 3.4.21.4., type XIII, treated with 1-(1-tosylamido-2-phenyl) ethyl chloroformate from Sigma). Guanidinium chloride (practical grade) was obtained from Acros Organics (Geel, Belgium) and was treated with activated charcoal and filtered through an Amicon PM-10 membrane before use. Synthetic peptides conjugated to keyhole limpet hemocyanin were purchased from Scandinavian Peptide Synthesis AB (Köping, Sweden). Sephacryl S-500 HR was from Amersham Pharmacia Biotech, Immobilion-P polyvinylidene difluoride membranes from Millipore, the ECL Western detection kit was from Amersham Pharmacia Biotech, and Ultrapure agarose was from Life Technologies, Inc. Digoxigenin-3-O-succinyl-e-amido-caproic acid hydrazide and alkaline phosphatase-conjugated anti-digoxigenin antibodies were from Roche Molecular Biochemicals. Horseradish peroxidase and alkaline phosphatase-conjugated swine anti-rabbit antibodies were bought from Dako (Glostrup, Denmark). Bovine serum albumin (fraction V, pH 7.0) was from Serva.

**Analytical Methods**—Density was determined by weighing aliquots. Total sialic acid was measured according to Davies *et al.* (26) with a further modification to allow alkali treatment (0.5 M NaOH for 5 min) of the samples on-line. Carbohydrate was determined by slot-blotting aliquots onto nitrocellulose membranes followed by staining with periodic acid-Schiff (PAS) (27) or by peroxidase-labeled swine anti-rabbit antibodies (swine anti-rabbit 1:2000 in blocking solution) was added to the reservoir washings and reservoir urine were concentrated over a PM-10 membrane (Amicon).

All fractions (sols, material solubilized with guanidinium chloride during extractions and material obtained after reduction/alkylation of the insoluble glycoprotein complexes) were dialyzed against 6 M guanidinium chloride, 1 M disopropylphosphofluoridate, 5 mM N-ethylmaleimide, 10 mM sodium phosphate buffer, pH 7.0, and left stirring at 4 °C overnight. After centrifugation (Beckman J2-MC centrifuge; JA 20 rotor; 5000 rpm; 10 °C; 60 min). Reservoir urine samples were obtained from the urine of patients with established neobladders. Sodium azide was added as above, and samples were stored at 4 °C until centrifugation (Beckman J2-MC centrifuge; JA 20 rotor; 18,000 rpm; 60 min) were performed to recover the mucous gel. Mucosal scrapings and the gel phase from reservoir washings and reservoir urine were gently dispersed with a Dounce homogenizer in 6 M guanidinium chloride, 1 M disopropylphosphofluoridate, 5 mM N-ethylmaleimide, 10 mM sodium phosphate buffer, pH 7.0, and left stirring at 4 °C overnight. After centrifugation (Beckman J2-MC centrifuge; JA 20 rotor; 18,000 rpm; 4 °C; 60 min), the soluble material was removed, and the pellets were re-extracted twice, as described above. The final extraction residues (insoluble glycoprotein complex) were solubilized by reduction in 6 M guanidinium chloride, 10 mM dithiotreitol, 5 mM EDTA, 0.1 M Tris-HCl buffer, pH 8.0, at 37 °C overnight and then alkylated with 25 mM iodoacetamide in the same buffer. After centrifugation (Beckman J2-MC centrifuge; JA 20 rotor; 18,000 rpm; 4 °C; 60 min), the supernatant was retained. The soluble phases (sols) obtained after the initial centrifugation to recover the mucous gel from the reservoir washings and reservoir urine were centrifuged over a PM-10 membrane (Amicon).

**FACS**

**Isolation and N-terminal Amino Acid Sequencing of a MUC2 C-terminal Cleavage Fragment**—Low density material reacting with the LUM2-4 antisera, partially separated from the major mucin band in the density gradients, was pooled (see Fig. 3D, peak II) and subjected to gel chromatography on a Sephacryl S-500HR column (16 × 50 cm) eluted with 4 M guanidinium chloride, 10 mM sodium phosphate buffer, pH 7.0, at a flow rate of 0.1 ml/min. Fractions (1 ml) were analyzed for density, absorbance at 280 nm, sialic acid (with and without pre-treatment with 0.5 M NaCl), PAS reactivity, and reactivity against the LUM2-2, LUM2-3, and LUM2-4 antisera.

**Enzyme-linked Immunosorbent Assay**—Samples (100 μl) were coated onto multwell assay plates overnight at 4 °C followed by blocking (200 μl) with PBS (0.15 M NaCl, 5 mM sodium phosphate buffer, pH 7.4) containing 0.5% bovine serum albumin (w/v) and 0.05% (v/v) Tween 20 (blocking solution) for 1 h at room temperature. The wells were incubated for 1 h with the LUM2-2, LUM2-3, or LUM2-4 antisera (1:1000) in blocking solution. Alkaline phosphatase-conjugated secondary antibody (swine anti-rabbit IgG) was added to the wells (1 h at room temperature) and reactivity was detected as absorbance at 405 nm and 2 h after blocking with naphthol-AS-BI phosphate (2 mg/ml in 1 M diethanolamine-HCl buffer, pH 9.8, containing 0.5 mM MgCl₂).
Coomassie-stained band from the 50 μg sample corresponding to that revealed by the LUM2-4 antisemur was cut out and subjected to Edman degradation using a pulsed liquid-phase sequencer model 457A (Applied Biosystems) equipped with a 120A analyzer for on-line detection of phenylthiohydantoin amino acids.

Isolation of Mucin Subunit Subfractions—Mucin subunits were pooled after density gradient centrifugation (see Fig. 3D, peak I) and chromatographed on a Sephacryl S-500HR column (1.6 × 50 cm) eluted with 4 mM guanidinium chloride buffer at a flow rate of 0.1 ml/min. Fractions (1 ml) were analyzed for sialic acid and reactivity against the LUM2-2, LUM2-3, and LUM2-4 antisera. In order to obtain mucin subfractions, the procedure was repeated 10 times, and material was pooled as indicated in Fig. 7A (sub-I, sub-II, and sub-III) and concentrated over a PM-10 filter (Amicon). In order to remove material that may have detached from the column matrix and would interfere with further analyses, the samples were finally subjected to density gradient centrifugation in CsCl/4 M guanidinium chloride (70.1 Ti rotor), and the mucin band was recovered.

Agarose Gel Electrophoresis and Western Blotting—Mucin subunits were dialyzed into 40 mM Tris acetate buffer, pH 8, containing 1 mM EDTA, and prior to electrophoresis, 10% (v/v) of a solution containing glycerol 30% (v/v), 1 mM EDTA, 1% SDS, 0.002% bromophenol blue in 40 mM Tris acetate buffer, pH 8, was added. Agarose gel electrophoresis was performed in a Bio-Rad DNA subcell for 18 h at 30 V and room temperature using 1% (w/v) agarose gels (15 × 15 cm) prepared in 40 mM Tris acetate buffer, pH 8, containing 1 mM EDTA and 0.1% SDS. After electrophoresis, gels were blotted onto nitro-cellulose membranes for 6 h at 40 mbar with an Amersham Pharmacia Biotech LKB VacuGene XL vacuum blotter using 0.6 M NaCl, 0.06 M sodium citrate buffer, pH 7.0.

After blotting, membranes were blocked with 0.5% (w/v) bovine serum albumin in PBS containing 0.05% (v/v) Tween 20 (blocking solution) for 1 h and incubated with the LUM2-2, LUM2-3, or LUM2-4 antisemur (1:1000) diluted in blocking solution for 1 h. Bound antibody was detected by incubation with horseradish peroxidase-conjugated swine anti-rabbit antibody (1:2000 in blocking solution) for 1 h followed by the ECL Western detection kit. All incubations were carried out at room temperature.

Isolation of High Molecular Mass Tryptic Glycopeptides—Mucin subunits (see Fig. 3D, peak I) were dialyzed against 0.1 M NH4HCO3, pH 8.0, treated with trypsin (50 μg/ml) overnight at 37 °C, dialyzed against 4 mM guanidinium chloride, 10 mM sodium phosphate buffer, pH 7, and subjected to gel chromatography on a Sephacryl S-500HR column (1.6 × 50 cm) eluted with the same buffer at a flow rate of 0.1 ml/min. Fractions (1 ml) were collected and analyzed for sialic acid. The material was pooled into two fractions, referred to as glycopeptides A and B (see Fig. 5), and subjected to density gradient centrifugation in CsCl/4 M guanidinium chloride as described for the mucin subunits in order to ensure complete removal of peptide material and contaminants from the column matrix. Finally, the two glycopeptide populations were dialyzed against water and freeze-dried. A larger pool of material was obtained using a column (2.6 × 93 cm) eluted with the same buffer at a flow rate of 0.27 ml/min.

Amino Acid Analysis of the High Molecular Mass Tryptic Glycopeptides—Amino acid analysis was carried out on approximately 2.5 μg of glycopeptides A and B according to Ref. 32 with modified gas-phase hydrolysis conditions. Samples were manually hydrolyzed in a vessel with 500 μl of 5.7 M HCl and a crystal of phenol at 110 °C for 24 h. The protein hydrolysate was dissolved in 250 mM borate buffer (pH 8.5), and analyzed with a GBC automated Aminomat HPLC system (GBC Scientific, Australia). Amino acids were derivatized with Fmoc and separated on a reversed-phase column (ODS-Hypersil, 5-μm packing, 150 × 4.6 mm internal diameter, Keystone Scientific) at 38 °C. Amino acid Fmoc derivatives were detected by fluorescence (LC1250 fluoro-detector system; excitation wavelength, 270 nm; emission wavelength, 316 nm), and peak analysis was controlled by the WinChrom chromatography data system (GBC), version 1.2.

Monosaccharide Analysis of the High Molecular Mass Tryptic Glycopeptides—Monosaccharide analysis was carried out on mucin aliquots from the same solution (2.5 μg/μl) as described above. Sialic acids were hydrolyzed from the mucin in 0.1 M trifluoroacetic acid at 100 °C for 40 min, followed by 30 min incubation for 5 min. Neutral and amino sugars were determined after hydrolysis in 2 M trifluoroacetic acid at 100 °C for 4 h and 4 mM HCl at 100 °C for 4 h, respectively. The hydrolysates were dried under vacuum, dissolved in 50 μl of water and analyzed by high pressure anion exchange chromatography ( Dionex, CarboPac PA1 column, 4 × 250 mm) combined with pulsed amperometric detection (Dionex DX 500). Neutral and amino sugars were separated isocratically at 1 ml/min in 12 mM NaOH, and N-acetyl neuraminic acid was eluted using a gradient of 0–200 mM sodium acetate in 250 mM NaOH over 15 min. No N-glycolyl neuraminic acid was detected even at a concentration of sodium acetate greater than 200 mM. Quanti-tation was performed using internal standards (2-deoxyxylulose for the neutral sugars and acetylated acetyl neuraminic acid for the sialic acids) with reference to the appropriate response factors.

Laser Light Scattering—Reduced mucin subunits obtained by density gradient centrifugation of the insoluble glycoprotein complex of reservoir urine after density gradient centrifugation (see Fig. 3D, peak I), were dialyzed against and then diluted with 6 M guanidinium chloride, 1 mM EDTA, 10 mM sodium phosphate buffer, pH 6.5. The intensity of light scattered at angles between 20° and 90° with a Malvern 4900 SM system equipped with a 25 mW HeNe laser. The data were plotted in according to Zimm (33), and a refractive index increment of 0.104 ml/g was used (34). The subunits were studied at 37.5, 75, 150, and 300 μg/ml.

Analytical Ultracentrifugation—Mucin subunits obtained by density gradient centrifugation of the reduced insoluble glycoprotein complex from reservoir urine (see Fig. 3D, peak I), subunit fractions (I, II, and III) obtained after gel chromatography (see Fig. 7A), and glycopeptides A and B (see Fig. 5) were dialyzed against 6 M ultrapure guani-dinium chloride. Sedimentation equilibrium experiments were performed in a Beckman Optima XL-A analytical ultracentrifuge using six-channel charcoal filled Epon centerpieces. Sedimentation distributions were, when possible, studied at several concentrations using absorption optics at 230 nm. The weight-average molecular weight of the entire distribution was determined by extrapolating the point-average value of M2 to the bottom of the cell (35), and the obtained values were then, when possible, extrapolated to zero concentration. Partial specific volumes of 0.66 and 0.65 ml/mg were used for subunits and high molecular mass glycopeptides, respectively (36).

Southern Blot Analysis—DNA-containing fractions obtained after density gradient centrifugation of the guanidinium chloride soluble material from the mucosal scrapings were pooled, desalted, and subjected to digestion with the restriction enzyme HindIII. The digest was fractionated on a 0.8% agarose gel, transferred to a nylon membrane and hybridized with 32P-labeled SMUC41 probe as described previously (10).

RESULTS

MUC2 Antibodies—The LUM2-2, LUM2-3, and LUM2-4 antisera were raised against three different sequences present in the C- and N-terminal regions of MUC2 (Fig. 1A). Each antisemur reacted only with the peptide used for immunization (Fig. 1B), and no reactivity with the peptides used to raise the other antibodies was observed (results not shown). All antisera reacted with mucin subunits obtained after density gradient centrifugation (Fig. 1C).

Immunohistochemistry—In human colon, staining with the LUM2-3 antisemur was seen over the goblet cells in the surface mucosa and along the entire length of the crypt (Fig. 2A). Surrounding tissues were not stained; however, reactivity was seen over secreted material at the luminal surface of the mu cosa. The LUM2-2 and LUM2-4 antibodies showed similar results (data not shown). Also in the small intestine, the LUM2-3 antibody stained the goblet cells over the surface mucosa and along the crypts, whereas the enterocytes showed no reactivity (Fig. 2B). Again, the other two antibodies provided similar results (data not shown).

Isolation of MUC2 Mucin Subunits—Repeated extractions of human colonic mucins with 6 M guanidinium chloride fractionated them into those that were soluble in this solvent and those that resisted extraction. Density gradient centrifugation of the guanidinium chloride soluble fraction (results not shown) showed that little material was present at a density expected for mucins, although some sialic acid-containing and PAS-reactive material banding at a density 1.29–1.38 g/ml was detected. However, this material did not react, or reacted only weakly, with the three MUC2 antisera. The sol fractions from the reservoir urine and reservoir washings also contained little MUC2.

Density gradient centrifugation of the reduced insoluble gly-
The MUC2 apoprotein and reactivity of the MUC2 antisera with synthetic peptides and mucin subunits. A, schematic picture of the MUC2 apoprotein. The open boxes represent the domains with a high degree of homology to the α-domains in the prepro-von Willebrand factor, and the filled boxes illustrate the Ser/Thr/Pro-rich 347-amino acid irregular repeat domain and the VNTR region with about 100 repeats of 23 amino acid sequence, rich in Thr and Pro. The locations of the sequences used to raise antibodies are indicated by arrowheads together with the name of the corresponding antibody. B, the synthetic peptides (0.5 μg/ml) used to raise the LUM2-2, LUM2-3, and LUM2-4 antisera were coated onto multilayer assay plates and incubated with the MUC2 antisera at different dilutions. C, mucin subunits (1.0 μg/ml) obtained after density gradient centrifugation were coated onto multilayer assay plates and incubated with LUM2-2 (□), LUM2-3 (■), or LUM2-4 (▲) at different dilutions. Bound antibody was detected as described under “Experimental Procedures” and expressed as absorbance at 405 nm.

coprotein complex revealed a single unimodal peak between 1.30 and 1.45 g/ml that reacted strongly with the sialic acid and PAS analyses (Fig. 3, A–C). The material was well separated from DNA and low buoyant density proteins that showed strong absorbance at 280 nm at 1.47 g/ml and at the top of the gradient respectively. Pretreatment with alkali before sialic acid analysis increased the color yield as much as 15-fold in some, but not all, preparations, suggesting the presence of O-acetylated sialic acids. Reactivity against the LUM2-2 and LUM2-4 antisera coincided with the mucin distribution as defined by sialic acid and PAS reactivity (Fig. 3, D–F). The major part (>95%) of the MUC2 mucin was present within this insoluble glycoprotein complex.

The reactivity with the LUM2-4 antiserum varied between preparations. In most, but not all, the reactivity appeared as two partially resolved peaks, one appearing at the same density as the mucin subunits, the second one on the lower buoyant density side of the main peak as would be expected for a less glycosylated protein (Fig. 3F). In some preparations, the reactivity against the LUM2-4 antibody appeared as a peak with a density only slightly lower than that of the major mucin band (Fig. 3, D and E), and in others, the reactivity followed that of the main mucin distribution.

MUC2 C-terminal Peptide—Low density material reacting with the LUM2-4 antibody (Fig. 3D, peak II) was pooled and subjected to gel chromatography on Sephacryl S-500HR. A peak reacting with the LUM2-4 antiserum was well included on the column (Fig. 4), and virtually no reactivity with the LUM2-2 or LUM2-3 antibodies was detected (results not shown). Similar results were obtained with corresponding material from several other preparations. The peak reacting with the LUM2-4 antiserum was pooled and subjected to SDS-polyacrylamide electrophoresis followed by Western blotting. A major band of approximately 120 kDa that reacted strongly with both Coomassie Blue and the LUM2-4 antibody was detected (Fig. 4, inset). This band was cut out and subjected to N-terminal amino acid sequence analysis. The sequence PHYVTFD was obtained (with the reservation that Y could possibly be V) corresponding to amino acids 292–298 present within the C-terminal part of MUC2 as defined by Gum et al. (37).

High Molecular Mass Tryptic Glycopeptides—When mucin subunits (Fig. 3D, peak I) were subjected to trypsin digestion followed by gel chromatography on Sephacryl S-500HR, two major peaks (referred to as glycopeptides A and B, respectively) were detected with the sialic acid assay (Fig. 5). Amino acid analysis was performed on the two glycopeptides from two different individuals and compared with the deduced compositions of the VNTR region and the irregular repeat domain (Table I). Glycopeptide A contains high amounts of threonine (approximately 45 mol %) and proline (approximately 20 mol %) but low amounts of serine (approximately 4 mol %), as expected for the VNTR region. The amino acid composition of glycopeptide B was close to that expected for the irregular repeat domain with high amounts of threonine (approximately 38 mol %), serine (approximately 10 mol %), and proline (approximately 28 mol %).

The results of carbohydrate analysis of glycopeptides A and B are shown in Table II. The monosaccharide composition of the two glycopeptides is similar, in keeping with a similar oligosaccharide substitution. Little mannose was found, suggesting that no N-linked oligosaccharides occur within these domains of the MUC2 mucin. N-acetyl neuraminic acid but no N-glycolyl neuraminic acid was present.

Analytical ultracentrifugation showed that the molecular weights of glycopeptides A and B are 930,000 and 180,000, respectively (Fig. 6). The ratio (5.2:1) of the molecular weights for the large and small glycopeptides is similar to that (5.9:1) expected from the differences in lengths of the VNTR region predicted from the most common allele (100 repeats) and the invariant irregular repeat domain, assuming a similar glycosylation of the two protein regions. The ratio of the yields of glycopeptides A and B (5.8:1) suggests that the two domains occur in equimolar proportions in the molecule.

MUC2 Subunit Populations—When MUC2 mucin subunits from the insoluble glycoprotein complex (Fig. 3D, peak I) were subjected to gel chromatography on Sephacryl S-500HR, a number of partially separated populations reacting with the LUM2-2, LUM2-3 and LUM2-4 antibodies were identified (Fig. 7A). The largest population eluted close to the void volume of the column and the smallest one appeared just ahead of the position for glycopeptide A. When individual chromatographic fractions of the partially resolved MUC2 populations were subjected to agarose gel electrophoresis, followed by Western blotting using the LUM2-3 antiserum, a number of distinct bands were obtained (Fig. 7A, inset). Molecules eluting first from the column had a lower mobility on the agarose gel than those that were more included and all bands reacted with all three MUC2 antisera (results not shown). Subunits from several different preparations were subjected to gel chromatography and agarose gel electrophoresis with similar or identical results.

Corresponding fractions from 10 gel chromatography runs were pooled as three populations referred to as sub-I, sub-II, and sub-III, respectively, according to the bars shown in Fig. 7A, concentrated, and recovered using density gradient centrifugation. The latter step was used to remove material that may
have detached from the column matrix and would interfere with further analyses (see below). Subsequent gel chromatography and agarose gel electrophoresis showed that the sub-III population contained mainly the “fastest” band, whereas the sub-I and sub-II populations were mixtures of the two “slowest” and the two fastest, respectively (Fig. 7B).

Size of the MUC2 Subunit Populations—MUC2 mucin subunits from the insoluble glycoprotein complex (Fig. 3D, peak I) were subjected to molecular weight determination using analytical ultracentrifugation and laser light scattering. A Zimm plot of the data obtained by laser light scattering (Fig. 8) reveals a molecular weight of 4.0 × 10⁶ for the entire subunit population. The results obtained using analytical ultracentrifugation (Fig. 9) for the same material provided a molecular weight of 3.8 × 10⁶, in good agreement with that given by light scattering.

The molecular weights for the sub-I, sub-II, and sub-III populations were determined using analytical ultracentrifugation, but due to the small amounts of material available, the analysis could only be carried out at a single concentration, and the concentration dependence could not be assessed. However, the value for the sub-II population (2.5 × 10⁶ at 0.20 mg/ml) falls on the line describing the concentration dependence for the entire distribution, and assuming the same concentration dependence, a molecular weight of 3.8 × 10⁶ is obtained. Using the same concentration dependence, the value for the sub-I population (3.4 × 10⁶ at 0.21 mg/ml) was extrapolated to 4.7 × 10⁶. Because the concentration dependence is usually more pronounced for larger molecules than for smaller ones, this value represents an underestimation rather than an overestimation. When the value for the sub-III population (1.4 × 10⁶ at 0.17 mg/ml) was corrected using the concentration dependence for the entire distribution and for glycopeptide A (Fig. 9), values of 2.4 × 10⁶ and 1.5 × 10⁶, respectively, were obtained. The two values are likely to represent over- and underestimations,

FIG. 2. Immunohistochemical staining of human colon and small intestine. Sections (4 μm) of paraffin-embedded colon (A) and small intestine (B) were stained with the LUM2-3 antiserum and counterstained as described in the text. The bars represent 100 μm.

FIG. 3. Isopycnic density gradient centrifugation of human colonic mucin subunits. Mucin subunits were obtained after reduction of the insoluble glycoprotein complex from reservoir urine (A and D), mucosal scrapings of colonic tissue (B and E), and reservoir washings (C and F) and subjected to density gradient centrifugation in 4 M guanidinium chloride/CsCl. Fractions were analyzed for density (——), sialic acid before (○) and after (●) pretreatment with NaOH, PAS (△), absorbance at 280 nm (——) (A–C), and reactivity with the LUM2-2(M), LUM2-3(f), and LUM2-4(Œ) antisera (D–F). The bars indicate material that was pooled and subjected to further analysis.

FIG. 4. Gel chromatography on Sephacryl S-500HR of the putative C-terminal cleavage peptide obtained after reduction of the insoluble glycoprotein complex. Low density, LUM2-4-reactive material obtained after density gradient centrifugation (Fig. 3D, peak II) was subjected to gel chromatography on a Sephacryl S-500 HR column (1.6 × 50 cm) eluted with 4 M guanidinium chloride, 10 mM sodium phosphate buffer, pH 7.0. Fractions (1 ml) were analyzed for absorbance at 280 nm (——) and reactivity against the LUM2-4 antiserum (●). The LUM2-4-reactive material was pooled (as shown) and subjected to SDS-polyacrylamide electrophoresis on a 4–16% gradient gel, blotted onto a polyvinylidene difluoride membrane followed by Coomassie Blue staining (C) or Western blotting (W) using the LUM2-4 antiserum (inset). The numbers indicate estimated molecular weights as determined using high and low molecular weight markers (electrophoresis calibration kits (Amersham Pharmacia Biotech)).
and the “true” value for the molecular weight of the sub-III population is expected to fall between the two.

Southern Blot Analysis and Comparison of the Purified MUC2 Subunits—In order to assess the length of the VNTR region in the various samples, DNA present in the density gradients of material soluble in guanidinium chloride was subjected to digestion with the HinII restriction enzyme followed by electrophoresis and Southern blot analysis using the SMUC41 probe. Eleven different individuals were tested. Eight of the samples contained a single band of approximately 7–8 kilobases or two closely spaced bands corresponding to the most common long allele(s) with a VNTR region containing of the order of 100 repeats of the 23 amino acid peptide unit. However, three individuals expressed both a long and a much shorter (≈4-kilobase) allele (Fig. 10). Agarose gel electrophoresis and Western blotting of the mucin subunits from individuals homozygous for two long alleles displayed a pattern of major bands similar to those shown above, whereas all individuals heterozygous for a long and a short allele showed a much more extensive ladder pattern (Fig. 10). The putative MUC2 monomers derived from the short and long alleles are referred to as α and β respectively, and αα and ββ are “homodimers” containing two short and two long monomers, respectively. The band appearing between the αα and ββ ones is interpreted as a heterodimer (αβ) consisting of one short and one long monomer.

DISCUSSION

The antibodies raised against the three different peptide sequences within the MUC2 apoprotein were shown to recognize both the cognate peptide and the purified fully glycosylated mucin subunits (Fig. 1, B and C). The peptide sequence recognized by the LUM2-2 antibody is located N-terminal to the irregular repeat domain, whereas LUM2-3 and LUM2-4 are directed to sequences C-terminal to the VNTR region (Fig. 1A), and together the three antibodies flank the entire central glycosylated domain of MUC2. All three antibodies stained the goblet cells in human colon and small intestine (Fig. 2) but not normal human gastric mucosa or the submucosal glands of normal human airway, suggesting that there is no cross-reactivity of the antisera with MUC5AC, MUC5B, and MUC6.

Mucosal scrapings and the gel phase of colonic secretions were subjected to extensive extractions using 6 M guanidinium chloride, and the putative insoluble glycoprotein complex was finally brought into solution using reduction. Density gradient centrifugation of the soluble material and the insoluble fraction after reduction revealed that most of the mucins were present in the insoluble glycoprotein complex, and virtually all of the MUC2 reactivity, as determined by using the three antibodies, was confined to this fraction. It is concluded that the major part of human intestinal MUC2 occurs as an insoluble glycoprotein complex. Fresh specimens of colonic tissue would be expected to provide the best source of colonic mucins, but sufficiently large samples of tissue are difficult to obtain. For this reason, secretions that can be obtained in relatively large amounts from artificial urinary bladders constructed from colonic segments, were also used. Previous studies have shown that such secretions contain a major population of mucins that is insoluble in guanidinium chloride but can be isolated as a distinct component following reduction (38). In all respects studied here, the MUC2 mucin from artificial bladders was similar to that obtained from mucosal scrapings showing that artificial bladders are a good source of colonic MUC2.

The large secreted mucus-forming mucins from stomach, respiratory tract, and cervix are long linear structures, formed by subunits linked end-to-end by disulfide bonds. The major part of the macromolecules is readily extracted using chaotropic salts such as guanidinium chloride. In contrast, mucins from rat and human intestine resist extraction with 6 M guanidinium but can be isolated as an insoluble glycoprotein complex following extensive treatment of the tissue with this solvent (21, 23). The fact that the mucins cannot be solubilized with guanidinium chloride suggests a larger involvement of covalent bonds in the formation of intestinal mucus than in, for example, gastric mucus and the gel phase of airway secretions. The physiological relevance of this is not known but may be related to the need to protect the intestinal surface with a highly resistant mucus gel that is at the same time porous enough to allow transport from the lumen over the epithelium.

In many, but not all, preparations, the reactivity with the LUM2-4 antisera in the density gradient of the subunits from the insoluble MUC2 complex was present at a lower buoyant density than the main mucin population as defined by carbohydrate analysis and the reactivity with the two other antibodies. The low density LUM2-4 reactive material was isolated using gel chromatography and SDS-polyacrylamide gel electrophoresis and shown to be a distinct component with an apparent molecular weight of approximately 120,000, similar to that (118,000) of the so-called “link” glycoprotein that was first identified after reduction of intestinal mucins (18, 19) and later shown to comprise the C-terminal part of the MUC2/Muc2 apoprotein (17). Edman degradation provided the sequence PHYVTFD for the 120-kDa component, corresponding to the N-terminal amino acids of the putative link protein. The 120-kDa component identified here thus corresponds to this structure and is, when cleaved, apparently at least in part left bound to the insoluble MUC2 complex via disulfide bonds. It is not known at present whether proteolytic cleavage in the C-terminal part of MUC2 has any biological significance or merely reflects proteolysis that occurs during the isolation procedure.

In order to assess the size of the MUC2 monomers and to investigate whether or not the insoluble glycoprotein complex contains other major components than MUC2, the subunit population obtained using density gradient centrifugation was subjected to trypsin digestion followed by gel chromatography. Two major glycopeptide populations (A and B) with similar carbohydrate compositions were identified, and the low amount of mannose present suggests that most of the oligosaccharides are O-linked structures. The amino acid compositions and the relative sizes of glycopeptides A and B are in good agreement with those expected from the VNTR region and the irregular repeat domain, respectively, strongly suggesting that they correspond to these two domains of MUC2 and that no other glycoproteins contribute significantly to the subunits obtained.
from the insoluble complex.

The molecular weights of glycopeptides A (930,000) and B (180,000) are in keeping with those estimated from structures containing 2300 amino acids (100 repeats of 23 amino acids as for the most common allele) and 347 amino acids, respectively, assuming approximately 70–80% carbohydrate, realistic values for epithelial mucins. It should be pointed out that the entire VNTR region has not been sequenced and that the amino acid composition predicted by the SMUC41 and SMUC42 clones (3) shows that lysine residues, and thus trypsin-sensitive sites, may appear in this part of the mucin. If so, the size of the VNTR region determined as the molecular weight of

| TABLE I | Amino acid composition of glycopeptides A and B isolated from two individuals together with the amino acid composition of the VNTR region and the irregular repeat domain |
|----------------------|----------------------|----------------------|----------------------|
| Amino acid            | Glycopeptide A | VNTRa | Glycopeptide B | IRb |
|----------------------|----------------|-------|----------------|-----|
|                      | mol %         |       | mol %          |     |
| Asp/Asn              | 1.9           | 1.8   | 0.5            | 2.2 |
| Glu/Gln              | 5.9           | 5.3   | 3.7            | 1.9 |
| Ser                  | 4.5           | 4.3   | 3.5            | 10.6|
| His                  | 0.3           | 0.5   | 0              | 0.4 |
| Gly                  | 5.3           | 5.3   | 4.0            | 2.5 |
| Thr                  | 46.6          | 43.9  | 54.2           | 38.2|
| Ala                  | 1.5           | 1.4   | 0.4            | 1.1 |
| Pro                  | 20.5          | 20.1  | 22.0           | 28.5|
| Tyr                  | 1.0           | 1.6   | 0              | 1.6 |
| Arg                  | 0.4           | 0.1   | 0              | 0.6 |
| Val                  | 5.0           | 4.4   | 4.7            | 1.5 |
| Met                  | 0.8           | 0.8   | 1.3            | 0.7 |
| Ile                  | 4.3           | 3.7   | 5.2            | 3.4 |
| Leu                  | 1.2           | 0.9   | 0.1            | 3.2 |
| Phe                  | 0.2           | 2.1   | 0              | 1.4 |
| Lys                  | 0.8           | 4.0   | 0.3            | 3.3 |
|                      |               |       | 1.0            |     |

a In the calculation of the VNTR region, amino acid sequences of the SMUC40, SMUC41, and SMUC42 clones published by Gum et al. (3) were used. One hundred repeats of the sequence were assumed, and amino acids up to the first trypsin site on either side of the VNTR region were included (3, 37).

b In the calculation of the irregular repeat domain (Ir), the amino acid sequence published by Gum et al. (37) was used, and amino acids up to the first trypsin site on either side of the irregular domain were included.

| TABLE II | Monosaccharide composition of glycopeptides A and B isolated from two individuals |
|----------------------|----------------------|----------------------|
| Monosaccharide        | Glycopeptide A | Glycopeptide B |
|----------------------|----------------|----------------|
| GalNAc               | 1.0            | 1.0            |
| GlcNAc               | 0.5            | 0.4            |
| Fuc                  | 0.3            | 0.2            |
| Gal                  | 0.4            | 0.3            |
| Man                  | 0.1            | 0              |
| NANA                 | 0.8            | 0.9            |

a Relative abundance of individual sugars are given with GalNAc as 1.0.

FIG. 6. Analytical ultracentrifugation of glycopeptides A and B. The molecular mass of glycopeptides A (A) and B (C) obtained after trypsin digestion and gel chromatography were determined using sedimentation equilibrium in the analytical ultracentrifuge at 20 °C for at least 70 h. Samples were monitored several times during the experiment to ensure that equilibrium was achieved. Glycopeptide A was studied at 0.4, 0.5, 0.6, 0.8, and 1.0 mg/ml at 3000 rpm, and glycopeptide B was studied at 0.5, 0.6, 0.8, and 1.0 mg/ml at 8000 rpm. The obtained values were extrapolated to zero concentration.

from the insoluble complex.

The molecular weights of glycopeptides A (930,000) and B (180,000) are in keeping with those estimated from structures containing 2300 amino acids (100 repeats of 23 amino acids as for the most common allele) and 347 amino acids, respectively, assuming approximately 70–80% carbohydrate, realistic values for epithelial mucins. It should be pointed out that the entire VNTR region has not been sequenced and that the amino acid composition predicted by the SMUC41 and SMUC42 clones (3) shows that lysine residues, and thus trypsin-sensitive sites, may appear in this part of the mucin. If so, the size of the VNTR region determined as the molecular weight of

FIG. 7. Gel chromatography on Sephacryl S-500 of reduced mucin subunits obtained from the insoluble glycoprotein complex. A, mucin subunits obtained after density gradient centrifugation (Fig. 3D, peak I) were subjected to gel chromatography on a Sephacryl S-500 HR column (1.6 × 50 cm) eluted with 4 M guanidinium chloride, 10 mM sodium phosphate buffer, pH 7.0, at 0.1 ml/min. Fractions (1 ml) were analyzed for sialic acid after pre-treatment with NaOH (●), absorbance at 280 nm (– – –), reactivity with the LUM2-2 (●), LUM2-3 (●), and LUM2-4 (●) antisera. Individual fractions containing mucin subunits were subjected to agarose gel electrophoresis followed by Western blotting using the LUM2-3 antiserum (inset). B, fractions were pooled into three populations (sub-I, sub-II, and sub-III) as indicated in A and re-run on the column as above; the fractions were analyzed for carbohydrate (sub-I, ——; sub-II, – – –; sub-III, · · · ) using the glycan detection method. Agarose gel electrophoresis followed by Western blotting using the LUM2-3 antiserum of the three individual subunit populations is shown in the inset.
urine, solubilized by reduction (Fig. 3). Mucin subunits were isolated after density gradient centrifugation (Fig. 3D, peak I) (≈10^6). Readings were taken at θ = 0 was performed from 20° to 45°.

Structure of Human Colonic MUC2

**FIG. 9.** Analytical ultracentrifugation of the mucin subunits and the three subpopulations thereof. Mucin subunits isolated after density gradient centrifugation (Fig. 3D, peak I) (●) and the three subfractions (sub-I, sub-II, sub-III) (○) obtained after gel chromatography (Fig. 7A) were studied with sedimentation equilibrium. Samples were run for approximately 70 h at 1500 rpm at 20 °C and monitored several times during the experiment to ensure that equilibrium was achieved. The unfraccionated subunits were studied at 0.15, 0.20, 0.22, 0.35, and 0.47 mg/ml, whereas sub-I was studied at 0.21, sub-II at 0.20, and sub-III at 0.17 mg/ml, respectively. The obtained values were extrapolated to zero concentration. For sub-I, sub-II, and sub-III, the same concentration dependence as for the unfraccionated material was used (— — —), for the sub-III population, also, the concentration dependence obtained for glycopeptide A (Fig. 6) was used (····).

FIG. 8. Zimm plot analysis of light scattering data for the mucin subunits. Mucin subunits were isolated after density gradient centrifugation of the insoluble glycoprotein complex from reservoir urine, solubilized by reduction (Fig. 3D, peak I). Readings were taken at θ = 20°, 25°, 30°, 35°, 40°, 45°, 60°, 75°, and 90°, and the subunits were studied at 37.5, 75, 150, and 300 μg/ml at 20 °C. Extrapolation to θ = 0 was performed from 20° to 45°.

Gel chromatography and gel electrophoresis were used to show that the MUC2 subunits comprise a mixture of well-defined populations. Gel chromatography would suggest that glycopeptide A may be an underestimation. However, the amino acid sequence for the irregular repeat domain is defined, and the size of the smallest possible tryptic fragment can thus be predicted. Because the ratios of the molecular weights (5.2:1) and of the predicted lengths of the protein cores (5.9:1) are similar, it appears likely—assuming a similar glycosylation of the two glycopeptides—that glycopeptide A represents the major part of the VNTR region, if not the entire region. A more detailed chemical characterization of these glycosylated regions is currently in progress. Finally, the ratio of the yields of the glycopeptides A and B (5.8:1) compared with that of the molecular weights (5.2:1) suggests that the two fragments occur in equimolar proportions as would be expected for the VNTR region and the irregular repeat domain. From the values of the molecular weights of glycopeptides A and B, the size of the C-terminal cleavage fraction obtained by SDS-polyacrylamide gel electrophoresis and the size of the N-terminal domain (approximately 1400 amino acids; molecular weight, 150,000), the molecular weight of the MUC2 monomer from human colon is predicted to be on the order of 1.5 × 10^6.

To further investigate the size of the MUC2 subunits, populations enriched in the various bands were isolated using gel chromatography and subjected to molecular weight determinations. The molecular weight for the most included, smallest (sub-III) population, which represents almost entirely the fastest moving band was estimated to be greater than 1.5 × 10^6 but less than 2.4 × 10^6; significantly smaller than that (approximately 4 × 10^6) for the entire population, but close to the value predicted for a MUC2 monomer (see above). The molecular weight for the sub-II (3.8 × 10^6) and sub-I (4.7 × 10^6) populations would accommodate two and three units of sub-III, respectively, considering the fact that they both represent mixtures of larger and smaller species. The interpretations are complicated by the fact that the concentration dependence of the subpopulations of the subunits can only be estimated. However, the size differences between the entire subunit population and that obtained for the smallest population is so large that it is safe to conclude that the MUC2 population obtained after reduction is not a true monomer but must comprise a series of oligomers joined by linkages that are not sensitive to reduction. This conclusion is strongly supported by the electrophoretic patterns obtained for heterozygous individuals with two MUC2 alleles of very different size, which displayed a much more extensive ladder pattern. The data obtained from such individuals indicate that both alleles are expressed and that both may

**FIG. 10.** Western blot analysis of MUC2 reduced subunits and Southern blot analysis of the VNTR region. A, mucin subunits obtained after density gradient centrifugation (mucosal scrapings) were subjected to agarose gel electrophoresis followed by Western blotting using the LUM2-3 antiserum. B, DNA isolated from the extraction-soluble fraction was subjected to digestion with HindIII and analyzed by agarose gel electrophoresis followed by probing with SMUC41. Data shown were obtained from one individual expressing both a long and a short allele (lane 1) and one homozygous for long alleles (lane 2). The putative structure of the isoforms is indicated alongside the gel, where α and β represent the products of the short and the long, allele respectively. The numbers indicate estimated size in kilobases as determined by comparison with Raoul molecular weight markers (Appligene, Durham, United Kingdom), the 1-kilobase ladder (Life Technologies, Ltd., Paisley, United Kingdom), and lambda HindIII digest (Life Technologies, Inc.).
form dimers (αα, αβ, and ββ) and possibly also higher oligomers in all possible permutations, as indicated in Fig. 10.

While this work was in progress, a paper was published that suggested the presence of reduction-insensitive MUC2 dimers in LS174T cells (39). After reduction of insoluble mucins from this source, two highly polydisperse bands were observed with agarose gel electrophoresis. The components corresponding to the two bands were partially separated using rate-zonal centrifugation in a sucrose gradient, and the differences in sedimentation rates were interpreted as being consistent with the behavior of a monomer and a dimer. The authors suggest that the smeared appearance of the individual bands following agarose gel electrophoresis reveals glycosylation differences; however, the range of migration rates covered by the smears is in many cases of the same order of magnitude as the differences in migration rate between the putative monomer and the surprisingly sharp dimer bands. Furthermore, the broad and overlapping distributions of the putative monomer and dimer revealed using rate-zonal centrifugation suggest that both populations encompass a range of s values that is actually larger than the difference in s value obtained from the “peak positions,” i.e., each population may contain molecules that differ in size by a factor of more than 2. It is thus not clear from these experiments, without the support of independent data, which differences in migration rates may be interpreted as glycosylation polydispersity and which as “oligomerization.” Nevertheless, the interpretations made by Axelsson et al. (39) are consistent with our findings.

In summary, we have shown that the MUC2 mucin from human colon is present as an insoluble glycoprotein complex and that this glycoprotein complex is mainly composed of MUC2. The complex is assembled from MUC2 “subunits” using disulfide bonds. The subunits obtained after reduction occur as a monomer and a series of oligomers apparently joined by a disulfide bond. The subunits obtained after reduction occur as a monomer and a series of oligomers apparently joined by a disulfide bond. The subunits obtained after reduction occur as a monomer and a series of oligomers apparently joined by a disulfide bond. The subunits obtained after reduction occur as a monomer and a series of oligomers apparently joined by a disulfide bond. The subunits obtained after reduction occur as a monomer and a series of oligomers apparently joined by a disulfide bond.