Glycoconjugate Journal (1996) **13:** 823–831

Molecular characterization of the large heavily glycosylated domain glycopeptide from the rat small intestinal Muc2 mucin

NICLAS G. KARLSSON[1], MALIN EV JOHANSSON[1], NOOMI ASKER[1], HASSE KARLSSON[1], SANDRA J. GENDLER[2], INGEMAR CARLSTEDT[3] and GUNNAR C. HANSSON[1]*

[1]*Department of Medical Biochemistry, Göteborg University, Medicinaregatan 9, S-413 90 Gothenburg, Sweden*
[2]*Mayo Clinic Scottsdale, 13400 East Shea Boulevard, Scottsdale, Arizona 85259, USA*
[3]*Department of Medical and Physiological Chemistry, PO Box 94, S-221 00 Lund, Sweden*

Received 10 October 1995, revised 31 January 1996

The largest high-glycosylated domain, glycopeptide A, of the 'insoluble' mucin complex of the rat small intestine has earlier been purified and characterized (Carlstedt *et al.*, 1993, *J Biol Chem* **268**: 18771–81). A rabbit antiserum raised against deglycosylated glycopeptide A was used to clone part of a mucin showing homology to the human MUC2 mucin (Hansson *et al.*, 1994, *Biochem Biophys Res Commun* **198**: 181–90). This serum specifically stained goblet cells (paranuclear) in the mouse small intestine. The size of the coding sequence of glycopeptide A was estimated by using reversed transcriptase PCR of mRNA from an inbred rat strain (GOT-W) using primers in the unique central and C-terminal parts of the proposed rat Muc2 sequences. The PCR and Southern blot of the PCR products showed a fragment of about 5.5 kb corresponding to about 1700 amino acids when the known Cys-rich sequences used for the primers were subtracted. This is slightly larger than the size estimated earlier by biochemical studies. The mRNA encoding the rat Muc2 was slightly smaller than the mRNA encoding the human MUC2 in a colorectal cell line. Although the size of glycopeptide A estimated from biochemical results and by PCR is not identical, the results obtained here further support that the 'insoluble' mucin of the rat small intestine is encoded by the Muc2 gene. Most of the oligosaccharides in glycopeptide A were either neutral (40%) or sialylated (40%). The remaining ones were sulfated with the sulfate group attached to C-6 of *N*-acetylglucosamine linked to C-6 of the *N*-acetylgalactosaminitol as revealed by tandem mass spectrometry of the perdeuteroacetylated oligosaccharides. Eighteen oligosaccharides were found of which fourteen were characterized and found to be mostly novel. Our findings thus expand the current knowledge of the core peptide of the rat intestinal goblet cell mucin and provide a relatively complete picture of the glycosylation of a defined mucin domain.

Keywords: mucin, rat small intestine, sulfated oligosaccharides, immunolocalization, goblet cell, Muc2

Abbreviations: gpA, glycopeptide A; α-gpA, antiserum against glycopeptide A; CID, collision induced dissociation; FAB, fast atom bombardment; Hex, hexose; HexNAc, *N*-acetylhexosamine; HexNAcol, *N*-acetylhexosaminitol

Introduction

Mucins are a group of highly glycosylated glycoproteins being part of the mucus gel protecting the mucosal cells. One of the largest surfaces in the body is the small intestine that is threatened by both endogenous digestive enzymes and exogenous agents. The major part of the intestinal mucins are produced by the intestinal goblet cells interspersed between the absorptive enterocytes. At least 80% of the intestinal mucins have been shown to be 'insoluble' in denaturing solvents such as guanidinium chloride when shear forces are avoided during preparation [1, 2]. When the high glycosylated domains from this 'insoluble' fraction were prepared, two glycopeptides with an estimated mass of 650 kDa and 335 kDa were obtained. An antiserum raised against the hydrogen fluoride

*To whom correspondence should be addressed. Tel: +46 31 7733488; Fax: +46 31 416108.

deglycosylated large glycopeptide (glycopeptide A, gpA) was used to isolate a cDNA clone (VR-1A) encoding both a cystein and a serine/threonine rich peptide [3]. Sequence comparisons of the cystein-rich sequence revealed high homology with the human MUC2 mucin [4–6] suggesting that the 'insoluble' mucin in rat was indeed encoded by the Muc2 gene [3]. Two other groups have also cloned what is proposed to be parts of the Muc2 mucin [7–9] as has been summarized [10]. The C-terminal part was cloned as part of a 'link'-peptide [7, 8] and the N-terminal part was cloned from rat lung [9].

The structural understanding of mucin oligosaccharides has increased substantially during recent years, but analyses of the glycosylation of defined mucin species with identical apoprotein has been limited. The neutral and sialic acid-containing oligosaccharides from the small and large glycopeptides (glycopeptide A and B) of rat small intestine were recently analysed in detail [2]. These oligosaccharides constituted 80% of all O-linked oligosaccharides spanning from one to seven sugar residues per oligosaccharide chain. In the present investigation, additional information is provided supporting the previous conclusion that the large high-glycosylated domain from the 'insoluble' mucins of rat small intestine is part of the rat Muc2 mucin. An antiserum to this structure recognize the goblet cells of the small intestine. Analysis of the glycosylation of glycopeptide A is also extended by studies of the sulfated oligosaccharides.

Materials and methods

Mucin glycopeptide and antiserum

The large trypsin resistant glycopeptide (glycopeptide A, gpA) from the 'insoluble' mucin complex of rat (GOT-W strain) small intestine was prepared as described [2]. In short, the mucosal scraping of the small intestine was extracted in 6 M guanidinium chloride and solubilized by reduction in 10 mM dithiotreitol. DNA and RNA were digested with DNase and RNase, and the highly glycosylated domains obtained by trypsin cleavage and purified on Sephacryl S-200 and S-500. Only the larger glycopeptide (gpA) was analysed here. GpA was deglycosylated using hydrogen fluoride [3] and an antiserum (α-gpA, PH897) raised in a rabbit [3].

Northern blot analysis

Total RNA and mRNA were prepared from mucosal scraping of rat (GOT-W) small intestine as well as LS 174T cells [3]. The electrophoresis, blotting and labelling of the probes were performed as described before [11].

PCR over the long Ser/Thr rich domain

The mRNA was prepared from 25 μg of total RNA with Dynabeads (dT)25 (Dynal, Oslo, Norway). Reverse transcription was made with Superscript II (Gibco BRL, Scotland) at 42 °C for 120 min using the provided buffer. The concentration of the reverse primer (5′-GAGTAGATGGACAGGACCAATTGAAGACC-3′, corresponding to nucleotides 67–30 in [7]) (Scandinavian Gene Synthesis AB, Köping, Sweden) was 1 μM. The reverse transcriptase was inactivated at 70 °C for 15 min, prior to a denaturation step at 95 °C, 5 min. The RNA was removed by incubation at 37 °C for 1 h with 5 μg of DNase free RNase A (Sigma, MO). One tenth of the obtained cDNA was used as template in PCR, for which the Expand Long Template PCR System (Boehringer Mannheim, Germany) was used. The final concentration of the reverse primer (the same as in the reverse transcription) and the forward primer (5′-GCAAAAATGCAGAACAGGGAATAGGAG-3′, corresponding to nucleotides 71–109 in [3]) was 300 nM of each. The supplied buffer number 3 (50 mM Tris-HCl, pH 9.2, 16 mM $(NH_4)_2SO_4$, 2.25 mM $MgCl_2$, 2% DMSO, 0.1% Tween 20) was used, and the dNTP concentration was 0.5 mM of each nucleotide. The amplification started with 93 °C for 1 min and was then performed for 30 cycles with 93 °C for 1 min, 66 °C for 1 min and 68 °C for 5 min. In the last 20 cycles the elongation time was extended with 20 s for each additional cycle. The PCR had a final extension at 68 °C for 7 min. A DNA Thermal Cycler (Perkin Elmer, CT) was used.

Agarose gel electrophoresis was performed with 1% agarose (Gibco BRL, Life Technologies, Scotland) in TBE-buffer [12]. Southern blots were performed as described [12]. The VR-1A insert [3] was labeled with ^{32}P using random priming (Megaprime, Amersham). The membrane was washed twice for 15 min in 2 × SSC, 0.1% SDS at 65 °C and three times for 15 min in 0.1 × SSC, 0.1% SDS at 65 °C.

Release, preparation and analysis of oligosaccharides from gpA

The preparation of acidic mucin oligosaccharides from rat small intestinal mucin glycopeptide A (5 mg) have been described elsewhere [2]. The neutral oligosaccharides were removed by applying the mixture of alkaline borohydride released mucin oligosaccharides on a DE23-cellulose column, eluting the neutral oligosaccharides with water, and the acidic ones with 1.0 M pyridinium acetate, pH 5.4. The acidic oligosaccharides were further fractionated into sialylated and sulfated species on 3 ml DEAE-Sephadex A-25 (Pharmacia, Uppsala) as described for porcine intestinal mucin oligosaccharides [13]. The sialic acid residues were methyl-esterified by three sequential incubations of the column with 0.6 ml DMSO: iodomethane (Fluka) 5:1 for 5 min each. The methyl esterified sialylated oligosaccharides were eluted with 50 ml methanol and the sulfated ones with 50 ml 1.0 M pyridinium acetate (pH 5.4) together with the iodide from the on-column esterification procedure. The iodide was removed

after concentrating the sulfated oligosaccharide fraction by desalting on a 80 ml column with Sephadex G-10, eluted with water:n-butanol 100:1 (10 ml h^{-1}) and collected in 1 ml factions. Fractions 20–57 were pooled, lyophilized and perdeuteroacetylated with 400 μl pyridine and 200 μl acetic acid anhydride-d_6 (Sigma, MO) for 12 h in room temperature before analysing the sample with negative ion FAB-MS/MS.

Analysis of sulfated oligosaccharides by tandem mass spectrometry

The FAB-MS/MS experiments were performed in the negative mode on a JEOL HX/HX 110A four sector tandem mass spectrometer equipped with JEOL MS-ADS II focal plane array detector for detection of the daughter ions in MS2. The mixture of peracetylated sulfated oligosaccharides (1–2 μl), dissolved in 100 μl methanol, was mixed with triethanolamine as matrix, ionized by fast atom bombardment using 6 keV Xe atoms, and the negative ions were analysed. The acceleration voltage was –10 kV and the resolution for MS1 and MS2 was set to 1000. $[M-H]^-$ of the sulfated oligosaccharides were selected as primary ions for collision induced dissociation (CID). The primary ion was attenuated by approximately 70% in the collision cell which was floating at 8 kV using helium as collision gas. CID-spectra were obtained from *m/z* 70 up to 30 amu below the molecular mass of the precursor ions. FAB-MS in the negative mode of the mixture of peracetylated sulfated mucin oligosaccharides prior to the MS/MS experiments was performed on a JEOL SX102A mass spectrometer which was scanned from m/z 100–3200 with a cycle time of 36 s using a linear magnet scan. The resolution was set to 1000 and the accelerating voltage was –8 kV.

Results

Extraction of rat intestinal mucosa with 6 M guanidinium chloride without high shear forces has revealed an 'insoluble' mucin complex [2]. This was solubilized by reduction of disulfide bonds and the protease resistant high-glycosylated domains were prepared after trypsin digestion [2].

Immunolocalization of gpA

To analyse the cellular origin of the 'insoluble' mucins in the rodent small intestine an antiserum raised against the HF-deglycosylated gpA (α-gpA) was used. The staining was very specific for the goblet cells with no staining of the enterocytes (Fig. 1). The antigen was only detected inside the cells and no staining of the upper parts of the goblet cell harboring the storage vesicles was observed. This was consistent with the antiserum being directed against the nonglycosylated apoprotein and that the epitopes were protected by glycosylation [14].

Northern blot analysis of the transcripts in rat small intestine and a human MUC2-producing cell line

To further substantiate that the clone (VR-1A) obtained by expression cloning using the α-gpA antiserum was the rat Muc2 the size of the transcripts in rat small intestine and a human colorectal cell line (LS 174T) was compared. As shown in Fig. 2 the VR-1A probe reacted with a single sharp band in the small intestine of the GOT-W strain. This band had an estimated size more than 13 kb and the human MUC2 probe from the large tandem repeat stained a band slightly larger (about 0.5 kb larger) in the human cell line. The human probe never reacted with the rat mRNA, but the VR-1A probe gave a variable weak staining of a band with similar size as the human probe in the LS 174T cells.

The size of the large Ser/Thr/Pro rich domain as revealed by PCR of rat intestinal mRNA

To analyse whether the length of gpA was compatible with the length of the mRNA encoding this part of the Muc2 gene, reverse transcription and PCR on mRNA from the rat small intestine were performed. The full sequence of this part of the rat Muc2 mucin is not known, but both Cys-rich sequences flanking the large Ser/Thr/Pro domain were available and used to select primer pairs. A band of about 5.5 kb was consistently found (see two reactions in lane 2 and 3 of Fig. 3A), despite technical difficulties in performing cDNA synthesis and PCR over long stretches of partially repeating mRNA. A major smaller band of about 1.4 kb was found, also in reactions containing only the reverse primer. The gel was blotted onto nitrocellulose and probed with the VR-1A cDNA (Fig. 3B). The larger band in both the identical PCR reactions was strongly stained. In addition, several smaller bands were stained in one of the reactions (lane 3, Fig. 3B). However, these bands were minor, probably incomplete PCR products, as these were not stained with ethidium bromide and only found in one PCR reaction out of three.

Analysis of the oligosaccharides of glycopeptide A

The oligosaccharides of glycopeptide A prepared from the GOT-W rat strain has been released by alkaline borohydride treatment. The oligosaccharides were separated into neutral, sialic acid-containing and sulfate-containing fractions and the relative amounts of these three fractions estimated to 40, 40 and 20%, respectively (Table 1) [2]. The oligosaccharide structures of the neutral and sialic acid-containing fractions have been analysed earlier [2].

Negative tandem ion mass spectometry of sulfated oligosaccharides

The presence of a sulfate group on perdeuteroacetylated mucin oligosaccharides made them suitable for analysis by negative ion FAB-MS. In Fig. 4, the mass spectrum of the

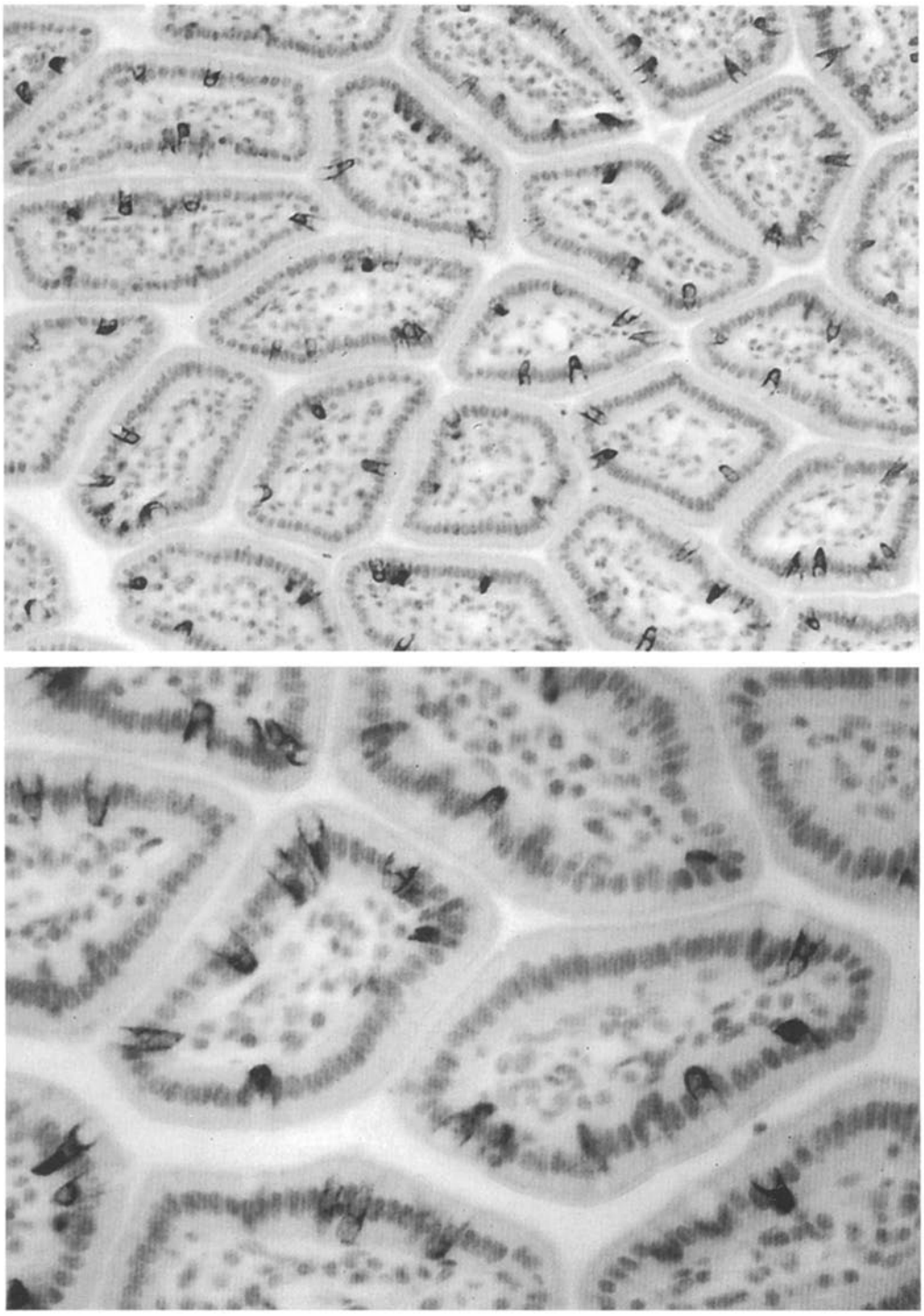

Figure 1. Immunostaining of normal mouse small intestine using the α-gpA antiserum. The upper photo is taken at 150× and the lower at 300× magnification.

mixture of the perdeuteroacetylated sulfated oligosaccharides is shown. The spectrum is dominated by the $[M-H]^-$-ions and no interpretable fragmentation is seen. Eighteen oligosaccharides were found and these were distributed from monosulfated trisaccharides up to low intense $[M-H]^-$-ions for monosulfated octasaccharides.

Analyses by MS/MS and interpretation of the collision induced dissociation (CID) spectra of the $[M-H]^-$-ions

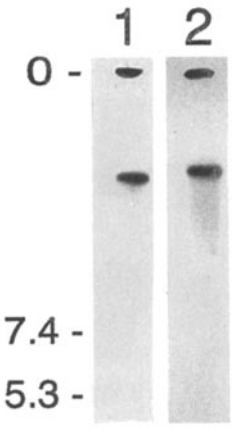

Figure 2. Northern blot of mRNA from normal rat small intestine (lane 1) and the human colorectal carcinoma cell line LS174T (lane 2). The VR-1A probe was used in lane 1 and a human MUC2 tandem repeat probe in lane 2. The application well in each lane is labelled by a marker and to the left with 0; standard mRNA bands are shown.

from the sulfated oligosaccharide fraction gave a total of fourteen assigned structures, all sharing the structural element of sulfation on C-6 of *N*-acetylglucosamine attached to C-6 of the *N*-acetylgalactosaminitol (Table 1). Low intense $[M-H]^-$-ions in the FAB-MS spectrum were not analysable by tandem mass spectrometry.

The assignment of fragment ions in the CID-spectra was done according to the nomenclature proposed by Domon and Costello [15] extended with the A_0-cleavages of the *N*-acetylgalactosaminitol residue indexed with α (C-6 branch) or β (C-3 branch) for describing which branch that retained the charge. All hexoses were assumed to be galactoses, *N*-acetylhexosamines attached to *N*-acetylhexosaminitols to be *N*-acetylglucosamines, deoxyhexoses to be fucoses, and the reduced *N*-acetylhexosaminitols to be *N*-acetylgalactosaminitols.

Characteristic ions in the CID-spectra were the intense $^{1,5}X$ fragment ions and the Z-OAc fragment ions, but information was also obtained from the A_0-cleavages of the *N*-acetylgalactosaminitol residue, locating the sulfate group to the C-6 branch. Characteristic features of the fragmentation are shown by the CID-spectra of the tetrasaccharide $[M-H]^- = 1370$ (Fig. 5). The intense fragment ions of *m/z* 80 and *m/z* 97 were characteristic of the sulfate group and the *m/z* 139 also present in this and all other spectra arose from the $^{0,4}A_{2\alpha}$-cleavages with the sulfate group at C-6 of the *N*-acetylglucosamine residue. The presence of *m/z* 495 from an $^{3,5}A_{2\alpha}$-cleavage in the sulfated *N*-acetylglucosamine residue was informative for placing the attached galactose residue to C-4 of the *N*-acetylglucosamine.

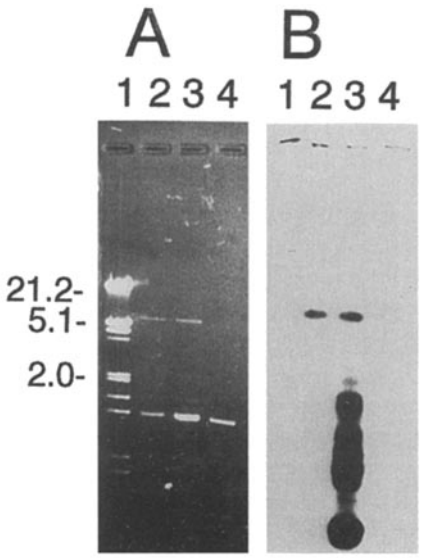

Figure 3. Ethidium bromide stained agarose gel electrophoresis of PCR products (A) and Southern blot analysis of the same gel (B). Agarose (1%) in TBE-buffer was used. Lane 1, molecular weight markers; lane 2, products of PCR with both reverse and forward primers included; lane 3, products of repeated PCR as in lane 2; lane 4, products of control PCR using only the reverse primer. Southern blot (B) was performed using the VR-1A clone as probe.

Discussion

About 80% of the mucins of the small intestine of rat was earlier found to be 'insoluble' in guaninidinium chloride. This insoluble 'complex' contained two trypsin-resistant domains, named glycopeptide A and B. A cDNA corresponding to a portion of the larger of these (gpA) was cloned and sequenced, part of which encoded a Cys-rich peptide showing high sequence homology with the central Cys-rich domain of the human MUC2 gene [3]. It was suggested that the 'insoluble' mucin of rat small intestine was encoded by the Muc2 gene [3]. In the present investigation we further strengthen this conclusion.

The major part of the intestinal mucins are produced by the goblet cells. By immunostaining it was shown that the glycopeptide A is only produced by goblet cells. This is the same localization as the human MUC2 mucin [16]. The α-gpA antiserum only stained the goblet cell in a paranuclear distribution, a finding consistent with the serum only recognizing the nonglycosylated form of the Muc2 mucin.

RT-PCR using two primers gave a distinct band of about 5.5 kb that hybridized with the VR-1A probe derived from a sequence within the predicted product. Another band consistently found at 1.4 kb was obtained with only the reverse primer and did not hybridize with the VR-1A probe. This probe contain 550 nucleotides (out of 705) encoding a Ser/Thr rich sequence and should probably also hybridize with related sequences if present. This propose that this band was derived from another

Table 1. Proposed structures of perdeuteroacetylated sulfated oligosaccharides in mucin glycopeptide A.

[M-H]⁻	*Structure*[a]	*Mol%*[b]
1068	HO_3S-6GlcNAc1-6 **GalNAcol** GlcNAc1-3	0.6
1072	HO_3S-6GlcNAc1-6 **GalNAcol** Gal1-3	3.0
1308	HO_3S-6GlcNAc1-6 **GalNAcol** Fuc1-2Gal1-3	2.4
1366	HO_3S-6GlcNAc1-6 **GalNAcol** GlcNAc1-Gal1-3	1.0
1366	Gal1-4(HO_3S-6)GlcNAc1-6 **GalNAcol** GlcNAc1-3	
1370	Gal1-4(HO_3S-6)GlcNAc1-6 **GalNAcol** Gal1-3	1.6
1602	HO_3S-6GlcNAc1-6 **GalNAcol** Fuc1-2Gal1-GlcNAc1-3	1.7
1602	Fuc1-2Gal1-4(HO_3S-6)GlcNAc1-6 **GalNAcol** GlcNAc1-3	
1606	Fuc1-2Gal1-4(HO_3S-6)GlcNAc1-6 **GalNAcol** Gal1-3	3.6
1606	Gal1-4(HO_3S-6)GlcNAc1-6 **GalNAcol** Fuc1-2Gal1-3	
1663	Gal1-4(HO_3S-6)GlcNAc1-6 **GalNAcol** Gal1-3GlcNAc1-3	0.5
1842	Fuc1-2Gal1-4(HO_3S-6)GlcNAc1-6 **GalNAcol** Fuc1-2Gal1-3	1.4
1898	HO_3S, Fuc, Gal_2, $HexNAc_2$, **GalNAcol**[c]	1.0
1960	Gal1-HexNAc1-Gal1-4(HO_3S-6)GlcNAc1-6 **GalNAcol** Gal1-3	0.6
2135	HO_3S, Fuc_2, Gal_2, $HexNAc_2$, **GalNAcol**[c]	0.7
2196	Fuc1-2Gal1-HexNAc1-Gal1-4(HO_3S-6)GlcNAc1-6 **GalNAcol** Gal1-3	1.2
2432	HO_3S, Fuc_2, Gal_3, $HexNAc_2$, **GalNAcol**[c]	0.4
2489	HO_3S, Fuc, Gal_3, $HexNAc_3$, **GalNAcol**[c]	0.3
Total amount sulfated[d]		20
Total amount neutral[d]		40
Total amount sialylated[d]		40

[a]Proposed structures were based on analyses by negative ion MS/MS. All hexoses were assumed to be Gal, all reduced *N*-acetylhexosamines to be GalNAcol, all deoxyhexoses to be Fuc linked in 1-2 as in the blood group H-determinates, and all *N*-acetylhexosamines (HexNAc) linked to the GalNAcol to be GlcNAc, relating to sugar analyses of the fraction showing only 10% GalNAc compared to GlcNAc, and in accordance with that no branched core structures with GalNAc attached to the protein linking GalNAc substituted both at C-3 and C-6 have been described. [3,5]A fragment ions was used to deduce the C-3 or C-4 linkages of substituents attached to the HO_3S-6GlcNAc residues.

[b]Calculations were made by normalizing the intensities of the $[M-H]^-$ ions in the FAB MS-spectrum with the background.

[c]Detected in the FAB MS-spectrum, but not analysable with MS/MS.

[d]The distribution of sulfated, neutral, and sialylated oligosaccharides from mucin glycopeptide A were from Carlstedt *et al.* [2].

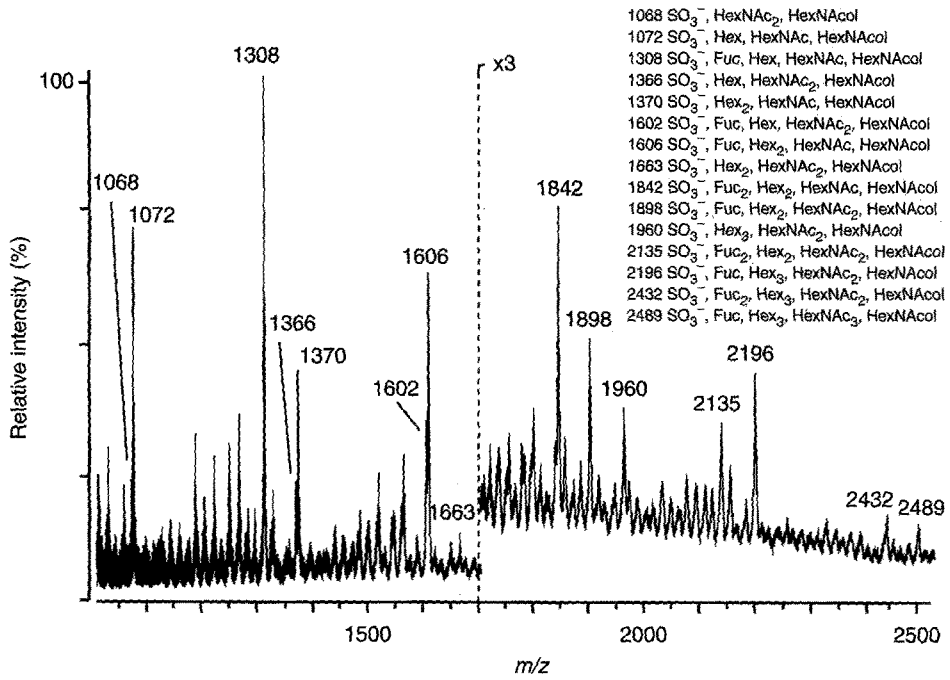

Figure 4. FAB-MS spectrum of the perdeuteroacetylated sulfated oligosaccharides from rat small intestinal mucin glycopeptide fraction A.

mRNA. The large PCR band was thus the only product in line with a predicted domain encoding the large Ser/Thr domain of Muc2 and shows at the same time that the sequenced central and N-terminal Cys-rich domains [3, 7, 8] indeed were derived from the same gene.

The human MUC2 mucin contains two sequences rich in hydroxy amino acids, probably being highly glycosylated (Fig. 6). The longer one of these is larger than that of the proposed rat homologue (gpA) as estimated from biochemical data [2]. The nucleotide sequence encoding gpA is not known and to address the question of the length of gpA the length of the cDNA encoding gpA was estimated using reverse transcriptase-PCR. To avoid possible problems of different allelic sizes in the same or different animals we used the same inbred rat strain as RNA source as was used for preparing gpA. The agarose gel of the PCR products revealed one larger band with an estimated size of about 5.5 kb corresponding to about 1800 amino acids. Parts of the flanking Cys-rich domains, not encoding for Ser or Thr, were used for the selected primers excluding about 90 nucleotides on the 5′ side and about 155 nucleotides on the 3′ side, leaving a region encoding about 1700 amino acids. This is larger than the biochemically estimated 1100 amino acids in gpA [2]. However, the trypsin sites that generated gpA is not known. In fact, this molecule was not homogeneously digested as Edman sequencing gave multiple amino acids in all cycles. Another explanation for this discrepancy in size could be that the nucleotide sequence of this Ser/Thr-region is not known and that additional stretches with low glycosylation allowing trypsin cleavage could be present. One such sequence has already been found on the C-terminal side of rat Muc2 [7]. The earlier estimated size of gpA using laser light scattering could also be influenced by problems related to determining the absolute concentration of the sample and that the value of dn/dc was not determined on the actual preparation.

The size of the transcript containing the sequence of the cloned fragment of gpA (named VR-1A) was estimated to be slightly smaller (0.5 kb) in the GOT-W rat strain than the human MUC2 transcript in the colorectal cell line LS 174T. A second inbred rat strain (GOT-BW) had a size similar to the human MUC2 (results not shown). The sizes were not accurately determined as no large standard mRNA markers were available. The human transcript of a common allele has been estimated to about 15 kb and to encode a protein of a little more than 5100 amino acids [6]. The rat Muc2 was only about 0.5 kb smaller proposing an almost identical apoprotein (about 5000 amino acids) assuming similar length of the nontranslated regions. The sequence of the N- and C-terminal portions of the proposed Muc2 is known [7–9], as is part of the central Cys-rich domain [3]. This give a total of more than 2300 known amino acids in the Cys-rich domains (Fig. 6). This propose an upper limit of 2700 amino acids in the Ser/Thr rich domains.

The biochemical data suggested a 1:1 molar relation between gpA and gpB [2] and the size of gpB was estimated to about 50% of that of gpA by both laser light scattering and electron microscopy. The number of amino

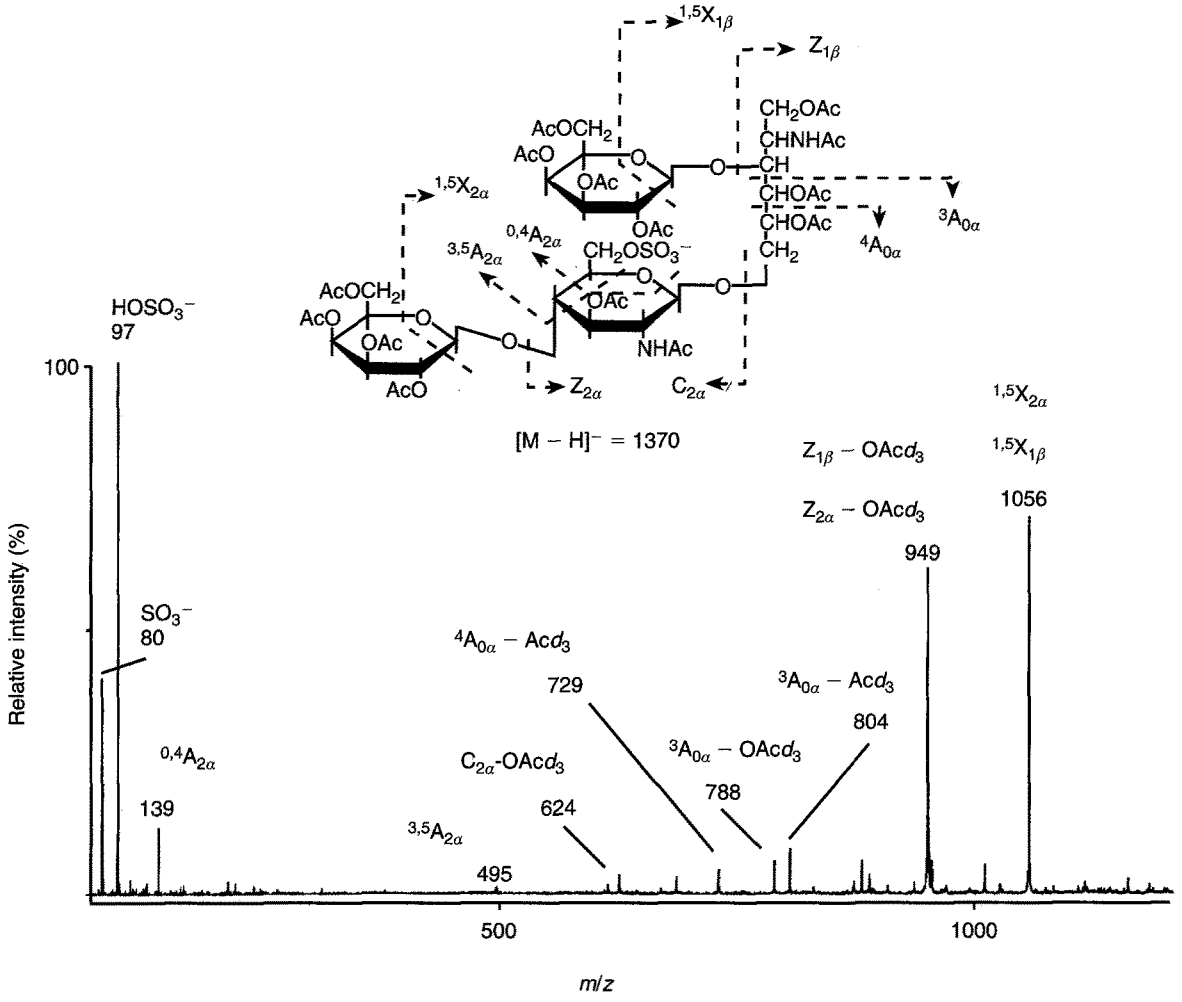

Figure 5. Negative ion CID-spectrum of the precursor ion $[M\text{-}H]^- = 1370$ of a monosulfated tetrasaccharide, with assignment of fragment ions according to Domon and Costello [15]. All acetyl groups in the figure are deuterated except for the *N*-acetyl groups of the amino sugar residues.

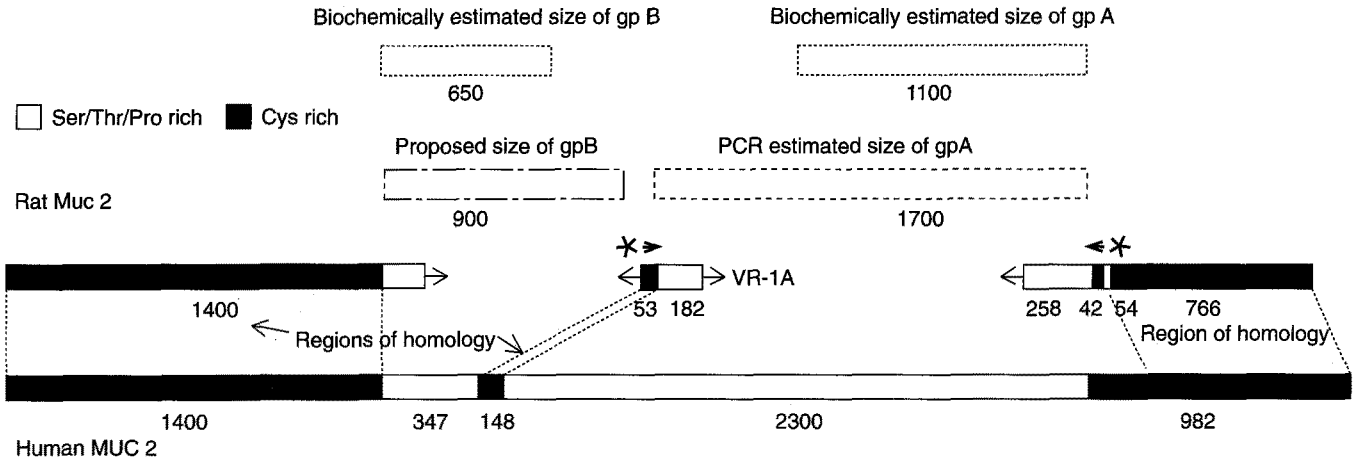

Figure 6. Outline of the rat Muc2 and human MUC2 mucin amino acid sequences. The estimated size of the large high-glycosylated domain, glycopeptide A, from biochemical data [2] and the present PCR data have been included. The position of the PCR primers have been marked by ∗-arrows. The numbers refer to the number of amino acids in each domain.

acids in gpA was here estimated to 1700 amino acids and the number in gpB could thus be proposed to be about 900. The length of the small human Ser/Thr is about 350 amino acids and 100 tandem repeats in the large domain give 2300 residues, giving a total sum of 2650 amino acids compared to the 2600 residues of the rat, a size compatible with the estimates based on mRNA size. The total size of the two mucin domains in rat and human are thus probably similar, although the individual lengths are different. An outline of the proposed rat Muc2 amino acid sequence is presented and compared to the human MUC2 in Fig. 6. Only parts of the large Ser/Thr domain of rat Muc2 are known, but the sequence is not similar to the large tandem repeat of human MUC2. These may suggest an evolutionary pressure for maintaining the total length of the mucin and the high-glycosylated domains, but not for its sequence or the individual lengths of the two high glycosylated domains.

The arguments presented here together with earlier results [2, 14] further strengthen the suggestion that the 'insoluble' mucins of rat small intestine are indeed encoded by the Muc2 gene. The intestinal mucins of the human large intestine are also 'insoluble' [17] and subunits purified from this complex have recently been shown to contain reactivity with the MUC2 apoprotein [18].

The oligosaccharides of the glycopeptide A from the GOT-W rat small intestine have been shown to contain about 40% neutral oligosaccharides, 40% sialic acid-containing oligosaccharides, the remaining 20% of the glycans being sulfated. The neutral and sialic acid-containing oligosaccharides were presented before [2], including an estimation of the number of individual oligosaccharides found on each glycopeptide. Here, information on the sulfated oligosaccharides from gpA are added. Eighteen oligosaccharide structures were found of which 14 were possible to characterize. The sulfate group was located on the C-6 of a *N*-acetylglucosamine residue linked to C-6 of the *N*-acetylgalactosaminitol in all the characterized oligosaccharides. This sugar residue was either terminal or further elongated with up to at least four residues. A sulfate group located at this position is prevalent in oligosaccharides of tracheobronchial mucins and hen ovomucin [19–21] and is a common feature of keratan sulfate oligosaccharides. However, of all the structures described here only the tetrasaccharide with two terminal galactoses ($[M-H]^-$ 1370, Table 1) has been described earlier [19–22]. The relative amounts of the individual oligosaccharides given in Table 1 are based on the intensity of the molecular ions ($[M-H]^-$) and thus only semiquantitative.

In summary, the gpA domain of rat 'insoluble' mucin complex is part of the major intestinal goblet cell mucin encoded by the Muc2 gene. The cDNA for this domain is only slightly smaller than the tandem repeat domain of the human MUC2 mucin, but DNA-hybridization results indicate poor sequence homology. The present characterization of sulfated oligosaccharides of gpA completes our goal to describe in detail the glycosylation of a defined high-glycosylated mucin domain.

Acknowledgements

This work was supported by grants from the Swedish Medical Research Council (No. 7461, 7902, and 10226), the Medical Faculty of Lund and the IngaBritt and Arne Lundbergs Stiftelse.

References

1. Carlstedt I, Karlsson H, Sundler F, Fransson L-Å (1982) *Adv Exp Med Biol* **144**: 155–57.
2. Carlstedt I, Herman A, Karlsson H, Sheehan J, Fransson L-Å, Hansson GC (1993) *J Biol Chem* **268**: 18771–81.
3. Hansson GC, Baeckström D, Carlstedt I, Klinga-Levan K (1994) *Biochem Biophys Res Commun* **198**: 181–90.
4. Gum JR, Hicks JW, Sack TL, Kim YS (1990) *Cancer Res* **50**: 1085–91.
5. Gum JR, Hicks JW, Toribara NW, Rothe EM, Lagace RE, Kim YS (1992) *J Biol Chem* **267**: 21375–83.
6. Gum JR, Hicks JW, Toribara NW, Siddiki B, Kim YS (1994) *J Biol Chem* **269**: 2440–46.
7. Xu G, Huan L, Khatri IA, Wang D, Bennick A, Fahim REF, Forstner GG, Forstner JF (1992) *J Biol Chem* **267**: 5401–7.
8. Xu G, Huan L, Khatri I, Sajjan US, McCool D, Wang D, Jones C, Forstner G, Forstner J (1992) *Biochem Biophys Res Commun* **183**: 821–28.
9. Ohmori H, Dohrman AF, Gallup M, Tsuda T, Kai H, Gum JR, Kim YS, Basbaum CB (1994) *J Biol Chem* **269**: 17833–40.
10. Gendler SJ, Spicer AP (1995) *Ann Rev Physiol* **57**: 607–34.
11. Baeckström D, Hansson GC (submitted) *Glycoconjugate J.*
12. Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular Cloning.* Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
13. Karlsson NG, Karlsson H, Hansson GC (1995) *Glycoconjugate J* **12**: 69–76.
14. Asker N, Baeckström D, Axelsson MAB, Carlstedt I, Hansson GC (1995) *Biochem J* **308**: 873–80.
15. Domon B, Costello CE (1988) *Glycoconjugate J* **5**: 387–409.
16. Chang SK, Dohrman AF, Basbaum CB, Ho SB, Tsuda T, Toribara NW, Gum JR, Kim YS (1994) *Gastroenterology* **107**: 28–36.
17. Mårtensson S, Lundblad A, Hansson GC, Carlstedt I (1988) *Scand J Clinical Invest* **48**: 633–40.
18. Carlstedt I, Herrmann A, Lindell G, Nordman H (1995) *Glycoconjugate J* **12**: 495.
19. Mawhinney TP, Landrum DC, Gayer DA, Barbero GJ (1992) *Carbohydr Res* **235**: 179–97.
20. Lo-Guidice JM, Wieruszeski JM, Lemoine J, Verbert A, Roussel P, Lamblin G (1994) *J Biol Chem* **269**: 18794–813.
21. Strecker G, Wieruszeski J-M, Martel C, Montreuil J (1987) *Glycoconjugate J* **4**: 329–37.
22. CarbBank (1995) University of Georgia, WA.