Vol. 198, No. 1, 1994
January 14, 1994

BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
Pages 181-190

MOLECULAR CLONING OF A cDNA CODING FOR A REGION OF AN APOPROTEIN FROM THE 'INSOLUBLE' MUCIN COMPLEX OF RAT SMALL INTESTINE

Gunnar C. Hansson*, Dan Baeckström*, Ingemar Carlstedt‡, and Karin Klinga-Levan§

*Department of Medical Biochemistry, University of Göteborg, Medicinaregatan 9, S-413 90 Gothenburg, Sweden

‡Department of Physiological Chemistry, University of Lund, P.O.Box 94, S-221 00 Lund, Sweden

§Department of Genetics, University of Göteborg, Guldhedsgatan 23, 413 90 Gothenburg, Sweden

Received November 24, 1993

The major part of rat small intestinal mucins occurs as an 'insoluble' glycoprotein complex unextractable in 6 M guanidinium chloride unless disulfide bonds are cleaved. One of the trypsin-resistant high-glycosylated domains of this complex (glycopeptide A) was recently isolated. We have now de-glycosylated it with HF, injected it into rabbits and the obtained antiserum was used for expression cloning providing a cDNA clone (VR-1A). This clone contained an open reading frame of 235 amino acids composed of two regions. The deduced N-terminal 53 amino acids included seven Cys residues and only one Ser, followed by a region of 182 residues with 64% Ser and Thr but devoid of Cys residues. Analysis of mRNA revealed a transcript of about 12 kb, identical in size to a band labelled with a probe based on the rat mucin-like protein (MLP/Muc2) cDNA. Pulsed-field gel electrophoresis of genomic rat DNA showed identical bands (380 and 500 kb) when blots were sequentially probed with the MLP/Muc2 probe and VR-1A. A panel of mouse x rat hybrids was used to localize the gene corresponding to both VR-1A and Muc2 to rat chromosome 1. The results strongly suggest that the 'insoluble' mucin complex of the rat small intestine is encoded by the Muc2 gene.

Mucins are a heterogeneous group of highly glycosylated glycoproteins with an apoprotein rich in Ser/Thr residues being glycosylated to at least 50% with O-linked oligosaccharides. The 'classical' mucins are gel-forming molecules found on mucosal surfaces (1). The biochemical understanding of these macromolecules has evolved relatively slowly because of great difficulties in obtaining pure preparations due to their heterogeneity and large size. The development has taken pace by improvement in purification techniques (reviewed in ref. 2) and by the cloning and sequencing of several mucin apoprotein genes (4-9). Most of these cDNA sequences are only partial, but it can be concluded that domains with tandem repeats rich in Ser/Thr are structural features where these regions probably correspond to the highly glycosylated domains found in mucins. However, the corresponding macromolecules have only been purified and characterized in a few instances (10, 11).

The mucins of the intestinal mucus layer are largely produced by the goblet cells and in contrast to cervical, respiratory, and gastric mucins the major parts of the intestinal mucus glycoproteins are not

Abbreviations used: kb, kilobase pairs; bp, base pairs; PCR, polymerase chain reaction; PFGE, pulsed field gel electrophoresis.

0006-291X/94 $5.00

extractable by slow stirring in denaturating solvents (12-14). Recently we have shown that only about 20% of the mucins from rat small intestine are extracted in 6.0 M guanidinium chloride unless disulfide bonds are cleaved (14). The major part (80%) of the mucins was obtained as an 'insoluble' glycoprotein complex and was shown to contain two highly-glycosylated trypsin-resistant domains (named glycopeptide A and B) with an M_r of 650 and 335 kDa, respectively. Electron microscopy revealed linear structures, where glycopeptide A had an average length of 230 nm and B 110 nm. The chemical compositions of the two glycopeptides were similar, having 17-19% protein. Of the oligosaccharide chains, 40% were neutral and 38-40% contained sialic acid, both showing complex, but identical patterns, when analyzed in detail by high-temperature gas chromatography-mass spectrometry (14). The major difference between the glycopeptides was thus the lengths of the apopeptides. Glycopeptides A and B are most likely derived from the same mucin as they occurred in about equimolar concentrations. An antiserum against deglycosylated glycopeptide A was obtained and used to isolate a cDNA clone. The corresponding genomic sequence was assigned to rat chromosome 1 and our evidence suggests that the cDNA clone is part of the rat Muc2 gene.

EXPERIMENTAL PROCEDURES

Isolation of Mucin Glycopeptides - High molecular mass mucin glycopeptides from the 'insoluble' mucin complex of rat small intestine were isolated from an inbred Sprague-Dawley strain (GOT-W) as described (14). In short, mucosal scrapings were reduced with dithiotreitol in guanidinium chloride, alkylated with iodoacetamide, digested with DNase and RNase and finally with trypsin. The glycopeptides were purified by gel filtration on Sephacryl S-200 and separated into glycopeptide A and B on Sephacryl S-500.

Deglycosylation and immunization - Glycopeptide A (10 mg) was dried under vaccum over P_2O_5 for 24 h at room temperature. Deglycosylation was performed essentially as described (15). Briefly, anhydrous methanol (1 ml) was added and samples transferred to an all-teflon HF apparatus (Peptide Institute Inc., Osaka, Japan). About 10 ml of HF was transferred and the vial was incubated for 3 h at room temperature while stirring using a magnetic bar. After evaporation of the HF and drying in vacuum, samples were dissolved in 50 mM pyridinium acetate, pH 5.4, dialyzed against this buffer and lyophilized. The amino acid compositions were determined as described (14).

HF-deglycosylated glycopeptide A (65 µg of protein) dissolved in 400 µl 0.9% NaCl was injected subcutaneously into New Zeeland White rabbits in an emulsion of Freund's complete adjuvant (600 µl). Two booster injections with the same amount of protein in Freund's incomplete adjuvant were given on day 21 and 42. Sera were collected weekly for 5 weeks.

Fluoroimmunoassays - HF-deglycosylated glycopeptide A (100 µl) was coated into polystyrene microtiter strips (Polysorp C-12, Nunc, Roskilde, Denmark) at a concentration corresponding to 5 mg of the original glycopeptide/ml over night at 4 °C. The strips were blocked with 2% bovine serum albumin in 50 mM Tris-HCl, pH 7.8, 0.15 M NaCl, 4 µM $CaCl_2$, 90 µM EDTA, 0.05% NaN_3 at 37 °C over night. The sera diluted in 1% bovine serum albumin-phosphate buffered saline were serially diluted and incubated for 2 h while shaking at room temperature. The strips were washed (16) and incubated for 2 h at room temperature with biotin-labelled swine anti-rabbit antibodies (DAKO, Glostrup, Denmark) diluted 1:2,000. After washing 5 times, europium-labelled streptavidin was added, incubated for 1 h, washed and the fluorescence measured in an Arcus fluorometer (Wallac, Turku, Finland) (16).

Screening of cDNA-library and Cloning - A Sprague Dawley rat jejunum cDNA library (oligo-dT and random primed, Stratagene, LaJolla, CA) in lambda ZAP II was kindly provided by Drs. Y. Kim and J. Gum, San Francisco, CA. The library was plated at 30,000 plaques/14 cm plate using XL-1 Blue bacteria. After an initial incubation at 42 °C until the plaques started to appear, the plates were overlayed with 10 mM isopropyl-β-D-thiogalactopyranoside-soaked (Boehringer-Mannheim) nitro-cellulose membranes (Schleicher & Schuell, Dassel, Germany). The membranes were blocked in 20% fetal calf serum and incubated with the antiserum in a 1:100 dilution for 1h. The membranes were washed in 20 mM Tris buffered saline, pH 7.5, containing 0.05% Tween-20 (Sigma, St. Louis, MO)

and incubated with alkaline phosphatase-conjugated goat anti-rabbit immunglobulins (DAKO) at a dilution of 1:2,000 for 1 h and developed in nitro blue tetrazolium (0.3 mg/ml) and 5-bromo-4-chloro-3-indolyl phosphate (0.15 mg/ml) (both from Sigma). Positives were cloned in three rounds, the phage DNA isolated (17), the inserts recovered by *EcoR*I digestion (Boehringer-Mannheim) and placed into the Bluescript-SK plasmid (Stratagene).

DNA Sequencing and Analyses - The plasmids were sequenced by the dideoxy chain termination protocol with the Sequenase 2.0 system (USB, Cleveland, OH) and [α-^{35}S]-ATP (Amersham, Buckinghamshire, England). Exonuclease III deletions were generated using Erase-a Base (Promega, Madison, WI). Additional internal sequencing primers were made by Scandinavian Gene Systems (Köping, Sweden). Both strands were entirely sequenced. The sequences obtained were analyzed with the Wisconsin Sequence Analysis Package (GCG) on a Vax computer and the sequences compared with the EMBL data base.

Northern Blots - Total RNA was prepared using the guanidine thiocyanate method (18) and mRNA was extracted with oligo-dT-magentic beads (Dynal, Oslo, Norway). The electrophoresis and blotting was performed as described before (19). The size of the RNA was estimated using the RNA molecular weight markers II (Boehringer-Mannheim).

Oligonucleotide probes - A 217 bp oligonucleotide probe, called MLP/Muc2, from the mucin-like protein (the probable rat Muc2 homologue) was generated with PCR from rat small intestinal mRNA using primers corresponding to the nucleotides 440-459 and 657-636 of the published sequence[1] (20,21). The sequence of this probe was verified by DNA sequencing using the PCR primers as sequencing primers. Oligonucleotides were labelled using the random priming technique (Megaprime kit) and [α-^{32}P]-CTP (both from Amersham).

Cell material and culture conditions. - The somatic cell hybrid panel used for rat gene mapping has been described earlier (22,23). Most hybrids (LB and GB series) were derived from fusions of cells from the mouse BWTG3 hepatoma cell line and adult rat hepatocytes (22), but one hybrid (BS511) came from the fusion of BWTG3 cells with rat skin fibroblasts (24). The hybrids were cultured in DHAT, i. e. Dulbecco's modification of Eagle's mimimal essential medium (DMEM) with hypoxanthine, aminopterin and thymidine (25). LB hybrids (designated TG) that had been backselected in thioguanine did not grow in DHAT, and were cultured in DMEM, as were the mouse parental cells (BWTG3 hepatoma cells). All media were supplemented with L-glutamine and 10 or 15% fetal calf serum.

DNA extraction and Southern blotting. - Genomic DNA was extracted from the cultured cells (26) using a DNA extractor (Applied Biosystems, Foster City, CA). Rat parental DNA was extracted either from fresh liver tissue, from the rat fibroblast line FR3T3, or from newly established skin fibroblast cultures of Sprague-Dawley rats. Human parental DNA was extracted from blood. The genomic DNA was restricted to completion with *BamH*I, *Bgl*II, *Dra*I, *EcoR*I, *Hinc*II and *Hind*III. Fifteen μg of digested DNA were loaded into each slot of 1% agarose gels and the fragments were separated by electrophoresis. The DNA was bound to nylon filters (Hybond-N+, Amersham) after alkaline transfer from the gels essentially as in the original protocol (27). Two kinds of filters were made, one with DNA restricted with *Bgl*II containing the hybrid panel cell lines, mouse hepatoma DNA (BWTG3), rat liver DNA (Sprague-Dawley) and a rat cell line (FR3T3). The other type contained DNA from rat liver (Sprague-Dawley), a rat cell line (FR3T3), mouse hepatoma (BWTG3), and human blood restricted with six different enzymes, *BamH*I, *Bgl*II, *Dra*I, *EcoR*I, *Hinc*II, and *Hind*III. Hybridization to the filters was performed in 0.5 M phosphate buffer, pH 7.2, with 7% SDS and 1 mM EDTA at 65 °C over night. The filters were washed in 100-200 mM phosphate buffer, pH 7.2, with 1% SDS for 5 min at room temperature, and then for 2 x 40 min at 65 °C.

Pulsed field gel electrophoresis (PFGE). - PFGE was performed as described (28). Briefly, high molecular mass DNA was prepared from cells immobilized in agarose plugs. The plugs were treated with proteinase K and rinsed several times in buffer supplemented with 0.5 mM phenylmethylsulfonyl fluoride. The DNA in the plugs was subsequently digested with *Mlu*I or *Sac*I and the fragments were

[1]Klinga-Levan, K., Gum,. J., Gendler, S., and Hansson, G.C., in preparation.

separated by PFGE in an LKB 2015 Pulsaphor system using 1% agarose gels prepared in 50 mM Tris, 50 mM boric acid and 1 mM EDTA. The DNA was transferred to filters and probed as described above.

RESULTS

Deglycosylation and antiserum characterization - Glycopeptide A from the 'insoluble' mucin complex of rat small intestine (14) was deglycosylated using hydrogen fluoride. The overall yield of amino acids was 57% and the amino acid composition of the deglycosylated glycopeptide was similar to the non-deglycosylated one (data not shown). Residual GalNAc that accounted for 8% of the GalNAc

```
   1  TTGTGTTCCAATGTTCCACTTGATCAGCTTGGACAAAAAGTGGTTTGTAATATAGAAGAT   60
      ---------+---------+---------+---------+---------+---------+
   1  L  C  S  N  V  P  L  D  Q  L  G  Q  K  V  V  C  N  I  E  D    20

  61  GGTCTAGTGTGCAAAAATGCAGAACAGGGAATAGGAGGGATTATTCCCATGCGCATGTGT  120
      ---------+---------+---------+---------+---------+---------+
  21  G  L  V  C  K  N  A  E  Q  G  I  G  G  I  I  P  M  R  M  C    40

 121  CTCAACTATGAAATCAATGTTTACTGCTGTATATGTATTACTTCAACCCCACCATCCACC  180
      ---------+---------+---------+---------+---------+---------+
  41  L  N  Y  E  I  N  V  Y  C  C  I  C  I  T  S  T  P  P  S  T    60

 181  ACCACGGAGATCCAAACTACAACATCAACAACAAAGACTTCCATTCCAACATCAACTGCC  240
      ---------+---------+---------+---------+---------+---------+
  61  T  T  E  I  Q  T  T  T  S  T  T  K  T  S  I  P  T  S  T  A    80

 241  ATGAAGACACCAAGCCCATCTCCTACAACTACAGTGACTCCAACTCCAGAAACAACCACC  300
      ---------+---------+---------+---------+---------+---------+
  81  M  K  T  P  S  P  S  P  T  T  T  V  T  P  T  P  E  T  T  T   100

 301  ACACAGATTTCAACTTCAACATCTACTACCACCAAGATCACAACCCCAACACCCATCACG  360
      ---------+---------+---------+---------+---------+---------+
 101  T  Q  I  S  T  S  T  S  T  T  T  K  I  T  T  P  T  P  I  T   120

 361  GAGACTTCCACTCCTATATCTACCACCAGCCAGACACCAAGTCCAGCCTCTACTACCACA  420
      ---------+---------+---------+---------+---------+---------+
 121  E  T  S  T  P  I  S  T  T  S  Q  T  P  S  P  A  S  T  T  T   140

 421  GTGACTCCAGTTACAACATCTACTACTACAGAGACCTCAATGTCATTATCTACTACTACA  480
      ---------+---------+---------+---------+---------+---------+
 141  V  T  P  V  T  T  S  T  T  T  E  T  S  M  S  L  S  T  T  T   160

 481  CAGACCACATCTGCAACACCTATAGTAACGGAGACATCTACTCCTAGATCTACCACTACT  540
      ---------+---------+---------+---------+---------+---------+
 161  Q  T  T  S  A  T  P  I  V  T  E  T  S  T  P  R  S  T  T  T   180

 541  CAGACACCAAGCACAGTCTCCACCACCACAGGGACTCCAACTACGACACCAACAATTGGA  600
      ---------+---------+---------+---------+---------+---------+
 181  Q  T  P  S  T  V  S  T  T  T  G  T  P  T  T  T  P  T  I  G   200

 601  AAGACGACAACCCCAACATCTACAGTGACAGAGACATCTACTCCTACATCAACCACTACT  660
      ---------+---------+---------+---------+---------+---------+
 201  K  T  T  T  P  T  S  T  V  T  E  T  S  T  P  T  S  T  T  T   220

 661  CAGACATCAAGTCCAACCCCTACAACCACAGGGACTCCACCCTCG  705
      ---------+---------+---------+---------+-----
 221  Q  T  S  S  P  T  P  T  T  T  G  T  P  P  S    235
```

Fig. 1. Nucleotide sequence of the VR-1A clone and the corresponding amino acid sequence. The first 158 nucleotides encode a Cys-rich region marked by a bold separator line. All Ser and Thr residues are bold. The cDNA sequence is directly flanked by the linkers used during cDNA construction. The amino acids are given with the one letter code.

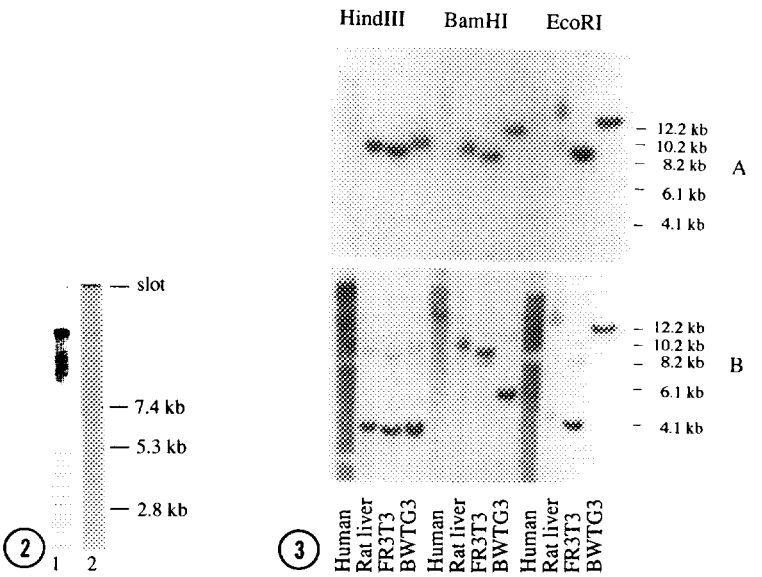

Fig. 2. Comparison of rat small intestine mRNA size stained with the VR-1A (lane 1) and the rat Muc-2 probe (MLP) (lane 2). Each lane contains mRNA from about 30 μg of total RNA. The numbers to the right are the sizes of the RNA marker.

Fig. 3. Southern blot analysis of genomic DNA. Sequential hybridization to the same blot first with the MLP/Muc2 probe in B and then, after deprobing of the filter, with the VR-1A probe (A). Filters contain genomic DNA from human leukocytes (marked human), rat liver, rat FR3T3 fibroblasts and mouse BWTG3 hepatoma cells, restricted with *Hind*II, *BamH*I or *Eco*RI as indicated.

in the original glycopeptide was found, but no GlcNAc. A rabbit antiserum that showed strong reactivity against HF-deglycosylated glycopeptide A was obtained.

cDNA cloning and sequencing - Screening of about 600,000 plaques with the antiserum against deglycosylated A glycopeptide gave one consistently reactive clone (named VR-1A). The VR-1A clone contained a 705 bp insert and the sequence revealed an open reading frame coding for 235 amino acids (Fig. 1). The first 53 amino acids contain 7 Cys and only one Ser/Thr whereas the remaining sequence lacks Cys and has 64% of the amino acids as Ser and Thr and 14% as Pro.

Size of corresponding mRNA - Probing rat small intestinal mRNA with VR-1A in Northern blots gave a hybridizing band with a size of about 12 kb (Fig. 2). A band of identical size was obtained when the rat mucin-like protein (MLP/Muc2) probe was hybridized to the same mRNA preparation.

Analysis of genomic DNA - The VR-1A insert and the MLP/Muc2 PCR fragment were used as probes for analyzing rat genomic DNA in Southern hybridizations. A comparison of the restriction patterns between the two probes is shown in Fig. 3 and summarized in Table I. No rat fragments were common to the two probes when rat genomic DNA cleaved with six restriction enzymes was analyzed. However, identical size hybridizing bands were found in pulsed field gel electrophoresis when fragments were generated by the two rare cutters, *Mlu*I (500 kb) and *Sal*I (380 kb). Southern analysis of mouse genomic DNA revealed two restriction fragments common to the two probes, one of 0.5 kb with *Bgl*II and one of >12 kb with *Eco*RI.

Chromosomal Localization in the Rat - The chromosomal assignment of VR-1A was done by using a rat x mouse somatic cell hybrid panel (22). The rat chromosomes present in each hybrid were

TABLE I
Lengths of rat genomic DNA restriction fragments (in kb) as probed with VR-1A and rat MLP/Muc2 after cutting with eight different restriction enzymes

Restriction enzyme	Rat DNA		Mouse DNA	
	VR-1A (kb)	MLP/Muc2 (kb)	VR-1A (kb)	MLP/Muc2 (kb)
*Bgl*II	4 2.5	12.0	8.3 3.0 0.5	 0.5
*Hinc*II	7.5	2.2	6.5	2.8
*Dra*I	>12	3.6	>12	3.1
*Eco*RI	8.5	4	>12	>12
*Bam*HI	8.5	9	12	5.8
*Hind*III	9	4.1	10	4.2
*Mlu*I	500	500	nd[a]	nd[a]
*Sal*I	380	380	nd	nd

[a]not determined.

determined by cytogenetic analysis in G-banded slides. Using *Bgl*II-restricted genomic DNA the rat-specific hybridizing fragments (2.5 and 4 kb) could be easily distinguished from the mouse fragments at 0.5, 3 and 8.3. The discordance analysis (Table II) unequivocally showed that both rat hybridizing fragments were localized on rat chromosome 1.

DISCUSSION

Expression cloning using an antiserum against HF-deglycosylated glycopeptide A from the 'insoluble' mucin complex of rat small intestine revealed a clone coding for an open reading frame, the deduced amino acid sequence of which shows several features of a mucin containing both a Cys-rich and a Ser/Thr-rich region. The fact that VR-1A codes for part of glycopeptide A is further substantiated by the predicted amino acid composition, especially the relative amounts of Ser, Thr and Pro, being similar to that of glycopeptide A. Moreover, the antiserum against HF-deglycosylated glycopeptide A reacts with goblet cells of intestine in mice[2]. The antiserum against HF-treated glycopeptide B did not stain the intestine and did not give any clones upon cDNA screening, suggesting that this serum did not contain any anti-apoprotein antibodies.

The 53 N-terminal amino acids of the VR-1A sequence contain seven Cys, one Ser and no Thr. The following region consists of 182 amino acids, is lacking Cys residues and is dominated by the three 'typical' mucin amino acids Ser/Thr/Pro. The Ser/Thr-rich region does not contain any conserved tandem repeats. However, several features of repeats can be found (Fig. 1) including, for example, 13 copies of the sequence TTT. The lack of typical tandem repeats is contradictory to the results by others

[2]Gendler, S., Carlstedt, I., and Hansson, G.C., in preparation.

TABLE II

Chromosomal assignment of the VR-1A clone by discordant analysis of rat x mouse hybrids

Cell hybrid	Rat chromosome*																					Hybridizing fragment	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	X	Y	4 kb	2.5 kb
LB20 D956	-	(+)	(-)	-	-	-	+	-	-	-	-	+	+	-	-	+	+	+	+	-	+	-	-	-
LB330, D906	-	(-)	+	+	-	+	-	-	-	+	-	+	-	-	-	(-)	(-)	-	(-)	-	+	-	-	-
LB510-5, D963	+	(+)	+	+	-	-	+	-	-	-	-	+	+	+	+	+	+	+	-	-	+	-	+	+
LB600, D991	-	-	-	+	-	-	-	-	+	-	+	-	+	+	+	+	-	+	+	-	+	-	-	-
LB780Z, D895	-	+	-	-	-	-	+	-	-	+	+	-	+	-	-	(+)	-	+	-	-	+	-	-	-
LB810, D945	-	+	+	+	-	+	+	+	-	+	+	+	+	+	+	+	+	-	+	+	+	-	-	-
BS511, D897	-	+	+	+	(-)	+	+	+	-	-	+	-	+	-	+	+	+	-	+	-	+	-	-	-
LB161, D950	-	+	+	+	(+)	+	+	-	+	+	-	+	(+)	+	+	+	+	+	+	-	+	-	-	-
LB1040TG5,D920	-	-	-	+	-	+	-	-	-	+	+	+	-	-	+	+	-	+	-	-	-	-	-	-
LB150-1, D901	-	-	+	+	-	-	+	-	+	(-)	+	(+)	+	-	-	(+)	(+)	+	(+)	-	+	-	-	-
LB150-2, D899	-	-	+	+	-	-	+	-	+	+	+	+	+	-	-	+	+	+	+	-	+	-	-	-
GB1, D900	(-)	(+)	-	+	(+)	+	+	-	+	+	+	+	+	+	+	+	-	+	+	-	+	-	+	+
LB210Z, D910	-	-	-	+	-	-	-	-	-	-	-	-	+	(-)	-	-	+	+	(-)	(-)	+	-	-	-
LB251, D909	-	-	-	+	-	-	-	-	-	+	-	-	+	-	-	(-)	+	-	+	-	+	-	-	-
LB260TG4, D939	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-
LB780-5, D944	-	+	-	-	-	-	-	-	-	+	+	-	-	-	-	-	-	-	-	-	+	-	-	-
Independent discordant	0	5	7	11	2	6	7	4	5	9	9	7	10	3	5	11	9	8	9	3	12	2		4 kb
clones	0	5	7	11	2	6	7	4	5	9	9	7	10	3	5	11	9	8	9	3	12	2		2.5 kb

*Designations:
+ Chromosomes present in more than 55% of the cells,
(+) Chromosomes present in 25-55% of the cells,
(-) Chromosomes present in less than 25% of the cells (not discorded in the discordancy analysis).
- Chromosomes absent.

(3-9), but can be explained by a tendency for rodents to have more degenerated repeats as suggested by a comparison of the mouse Muc-1 and the human MUC1 sequences (29).

A putative 'link' peptide, proposed to join mucus glycoprotein subunits into larger structures, has been inferred in several tissues including the small intestine (30). Using the partial peptide sequence of the 118 kDa peptide from rat small intestine, a cDNA was isolated coding for the carboxy terminal end of a large mucin-like protein, MLP (20). This sequence was later shown to have a high degree of homology with the C-terminal 800 amino acids of the human MUC2 (4,21) making MLP very likely to be the rat Muc2 mucin.

Analysis of the size of mRNA probed by the VR-1A in rat small intestine revealed a relatively sharp band about 12 kb in size (Fig. 2). This band was identical in size to the band stained by the MLP/Muc2 probe, suggesting that the two probes correspond to the same mRNA. There is no sequence overlap between VR-1A and the published partial sequences of the rat Muc2 (MLP) (20, 31), but there are general sequence motifs of the Ser/Thr-rich region of VR-1A that are similar. When the two probes were sequentially hybridized to filters obtained after PFGE of rat genomic DNA restricted with the rare-cutters *Mlu*I and *Sal*I, both probes were shown to hybridize to identical size fragments for each of the enzymes (Table I), supporting the notion that they correspond to the same gene. However, when the two probes were used in traditional Southern analysis they hybridized to different size fragments in rat DNA (Table I), one band for all enzymes except *Bgl*II that gave two due to a cleavage site in VR-1A. The MLP/Muc2 probe was first hybridized to the filters, which were subsequently deprobed before the VR-1A probe was applied. This excluded the possibility that the slight crosshybridization (Fig. 3) of the MLP/Muc2 probe to the fragments that in the subsequent VR-1A

```
VR-1A        1 LCSNVPLDQLGQKVVCNIEDGLVCKNAEQG 30
                ::.:||::|||||.|||::. ||:|||.:|
MUC2      1202 MYPDVPIGQLGQTVVCDVSVGLICKNEDQK 1231

VR-1A       31 IGGIIPMRMCLNYEINVYCCICIT 54
                 ||:||| :|||||||| || |:|
MUC2      1232 PGGVIPMAFCLNYEINVQCCECVT 1255
```

Fig. 4. Amino acid sequence similarity between the VR-1A clone and the human MUC2 sequence. The amino acid sequences are given with the one letter code having Cys residues marked in bold. The MUC2 sequence is from the N-terminal portion just before the large tandem repeat domain. The amino acid numbers are those used in Fig. 3 of ref. 3.

hybridization became the major hybridization band was a technical artefact. The reason for this weak staining is probably that the MLP/Muc2 probe contains a small part that is homologous to VR-1A. In fact, there is a stretch of 49 bp with a 64% identity in common for the two probes. This analysis showed that if the sequences corresponding to the two probes are adjacent in rat genomic DNA they must be separated by stretches containing restriction sites for each of the enzymes tested. The results for the mouse were similar, although in two cases the two probes hybridized to fragments of identical size. This may be fortuitous, but could also mean that sequences corresponding to the two probes are found closer together in the mouse than in the rat.

The VR-1A sequence was localized to the rat chromosome 1. During a project to localize other mucin genes in rat[2] it was found that a probe obtained by PCR based on the sequence information of the C-terminal part of the probable rat Muc2 gene (MLP) (20, 21) also localized to rat chromosome 1. The presence on the same chromosome, the identical size of mRNA, the common amino acid sequence motifs, and the identical genomic fragments in both rat and mouse strongly suggests that the VR-1A clone, encoding a portion of the glycopeptide A, is part of the same gene, most likely the rat Muc2 gene. This conclusion is further substantiated by comparing the sequence of VR-1A clone with that of the human MUC2 (4). As shown in Fig. 4 there is a 59% similarity within the VR-1A Cys-rich region and six out of seven Cys residues are located at the same positions. The homologous part of the MUC2 mucin is the Cys-rich region flanking the large tandem repeat domain at the N-terminal side, see Fig. 5. The published sequences of MLP/Muc2, start from the C-terminal end and go upstream into a Ser/Thr-rich region (20,31). The C-terminal part is homologous to human MUC2 as indicated in Fig. 5.

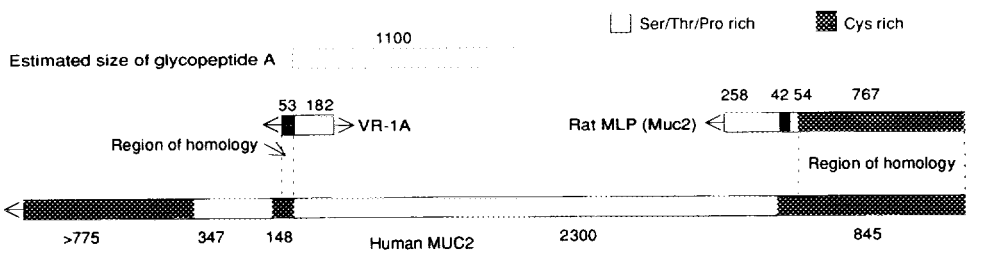

Fig. 5. Schematic picture showing a possible alignment of the amino acid sequence of human MUC2, VR-1A, rat MLP/Muc2 (mucin-like protein), and glycopeptide A. The number of amino acids in each Ser/Thr/Pro- and Cys-rich domain are given below or above the corresponding region. The arrows show that the sequence is unknown and extending in the direction indicated. The size of glycopeptide A is also indicated (14). The sequence of human MUC2 is from ref. 4 and of rat MLP/Muc2 from refs. 20 and 31.

The human MUC2 cDNA sequence, not yet fully completed, has revealed two Ser/Thr-rich domains (4), see Fig. 5. The deduced amino acid composition of the large tandem repeat domain is almost lacking Ser residues and is thus not similar to that of VR-1A. Instead, the deduced amino acid composition of VR-1A shows a relative abundance of Ser/Thr/Pro similar to the small 347 amino acid N-terminally located tandem repeat of MUC2. Searching the EMBL data base using the GCG-TFASTA algorithm with the Ser/Thr-rich part of VR-1A sequence as query gave the small repeat of MUC2 the highest score. The amino acid compositions of glycopeptide A and B (14) are similar, suggesting that both sequences resemble the MUC2 347-amino acid repeat. It is presently not known whether glycopeptide B is part of the same mucin as glycopeptide A, but is very likely (14). If the alignment of the different sequences of Fig. 5 is correct one might speculate that glycopeptide B is located on the N-terminal side of the VR-1A sequence.

Whereas the rat intestine yields two large trypsin-resistant glycopeptides (12, 14), the human only gives rise to one[3] (13) which could be the highly glycosylated large tandem repeat part of MUC2. The smaller repetitive region of the human MUC2 only contains 347 amino acids and may therefore be too small, even after glycosylation, to be detected with the separation protocols used. Thus, despite the difference in biochemical characteristics of rat and human 'insoluble' mucin complex, the available information is consistent with the MUC2 mucin being a major component also of the human 'insoluble' mucin.

The results presented here have shown that the gene encoding the apoprotein for the highly glycosylated mucin glycopeptide A of rat small intestine most likely is the rat Muc2. Glycopeptide A has recently been characterized in detail (14) which gives this mucin both a biochemical and a genetic identity. It can also be concluded that the 'insoluble' mucin complex of rat small intestine contains the Muc2 mucin at least as one of its major components. The published sequence of the human MUC2 (4), containing several Cys-rich regions and homologies to the von Willebrand factor in both the N- and C-terminal ends, admits the possibility of a mucin structure based on linear polymers. This is the general structure proposed for mucins of the lung, stomach, and cervix (2, 32), but these mucins are not 'insoluble' as the intestinal ones analyzed here. As almost no sequence information is available for respiratory, gastric and cervical mucins no conclusion on the differences in organization of these in comparison with the 'insoluble' intestinal ones can be drawn. However, the MUC2 sequence has, in addition to the von Willebrand factor homologies, a third Cys-rich region in its central part making additional disulphide bonds between subunits a possibility for its 'insolubility'. Although the present results do not provide any explanation for the 'insolubility' observed, they will focus our attention on the MUC2 mucin subunit, its properties and possibilities for multimerization when trying to understand this phenomenon.

Acknowledgments - We are indebted to Dr. Jim Gum for valuable advice, help with phage DNA purification and primer selection for PCR. The help of Dr. Salo Arielly, Chromogenix, Gothenburg, in doing the HF-treatment is gratefully acknowledged. The qualified technical assistance of Ms. Helena Alstermark and Malin Ek is appreciated. We thank Ms. Kerstin Montelius-Alatalo for her skilful assistance with the PFGE runs. This work was supported by the Swedish Medical Research Council

[3]Carlstedt, I., unpublished observation.

Grants 7461, 10226, and 7902; the Swedish Research Council for Engineering Sciences, the Swedish Cancer Society, IngaBritt and Arne Lundbergs Stiftelse, CANCIRCO, BioVäst, the Medical Faculty of Lund, and Gabrielsson's Research Fund.

REFERENCES

1. Neutra, M., and Forstner, J. F. (1987) in: Physiology of the Gastrointestinal Tract (Johnson, L.R., ed.) pp. 975-1009, Raven, New York.
2. Carlstedt, I., and Sheehan, J. K. (1984) CIBA Found. Symp. 109, 157-172.
3. Gendler, S. J., Lancaster, C. A., Taylor-Papadimitriou, J., Duhig, T., Peat, N., Burchell, J., Pemberton, L., Lalani, E.-N., and Wilson, D. (1990) J. Biol. Chem. 265, 15286-15293.
4. Gum, J. R., Hicks, J. W., Toribara, N. W., Rothe, E.-M., Lagace, R. E., and Kim, Y. S. (1992) J. Biol. Chem. 267, 21375-21383.
5. Gum, J. R., Hicks, J. W., Swallow, D. M., Lagace, R. L., Byrd, J. C., Lamport, D. T. A., Siddiki, B., and Young, Y. S. (1990) Biochem. Biophys. Res. Commun. 171, 407-415.
6. Porchet, N., Van Cong, N., Dufosse, J., Audie, J.P., Guyonnet-Duperat, V., Gross, M. S., Denis, C., Degand, P., Bernheim, A., and Aubert, J. P. (1991) Biochem. Biophys. Res. Commun. 175, 414-422.
7. Toribara, N. M., Roberton, A. M., Ho, S. B., Kuo, W.-L., Gum, E., Hicks, J. W., Gum, J. R., Byrd, J. C., Siddiki, B., and Kim, Y. S.(1993) J. Biol. Chem. 268, 5879-5885.
8. Eckhardt, A. E., Timpte, C. S., Abernaethy, J. L., Zhao, Y., and Hill, R. L. (1991) J. Biol. Chem. 266, 9678-9686.
9. Gum, J. R., Hicks, J. W., Lagace, R. E., Byrd, J. C., Toribara, N. W., Siddiki, B., Fearney, F. J., Lamport, D. T. A., and Kim, Y. S. (1991) J. Biol. Chem. 266, 22733-22738.
10. Hilkens, J., and Buijs, F. (1988) J. Biol. Chem. 263, 4215-4222.
11. Eckhardt, A. E., Timpte, C. S., Abernathy, J.L., Toumadje, A., Johnson, W. C., and Hill, R. L. (1987) J. Biol. Chem. 262, 11339-11344.
12. Carlstedt, I., Karlsson, H., Sundler, F., and Fransson, L.-Å. (1982) Adv. Exp. Med. Biol. 144, 155-157.
13. Mårtensson, S., Lundblad, A., Hansson, G. C., and Carlstedt, I. (1988) Scand. J. Clin. Invest. 48, 633-640.
14. Carlstedt, I. , Herrmann, A., Karlsson, H., Sheehan, J., Fransson, L.-Å., and Hansson, G. C. (1993) J. Biol. Chem. 268, 18771-18781.
15. Byrd, J. C., Lamport, D. T. A., Siddiqui, B., Kuan, S.-F., Erickson, R., Itzkowitz, S. H., and Kim, Y. S. (1989) Biochem. J. 261, 617-625.
16. Baeckström, D., Hansson, G. C., Nilsson, O., Johansson, C., Gendler, S. J., and Lindholm, L. (1991) J. Biol. Chem. 266, 21537-21547.
17. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
18. MacDonald, R. J., Swift, G. H., Przybyla, A. E., and Chirgwin, J. M. (1987) Metods Enzymol. 152, 219-227.
19. Baeckström, D., Nilsson, O., Price, M. R., Lindholm, L., and Hansson G. C. (1993) Cancer Res. 53, 755-761.
20. Xu, G., Huan, L.-J., Khatri, I. A., Wang, D., Bennick, A., Fahim, R. E. F., Forstner, G. G., and Forstner, J. F. (1992) J. Biol. Chem. 267, 5401-5407.
21. Xu, G., Huan, L., Khatri, I., Sajjan, U. S, McCool, D., Wang, D., Jones, C., Forstner, G., and Forstner, J. (1992) Biochem. Biophys. Res. Commun. 183, 821-828.
22. Szpirer, J., Levan, G., Thörn, M., and Szpirer, C. (1984) Cytogenet. Cell Genet. 38, 142-149.
23. Szpirer, C., Rivière, M., Szpirer, J., Genet, M., Drèze, P., Islam, M. Q. and Levan, G. (1990) Genomics 6, 679-684.
24. Szpirer, C., and Szpirer, J. (1980) Exp. Cell Res. 125, 305-312.
25. Littlefield, J. W. (1964) Science 145, 709-710.
26. Blin, N., and Stafford, D. W. (1976) Nucleic Acid Res. 3, 2303-2308.
27. Southern, E. M. (1975) J. Mol. Biol. 98, 503-517.
28. Ståhl, F., Wettergren, Y., and Levan, G. (1993) Hereditas 118, 121-130.
29. Spicer, A. P., Parry, G., Patton, S., and Gendler, S. J. (1991) J. Biol. Chem. 266, 15099-15109.
30. Mantle, M., Forstner, G. G., and Forstner, J. F. (1984) Biochem. J. 224, 345-354.
31. Huan, L. J., Xu, G., Forstner, G., and Forstner, J. (1992) Biochim. Biophys. Acta 1132, 79-82.
32. Carlstedt, I. and Sheehan, J. K. (1984) Biochem. Soc. Trans. 12, 615-617.