Molecular Cloning of Human Intestinal Mucin cDNAs

SEQUENCE ANALYSIS AND EVIDENCE FOR GENETIC POLYMORPHISM*

(Received for publication, August 8, 1986)

James R. Gunz‡, James C. Byrd‡, James W. Hicks‡, Neil W. Toribara‡‡, Derek T. A. Lamport¶, and
Young S. Kim‡‡
From the 3Gastrointestinal Research Laboratory (151M2), Veterans Administration Medical Center, San Francisco,
California 94121, the Department of Medicine, University of California, San Francisco, California 94143, and the 4Department
of Energy Plant Research Laboratory and Department of Biochemistry, Michigan State University, East Lansing, Michigan 48824

A human small intestine $\lambda 11$ cDNA library was screened using antisera prepared against the deglycosylated
protein backbone of human colon cancer xenograft mucin. Three cDNAs were isolated from this
screening, designated SMUC 40–42. These cDNAs were all found to contain tandem repeats of 69
nucleotides which encoded a threonine- and proline-rich protein consensus sequence of
PTTPTPTTTTTVPTPTPTGTQT. RNA blots probed with one of these cDNAs, SMUC 41, exhibited large,
polydisperse hybridization bands at ~7,600 bases. Band intensities were strongest when human small
intestine, colon, and colon cancer poly(A)$^+$ RNA was used. In vitro translation of poly(A)$^+$ RNA from human
small intestine, colon, and colon cancer cells produced a 162,000-dalton peptide that was immunoprecipitated
with antibodies to deglycosylated mucin. SMUC 41 was also used to probe DNA blots, which indicated
the presence of restriction fragment length polymorphisms in the intestinal mucin gene. These findings
may be important in assessing the abnormal mucins found associated with several human diseases.

Human intestinal mucin is a viscous gel that lubricates and protects the delicate epithelium of the digestive tract. This
substance derives its characteristic fluid mechanical properties from its content of mucins, which are large glycoproteins
($M_r > 250,000$) consisting of ~75% carbohydrate, ~20% protein, and trace quantities of other compounds (1–4). The
oligosaccharides that account for most of the mass of mucins are heterogeneous and frequently branched, consisting of as
many as 20 individual sugar residues/chain. Mucin oligosaccharides are bound to serine and threonine residues in the
protein backbone by a terminal GalNAc residue. The protein backbone itself appears to be covered with these $O$-linked
oligosaccharides as only ~10% of it is susceptible to proteolysis (1–3). Although the oligosaccharide moieties of normal
colic mucin have been recently characterized (5, 6), little is known about the structure and amino acid sequence of the protein
core of this high molecular weight glycoconjugate.

Several human diseases have been observed to be associated with alterations in intestinal mucins. These include cystic
fibrosis, familial polyposis coli, ulcerative colitis, and colon cancer. Patients with cystic fibrosis produce excessive
amounts of mucin in their gastrointestinal, respiratory, and reproductive tracts whereas patients with familial polyposis
colii, ulcerative colitis, and colon cancer produce mucins that are abnormally glycosylated (2–4, 7, 8). Hence, a better
understanding of the molecular genetics and biosynthesis of mucin may provide insight into the pathogenesis, diagnosis,
and treatment of several important human diseases.

In order to examine further the structure, biosynthesis, and genetics of intestinal mucin, we sought to clone cDNAs that
encode the mucin protein backbone. This was achieved in the present study using antibodies to deglycosylated colon cancer
xenograft mucin and a small intestine $\lambda 11$ expression library. The resulting cDNAs indicate that this mucin contains
threonine- and proline-rich regions consisting of tandem repeats of 23 amino acids each. Furthermore, these cDNAs
enabled us to identify the mucin message produced in various cell lines and tissues and to determine that the intestinal
mucin gene is genetically polymorphic.

MATERIALS AND METHODS

Purification of Mucin and Production of Antibodies to Deglycosylated Mucin—Mucin was purified from LS174T human colon cancer
cell tumors (grown in nude mice) using gel filtration and CsCl density gradient centrifugation. This mucin had an amino acid
composition that was 29% threonine, 14% serine, and 15% proline, similar to that found previously for human intestinal mucin (1–3). Details of the
LS174T cell mucin purification and characterization are published elsewhere (9). The purified mucin was deglycosylated by treat-
ment with hydrogen fluoride under anhydrous conditions for 1 h at room temperature (to give HFA)$^1$ or 3 h at room temperature (to give HFB) (10).
Compositional analysis indicated that almost all (~98%) of the sugar had been removed from HFB but that HFA still contained ~16% of its
original content of GalNAc and ~75% of its GaINAc content. Antibodies were prepared in New Zealand White rabbits against HFA,
HFB, or native mucin using three or four subcutaneous injections of 50–100 $\mu$g of antigen. Enzyme-linked immunosorbent assays indici-
cated that all immunogens elicited antibodies (10).

Library Screening and Lysogen Preparation—A human jejunal cDNA library constructed in the $\lambda 11$ expression vector was obtained
from Dr. Yvonne Edwards (Medical Research Council, Human Bio-

* This work was supported by the Veterans Administration Medical Research Service, by National Cancer Institute Grant CA47551 (to
Y. S. K.), and by Department of Energy contract DE-AC 02-76 ER
01 538 (to D. T. A. L.). 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.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank*®/EMBL Data Bank with accession number(s)
J06458.

‡ Associate Investigator of the Veterans Administration.
¶ Medical Investigator of the Veterans Administration. To whom
reprint requests should be addressed.

1 The abbreviations used are: HFA and HFB, preparations of
LS174T xenograft mucin deglycosylated with hydrogen fluoride as
described in the text; MRP, a synthetic peptide with the mucin repeat
sequence; BSA, bovine serum albumin; bp, base pairs; kb, kilobase
pairs; SSC, standard saline citrate; pfu, plaque-forming unit.

6480
chemical Genetics Unit, University College London, London, United Kingdom) (11). This library was plated in soft agar at a density of 25,000 plaques/150-mm plate as described (12). The plates were incubated at 37 °C until plaques began to appear and were then overlaid with isopropyl β-D-thiogalactopyranoside-saturated nitrocellulose membranes and incubated for an additional 5 h. The membranes were then removed and immunoscreened using anti-HFB serum at a 1:50 dilution and horseradish peroxidase-conjugated goat anti-rabbit IgG (Tago) using previously described methods (13). Positives were purified to clonality by successive rounds of rescreening, phage DNA was isolated, and inserts were recovered by EcoRI digestion. Lysogenation of Escherichia coli strain Y1089(r−) (Promega overlayed with isopropyl β-D-thiogalactopyranoside-saturated nitrocellulose membranes were then removed and immunoscreened using anti-HFB polymerase (United States Biochemical Corp.) (14). For each cDNA, both strands were sequenced in their entirety. DNA sequences were assembled and analyzed using DNA and Protein Sequence Analysis software purchased from International Biotechnologies, Inc.

**DNA, RNA, and Protein Blot Analysis and in Vitro Translation**—RNA purification and poly(A)+ RNA isolation, gel electrophoresis, transfer to nylon membranes, and hybridization probe analysis was conducted as described (15). Protein immunoblots were performed using a 1:50 dilution of antibody (15). High molecular weight DNA was prepared using proteinase K, RNase A, and phenol as described (16). This material was digested with restriction enzymes, and the fragments were separated by electrophoresis in 1% agarose gels using a buffer containing 100 mM Tris, 100 mM boric acid, and 2 mM EDTA, pH 8.0 (17). The gels were soaked for 30 min in 1.5 M NaCl, 0.5 M NaOH, and then in 3 M NaOAc, pH 5.5 for the same period of time. Transfer to nylon membranes, hybridization, and washing then proceeded as described above for RNA blots. In vitro translations and immunoprecipitations were performed as described (15) except that (a) [35S]cysteine (Amersham Corp.) was used as the radioactive amino acid in conjunction with a cysteine-free translation mixture, (b) 0.45 μg of poly(A)+ RNA was used in 39 final reaction volumes, (c) the carrier lysate for immunoprecipitation also contained 10 mg/ml of added HFB, and (d) 5 μl of control serum or 3 μl of anti-HFB serum was used.

**Synthetic Peptide (MRP) and Antibody Preparation**—A peptide with the sequence KYPITTPTSTTTMTPTPTGQT was prepared using an Applied Biosystems model 430A peptide synthesizer by Joel Boymel of the National Jewish Center for Immunology and Respiratory Medicine, Denver, CO. The final 23 residues of this peptide represent the sequence of the first repeat of SMUC 40; the initial K and Y residues were added to allow glutaraldehyde conjugation and radioiodination (for future studies), respectively. For antibody production, 1 mg of peptide was emulsified in complete Freund’s adjuvant and injected intradermally at multiple sites into a female New Zealand rabbit (18). Three weeks later a second set of injections using 0.5 mg of peptide in incomplete adjuvant was administered, and the rabbit was bled 12 days later and serum prepared.

**RESULTS**

**Isolation of Intestinal Mucin cDNAs**—Because the protein backbone of intestinal mucin is so heavily laden with oligosaccharide chains it is difficult to characterize biochemically. The conditions required to remove the carbohydrate result in breakage of the protein backbone. Thus, it is impractical to obtain information pertaining to the primary structure of intestinal mucin by conventional peptide sequencing. In order to acquire this structural information, we therefore decided to clone and sequence intestinal mucin cDNAs.

Antibodies prepared against HFB were used to screen the intestinal cDNA library and three positives were obtained from a screening of 230,000 recombinant plaques. These clones, which were designated SMUC 40, SMUC 41, and SMUC 42, were purified and tested for antigenicity using anti-HFA, anti-HFB, and anti-native mucin as shown in Fig. 1. Only antisera against the completely deglycosylated HFB produced positive plaques in this experiment. Fig. 2 shows immunoblot analysis of the β-galactosidase fusion proteins produced by lysogens of these recombinants. Anti-HFB reacts strongly with the fusion proteins produced by SMUC 40–42. These experiments indicate that these clones produce recombinant fusion proteins that are recognized by antisera against deglycosylated mucin but not by antisera against native mucin. Thus, the fusion proteins apparently contain epitopes that do not function as immunogens when mucin is injected into rabbits unless the mucin is first deglycosylated, providing evidence that these cDNAs encode the normally covered mucin protein backbone.

**Sequence Analysis of Mucin cDNAs**—The recombinant phage DNA was digested with EcoRI, and each clone was found to contain a single, unique insert. Sequence analysis of the terminal regions of these clones indicated that they all contained repetitive sequences. Exonuclease III was then used to generate partially deleted clones for sequence analysis of the interior regions of these cDNAs (19). This made it possible to correlate the region of the cDNA sequenced from each template with the length of the deletion, information necessary to avoid confusion caused by not knowing which repeat unit was being sequenced (20). Details of the sequencing strategy used are given in Fig. 3.

Each of these clones was found to contain tandem repeats of 69 nucleotides (Fig. 4). In fact, only the 5′-terminal 71 nucleotides of SMUC 42 and the 3′-terminal 471 nucleotides of SMUC 41 can be clearly identified as not consisting of these repeat units. Thus, anti-HFB serum appears to be
Fig. 2. Immunoblot of lysates prepared from lysogens of SMUC 40-42. Lysates containing 100 µg of lysogen protein were subjected to sodium dodecyl sulfate-polyacrylamide gel electrophoresis and transferred electrophoretically to nitrocellulose. Immunoblot analysis was conducted using a 1:50 dilution of anti-HFB. Lanes 40, 41, and 42 are from lysogens of SMUC 40–42, respectively. Lane C is a nonrecombinant λgt11 lysogen control.

Fig. 3. Sequencing strategy for clones SMUC 40–42. The arrows represent the length and direction of individual sequencing reactions. The terminal regions of SMUC 40–42 were sequenced using templates derived from vectors containing each clone in its entirety. Most of the sequencing of the interior regions of SMUC 40 and 41 was performed using exonuclease III-deleted clones. In a few cases, restriction fragments obtained from TqI and MspI digests of SMUC 40 and 42 were force cloned into Accl- and EcoRI-digested M13mp18 and used to generate templates. Sequencing done using this latter method is indicated using dashed arrows.

strongly immunoreactive with the protein encoded by the 69-bp repetitive element. The amino acid sequences deduced for each of these tandem repeats are shown in Fig. 5. The 23-amino acid consensus sequence of these repeat units contains 14 threonine and 5 proline residues, including a group of five consecutive threonines and a stretch containing three threonine-proline direct repeats. The 14 repetitive units contained in the three partial cDNA clones isolated in this study have 90% overall sequence identity with the consensus sequence shown in Fig. 5. Even more conserved is the 12-amino acid stretch enclosed in the box in Fig. 5, which exhibits 98% overall sequence identity with the consensus sequence. Only 11 serine residues are found dispersed among these 14 tandem repeats and nine of them occur as substitutions for threonine in the consensus sequence. On the other hand, the carboxyterminal 187-amino acid region deduced from the 3'-terminal 471 nucleotides of clone SMUC 41 (which does not consist of the tandem repeats) contains 25 serine residues. Hence, it appears that the majority of serine residues in intestinal mucin are clustered in regions other than the tandem repeats. The 3'-terminal region of SMUC 41 also contains the only cysteine, present as a cystycys dipeptide, and most of the aromatic amino acids. Two potential N-glycosylation recognition sites are encoded in the sequences presented here, one in the last repeat unit of SMUC 40 and one near the 3'-terminal of SMUC 41.

Reactivity of Antibodies against the MRP with HFB—As shown in panel A of Fig. 6, antibodies against HFB reacted with both HFB and BSA conjugated with the MRP but not with partially deglycosylated mucin (HFA) or unconjugated BSA. The broad smear of antibody reactive protein in the HFB sample is indicative of the cleavage of the mucin backbone that occurs during deglycosylation (10). MRP-conjugated BSA exhibits polydispersity, on the other hand, due to irregular conjugation of BSA with itself and the peptide. Antibodies against the MRP had a specificity similar to anti-HFB. Again, reactivity was apparent with HFB and MRP-conjugated BSA but not with HFA or unconjugated BSA (Fig. 6, panel B). Thus, antibodies prepared against a synthetic peptide made using the deduced sequence of a mucin repeat unit were reactive with HFB, providing additional evidence that these cDNAs are actually derived from mucin messages.

RNA Blot Analysis and in Vitro Translation of Mucin mRNAs—Poly(A)+ RNA from a number of human cell lines and tissues was subjected to RNA blot analysis using SMUC 41 cDNA as a probe (Fig. 7). The messages that hybridized to SMUC 41 were large and polydisperse, averaging 7600 bases in length. In addition, a distinct but faint band at 1850 bases was sometimes detectable (Fig. 10). The strongest hybridization signals observed in these experiments were expressed by colon, colon tumor, and small intestine RNA. HM-7 and H498, two high mucin-producing human colon cancer cell lines (22, 23), also contained high levels of message. LM-12, a low mucin-producing variant of LS174T cells (22), exhibits only a faint hybridization signal as does RNA from LS-G and SW1116 cells. No detectable signal was obtained with either placenta or the thyroid tumor poly(A)+ RNA used here.

In vitro translation and immunoprecipitation with anti-HFB serum was used in an attempt to identify the intestinal mucin primary translation product (Fig. 8). When the in vitro translation reactions were programmed with poly(A)+ RNA from human small intestine, colon, H498 cells, or HM-7 cells a single discernible protein of 162,000 daltons was specifically immunoprecipitated. This band was fainter but detectable when LM-12 poly(A)+ RNA was used to program the reactions.

**Human Intestinal Mucin**

![Image](image_url)

**Fig. 2.** Immunoblot of lysates prepared from lysogens of SMUC 40-42. Lysates containing 100 µg of lysogen protein were subjected to sodium dodecyl sulfate-polyacrylamide gel electrophoresis and transferred electrophoretically to nitrocellulose. Immunoblot analysis was conducted using a 1:50 dilution of anti-HFB. Lanes 40, 41, and 42 are from lysogens of SMUC 40–42, respectively. Lane C is a nonrecombinant λgt11 lysogen control.

**Fig. 3.** Sequencing strategy for clones SMUC 40–42. The arrows represent the length and direction of individual sequencing reactions. The terminal regions of SMUC 40–42 were sequenced using templates derived from vectors containing each clone in its entirety. Most of the sequencing of the interior regions of SMUC 40 and 41 was performed using exonuclease III-deleted clones. In a few cases, restriction fragments obtained from TqI and MspI digests of SMUC 40 and 42 were force cloned into Accl- and EcoRI-digested M13mp18 and used to generate templates. Sequencing done using this latter method is indicated using dashed arrows.

strongly immunoreactive with the protein encoded by the 69-bp repetitive element. The amino acid sequences deduced for each of these tandem repeats are shown in Fig. 5. The 23-amino acid consensus sequence of these repeat units contains 14 threonine and 5 proline residues, including a group of five consecutive threonines and a stretch containing three threonine-proline direct repeats. The 14 repetitive units contained in the three partial cDNA clones isolated in this study have 90% overall sequence identity with the consensus sequence shown in Fig. 5. Even more conserved is the 12-amino acid stretch enclosed in the box in Fig. 5, which exhibits 98% overall sequence identity with the consensus sequence. Only 11 serine residues are found dispersed among these 14 tandem repeats and nine of them occur as substitutions for threonine in the consensus sequence. On the other hand, the carboxyterminal 187-amino acid region deduced from the 3'-terminal 471 nucleotides of clone SMUC 41 (which does not consist of the tandem repeats) contains 25 serine residues. Hence, it appears that the majority of serine residues in intestinal mucin are clustered in regions other than the tandem repeats. The 3'-terminal region of SMUC 41 also contains the only cysteine, present as a cystycys dipeptide, and most of the aromatic amino acids. Two potential N-glycosylation recognition sites are encoded in the sequences presented here, one in the last repeat unit of SMUC 40 and one near the 3'-terminal of SMUC 41.

Reactivity of Antibodies against the MRP with HFB—As shown in panel A of Fig. 6, antibodies against HFB reacted with both HFB and BSA conjugated with the MRP but not with partially deglycosylated mucin (HFA) or unconjugated BSA. The broad smear of antibody reactive protein in the HFB sample is indicative of the cleavage of the mucin backbone that occurs during deglycosylation (10). MRP-conjugated BSA exhibits polydispersity, on the other hand, due to irregular conjugation of BSA with itself and the peptide. Antibodies against the MRP had a specificity similar to anti-HFB. Again, reactivity was apparent with HFB and MRP-conjugated BSA but not with HFA or unconjugated BSA (Fig. 6, panel B). Thus, antibodies prepared against a synthetic peptide made using the deduced sequence of a mucin repeat unit were reactive with HFB, providing additional evidence that these cDNAs are actually derived from mucin messages.

RNA Blot Analysis and in Vitro Translation of Mucin mRNAs—Poly(A)+ RNA from a number of human cell lines and tissues was subjected to RNA blot analysis using SMUC 41 cDNA as a probe (Fig. 7). The messages that hybridized to SMUC 41 were large and polydisperse, averaging 7600 bases in length. In addition, a distinct but faint band at 1850 bases was sometimes detectable (Fig. 10). The strongest hybridization signals observed in these experiments were expressed by colon, colon tumor, and small intestine RNA. HM-7 and H498, two high mucin-producing human colon cancer cell lines (22, 23), also contained high levels of message. LM-12, a low mucin-producing variant of LS174T cells (22), exhibits only a faint hybridization signal as does RNA from LS-G and SW1116 cells. No detectable signal was obtained with either placenta or the thyroid tumor poly(A)+ RNA used here.

In vitro translation and immunoprecipitation with anti-HFB serum was used in an attempt to identify the intestinal mucin primary translation product (Fig. 8). When the in vitro translation reactions were programmed with poly(A)+ RNA from human small intestine, colon, H498 cells, or HM-7 cells a single discernible protein of 162,000 daltons was specifically immunoprecipitated. This band was fainter but detectable when LM-12 poly(A)+ RNA was used to program the reactions.
and was absent when LS-G poly(A)* RNA was used. A protein of 162,000 daltons would require an mRNA of approximately 5,000 bases or larger, depending on the length of the 5'- and 3'-untranslated regions. Hence, the molecular weight of the immunoprecipitated protein is in good agreement with the message size determined in Fig. 7.

Genomic DNA Blot Analysis—Genomic DNA was isolated from the lymphocytes of two human donors and five colon cancer cell lines, restriction endonuclease-digested, and subjected to electrophoresis and hybridization blot analysis using the SMUC 41 probe (Fig. 9). Six of these DNA samples were cleaved with EcoRI and all exhibited a single hybridization band that was larger than the 23.1-kb standard (Fig. 9A). This demonstrates that a polymorphism exists in or around the gene that encodes SMUC 41. Further evidence for polymorphism in this gene is shown in Fig. 9C. Sau3A digestion of these DNA samples revealed a different set of hybridization bands for three of the four samples tested. Thus, both HinfI and Sau3A identify

**Fig. 4.** Nucleotide and deduced amino acid sequence of intestinal mucin cDNA clones. Nucleotide position is indicated by the numbers at the right and amino acid position by those at the left. Asterisks appear every 10 nucleotides. The repeat units are indicated by the arrows and two putative N-glycosylation sites are underlined.
**Human Intestinal Mucin**

**FIG. 5.** Repetitive sequences in small intestine mucin cDNA clones. The numbers indicate the amino acid residues at the beginning and end of each tandem repeat. Lowercase letters indicate differences from the consensus sequence. The amino acids enclosed in the box are especially conserved, as described in the text.

**FIG. 6.** Immunoblots using anti-HFB and anti-MRP. Blots were prepared containing 5 μg/lane of HFA, HFB, BSA, or MRP-conjugated BSA and probed using anti-HFB (panel A) or anti-MRP (panel B). MRP was conjugated to BSA using glutaraldehyde as described by Avrameas (21).

**FIG. 7.** RNA blot analysis. Poly(A)⁺ RNA samples (0.5 μg) were analyzed as described under “Materials and Methods” using cDNA clone SMUC 41 as a probe. LS-G is a substrain of colon cancer cell line LS174T which contains high levels of butyrate-inducible alkaline phosphatase (15). LM-12 and HM-7 are LS174T cell variants which contain a low and high content of mucin, respectively (22). SW1116 is a colon cancer cell line which, like LS-G, is uncharacterized in terms of its mucin content (24). H498 is a recently isolated colon cancer cell line which produces and secretes high levels of mucin (23). The placenta, small intestine, and colon samples used here were from normal individuals. S. J. tumor and colon were from tissue surgically removed from a patient with colon cancer. The last lane contained poly(A)⁺ RNA isolated from a thyroid tumor. The 28 S and 18 S rRNA subunits were used as size markers (5400 and 2100 bases, respectively).

**FIG. 8.** In vitro translation and immunoprecipitation of the putative mucin primary translation product. Poly(A)⁺ RNA was translated in rabbit reticulocyte lysates and immunoprecipitated with anti-HFB as described under “Materials and Methods.” Controls performed with serum from a nonimmunized rabbit are indicated with (−) and immunoprecipitations by (+). The arrowhead marks the location of the only specifically immunoprecipitated band at Mr = 162,000.
Fig. 9. Genomic DNA blots. Genomic DNA blots were prepared and analyzed using SMUC 41 as a probe. Panel A, DNA prepared from the lymphocytes of two healthy donors (PS and VE) and from four colon cancer cell lines was digested with EcoRI prior to electrophoresis. Panels B and C, DNA prepared from colon cancer cell lines was digested with HinfI and Sau3A, respectively. Eight µg of DNA were used per lane.

Discussion

In the present study we used antibodies prepared against the deglycosylated protein core of mucin to identify mucin cDNA clones in a human small intestine λgt11 library. This approach has proven successful in the past for the isolation of cDNAs encoding the mucins expressed by porcine submaxillary gland (20) and human mammary epithelium (26, 27). It is reasonable to expect, therefore, that this method may be generally applicable to the cloning of cDNAs for additional mucins or other heavily glycosylated proteins.

All three of the cDNA clones isolated in this study contain 69-bp tandem repeats. The deduced consensus amino acid sequence for these repeats are high in threonine and proline and low in serine content. In this respect, the repeat units are more similar to the acidic fraction of intestinal mucin isolated by Wesley et al. (3) than to the more abundant neutral fraction. The significance of this is not yet clear, however, as cDNAs representing the entire mucin protein backbone have not yet been isolated. Since the number of GalNAc residues in intestinal mucin approaches the number of threonine residues, it is very likely that a sizable percentage of the threonine present in the tandem repeats contains O-linked carbohydrate (1-3). Furthermore, studies of UDP-GalNAc:polypeptide GalNAc transferase have indicated a preference for one or more proline residues in the vicinity of an O-glycosylation site (28, 29), thus further supporting the conclusion that the deduced tandem repeats are O-glycosylated. Synthetic peptide acceptors modeled after the consensus sequence of the repeating units may be useful in defining the O-glycosylation sites more precisely.

The deduced amino acid sequences described here contain two potential N-glycosylation sites. This was not expected in view of previously published carbohydrate analyses of human intestinal mucin which reveal its major glycoprotein component to be devoid of mannose (2). Other mucins, however, have been shown to contain detectable quantities of N-linked carbohydrate. These include mouse submandibular mucin which has both N- and O-glycosidically bound carbohydrate chains present (30, 31). Also, the nascent protein precursor of human mammary mucin has recently been shown to undergo N-glycosylation (32, 33), and the deduced amino acid sequence for porcine submaxillary gland mucin contains an N-glycosylation site (at amino acid position 418, Ref. 20). Thus, it is not unprecedented for mucin-type glycoproteins to contain some N-linked oligosaccharide chains in addition to their O-linked glycans. The significance of the N-linked oligosaccharides is currently unclear. Since N-glycosylation occurs cotranslationally it is possible that the N-glycans stabilize the nascent apomucin prior to O-glycosylation or that they play a role in the intracellular targeting of this macromolecule. On the other hand, the N-linked glycans may be important in the functioning of mature mucin. Further studies are needed to define the role of the N-linked carbohydrate chains in mucin.
As mentioned above, cDNAs for porcine submaxillary gland mucin and human mammary mucin have been recently isolated (20, 26, 27). Comparison of the nucleotide sequence of the clones isolated in this study to the sequences of these other types of mucin failed to reveal any significant homology. This is not surprising as these three mucins have substantially different amino acid compositions (1-3, 34, 35). A similarity exists, however, in the sense that all three of these mucins contain tandem repeats. Porcine submaxillary gland mucin contains at least eight tandem repeats of 243 bp each, and human mammary mucin contains an unknown number of repeats of 60 bp each (20, 26, 27). This suggests that selective evolutionary pressure exerted on several types of mucin genes has favored events such as gene duplications or unequal crossovers which have yielded tandem repeats.

Another similarity that exists, at least between mammary mucin and intestinal mucin, is that both are encoded by genes that are genetically polymorphic (Fig. 9, see Refs. 26, 36, 37). In the case of mammary mucin, a length polymorphism is observed; i.e. different alleles are thought to have a variable number of repeat units which leads to differences in the size of the mucin protein expressed (37). At the present time, we do not have sufficient data to determine whether the intestinal mucin gene also exhibits length polymorphism or if the different restriction fragments observed are due to differences in the sequence (in either introns or exons). In any event, this polymorphism may contribute to the polydispersity observed with regard to intestinal mucin message size (Fig. 7). Furthermore, these polymorphisms could reflect significant differences in the mucin gene-coding sequence and/or in regions of the gene that effect the level of its expression. The isolation and sequencing of genomic DNA clones containing different restriction fragment length polymorphism-defined mucin alleles may reveal variations in intestinal mucin structure and expression that occur within the human population.

Acknowledgments—We express our appreciation to Nona Baker and Dr. Rajvir Dahiya for some of the poly(A)⁺ RNA samples used, to E. P. Muldoon for performing the mucin deglycosylations, to Rita Burns for assistance in preparation of the manuscript, and to Dr. Don M. Carlson for helpful discussions.

REFERENCES
1. Mantle, M., Forstner, G. O., and Forstner, J. F. (1984) Biochem. Biophys. Res. Commun. 127, 159-167
2. Mantle, M., Forsner, G. O., and Forstner, J. F. (1984) Biochem. J. 224, 345-354
3. Wessely, A., Mantle, M., Man, D., Qureshi, R., Forstner, G., and Forstner, J. (1985) J. Biol. Chem. 260, 7955-7959
4. Neutra, M., and Forstner, J. F. (1987) in Physiology of the Digestive Tract (Johnson, L. F., ed) Second Ed., pp. 975-1009, Raven Press, New York
5. Podolsky, D. K. (1985) J. Biol. Chem. 260, 8262-8271
6. Podolsky, D. K. (1985) J. Biol. Chem. 260, 15510-15515
7. Boland, C. R., Montgomery, C. K., and Kim, Y. S. (1982) Proc. Natl. Acad. Sci. U. S. A. 79, 2051-2055
8. Podolsky, D. K., and Isselbacher, K. J. (1984) Gastroenterology 87, 991-996
9. Nardelli, J., Byrd, J. C., Ho, J., Fearney, F. J., and Kim, Y. S. (1988) Pancreas 3, 631-641
10. Byrd, J. C., Lamport, D. T. A., Siddiqui, B., Kuan, S. F., Erickson, R., Itzkowitz, S. H., and Kim, Y. S. (1989) Biochem. J., in press
11. Green, F., Edwards, Y., Hauri, H., Povey, S., Ho, M. W., Pinto, M., and Swallow, D. (1987) Gene (Amst.) 57, 101-110
12. Huynh, T., Young, R. A., and Davis, R. W. (1985) in DNA Cloning Techniques: A Practical Approach (Glover, D., ed) pp. 49-78, IRL Press, Oxford
13. Young, R. A., and Davis, R. W. (1985) in Genetic Engineering (Setlow, J., and Hollaender, A., eds) Vol. 5, pp. 29-41, Plenum Publishing Corp., New York

Fig. 10. RNA and DNA blots probed using 5' and 3' segments of SMUC 41. Blots prepared as described above were probed with 3'-460 as discussed in the text and autoradiography was conducted. The probe was then removed by twice incubating the blots at 65 °C in 10 mM sodium phosphate, pH 6.5, and 50% formamide for 1 h. Following a rinse in 2 x SSC, and 0.1% sodium dodecyl sulfate the blots were probed again using 5'–370. Panel A, blot was prepared using 0.5 μg of poly(A)⁺ RNA from human small intestine and colon. Panel B, 8 μg of DNA samples digested with the indicated restriction enzymes were analyzed.
14. Tabor, S., and Richardson, C. C. (1987) *Proc. Natl. Acad. Sci. U. S. A.* **84**, 4767–4771
15. Gum, J. R., Kam, W. K., Byrd, J. C., Hicks, J. W., Sleisenger, M. H., and Kim, Y. S. (1987) *J. Biol. Chem.* **262**, 1092–1097
16. Blin, N., and Stafford, D. W. (1976) *Nucleic Acids Res.* **3**, 2303–2314
17. Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York
18. Vaitukaitus, J. L. (1981) *Methods Enzymol.* **73**, 46–52
19. Heinikoff, S. (1984) *Gene (Amst.*) **28**, 351–359
20. Timpte, C. S., Eckhardt, A. E., Abernethy, J. L., and Hill, R. L. (1988) *J. Biol. Chem.* **263**, 1081–1088
21. Avrameas, S. (1969) *Immunocytochemistry* **6**, 43–52
22. Kuan, S. F., Byrd, J. C., Basbaum, C. B., and Kim, Y. S. (1987) *Cancer Res.* **47**, 5716–5724
23. Park, J. G., Oie, H. K., Sugarbaker, P. H., Henslee, J. G., Chen, T. R., Johnson, B. E., and Gazdar, A. (1987) *Cancer Res.* **47**, 6710–6718
24. Leibovitz, A., Stinson, J. C., McCombs, W. B., McCoy, C. E., Mazur, K. C., and Mabrey, N. D. (1976) *Cancer Res.* **36**, 4562–4569
25. Thompson, J. A., Pande, H., Paxton, R. J., Shively, L., Padma, A., Simmer, R. L., Todd, C. W., Riggs, A. D., and Shively, J. E. (1987) *Proc. Natl. Acad. Sci. U. S. A.* **84**, 2965–2969
26. Gendler, S. J., Burchell, J. M., Duhig, T., Lamport, D., White, R., Parker, M., and Taylor-Papadimitriou, J. (1987) *Proc. Natl. Acad. Sci. U. S. A.* **84**, 6060–6064
27. Siddiqui, J., Abe, M., Hayes, D., Shani, E., Yunis, E., and Kufe, D. (1988) *Proc. Natl. Acad. Sci. U. S. A.* **85**, 2320–2323
28. Brand, J. P., Andrews, S. F., Jr., Cahill, E., Conway, N. A., and Young, J. D. (1981) *J. Biol. Chem.* **256**, 12205–12207
29. Hanover, J. A., Lennarz, W. J., and Young, J. D. (1980) *J. Biol. Chem.* **255**, 6713–6716
30. Denny, P. A., and Denny, P. C. (1982) *Carbohydrate Res.* **110**, 305–314
31. Amerongen, A. V. N., Oderkerk, C. H., Roukema, P. A., Wolf, J. H., Lieman, J. J. W., and Overdijk, B. (1983) *Carbohydrate Res.* **115**, C1–C5
32. Hilkens, J., and Buija, F. (1988) *J. Biol. Chem.* **263**, 4215–4222
33. Linsley, P. S., Kallestad, J. C., and Horn, D. (1988) *J. Biol. Chem.* **263**, 8390–8397
34. Eckhardt, A. E., Timpte, C. S., Abernethy, J. L., Toumacie, A., Johnson, W. C., Jr., and Hill, R. L. (1987) *J. Biol. Chem.* **262**, 11339–13344
35. Burchell, J., Gendler, S., Taylor-Papadimitriou, J., Girling, A., Millis, R., and Lampert, D. (1987) *Cancer Res.* **47**, 5476–5482
36. Karlson, S., Swallow, D. M., Griffiths, B., Corney, G., and Hopkinson, D. A. (1983) *Annu. Hum. Genet.* **47**, 263–269
37. Swallow, D. M., Gendler, S., Griffiths, B., Corney, G., Taylor-Papadimitriou, J., and Bromwell, M. E. (1987) *Nature* **328**, 82–84