Cloning and characterization of mouse intestinal MUC3 mucin: 3′ sequence contains epidermal-growth-factor-like domains

Laurie L. SHEKELS*, Denise A. HUNNINGHAKE*, Ann S. TISDALE§, Ilene K. GIPSON§, Marcia KIELISZEWSKI†, Christine A. KOZAK‡ and Samuel B. HO*[1]

*Department of Medicine, University of Minnesota and VA Medical Center, Minneapolis, MN 55417, U.S.A., †Department of Chemistry, Ohio University, Athens, Ohio 45701, U.S.A., ‡National Institutes of Health, Bethesda, MD 20892, U.S.A., and §Schepens Eye Research Institute, Boston, MA 02114, U.S.A.

Mucin glycoproteins are a heterogeneous family of high-molecular-mass, heavily glycosylated proteins differentially expressed in epithelial tissue of the gastrointestinal, reproductive and respiratory tracts. We report here the cloning of a mouse caecal mucin (MCM). Amino acid analysis of purified MCM revealed a high content of serine (10.8 %) and threonine (25.1 %). Antibodies against deglycosylated MCM were prepared for immunohistochemical analysis and for screening a mouse caecal cDNA library. Immunohistochemical analysis showed strong staining of goblet cells and patchy staining of surface columnar cells in the duodenum, small intestine, caecum, colon and rectum. Screening of a mouse caecal cDNA library yielded clones containing tandem repeats of 18 bp with two predominant peptide sequences of TTTADV and TTTVVV. The tandem repeat domain is followed by 1137 bp of non-repetitive sequence and 521 bp of 3′ untranslated sequence prior to the poly(A) tail. Two cysteine-rich regions lie within the 3′ non-repetitive domain. The arrangement of the cysteines within these regions corresponds to epidermal growth factor-like domains. Following the second cysteine-rich region is a stretch of 19 hydrophobic amino acids which may act as a transmembrane domain or allow for interaction with hydrophobic molecules. Northern blot analysis indicates the mRNA is approximately 13.5 kb with greatest expression in the caecum and lesser amounts in the colon and small intestine. No MCM message is found in mouse stomach, trachea, lung, kidney, oesophagus or pancreas. *In situ* hybridization studies show that MCM message is expressed at the tips of villi in the intestine and in the upper crypts and surface cells of the caecum and colon. Chromosomal analysis assigns this gene to mouse chromosome 5 in a region of conserved linkage with human chromosome 7, the location of the human MUC3 gene. We conclude that we have identified a mouse caecal mucin which represents the mouse homologue of human MUC3. The mouse MUC3 cDNA sequence suggests that it is a novel non-polymerizing mucin which may participate in membrane or intermolecular interactions through its 3′ non-repetitive region.

INTRODUCTION

Mucins are high-molecular-mass glycoproteins present in the mucus coating the epithelial surfaces of the gastrointestinal, respiratory and reproductive tracts [1]. Mucins are decorated by oligosaccharide chains O-linked to serine and threonine residues in the tandem repeat domain of the polypeptide backbone. Synthesized by epithelial tissues as membrane-bound or secreted proteins, mucins provide a protective gel and lubricating coat for epithelial surfaces. In the gastrointestinal tract, mucins are thought to provide a barrier to gastric acid, prevent dehydration, protect the gut wall by excluding microorganisms and parasites and to protect mucosa from digestive proteases. Alterations in this protective mucus coat and in mucin expression have been noted in various human diseases, such as gastric and colon cancers, ulcerative colitis and cystic fibrosis [2,3].

Currently nine unique human mucins have been identified, designated MUC1, 2, 3, 4, 5AC, 5B, 6, 7 and 8. The tandem repeat of each mucin is distinct with respect to length and amino acid sequence. MUC1 belongs to a family of membrane-bound mucins and is ubiquitously expressed [4,5]. Full-length mouse and human MUC1 clones have been obtained and, while their tandem repeat sequences differ significantly, the human and mouse MUC1 mucins share high sequence similarity in their non-repetitive and promoter regions [6]. The other mucins are characterized by tissue-specific expression. In brief, human MUC2 is expressed in the small intestine and colon and MUC3 is primarily expressed in small intestine, colon and gallbladder [7,8]. MUC4 is observed primarily in colon and bronchial tissue [8,9]. MUC5AC is found in bronchial, gastric tissues, gallbladder and endocervix [10,11]. MUC6 is primarily noted in stomach, gallbladder, endocervix, seminal vesicles, pancreas and Brunner's glands [8,10,12,13]. MUC5B is found in gallbladder, trachea and endocervix [14–16]. MUC7 was isolated from salivary glands [17] and MUC8 encodes a tracheobronchial mucin [18]; however, complete organ distribution studies of these mucins have not been completed. The functional significance of the presence of multiple mucins within a single tissue is unknown.

Emerging data from sequence analysis of cDNA clones enables one to hypothesize different roles for the mucins. For example, MUC2 contains cysteine-rich regions on either side of the tandem repeat domain which resemble the D domains of von Willebrand clotting factor [19]. By analogy to von Willebrand factor, the cysteine-rich regions of MUC2 function in end-to-end polymerization of MUC2. MUC5AC, MUC5B and MUC6 share similar cysteine-rich regions in their 3′ non-repetitive regions

Abbreviations used: MCM, Mouse caecal mucin; MGM, mouse gastric mucin; PMSF, phenylmethylsulphonyl fluoride; SDS/PAGE, sodium dodecyl sulphate polyacrylamide gel electrophoresis; PBS, phosphate-buffered saline containing 0.1 M phosphate and 0.15 M sodium chloride, pH 7.4; IPTG, isopropylthio-β-D-galactoside; KLH, keyhole limpet haemocyanin; TBS, Tris-buffered saline containing 25 mM Tris, pH 7.4, 140 mM NaCl, 3 mM KCl; GAPDH, glyceraldehyde phosphate dehydrogenase; EGF, epidermal growth factor; PEG, poly(ethylene glycol).

[1] To whom correspondence should be addressed. Present address: Department of Gastroenterology (111-D), VA Medical Center, 1 Veterans Drive, Minneapolis, MN 55417, U.S.A.

The sequence of mouse gastric mucin depicted in Figure 1 has been submitted to the GenBank™ Nucleotide Sequence Database under the accession number AF027131.

suggesting that they may also polymerize in an end-to-end fashion [20,24].

In order to delineate the MUC3 structure and provide insight into the function of MUC3, we have cloned the mouse homologue of MUC3. Because sequence data alone can not fully elucidate function, the availability of a mouse MUC3 will lead to the development of transgenic and knockout mice to further establish MUC3 function.

EXPERIMENTAL

Materials

Peroxidase-conjugated rabbit anti-chicken IgY antibody, 4-chloro-1-naphthol, diaminobenzidine and isopropylthio-β-D-galactoside were obtained from Sigma (St. Louis, MO, U.S.A.). Biotinylated rabbit anti-chicken antibody was purchased from Zymed Laboratories (South San Francisco, CA, U.S.A.). [α-^{32}P]dCTP and [α-^{35}S]dATP were purchased from Amersham.

Mucin purification

Mucin was purified by gel filtration and CsCl density gradient centrifugation from the soluble fraction of mouse caecal mucosa as has been described previously for the purification of mucin from mouse stomachs [21]. Briefly, the mucosa of freshly harvested mouse caecums was scraped into 0.1 M NH_4HCO_3, 0.5 M NaCl, 0.1 mM PMSF on ice. The sample was then homogenized followed by centrifugation for 45 min at 45000 ***g***. After removing the lipid layer, the supernatant was centrifuged as before and then dialysed overnight against 10 mM Tris, pH 8.0. The protein was size fractionated on a 2.5 cm × 70 cm Sepharose CL-4B column equilibrated in 10 mM Tris, pH 8.0. The void volume of the column was collected and dialysed against water, lyophilized and digested for 2 h at room temperature with RNase A and DNase I (1:100 protein ratio) in PBS, 1 mM $MgSO_4$, 0.1 mM PMSF, 0.2% NaN_3. Following digestion, the sample was centrifuged and the supernatant dialysed overnight against PBS. CsCl was added to the dialysed supernatant to a final concentration of 0.54 g/ml and then the sample was centrifuged for 72 h at 160000 ***g***. Fractions of 1 ml were collected and the density and protein and hexose content of each fraction was measured. Fractions of high density (> 1.35 g/ml) and with a hexose:protein ratio of 2:1 were pooled, dialysed against water and then applied to a 1 × 43 cm Sepharose CL-4B column equilibrated in 10 mM Tris, pH 8.0. The void volume containing the purified mucin was dialysed against water and lyophilized. The purity of the mucin was assessed by SDS/PAGE [25] and periodic acid/silver nitrate staining [26].

Amino acid analysis and mucin deglycosylation

Amino acid analysis of the purified mouse caecal mucin (MCM) was performed by the University of Minnesota Microchemical Facility. An aliquot of the purified protein was deglycosylated by treatment with anhydrous HF. Deglycosylation was performed in 2 ml Sarstedt screw cap microtubes with anhydrous HF at a concentration of 20 μg protein: 1 μl HF and containing 10% methanol. After 3 h at room temperature the reaction mixture was quenched by freezing it in liquid N_2 and by adding ice-cold water to bring the final HF concentration to 10%. The sample was dried under nitrogen, washed three times with water and again dried under nitrogen to concentrate the protein and remove the residual HF. The deglycosylated protein was resuspended in water and lyophilized.

Antibody production

White leghorn chickens, age 22–24 weeks, were injected with 15 μg deglycosylated MCM emulsified in complete Freund's adjuvant. Two boosters of 7.5 μg deglycosylated MCM emulsified in incomplete Freund's adjuvant were given at 2-week intervals. Ten days following the final booster, eggs were collected for antibody purification. Polyclonal chicken IgY was purified as described by Goueli et al. [27]. Briefly, the egg yolks were separated and mixed with an equal volume of buffer A (10 mM potassium phosphate, pH 7.5, 0.1 M NaCl and 0.1% NaN_3). A volume of 10.5% poly(ethylene glycol) (PEG) in buffer A equivalent to the total egg yolk volume was added and stirred for 30 min at room temperature. The mixture was centrifuged at 12000 ***g*** for 20 min. Lipids were removed by addition of silicon dioxide to the supernatant to a final concentration of 5 g/100 ml. The mixture was stirred for 20 min and then allowed to stand at room temperature for 10 min before centrifugation. A solution of 42% PEG in buffer A was added to yield a final concentration of 12% PEG and stirred for 30 min at 4 °C. The precipitated proteins were collected by centrifugation and the pellet was dissolved in a minimal volume of buffer A. An equivalent volume of 4 M $(NH_4)_2SO_4$, pH 7.0 was added and the sample was stirred for 30 min at 4 °C. The antibodies were collected by centrifugation.

ELISA analysis

Antigens (10 ng) were plated on 96-well ELISA plates for 2 h at room temperature. Plates were blocked with 5% BSA in TBS overnight at 4 °C. The plates were washed with 0.02% Tween-20 in TBS and incubated with the primary antibody for 3 h at room temperature. Following washing as before, peroxidase-conjugated rabbit anti-chicken IgY (1:2000) was added for 1.5 h. Colour development was performed with 3,3′,5,5′-tetramethylbenzidine and quenched with 2 M H_2SO_4. Bound antibody was quantified by measuring the absorbance at 450 nm with a TitreTek spectrophotometer. Preimmune antibodies were used as negative controls.

Immunohistochemistry

Localization of mucin protein expression was determined using the streptavidin-peroxidase technique on formalin-fixed specimens of normal mouse tissues [7]. Antibody against deglycosylated mouse caecal mucin was reactive with both formalin-fixed and frozen-ethanol-fixed sections. Prior to staining, a high-pressure heat-induced antigen retrieval treatment was performed [28]. Briefly, tissue sections were deparaffinized and rehydrated followed by 3 M urea treatment and heated under pressure (15 lbf/in², 104 kPa) for 5 min. The slides were washed in PBS and incubated with fresh 3% hydrogen peroxide in methanol for 10 min followed by another PBS wash. Normal rabbit serum (5%) and 1% BSA and 10% skim milk in PBS was applied for 20 min and removed by blotting. Next the sections were incubated with the primary antibody for 90 min at a dilution of 1:2500 (2 μg/ml). The sections were then washed and incubated with the biotinylated rabbit anti-chicken antibody (1:75 dilution in PBS) for 20 min. After washing, the sections were incubated with streptavidin-peroxidase conjugate (10 μg/ml) for 30 min followed by repeated washing. Next the sections were incubated with diaminobenzidine in 0.03% hydrogen peroxide for 10 min, washed, counterstained with haematoxylin, rinsed in tap water, and mounted. Preimmune chicken IgY (2 μg/ml) was substituted for the primary antibodies as a negative control.

Isolation of a mouse caecal mucin cDNA

Mouse caecal RNA was isolated by the acid guanidinium thiocyanate/phenol/chloroform extraction method [29]. Caecal poly(A) RNA was isolated from the total RNA preparation by oligo(dT) chromatography and used for the construction of a cDNA library in the bacteriophage λZAPII by Stratagene. Identification of MCM clones was performed by screening the expression library with anti-deglycosylated MCM. Positive clones were visualized with peroxidase-conjugated rabbit anti-chicken antibody using 4-chloro-1-naphthol as the substrate. Hybridizing plaques were purified by successive rounds of screening.

DNA sequencing and sequence analysis

The cDNAs were sequenced by the Sanger dideoxy-mediated chain termination method [30] using Sequenase Version 2.0 (United States Biochemical Corporation). Both strands were sequenced by taking advantage of restriction sites (*Eco*RI at bp 468, *Pst* sites at bp 902, 924, 1461 and 1521) and subcloning smaller fragments. In addition to using universal sequencing primers, MCM specific primers were made: MCM1 (351–364, tgtgaggaactggt), MCM9 (1025–1011, tccatacaccaggct) and MCM7 (1722–1708, agtgcttgccatgga). The University of Wisconsin Genetics Computer Group software was used to analyse DNA sequence information [31].

RNA and DNA analysis

RNA was isolated from various mouse and rat tissues by the acid guanidinium thiocyanate-phenol-chloroform extraction method [29]. A 10 μg portion of each RNA was separated on 1.2 % agarose gels. The gels were stained with ethidium bromide to assess RNA integrity and then the RNA was transferred to Nytran nylon membranes. Following prehybridization, the filters were hybridized in the presence of radiolabelled cDNA probes which had been prepared by the random primer method [32]. In addition to the MCM clones, mouse gastric MUC5AC mucin [21] and rat glyceraldehyde phosphate dehydrogenase [33] cDNAs were also used as probes. The membranes were washed twice with 2 × SSC, 0.1 % SDS at room temperature for 30 min, once with 0.1 × SSC, 0.1 % SDS for 1 h at room temperature and finally with 0.1 × SSC, 0.1 % SDS at 55 °C (1 × SSC is 0.15 M NaCl/0.015 M sodium citrate).

DNA was isolated from mouse, hamster, rabbit, rat and human using the Puragene DNA isolation kit for human and animal tissue (Gentra Systems, Inc.). For Southern analysis, 10 μg of purified DNA was digested with *Bam*HI or *Pst*I overnight at 37 °C and the digested DNA was separated on a 1.2 % agarose gel. Following denaturation of the gel, the DNA was transferred to a Nytran nylon membrane. The membrane was prehybridized, hybridized and washed as described for Northern analysis of RNA.

Chromosomal localization

The MCM gene was mapped by analysis of the progeny of two genetic crosses: (NFS/N or C58/J × *M. m. musculus*) × *M. m. musculus* [34] and (NFS/N × *M. spretus*) × C58/J [35]. DNA from parental mice and the progeny of both crosses were typed by Southern blotting for restriction enzyme polymorphisms using MCM as the probe. The progeny of these crosses was also typed for inheritance of over 1000 markers which map to all 19 autosomes and the X chromosome including the Chr5 markers: *Gus*, *Ncf1*, *Zp3* and *Ccnb1-rs1* as reported previously [36,37]. Percentage recombination and standard errors between loci were calculated as described by Green [38]. Data was stored and analysed using the program LOCUS prepared by C. E. Buckler (NIAID, NIH, Bethesda, MD U.S.A.).

In situ hybridization

To determine the cellular distribution of MCM, *in situ* hybridization was performed using a ^{35}S-labelled cDNA probe corresponding to the tandem repeat region of MCM. Antisense and sense riboprobes to MCM were prepared by reverse transcription using T3 or T7 RNA polymerase (Boehringer–Mannheim). Six-micrometre paraffin sections of mouse tissue were deparaffinized and treated with proteinase K for 30 min at 37 °C. Acetylation was carried out in 0.5 % acetic anhydride, 0.1 M triethanolamine (pH 8) for 10 min followed by hybridization overnight at 52 °C with (2–5) × 10^6 c.p.m./ml ^{35}S-labelled probe in 50 % formamide, 20 % dextran sulphate, 0.3 M NaCl, 10 mM Tris (pH 7.6), 5 mM EDTA, 0.02 % Ficoll 400, 0.02 % polyvinylpyrrolidone, 0.02 % BSA, 0.1 M dithiothreitol and 0.02 mg/ml tRNA. Sections were washed twice with 2 × SSC/50 % formamide, 14 mM β-mercaptoethanol at 57 °C for 30 min and then for 15 min. Sections were then treated with 20 μg/ml RNase A in 10 mM Tris, 1 mM EDTA at 37 °C for 30 min. Two additional washes were carried out for 15 min each followed by washing at 1 × SSC then 0.1 × SSC at 37 °C for 30 min each. Slides were dipped in Kodak NTB-2 emulsion (Eastman Kodak), developed after 1 week and counterstained with haematoxylin and eosin.

RESULTS

Mucin purification and antibody production

MCM was purified by gel filtration and CsCl density gradient centrifugation. The caecal portions of colons from 336 mice were processed and yielded 392 μg of purified mucin. Analysis of the purified protein demonstrated a high carbohydrate content, with a hexose:protein ratio of at least 3:1. Amino acid analysis revealed that the protein contained high concentrations of threonine (25.1 %) and serine (10.8 %) which is consistent with other mucins (Table 1). The MCM serine/threonine composition

Table 1 Amino acid analysis of mouse caecal, rat intestinal and mouse gastric mucins

The emboldened material refers to characteristic mucin amino acids.

Amino acid	Mucin ...	Concn. (%) Mouse caecal	Rat intestinal	Mouse gastric
Asx		7.3	6.8	9.6
Thr		**25.1**	**30.8**	**13.5**
Ser		**10.8**	**10.6**	**12.2**
Glx		8.9	6.7	10.4
Pro		**13.0**	**11.2**	**7.4**
Gly		6.3	5.9	8.4
Ala		3.9	6.7	4.3
Val		5.4	6.4	6.4
Met		0.4	1.1	1.4
Ile		4.7	2.8	3.2
Leu		4.0	3.6	5.4
Tyr		0.9	N.D.	2.5
Phe		1.9	N.D.	2.6
His		1.8	4.0	2.3
Lys		3.4	1.7	3.9
Arg		1.7	1.6	2.3

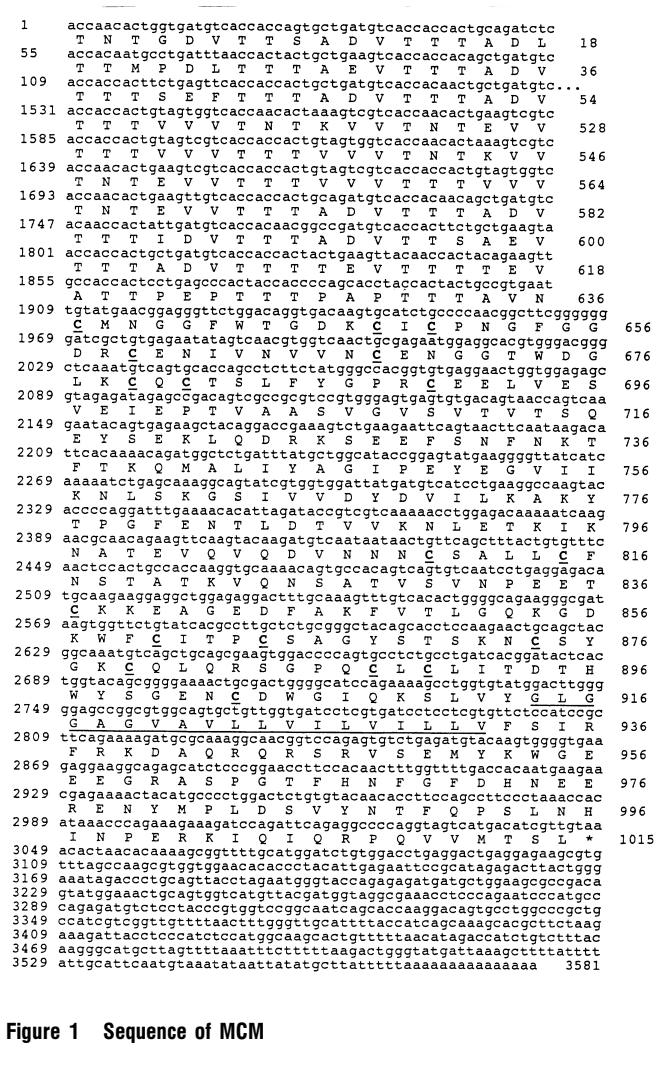

Figure 1 Sequence of MCM

The nucleotide and one-letter amino acid sequence are shown for the tandem repeat domain and 3′ non-repetitive region of MCM. The cysteine residues are in bold and underlined and the hydrophobic stretch of amino acids is underlined. The nucleotides are numbered on the left and the amino acids are numbered on the right side of the figure.

is similar to that of rat intestinal mucin [39]; however, it differs from the MCM by having a greater threonine percentage and slightly different concentrations of serine and proline.

cDNA isolation and sequencing

Approx. 100000 recombinants were screened with the anti-deglycosylated MCM antibody from which two positive clones were identified. MCM1 and MCM2 contained inserts of approximately 1800 and 2000 bp, respectively and sequencing the ends of each clone gave 18 bp tandem repeats. The tandem repeat sequence contains a conserved *Mae*III restriction enzyme site. Digestion with *Mae*III results in complete digestion of the clones indicating that they consist entirely of tandem repeats. Therefore, no additional sequencing of MCM1 and MCM2 was done. These clones were used for further screening of the mouse caecal cDNA library and initially only additional tandem repeat clones were isolated. The tandem repeats encode for a six amino acid peptide. Distribution analysis of the tandem repeat units sequenced shows high conservation of the first three amino acids. The first three amino acids of MCM are virtually always

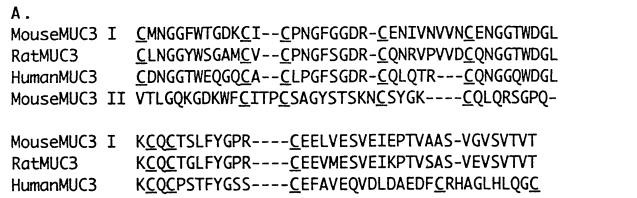

Figure 2

(**A**) Sequence similarity between the 3′ non-repetitive region of the mouse, rat and human intestinal MUC3 mucins. Cysteines are underlined. The mouse sequence is shown for the first cysteine-rich region amino acids 635–714; the rat sequence corresponds to amino acids 356–447 [39], the human sequence corresponds to amino acids 852–927 [41] and the second mouse cysteine-rich region for amino acids 849–928. (**B**) Alignment of the cysteine residues within the cysteine-rich region corresponding to the EGF-like domain. Mouse MUC3 I and mouse MUC3 II, the first and second cysteine-rich regions in mouse MUC3, respectively.

threonine, which follows the amino acid analysis data indicating a higher prevalence of threonine in comparison to serine. The last three amino acid positions show more variability. The last amino acid is usually valine. The fourth residue seems to be equally distributed between valine and alanine and at the fifth position aspartic acid and valine are predominant. Interestingly, the last three amino acids always appear in the combination ADV or as three valines suggesting that two consensus tandem repeats may be present – TTTADV and TTTVVV. Within individual clones the two different consensus sequences cluster together giving a type of subdomain structure with regions of tandem repeats containing the hydrophobic second half and regions with a charged residue in the second half. The subdomain-type structure is an intriguing observation, however, it remains to be determined how the two consensus sequences act in mucin structure, glycosylation or function. The MCM consensus sequences are very similar to the consensus sequence for the rat intestinal mucin (consensus sequence TTTPDV) [39]. Of particular interest is the virtual absence of proline and serine residues in the MCM tandem repeats. Proline residues are important in recognition of sites to be O-glycosylated by glycosyltransferases [40] and the difference in their occurrence in the rat versus mouse tandem repeats may reflect species differences in glycosylation patterns.

To obtain the MCM3′ non-repetitive sequence, the mouse caecal cDNA library was consecutively screened with MCM1 and then with 3′ non-repetitive subclones. The sequence 3′ to the tandem repeat domain consists of 1137 bp non-repetitive sequence followed by 521 bp of 3′ untranslated sequence adjacent to the poly(A) tail (Figure 1). The sequence for rat intestinal mucin cDNA, RMUC176, has 92 amino acids of non-repetitive sequence at the 3′ end of its tandem repeat domain [39]. Within this region MCM is 76% identical to the rat intestinal mucin sequence (Figure 2A). Further sequence analysis indicated that MCM shares 41% identity to the human intestinal MUC3 mucin in a 92 amino acid region of the human MUC3 3′ non-repetitive sequence [41]. Of particular interest is the conservation of eight cysteine residues within this region between the mouse, rat and human mucins. The last six cysteine residues conform to an EGF-like domain (Figure 2B). Adjacent to the first MCM

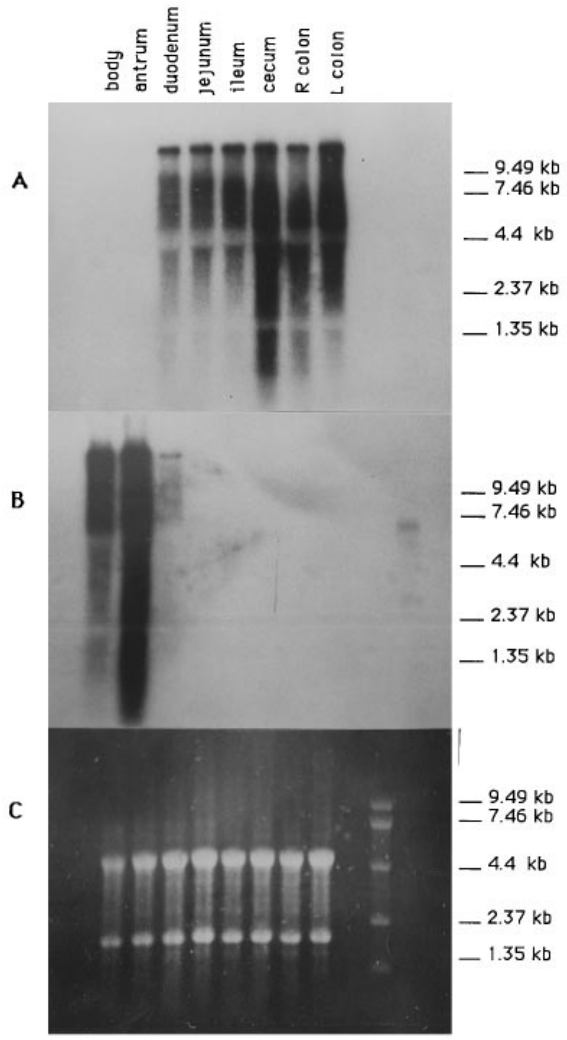

Figure 3 Northern blot analysis of RNA from mouse tissues

RNA isolated from the indicated tissues was hybridized to (**A**) MCM and (**B**) MGM (mouse MUC5AC); (**C**) the ethidium bromide staining of the RNA gel showing intact, non-degraded RNA.

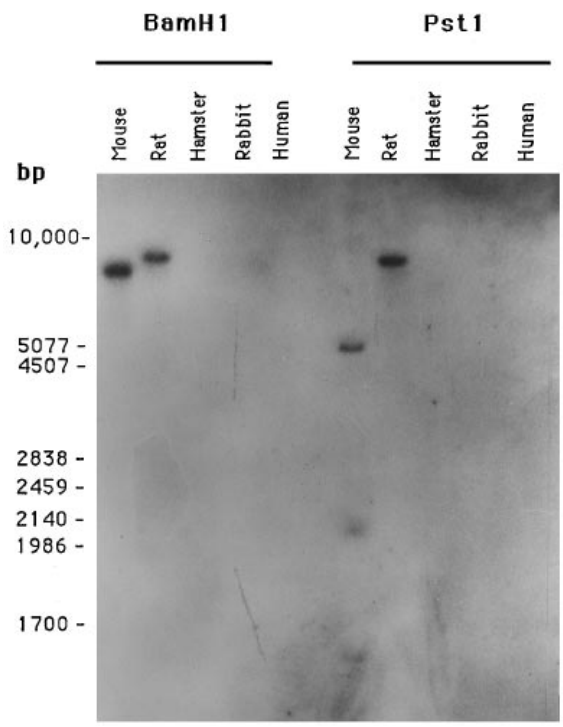

Figure 4 Southern blot analysis

Spleen DNA (10 μg) from the indicated species was digested with either *Bam*HI or *Pst*I and separated on an agarose gel. The DNA was transferred to a nylon membrane and hybridized to an MCM tandem repeat probe. Markers along the side indicate size in bp.

cysteine-rich EGF-like domain are 119 amino acids of non-repetitive sequence containing no cysteine residues and few serine or threonine residues. This region is followed by a second cysteine-rich domain containing ten cysteine residues. The arrangement of the last six cysteine residues within this second cysteine-rich region also aligns with the cysteines found in an EGF-like domain. The two MCM cysteine-rich regions share only 14% sequence similarity, which primarily results from conservation of cysteine residues. Neither MCM cysteine-rich region resembles the von Willebrand factor D domains found in other secretory mucins. A stretch of 19 hydrophobic residues lies adjacent to the second cysteine-rich region followed by 83 amino acids containing no cysteine residues and few serine or threonine residues. The 3′ untranslated region contains 521 bp prior to the poly(A) tail.

RNA and DNA analysis

Analysis of the tissue and species specificity of MCM was performed by Northern and Southern blot analysis. When mouse RNA is probed with a MCM tandem repeat clone, expression is observed only in the caecum, small intestine and large intestine (Figure 3A). No hybridization is observed with mouse stomach RNA. The MCM message is polydisperse as described previously for other mucins [21,42] beginning at 13.5 kb and does not indicate degraded RNA as evidenced by ethidium bromide staining (Figure 3C) and hybridization to a rat GAPDH probe (data not shown). The level of expression appears highest in the caecum followed by the left and right colon and then the small intestine. This pattern of expression is in contrast to the MGM (MUC5AC) which is found primarily in the mouse antrum and fundus with some hybridization in the duodenum (Figure 3B).

Southern blot analysis of mouse DNA digested with *Bam*HI reveals a band of less than 10 kb when using a MCM tandem repeat probe (Figure 4). Three hybridizing bands are observed in mouse DNA digested with *Pst*I. A related gene is present in rat as evidenced by a single band in DNA digested with *Bam*HI or *Pst*I. No hybridization was observed in hamster, rabbit or human DNA under the indicated hybridization and wash conditions.

Localization of MCM expression

Caecum mucin cellular location and distribution were determined by performing immunohistochemical analysis of various mouse tissues (Table 2). The antibody against deglycosylated MCM was found to strongly stain goblet cells of the duodenum, small intestine, caecum, colon and rectum (Figure 5). This antibody did not stain mouse pancreas, oesophagus, kidney, lung or liver tissues. This goblet cell reactivity is similar to the location of human MUC2 (Figures 5B and 5C). This polyclonal antibody

Table 2 Immunohistochemical distribution of MCM and MGM epitopes

−, Negative reactivity; ±, trace positive; +, positive; ++/+++, strong positive.

Tissue	Anti-deglycosylated MCM	Anti-MGM tandem repeat
Stomach	—	++ Surface mucous cells
		++ Antral glands
Small intestine	++ Goblet cells, crypt bases, villi	− Columnar cells
		± Glands at crypt base
	+ Some villi of columnar cells	+++ Brunner's glands
	− Brunner's glands	
Caecum	++ Supranuclear cytoplasm goblet cells	± Goblet cells
	++ Surface columnar cells	− Columnar cells
Colon/rectum	+ Goblet cells	—
	+ Surface columnar cells	
Kidney	—	—
Liver	—	—
Bronchus/lung	—	+ Glands
Pancreas	—	—
Gallbladder	+ Columnar cells	+ Columnar cells

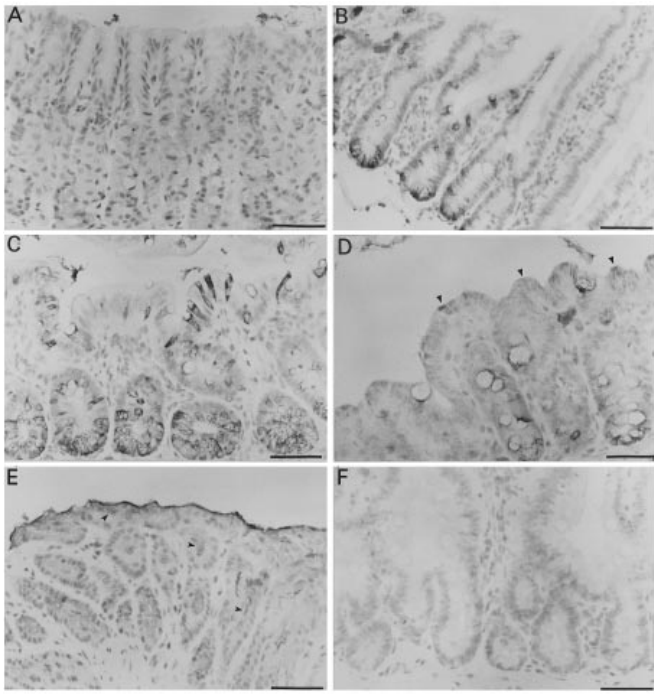

Figure 5 Immunohistochemistry of mouse gastrointestinal tissues

(**A**) Stomach stained with antibody against deglycosylated mouse caecal mucin (anti-MCM) showing no staining. (**B**) Duodenum stained with anti-MCM. (**C**) Ileum stained with anti-MCM. (**D**) Colon stained with anti-MCM. Arrowheads indicate staining of surface columnar cells, and staining is also present in crypt goblet cells. (**E**) Stomach stained with MGM tandem repeat antibody (anti-MGM). Arrowheads indicate positive staining on surface cells and within glandular lumen. (**F**) Intestine stained with anti-MGM reveals negative staining. Bars = 50 μm.

appears to recognize epitopes residing in locales of both MUC2 and MUC3. Patchy staining of the columnar cells of the villous tips in the small intestine and in the surface columnar cells of the colon was noted (Figure 5D). This is the location of expression of the human intestinal MUC3 mucin [43]. In addition, this antibody stained columnar cells of the gallbladder which in the human are also known to express MUC3.

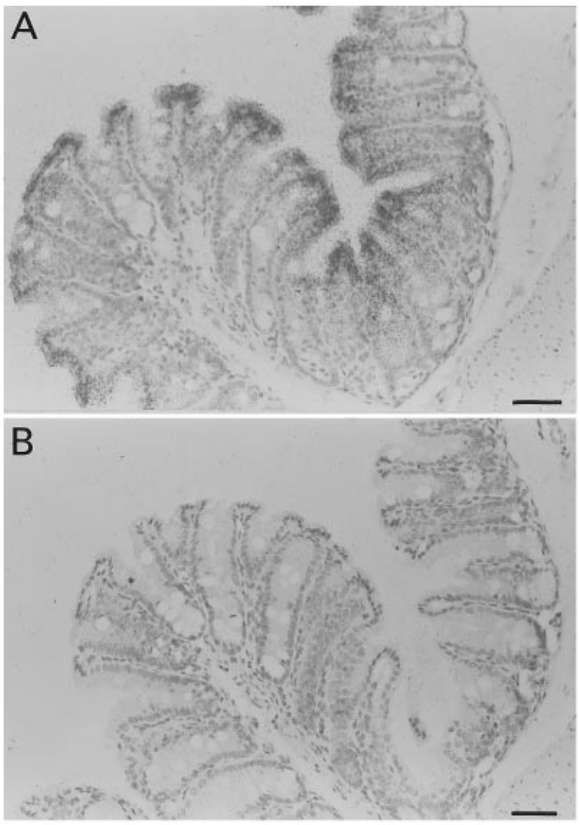

Figure 6 *In situ* hybridization

(**A**) Mouse caecum hybridized to an antisense MCM tandem repeat probe. (**B**) Mouse caecum hybridized to a sense MCM tandem repeat probe showing no hybridization. Bars = 50 μm.

The distribution of deglycosylated caecal mucin immunoreactivity differs from that of the gastric mucin as evidenced by staining with an antibody against the MGM tandem repeat (mouse muc5ac). The anti-MGM antibody demonstrates strong staining of the surface and neck mucous cells of the fundus, surface mucus cells and antral glands of the antrum and cardia and Brunner's glands in the duodenum (Figure 5E). No staining with anti-MGM was noted in the small intestine, caecum, colon or rectum (Figure 5F) [21].

In situ hybridization demonstrated hybridization of the MCM tandem repeat probe in cells at the tips of the villi in the small intestine and most strongly in cells of the upper crypt and surface of the caecum and colon (Figure 6A). No hybridization was seen in stomach tissue. Using the sense probe as the negative control showed no binding (Figure 6B). This hybridization pattern is similar to that seen for *in situ* hybridization analysis of human MUC3 in the small intestine and colon [43].

Chromosomal localization

Southern blot analysis identified a 23.0 kb *Apa*I fragment in NFS/N and a 28 kb fragment in *M. m. musculus*. *Sca*I digestion produced 15.5, 12.8 and 2.7 kb fragments in NFS/N and a 13.5 kb fragment in *M. spretus*. Inheritance of the inbred strain fragment and the *M. spretus* fragment was followed in the progeny of two genetic crosses and compared with inheritance of over 1000 markers in the two sets. The gene for mouse intestinal mucin *Muc3* was linked to markers on distal mouse Chr 5. No

recombinants were found between *Muc3* and *Ncf1* and *Zp3* in 160 mice indicating that these genes are within 1.9 cM at the upper limit of the 95% confidence level. This places mouse *Muc3* at or near the end of a region of conserved linkage with human chromosome 7 on which human *MUC3* is found at 7q22 [44].

DISCUSSION

Various rat intestinal mucins have been cloned in previous studies [39,45–47] but to our knowledge this report represents the first intestinal mucin described in mouse. Based on data from cDNA sequencing, Northern and Southern blot analysis, chromosomal location and *in situ* hybridization, we conclude that we have identified a tissue-specific mucin representing the mouse homologue of MUC3. Support that our clone represents a mucin is provided by the presence of a threonine-rich tandem repeat sequence. A repeated domain rich in serine and/or threonine is a key feature of mucin glycoproteins. This region provides multiple sites for O-linked glycosylation through the hydroxyl group of serine or threonine. A polydisperse hybridization pattern is observed with Northern blot analysis, which is observed with other mucin transcripts and may result from rapid turnover of the mucin message, degradation of the long transcript due to instability of the extended repeat domain, or incomplete or alternative splicing [2,42]. Northern blot analysis shows MCM is strictly expressed in the intestinal tract. Immunohistochemical and *in situ* analysis localized MCM to small intestine, colon and gallbladder, a pattern distinct from the MGM MUC5AC. Expression of MCM in villous tips of the small intestine and surface and upper crypt cells of the caecum and colon is identical to the distribution of the human MUC3.

The rat intestine has been shown to express at least two different mucins [46]. The human intestine expresses multiple mucins, including MUC2, MUC3, and MUC4. The rat MUC2 mucin has been cloned [45,46,48] and shares striking similarity to human MUC2 in the non-repetitive regions. The rat MUC2 shows no sequence similarity to MCM. Two other partial rat intestinal mucin cDNA clones, RMUC176 and M2, have been isolated which are distinct from rat MUC2 [39,46]. The tandem repeat sequences of RMUC176 and M2 are similar to the MCM repeat sequence. The available 3′ sequence of RMUC176 encodes for 92 amino acids of non-repetitive sequence which is 76% identical to MCM and shares the conserved cysteine residues in an EGF-like domain. This differs from M2 which has a very hydrophobic domain comprised of 82 amino acids immediately adjacent to its tandem repeat domain. Khatri et al. [46] argue that it is unlikely that the M2 hydrophobic domain represents a membrane-spanning segment due to its extended length which contrasts greatly to the size of the transmembrane regions of two

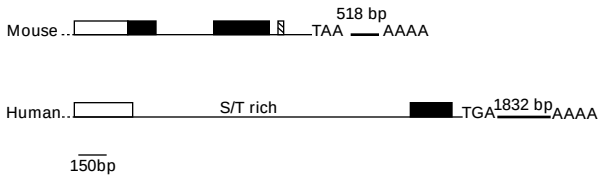

Figure 7 Schematic diagram of mouse and human MUC3 mucins

The mouse and human MUC3 mucins illustrate the structural domains of the 3′ region as determined by cDNA sequencing. White box, tandem repeat domain; black box, EGF-like cysteine-rich domain; hatched box, hydrophobic segment; TAA, TGA, stop codons and AAAA, poly(A) tail. The coding region is drawn to scale; the 3′ untranslated region is not drawn to scale.

membrane-bound mucins MUC1 (31 amino acids) [4] and rat mammary carcinoma mucin component ASGP-2 (26 amino acids) [49]. Alternatively, this region of M2 may interact with other hydrophobic molecules such as lipids and mutagens. The 3′ end of MCM contains a hydrophobic domain of 19 amino acids which could function to insert MCM into the membrane or maintain hydrophobic interactions as suggested for M2. Experiments are in progress to define the function of the MCM hydrophobic segment.

Despite the lack of homology between the human and mouse mucin tandem repeats, the mouse intestinal mucin clone does share significant sequence similarity with a portion of the recently published 3′ non-repetitive region of human MUC3 [41]. Within this region, MCM shares 41% similarity to the human MUC3. The mouse and human MUC3 mucins differ in their overall structural domains (Figure 7). Immediately adjacent to the mouse tandem repeat domain lies a cysteine-rich region, however, the human MUC3 has a serine/threonine-rich region between its tandem repeat domain and its cysteine-rich region. The mouse MUC3 has a second cysteine-rich region separated from the first cysteine-rich region by 119 amino acids. A recently published preliminary sequence analysis of a rat MUC3 cDNA obtained by 3′-RACE techniques indicates that it also contains two cysteine-rich regions [50]. This second cysteine-rich region is absent in the human MUC3 which ends 30 amino acids beyond its only 3′ cysteine-rich region. The shared cysteine-rich regions of the mouse, rat and human MUC3 mucins show noted conservation of the cysteine residues. The spacing of the MCM and human MUC3 cysteines is unlike that found in the cysteine-rich regions of MUC2, MUC5AC, MUC5B, MUC6 and von Willebrand factor D domains. It has been suggested that the von Willebrand D domain spacing of the cysteines in these molecules enables them to form intermolecular disulphide bonds creating end-to-end mucin polymers [14,19,22–24]. The differences in cysteine motifs may provide insight into the individual roles of the different mucins.

Interestingly, within the MCM cysteine-rich regions are found EGF-like motifs. The six cysteines found in EGF-like domains form three intramolecular disulphide bonds creating a structural domain which is important in maintaining protein–protein interactions or perhaps protein–membrane interactions [51]. EGF-like domains are found in several growth factors as well as in numerous extracellular proteins involved in formation of the extracellular matrix, cell adhesion, chemotaxis and wound healing. By structural analogy, one could propose that MCM may have a related function. There is some evidence suggesting that mucin does aid in wound healing. The secretion of mucin from goblet cells has been shown to occur in conjunction with secretion of trefoils, small proteins which facilitate repair after mucosal injury. Recent reports describe the ability of mucin both alone and in combination with trefoil proteins to promote restitution of wounded epithelium [52]. However, the mucin preparations used were heterogeneous, i.e. the preparations contained more than one mucin species, preventing the assignment of a wound-healing role to a specific mucin. In addition to simply forming a protective mucus layer, mucins have been postulated to selectively bind to substances to facilitate uptake by epithelial cells [1]. For example, the acidic mucin fraction has been proposed to aid fatty acid uptake. MUC3 mucins fractionate with the acidic mucin fraction [53] and thus may not polymerize but rather interact with other molecules. Suggestive functional implications for MUC3 may be inferred from a recent report describing linkage between inflammatory bowel disease and regions on human chromosomes 12, 7q22 and 3 [54]. Given the fact that MUC3 resides near human chromosome 7q22, MUC3 has been sug-

gested to be a possible candidate inflammatory bowel disease susceptibility gene [54].

We report here the cloning of a mouse intestinal mucin. Based on expression pattern, sequence similarity to the human and rat intestinal MUC3 mucins and chromosomal location, MCM represents the mouse homologue of MUC3. The presence of an EGF-like domain and a hydrophobic domain within the mouse MUC3 3′ non-repetitive domain implies a role for MUC3 distinct from the MUC2 intestinal mucin. MUC3 may participate in membrane or intermolecular interactions and may not form gels by end-to-end polymerization as the MUC2, MUC5AC, MUC5B and MUC6 secretory mucins are believed to do. The availability of a mouse MUC3 cDNA will allow for studies defining the functional role for MUC3 as well as determining the regulatory elements responsible for its expression.

This work was supported by a National Research Service Award (L. L. S.), a Veterans Affairs Merit Review Award (S. B. H.) and the research service of the Veterans Affairs Medical Center. We would like to thank Dr. James Gum for helpful discussions regarding the human MUC3.

REFERENCES

1 Strous, G. J. and Dekker, J. (1992) Crit. Rev. Biochem. Mol. Biol. **27**, 57–92
2 Rose, M. (1992) Am. J. Physiol. **263**, L413–L429
3 Neutra, M. and Forstner, J. F. (1987) in Physiology of the Digestive Trace (Johnson, L., ed.), pp. 975–1009, Raven Press, New York
4 Gendler, S. J., Lancaster, C. A., Taylor-Papadimitriou, J. et al. (1990) J. Biol. Chem. **265**, 15286–15293
5 Shimizu, Y. and Shaw, S. (1993) Nature (London) **366**, 630–631
6 Spicer, A. P., Parry, G., Patton, S. and Gendler, S. J. (1991) J. Biol. Chem. **266**, 15099–15109
7 Ho, S. B., Niehans, G. A., Lyftogt, C. et al. (1993) Cancer Res. **53**, 641–651
8 Audie, J. P., Janin, A., Porchet, N. et al. (1993) J. Histochem. Cytochem. **41**, 1479–1485
9 Porchet, N., Cong, N. V., Dufosse, J. et al. (1991) Biochem. Biophys. Res. Comm. **175**, 414–422
10 Ho, S. B., Roberton, A. M., Shekels, L. L. et al. (1995) Gastroenterology **109**, 735–747
11 Guyonnet Duperat, V., Audie, J.-P., Debailleul, V. et al. (1995) Biochem. J. **305**, 211–219
12 Toribara, N. W., Roberton, A. M., Ho, S. B. et al. (1993) J. Biol. Chem. **268**, 5879–5885
13 deBolos, C., Garrido, M. and Real, F. (1995) Gastroenterology **109**, 723–734
14 Keates, A., Nunes, D., Afdhal, N., Troxler, R. and Offner, G. (1997) Biochem. J. **324**, 295–303
15 Gipson, I., Ho, S. B., Spurr-Michaud, S. et al. (1997) Biol. Reproduction **56**, 999–1011
16 Desseyn, J.-P., Guyonnet-Duperat, V., Porchet, N., Aubert, J.-P. and Laine, A. (1997) J. Biol. Chem. **272**, 3168–3178
17 Bobek, L. A., Tsai, H., Biesbrock, A. R. and Levine, M. J. (1993) J. Biol. Chem. **268**, 20563–20569
18 Shankar, V., Pichan, P., Eddy, R. et al. (1997) Am. J. Resp. Cell and Mol. Biol. **16**, 232–241
19 Gum, J., Hicks, J., Toribara, N., Siddiki, B. and Kim, Y. (1994) J. Biol. Chem. **269**, 2440–2446
20 Meerzaman, D., Charles, P., Daskal, E. et al. (1994) J. Biol. Chem. **269**, 12932–12939
21 Shekels, L. L., Lyftogt, C. T., Kieliszewski, M. et al. (1995) Biochem. J. **311**, 775–785
22 Toribara, N. W., Ho, S. B., Gum, E. et al. (1997) J. Biol. Chem. **272**, 16398–16403
23 Desseyn, J.-L., Aubert, J.-P., Van Seuningen, I., Porchet, N. and Laine, A. (1997) J. Biol. Chem. **272**, 16873–16883
24 Perez-Vilar, J., Eckhardt, A. E. and Hill, R. L. (1996) J. Biol. Chem. **271**, 9845–9850
25 Laemmli, U. (1970) Nature (London) **227**, 680–685
26 Dubray, G. and Bezard, G. (1982) Anal. Biochem. **119**, 325–329
27 Goueli, S. A., Hanten, J., Davis, A. and Ahmed, K. (1992) Biochem. Int. **21**, 685–694
28 Norton, A. (1994) J. Pathol. **173**, 371–379
29 Chomczynski, P. and Sacchi, N. (1987) Anal. Biochem. **162**, 156–159
30 Sanger, F., Nicklen, S. and Coulson, A. (1977) Proc. Natl. Acad. Sci. U.S.A. **74**, 5463–5467
31 Devereux, J., Haeberli, P. and Smithies, O. (1984) Nucl. Acids Res. **12**, 387–395
32 Feinberg, A. P. and Vogelstein, B. (1983) Anal. Biochem. **132**, 6–13
33 Tso, J. Y., Sun, X. H., Kao, T. H., Reece, K. S. and Wu, R. (1985) Nucl. Acids Res. **13**, 2485–2502
34 Kozak, C. A., Peyser, M., Krall, M. et al. (1990) Genomics **8**, 519–524
35 Adamson, M. C., Silver, J. and Kozak, C. A. (1991) Virology **183**, 778–781
36 Hanley-Hyde, J., Muchinski, J., Sadofsky, M. et al. (1992) Genomics **13**, 1018–1030
37 Jackson, S., Malech, H., Kozak, C. et al. (1994) Immunogenetics **39**, 272–275
38 Green, E. L. (1991) in Genetics and Probability in Animal Breeding Experiments, pp. 77–113, Macmillan, New York
39 Gum, J. R., Hicks, J. W., Lagace, R. E. et al. (1991) J. Biol. Chem. **266**, 22733–22738
40 Briand, J. P., Andrews, S. P., Cahill, E., Conway, N. A. and Young, J. D. (1991) J. Biol. Chem. **256**, 12205–12207
41 Gum, J., Ho, J., Pratt, W. et al. (1997) J. Biol. Chem. **272**, 26678–26686
42 Gum, J. R. (1992) Am. J. Respir. Cell Mol. Biol. **7**, 557–564
43 Chang, S.-Y., Dohrman, A. F., Basbaum, C. B. et al. (1994) Gastroenterology **107**, 28–36
44 Fox, M., Lahbib, F., Pratt, W. et al. (1992) Ann. Hum. Genet. **56**, 281–287
45 Xu, G., Huan, L.-J., Khatri, I. A. et al. (1992) J. Biol. Chem. **267**, 5401–5407
46 Khatri, I., Forstner, G. and Forstner, J. (1993) Biochem. J. **294**, 391–399
47 Hansson, G., Baeckstrom, D., Carlstedt, I. and Klinga-Levan, K. (1994) Biochem. Biophys. Res. Commun. **198**, 181–190
48 Ohmori, H., Dohrman, A. F., Gallup, M. et al. (1994) J. Biol. Chem. **269**, 17833–17840
49 Sheng, Z., Wu, K., Carraway, K. and Fregien, N. (1992) J. Biol. Chem. **267**, 16341–16346
50 Khatri, I., Forstner, G. and Forstner, J. (1997) Biochim. Biophys. Acta **1326**, 7–11
51 Baron, M., Norman, D. and Campbell, I. (1991) Trends Biochem. Sci. **16**, 13–17
52 Kindon, H., Pothoulakis, C., Thim, L., Lynch-Devaney, K. and Podolsky, D. K. (1995) Gastroenterology **109**, 516–523
53 Gum, J. R., Hicks, J. W., Swallow, D. M. et al. (1990) Biochem. Biophys. Res. Commun. **171**, 407–415
54 Satsangi, J., Jewell, D. and Bell, J. (1997) Gut **40**, 572–574

Received 11 September 1997/26 November 1997; accepted 4 December 1997