Characterization of Human Mucin MUC17

COMPLETE CODING SEQUENCE AND ORGANIZATION

Received for publication, January 11, 2006, and in revised form, May 8, 2006 Published, JBC Papers in Press, May 31, 2006, DOI 10.1074/jbc.M600302200

Nicolas Moniaux1, Wade M. Junker1, Ajay P. Singh, Andria M. Jones, and Surinder K. Batra2

From the Department of Biochemistry and Molecular Biology, University of Nebraska Medical Center, Omaha, Nebraska 68198

With increasing interest on mucins as diagnostic and therapeutic targets in cancers and other diseases, it is becoming imperative to characterize novel mucins and investigate their biological significance. Here, we present the completed coding sequence and genomic organization of the previously published partial cDNA sequence of MUC17. Rapid amplification of cDNA ends with PCR, sequences from the Human Genome databases, and in vitro transcription/translational assays were used for these analyses. The MUC17 gene is located within a 39-kb DNA fragment between MUC12 and SERPINE1 on chromosome 7 in the region q22.1. The full-length coding sequence of MUC17 transcribes a 14.2-kb mRNA encompassing 13 exons. Alternate splicing generates two variants coding for a membrane-anchored and a secreted form. The canonical variable number of tandem repeats polymorphism of the central tandem repeat domain of the MUC genes is not significantly detected in the MUC17 gene. In addition, we show the overexpression of MUC17 by Western blot and immunohistochemical analyses in pancreatic tumor cell lines and tumor tissues compared with the normal pancreas. The expression of MUC17 is regulated by a 1,146-bp fragment upstream of MUC17 that contains VDR/RXR, GATA, NFkB, and Cdx-2 response elements.

Mucins, the main components of the mucus, are high molecular weight O-glycoproteins expressed and secreted by epithelial cells and, in some cases, by endothelial cells (1, 2). Their principal function is to protect and lubricate the epithelial surfaces; however (1), recent reports also demonstrate that mucins and, more specifically, membrane-bound mucins have a role in the signal transduction and oncogenic processes (1, 3, 4). Presently, 20 mucin (MUC)3 genes have been identified and are named MUC1–2, MUC3A/B, MUC4, MUC5AC, MUC5B, MUC6–13, and MUC15–20 (2, 5–8). These mucins have been grouped into two subfamilies, the secreted and the membrane-bound. The secreted mucins are exclusively expressed by specialized epithelial cells and exhibit a restricted pattern of expression within the human body (1, 2). The membrane-bound mucins are expressed at the apical region of epithelial cells under normal conditions and have a wide expression (1, 2). Moreover, alternative splicing and proteolytic cleavage can lead to the generation of three distinct forms of the transmembrane mucins, such as soluble (proteolytic cleavage of the membrane-bound form), secreted (alternatively splice variants), and one lacking the tandem repeat domain (alternatively splice variants) (9–12). The ratio of one form to another shows tissue specificity and is associated with the physiologic condition (13, 14).

MUC17, a membrane-bound mucin, was recently identified and located in the mucin cluster at the chromosomal locus, 7q22 along with MUC3A/B, MUC11, and MUC12 mucins (5, 15, 16). The first partial length cDNA sequence, now known to correspond to MUC17, was identified by Van Klinken et al. (17), who reported five tandem repeats, each encoding 59 amino acid residues, located upstream of the 17 tandem repeat residues of MUC3. Both sequences, repeated in tandem, were identified on the same cDNA fragment. However, after the characterization of the full-length sequence of MUC3A and MUC3B, the clone isolated by Van Klinken appeared as a chimera cDNA fragment, composed with an unknown gene sequence fused to the MUC3 tandem repeat sequence. In 2002, driven with the hypothesis that the five 59-amino acid residue tandem repeat sequences were part of a new unidentified mucin, Gum et al. (5) screened the public GenBankTM data base and the proprietary Lifeseq Gold data base (Incyte Genomics, Inc., Palo Alto, CA) and identified the 59-amino acid tandem repeat downstream sequence. The authors reported a partial cDNA fragment of 3,803 bp (accession number AF430017) composed of five repetitions of a 177-bp motif upstream of a non-repetitive sequence. The deduced amino acid sequence presented characteristics of membrane-bound mucin with the presence of five repetitions of the 59-amino acid residue motif, followed by two EGF-like domains, an/a SEA domain, a hydrophobic transmembrane domain, and an 80-amino acid long cytoplasmic tail. The new mucin gene, called MUC17, was localized to chromosome locus 7q22 along with MUC3A/B and MUC11/12.

Herein, we report the complete characterization of the MUC17 gene and its transcripts along with the deduced structural organization of the protein. Our study shows that MUC17...
is expressed in at least two alternatively spliced forms encoding for membrane-bound and -secreted forms. Moreover, interindividual VNTR polymorphism is also observed, giving rise to three allelic forms. Furthermore, we report that the intergenic region (1146 bp) between MUC12 and MUC17 possesses both basic and enhancer regulatory elements and may be responsible for cell-specific regulation of MUC17. A differential expression profile of MUC17 was observed in pancreatic tumors compared with the normal pancreas.

MATERIALS AND METHODS

**Tissue Specimens and Cell Lines**—A total of 24 established cancer cell lines (pancreas, colon, and breast) were used as sources of genomic DNA. Additionally, four genomic DNA samples were extracted from peripheral blood mononuclear cells of healthy individuals to validate the results obtained using the cancer cell lines. Samples were collected under protocol approved by the Institutional Review Board at the University of Nebraska Medical Center, Omaha, NE. Informed consent was obtained from all subjects.

*5’-Rapid Amplification of cDNA Ends PCR*—The 5’-RACE kit (Roche Applied Science) was used to synthesize first-strand cDNA from total AsPC-1 cell line RNA (2 μg) with specific MUC17 primer (RACE 171, GTGATAGCTCTGAACTGGCC). Terminal transferase was used to add a poly(dA) tail to the 5’-end of the cDNA. RACE-PCR experiments were performed in 50-μl reaction volumes containing 5 μl of 10× buffer (100 mM Tris/HCl/15 mM MgCl₂/500 mM KCl, pH 8.3), 5 μl of 10 mM deoxynucleoside triphosphates, 0.2 μM of each primer (MUC17-specific RACE 172, CATGGTGCTGGCAGGCATACT), oligo(dT)-anchor primer (provide by the RACE Kit supplier), and 2 units of Taq DNA polymerase (MBI Fermentas, Hanover, MD). The mixture was denatured at 94 °C for 2 min, followed by 30 cycles at 94 °C for 30 s, 60 °C for 1 min, and 72 °C for 2 min. The final elongation step was a 15-min extension. A 1-μl amplification product was generated using forward primer: GATATGGAC-CTGTTTGCTGGCAGGCATACT, reverse primer RACE 171 (provided by the RACE Kit supplier), and 2 units of Taq DNA polymerase (MBI Fermentas, Hanover, MD). The mixture was denatured at 94 °C for 2 min, followed by 30 cycles at 94 °C for 30 s, 60 °C for 1 min, and 72 °C for 2 min. The final elongation step was a 15-min extension. A 1-μl amplification product was generated by a second PCR reaction using a MUC17-specific nested primer (RACE 173, GTAGGAGATGAACTTGGCCTGA) and the PCR anchor primer (provided by the supplier). PCR products were electrophoretically resolved on 1% agarose gels and stained with ethidium bromide. Photographs were taken under UV light, using the GelExpert software (Nucleotech, San Carlos, CA). Amplification products were excised and purified with the QIAquick® gel extraction kit (Qiagen, Valencia, CA), cloned into pCR®2.1 vector (Invitrogen), and sequenced.

**Expand Long Template PCR**—To identify potential MUC17 splice variants in the 3’-extremity, an RT-PCR reaction was performed using the Expand Long Template PCR system (Roche Applied Science) with sense (5’-CTGTGGCCAAGAACACACAAATC'), and antisense primers (5’-CTCCTCCTACTTCCAGGACTTCTC'). Expand Long Template PCR was performed in 50-μl reaction volumes containing 5 μl of AsPC-1 cDNA, 5 μl of 10× buffer 3, 2.5 μl of 40 mM deoxynucleoside triphosphates, 0.2 μM of each primer, 0.75 mM MgCl₂, and 2.5 units of polymerase mixture (Roche Applied Science). The reaction mixture was denatured at 94 °C for 2 min, followed by 30 cycles at 94 °C for 30 s, 60 °C for 1 min, and 68 °C for 4 min with the elongation time of the last 20 cycles extended 40 s for each cycle. The final elongation step was extended for an additional 30-min period. The amplification product was directly cloned into the pCR®2.1 vector (Invitrogen), amplified, and sequenced.

**In Vitro Transcription and Translation Assays**—An amplification product was generated using forward primer 5’-GCC-AGCTCTCTGAGGGTGCAC-3’ and reverse primer RACE 171 (described previously). The product was cloned in pCR®2.1 under control of the T7 promoter. The DNA contained a coding region for a peptide with a predicted size of 36 kDa preceded by a putative Kozak sequence, followed by an ATG as well as 25-residue N-terminal signal sequence. Transcription and translation experiments were performed with the TnT® Quick Coupled Transcription/Translation System (Promega, Madison, WI) according to the manufacturer’s instructions. The amino acid mixture containing [35S]methionine (1000 Ci/mmol) was used for in vitro translation, and the product was analyzed by SDS-PAGE. Negative controls consisted of a MUC17 sequence cloned in the opposite direction and an empty vector. The β-lactamase gene was used as a positive control as recommended by the supplier.

**Southern Blot Analysis**—Genomic DNA, isolated from the 24 human tumor cell lines and peripheral blood mononuclear cells, from four healthy individuals, was digested with EcoRI and HindIII restriction endonucleases. Digested products were resolved by electrophoresis in 0.8% agarose gels and transferred to nylon membranes. Membranes were hybridized with a MUC17 tandem repeat probe. The probe (3 kb) was generated by PCR amplification using the MUC17-TR forward: GATATGACACACCTCTGACC and MUC17-TR reverse: ATGGTTGGTTCTTGGCCACAG primers, cloned in pCR®2.1, and sequenced before use. The probe was radiolabeled using the Random Primers DNA Labeling System (Invitrogen) and α-[32P]dCTP (BM Biomedicals, Irvine, CA).

**Assay of Transcriptional and Enhancer Activities**—Transient transfections of MUC17 promoter/enhancer luciferase reporter constructs were performed using the pGL3-basic and -enhancer vectors (Promega). Five constructs overlapping exon 11 of MUC12 to the MUC17 5’-untranslated region were prepared using AsPC-1 DNA. For transfections, cells were seeded into 6-well plates, in Dulbecco’s modified Eagle’s medium supplemented with 10% fetal bovine serum. Transfections were carried out using Lipofectamine (Invitrogen) according to the manufacturer’s instructions. Transfection conditions were optimized for each cell line. Cells were co-transfected with pSV-βGal vector (Promega) to control for variation in the transfection efficiency by assaying and normalizing data for β-galactosidase activity. After 2 days, cell lysates were made in reporter lysis buffer (Promega), and luciferase activity was determined using a Luciferase Reporter Assay System (Promega). Results were presented as -fold changes in luciferase activity as compared with the empty vector-transfected control cells.

**RESULTS**

**Identification and Characterization of the Central Repetitive Domain of MUC17**—To extend the MUC17 sequence toward its 5’-extremity, the known repeated 177-bp motif (character-
amplification product was, however, detected with a size of 10.8 kb, which is the expected size for the full-length tandem repeat domain.

Analysis of the BAC RP11–395B7 sequence suggested the presence of a HindIII site at 5434 bp upstream of the tandem repeat array and an EcoRI site at 1128 bp downstream of the repetitive sequence, encompassing a fragment of 17.4 kb (Fig. 1A).

Genomic DNA, purified from 24 cancer cell lines (pancreas, colon, and breast) and from four healthy individuals, was digested with HindIII and EcoRI endonucleases and probed with a MUC17 tandem repeat-specific probe. The 3-kb amplification product shown in Fig. 1B was cloned in PCR2.1 vector, sequenced, and used as a probe in Southern blot analysis on 28 genomic DNA samples. The probe contained 17 repetitions of the 177-bp motif. A low degree of VNTR polymorphism was observed, with only three bands detected with an approximate size of 17 kb (Fig. 1C). The size of the bands detected was consistent with the size of tandem repeats observed in the BAC clone (RP11–395B7) sequence which contained 60 repetitions of the 177-bp motif. These three bands were considered to result from various allelic forms of MUC17 (alleles A, B, and C), exhibiting VNTR polymorphism. Of the 24 cancer cell lines and 4 control DNA samples, 22 were homozygous for MUC17.

Two cell lines have the allelic form A (higher molecular weight), whereas the remaining 20 cell lines have the allelic form B (intermediate molecular weight). None of the cell lines have the allelic form C. Of the remaining six samples heterozygous for MUC17, allelic form B was common in all samples. Four of six samples were heterozygous with the allelic forms A and B, whereas the two other samples were heterozygous with the allelic forms B and C (lower molecular weight). The frequency of the allelic forms was 14.3% for the allele A, 82.1% for the allele B, and 3.6% for the allele C.

The average size of each allelic form (~17.0 kb) made it difficult to precisely measure which one is composed of 60 repetitions of the 177-bp motif. To answer this question, sequences within the NCBI, ENSEMBL, and USCS browser were analyzed. The MUC17 sequence reported in both NCBI and ENSEMBL sequence is based on the sequence with accession number AJ606307 as referred to in the current report. There-
fore, we cannot precisely assign one of the three allelic forms to the allele containing the 60 repetitions of 177 bp.

Identification and Characterization of the 5′-Extremity and Genomic Organization of MUC17—Combining the information deduced from the BAC clone RP11–395B7 sequence and the partial cDNA sequence of MUC17 (AF430017), the MUC17 gene was precisely located at chromosome 7 in the region q22.1, oriented from centromere to telomere, between the MUC12 and the SERPINE1 (serine proteinase inhibitor) genes (Fig. 2A). The 5′-RACE-PCR was performed on total RNA from AsPC-1, a pancreatic adenocarcinoma cell line that expresses a high level of MUC17, using three antisense primers localized in the degenerate sequence upstream of the tandem repeat array. Several amplification products were detected with size varying from 100 to 1000 bp for the first PCR, and from 200 to 700 bp after nested PCR (Fig. 3A). The detection of several amplification products during the 5′-RACE-PCR was expected for two main reasons. First, the MUC17-specific primers were localized in the degenerate sequence upstream of the tandem repeat array. Second, the size of the amplification products is directly dependent on the reverse transcription efficiency that gives rise to a multitude of cDNAs (partial or full copy of the mRNAs). Of these fragments, the largest cloned cDNA fragment (653 bp size) was sequenced. Its 3′-end overlapped the 5′-extremity of the degenerate repetition located upstream of the tandem repeat array. Comparison of the 5′-end of the RACE-PCR product with sequence of the BAC RP11–395B7 clone led to the identification of two new exons. The compiled nucleotide sequences of the RACE-PCR clone, the 177-bp tandem repeat of the BAC RP11–395B7, and the sequence identified and characterized by Gum et al. (AF430017) (5) allowed us to establish the complete organization of MUC17 (Fig. 2, B and C). This sequence has been deposited in EMBL (European Molecular Biology Laboratory’s Heidelberg, Germany) data base (AJ606307).

Altogether, the MUC17 gene (39 kb) encodes an mRNA of up to 14,221 bp in size (due to the VNTR polymorphism) after splicing of 13 exons ranging in size from 61 to 12,185 bp (Table 1). The size of the introns ranges from 121 to 10,902 bp. The largest exon, E3, encodes the central domain and is composed of 60 repetitions of a 59-amino acid motif repeated in tandem with residues rich in serine, threonine, and proline. A 25-amino acid signal peptide is found in the N-terminal extremity. The middle part of C presents the different sequences that were aligned to get the full-length sequence of MUC17. D, an alternative splice event occurs and skips exon 7 coding for a secreted form of MUC17, MUC17/SEC. MUC17/SEC lacks the MUC17 non-repetitive sequence located just upstream of the second EGF-like domain, as well as the second EGF-like domain, transmembrane domain, and cytoplasmic tail. The last 21 residues of the splice variant are specific to MUC17/SEC.
The Membrane-bound Mucin MUC17

A 5′-RACE and nested PCR

B Transcription/translation in vitro

C MUC17 alternatively spliced variants

**FIGURE 3.** Characterization of MUC17 N-terminal sequence and the alternatively spliced form. 

**A**, AsPC-1 RNA was reverse-transcribed using RACE171 primer and was 5′-poly(dA)-tailed using terminal transferase. Amplification used MUC17-specific primers and oligo(dT)-anchor primer as described under “Materials and Methods.” The 5′-RACE and nested RACE-PCR products were run on 1% agarose gels. Amplification products ranging from 100 to 1000 bp were detected. The most intense product was subcloned in the proper orientation. The 5′-untranslated region and sequence coding for the MUC17 signal peptide, MUC17 genes are transcriptionally independent. Exon1 of MUC17 contains the 5′-untranslated region and sequence coding for the MUC17 signal peptide.

**Full-length Coding Sequence of MUC17**—The translation initiation codon is located at position 54 in the sequence (AJ606307) and is preceded by non-consensus Kozak (18) sequence (AGAGCTCCGATG). A Kyte-Doolittle (19) hydrophathy plot of the N-terminal extremity of MUC17 shows that the initial 25 residues encoded by exon 1 are very hydrophobic. Analysis using SignalP V1.1 software from the Center for Biological Sequence Analysis (Technical University of Denmark) predicted the presence of a signal peptide of 25 amino acids with a cleavage site located between positions 25 and 26 (AAA-EQ). Fig. 2C shows a schematic representation of the MUC17 deduced amino acid sequence.

To confirm the functionality of the potential translation initiation site, the region upstream of the tandem repeat of MUC17 was amplified by PCR and subcloned in sense orientation downstream of the T7 promoter. The resulting construct was used for an in vitro transcription/translation assay (Fig. 3B). An expected 36-kDa protein corresponding to the MUC17 N-terminal region was detected. No protein product was detected for both negative controls. As a positive control, the N-terminal extremity of the β-lactamase gene was used (provided by supplier), which produced an expected 30-kDa protein. The structure of the 5′-extremity of human MUC17 is similar to the structure of rodent Muc3 (20). However, the N-terminal domain for MUC17 is coded by two exons, whereas for MUC3, it is coded by a single exon. Gum et al. (5) showed that the degree of sequence homology between the carboxyl-extremity of MUC17 and mMuc3 was higher than between MUC3 and mMUC3. It is suggested that MUC17 is the structural homologue of mMuc3. Fig. 4 presents an alignment of the N termini of MUC17, MUC3, and mMuc3. No similarity was detected between MUC3 and mMuc3, but a high degree of identity exists between MUC17 and mMuc3. Their similar structural organization and high degree of identity show that MUC17 is the human homologue of mMuc3.

**Alternative Splicing of the MUC17 Transcript**—The presence of one or more alternative splice events in the 3′-extremity of MUC17 was investigated by RT-PCR. For this purpose, a forward primer was chosen in exon 3 (tandem repeat domain) and a reverse primer in the 3′-untranslated region (position and sequence given under “Materials and Methods”). Using these primers, RT-PCR was carried out on AsPC-1 cDNA. The generated amplification products were cloned into pCR® 2.1 and sequenced. Two distinct fragments were identified through sequencing (Fig. 3C). One of the fragments was 100% identical to the previously referred sequence of MUC17 (accession number AJ606307). The second product revealed the occurrence of an alternative splice event that resulted from the skipping of exon 7. This alternative splice event generated a frameshift coding for 21 MUC17/SEC-specific amino acid residues and introduced by PCR on AsPC-1 cDNA. Two bands were detected; the higher molecular weight band corresponded to the full-length MUC17 sequence, whereas the lower one corresponded to a newly alternatively spliced form that lacks the second EGF domain, the transmembrane domain, and cytoplasmic tail (MUC17/SEC).
TABLE 1
Characteristics of the exon-intron junctions of the MUC17 gene

| Protein domain | Exon | 5'-Splice donor | Intron | 3'-Splice acceptor |
|----------------|------|-----------------|--------|--------------------|
|               | No.  | Name            | Size   | bp                 |
| 5'-UTR, leader sequence | 1 | ACAAGG | 136  | 1 136-137 10902 3 | tctcttcagACCTCA |
| N-terminal     | 2   | GGACAGG | 101  | 2 237-238 379 2 | tcctaaagFTTCTG |
| Central        | 3   | CAACAGG | 12,185 | 3 12456-12457 4163 2 | ttcacaggAGCTTG |
| EGF1           | 4   | CAATAGG | 132  | 4 12587-12588 729 2 | ccgcttcagGCCGAC |
| EGF1, SEA      | 5   | GGACAGC | 129  | 5 12715-12716 351 1 | tgctttcagATGGAAT |
| SEA            | 6   | GGTAGGC | 611  | 6 12775-12776 1101 3 | ccctctctcagCTTCTG |
| SEA, EGF2      | 7   | GTGCGAGG | 153  | 7 12996-12997 121 2 | ccctctctcagAGGACT |
| EGF2           | 8   | GTGCGGCG | 99  | 8 13156-13157 1023 3 | ccctctctcagCTGCTG |
| Transmembrane sequence | 9  | GAACAGG | 163  | 9 13318-13319 1913 2 | ccctctctcagGCAAAC |
| Cytoplasmic tail | 10 | GCAAGG | 11  | 10 13492-13493 1730 3 | ctcttttcagATGCCA |
| Cytoplasmic tail, 3'-UTR | 11  | GCAAGG | 77  | 11 136-137 10902 3 | tctcttcagACCTCA |

duced a stop codon positioned 66 nucleotides after the alternative splice site junction. The resulting protein is the secreted form of MUC17 (accession number AJ606308, MUC17/SEC), lacking the second EGF domain, the transmembrane domain and cytoplasmic tail. Several sets of primers were assayed along the 3'-extremity of MUC17 and RT-PCR carried out in four distinct cell lines (pancreatic AsPC-1 or colonist LS 174T, CaCo-2, and LS 180) (Fig. 3C). Two amplification products were detected. Sequencing of the major amplification product identified it as the MUC17 sequence described by Gum et al. (5) with the accession number AF430017, whereas the other amplicon (minor) corresponded to an alternatively spliced (skipping of exon 7) secreted form (AJ606308) of MUC17 (MUC17/SEC). The expression of MUC17/SEC was low in the cells investigated as compared with the major MUC17 membrane-bound form. The MUC17/SEC sequence was submitted to EMBL database (AJ006308) and also appears in the GenBank™ database.

Expression Analyses of MUC17 in Normal versus Diseased Pancreas—MUC17 expression was investigated in a panel of 2 normal pancreata, 8 pancreatitis, and 11 pancreatic adenocarcinoma tissue samples by RT-PCR. MUC17 was expressed in 81% of the tumor samples, whereas it was not detectable in either the normal pancrea or in the pancreatitis samples (Fig. 5A).

To confirm that MUC17 is detectable at the protein level, the 60 units of repetition composing the central domain of MUC17 were aligned to establish a consensus motif (Fig. 5B). A Hopp and Woods (21) hydrophilicity plot of this consensus sequence was carried out to delineate the antigenic region within these 59 amino acid residues. The synthetic peptide (showing the highest level of antigenicity), Pro-Thr-Thr-Ala-Glu-Gly-Thr-Ser-Met-Pro-Thr-Ser-Thr-Pro-Ser-Glu, was synthesized. An additional cysteine was added to the C terminus to boost antigenicity. Keyhole limpet hemocyanin was conjugated and served as a carrier for immunization of rabbits. Serum from the immunized rabbits showed a high antibody titer and good specificity by enzyme-linked immunosorbent assay, Western blot, and immunohistochemistry. Protein lysate from LS 174T and AsPC-1 cells served as a positive control (5), whereas lysate from PANC-1 cells served as a negative control. Immunoblot analysis revealed the presence of a single intense band for both AsPC-1 and LS 174T protein lysates, with an apparent molecular mass consistent with a protein of ~500 kDa (Fig. 5C). The peptide blocking study confirmed the specificity of the anti-MUC17 antibody (Fig. 5D). Using this polyclonal antibody, frozen sections of pancreatic adenocarcinomas were analyzed by immunohistochemistry (Fig. 5E). The staining of two tissue samples is shown in Fig. 5E. One tissue sample is strongly positive for MUC17 expression (Fig. 5E, panel 1), and another tissue sample is weakly positive (Fig. 5E, panel 2). In both cases, the malignant duct-like structures (main feature of well differentiated adenocarcinoma of the pancreatic gland) stained for MUC17 expression. The staining appeared to be concentrated at the luminal surfaces of the cells, consistent with the sub-cellular localization of MUC17 at the apical face of epithelial cells.

Characterization of the MUC17 Gene Regulatory Sequence—Having shown that MUC17 is abnormally expressed in pancreatic adenocarcinoma, further investigation of the MUC17 5'-flanking region DNA sequence was conducted to identify the regulatory sequence. No consensus promoter sequence was identified within this region with computer-based analysis, using both the PromoterInspector browser from Genomatix Suite 3.4.1 software (Genomatix, Munich, Germany) and the PROSCAN Version 1.7 program from the Center for Information Technology, National Institutes of Health. The 1146-bp
The Membrane-bound Mucin MUC17

DNA fragment located in between MUC12 and MUC17 was cloned in the pGL3-basic and -enhancer vectors. The fragment was generated by PCR using a forward primer that overlapped the last 10 nucleotides of MUC12 gene and a reverse primer overlapping the 11 first nucleotides of MUC17 gene. The corresponding amplification product was 1167 bp long. Four additional constructs were made that comprised the 775 bp fragment cloned in the pGL3-basic vector and was used in the assays for transcriptional activity (Fig. 6) (19). A 19-fold activation was detected for the 1167-bp intergenic region cloned in the basic vector in AsPC-1 cells, whereas 1.1- and 1.3-fold activations were detected for HPAF and PANC-1 cells, respectively. Hence, it can be inferred that the intergenic region possesses the basic promoter activity, which seems to be cell-specific. Interestingly, the intergenic region showed a very strong enhancer activity, with 300- and 110-fold activations measured for the AsPC-1 and HPAF cells, respectively. No enhancing activity was, however, detected in the PANC-1 cells for this region. The 2429-bp full-length fragment cloned in the basic vector also presented a 15-, 1.76-, and 1.12-fold basic promoter activity in the AsPC-1, HPAF, and PANC-1 cells, respectively. These results confirmed the presence of a promoter within the intergenic fragment. The results obtained using the MUC3 fragment were consistent with those described by Gum et al. (22), with a basic and enhancing activity detected.

DISCUSSION

Over the years, interest has increased in the study of the different mucin family members. When first investigated, mucins were identified to be associated with respiratory obstruction linked to the common cold or the flu. Now it is clear that the overexpression of mucins (23, 24) as well as the modification of the rheologic properties of the mucus (25) are also responsible for respiratory obstruction in patients with cystic fibrosis (26, 27). Currently, mucins and their implications in numerous disorders are widely being investigated for the development of early diagnostics (28) and/or therapeutics such as vaccines (29–32). For instance, the role of mucin members during malignant development and progression is starting to be well documented (3, 4). Mucins, specifically membrane-bound mucins, are thought to act through the tyrosine-kinase receptor to promote proliferation, with MUC1 acting with EGF receptor to promote proliferation, with MUC1 acting with EGF
receptor (33) and MUC4 through HER2 (34). CA125, the marker used to diagnose ovarian cancer, is a membrane-tethered mucin MUC16 (8). The full comprehension of mucin functionality in the development and progression of these diseases and their use for diagnostic and prognostic purposes requires the identification and characterization of the full-length mucin sequences.

One of the membrane-anchored mucins, MUC17, was identified by Gum et al. in 2002 (5). In the present study, the previously known MUC17 sequence was extended toward its 5' extremity to complete the sequence and localize the promoter and regulatory elements. MUC17 presents the classic architecture of the membrane-bound family members. MUC17 encompasses 13 exons. Its N-terminal domain is coded by 2 exons and possesses a leader sequence followed by a short unique sequence of 34 amino acid residues. Surprisingly, this sequence is not rich in serine, threonine, and proline but contains three cysteine residues. A BLAST search of this sequence using the Prosite recognition domains failed to identify any known functional domain within these 34 residues.

MUC17 translation initiation codon is surrounded by a non-consensus Kozak sequence with a C for an A/G in the position −3 (18, 35) and a C for a G in the +4 position. This type of non-consensus Kozak sequence is found in <1% of all known mammalian genes and may be 10-fold less efficient in initiating translation in comparison to the consensus Kozak sequence (36). Other mucins, such as the MUC3A and MUC3B, clustered with MUC17 on chromosome 7q22, also have similar non-consensus Kozak sequences (22).

MUC17 contains a large domain composed of at least 60 repetitions of a 59-amino acid residue motif in its central region, which is followed by a sequence of degenerated repeats. The C-terminal sequence is composed, as described by Gum et al. (5), of two EGF-like domains, a SEA (sea-urchin sperm protein, enterokinase, and agrin) domain (37, 38), a transmembrane sequence, and an 81-amino acid residue cytoplasmic tail. A comparison of the MUC17 N-terminal amino acid sequence with the MUC3 and mMuc3 counterparts reveals a high degree of identity between MUC17 and mMuc3 (over 55%). However, no homology was detected between MUC17 and the recently identified MUC3 amino acid sequence (22). At the gene level, MUC3 possesses a single exon at the 5’-extremity, whereas MUC17 and mMuc3 have two exons. Gum et al. (5) reported that the C-terminal sequence of MUC17 is more similar to rat Muc3 and mouse Muc3 than to any other known human protein. Because of the high degree of similarity between both the proteins, and their identical genomic organization, we suggest that rodent Muc3 should be referred to as rat Muc17 and mouse Muc17. We will use the name of Muc17 for clarity reason within the remaining of the text.

Additionally, we examined the expression profile of MUC17...
The Membrane-bound Mucin MUC17

in (Fig. 5) by utilizing polyclonal antibodies generated against the tandem-repeat domain. The advantage of choosing a sequence within the repetitive domain is that the antibodies will recognize multiple epitopes on the MUC17 protein resulting in enhanced sensitivity. Conversely, the limitation with such a choice is that the epitope(s) can be masked in vivo due to glycosylation. However, our previous experience with antibody generated against the tandem repeat region of MUC4, which exhibits high sensitivity (39), encouraged us to use the MUC17 tandem repeat peptide as an immunogen. Moreover, several studies have shown that, in cancer conditions, MUC1 and MUC4 are present in underglycosylated forms at the cell surface (40, 41). Furthermore, the anti-tandem repeat MUC antibody has been shown to recognize mucins in normal physiological conditions as well (39). It is strongly suggested that rodent Muc17, like its human counterpart, goes through a recycling step to be fully glycosylated. The existence of membrane-bound mucins under two glycoforms (hypo- and fully glycosylated) allows us to use the sequence with the tandem repeat array to design a mucin-specific antibody. Even if the antibodies cannot recognize the mature, fully glycosylated mucins, they should react against the hypoglycosylated forms.

In 2001, Khatri et al. (42) showed that rodent Muc17 (or Muc3) was expressed under two forms, a membrane-associated and a soluble form. They showed that the soluble form was produced after proteolytic cleavage of rodent Muc17 in between the second EGF and the transmembrane domains (42). Furthermore, they reported that the columnar cells predominantly expressed the membrane-bound form, whereas the goblet cells uniquely expressed the soluble form of Muc17. The differential pattern of expression of transmembrane and soluble forms of Muc17 was proposed to result from alternative splicing of Muc17 transcript. Our data are consistent with the proposed hypothesis revealing the transcript for the soluble form of MUC17 (MUC17/SEC), however, MUC17/SEC appears to be expressed at a very low level in the cells and needs further examination.

Polycistronic mRNAs are very rare in eukaryotic cells but have been identified, for instance, in humans (43), mice (43), and Drosophila (44). In humans and mice, both UOG-1 and GDF-1 are clustered on the same chromosomal locus, separated by 269 bp (human) and 404 bp (mouse) (43). Interestingly, the first exon of MUC17 was located 1146 bp from the last exon of MUC12. Because the gap between MUC12 and MUC17 was of only 1146 bp, PCR was conducted on genomic DNA and cDNA derived from the total RNA in the MUC3, MUC12, MUC17-positive cell line AsPC-1. The expected amplification product was detected in reactions that used genomic DNA; however, no amplification product was detected with cDNAs as templates. This observation ruled out the extreme possibility that MUC12 and MUC17 might be transcribed as a polycistron.

Having shown that both MUC12 and MUC17 genes are transcribed as independent mRNAs, we searched for a consensus promoter with in the 1146-bp intergenic region. No consensus promoter region was identified, and the only sequence recognized as a potential promoter for MUC17 was localized within the last exon of MUC12. The 1146-bp intergenic fragment presented both basic promoter and enhancer activities, whereas the MUC12 intron 11 did not. As depicted in Fig. 6, the MUC17 fragment possessing promoter activity harbors numerous transcription factor binding elements, including GATA, VDR/RXR, Cdx-2, NFκB, and Pdx-1. Interestingly, intron 11 of MUC12 is composed of at least 21 repetitions of the same sequence repeated in tandem, which consist mainly of VDR/RXR, NFκB, and Gut-enriched Kruppel-like transcription factor recognition sequences. Further experiments need to be performed to show if this fragment carries enhancer activity for MUC17 after stimulation with vitamin D.

MUC12 was first identified by Williams et al. (15) by a differential display as down-regulated in colorectal cancer. MUC12 is strongly expressed in the normal colon but down-regulated in colon cancer. As was the case for MUC17, just the C-terminal sequence of MUC12 is known and presents MUC12 as a membrane-bound protein with two EGF-like domains. Altogether, the ratio of the 7q22 mucin could be an indicator of the physiological condition of a given tissue, normal versus malignant.

In conclusion, our study provides the completed coding sequence for the MUC17 mucin. MUC17 is expressed in two forms, membrane-associated and secreted. Unlike other MUC genes, there is a very low level of polymorphism in the VNTR region of MUC17. The overexpressed MUC17 is detected in pancreatic cancer cells compared with the normal pancreas. In the future, the regulation of MUC17 expression will be explored for investigating its significance in cancer and under normal conditions.

Acknowledgments—We thank Erik Moore and Allison Rahde for excellent technical support and Kristi Berger for editing the manuscript. We also thank the director of the Tissue Bank at University of Nebraska Medical Center, Dr. Julia Bridge, for providing tissue samples. The Comparative Human Tissue Network, Western Division, Case Western Reserve University, Columbus, OH, is greatly acknowledged for providing the additional tissues for this study.

REFERENCES

1. Hollingsworth, M. A., and Swanson, B. J. (2004) Nat. Rev. Cancer 4, 45–60
2. Moniaux, N., Escande, F., Porchet, N., Aubert, J. P., and Batra, S. K. (2001) Front. Biosci. 6, D1192–D1206
3. Moniaux, N., Andrianifahanana, M., Brand, R. E., and Batra, S. K. (2004) Br. J. Cancer 91, 1633–1638
4. Singh, A. P., Moniaux, N., Chauhan, S. C., Meza, J. L., and Batra, S. K. (2004) Cancer Res. 64, 622–630
5. Gum, J. R., Jr., Crawley, S. C., Hicks, J. W., Szymkowski, D. E., and Kim, Y. S. (2002) Biochem. Biophys. Res. Commun. 291, 466–475
6. Higuchi, T., Orita, T., Nakanishi, S., Katsuya, K., Watanabe, H., Yamashita, Y., Waga, I., Nanayama, T., Yamamoto, Y., Munger, W., Sun, H. W., Falk, R. J., Jennette, J. C., Alcorta, D. A., Li, H., Yamamoto, T., Saito, Y., and Nakamura, M. (2004) J. Biol. Chem. 279, 1968–1979
7. O’Brien, T. J., Beard, J. B., Underwood, L. J., Dennis, R. A., Santin, A. D., and York, L. (2001) Tumour Biol. 22, 348–366
8. Yin, B. W., and Lloyd, K. O. (2001) J. Biol. Chem. 276, 27371–27375
9. Baruch, A., Hartmann, M., Zrihan-Licht, S., Greenstein, S., Burstein, M., Keydar, I., Weiss, M., Smorodinsky, N., and Wreschner, D. H. (1997) Int. J. Cancer 71, 741–749
10. Choudhury, A., Moniaux, N., Winpenny, J. P., Hollingsworth, M. A., Aubert, J. P., and Batra, S. K. (2000) J. Biochem. (Tokyo) 128, 233–243
11. Choudhury, A., Moniaux, N., Ringel, J., King, J., Moore, E., Aubert, J. P., and Batra, S. K. (2001) Teratog. Carcinog. Maten. 21, 83–96
12. Crawley, S. C., Gum, J. R. J., Hicks, J. W., Pratt, W. S., Aubert, J. P., Swallow,
