Initiation of Transcription of the MUC3A Human Intestinal Mucin from a TATA-less Promoter and Comparison with the MUC3B Amino Terminus*

Received for publication, June 2, 2003, and in revised form, September 3, 2003
Published, JBC Papers in Press, September 4, 2003, DOI 10.1074/jbc.M305769200

James R. Gum, Jr.,‡‡, James W. Hicks‡, Suzanne C. Crawley*, Christine M. Dahl‡, Stacey C. Yang‡, Anthony M. Roberton**‡, and Young S. Kim‡‡

From the *Gastrointestinal Research Laboratory, Department of Veterans Affairs Medical Center, San Francisco, California 94121 and the Departments of ‡Anatomy, ‡‡Medicine, and ‡¶Pathology, University of California at San Francisco, San Francisco, California 94121

Human intestinal mucin genes MUC3A and MUC3B are members of a membrane mucin gene family residing at chromosome 7q22. In this paper, we utilized genomic and cDNA cloning to elucidate the sequence of the 5’-region of the MUC3A gene including the gene promoter and the amino terminus coding sequences. Following its 21-residue signal peptide, the amino terminus of the mucin consists of a 233-residue Thr-, Ser-, and Pro-rich nonrepetitive sequence that is contiguous with its hypervariable domain of 375-residue repeats. RNase protection analysis and 5’-GeneRacer PCR indicated that MUC3A gene transcripts initiate from multiple start sites along a region spanning ~180 bases. The 5’-flanking region of the gene had promoter activity when fused to a luciferase reporter gene in all of the tested cell lines. This region contained binding sites for several transcription factors, including those implicated in the regulation of intestinal genes, but lacked a cognate TATA box. These features of the gene promoter may enable the gene to be expressed at variable levels in several cell types with different repertoires of transcription factors. We also utilized 5’-GeneRacer PCR to determine the sequence of the 5’-terminus of the MUC3B message. The amino termini of the MUC3A and MUC3B mucins are 91% conserved at the amino acid level. Thus, MUC3A and MUC3B have highly conserved amino and carboxyl termini, suggesting a recent duplication of the entire ancestral gene. It remains to be determined whether other members of the 7q22 membrane mucin gene family have amino-terminal domains similar to MUC3A and MUC3B.

Membrane mucins are an important class of glycoproteins with diverse structures and functions. These molecules contain extracellular domains that serve as a scaffold for O-glycosylation. The O-glycans associated with membrane mucins are thought to function in cytoprotection and have been demonstrated to confer anti-adhesion properties upon cells (1). This latter characteristic may play a role in the dissemination and spread of cancer cells. In addition to conferring these electrostatic/physical properties upon cells, membrane mucins can anchor carbohydrate moieties with specific functions. Selectin ligands associated with membrane mucin glycans, for example, play a role in cancer cell extravasation during metastases (2). Certain membrane mucins function in signal transduction as well (3–5). Several membrane mucins also serve as clinically important tumor antigens (6, 7).

All membrane mucins contain O-glycosylated extracellular domains; however, the size of these domains varies widely, from a few hundred amino acid residues to several thousand residues (8–11). The composition and structure of other domains of these molecules also varies, suggesting different functions. Moreover, the alternative splicing of several membrane mucins leads to forms with deleted domains and to truncated forms lacking membrane-binding domains, rendering them soluble (12–15).

The membrane mucin gene cluster located at human chromosome 7q22 contains at least four members including MUC3A, MUC3B, MUC17, and MUC11/12 (which may be one or two genes) (16, 17). These mucins contain large extracellular glycosylation domains that exhibit variable number of tandem repeat (VNTR) polymorphism. They also contain extracellular EGF1-like domains, a transmembrane domain, and a cytoplasmic domain. We have focused much attention upon the MUC3 mucins, MUC3A and MUC3B (12, 16, 18, 19). The extracellular regions of these glycoproteins both contain two polymorphic tandem repeat domains, one containing 375-residue tandem repeats and another containing 17-residue tandem repeats (19). The carboxyl termini of these mucins contain a mucin-like (Ser/Thr-rich) unique sequence, two EGF-like domains separated by a SEA module, a transmembrane domain, and a cytoplasmic domain (12, 16). The exonic and intronic sequences of the carboxyl termini of MUC3A and MUC3B are ~95% identical, suggesting a recent gene duplication (16).

In this report we describe the sequences of the MUC3A gene 5’-flanking region and amino terminus, thus completing the MUC3A sequence. The coding sequence of the 5’-region of MUC3A consists of a signal peptide and a mucin-like but non-repetitive amino terminus domain. Transcription is initiated from a TATA-less promoter at multiple start sites. We also describe the homologous coding region for MUC3B, which is highly conserved, indicating that the entire ancestral gene was duplicated.

* This work was supported by the Department of Veterans Affairs Medical Research Service. 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.

‡ To whom correspondence should be addressed. Tel.: 415-750-2095; Fax: 415-750-6972; E-mail: jgum@maelstrom.ucsf.edu.

** Present address: University of Auckland, Private Bag 92019, Auckland, New Zealand.

† The abbreviations used are: EGF, epidermal growth factor; contig, group of overlapping clones; SEA, domain found in sea urchin sperm protein, enterokinase, and agrin; GSS, genome survey sequence.
**EXPERIMENTAL PROCEDURES**

In this paper, the term MUC3 refers to a clone derived from MUC3A or MUC3B or a hybridization probe or procedure capable of detecting both genes and their derived products. This terminology is used because almost all clones obtained in previous studies could have been derived from either gene and will hybridize to either gene (16). The human small intestinal cDNA library used previously to obtain MUC2 and MUC3 clones (18–20) was screened using SIB227 as a hybridization probe. SIB227 consists entirely of sequence derived from a 1125-bp MUC3 tandem repeat domain (19). Hybridizing phage were plaque-purified, DNA was isolated, and inserts were retrieved by EcoRI digestion. DNA was sequenced following its cloning into pBluescript. All of the exonic sequence and most of the noncoding sequence were obtained using manual sequencing methods; however, some of the noncoding sequence was determined by automated sequencing performed by the University of California at San Francisco Cancer Center Genome Analysis Facility.

RNA was extracted from human LS174T colon cancer cells and from normal duodenal biopsy samples pooled from 10 patients using TRI reagent as described (21). DNA was extracted from the lymphocytes of individual donors using previously described procedures (21). RNA blot analysis was performed using SIB124 as a probe (18). Reverse transcription and PCR were performed as described (22).

Clones representative of 5′-termini of MUC3A and MUC3B messages were obtained using a GeneRacer kit purchased from Invitrogen. The protocol accompanying the kit was used essentially without modification. Briefly, 2 μg of total RNA was treated with calf intestinal alkaline phosphatase and subsequently with tobacco acid pyrophosphatase. This procedure is designed to leave a 5′-phosphate moiety only on RNAs that contain a cap structure, thus selecting for 5′-full-length messages. This material was then joined to the Gene Racer 5′-RNA oligonucleotide (5′-GACUGGAGGACGAGACACAGACUGACUGAACGAGUGAGAA) using RNA ligase. The ligated mRNA was then reverse transcribed using Superscript II reverse transcriptase and a gene-specific deoxyribonucleotide primer, 5′-GGTTGGAGTGGTGGTCCAAC (reverse complement of bases 2558–2576 in Fig. 1B). The initial PCR utilized GeneRacer 5′-primer (5′-CGACTGGACGAGACGACGATCGA) and 5′-TGAAGGCAAACCTGAG as gene-specific reverse primer (reverse complement of bases 2538–2555 in Fig. 1B). PCR, when performed, utilized GeneRacer 5′-nested primer (5′-GACACTGGACGAGACGACGATCGA AGGACTA) and 5′-AGTGTTGGACCTGAGCTGG as gene-specific nested primer (reverse complement of bases 2507–2355 in Fig. 1B). PCR products were cloned into pCR4-TOPO vector for sequencing.

RNA protection mapping of transcription start sites was performed using kits (RPA II and MAXIScript) purchased from Ambion. Templates for riboprobes were prepared by PCR using deoxyribonucleotide primers 5′-GAGGAAACGCGACCTTCCGAG (bases –278 to –259 in Fig. 1D) and 5′-ACGGGGGAGGGGCTGG (reverse complement of bases 35–51 in Fig. 1B). DNA isolated from human chromosome 7 BAC clone CTC-786K12 (purchased from Research Genetics, Inc.) was used in this reaction. The ampiclon was ligated into pCR2.1, and clones containing insert in both orientations were identified by sequence analysis. The templates were linearized by digestion with BamHI, and the reactions were stopped and made RNAse-free by digestion with proteinase K. The material was then extracted with chloroform/phenol and then chloroform and precipitated with sodium acetate/ethanol. After drying, the templates were dissolved in diethylypyrocarbonate-treated water. The riboprobes were synthesized from 1 μg of template using the protocol described in the MAXIScript kit. To achieve high specific activity labeling, 2.0 μCi unlabeled CTP was used in conjunction with 100 μCi of [α-32P]CTP (specific activity > 3000 Ci/mmol; ICN) in a final volume of 20 μl. The ribonuclease protection reactions were conducted using 25 μg of total RNA, 100,000 cpm riboprobe, and the protocol described in the RPA II kit. The final products were subjected to electrophoresis on a standard sequencing gel, using the products derived from a sequencing reaction as size standards as described (20).

**RESULTS**

The MUC3A Amino Terminus—To obtain MUC3A or MUC3B amino terminus clones, the human intestinal cDNA library was first screened using SIB227 as a hybridization probe. Approximately 20 clones were obtained, successfully plaque-purified, and sequenced. Many of these clones consisted entirely of elements of the 1125 base tandem repeat domain as does SIB227. Some were chimeras with unrelated DNA fragments as commonly occurs when cloning repetitive cDNAs (19). However, one clone designated SIB275, encoded a mucin-like unique sequence upstream of and in frame with SIB227. Thus, SIB275 was considered a likely candidate as an amino terminus MUC3A or MUC3B cDNA clone.

We obtained a genomic clone containing MUC3A sequence by BLAST searching the public data base using SIB227 as query sequence. BAC clone CTC-786K12 was identified as containing this sequence at one terminus, reading in a 3′ to 5′ direction. The other terminus of this clone contained sequence identifiable as part of the zonadhesion gene. Thus, we concluded that this clone should contain the 5′-portion of the at least one of the MUC3 genes and associated flanking sequencing. (Later work identified this clone as containing MUC3A (16).) This clone was therefore purchased from Research Genetics, Inc. to enable additional sequencing. We sequenced a 14-kb segment of this clone initiating from the terminus containing SIB227 sequence, a portion of which is shown in Fig. 1. Later, clones and contigs confirming and extending this sequence became available in public data bases (GenBankTM accession numbers AC118759 and BL000001).

The clone CTC-786K12 sequence obtained confirmed the sequence of cDNA clone SIB275. By also extending the sequence, it revealed the presence of a possible initiation codon and signal peptide (Fig. 1). A single intron of 2167 bases was present in the genomic sequence, beginning at the termination point of...
FIG. 1. Sequence of the amino terminus of MUC3A. A, diagram depicting the cloning strategy employed to enable sequencing of the 5'-region of the MUC3A gene. Both the MUC3A message and gene are represented (at different scales). The positions of cDNA clones SIB275, SIB236, SIB324, SIB321, and SIB327 are represented by solid lines above the message. Different domains of the MUC3A message are depicted. The 5'-untranslated region initiates from multiple start sites as described in the text and therefore varies in length. The BAC clone CT786K12 initiates in the second exon of MUC3A and contains the 5'-portion of the gene and its 5'-flanking region, which resides on chromosome 7q22. The location of the primers used to perform GeneRacer PCR to obtain message sequence is also given. The initiation points for the clones obtained are indicated by boxes above the sequence, and those derived from butyrate-treated LS174T cell RNA are indicated by circles. The number of clones obtained initiating from a given base is indicated by the number within the box or circle. Solid dots above the sequence indicate major transcription start sites as determined by RNase protection.

B. R.T., reverse transcription primer. B, genomic DNA and message sequence of the 5'-region of MUC3A. Potential transcription factor binding sites were determined using TFSEARCH and MatInspector web sites (150.82.196.184/research/db/TFSEARCH.html and www.genomatix.de; Refs. 23 and 24). The initiation points for GeneRacer clones derived from butyrate-treated LS174T cell RNA are indicated by boxes above the sequence, and those derived from duodenal RNA are indicated by circles. The number of clones obtained initiating from a given base is indicated by the number within the box or circle. Solid dots above the sequence indicate major transcription start sites as determined by RNase protection.
Double underlines indicate a signal sequence (25). The start site of cDNA clone SIB275 is indicated by an arrow; the sequence terminates at the end of this cDNA clone. Splice sites demarcating the single intron are also indicated by arrows, as are the boundaries of the mucin-like unique sequence and the 375-amino acid tandem repeat array. The sequence was submitted to GenBank with accession number AY307930.
the putative signal peptide. It was not certain from this data, however, that the MUC3A amino terminus had been reached. A first approach in examining this further was to screen the human intestinal cDNA library with a probe made by PCR from the first 250 bases of clone SIB275. Several clones were obtained, and the ends were sequenced (Fig. 1A). Two of these clones, SIB324 and SIB326, extended the SIB275 sequence read by 64 and 37 bases in the 5'-H11032-direction, respectively. The fact that cDNA clones derived from further upstream genomic sequences were not obtained in this screening suggested that these clones may represent the 5'-terminal region of the message. To examine this still further, 5'-GeneRacer PCR was conducted. Previous work demonstrated that duodenal RNA and butyrate-treated LS174T cell RNA contained high levels of MUC3 message (17); thus, these were used for this study. We were able to obtain product visible on an ethidium bromide-stained agarose gel after initial PCR with duodenal RNA; however, butyrate-treated LS174T cell RNA required that nested PCR be employed. A total of 16 clones were obtained with butyrate-treated LS174T cell RNA, of which 15 were similar to sequence shown in Fig. 1B. Interestingly, however, these clones initiated in different places along a 179-base segment as shown in Fig. 1B. A total of 15 clones were obtained from the duodenal RNA. Five of these clones had slightly different sequences and appear to be derived from MUC3B, as will be described later. The other 10 were derived from MUC3A, and the points at which they initiate are depicted in Fig. 1B. Three of these clones were derived from initial PCR, and seven were from nested PCR. Multiple start sites were also observed with duodenal RNA.

The experimentation described above suggested that MUC3A transcripts initiate from multiple start sites. This possibility was tested using RNase protection analysis. We have used this technique previously to map the start sites of MUC2 (20), DPPIV (26), and a MUC2 promoter/growth hormone transgene (21). Preliminary results indicated that the technique was not sensitive enough to map protected fragments utilizing butyrate-treated LS174T cell RNA (data not shown). The fact that GeneRacer PCR was successful after initial PCR with duodenal RNA but required nested PCR with butyrate-treated LS174T cell RNA suggested that duodenum is the richer source of MUC3A message. We therefore directed our efforts entirely upon utilizing duodenal RNA to map start sites by RNase protection analysis. We were able to detect protected fragments utilizing this technique with 25 µg of duodenal RNA and the high specific activity probe described under “Experimental Procedures.” Multiple protected fragments were obtained with the antisense probe when duodenal RNA was used but not with the tRNA control (Fig. 2). In contrast, no bands were observed with the sense probe with RNA from either source. The actin control provided with the kit gave a single major band at the expected location (Fig. 2). The protected bands observed with duodenal RNA indicated the presence of eight major start sites for MUC3A transcription. Their locations, which span 141 bases, are indicated in Fig. 1B. In addition, many bands of lower intensity are evident as well, suggesting that MUC3A transcription can initiate from many start sites, in addition to the major ones. Thus, RNase protection analysis confirms the notion drawn from GeneRacer PCR that MUC3A transcripts initiate from multiple start sites.

Promoter Activity of the MUC3A Gene 5'-Flanking Sequence—The genomic region flanking the MUC3A transcription start sites contains the consensus sequences for many transcription factor-binding sites as depicted in Fig. 1B. Many of these elements including Cdx-2, GATA, USF, CACCC, and HNF-1 have been shown to be present within the promoters of genes expressed in the intestine (26–29). Noticeably absent, however, is a cognate TATA box, an important element that binds transcription factor IID, orienting and nucleating the transcription initiation complex of many tissue-specific genes.
sites should produce luciferase activity in transient transfection assays. Constructs terminating upstream of the first transcription start site would not be expected to be active, however. This follows from the observation that transcripts are not initiated from upstream sequences in vivo.

These constructs were then transfected into four pancreatic and three colon cancer cells, and luciferase was assayed as a measure of promoter activity (Fig. 3B). The constructs were active in all of the tested cell lines; however, the relative levels of luciferase activity obtained varied markedly in magnitude. The pattern of activity elicited by the various constructs was remarkably similar in all tested lines. The constructs that terminated downstream of the most distal start site (−775/+57 and −623/+57) produced high levels of luciferase activity. Those that terminated within the region of start sites (−775/−62 and −623/−62) had intermediate levels of activity. Most importantly, however, constructs that terminated upstream of the first known MUC3A start site (−775/−242 and −623/−242) produced only minimal levels of luciferase activity. Little difference was observed whether the constructs initiated at base −775 or base −623. These results demonstrate that the MUC3A 5'-flanking region has functional promoter activity, but for this to be manifested a region containing transcription start sites must be included.

Comparison of MUC3A and MUC3B Amino Terminus Sequences—As described above, five clones from 5'-GeneRacer PCR with duodenal RNA were noted that had slightly different sequences from the MUC3A sequence described in Fig. 1B. We hypothesized that these clones may represent MUC3B sequence but also considered the possibility that they could be an allelic variant of MUC3A. As a test of these possibilities, we posed the question: Is the putative MUC3B sequence invariably present in the DNA of individual donors? If so, then this would indicate that the sequence represented a second gene rather than an allele of MUC3A. To amplify the sequences, we chose forward and reverse PCR primers from DNA segments in which both sequences were identical (forward, 5'-CTTTATC CACGCCCAATCC (bases 2289–2308 in Fig. 1B) and reverse, 5'-AGGCAAAACGTGACGGGG (reverse complement of bases 2533–2552 in Fig. 1B)). Genomic DNA samples from the lymphocytes of eight donors were used for the reactions. The products were then digested with a series of restriction enzymes chosen because they digest the MUC3A sequence or the putative MUC3B sequence but not both sequences (Fig. 4A). In this way we were able to test whether the individual samples contained both sequences. Results from two representative samples are shown in Fig. 4B. The size of the uncut PCR product was 263 bases. BfaI and HgaI are expected to digest the MUC3B sequence but not the MUC3A sequence, as described in Fig. 4A. This occurred in all of the tested samples, examples shown in Fig. 4B. HhaI, on the other hand, is expected to digest MUC3A but not MUC3B. This result was also obtained with all of the samples. As a control, we digested DNA isolated from BAC clone CTC-7866K12 with these enzymes. This was digested with HhaI but not with BfaI or HgaI (Fig. 4B), the expected result because this clone contains MUC3A sequence. These results indicate that all eight tested genomic DNA samples contain both the MUC3A and MUC3B sequences. This result is consistent with the MUC3B sequence representing a separate gene, because it is highly unlikely that all samples would contain two alleles of the same gene, even if the gene frequency was 0.5 for both alleles (probability = 0.5^2 < 0.004).

We performed reverse transcription-PCR to amplify the complete MUC3B amino terminus for sequencing. Duodenal RNA was used for reverse transcription in conjunction with random primers. For the forward PCR primer we chose a sequence

![Figure 2](image-url)
present in MUC3B but not MUC3A (5'-GATGCTCAAGTC-CTCCCCAG). For the reverse primer we chose a highly conserved segment within the 1125-base tandem repeat domain (5'-GGTGGTCTCAGATGTAGGTGTAG). This generated a 1.0-kb fragment, which was ligated into pCR2.1 for sequence analysis. The sequence confirmed and extended that of the GeneRacer PCR clones (Fig. 5). There are four changes in the 21-residue signal peptide. The 234-residue mucin-like unique sequence that constitutes the amino terminus domain of MUC3B is 91% conserved with that of MUC3A, with a single 1-residue insertion being present. The MUC3A and MUC3B amino termini are both contiguous with 375-residue tandem repeat domains described earlier (19). Thus, the amino termini of MUC3A and MUC3B are highly conserved, as are the carboxyl termini of the two glycoproteins (16).

**DISCUSSION**

MUC3A and MUC3B are highly conserved members of the membrane mucin gene family residing on chromosome 7q22. The initial cDNAs encoding these genes were thought to represent a single, secretory mucin (18, 19). Later work demonstrated existence of the two separate, closely related genes (16). Moreover, the most common splice variants of these genes were shown to encode membrane-bound mucins (12, 32). These
genes contain repetitive sequences that often rearrange or form chimeras during cloning (19). This has hampered efforts to obtain their sequences by both manual and automated methods. Only recently has the sequence of a genomic clone containing the amino terminus of \( MUC3A \) become available in the public data bank (BAC clone RP13-650G11; GenBank\textsuperscript{TM} accession number AC118759). This sequence contains a large deletion encompassing both the 1125-base and the 51-base tandem repeats domains. Errors in mucin sequences in genomic data banks may be due to the software employed by the human genome project, which often handles repetitive sequences by deleting them from assembled contigs (33). There are still no clones available in the public data bank for the \( MUC3B \) amino terminus or promoter. Thus, the sequencing of these genes has been difficult, and the sequences presented here for the \( MUC3A \) and \( MUC3B \) amino termini represent the first complete coding sequence available for 7q22 family mucins.

Several types of evidence support the argument that the conceptual translations shown in Figs. 1 and 5 represent the true amino termini of \( MUC3A \) and \( MUC3B \): (i) 5'-GeneRacer PCR revealed no sequences initiating upstream of those shown in Fig. 1. (ii) There is an in-frame stop codon at base +19 of the sequence, 40 bases upstream of the putative initiation codon. There is also a potential 3'-splice site at base +38, and we considered the possibility that this may be functional. This does not appear to be the case, however, because none of the 30 5'-GeneRacer \( MUC3A \) and \( MUC3B \) clones that we sequenced was spliced at this site. (iii) Attempts to isolate upstream clones by screening the intestinal cDNA library with a probe derived from the upstream region of SIB275 were unsuccessful. (iv) No ATG triplets occur in the \( MUC3A \) message sequence upstream of the proposed initiation codon. (v) The translated sequences were subjected to analysis for the presence of signal peptides using the SignalP program (www.cbs.dtu.dk; Ref. 25). This analysis strongly indicated the presence of 21-residue signal peptides for both \( MUC3A \) and \( MUC3B \). Downstream of the signal peptide, the \( MUC3A \) and \( MUC3B \) amino termini unique sequences are mucin-like, containing more than 40% Thr, Ser, and Pro residues. They do not, however, contain repetitive elements, nor do they have sequence similarity with any other known protein. It is interesting that the initiation codons for both \( MUC3A \) and \( MUC3B \) occur in the same DNA context, i.e. GGCCCATGC, which deviates at the -3 position (underscored) from the consensus sequences for initiation codons. Only 3% of eukaryotic messages have pyrimidines at this position (34). The significance of this deviation from the consensus in \( MUC3A \) and \( MUC3B \) messages is not known.

The \( MUC3A \) upstream sequence contains binding sites for FIG. 4. Presence of both \( MUC3A \) and \( MUC3B \) sequences in the DNA of multiple human donors. A, size of restriction bands predicted in the 263-base PCR fragment amplified from either the \( MUC3A \) or \( MUC3B \) genes. B, restriction digestion of the PCR amplicon from the DNA of two representative donors. Digestion of the amplicon obtained using DNA isolated from BAC clone CTC-786K12, which contains the \( MUC3A \) sequence only, is included as a control.

**Fig. 5.** Comparison of the amino termini of \( MUC3A \) and \( MUC3B \). The MacVector 6.5 ClustalW routine was used to perform the alignment; identical residues are boxed. Arrows demarcate the boundaries of the signal peptides, the mucin-like unique sequences, and the 375-residue tandem repeat array. The \( MUC3B \) cDNA sequence was deposited in GenBank\textsuperscript{TM} with accession number AY307931. AA, amino acid.
many transcription factors, although it lacks a TATA box. This sequence has functional promoter activity in all of the tested cell lines and initiates transcription from a region spanning ~180 bases (Figs. 1–3). Genes with promoters lacking TATA boxes often have multiple transcription start sites (35–37). Other mucin genes including MUC2 and MUC5AC have promoters with TATA boxes and initiate transcription from a region spanning only a few bases (20, 38). However, these genes have very restricted tissue- and cell type-specific patterns of expression. Conversely, MUC3 transcripts can be found in many disparate tissues including small and large intestine, thymus, liver, lymph nodes, and heart (17, 39). Because the probes used to detect MUC3 expression in most previous studies would recognize either MUC3A or MUC3B, it is possible that the wide range of tissues MUC3 is detected in is due to either one or the other gene being expressed. However, this cannot entirely explain the wide range of tissues expressing MUC3. Previous work has demonstrated that both MUC3A and MUC3B are expressed in adult and fetal small and large intestine and Caco-2 cells (16). Data from this work indicate that both genes are expressed in duodenum. It seems likely that the wide range of transcription initiation sites for MUC3A transcripts noted in this study may allow the gene to be active in different tissues with different repertoires of transcription factors. It may also be significant that the region encompassing the first 450 bases of the MUC3A promoter is GC-rich, i.e. it contains 62% G and C residues. This region contains 16 CpG dinucleotides, and it is possible that methylation-mediated gene silencing plays a role in regulating this gene.

There is some discrepancy between the mapping of transcription start sites by GeneRacer PCR and RNase protection analysis (Figs. 1 and 2). Some of this may be due to cloning biases. This possibility is suggested by the observation that several instances of multiple clones initiating at the same base were noted. This occurred with both LS174T cell and duodenal RNA. RNase protection analysis revealed the presence of eight major and many minor start sites (Fig. 2). Some of the faint bands may be due to nicking of the hybridized riboprobe during RNase digestion or at some other step of the procedure. The presence of MUC3B messages may complicate this analysis as well. The procedure did yield a major band with the actin control provided with the kit, however, and we obtained a narrow range of protected fragments when we applied RNase protection analysis to map the start of MUC2 and MUC2/hGH transgene transcription (20, 21). Thus, it seems highly probable that the MUC3A promoter initiates transcription from multiple start sites. Interestingly, the MUC3B 5′-GeneRacer clones obtained all initiated from the same base. Since only five such clones were obtained, it is difficult to draw conclusions, but this may indicate that the MUC3B promoter is fundamentally different from the MUC3A promoter.

The structure of MUC3A and MUC3B is compared with other fully sequenced human membrane mucins in Fig. 6. The MUC3 mucins are large, with mucin-like unique sequences in their amino termini. They contain two separate Thr-, Ser-, and Pro-rich tandem repeat domains that exhibit VNTR polymorphism within the human population (19). The carboxyl termini of the MUC3 mucins are alternatively spliced, with the most abundant splice variants containing two EGF-like domains, a transmembrane domain, a cytoplasmic domain, and a SEA module domain (12, 16, 32). This latter domain is often associated with the extracellular region of O-glycosylated proteins (40) and contains a conserved proteolytic processing cleavage site (41). MUC4 is also a large membrane-associated mucin; in fact most alleles of MUC4 are larger than those of MUC3A and MUC3B (11). The amino termini of MUC4 contain three 126–130-residue degenerate tandem repeats and a mucin-like unique sequence. The carboxyl termini lacks a cognate SEA module domain, but it does contain a proteolytic cleavage site (10). MUC4 also contains a cognate NIDO domain, an AMOP domain, and a vWF-D domain (42). The function of these conserved structural domains in MUC4 is unknown. The EGF-like domains of the MUC4 carboxyl terminus may be functionally important, however, because the EGF-like domains of the rat homolog of MUC4 binds to ErbB2 and modulates its phosphorylation and interaction with ErbB3 (3, 43). MUC4 is highly expressed in airways and colon; however, appreciable levels are also found in stomach, lung and prostate (17). Thus, similar to the MUC3 mucins, MUC4 is expressed in several tissues. MUC1, which is associated with normal breast and pancreas and many types of carcinomas (44), is smaller than the MUC3 mucins and MUC4. This mucin lacks cognate extracellular structural domains other than its set of tandem repeats; however, its cytoplasmic domain is evolutionarily conserved and can function as a scaffold for signal transduction (4, 5). The MUC13 (8) and MUC15 (9) mucins are very small in size; however, the carboxyl terminus of MUC13 has a domain structure similar to that of MUC3A and MUC3B, with EGF-like domains and a SEA module.

Membrane mucins vary tremendously in size and sequence...
but often share conserved structural domains. They function as a scaffold for glycans that can alter the electrostatic properties of cells or serve as specific ligands. Certain membrane mucins have been shown to function in signal transduction, and others may as well. The challenge now is to understand how the expression of these molecules affects cell function in both normal tissues and in various pathological states in which their expression is often altered.

REFERENCES

1. Hilkens, J., Lichtenberg, M. J., Vos, H. L., and Litvinov, S. V. (1992) Trends Biochem. Sci. 17, 359–363
2. Mannori, G., Crochet, P., Ceconi, O., Hanasaki, K., Aruffo, A., Nelson, R. M., Varki, A., and Bevilacqua, M. P. (1995) Cancer Res. 55, 4425–4431
3. Carraway, K. L., Rossi, E. A., Komatsu, M., Prie-Schiavi, S. A., Huang, D., Ren, J., Li, Y., and Kufe, D. (2002) J. Biol. Chem. 277, 881–889
4. Parry, S., Silverman, H., McDermott, K., Willis, A., Hollingsworth, M. A., and Brunak, S. (1994) Proteomics 4, 1159–1165
5. Zaretsky, J., Kotkes, P., Weiss, M., Lathe, R., Dion, A., and Keydar, I. (1990) J. Biol. Chem. 265, 13057–13064
6. Yin, B. W. T., and Lloyd, K. O. (2000) Bioessays 22, 659–665
7. Taylor-Papadimitriou, J., Burchell, J., Pemberton, L., Lalani, E., and Wilson, D. (1990) J. Mammary Gland Biol. Neoplasia 25, 299–309
8. Williams, S. J., Wroblewski, D. H., Tran, M., Eyre, H. J., Sutherland, G. R., and M. A. M. (2001) J. Biol. Chem. 276, 18327–18336
9. Pallesen, L. T., Bengtsson, L., Rasmussen, L. K., Petersen, T. E., and Rasmussen, J. T. (2002) Eur. J. Biochem. 269, 3575–3583
10. Moniaux, N., Vos, H. L., Litvinov, S. V., and Porchet, N. (2002) J. Biol. Chem. 277, 4161–4169
11. Debailleul, V., Laine, A., Huet, G., Mathon, P., d’Hoooge, M. C., Aubert, J. P., and Kim, Y. S. (1992) Biochem. J. 283, 466–475
12. Moniaux, N., Escande, F., Batra, S. K., Laine, A., and Aubert, J. P. (2000) Eur. J. Biochem. 267, 4536–4544
13. Moniaux, N., Escande, F., Batra, S. K., Laine, A., and Aubert, J. P. (2000) Eur. J. Biochem. 267, 4536–4544
14. Wroblewski, D. H., Hareven, M., Tsefla, H., Smorodinsky, N., Horev, J., Zaretsky, J., Kotles, P., Weiss, M., Rathe, L., Dion, A., and Keydar, I. (1990) Eur. J. Biochem. 200, 463–475
15. Hinojosa-Kurtzberg, A. M., Johansson, M. E., Madsen, C. S., Hansson, G. C., and Gendler, S. J. (2003) Am. J. Physiol. 284, G853–G862
16. Pratt, W. S., Carraway, K. L., Ho, J., J., Nash, M., Kim, Y. S., and Swallow, D. M. (2000) Biochem. Biophys. Res. Commun. 275, 916–923
17. Gum, J. R., Carraway, C. A., and Kim, Y. S. (2002) Biochem. Biophys. Res. Commun. 291, 466–475
18. Gum, J. R., Hicks, J. W., Swallow, D. M., Lagace, R. E., Byrd, J. C., Lampert, D. T. A., Siddiki, B., and Kim, Y. S. (1990) Biochem. Biophys. Res. Commun. 171, 407–415
19. Gum, J. R., Ho, J. L. L., Pratt, W. S., Hicks, J. W., Hill, A. S., Vinall, L. E., Roberton, A. M., Swallow, D. M., and Kim, Y. S. (1997) J. Biol. Chem. 272, 26674–26686
20. Gum, J. R., Hicks, J. W., Toribara, N. W., Siddiki, B., and Kim, Y. S. (1994) J. Biol. Chem. 269, 2440–2446
21. Gum, J. R., Hicks, J. W., Gillespie, A. M., Carlson, E. J., Kamnues, L., Karrnik, S., Hong, J. C., Epstein, C. J., and Kim, Y. S. (1999) Am. J. Physiol. 276, G666–G676
22. Gum, J. R., Hicks, J. W., Toribara, N. W., Rothe, E. M., Lagace, R. E., and Kim, Y. S. (1992) J. Biol. Chem. 267, 21375–21383
23. Heinemeyer, T., Wingender, E., Reuter, I., Hermjakob, H., Kel, A. E., Kel, O. V., Ignatieva, E. V., Ananko, E. A., Podkolodnaya, O. A., Kolpakov, F. A., Podkoleny, N. L., and Kolchanov, N. A. (1998) Nucleic Acids Res. 26, 362–367
24. Quandt, K., Frech, K., Karas, H., Wingender, E., and Werner, T. (1995) Nucleic Acids Res. 23, 4878–4884
25. Nielsen, H., Engelbrecht, J., Brunak, S., and von Heijne, G. (2001) J. Mol. Biol. 314, 1–6
26. Bohn, S. K., Gum, J. R., Erickson, R. H., Hicks, J. W., and Kim, Y. S. (1995) Biochem. J. 311, 833–843
27. Traher, P. G., and Silberg, D. G. (1996) Annu. Rev. Physiol. 58, 275–297
28. Eisenberg, R. L., H. R., Lottermann, C., and Kim, Y. S. (2000) Gene (Amst.) 258, 375–388
29. Gum, J. R., Hicks, J. W., and Kim, Y. S. (1997) Biochem. J. 324, 259–267
30. Verrijzer, C. P. (2001) Science 289, 2010–2011
31. Orphanides, G., Lagrange, T., and Reinberg, D. (1996) Genes Dev. 10, 2657–2663
32. Williams, S. J., Munster, D. J., Quin, R. J., Gotley, D. C., and McGuckin, M. A. (1999) Biochem. Biophys. Res. Commun. 261, 83–89
33. Marshall, E. (2001) Science 292, 1982–1983
34. Kosak, M. (1991) J. Biol. Chem. 266, 19867–19870
35. Deng, T. L., Li, D. W., Jenh, C. H., and Johnson, L. F. (1986) J. Biol. Chem. 261, 16000–16005
36. Lu, J., Lee, W., Jiang, C., and Keller, E. B. (1994) J. Biol. Chem. 269, 5391–5402
37. Liu, W., Wang, G., and Yokovlev, A. G. (2002) J. Biol. Chem. 277, 8273–8278
38. Li, D., Gallup, M., Fan, N., Szymkowski, D. E., and Basbaum, C. B. (1998) J. Biol. Chem. 273, 6812–6820
39. Kyo, K., Mato, T., Nagawa, H., Lathrop, G. M., and Nakamura, Y. (2001) J. Hum. Genet. 46, 5–20
40. Wroblewski, D. H., McGuckin, M. A., Williams, S. J., Baruch, A., Yoeli, M., Ziv, R., Okun, L., Zaretsky, J., Smorodinsky, N., Keydar, I., Neophyto, P., Stacey, M., Lin, H. H., and Gordon, S. (2002) Protein Sci. 11, 698–706
41. Parry, S., Silverman, H., McDermott, K., Willis, A., Hollingworth, M. A., and Harris, A. (2001) Bioessays 23, 715–720
42. Dekker, J., Rossen, J. W., Buller, H. A., and Einerhand, A. W. (2002) Trends Biochem. Sci. 27, 126–131
43. Carraway, K. L., Ramaswaur, V. P., Haq, B., and Carothers Carraway, C. A. (2003) Bioessays 25, 66–71
44. Gendler, S. J., and Spicer, A. P. (1995) Annu. Rev. Physiol. 57, 607–634
45. Gendler, S. J., Lancaster, C. A., Taylor-Papadimitriou, J., Duhig, T., Peat, N., Burchell, J., Pemberton, L., Lalani, E., and Wilson, D. (1990) J. Biol. Chem. 265, 15286–15293
Initiation of Transcription of the MUC3A Human Intestinal Mucin from a TATA-less Promoter and Comparison with the MUC3B Amino Terminus

James R. Gum, Jr., James W. Hicks, Suzanne C. Crawley, Christine M. Dahl, Stacey C. Yang, Anthony M. Roberton and Young S. Kim

J. Biol. Chem. 2003, 278:49600-49609.
doi: 10.1074/jbc.M305769200 originally published online September 4, 2003

Access the most updated version of this article at doi: 10.1074/jbc.M305769200

Alerts:
• When this article is cited
• When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 42 references, 19 of which can be accessed free at http://www.jbc.org/content/278/49/49600.full.html#ref-list-1