Identification and characterization of the *MUC2* (human intestinal mucin) gene 5'-flanking region: promoter activity in cultured cells

James R. GUM, JR., James W. HICKS and Young S. KIM

Gastrointestinal Research Laboratory (151M2), Department of Veterans' Affairs Medical Center and Departments of Anatomy, Medicine and Pathology, University of California, San Francisco, CA 94121, U.S.A.

The initiation point for *MUC2* gene transcription is located within a 7000-base GC-rich region of the mucin gene cluster found on chromosome 11p15.5. The promoter activity of the 5'-flanking region of the *MUC2* gene was examined following its cloning into the luciferase-producing pGL2-Basic reporter vector. A short segment comprising bases –91 to –73 relative to the start of transcription was found to be important for basal promoter activity in all cell lines tested. Electrophoretic mobility shift assays demonstrated nuclear protein binding to this region, which contains the consensus CACCC motif (5'-GCCACACCC). This element has been shown to be functionally important in several promoters that are active in diverse cell types. Competition experiments using an Sp1 oligonucleotide and antibody supershift experiments indicated that both Sp1 and other Sp1 family members bind to this element. Inclusion of the region

between bases –228 and –171 in pGL2-Basic constructs increased normalized luciferase reporter activity by almost 3-fold in C1a cells, which produce relatively high levels of *MUC2* mRNA. Significantly lower levels of normalized luciferase activity resulted when the same construct was transfected into cultured cell lines that express low or undetectable levels of *MUC2*, suggesting a possible role for this region in conferring cell-type specificity of expression. We also demonstrate, using actinomycin D, that the *MUC2* mRNA is long-lived, at least in cultured cells. Moreover, no evidence was found that the *MUC2* mRNA turned over more rapidly in LS174T cells, which produce relatively low levels of *MUC2* mRNA, as compared with C1a cells, which produce high levels of mRNA. Thus a long mRNA half-life appears to be an important mechanism involved in achieving elevated levels of *MUC2* mRNA.

INTRODUCTION

The human *MUC2* gene encodes a large, highly glycosylated, polymeric mucin that is abundantly expressed by the intestinal epithelium [1–5]. *In situ* hybridization analysis and immunohistochemistry have demonstrated that the *MUC2* gene is expressed solely in goblet cells in the small and large intestine, whereas genes encoding two other mucins, *MUC3* and *MUC4*, are expressed in both goblet cells and enterocytes [6,7]. Other tissues of the gastrointestinal tract, including stomach, oesophagus and gall bladder, fail to produce significant quantities of *MUC2* [8,9]. Human colon cancer specimens usually contain only relatively low levels of *MUC2* mRNA [10,11], thus implying that *MUC2* expression is dependent upon maintenance of the high state of differentiation characteristic of the normal intestinal epithelium. *MUC2* is expressed in some other tissues as well. At least low-level expression of *MUC2* is observed in the epithelium and glands of the endocervix and in the airways [6,12,13]. Moreover, in airways *MUC2* is induced by certain pathological conditions [13,14].

The gut represents an ideal system in which to study tissue- and cell-specific gene expression. The normal pattern of development in the gut has been well characterized and the cell lineages present in the epithelium are defined and easily recognized. Moreover, cell lines derived from the intestinal epithelium are readily available. Gordon and colleagues [15,16] and Traber and colleagues [17,18] have performed elegant studies elucidating the elements and factors required for enterocyte-specific expression of intestinal fatty acid binding protein and sucrase-isomaltase respectively. Less is known about genes expressed in goblet cells, although Podolsky and colleagues have recently described the sequences necessary for activity of the intestinal

trefoil factor gene promoter [19]. The identification of mechanisms responsible for the regulation of several intestine-specific genes will be important not only to facilitate our understanding of the regulation of individual genes, but also because it may provide clues to mechanisms that control differentiation and normal and malignant development in these organs.

Our laboratory has long been interested in intestinal mucins; their structure, regulation and altered expression in disease states, especially cancer [5,20]. In order to better understand the factors that control mucin gene expression, we have cloned and characterized the 5'-flanking region of the human *MUC2* gene. Luciferase reporter gene constructs and transient transfection experiments were used to characterize the promoter activity of this region in a variety of cell types. Moreover, we demonstrate the binding of Sp1 and Sp1 family members to a CACCC element found in the proximal portion of the *MUC2* promoter. In addition, experiments using actinomycin D to inhibit RNA synthesis followed by blot analysis indicate that the *MUC2* mRNA is long-lived in cultured cells. Thus the present paper describes the structure and activity of the human *MUC2* mucin gene promoter and further indicates that low turnover rates are important for maintaining *MUC2* mRNA levels in expressing cells.

EXPERIMENTAL

DNA sequence analysis, cell lines and culture, RNA isolation and RNA blot analysis

An 11.3 kb *SalI/EcoRI* fragment of clone GMUC 46 [4] was sequenced manually using Sequenase (United States Biochemicals) following subcloning of suitable restriction fragments

Abbreviations used: CAT, chloramphenicol acetyltransferase; EMSA, electrophoretic mobility shift assay.

The nucleotide sequence data reported have been submitted to the GenBank/EMBL/DBJ Nucleotide Sequence Databases under accession number U67167.

Figure 1 Construction of MUC2 promoter/reporter constructs

In step 1, a GMUC 46 fragment comprising nucleotides -8111 to $+806$ was cloned into the *SalI* site of pBluescript SK $-$. The angled arrow represents the *MUC2* transcription start site, and relevant restriction sites are indicated. The terminal *SalI* site of the resulting construct was destroyed (enclosed in parentheses) as an *XhoI* cohesive terminus was ligated in. This construct was then digested with *NarI* and *SalI*, treated with Klenow fragment and re-ligated to give the first construct indicated in step 2. PCR was utilized to generate a fragment containing bases -233 to $+19$ with a *HindIII* site introduced at its 3'-terminus. This fragment and the construct were digested with *HindIII* and *BsmI* and ligated to give the second construct shown in step 2. The *MUC2* 5'-flanking region was then retrieved using *XhoI* and *HindIII* and cloned into the same sites of pGL2-Basic to yield the $-2864/+19$ construct indicated in step 3. This places the luciferase structural gene (*luc*) under the control of the *MUC2* 5'-flanking sequence. The bases generated by PCR were sequenced to ensure that no mutations had been introduced. Constructs $-1309/+19$ and $-343/+19$ were generated by digesting the $-2864/+19$ construct with *PmlI* and *XhoI* or *NcoI* and *XhoI*, treating with Klenow fragment and re-ligating. Inserts for the smaller constructs and the reverse construct were produced by PCR as described in the Experimental section.

into M13 vectors. Almost all sequencing templates were produced using the exonuclease III method [21]; 100% of the sequence was determined in both directions. LS174T cells, originally derived from a colon carcinoma, were a gift from Dr. Barry Kahan (University of Texas Medical School at Houston, TX, U.S.A.) [22]. C1a cells, produced in our laboratory, are an adherent monolayer variant of the mucinous colon carcinoma cell line 5583s [23], obtained from Dr. Fred Bosman (Rotterdam Cancer Institute, The Netherlands). MGC80-3 cells were derived from a gastric carcinoma and were obtained from Dr. Kai-Hua Wang (Peking Union Medical College, Beijing, China) [24]. All other cell lines used, including L cells (mouse fibroblasts), AGS cells (gastric carcinoma) and HT-1080 cells (fibrosarcoma), were obtained from the American Type Culture Collection. All cell lines except L cells are of human origin. Cells were maintained in Dulbecco's modified Eagle's medium as described [25], except that 10% (v/v) fetal bovine serum was used. Total RNA was prepared using Tri Reagent (Molecular Research Center), and RNA blot analysis was performed as described previously [1,25].

Construction of MUC2 promoter/luciferase reporter constructs

The pGL2-Basic vector (Promega), which contains a luciferase structural gene located immediately downstream of a polylinker, was used for this purpose. A *SalI/XhoI* fragment of GMUC 46 spanning nucleotides -8111 to $+806$ was cloned into the *SalI* site of pBluescript SK $-$ (Stratagene), oriented such that the original *XhoI* site (inactivated upon cloning into the *SalI* site) was positioned adjacent to the *HindIII* site of the vector (Figure 1). This construct was digested with *NarI* and *SalI*, treated with Klenow fragment and re-ligated. This gave a construct stretching from nucleotides -2864 to $+806$, with the *SalI* site being maintained.

PCR with primer 65 (5'-CCATGGAAGCTTGCAGGGGC-GGTGTGGG; emboldened bases altered to introduce a *HindIII* site) and primer 69 (5'-TAAGGAGCCTGACCAGACTTGCT-TCTGGC) was carried out to produce a 364 bp fragment containing the terminal bases of the *MUC2* 5'-flanking sequence. This fragment was ligated into the construct described above

following digestion with *HindIII* and *BsmI*, resulting in an insert containing bases -2864 to $+19$ of the *MUC2* sequence. The -2864 to $+19$ fragment was retrieved by digestion with *XhoI* and *HindIII* and cloned into the same sites of pGL2-Basic. The construct was then sequenced from the *BsmI* to the *HindIII* site (bases -233 to $+19$) to ascertain that no errors had resulted from the PCR used to generate the construct. This construct ($-2864/+19$) was digested with *XhoI* and *PmlI* or *XhoI* and *NcoI*, followed by S1 nuclease and Klenow fragment treatment and ligation to give the $-1308/+19$ and $-343/+19$ constructs respectively. Constructs originating subsequent to base -343 were generated using PCR with primer 65 and a gene-specific primer followed by cloning into pGL2-Basic using the *SmaI* and *HindIII* restriction sites. The *MUC2* portions of all constructs produced in this manner were sequenced. The reverse construct containing bases $+17$ to -331 was generated using gene-specific primers that introduced *HindIII* and *XbaI* sites into the appropriate residues for cloning into pGL2-Basic in the reverse orientation ($5'$ -GCCCATGGCAAGCTTAGGAGCCTGACC and $5'$ -CCATGGTGTCTAGAAGGGGCGGTTGTGGG).

Transient transfections and assays for luciferase and chloramphenicol acetyltransferase (CAT)

All plasmids used for eukaryotic cell transfections were purified using caesium chloride density gradients followed by Sepharose 4B chromatography. pCAT-Control vector (Promega) was used as a CAT-producing internal standard in all transfections. Cells were plated at 1.5×10^6 per 35 mm tissue culture dish. On the following day, transfections were initiated by mixing $5 \mu\text{g}$ of experimental construct plus $5 \mu\text{g}$ of internal standard vector with $10 \mu\text{l}$ of Lipofectin (Gibco BRL) in a total volume of $200 \mu\text{l}$ of serum-free medium. Following a 15 min incubation at room temperature, $800 \mu\text{l}$ of serum-free medium was added, and the cell cultures were rinsed once with PBS and overlaid with the mixtures. The cultures were returned to incubation at 37°C for 24 h, then the medium was changed to standard growth medium and incubation was continued for a further 24 h. Cells were harvested after washing three times with PBS by sonicating in $500 \mu\text{l}$ of Reporter Lysis Buffer (Promega) followed by a 5 min centrifugation in a Microfuge.

The luciferase activity of the supernatants was assayed in duplicate using luciferase assay reagent (Promega) as photons released in 30 s using a Monolight 2010 luminometer (Analytical Luminescent Laboratory). CAT activity was measured in duplicate as described previously using butyryl-CoA and [^3H]chloramphenicol (New England Nuclear) in a 1 h incubation [26]. Luciferase activities thus obtained were corrected for transfection and harvesting efficiencies by dividing by CAT activities. Results were normalized to percentages of the value obtained using pGL2-Control vector (Promega), which directs the synthesis of luciferase using the simian virus 40 promoter and enhancer. Typically, assay results from cultures transfected with the pGL2-Control vector measured between 10000 and 100000 luciferase units (photons released in 30 s), depending on the cell line.

Electrophoretic mobility shift assays (EMSAs)

EMSAs were conducted using reagents contained in BandShift kits obtained from Pharmacia Biotech. Portions of $20 \mu\text{l}$ of reaction mixture contained 40 mM Tris/HCl (pH 7.5), 200 mM NaCl, 5 mM MgCl_2 , 2 mM dithiothreitol, 10% glycerol, $1 \mu\text{g}$ of poly(dI-dC) and 0.05% Nonidet P-40. Nuclear protein extracts were prepared as described [27] and used at $5 \mu\text{g}$ per reaction. In the experiments presented here, each binding reaction contained 1 ng of double-stranded CACCC probe ($5'$ -GCAGATGCCAC-

ACCCACCCTTGGCCCTCCC; bases -94 to -65 of Figure 2) or Sp1 probe ($5'$ -ATTCGATCGGGGCGGGGCGAGC). The probes were labelled with T4 polynucleotide kinase and [γ - ^{32}P]ATP to a specific radioactivity of approx. 800000 c.p.m./ng. Non-radioactive competitor was used at a concentration of 100 ng per reaction, except where noted. The CACCC and Sp1 sequences described above were used as competitors, as were double-stranded oligonucleotides based upon bases 228–207 (228/207) and 162–139 (162/139) of the *MUC2* sequence shown in Figure 2.

Purified IgG prepared against a synthetic peptide corresponding to amino acids 436–454 of human Sp1 (Santa Cruz Biotechnology) was used for antibody supershift experiments at $4 \mu\text{g}$ per reaction. Binding reactions were incubated at room temperature for 20 min, and $2 \mu\text{l}$ of IgG was added if required followed by incubation for another 30 min. Then $2 \mu\text{l}$ of loading dye was added (250 mM Tris/HCl, pH 7.5, 0.2% Bromophenol Blue, 0.2% Xylene Cyanol and 40% glycerol) and the mixtures were loaded on to a pre-run 4% polyacrylamide gel (58 parts acrylamide to 1 part *N,N'*-methylenebisacrylamide) prepared in 89 mM Tris/borate (pH 8.0), 89 mM boric acid and 1 mM EDTA. Following electrophoresis for 4 h at 100 V, the gel was dried and autoradiography was conducted for 2–8 h.

RESULTS

Sequence of the 5'-region of the *MUC2* gene

An 11.3 kb *SalI/EcoRI* fragment of GMUC 46 [4] was sequenced. This region stretched from base -8111 to base $+3152$, encompassing the 5'-flanking region of the *MUC2* gene as well as a segment of the structural gene to the end of exon 4 (Figure 2). The region between bases -4000 and $+3152$ has greater than 60% GC content, whereas upstream of base -4000 the sequence gradually becomes AT-rich. Repetitive sequences are evident throughout this region. Specifically, the tetranucleotide motif ATCC is repeated 101 times between bases -5232 and -4330 , and a region of the second intron between bases 1545 and 2021 consists of 33–38 base repeats with 70% GC content. A 90-nucleotide segment between bases -3320 and -3231 is repeated with 90% similarity between bases -2074 and -1985 , and the pentanucleotide CCTGC is repeated 26 times between bases -3450 and -1 . Also, bases between the 5'-terminus and nucleotide -8042 and between bases -7301 and -7022 contain the *Alu* repeat [28]. Near the transcription start site there is a 12 nt stretch that repeats with one base substitution. Also, a TATA box is located 25 nt upstream of the transcription start site, and a CACCC box is located between bases -88 and -80 .

Promoter activity of the *MUC2* 5'-flanking region

The promoter activity of the *MUC2* 5'-flanking region was analysed using cultured cell lines that express various levels of *MUC2*. As judged using RNA blot analysis (Figure 3), C1a cells expressed relatively high levels of *MUC2*, roughly 10 times those found in LS174T cells (densitometry performed using NIH Image software). The other cell lines tested expressed no detectable *MUC2*, even when the blots were overexposed. These cell lines and a mouse fibroblast line (L cells) were examined using various lengths of the *MUC2* 5'-flanking sequence cloned upstream of a luciferase reporter gene in pGL2-Basic. Table 1 shows results obtained with constructs initiated from various points in the *MUC2* sequence from base -2864 to base -91 . It is clear from these data that the *MUC2* 5'-flanking sequence expressed luciferase activity in all of these cell lines, and that this activity generally increased as the size of the construct increased.

Figure 2 Sequence of the 5'-region of the human *MUC2* gene

Nucleotides —3450 to +159 relative to the start of transcription [4] are shown. The overlined sequences represent a 90-base segment repeated one time with 90% similarity. The double-underlined sequences are repeated 26 times (in both directions) between bases —3450 and —1. The arrow above the sequence indicates the transcription start site [4], and the arrow below the sequence demarcates the first intron. Consensus TATA and CACCC motifs are boxed, as is a segment containing an overlapping direct repeat. The sequence was compiled and analysed using MacVector and Intelligenetics Suite software.

Figure 3 RNA blot analysis of the human cell lines used for transfections

Aliquots of 10 μ g of total RNA were analysed using SMUC 41 as a probe [1]. Mouse L cells, a fibroblast line that was also used for transfections, is derived from a species with DNA that does not hybridize to SMUC 41 [29] and therefore was not included in this analysis.

Close examination of Table 1 reveals, however, that only C1a cells exhibited high levels of activity with construct —343/+19. A possible explanation for this is that C1a cells may express

factors that specifically interact with elements located in the proximal region of the *MUC2* 5'-flanking sequence, thereby activating transcription. The longer constructs may express activity in the other cell lines because they are able to spuriously initiate transcription from various points along the 5'-flanking sequence. Indeed, pGL2-Basic has a polyadenylation signal located immediately upstream of its polylinker to eliminate spurious transcription starts from within the vector, thus yielding extremely low background transcription (Promega product literature). Cloning of long DNA fragments into this vector may thwart the background reduction feature, yielding aberrantly high luciferase values.

We next attempted to characterize further the promoter activity of the proximal portion of the *MUC2* 5'-flanking sequence (Figure 4). Constructs initiating from bases —343, —228, —171, —91 and —73, and a reverse construct, were tested. A significant increase in normalized luciferase activity was observed when results with the —91/+19 construct were compared with those obtained with the —73/+19 construct (average increase of 6.2-fold). The region between bases —91 and —73 contains the motif 5'-GCCACACCC, previously described as the CACCC box, the GT box or the β -globin consensus sequence. This element has been shown to be an integral part of several promoters [30–35]. No further increase in activity was observed using the —171/+19

Table 1 Normalized luciferase activities of various cell lines transfected with *MUC2* promoter/reporter constructs

Transfections and analysis were performed as described in the Experimental section. Values are means $\pm$ S.D. The numbers of individual cultures analysed are given in parentheses. N.D., not determined.

Cell line	Construct ...	Luciferase activity (units)				
		−91/+19	−114/+19	−343/+19	−1308/+19	−2864/+19
LS174T		1.39 $\pm$ 0.78 (14)	2.65 $\pm$ 1.38 (12)	1.94 $\pm$ 1.15 (10)	5.92 $\pm$ 1.94 (6)	5.27 $\pm$ 1.87 (6)
L cells		4.22 $\pm$ 1.91 (4)	N.D.	3.95 $\pm$ 2.55 (4)	8.15 $\pm$ 5.60 (4)	9.52 $\pm$ 7.31 (4)
AGS		4.85 $\pm$ 2.65 (4)	N.D.	3.90 $\pm$ 0.84 (4)	3.68 $\pm$ 0.84 (4)	2.00 $\pm$ 0.08 (4)
HT-1080		2.59 $\pm$ 1.91 (10)	4.40 $\pm$ 0.85 (2)	3.92 $\pm$ 1.64 (6)	14.92 $\pm$ 6.07 (4)	17.00 $\pm$ 11.20 (4)
MGC80-3		3.64 $\pm$ 2.23 (12)	1.82 $\pm$ 0.45 (4)	4.46 $\pm$ 2.38 (8)	14.75 $\pm$ 2.60 (4)	10.10 $\pm$ 2.38 (4)
C1a		7.01 $\pm$ 4.60 (15)	9.06 $\pm$ 6.88 (8)	16.30 $\pm$ 4.62 (11)	17.68 $\pm$ 1.00 (4)	26.88 $\pm$ 1.10 (4)

Figure 4 Promoter activity of promoter/reporter constructs in various cell lines

Transfections were conducted with the constructs indicated, and normalized luciferase activity was calculated as described in the Experimental section. The results are means $\pm$ S.D., and the numbers of individual cultures analysed are indicated.

construct as compared with the −91/+19 construct. A large increase in normalized luciferase activity was observed, however, with the −228/+19 construct, at least in C1a cells (increase from 6.8 % to 19.4 % of simian virus 40-directed positive control expression). The other cell lines expressed either little or no increased activity with this construct. This suggests that the region between bases −228 and −171 may be important for both the activity and the specificity of the *MUC2* promoter. No additional increases in activity were observed when the −343/+19 construct was used. Finally, the +17/−331 reverse construct produced only low (in fact, barely detectable) levels of luciferase activity.

EMSAs

The transient transfection experiments described above indicated a possible important role for the CACCC box found in the proximal portion of the *MUC2* promoter. This 9-base motif is

GC-rich (5'-GCCACACCC), and not dissimilar to the Sp1 motif 5'-CCCGCC. Evidence has been presented by several groups that Sp1 binds to the CACCC motif of various promoters and thereby participates in the initiation of transcription [32,36,37]. In other promoters, however, the CACCC motif appears to bind a transcription factor that is different from Sp1 [33,38]. Mammalian factors with sequence similarity to the *Drosophila Krüppel* family of nuclear proteins have been cloned and their binding to CACCC motifs has been demonstrated [39,40]. Thus we sought to determine possible nuclear protein binding to this site.

Figure 5 shows EMSAs that were performed using a labelled probe containing the *MUC2* CACCC box and with CACCC- and Sp1-containing oligonucleotides as non-radioactive competitors. Inclusion of nuclear proteins isolated from any of the four cell lines tested resulted in one intense band and several faint bands of low mobility. The addition of a 100-fold excess of non-radioactive CACCC oligonucleotide dramatically diminished the intensities of these bands, demonstrating the specific nature of

Figure 5 EMSA analysis using a labelled CACCC probe

Assays were conducted using nuclear proteins isolated from four cell lines used for transfections. Non-radioactive competitor was absent (—) or was either the CACCC oligonucleotide (CAC) or the Sp1 oligonucleotide (Sp1).

Figure 6 Specificity of binding of the CACCC probe

Binding of the labelled CACCC probe was challenged with various amounts (ng) of unlabelled CACCC (CAC) or Sp1 competitor as indicated. The last two lanes contained 100 ng of either the 228/207 or the 162/139 oligonucleotide as an irrelevant sequence control. All reactions contained C1a nuclear proteins except for the control (no protein).

the binding. Surprisingly, however, the Sp1 non-radioactive oligonucleotide diminished the intensities of the bands to a greater extent than did the non-radioactive CACCC oligonucleotide (Figure 5). This effect can be visualized more clearly in the experiment shown in Figure 6, in which various concentrations of the CACCC and Sp1 oligonucleotides were used as competitors for CACCC probe binding. This indicates that the Sp1 oligonucleotide is an even more effective inhibitor of CACCC probe binding than is the CACCC oligonucleotide itself. Both competitor oligonucleotides inhibited binding of all bands, not just the major band. Two other double-stranded oligonucleotides proved ineffective as competitors, demonstrating the specific nature of the binding (Figure 6).

We next performed electrophoretic mobility supershift experiments using anti-Sp1 antibody (Figure 7). Anti-Sp1 supershifted the mobility of the major band in every instance. The mobilities of the minor bands were unaltered, despite the fact that they were all effectively competed by the unlabelled Sp1 oligonucleotide. This suggests that the major band represents binding of the CACCC oligonucleotide by Sp1, whereas the minor bands

Figure 7 Anti-Sp1 EMSA supershift experiments

Binding reactions in the various cell lines were incubated in either the absence (—) or the presence (+) of anti-Sp1 for 30 min prior to electrophoresis. Only the major band was shifted to a lower mobility by the presence of the antibody.

represent binding by other Sp1 family members not recognized by the antibody. Thus Sp1 and other Sp1 family members clearly bind to the proximal portion of the *MUC2* promoter. Finally, we examined the binding of nuclear proteins from these cell lines to the labelled Sp1 probe (Figure 8). Specific binding was again apparent, with a pattern similar to that seen with the labelled CACCC probe; as expected, the non-radioactive Sp1 oligonucleotide was a more effective competitor than was the non-radioactive CACCC oligonucleotide.

MUC2 mRNA turnover

Analysis of *MUC2* mRNA turnover in C1a and LS174T cells was conducted using actinomycin D to inhibit RNA synthesis in cell culture. It is possible that *MUC2* mRNA levels are regulated in cultured cells, as well as *in vivo*, by the rate of degradation. To examine this, we treated cultures of C1a and LS174T cells with actinomycin D for various lengths of time up to 48 h to inhibit RNA synthesis and then analysed *MUC2* mRNA levels by blot analysis (Figure 9). Actinomycin D is toxic to these cells over this

Figure 8 EMSA analysis using labelled Sp1 probe

Binding reactions in the various cell lines were conducted with no competitor, with CACCC competitor (CAC) or with Sp1 competitor. The last lane of each series (+) shows the mobility supershift resulting from incubation with anti-Sp1.

Figure 9 *MUC2* mRNA stability in C1a and LS174T cells

Plates (10 cm diam.) were seeded at 1×10^7 cells per plate using standard culture conditions. After 2 days, four plates for each line (approx. 80% confluent) were treated with medium containing 10 μ g/ml actinomycin D (Sigma Chemical Co.). Total RNA was isolated after treatment for the indicated time periods and subjected to blot analysis using SMUC 41 as a probe. Lanes containing C1a cell RNA were exposed to film for 16 h; those containing LS174T cell RNA were exposed for 96 h. An identically prepared blot was analysed using a glyceraldehyde-3-phosphate dehydrogenase probe (GAPDH) and exposed to film for 3 h.

time period, perhaps more so to C1a cells than to LS174T cells. For example, the yield of total RNA harvested fell to 22% and 47% of that in the untreated cultures at the 48 h time point in C1a cells and LS174T cells respectively. Even under these conditions, however, with the cells rapidly deteriorating, the large *MUC2* mRNA was stable for at least 12 h in C1a cells, and perhaps for even longer in LS174T cells. In both cell lines, *MUC2* mRNA degradation paralleled that of the glyceraldehyde-3-phosphate dehydrogenase mRNA used as a control. Thus the *MUC2* mRNA appears to be reasonably long-lived in cultured cells, and no evidence was found that this mRNA was degraded more rapidly in LS174T cells than in C1a cells.

DISCUSSION

The *MUC2* gene is located on chromosome 11p15.5, within a 400 kb segment containing genes for at least three other secretory mucins [41]. The initiation point for *MUC2* gene transcription lies within a 7000-base GC-rich region. Like many genes exhibiting tissue-specific expression, it contains the TATA element located 25 bases from its initiation point [42]. Thus the *MUC2* gene 5'-flanking sequence contains this important feature of cell-type-specific transcriptional regulation. Transient transfections of a pGL2-Basic reporter construct containing nucleotides -73 to +19 of the *MUC2* sequence resulted in only low levels of luciferase activity being produced, indicating that this segment by itself does not constitute an efficient promoter. Inclusion in the construct of the next 18 nucleotides up to base -91, however, resulted in significant increases in promoter activity in all cell lines tested. This increase, which averaged 6.2-fold, was observed both in cell lines that express *MUC2* mRNA and in those that do not. Computer analysis of the segment between bases -91 and -73 revealed the presence of a CACCC box motif [30,31]. No increase in activity was observed with the construct initiating at base -171 compared with that initiating at base -91. The inclusion of the next 57 bases in the -228 to +19 construct, however, resulted in an increase of almost three times in the

normalized luciferase activity in C1a cells. The other cell lines tested, which produce little or no *MUC2* mRNA, exhibited little or no increase in normalized luciferase activity with this construct. These data are consistent with the notion that the region between bases -171 and -228 contains an element or elements important for specific expression from the *MUC2* promoter. Because of its possible importance in regulation, we searched the -171 to -228 region for cognate nuclear protein-binding sites using both MacVector and Intelligenetics Suite software, but failed to find any matches that appeared biologically significant. Use of the -343 to +19 construct produced at best only a small increase in luciferase activity in any of the cell lines tested.

The CACCC box found in the proximal portion of the *MUC2* promoter is an interesting element that appears to be important for *MUC2* gene transcription. Work in several laboratories has indicated the importance of this element in facilitating gene transcription in both erythroid and non-erythroid cell lineages [30-35]. Our EMSA experiments demonstrated that Sp1 is the most abundant binding factor for this motif. In addition, this element appears to be able to bind other factors as well, as not all bands were supershifted with anti-Sp1. Likely candidates for additional factors able to bind this element include Sp2, Sp3 and Sp4, all of which are zinc-finger proteins with sequence similarity to Sp1 [34,35]. The CACCC oligonucleotide was actually less effective than the Sp1 oligonucleotide as a competitor for binding of the labelled CACCC probe (Figures 5 and 6). This may indicate that factors previously described as having similarity to the *Drosophila Krüppel* family of nuclear proteins and demonstrated to have greater affinity for CACCC boxes than for GC boxes are relatively unimportant in the context of the *MUC2* promoter [39,40].

How Sp1, and perhaps other Sp1 family members, function in concert with other transcription factors to regulate *MUC2* gene transcription is an important question requiring further study. Recent work with the epithelial cell membrane mucin gene *MUC1* promoter suggests that the relative activities of two Sp1 family members may play a significant role in regulating *MUC1* gene transcription [43]. A similar mechanism may play a role in regulating the *MUC2* promoter, but additional regulatory mechanisms are likely to be present as well. A role for the region between nucleotides -171 and -228 is suggested by transient transfection experiments. It is also possible, however, that regions downstream of the transcription initiation site may be important in this regard, as has been shown for a growing number of mammalian reporters [44-46]. The importance of the proximal 5'-flanking region in regulating transcription has been demonstrated for several intestinal enterocyte-specific genes [15-18]. In goblet cells, however, less is known, although the importance of the 5'-flanking region in goblet-cell-specific regulation has been demonstrated for intestinal trefoil factor [19].

Problems arise in interpreting the results of promoter/reporter transfections and DNA/protein binding experiments using cultured cells. For example, the regulatory mechanisms that control mucin gene expression in goblet cells residing in the highly differentiated intestinal epithelium may differ substantially from those of cultured cells. This problem is underscored by the differences in mRNA levels present in the two cell types. Comparison of RNA blots from several studies reveals that intestinal tissue samples contain substantially higher levels of *MUC2* mRNA than are found in cultured cells [1,11,47]. We attempted to quantify this by comparing *MUC2* mRNA levels in the C1a cells used in the present study with those in a sample of human small intestinal mucosa RNA using standard blot analysis. It was found that the cultured cells contained approximately one-tenth of the *MUC2* mRNA found in the intact tissue (results

not shown). Furthermore, only a small percentage of intestinal mucosal cells are goblet cells [48,49]. Thus it can be conservatively estimated that cultured-cell models differ from authentic goblet cells in terms of steady-state *MUC2* mRNA levels by at least one, and quite possibly by more than two, orders of magnitude.

As part of these studies, we sought to gain some insight into the stability of the *MUC2* mRNA. We treated cultures of C1a and LS174T cells with the RNA synthesis inhibitor actinomycin D and measured the decay of *MUC2* mRNA levels over time by RNA blot analysis. We found no detectable change in *MUC2* mRNA levels in either cell line over the 12 h time frame that was allowed by the overall deterioration of the cells brought about by actinomycin D treatment. This implies that the half-life of the *MUC2* mRNA is substantially longer than 12 h. Moreover, this approach produced no evidence that the *MUC2* mRNA has a shorter half-life in LS174T cells than in C1a cells, in spite of the observed 10-fold difference in steady-state mRNA levels between these cell lines. We attempted to measure relative *MUC2* transcription rates in the two cell lines using nuclear run-on assays [50,51]. Despite repeated attempts to optimize conditions, including the use of both tandem repeat (SMUC 41 [1]) and unique (SMUC 313 [52]) sequence probes, we never observed an adequate signal-to-background ratio for *MUC2* as would be needed for quantification with any degree of confidence. The use of an actin probe as a positive control in these experiments indicated that the procedure was functioning properly in our laboratory (results not shown). Thus the results of the actinomycin D inhibition experiment and the nuclear run-on assays are consistent with the relatively high steady-state levels of *MUC2* mRNA found in certain cultured cell lines being achieved by a long mRNA half-life, in spite of low rates of transcription. Whether similar mechanisms are operative in intestinal goblet cells is an important question. Velich and Augenlicht [53] found that *MUC2* mRNA levels were increased in HT-29 colon cancer cells by inducers of protein kinases C and A, while transcription rates remained low but constant. This suggests the possibility that the *MUC2* gene is not simply transcriptionally regulated in HT-29 cells, but the relevance of this to regulation *in vivo* is unclear.

In summary, we have cloned and characterized the 5'-region of the *MUC2* gene. This segment is GC-rich and contains several repetitive elements. Transient transfections of promoter/reporter constructs indicated that the *MUC2* 5'-flanking region has promoter activity and that bases -91 to -73 are important for this activity in several cell types. This region contains the CACCC motif present in many promoters and was shown to bind Sp1 and other Sp1 family members present in several cell lines. Inclusion of the segment containing bases -228 to -171 in promoter/reporter constructs increased promoter activity in transiently transfected C1a cells to a significantly higher degree than in cell lines that produce low or undetectable levels of *MUC2* mRNA. This suggests a role for this segment in conferring cell-type specificity upon *MUC2* gene transcription. Finally, *MUC2* mRNA was shown to be long-lived, at least in cultured cells. Thus the present study provides an initial description of the *MUC2* promoter and indicates segments likely to be important to its function. Hopefully, thorough characterization of intestine-specific promoters, including that of *MUC2*, will elucidate not only the elements and factors important in the regulation of specific genes but also mechanisms responsible for maintaining the exquisite degree of differentiation exhibited by the intestinal mucosa.

This work was supported by the Department of Veterans' Affairs Medical Research Service, and by grant CA 24321 from the National Cancer Institute.

REFERENCES

- Gum, J. R., Byrd, J. C., Hicks, J. W., Toribara, N. W., Lampert, D. T. A. and Kim, Y. S. (1989) *J. Biol. Chem.* **264**, 6480-6487
- Toribara, N. W., Gum, J. R., Culhane, P. J., Lagace, R. E., Hicks, J. W., Petersen, G. M. and Kim, Y. S. (1991) *J. Clin. Invest.* **88**, 1005-1013
- 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
- Gum, J. R., Hicks, J. W., Toribara, N. W., Siddiki, B. and Kim, Y. S. (1994) *J. Biol. Chem.* **269**, 2440-2446
- Gum, J. R. (1995) *Biochem. Soc. Trans.* **23**, 795-799
- Audie, J. P., Janin, A., Porchet, N., Copin, M. C., Gosselin, B. and Aubert, J. P. (1993) *J. Histochem. Cytochem.* **43**, 1479-1485
- Chang, S. K., Dohrman, A. F., Basbaum, C. B., Ho, S. B., Tsuda, T., Toribara, N. W., Gum, J. R. and Kim, Y. S. (1994) *Gastroenterology* **107**, 28-36
- Baeckstrom, D., Karlsson, N. and Hansson, G. C. (1994) *J. Biol. Chem.* **269**, 14430-14437
- Toribara, N. W., Robertson, A. M., Ho, S. B., Kuo, W. L., Hicks, J. W., Gum, E., Hicks, J. W., Gum, J. R., Byrd, J. C., Siddiki, B. and Kim, Y. S. (1993) *J. Biol. Chem.* **268**, 5879-5885
- Ho, S. B., Niehans, G. A., Lyftogt, C., Yan, P. S., Cherwitz, D. L., Gum, E. T., Dahiya, R. and Kim, Y. S. (1993) *Cancer Res.* **53**, 641-651
- Ogata, S., Vehara, H., Chen, A. and Itzkowitz, S. H. (1992) *Cancer Res.* **52**, 5971-5978
- Jany, B. H., Gallup, M. W., Yan, P., Gum, J. R., Kim, Y. S. and Basbaum, C. B. (1991) *J. Clin. Invest.* **87**, 77-82
- Gerard, C., Eddy, R. L. and Shows, T. B. (1990) *J. Clin. Invest.* **86**, 1921-1927
- Dohrman, A., Young, C., Gallup, M., Ohmori, H., Massion, P., Nadel, J., Escudier, E., Chapelin, C., Coste, A. and Basbaum, C. (1995) *Am. J. Res. Crit. Care Med.* **151**, A160
- Simon, T. C. and Gordon, J. I. (1995) *Curr. Opin. Genet. Dev.* **5**, 577-586
- Gordon, J. I., Schmidt, G. H. and Roth, K. A. (1992) *FASEB J.* **6**, 3039-3050
- Markowitz, A. J., Wu, G. D., Bader, A., Cui, Z., Chen, L. and Traber, P. G. (1995) *Am. J. Physiol.* **269**, G925-G939
- Suh, E., Chen, L., Taylor, J. and Traber, P. G. (1994) *Mol. Cell. Biol.* **14**, 7340-7351
- Sands, B. E., Ogata, H., Lynch-Devaney, K., deBeaumont, M., Ezzell, R. M. and Podolsky, D. K. (1995) *J. Biol. Chem.* **270**, 9353-9361
- Kim, Y. S. (1992) *J. Cell. Biochem. Suppl.* **16G**, 91-96
- Henikoff, S. (1984) *Gene* **28**, 351-359
- Tom, B. H., Rutky, L. P., Jakstys, M. M., Oyasu, R., Kay, C. I. and Kahan, B. D. (1976) *In Vitro* **12**, 180-191
- Verstijnen, C. P., Arends, J. W., Moerkerk, P. T., Geraedts, J. P., Sekikawa, K., Uitendaal, M. P. and Bosman, F. T. (1987) *Virchows Arch. B* **53**, 191-197
- Wang, K. H. (1983) *Acta Biol. Exp. Sin.* **16**, 257-261
- Gum, J. R., Kam, W. K., Byrd, J. C., Hicks, J. W., Sleisenger, M. H. and Kim, Y. S. (1987) *J. Biol. Chem.* **262**, 1092-1097
- Seed, B. and Sheen, J. Y. (1988) *Gene* **67**, 271-277
- Andrews, N. C. and Faller, D. V. (1991) *Nucleic Acids Res.* **19**, P2499
- Schmid, C. W. (1996) *Prog. Nucleic Acid Res. Mol. Biol.* **53**, 283-319
- Gum, J. R., Hicks, J. W., Lagace, R. E., Byrd, J. C., Toribara, N. W., Siddiki, B., Fearney, F. J., Lampert, D. T. A. and Kim, Y. S. (1991) *J. Biol. Chem.* **266**, 22733-22738
- Myers, R. M., Tilly, K. and Maniatis, T. (1986) *Science* **232**, 613-618
- Weiher, H., Konig, M. and Gruss, P. (1983) *Science* **219**, 626-631
- Dittmer, J., Gegonne, A., Gitlin, S. D., Ghysdael, J. and Brady, J. N. (1994) *J. Biol. Chem.* **269**, 21428-21434
- DeVack, C., Lupp, B., Nichols, M., Kowenz-Leutz, E., Schmid, W. and Schutz, G. (1993) *Eur. J. Biochem.* **211**, 459-465
- Kingsley, C. and Winoto, A. (1992) *Mol. Cell. Biol.* **12**, 4251-4261
- Hagen, G., Muller, S., Beato, M. and Suske, G. (1992) *Nucleic Acids Res.* **20**, 5519-5525
- Gegonne, A., Bosselut, R., Bailly, R. A. and Ghysdael, J. (1993) *EMBO J.* **12**, 1169-1178
- Fischer, K. D., Haese, A. and Nowock, J. (1993) *J. Biol. Chem.* **268**, 23915-23923
- Hartzog, G. A. and Myers, R. M. (1993) *Mol. Cell. Biol.* **13**, 44-56
- Miller, I. J. and Bieker, J. J. (1993) *Mol. Cell. Biol.* **13**, 2776-2786
- Anderson, K. P., Kern, C. B., Crable, S. C. and Lingrel, J. B. (1995) *Mol. Cell. Biol.* **15**, 5957-5965
- Pigny, P., Guyonnet Duperat, V., Hill, A. S., Pratt, W. S., Gallegue-Zouitina, S., Collynn D'Hooge, M., Laine, A., Van Seuning, I., Gum, J. R., Kim, Y. S., et al. (1996) *Genomics* **38**, 340-352
- Zawel, L. and Reinberg, D. (1995) *Annu. Rev. Biochem.* **64**, 533-561
- Kovarik, A., Lu, P. J., Peat, N., Morris, J. and Taylor-Papadimitriou, J. (1996) *J. Biol. Chem.* **271**, 18140-18147
- Bisaha, J. G., Simon, T. C., Gordon, J. I. and Breslow, J. L. (1995) *J. Biol. Chem.* **270**, 19979-19988

-
- 45 Lee, S. Y., Nagy, B. P., Brooks, A. R., Wang, D. M., Paulweber, B. and Levy-Wilson, B. (1996) *J. Biol. Chem.* **271**, 707–718
- 46 Smith, A. N., Barth, M. L., McDowell, T. L., Moulin, D. S., Nuthall, H. N., Hollingsworth, M. A. and Harris, A. (1996) *J. Biol. Chem.* **271**, 9947–9954
- 47 Gum, J. R., Hicks, J. W., Swallow, D. M., Lagace, R. E., Byrd, J. C., Lampton, D. T. A., Siddiki, B. and Kim, Y. S. (1990) *Biochem. Biophys. Res. Commun.* **171**, 407–415
- 48 Cheng, H. (1974) *Am. J. Anat.* **141**, 481–501
- 49 Cheng, H. and Leblond, C. P. (1974) *Am. J. Anat.* **141**, 537–561
- 50 Greenberg, M. E. and Ziff, E. B. (1984) *Nature (London)* **311**, 433–438
- 51 Celano, P., Berchtold, C. and Casero, R. A. (1989) *Biotechniques* **7**, 942–944
- 52 Ohmori, H., Dohrman, A. F., Gallup, M., Tsuda, T., Kai, H., Gum, J. R., Kim, Y. S. and Basbaum, C. B. (1994) *J. Biol. Chem.* **269**, 17833–17840
- 53 Velcich, A. and Augenlicht, L. H. (1993) *J. Biol. Chem.* **268**, 13956–13961
-

Received 19 November 1996/10 March 1997; accepted 17 March 1997