Organization and Regulatory Aspects of the Human Intestinal Mucin Gene (MUC2) Locus*

(Received for publication, September 4, 1996, and in revised form, January 10, 1997)

Anna Velcich‡‡, Lisa Palumbo‡, Licia Selleri¶, Glen Evans***‡‡, and Leonard Augenlicht‡

†† Supported by National Institutes of Health Grant NCHGR HG00202.

From the ‡Department of Oncology, Montefiore Medical Center/Albert Einstein Cancer Center, Bronx, New York 10467, the ¶Department of Pathology, Stanford University School of Medicine, Stanford, California 94305, and the **McDermott Center for Human Growth and Development, University of Texas Southwestern Medical Center, Dallas, Texas 75235

THE JOURNAL OF BIOLOGICAL CHEMISTRY Vol. 272, No. 12, Issue of March 21, pp. 7968–7976, 1997
© 1997 by The American Society for Biochemistry and Molecular Biology, Inc.

The human MUC2 gene maps to chromosome 11p15, where three additional mucin genes have been located, and encodes the most abundant gastrointestinal mucin normally expressed in the intestinal goblet cell lineage. However, in pathological conditions, including colorectal cancer, MUC2 can be abnormally expressed. Therefore, it is of considerable interest to understand the regulation of the MUC2 gene and how the mechanism is altered in colon cancer. Toward this goal, we have isolated a group of overlapping clones (contig) spanning 85 kilobases harboring the entire MUC2 locus, including sequences located upstream of the gene. Detection of two DNase I-hypersensitive sites in the 5′ region of the MUC2 gene suggests the presence of DNA regulatory elements. To better characterize this region, we have sequenced 12 kilobases of the upstream region and analyzed it for functional activity by cloning portions of it into a luciferase reporter vector and assaying for promoter/enhancer activity using a transient transfection assay. A fragment from the AUG translational initiation codon +1 to −848 confers maximal transcriptional activity in several intestinal cell lines. Elements located further upstream exert a negative effect on the expression of the reporter gene when tested in conjunction with homologous or heterologous promoters. The same pattern of expression is observed when the MUC2/luciferase constructs are transfected into HeLa cells, which do not express the endogenous MUC2 gene. However, the level of activity in HeLa cells is at least an order of magnitude higher, suggesting that additional sequences singularly or in combination are responsible for the tissue- and cell lineage-specific expression of MUC2. Finally, we have identified an additional mucin-like gene (MUCX), located upstream of MUC2. We show that this MUCX gene, that is transcribed in opposite orientation to that of MUC2, is expressed with a pattern distinct from that of MUC2, yet similar to that of MUC5B and MUC6, two additional mucin genes located at chromosome 11p15. Recent information on the order of the mucin genes at chromosome 11p15 suggests that MUCX may be MUC6, one of the already identified mucin genes, or a novel one, yet to be fully characterized.

Mucins are the major components of mucus, the visco-elastic substance that protects and lubricates epithelial mucosa, including that of the gastrointestinal tract. They are highly glycosylated molecules, and up to 80% of their mass consists of O-linked glycosyl residues. Recently, the cloning of full-length or partial cDNA sequences of mucins expressed in different tissues has greatly facilitated investigations of the polypeptide moieties (reviewed in Refs. 1–3).

In the intestinal epithelium the major mucin is MUC2, and the corresponding gene has been mapped to human chromosome 11p15 (4). cDNAs, most likely corresponding to three distinct mucin genes, MUC6 (5), MUC5AC, and MUC5B (6, 7), encoding for a gastric and two tracheobronchial mucins, respectively, have been mapped to this same band. Although only the MUC2 cDNA has been characterized completely (8), the partial cDNA sequences of the gastric and tracheobronchial mucins suggest that these are distinct genes. Thus, chromosome 11p15 may be a locus that contains a cluster of mucin genes.

The expression of individual mucin genes is relatively organ-specific (9–11) and in the case of MUC2 is also cell type-specific, MUC2 being expressed almost exclusively in goblet cells (10, 12). In the intestine, this lineage, as well as columnar, enteroendocrine and Paneth cells most likely arise from a common precursor, the stem cell located near the bottom of the crypt. Stem cells differentiate as they migrate upwards to the crypt surface where they are exfoliated into the intestinal lumen (13). The molecular mechanisms that regulate the spatial and temporal differentiation of normal colonic epithelial cells are poorly understood. Moreover, the process is most likely under the influence of many external signals, some of which may be recapitulated in in vitro studies. For example, we have shown that in HT29 cells, a human adenocarcinoma cell line that is considered multipotent since it can express distinct cell lineage-specific markers upon exposure to appropriate inducers, MUC2 gene expression can be modulated through protein kinase C- and protein kinase A-dependent signal transduction pathways (14).

Although alterations of the level of mucin glycosylation have been well studied as a characteristic of colon cancer (15–17), only recently, with the availability of cDNA probes, has it been possible to study alterations in apomucin expression. Both up-regulation and down-regulation of mucin gene expression have been reported in cancer cells (10, 11, 18). Indeed, we have shown that in several cell lines that are derived from human mucinous tumors, which are characterized by the synthesis of large quantities of mucins, there is a constitutive high level of expression of the MUC2 gene, suggesting that in this subset of tumors MUC2 is deregulated (14).

Thus, to investigate the mechanisms that govern MUC2 expression during the differentiation of the goblet cell lineage and determine the causes of its abnormal expression in colon cancer, we undertook cloning the MUC2 locus to identify DNA
elements capable of directing tissue and cell lineage-specific expression of MUC2.

In this report we present a partial characterization of a contig isolated by chromosome walking from a human chromosome 11 cosmId library. The group of overlapping clones contains the entire MUC2 locus, including several thousand nucleotides of DNA extending from the 5′ terminus of MUC2, for which we have determined the DNA sequence. Detection of DNase I-hypersensitive sites in the untranscribed 5′-flanking sequence of MUC2 suggested the presence of DNA regulatory elements. To further characterize the potential cis-acting regulatory elements in this region we tested the functional activity of distinct segments of the MUC2 upstream region by performing in vitro transient transfection assays in several epithelial cell lines. We present evidence that a region of 850 nucleotides extending upstream from the MUC2 initiation of translation confers maximal activity to a luciferase reporter gene, while sequences that are located further upstream exert an inhibitory effect both on homologous and heterologous promoters. These data suggest that the regulation of the MUC2 gene most likely depends on the interaction of both positive and negative regulatory elements.

Finally, we have identified an additional mucin-like gene (MUCX) located upstream of MUC2. The pattern of expression of MUCX, which is transcribed in the opposite direction compared with MUC2, is also presented. Comparative analysis of DNA and deduced amino acid sequences, the pattern of expression of MUCX and the information on the physical order of the mucin genes clustered at 11p15 (19) suggest that the genomic sequences of MUC2 may correspond to a portion of the amino terminus of MUC6 or to a novel, yet unidentified, mucin gene.

MATERIALS AND METHODS

Oligonucleotides and DNA Probes—Oligonucleotides corresponding to the antisense sequence of portion of the tandem repeats in the MUC2 (5), MUC5AC (20), and MUC5B (21) were obtained from the oligosynthesis facility at the Albert Einstein Cancer Center and they are as follows: MUC6/LA184, 5′- AAGCTGGGACGCTGAGTGGAGTGT-GGT 3′ (5); MUC5AC/LA175, 5′ TGGAGTAGAGGTTGTGCTGGTTGT 3′ (20); MUC5B/AV-1, 5′ GGCTGTGGTGTCAGCACTGTGAGGGT-TGGAGTAGAGGTTGTGCTGGTTGT 3′ (20); MUC5B/AV-2, 5′ GGCTGTGGTGTCAGCACTGTGAGGGT-TGGAGTAGAGGTTGTGCTGGTTGT 3′ (21). B02 corresponds to the antisense of the MUC2 cDNA sequence spanning nucleotides 82–101 according to the sequence published in Gun et al. (8). CI2B is a cDNA representing a portion of the MUC2 tandem repeats (14). Probe C was generated from human placenta DNA by polymerase chain reaction amplification using primers 2S and 2A based on the cDNA sequence of human mucin-like protein (H-MLP) (22). The product is a 1-kb fragment containing sequences spanning from nucleotide 13486 to nucleotide 13911 of the MUC2 cDNA. Glyceraldehyde-3-phosphate dehydrogenase was analyzed by a rat probe. All DNA probes were purified inserts labeled by random priming using a Random Primers DNA labeling system from Life Technologies, Inc.

Genomic Cloning, Characterization, and Sequencing—High density filters of a human chromosome 11 cosmId library (SRH library) (23) were screened initially with CI2B in 0.5 M sodium phosphate buffer, pH 7.2, 8% SDS, 100 μg/ml of sonicated single-stranded salmon sperm DNA, at 65 °C. The 5′ and 3′ end fragments of the initial clone were isolated and used to screen the same high density filters to isolate overlapping clones. Partial restriction map analysis determined the extent of overlapping. For sequence analysis, restriction fragments, covering the contig, were subcloned into pBluescript SK+ vector (Stratagene), and double-stranded DNA templates were sequenced with both T3 and T7 promoter primers and internal primers using the dyeoxy nucleotide chain termination method with Sequenase kit, version 2 (U. S. Biochemical Corp.). When artifact banding due to the presence of G + C-rich areas was a problem, termination reactions were carried out in the presence of terminal deoxynucleotidyl transferase and excess dNTPs (24). Analysis of nucleic acid and protein sequence data was performed using PC/Genet (Intelligenetics) and Wisconsin Sequence Analysis Package GCG (Genetics Computer Group, Madison, WI) software.

MUC2/Luciferase Plasmid Construction—The vectors used were pGL2 basic, pGL2 enhancer, and pGL2 promoter, which contain no regulatory elements, the SV40 enhancer, and the SV40 promoter, respectively (Promega). Distinct sequence portions of the 12-kb 5′-flanking sequence of MUC2 were cloned in these vectors. A 525-nucleotide BamHI-NcoI fragment was blunt-ended at the NcoI site, corresponding to the AUG initiation codon (8). HindIII linkers were added, and the fragment was then cloned into the HindIII-BglII sites of pGL2basic and enhancer. The resulting construct is the MUC2/luciferase 0.5 plasmid. Plasmid 0.5 was linearized with NcoI, which cuts at a second internal NcoI site in the MUC2 sequence, and XhoI, in the pGL2 polylinker; a different NcoI-XhoI fragment, containing the same sequence present in the original NcoI-BglII region of the 0.5 fragment plus 375 nucleotides of the BamHI fragment that is located upstream and contiguous to the 0.5 fragment, was then inserted, generating the 0.8 plasmid. Plasmid 0.8 was partially digested with BamHI and a BamHI-SacI fragment of 4.9 kb, containing sequences upstream of the 0.9 fragment was cloned into the 0.8 basic and enhancer constructs, generating the 5.0-kb plasmid. A 6-kb SacI-KpnI fragment, carrying the region further upstream of MUC2, was isolated from the 15-kb EcoRI fragment, previously cloned into pBSKK+, and subcloned in the MUC2/luciferase 0.5 plasmid cut with SacI and Kpnl. The resulting construct is the 9.0-kb MUC2/luciferase construct. Plasmid p4.3 was generated by subcloning a 4556-nucleotide BamHI-SacI fragment, derived by partial BamHI digestion of a 5.2-kb SacI clone in pBSKK+, which excludes the 0.5 fragment, into the promoter/luciferase vector cut with BglII and SacI. A schematic representation of these plasmids is shown in Fig. 1B. The identity of each of the MUC2/luciferase plasmids was verified by restriction enzyme and partial sequence analyses.

Cell Culture and Transient Transfection Assays—HT29, LS174T, and HeLa cells were maintained in minimal essential medium supplemented with nonessential amino acids and 10% fetal calf serum. HT29 cells were induced with TPA and forskolin as described previously (14). For transfection, cells were seeded at 5 × 10⁵well in 24-well plates, and liposome-mediated transfections were performed 3 days later. The intact cell lines were transfected with tfx50 (Promega) using a ratio of txfDNA of 3:1. Typically, 1 μg of plasmid DNA containing 0.1 μg of test plasmid DNA, 0.2 μg of CMV-βgal, to correct for transfection efficiency, and 0.8 μg of carrier DNA were mixed with tfx50 (Promega) using a ratio of txfDNA of 3:1. Typically, 1 μg of plasmid DNA containing 0.1 μg of test plasmid DNA, 0.2 μg of CMV-βgal, to correct for transfection efficiency, and 0.8 μg of carrier DNA were mixed with tfx50 in 200 μl of serum and antibiotic-free medium. Cells were incubated with the transfection mixture for 2 h at 37 °C followed by the addition of 1 ml of complete medium. Cells were harvested 48 h later. HeLa cells were transfected with lipofectAMINE (Life Technologies, Inc.) following the supplier’s recommendations.

For the luciferase assay, cells were lysed on plates using the reporter lysis buffer (Promega), and cell extracts were prepared following supplier instructions. 5–20 μl of cell extract were used to determine luciferase activity, using the Promega detection kit and a Turner TD-20-e luminometer. The β-galactosidase activity was measured using 20–40 μl of cell extract. The luciferase activity of test plasmids is expressed as fold of induction of the test plasmid activity compared with that of the corresponding control, after correction for transfection efficiency as measured by the β-galactosidase activity.

DNase I-hypersensitive Site Detection—Nuclei from the different cell lines were prepared according to the protocol of Enver et al. (25). The nuclear pellet was resuspended at 500 μg of DNA/ml. 20–40 μl aliquots were incubated at 37 °C for 10 min with increasing amounts of DNase I, ranging from 0.05 to 0.8 unit of enzyme. DNase I digestion was blocked by the addition of EDTA, pH 8, at a final concentration of 25 mM, followed by incubation with RNase A, at 20 μg/ml, for 30 min at 37 °C. Samples were digested overnight with proteinase K, at 50 μg/ml in NTE buffer (0.1 M NaCl, 0.05 M Tris, pH 8, 1 mM EDTA) containing 0.25% SDS, final concentration. DNA was extracted twice with phenol:chloroform and ethanol-precipitated. DNA, digested with EcoRI, was analyzed by Southern blot using a 2.6-kb PstI fragment located downstream of the first exon of the MUC2 gene as a probe (8).

RNA Isolation and Analysis—Total RNA was isolated and analyzed as described previously (14).

RESULTS

Partial Characterization of a Contig Encompassing the Human MUC2 Locus and an Additional Mucin Gene—We used a cDNA probe (CI2B), corresponding to a portion of the repetitive region of the MUC2 gene, to screen a human chromosome 11 cosmId library and isolated a contig of approximately 85 kb.
Results from partial sequence analysis, Southern blot hybridization of restriction enzyme-digested DNA of the contig with probes corresponding to unique regions in the 5′-flanking region of approximately 12 kb by the heavy dotted box and the MUC2 5′-flanking region of approximately 12 kb by the black box. Other features of the MUC2 gene are as follows: start sites (8) are indicated by the arrow, BO2 denotes the localization of the corresponding oligonucleotide identifying the first exon, TRs represents the region of the tandem repeats, and C indicates portion of the unique carboxy-terminal region localized by the C probe (details under “Materials and Methods”). The light dotted box depicts the relative location of the putative mucin-like gene MUCX, identified by the 13-kb NotI fragment (probe X) and the 2.1-kb HindIII/NotI subfragment (probe X1). The arrow indicates the direction of MUCX transcription, which is opposite to that of MUC2. B is an expansion of the 15-kb EcoRI fragment indicated in A. The location, relative to the AUG (+1) translation initiation codon, and sizes of the different fragments subcloned in the pGL2 series vectors, as detailed under “Materials and Methods,” are also shown. The resulting MUC2/luciferase plasmids were tested in transient transfection assay for functional activity.

To investigate the nature of the mRNA detected in LS174T by the 13-kb NotI fragment, which we refer to as MUCX, we determined whether the MUC5AC, MUC5B, and MUC6 genes had the same pattern of expression as the mRNA detected by probes X and X1 in LS174T cells and HT29 cells stimulated with forskolin and TPA. As probes we utilized oligonucleotides, described under “Materials and Methods,” derived from published partial cDNA sequences corresponding to the tandem repeat regions of these other mucin genes. Fig. 3 shows the results of this analysis using probes corresponding to MUC5AC (A), MUC5B (B), and MUC6 (C). The pattern of expression of the different MUC genes is summarized in Table I.

The data presented in Fig. 3 and Table I show that the MUCX gene shares the same pattern of expression of both MUC5B and MUC6. In addition, both HeLa and HL60 are negative for the expression of any of the mRNA species detected by these probes.

To characterize further the MUCX gene, we determined the sequence of approximately 4 kb of DNA located at the 3′ end of the contig, and homology searches were done with the BLAST network service at National Center for Biotechnology Information. The MUCX gene showed three areas of homology with the hMUC2 gene at the nucleotide level. Most important, the homology was maintained between the derived amino acid sequences of the regions in the MUCX DNA identified in the search and the corresponding portions in the MUC2 protein (Fig. 4, A and B). These regions of similarity lay in the D2 and D3 domains of the MUC2 protein. The D domains (D1–D4) in MUC2 are characterized by a high degree of sequence similarity to four D-domains in prepro-von Willebrand factor and by the presence of cysteine residues whose position in MUC2 and other mucin sequences is maintained invariant (26). In addition, one of the MUCX-derived polypeptides also exhibits homology to a portion of the HGM-1 (gastric mucin) sequence (Ref. 27). Although the homology among the distinct portions of MUCX and MUC2 is only about 50%, it is noteworthy, as shown in Fig. 4, A and B, that the position of the Cys residues is perfectly conserved in the three sequences. It has
structure and Regulation of the Human MUC2 Locus

TABLE I

Pattern of expression of different mucin genes

Data are derived from Figs. 2 and 3.

| Gene   | LS174T | HT29 |
|--------|--------|------|
| MUCX   | C      | F    | TPA |
| MUC6   | +      | –    | –   |
| MUC5B  | +      | +    | +   |
| MUC5AC | +      | +    | +   |
| MUC5B  | –      | –    | –   |

The presence of DNase I-hypersensitive sites and several putative cis-regulatory motifs in the 5' region of the MUC2 gene is consistent with its role as promoter for MUC2. To explore this further, we tested for the presence of promoter/enhancer activity in the MUC2 5' flanking sequence. As shown in Fig. 1B, distinct portions of the 15-kb EcoRI fragments were cloned into the pGL2 vector series harboring the luciferase reporter gene. These plasmids contain DNA segments starting from the MUC2 translation initiation site (+1), which is located 25 nucleotides downstream from the mRNA cap sites and extending to −364 in the 0.3 construct; −516 in 0.5; −848 in 0.8; −513 in 5.0 kb; and −9062 in 9.0-kb construct, respectively (Fig. 1B). These fragments were cloned into the basic vector, which does not contain any regulatory element, as well as into the enhancer vector, which harbors the SV40 enhancer sequence downstream of the luciferase gene. The resulting MUC2/luciferase plasmids are labeled "b" (basic) and "enh" (enhancer) to indicate the vector background.

We transiently transfected the MUC2/luciferase plasmids into two different human intestinal cell lines that show distinct patterns of expression of the endogenous MUC2 gene. HT29 cells do not express MUC2 unless stimulated by any one of several agents, including forskolin and TPA. In contrast, LS174T cells have a constitutive very high level of MUC2 mRNA expression. Although a modest increase in luciferase activity of the MUC2 reporter constructs, ranging between 2- and 3-fold compared with control vectors, was detected, a general pattern of expression emerged. When the MUC2 fragments were tested for transcriptional activity in the enhancer background (Fig. 7A), maximal activity was associated with fragment 0.8, that extends up to −848 relative to the AUG,
both in HT29 and LS174T cells. In HT29 cells the luciferase activity associated with both enh0.5 and -0.8 plasmids was significantly greater than that of control \((p < 0.05, \text{signed rank test})\). A reduced activity was associated with sequences located further upstream \((\text{fragments 5.0 and 9.0 in Fig. 1B})\) when tested in the enhancer background. However, in the basic background, the 0.8 and 5.0 fragments were equally active, in both HT29 and LS174T cells. In HT29 cells the luciferase activity associated with both enh0.5 and -0.8 plasmids was significantly greater than that of controls \((p < 0.05)\). Other aspects of expression in HeLa cells are similar to that seen in the intestinal cell lines, including the dramatic inhibition of luciferase activity associated with the MUC2 sequence spanning \(-848\) to \(-5183\) \((\text{fragment 4.3, Fig. 1B})\) in the p4.3 construct.

**DISCUSSION**

The human MUC2 gene encodes the major mucin peptide expressed in the intestine \((10, 12)\). In normal intestine, MUC2 is almost exclusively localized in goblet cells, thus it is an important marker for the study of differentiation of this cell lineage. In addition, altered expression of apomucins has been reported to occur in cancer \((10, 11, 18)\). Thus, to investigate the molecular mechanisms that govern the expression of MUC2 during the differentiation process of goblet cells and the alterations that occur in malignant transformation, we undertook cloning of the MUC2 locus. In this paper we report the isolation of a contig spanning the entire MUC2 locus and a partial physical and functional characterization of the MUC2 5'-flanking sequence. In addition, the analysis of the DNA sequence upstream of MUC2 revealed the presence of an additional mucin-like gene, that we refer to as MUCX. This is not a surprising observation, since on chromosome 11p15, at the same general band where MUC2 resides, at least three addi-
tional mucin genes (MUC6, MUC5AC, and MUC5B) have been localized (4–6). The evidence we present for the presence of an additional mucin gene close to MUC2 is 2-fold. First, a contig fragment (probe X, Fig. 1A), used as a probe, detected a mRNA with typical physical characteristics of some of the other mucin mRNAs, namely polydispersity (1), in a mucinous cell line. Second, three areas in probe X showed homology with distinct portions of the MUC2 cDNA, both at the nucleotide and corresponding amino acid level. These regions of similarity lay in the D2 and D3 domains located in the amino-terminal portion of the MUC2 protein, which are characterized by the presence of cysteine residues and by a high degree of sequence similarity to four D-domains in prepro-von Willebrand factor (26). The conservation of number and position of Cys residues is a hallmark of apomucins, and based on analogy with the von Willebrand factor model, it has been suggested that these Cys residues are critical for promoting mucin oligomerization (26).

The identity of MUCX is not clear, but our data and a recent report (19) on the organization of mucin genes at 11p15 narrow the possibilities. The pattern of expression of mRNA detected with the MUCX probe (probe X in Fig. 1A) in HT29 and LS174T cells corresponds to that of both MUC6 and MUC5B, all of which were expressed exclusively in LS174T cells (Table I). In contrast, MUC5AC expression was induced in HT29 cells by FIG. 6. Sequence of the proximal 5'-flanking region of the human MUC2 gene. The sequence of 2600 nucleotides, proximal to the translation start site in the 12-kb 5'-flanking region of the MUC2 gene, is shown. The sequence of the entire 12-kb region has been submitted to the GenBank™ sequence data bank and is available under accession number U68061. The numbers on the left refer to the positions in the sequence relative to the start of translation (+1). The TATA box (in bold, underlined) and the multiple initiation sites (asterisks) as reported by Gum et al. (8) are indicated. Putative recognition sequences for transcription factors are underlined.
FIG. 7

Structure and Regulation of the Human MUC2 Locus

A

HT29

LS174T

Hela

B

HT29

LS174T

Hela

C

HT29

Hela

FIG. 7
forskolin and TPA. Similar results for the expression of MUC6 and MUC5AC have been reported recently by others (35). Thus, MUC5AC is eliminated as a candidate, even though it flanks MUC2 (19). MUC6 has also been reported to flank MUC2, and thus, MUC6 is a stronger candidate for MUCX than is MUC5B.

Our sequence data on MUCX do not yet resolve this issue. Identity between MUCX and MUC6 (5) could not be established, since the partial cDNA sequences for MUC6 in the literature correspond to the tandem repeats and 3′ unique regions of the gene, while the MUCX sequence reported herein is, based on homology to MUC2, most likely in the 5′ end portion of a mucin gene. That MUCX represents a 5′ region is consistent with the good homology found between MUCX and HMG-1, a potential cDNA spanning a more 5′ unique region of the MUC5AC gene (27), but again the lack of sequence identity between MUC5AC and MUCX reinforces the conclusion from the expression data that MUCX is not MUC5AC. Thus, it seems that the genomic sequence of MUCX that we have isolated may correspond to the 5′ portion of MUC6 or of a novel mucin gene, not yet fully characterized. The definite identification of the MUCX gene will require the isolation and characterization of the corresponding cDNA.

The MUCX gene is transcribed in the opposite direction as compared with MUC2, raising the possibility that the two genes may share regulatory elements and pattern of expression. However, our data on the expression of MUCX suggest otherwise. In fact, MUCX is exclusively expressed in LS174T and undetected in HT29 cells, whereas MUC2 expression is high in LS174T and can be induced by TPA and forskolin in HT29 cells. These data suggest that the two genes are independently regulated. This notion is consistent with the specific tissue distribution of mucins whose genes have been mapped on chromosome 11p15 (9).

The contig we have isolated contains 50 kb of DNA upstream of the MUC2 gene. We have sequenced 12 kb of this 5′-flanking region of MUC2 where elements which impart MUC2 tissue-specific and differentiation-dependent regulation may reside. Furthermore, functional analysis was performed using distinct portions of the 12-kb upstream sequence linked to the luciferase reporter gene and transfection of these MUC2/luciferase plasmids into two different intestinal cell lines characterized by a unique pattern of expression of the endogenous MUC2 gene. Our data indicate that elements, which impart promoter activity, are present in the 5′ region of the MUC2 gene. Although two different lines (LS174T and HT29 cells) are characterized by either constitutive high or well inducible levels of MUC2 expression, and they correspond to mucinous and stem-like cells, respectively, there is a clear consistency in the activity associated with the different segments of the MUC2 promoter in these cell lines. A fragment extending from the AUG to nucleotide −848 (fragment 0.8 in Fig. 1B) gives the maximal transcriptional increase over the corresponding control plasmid. This increase is modest, varying between 2- and 5-fold and was detected both when the 0.8 fragment was inserted in an enhancer or basic background. This modest increase is consistent with our previous data suggesting that the MUC2 gene is transcribed at a low rate (14). Moreover, in a basic background, the 5.0 fragment conferred an activity similar to that associated with the 0.8 fragment. Consistent with our results it has been reported that sequences located between −1308 and −641, relative to the cap site, are important for promoter activity when linked to a reporter gene (36). In an enhancer background, however, this fragment exerts a negative effect. Further analysis of this region will establish whether there may be competition for transacting factors between sequences located between −848 and −5183 in the MUC2 5′ region and the SV40 enhancer, as the data suggest.

The down-modulation of the luciferase gene by MUC2 DNA sequences extending further upstream from the 0.8-kb fragment is further documented, in all the cell lines we have tested, by the consistent repression of the luciferase gene when it is driven by the SV40 promoter in the presence of the 4.3-kb MUC2 fragment located between −516 and −5183. The repressor activity of the 4.3-kb fragment is much greater when tested in conjunction with the SV40 promoter. We are investigating whether this repressor activity shows promoter specificity by testing fragment 4.3 in conjunction with additional promoters. Although our data do not distinguish between nonspecific mechanisms due to competition for transcription factors, which are important for MUC2 expression, and the presence of specific negative elements, it is noteworthy that several intestinal genes have been shown to be regulated by the combined action of both positive and negative elements, which ensure both cell lineage-specific expression and proper temporal and spatial expression of the gene in the intestine (37, 38).

Localized to the 0.8-kb fragment is one of the two DNase I-hypersensitive sites that we have mapped in the MUC2 promoter at approximately −600 and −1600 from the AUG. These two hypersensitive sites are present both in unstimulated and stimulated HT29 cells, consistent with our previous data indicating that MUC2 induction in HT29 cells occurs mainly through a post-transcriptional mechanism. The same sites are present in LS174T cells that have very high levels of MUC2 mRNA. A specific role of these hypersensitive sites in the expression of MUC2 in intestinal cells is suggested by the presence of a single different site in the chromatin of HL60 cells, an unrelated cell line that does not express MUC2.

HT29 cells are considered equivalent to stem cells and do not express markers of differentiation, yet they have the same DNase I-hypersensitive sites as LS174T cells, which express high levels of MUC2. This observation suggests that HT29 cells are already committed to differentiation, although not yet lineage-restricted. Indeed, we have shown that, depending on the stimuli, these cells can express simultaneously markers that in the mature cells are cell lineage-restricted (39). Whether the process of lineage restriction is accompanied by a reorganization of the chromatin of those genes that are not expressed in the mature cells is presently not known.

We found that in HeLa cells the pattern of transcriptional activity for each MUC2/luciferase plasmid was similar to that observed in the intestinal cells. However, the level of activity was much greater, being at least 20-fold higher than that detected in intestinal cells, upon correction for transfection efficiency. Although surprising, since HeLa cells do not express the MUC2 gene under any tested conditions (data not shown), it is possible that the activity detected in HeLa cells reflects the

Fig. 7. Functional analysis of the 5′-flanking region of the MUC2 gene. A, the activity of each MUC2/luciferase construct, as schematically represented in Fig. 1B, in the enhancer background, is presented for each of the cell lines tested. B, the activity of constructs, harboring the same MUC2 fragments assayed in A, in the basic background is presented. C shows the repressor activity of a 4.3-kb MUC2 fragment when tested in conjunction with the SV40 promoter. The luciferase activity for each is calculated relative to the activity of the corresponding control plasmid (fold of induction) after correction for transfection efficiency, as determined by the activity of a CMV-β-gal reference plasmid. Names of the plasmids correspond to the sizes of the fragment tested, as illustrated in Fig. 1B. Prefixes enh, b, and p, indicate the background of the corresponding plasmids: enhancer, basic, and promoter, respectively. Each construct was tested in at least three separate experiments in duplicate. Standard errors are indicated with bars.
absence from the transfected DNA of the proper chromatin structure associated with regulatable MUC2 gene expression. For example, we have previously reported that MUC2 mRNA is induced by forskolin and TPA in HT29 cells predominantly at a post-transcriptional level (14). Accordingly, the activity of the MUC2/luciferase plasmids was not significantly modulated in HT29 cells by treatment with these agents, confirming that transcription does not play a prominent role in the regulation of MUC2 in these colonic cells (data not shown). However, the same constructs could be modestly induced by TPA in HeLa cells (data not shown). Inspection of the first 2600 nucleotides upstream from the MUC2 AUG reveals the presence of several consensus AP1 binding sites, the TPA-responsive elements, which can mediate TPA induction, as well as an AP2 consensus site, which also can confer TPA responsiveness (40). Most likely, these sites mediate TPA induction of the MUC2/luciferase constructs in HeLa cells, while they are inactive in HT29 cells, suggesting that, in the context of the MUC2 promoter, these sites are not functional.

The presence in the 5′-flanking region of MUC2 of a number of consensus sequences for transcription factors is not unexpected, and their functional significance requires further investigation. Nonetheless, comparison with the promoter of other genes expressed in the intestine may provide insight into the regulation of MUC2 during differentiation and transformation. Recently, it has been reported that a Sp1-like factor binds to a GC box in the MUC5B gene (41) and may play a role in its regulation in HT29-MTX cells, a methotrexate-resistant clone of HT29 cells characterized by the expression of several mucin genes (42). Whether this factor has any general relevance for the regulation of additional mucin genes, including MUC2 that contains several GC boxes in its 5′ region, has yet to be investigated.

A Sp1-like factor has also been implicated in the regulation of the carcinoembryonic antigen gene (CEA) (29, 30), that can be expressed in goblet cells, as is MUC2, as well as in the absorptive cell lineage. The partial overlapping pattern of expression of CEA and MUC2 (39) suggests that these genes may share some common regulatory elements. Moreover, cell lines derived from mucinous tumors and expressing high levels of MUC2 are also characterized by very high levels of CEA secretion (43). Mucinous tumors generally have a higher frequency of c-myc amplification as compared with more common colorectal tumors (44), although no association between c-myc alterations and MUC2 or CEA deregulation has been established. However, it is worth pointing out that a consensus c-myc binding site, an E box, is located at −1329 from the AUG of MUC2 and that a similar site, present in the CEA promoter, binds USF (29), a member of the basic helix-loop-helix leucin zipper proteins (45), as is c-Myc (46). All the Myc family members share an identical core sequence CACGTG and bind to the corresponding site as homo- or heterodimers (46–48). Whether the USF binding site in the CEA promoter can be modulated by c-Myc in conjunction with any of its partners is not known. However, it is tempting to speculate that MUC2 and CEA may be coordinately deregulated via alterations of common transcription factors, resulting in the progression of the malignant phenotype.