Cloning of the Amino-terminal and 5’-Flanking Region of the Human MUC5AC Mucin Gene and Transcriptional Up-regulation by Bacterial Exoproducts*

Daizong Li, Marianne Gallup, Nancy Fan, David E. Szymkowski, and Carol B. Basbaum

From the Department of Anatomy and Cardiovascular Research Institute, University of California, San Francisco, California 94143 and Roche Bioscience, Palo Alto, California 94304

To obtain gene regulatory sequence for the mucin gene MUC5AC, we have isolated the MUC5AC amino terminus cDNA and 5’-flanking region. This was possible through the use of rapid amplification of cDNA ends-polymerase chain reaction (RACE-PCR) in which the 5’ sequence of the human gastric mucin cDNA HGM-1 (1) was used to design the first MUC5AC-specific primer. Primers for subsequent rounds of RACE were designed from the 5’-ends of amplified RACE products. After five rounds of RACE-PCR, we could no longer generate upstream extensions of the cDNA and hypothesized that we had reached the 5’-end. Primer extension and RNase protection analysis confirmed this. Combined nucleotide sequence for the RACE-PCR products was 3.3 kb with an open reading frame encoding 1100 amino acids. A putative translation start site was found at nucleotide +48. This was followed by a 45 nucleotide putative signal sequence. This amino-terminal sequence contains no tandem repeats but is >60% similar to the amino-terminal nucleotide sequence of MUC2. The positions of cysteine residues in this MUC2-similar region are almost 100% conserved between the two genes. Northern analysis showed expression of cognate RNA in the stomach and airway but not muscle and esophagus. This pattern was the same as that obtained using previously reported 3’-MUC5AC sequences. We have cloned approximately 4 kb of genomic DNA upstream of the transcription start site and have sequenced 1366 nucleotides containing a TATA box, a CACCC box, and putative binding sites for NFkB and Sp 1. Within 4 kb of the transcription start site are elements mediating transcriptional up-regulation in response to bacterial exoproducts.

Mucin is a glycoprotein secreted from epithelial cells at many body surfaces. In the airways, mucin interacts with cilia to trap and clear pathogens and irritants. This mucociliary mechanism is impaired when mucin is produced excessively as in cystic fibrosis, chronic bronchitis, and asthma. Mucociliary impairment leads to airway mucus plugging, which promotes chronic infection, airflow obstruction, and sometimes death.

Nine mucin genes are known to be expressed in man: MUC1-4, MUC5AC, MUC5B, MUC6-8 (2–12). The mRNAs encoding two of them, MUC2 and MUC5AC, have been shown to be up-regulated in cystic fibrosis airways (13, 14) and likely contribute to the airway mucus plugging characteristic of this disease. Insofar as DNA-RNA transcription is controlled by mechanisms amenable to pharmaceutical intervention, an understanding of mucin transcription may suggest ways of inhibiting mucin overproduction.

Both MUC2 and MUC5AC map to chromosome 11p15.5 and may have arisen from a common ancestral gene. The structure of MUC2 is known. Its central region, comprising >50% of the polypeptide, contains two tandem repeat sequences rich in threonine, serine, and proline (4, 17); this is flanked up- and downstream by cysteine-rich regions (17, 18). The threonine and serine residues represent O-glycosylation sites, whereas the cysteine residues are thought to mediate intermolecular interactions underlying mucin gel formation. The isolation of the amino terminus of the MUC2 cDNA by anchor PCR provided sequence for probing a genomic library to obtain the 5’-flanking sequence (17). Using portions of this sequence in luciferase vectors, we identified DNA elements controlling the MUC2 response to the common cystic fibrosis pathogen Pseudomonas aeruginosa (14).

Much less information is available regarding MUC5AC. Understanding the transcriptional control of this gene will require isolation of the amino terminus and 5’-flanking region. To date, MUC5AC amino-terminal cDNAs have not been reported. Although PCR-based techniques can in principle extend existing cDNA fragments over long distances, the large size of the MUC5AC mRNA (10–12 kb) (8, 9), and the potential presence of a central repetitive region present obstacles to extending the existing cDNA sequences to the 5’-end.

A significant aid in this regard was provided by publication of the sequence of cDNA HGM-1 cloned from the human stomach (1). This cDNA likely derives from MUC5AC as nucleotides 1942–2281 are 99% similar to the MUC5AC clone JUL 32 (19) and nucleotides 2190–2541 are 92% similar to the 5’-end of MUC5AC clone NP3a (1). This cDNA was used to design the first set of primers for subsequent rounds of RACE-PCR, in which the 5’-end of the cDNA and putative binding sites for NFkB and Sp 1. Within 4 kb of the transcription start site are elements mediating transcriptional up-regulation in response to bacterial exoproducts.

1. Dohrmann, A., Miyata, S., Gallup, M., Li, J.-D., Chapelin, C., Coste, A., Escudier, E., Nadel, J., and Basbaum, C. (1998) Biochim. Biophys. Acta, in press.

2. The abbreviations used are: PCR, polymerase chain reaction; RACE, rapid amplification of cDNA ends; HGM, human gastric mucin; RPA, RNA protection assay; kb, kilobase(s); bp, base pair(s).

* This work was supported by National Institutes of Health Public Health Service Grants HL 24136 and HL 45762 and a grant from the state of California Tobacco Research and Development Program. 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.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBankTM/EBI Data Bank with accession number(s) AF015521, AF016834.

† To whom correspondence should be addressed. Tel.: 415-476-3835; Fax: 415-476-4845; E-mail: cba@itsa.ucsf.edu.

This paper is available on line at http://www.jbc.org
within 3 kb of the MUC2 transcription start site (17). Hypotheti-
cizing that HGM-1 is present near the MUC5AC 5′-end, we used an HGM-1 sequence as the first gene-specific primer in repetitive 5′-RACE-PCR reactions. This approach ultimately permitted amplification of a 3.3-kb upstream extension of HGM-1, which we call MUC5AC 5′-RACE product (MUC5AC-5′RP). Primer extension, RNase protection assays, and the presence of a translation start site and putative signal se-
quence indicate that this sequence is at the gene’s 5′-end. Genomic DNA immediately upstream of MUC5AC-5′RP has the structural properties of a promoter and contains elements mediating transcriptional up-regulation in response to bacte-
rial exoproducts. We conclude that our cloned sequences are the amino-terminal and 5′-flanking region of MUC5AC. The availability of these sequences should aid identification of the elements controlling MUC5AC overexpression in disease.

MATERIALS AND METHODS

Cell Culture—The human lung epithelial carcinoma cell line NCIH292 was grown in RPMI 1640 medium supplemented with 10% heat-inactivated fetal calf serum (Life Technologies, Inc.). The human colon carcinoma cell line H93 (4) was grown in Dulbecco’s modified Eagle’s medium with high glucose and 10% fetal calf serum. In some experi-
ments, cells were exposed for 6 or 24 h to P. aeruginosa.

Bacterial Culture and Preparation of Cell-free Supernatants—P. aeruginosa strain PA01 was grown in M9 buffer (20) for 72 h at 37 °C (to late log phase). Cell-free supernatant was obtained by centrifugation at 10,000 rpm for 60 min at 4 °C and by filtration through a 0.22-μm filter (Corning). Supernatant was aliquoted and stored at −80 °C until used.

Exposure of Tissues and Cells to Bacterial Cell-free Filtrates—To look at the effects of P. aeruginosa on MUC5AC steady state mRNA, incuba-
tion was as described (21). Briefly, cells were washed twice with phosphate-buffered saline at 37 °C. Samples were then incubated with bacterial supernatant or buffer (M9) diluted 1:4 with mammalian cell culture medium for 6 h. Total RNA was obtained from pelleted cells scraped from the culture dish (22). Lactate dehydrogenase release was measured (LDH 320, Sigma) to detect any cell lysis.

cDNA Synthesis and 5′-RACE-PCR—Sources known to contain abundant MUC5AC mRNA (P. aeruginosa-exposed NCIH292 cells or human stomach) were subjected to RNA extraction (22). Total RNA (3 μg) was used to generate double-stranded cDNA following the Gene-amp cDNA Amplification kit (CLONTECH). The double-stranded cDNA was ligated with the Marathon cDNA adaptor and purified on a chromaspin +TE-1000 column (CLONTECH) in a total volume of 100 μl. 5′-RACE was performed using the double-stranded cDNA template with one HGM-1 gene-specific primer (Gm1) and the adaptor primer AP1 or AP2. Additional gene-specific primers (Gm5, Gm9, Gm9G, and GmH9) were generated based on the sequences of progressively amplified 5′-RACE products.

Northern Blot Analysis of Tissue Distribution of MUC5AC mRNA, Results Using MUC5AC-5′RP and NP3a Probes—Total RNA was ex-
tracted from human tissues according to previously described methods (22). RNA samples (20 μg) were separated on 1.0% agarose gels con-
taining 2.2 μg formaldehyde and then transferred to a positively charged nylon membrane (Genene Screen, NEN Life Science Products). cDNA probes were labeled with [α-32P]dCTP using a Life Technologies, Inc. random prime labeling kit. For the MUC5AC 3′-end, a cDNA fragment was amplified from tissue mRNA using primers NP3a5′ and NP3a5. The insert for the probe was gel-purified from a construct made by TA ligating with the Marathon cDNA adaptor and purified on a chromaspin filtrate (Corning). Supernatant was aliquoted and stored at −20 °C. In the absence of the filter (Corning). Supernatant was aliquoted and stored at −80 °C. In some experi-
ments, cells were exposed for 6 or 24 h to P. aeruginosa.

Bacterial Culture and Preparation of Cell-free Supernatants—P. aeruginosa strain PA01 was grown in M9 buffer (20) for 72 h at 37 °C (to late log phase). Cell-free supernatant was obtained by centrifugation at 10,000 rpm for 60 min at 4 °C and by filtration through a 0.22-μm filter (Corning). Supernatant was aliquoted and stored at −80 °C until used.

Exposure of Tissues and Cells to Bacterial Cell-free Filtrates—To look at the effects of P. aeruginosa on MUC5AC steady state mRNA, incuba-
tion was as described (21). Briefly, cells were washed twice with phosphate-buffered saline at 37 °C. Samples were then incubated with bacterial supernatant or buffer (M9) diluted 1:4 with mammalian cell culture medium for 6 h. Total RNA was obtained from pelleted cells scraped from the culture dish (22). Lactate dehydrogenase release was measured (LDH 320, Sigma) to detect any cell lysis.

cDNA Synthesis and 5′-RACE-PCR—Sources known to contain abundant MUC5AC mRNA (P. aeruginosa-exposed NCIH292 cells or human stomach) were subjected to RNA extraction (22). Total RNA (3 μg) was used to generate double-stranded cDNA following the Gene-amp cDNA Amplification kit (CLONTECH). The double-stranded cDNA was ligated with the Marathon cDNA adaptor and purified on a chromaspin +TE-1000 column (CLONTECH) in a total volume of 100 μl. 5′-RACE was performed using the double-stranded cDNA template with one HGM-1 gene-specific primer (Gm1) and the adaptor primer AP1 or AP2. Additional gene-specific primers (Gm5, Gm9, Gm9G, and GmH9) were generated based on the sequences of progressively amplified 5′-RACE products.

Northern Blot Analysis of Tissue Distribution of MUC5AC mRNA, Results Using MUC5AC-5′RP and NP3a Probes—Total RNA was ex-
tracted from human tissues according to previously described methods (22). RNA samples (20 μg) were separated on 1.0% agarose gels con-
taining 2.2 μg formaldehyde and then transferred to a positively charged nylon membrane (Genenc Screen, NEN Life Science Products). cDNA probes were labeled with [α-32P]dCTP using a Life Technologies, Inc. random prime labeling kit. For the MUC5AC 3′-end, a cDNA fragment was amplified from tissue mRNA using primers NP3a5′ and NP3a5. The insert for the probe was gel-purified from a construct made by TA ligating the PCR product into pCRII vector (Invitrogen). For the new sequence MUC5AC-5′RP, probes were made from amplified fragments using primers TER and GM9. Labeled probe was added to 10 ml of hybridization buffer containing 50% formamide, 10% dextran sulfate, 0.2% Denhardt’s, 50 μM TRIS-HCl, pH 7.5, 1 mM NaCl, and 0.1% sodium pyrophosphate to give a concentration of 2–5 × 106 cpm/ml. Membrane hybridization and washing were performed using conditions described previously (23).

Primers—Primers used for 5′-RACE, construction of Northern blot probes, genomic library screening and DNA walking, primer extension, and RNase protection assays are shown in Table I.

DNA Cloning and Sequencing—After RACE-PCR, amplified frag-
ments were purified by low-melting point agarose gel electrophoresis, cut with appropriate restriction enzymes and cloned into pBluescript II SK− (Stratagene) or sequenced directly. Escherichia coli (SURE strain, Stratagene) was transformed with plasmids containing these fragments. Transformants were grown at 37 °C or 30 °C. Both sense and antisense strands were sequenced. Sequencing reactions were car-
ried out using SequiTherm Long-Read cycle sequencing kits (Epigenetic Technologies) and Thermo Sequenase fluorescent labeled primer cycle

Human MUC5AC Gene Amino Terminus and Promoter

6813
assayed for luciferase activity using LucLite reagent (Packard) and a TopCount luminometer (Packard).

RESULTS

RACE-PCR, cDNA Cloning and Sequence Determination of the MUC5AC Amino Terminus—Based on the 99% sequence identity between MUC5AC clone JUL 32 (19) and HGM-1 nucleotides 1947–2278 (1), we hypothesized that HGM-1 is a part of MUC5AC. Based on >60% similarity between HGM-1 and the amino-terminal cysteine-rich domain of MUC2 (D-domain 3), we hypothesized that HGM-1 is an amino-terminal sequence. This led us to initiate 5' RACE-PCR experiments aimed at extending HGM-1 to the MUC5AC transcription start site (Fig. 1).

In the first round of 5'-RACE-PCR, we used an HGM-1-specific primer (GM1) and an adaptor primer (AP1). This yielded a 900-bp PCR fragment. Sequence data showed that this fragment was the 5' extension of human gastric mucin (HGM) (1) and was >65% similar to the MUC2 D-domain 3 just 5' to the central repeat region. Primer GM5 was designed based on the 5'-end of this fragment and was used in a second round of 5'-RACE-PCR. This generated an 1100-bp PCR fragment whose 5'-end was used to design primer GM9. When used in a third round of 5'-RACE-PCR, GM9 generated a 700-bp fragment. Primer GM9G was designed based on the 5'-end of this fragment and was used in a fourth round of RACE-PCR to generate a 600-bp fragment. Primer GM9H, 103-bp downstream of the 5'-end of the fourth round RACE-PCR product, was used in a fifth round of RACE-PCR and generated a 110-bp product. Repeated efforts to generate larger products with primer GM9H from both gastric tissue and NCIH292 (airway) cell cDNA yielded PCR products with identical sequence that were ~100 bp in length. This suggested that GM9H was approximately 100 bp from the 5'-end of the mRNA as processed in both gastric tissue and NCIH292 cells.

The overall cDNA sequence obtained by 5'-RACE-PCR was 65% similar to the MUC2 amino-terminal cysteine-rich domain (D-domain 3). At +48 this is an ATG codon embedded in a Kozak consensus sequence (24). This is a putative translation start site. Following this is a putative secretory protein signal sequence. The entire open reading frame encodes 1100 amino acids. The nucleotide sequence is approximately 65% similar to the MUC2 amino-terminal sequence (1–3500, Fig. 3A). No random repeat sequence is present, but there are three cysteine-rich domains (D1-D3) in which the cysteine positions correspond almost exactly to those previously described for the amino terminus of human MUC2 (Fig. 3B).

Northern Blot Analysis of Tissue Distribution of RNA Corresponding to Newly Cloned Sequence—Our interpretation that the 5' extension of HGM-1 (MUC5AC-5'RP) is at the 5'-end of MUC5AC rests primarily on the 99% similarity between a portion of HGM-1 and the MUC5AC cDNA JUL 32 (19). Further confirmation of the identity between our new sequence and MUC5AC was provided by Northern blot analysis in which we observed that a probe from our new sequence showed tissue-specific hybridization identical to that obtained using a probe from the previously described MUC5AC C-terminal cDNA NP3a (8) (Fig. 4).

| Name | Sequence | Location |
|------|----------|----------|
| GM1  | GGCCCTGGAGGATGCTGACTGCTTCT | 3348/3322 |
| GM5a | GATGAGCGAACACTGTGAGGTTGCCC | 2255/2227 |
| GM9  | AGAATTGGAGCCTCACAAGCAAGACCC | 1374/1348 |
| GM9G | AAGTCCCCAGAGCCACAGCTCTGTTG | 698/668 |
| GM5H  | GTGTGCGCGAGCAAGCCAGGGAAGG | 119/90 |
| AP1  | ACTCCTATTAGAGCTGACGCGC | Adaptor |
| AP2  | GTCATAGATGCTGACGCGC | Adaptor |

* Also used to generate the Northern probe.  
* Also used for primer extension.  
* Numbered according to Meezaman et al. (8) and also used for Northern probes.
The positions of nucleotides are indicated by the numbers on the left and amino acids on the right. The hydrophobic putative secretory protein signal sequence is underlined.

Fig. 2. Nucleotide and deduced amino acid sequence of MUC5AC-5'-RP. The positions of nucleotides are indicated by the numbers on the left and amino acids on the right. The hydrophobic putative secretory protein signal sequence is underlined.
Chromosome Mapping—Human chromosome 11p15 contains a mucin gene cluster currently known to include MUC5AC as well as MUC5B, MUC6 and MUC2. To obtain further supporting evidence that the newly cloned RACE-PCR sequence is part of MUC5AC, we performed chromosomal mapping experiments. As shown in Fig. 5, MUC5AC-5' primers amplified a product from mouse-human hybrid cell line 1049, but not from cell line 1079. As both cell lines contained DNA from chromosome 5 but only 1049 contained DNA from chromosome 11, the results clearly show that our RACE product MUC5AC-5' RP maps to chromosome 11. This is consistent with identification of this product as part of MUC5AC.

Primer Extension and RNase Protection Analysis-MUC5AC—Fig. 3. Similarity matrix of cDNA and alignment of protein for MUC5AC-5'RP and MUC2. A, the cDNAs were compared using the MacVector similarity matrix routine with a cutoff of 65% similarity. Lines indicate regions of similarity. B, alignment of amino acids was done using MacVector alignment program. Solid lines represent identity and double dots (*) indicate conservative substitutions. Conserved cysteine residues are indicated by asterisks (*) and the D-domains are indicated by bent lines.
The primer extension reaction yielded a product of 114 bp (Fig. 6). This was consistent with the 3’-end (Fig. 2) suggesting that its 5’-end is at or near the transcription start site. To investigate this, we performed primer extension and RNase protection analysis. For primer extension, we used primer GM9H, which is approximately 100 bp upstream of the 5’-end of our RACE-PCR product as estimated from agarose gels. The primer extension reaction yielded a product of 114 bp (Fig. 6A) when RNA from gastric tissue or airway cells was used as a template, supporting the view suggested by RACE-PCR that the transcription start site was approximately 100 bp upstream of primer GM9H.

For RNase protection analysis, we used as probe a portion of MUC5AC RNA in epithelial cells exposed to P. aeruginosa or its exoproducts in cell-free supernatants (Fig. 8A). That this was controlled at the transcriptional level was indicated by 15–20-fold induction of transcriptional activity in epithelial cells stably transfected with MUC5AC-luciferase reporter constructs and exposed to P. aeruginosa (Fig. 8B). These findings indicate the presence of elements responsive to P. aeruginosa in the 4-kb DNA fragment immediately upstream of the MUC5AC transcription start site. Analysis of deletion mutants will permit precise identification of these elements and open the way to identification of cognate transcription factors.

**DISCUSSION**

In this series of studies, we isolated the amino terminus and 5’-flanking region of the MUC5AC mucin gene as a first step toward understanding the dysregulation of mucin mRNA production in the airways of cystic fibrosis patients. Hypothesizing that the previously reported cDNA HGM-1 was relatively upstream in the MUC5AC sequence, we performed progressive RACE-PCR amplifications that eventually reached the transcription start site. We used a similar approach to isolate the 5’-flanking region from genomic DNA.

Evidence That HGM-1 and Its 5’-RACE Extended Product Are Part of MUC5AC—HGM-1 is a human gastric mucin cDNA (1) containing cysteine clusters interspersed with threonine-, serine-, and proline-rich domains (1). Cysteine-rich domains are considered to be typical of mucin sequences, having been reported in many mucins including MUC2 (18, 19), MUC5AC (8), MUC5B (7), and MUC6 (10) as well as in rat (23), pig (26), cow (27), and frog (28) mucins. The cysteine-rich domains in mucins show varying degrees of similarity to the D-domains of von Willebrand factor.

The evidence that HGM-1 is part of MUC5AC essentially rests on the observation that HGM-1 nucleotides 1942–2281 are 99% similar to the MUC5AC clone JUL 32 (19) and nucleotides 2190–2541 are 92% similar to the 5’-end of MUC5AC clone NP3a (1). By the same reasoning, HGM-1’s extended 5’-sequence MUC5AC-5’RP is at the 5’-end of the mRNA and that the designated transcription start site, +1 is accurate. Present in the putative promoter region are NFκB, Sp-1, GRE, AP-2, and CACCC box sites. Up-regulation of MUC5AC Transcriptional Activity by P. aeruginosa—Availability of the upstream regulatory region permits analyses of potential abnormalities in MUC5AC transcription in disease models. We observed large inductions of MUC5AC RNA in epithelial cells exposed to P. aeruginosa or its exoproducts in cell-free supernatants (Fig. 8A). That this was controlled at the transcriptional level was indicated by 15–20-fold induction of transcriptional activity in epithelial cells stably transfected with MUC5AC-luciferase reporter constructs and exposed to P. aeruginosa (Fig. 8B). These findings indicate the presence of elements responsive to P. aeruginosa in the 4-kb DNA fragment immediately upstream of the MUC5AC transcription start site. Analysis of deletion mutants will permit precise identification of these elements and open the way to identification of cognate transcription factors.
structure of the MUC5AC carboxyl terminus has been known since the cloning of NP3a, a cDNA isolated from a nasal polyp library. Its identification as part of MUC5AC rests on the fact that cDNAs containing part of the NP3a sequence had previously been designated as MUC5 (25) and were later designated as MUC5AC (19). The recognition that NP3a comprises the gene’s 3′-end rests on its containing a polyadenylation signal and poly(A) tail. It also contains a homologue of the MUC2 D-domain 4 (8). A similar cDNA, L31, was isolated from an HT29 (colonic carcinoma) cell library (9).

Other than the positioning of NP3a and L31 at the 3′-end, it has not been possible to assign any of the other known MUC5AC cDNAs to particular positions in the coding sequence. cDNAs containing threonine/serine/proline-rich repetitive sequences (JER 47, JER 58, Mar 2, 10, 11, and CEL 2) from a tracheobronchial library (19) and cDNA 4F from a stomach library (15) are assumed to occupy positions in a central part of the gene based on comparisons with MUC2. Our Northern blots (Fig. 4) suggest that the size of the full-length MUC5 AC mRNA is 12–14 kb and possibly larger. Subtracting the amount of currently known sequence at the 5′- and 3′-ends, we estimate the size of the repeat region in MUC5AC to be at least 6 kb. The size of the repeat region in MUC2 is 8.4 kb.

Despite considerable interest in the gene 5′-end and promoter, prior to this report no 5′ cDNAs had been conclusively identified. Klomp et al. (1) had noted, however, that HGM-1, a cDNA isolated from a gastric cDNA library, is 60% similar to the MUC2 D-domain 3, which is approximately 3 kb downstream of the MUC2 5′ transcription start site. Taken together with evidence that HGM-1 is part of MUC5AC (see above), its similarity to this upstream region of MUC2 suggested that HGM-1 might comprise an upstream region of MUC5AC. Our 5′-RACE-PCR studies yielded a 3.3-kb 5′ extension of HGM-1. Despite considerable interest in the gene 5′-end and promoter, prior to this report no 5′ cDNAs had been conclusively identified. Klomp et al. (1) had noted, however, that HGM-1, a cDNA isolated from a gastric cDNA library, is >60% similar to the MUC2 D-domain 3, which is approximately 3 kb downstream of the MUC2 5′ transcription start site. Taken together with evidence that HGM-1 is part of MUC5AC (see above), its similarity to this upstream region of MUC2 suggested that HGM-1 might comprise an upstream region of MUC5AC. Our 5′-RACE-PCR studies yielded a 3.3-kb 5′ extension of HGM-1. Within the MUC5AC sequence reported here are homologues of MUC2 D-domains 1 and 2 and the 5′-end of D-domain 3. Upstream of D-domain 1 is a signal peptide and translation start site with considerable interest in the gene 5′-end and promoter, prior to this report no 5′ cDNAs had been conclusively identified. Klomp et al. (1) had noted, however, that HGM-1, a cDNA isolated from a gastric cDNA library, is >60% similar to the MUC2 D-domain 3, which is approximately 3 kb downstream of the MUC2 5′ transcription start site. Taken together with evidence that HGM-1 is part of MUC5AC (see above), its similarity to this upstream region of MUC2 suggested that HGM-1 might comprise an upstream region of MUC5AC. Our 5′-RACE-PCR studies yielded a 3.3-kb 5′ extension of HGM-1. Within the MUC5AC sequence reported here are homologues of MUC2 D-domains 1 and 2 and the 5′-end of D-domain 3.
ease. Although MUC5AC has been recognized for some time to encode an airway mucin, it was only recently discovered that its expression is up-regulated in airway disease. The work reported here is the first to establish that this up-regulation is controlled at the transcriptional level and that key cis- and trans-activating factors operate within 4 kb upstream of the transcription start site. Availability of the newly cloned sequence will permit precise identification of transcriptional control mechanisms and will facilitate elucidation of upstream signaling pathways as well.

**Acknowledgments**—We thank Jian-Dong Li, M.D., Ph.D. for helpful discussion and providing the P. aeruginosa supernatants.

**REFERENCES**

1. Klomp, L., van Rens, L., and Strous, G. (1995) *Biochem. J.* 308, 831–838
2. Gendler, S. J., Lancaster, C. A., Taylor-Papadimitriou, J., Dubig, T., Peat, N., Burchell, J., Pemberton, L., Lalani, E. N., and Wilson, D. (1990) *J. Biol. Chem.* 265, 15286–15293
3. Gendler, S., Spicer, A., Lalani, E. N., Dubig, T., Peat, N., Burchell, J., Pemberton, L., Boshell, M., and Taylor-Papadimitriou, J. (1991) *Am. Rev. Respir. Dis.* 144, S42–S47
4. Gum, J. R., Byrd, J. C., Hicks, J. W., Toribara, N. W., Lamport, D. T. A., and Kim, Y. S. (1989) *J. Biol. Chem.* 264, 6480–6487

**FIG. 8.** Up-regulation of MUC5AC promoter activity in response to bacterial exoproducts. A, RNase protection assay showing up-regulation of the endogenous MUC5AC mRNA in response to *P. aeruginosa*. In both the airway epithelial cell line NCIH292 and the colon epithelial cell line HM3, MUC5AC-protected mRNA fragments are more abundant in the presence than in the absence of *P. aeruginosa*. All cells were studied after 6 h of incubation. Lane 1, low serum medium; lane 2, PAO1 supernatant; low serum medium, 1:4; lane 3, PAO1 live bacteria. GAPDH- and cyclophilin-protected RNA fragments reflect the amount of RNA incubated per lane. B, luciferase reporter assays. The MUC5AC upstream flanking region (~4000/+681) and a luciferase reporter were cloned into pGL3 basic vector and transfected into HM3 cells. G418-selected colonies were pooled and expanded before 24 h of exposure to *P. aeruginosa* supernatant at the concentrations shown on the x-axis. Fold induction (mean of data from three separate bacterial colonies ± S.E.) in luciferase activity is shown. The data indicate that the 4-kb flanking region contains one or more elements responsive to *P. aeruginosa* exoproducts.

**FIG. 9.** Diagram comparing elements of the MUC2 structure with those known so far of the MUC5AC structure. Positions of representative MUC5AC cDNAs are shown. The four cysteine-rich D-domains are >60% similar between MUC2 and MUC5AC. The size of the MUC2 repeat region is 8.4 kb. The corresponding region for MUC5AC is estimated at ≥6 kb (see "Discussion"). Regions with vertical hatching are threonine/serine/proline-rich repeats, and regions with oblique hatching are cysteine-rich domains.
Human MUC5AC Gene Amino Terminus and Promoter

5. Gum, J. R., Hicks, J. W., Swallow, D. M., Lagace, R. L., Byrd, J. C., Lamport, D. T. A., Siddiki, B., and Kim, Y. S. (1990) *Biochem. Biophys. Res. Commun.* **171**, 407–415

6. Porchet, N., Van Cong, N., Dufosse, J., Audie, J. P., Guyonnet-Duperat, V., Gross, M. S., Denis, C., Degand, P., Bernheim, A., and Aubert, J. P. (1991) *Biochem. Biophys. Res. Commun.* **175**, 414–422

7. Dusseyn, J.-L., Guyonnet-Duperat, V., Porchet, N., Aubert, J.-P., and Laine, A. (1997) *J. Biol. Chem.* **272**, 3168–3178

8. Meeezaman, D., Charles, P., Daskal, E., Polymeropoulos, M., Martin, B., and Rose, M. (1994) *J. Biol. Chem.* **269**, 12952–12959

9. Lesuffleur, T., Roche, F., Hill, A., Larasa, M., Fox, M., Swallow, D., Zweibaum, A., and Real, F. (1995) *J. Biol. Chem.* **270**, 13665–13673

10. Toribara, N., Ho, S. B., Gum, E., Gum, J. R., Jr., Lau, P., and Kim, Y. S. (1997) *J. Biol. Chem.* **272**, 16398–16403

11. Bobek, L., Tsai, H., Biesbrock, A., and Levine, M. (1993) *J. Biol. Chem.* **268**, 20563–20569

12. Shankar, V., Gilmore, M., Elkins, R., and Sachdev, G. (1994) *Biochem. J.* **300**, 295–298

13. Gendler, S., Madsen, C., Aust, M., Yankaskas, J., Jennings, J., and Kasperbauer, J. (1996) *Pediatric Pulmonology* **13S**, 290 (abstr.)

14. Li, J.-D., Dehrman, A., Gallup, M., Miyata, S., Gum, J., Kim, Y., Nadel, J., Prince, A., and Basbaum, C. (1997) *Proc. Natl. Acad. Sci. U. S. A.* **94**, 967–972

15. Ho, S., Roberton, A., Shekele, L., Lyftogt, C., Niehans, G., and Toribara, N. (1995) *Gastroenterol.* **109**, 735–747

16. Deleted in proof

17. Gum, J. R. Jr., Hicks, J., Toribara, N., Siddiki, B., and Kim, Y. (1994) *J. Biol. Chem.* **269**, 2440–2446

18. Gum, J., Hicks, J., Toribara, N., Rothe, E., Lagace, R., and Kim, Y. (1992) *J. Biol. Chem.* **267**, 21375–21383

19. Guyonnet-Duperat, V., Audie, J., Debaillieul, V., Laine, A., Buisine, M., Gallegue-Zouitina, S., Figny, P., Degand, P., Aubert, J.-P., and Porchet, N. (1995) *Biochem. J.* **311–219

20. Jackowski, J., Szepfalusi, Z., Wanner, D., Seybold, Z., Sielezak, M., Laurende, I., Adams, T., Abraham, W., and Wanner, A. (1991) *Am. J. Physiol.* **260**, L61–L67

21. Massion, P., Inoue, H., Richman-Eisenstat, Grunberger, D., Jorens, P., Houset, B., Pittet, J.-F., Wiener-Kronisch, J., and Nadel, J. (1994) *J. Clin. Invest.* **93**, 26–32

22. Chomczynski, P., and Sachi, N. (1987) *Anal. Biochem.* **162**, 156–159

23. Ohmori, H., Dohrmann, A. F., Gallup, M., Tsuda, T., Kii, H., Gum, J. R., Jr., Kim, Y., and Basbaum, C. (1994) *J. Biol. Chem.* **269**, 17833–17840

24. Kozak, M. (1991) *J. Biol. Chem.* **266**, 19867–19870

25. Aubert, J.-P., Porchet, N., Crepin, M., Duterque-Coquillaud, M., Verges, G., Mazruca, M., Debuire, B., Petiprez, D., and Degand, P. (1991) *Am. J. Respir. Cell Mol. Biol.* **5**, 175–185

26. Eckhardt, A. E., Timpte, C. S., Abernethy, J. J., Zhao, Y., and Hill, R. L. (1991) *J. Biol. Chem.* **266**, 9678–9686

27. Bhargava, A. K., Witten, J. T., Davidson, E. A., and Bhavanandan, V. P. (1990) *Proc. Natl. Acad. Sci. U. S. A.* **97**, 6798–6802

28. Probst, J., Gertz, E.-M., and Hoffmann, W. (1990) *Biochemistry* **29**, 6240–6244
Cloning of the Amino-terminal and 5′-Flanking Region of the Human MUC5AC Mucin Gene and Transcriptional Up-regulation by Bacterial Exoproducts
Daizong Li, Marianne Gallup, Nancy Fan, David E. Szymkowski and Carol B. Basbaum

J. Biol. Chem. 1998, 273:6812-6820.
doi: 10.1074/jbc.273.12.6812

Access the most updated version of this article at http://www.jbc.org/content/273/12/6812

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 26 references, 15 of which can be accessed free at http://www.jbc.org/content/273/12/6812.full.html#ref-list-1