BIOCHEMICAL LETTERS

Further evidence that the human MUC2 gene transcripts in the intestine and trachea are identical

The human MUC2 mucin has been the only well-characterized intestinal mucin to date. It has an enormous protein backbone of greater than 5000 amino acids that can be divided into three major structural domains. The longest of these domains is the central tandem repeat array, flanked on either side by Cys-rich regions. The Cys-rich regions of MUC2 consist, in large part, of a repetitive element with a high degree of sequence similarity to the D domains of prepro-von Willebrand factor [1,2]. The complete cDNA sequence is to be found in the GenBank and EMBL data bases with the accession number L2998.

Based on Northern-blot experiments with the MUC2 tandem repeat probe that produces a polydisperse signal with mRNA from both organs, previous studies [3–5] have shown that the MUC2 intestinal gene is also expressed in the airways. Also, Gerard et al. [3] isolated a partial cDNA from a cystic fibrosis tracheal library that consisted entirely of repetitive sequences and showed remarkable identity with the tandem repeats present in the MUC2 transcript in the intestine. Polydispersity is typical of mucin mRNA and may be related to the effect of alternative splicing or partial degradation during turnover of the large mucin mRNA [6]. Since tissue-specific alternative splicing or tissue-specific processing of a single message could lead to similar, but not identical, transcripts in different tissues, we investigated whether the MUC2 mucin sequence in the airways, outside of the 69 bp tandem repeat domain, is identical with that in the intestine. Human tracheal mRNA was used to construct a cDNA library in the Uni-ZAP vector (Stratagene, La Jolla, CA, U.S.A.) using an oligo(dT) linker-primer to initiate first-strand synthesis as per the manufacturers' protocol. The library was screened with the cDNA probe MP-4 of 598 bp, corresponding to nucleotide

positions 14350–14947 at the C-terminus of the intestinal MUC2 transcript obtained from Dr. James R. Gum, Jr., University of California, San Francisco, CA, U.S.A. [2]. A single plaque that hybridized to the probe was isolated, subcloned into pBluescript vector and sequenced on both strands of the DNA, by the dideoxy-chain-termination method using [α - 35 S]dATP and Sequenase 2.0 (U.S. Biochemicals, Cleveland, OH, U.S.A.), in their entirety. Sequence analyses revealed a 1842 bp insert with exact match with the sequence published for the intestinal mucin transcript (bases 13879–15720), except for base 14765, which in our clone was T instead of C (Figure 1, region A). This sequence difference, however, reflects a third position substitution which does not change the amino acid encoded, and could be due to a polymorphism at this site. The PCR approach was used to amplify specific regions of the MUC2 transcript. First-strand cDNA synthesis for reverse-transcription PCR was carried out by using 100 ng of human tracheal mRNA (Clontech Laboratories, Palo Alto, CA, U.S.A.) and a mixture of oligo(dT) and random hexamer primers. Superscript II (Life Technologies, Gaithersburg, MD, U.S.A.) was used for reverse transcription in a typical reaction volume of 20 μ l as specified by the manufacturer. The final reaction products were diluted to 100 μ l, and 5 μ l samples were used for PCR amplification.

PCR amplifications (30 cycles) were carried out in 50 μ l volumes using *Taq* polymerase (Perkin-Elmer Corp., Norwalk, CT, U.S.A.). Each cycle consisted of 1 min denaturation at 94 °C, 1 min annealing at 65 °C and 2 min extension at 72 °C, with a final extension at 72 °C for 10 min. The reaction products were analysed by electrophoresis, and bands of interest were excised from the gel, purified using Gene-Clean II (Bio-101 Inc., La Jolla, CA, U.S.A.) and cloned into a modified pBluescript/T-vector. The nucleotide sequences of the cDNA inserts from three independent clones were determined as described above, either by using nested deletions obtained by using exonuclease III and S1 nuclease or by subcloning restriction-enzyme-digested

Figure 1 Schematic representation of the identity between the MUC2 gene transcripts in the intestine (upper block) and trachea (lower segment)

The bracketed regions indicate the various domains of the full-length MUC2 cDNA along with the nucleotide position where each domain ends. 'A' denotes the 1842 bp cDNA isolated from the tracheal cDNA library by conventional screening. 'B' and 'C' are the 308 bp and 465 bp cDNA clones, respectively, amplified by PCR from tracheal mRNA. 'D' denotes the 1609 bp fragment spanning the region of imperfect repeats 5' to the 69 bp tandem repeat region. The arrows indicate the location of the various primers used for PCR amplifications. The broken line shows the sequence identity in the conserved 69 bp tandem repeat domain reported previously [3,4].

fragments. When necessary, dITP was substituted for dGTP to resolve band compressions and ambiguities.

To investigate whether the region 5' to that of the tandem repeat portion was conserved, we conducted PCR amplifications using the primers MUC2A (5' > GTGGGTTGGGTGACACACTC < 3') and MUC2S (5' > CCATGCGTGCCTCTCTGCAA < 3') initially. This resulted in the specific amplification of an expected 308 bp fragment, and subsequent sequencing indicated an identical match of the sequence (B in Figure 1) with the sequence published for the intestinal MUC2 transcript (bases 5368–5675). To determine further whether the 308 bp non-repeat region was immediately upstream of the 69 bp tandem repeat domain, as in the MUC2 intestinal transcript, we performed PCR amplifications on human tracheal first-strand cDNA by using the MUC2S primer and a third primer MUC2CFT (5' > TGTTGGGGTTGGGGTCACCGTGGTGGTGGT < 3'). The reaction products when electrophoresed on an agarose gel showed a series of bands in a ladder-like fashion, as expected. A prominent band around 450 bp was excised from the gel, subcloned and sequenced. The sequence (C in Figure 1) once again exactly matched that reported previously [2] in the MUC2 intestinal transcript (bases 5368–5832).

To investigate if the region of imperfect repeats 5' to the 69 bp tandem repeat domain was similar in both intestinal and tracheal MUC2 transcripts, we attempted to amplify across the region of imperfect repeats, using the primers MUC2S1 (5' > GCTTGGAGCAGCATGGCCAG < 3') and MUC2A. The reaction products when analysed on an agarose gel showed a discrete band around 1.6 kb, which was also subcloned. After confirming on Southern blots that the insert DNA from positive colonies reacted strongly with the MUC2S internal primer, nested deletions were generated from plasmid DNA by using exonuclease III and S1 nuclease and were sequenced. The resulting

sequence indicated a 1609 bp fragment (D in Figure 1), which matched exactly that of the MUC2 intestinal transcript from bases 4067–5675.

In summary, although previous reports have suggested that the MUC2 gene products in the intestine and the trachea are perhaps similar, our data provide further evidence regarding the sequence identity of MUC2 transcripts in both the intestine and the trachea in the region of imperfect repeats, sequence immediately upstream of the 69 bp tandem repeats and the extreme C-terminus. It remains possible, however, that alternative transcripts are present as well, and differences may occur in the extreme N-terminus of the MUC2 gene transcripts in the intestine and trachea.

This work was supported, in part, by a NIH grant HL34012.

VISWANATHAN SHANKAR*, MICHAEL S. GILMORE† and GOVERDHAN P. SACHDEV*‡

*College of Pharmacy and †College of Medicine, University of Oklahoma Health Sciences Center, P.O. Box 26901, Oklahoma City, OK 73190, U.S.A.

‡ To whom correspondence should be addressed.

- 1 Gum, J. R., Jr., Hicks, J. W., Toribara, N. W., Rothe, E. M., Lagace, R. E. and Kim, Y. S. (1992) *J. Biol. Chem.* **267**, 21375–21383
- 2 Gum, J. R., Jr., Hicks, J. W., Toribara, N. W., Siddiki, B. and Kim, Y. S. (1994) *J. Biol. Chem.* **269**, 2440–2446
- 3 Gerard, C., Eddy, R. L., Jr., and Shows, T. B. (1990) *J. Clin. Invest.* **86**, 1921–1927
- 4 Jany, B. H., Gallup, M. W., Yan, P. S., Gum, J. R., Kim, Y. S. and Basbaum, C. B. (1991) *J. Clin. Invest.* **87**, 77–82
- 5 Shankar, V., Tan, S., Gilmore, M. S. and Sachdev, G. P. (1992) *Biochem. Biophys. Res. Commun.* **189**, 958–964
- 6 Gum, J. R., Jr. (1992) *Am. J. Respir. Cell Mol. Biol.* **7**, 557–564