Biochemical and Biophysical Research Communications **261,** 83–89 (1999)
Article ID bbrc.1999.1001, available online at http://www.idealibrary.com on IDEAL®

The MUC3 Gene Encodes a Transmembrane Mucin and Is Alternatively Spliced

Stephanie J. Williams,*[,1] David J. Munster,* Rachel J. Quin,*
David C. Gotley,† and Michael A. McGuckin*

**Mater Medical Research Institute, Level 3, Aubigny Place, Mater Misericordiae Hospitals, Raymond Terrace, South Brisbane, Q4101 Australia; and †Department of Surgery, Princess Alexandra Hospital, Ipswich Road, Woolloongabba, Brisbane, Q4101 Australia*

Received June 11, 1999

Epithelial mucins are a family of secreted and cell surface glycoproteins expressed by epithelial tissues and implicated in epithelial cell protection, adhesion modulation and signaling. The gene encoding human MUC3 (hMUC3), localised to chromosome 7q22, is most highly expressed in the small intestine. It has previously been reported to be a non-transmembrane mucin with minimal homology to its suggested orthologues from rat (rMuc3) and mouse (mMuc3). RT-PCR was performed to investigate the carboxyl terminus of the published sequence of hMUC3 from normal colon and small intestine tissues and also from a series of 10 colorectal cancer cell lines. Two distinct PCR products were identified. In contrast to the previously published hMUC3 sequence, which terminates shortly after a single cysteine-rich EGF-like domain, conceptual protein translation of the dominant and largest PCR product identified two extracellular cysteine-rich EGF-like domains separated by an N-glycosylation-rich domain and a potential coiled-coil region, followed by a putative transmembrane region and a 75 amino acid cytoplasmic tail. The smaller of the two PCR products was found to be an alternative splice variant of *MUC3* including the first EGF-like domain but lacking part of the second EGF-like domain and the transmembrane region. Nine out of 10 colorectal cancer cell lines were found to express *MUC3*. Interestingly, one of the cell lines, LoVo, expressed predominantly the alternative splice form lacking a transmembrane domain. Structural homology of the new protein sequence of hMUC3 with rMuc3 and mMuc3 indicates it is closely related to the rodent proteins and is likely to be involved in ligand-binding and intracellular signaling. The new finding that *MUC3* encodes a transmembrane molecule presents a new paradigm for the structure of this mucin and the manner in which it may function. © 1999 Academic Press

Epithelial mucins are large glycoproteins characterized by a central polymorphic tandem repeat structure with a large number of O-linked carbohydrate side chains. Ten human epithelial mucin genes have been identified, though due to their large size and tandem repeat nature, full-length cDNA clones are available for only 5 of these genes (1).

Many human mucin genes appear to encode secreted proteins which protect and lubricate epithelial tissues by forming a layer of viscoelastic gel. However, two human membrane-anchored epithelial mucins have been identified to date. The MUC1 mucin is expressed by almost all human glandular epithelial cells and has a distinct role in adhesion modulation and intracellular signaling (2). The second transmembrane mucin to be reported, MUC4, is also widely distributed in human epithelial tissues. MUC4 and its rat orthologue, ASGP1/2, contain two extracellular cysteine-rich EGF-like domains after a large mucin-like tandem repeat domain (3, 4). The cysteine-rich region is followed by a transmembrane domain and a cytoplasmic tail containing a potential tyrosine phosporylation site. The functional significance of EGF-like domains is unclear, however, they may allow exposure of ligand-binding sites on the exterior regions of a tri-lobed structure formed by disulfide bridging of the core 6 cysteine residues. Such motifs are found in several growth factors and in numerous extracellular proteins involved in formation of the extracellular matrix, cell adhesion, chemotaxis and wound healing (5). Association of rat ASGP1/2 with the c-erbB-2 growth factor receptor and consequent modulation of receptor function has been demonstrated (6), implicating the EGF-like domain-

[1] To whom correspondence and reprint requests should be addressed at Mater Medical Research Institute, Level 3, Aubigny Place, Mater Misericordiae Hospitals, Raymond Terrace, South Brisbane, Q4101 Australia. Fax: +61-7-38402550. E-mail: 6mmrap@mater.org.au.

0006-291X/99 $30.00

containing transmembrane mucins in growth modulation.

MUC3, located on human chromosome 7q22, encodes an intestinal mucin expressed by goblet and absorptive cells. It has previously been reported to contain a single cysteine-rich EGF-like domain at its carboxyl-terminus and to lack a transmembrane domain, suggesting it could be a secreted mucin (7). Two rodent genes, encoding rMuc3 and mMuc3, have been designated *Muc3* due to their location on the syntenic regions to human chromosome 7, their expression in intestine and their weak homology with *MUC3*. Although the EGF-like domain of hMUC3 is similar to one of the EGF-like domains of the rodent proteins, overall the hMUC3 carboxy terminus shows limited homology to the rodent proteins (34%), suggesting that hMUC3 may not be the orthologue of rat and mouse Muc3.

We have recently cloned a novel mucin cDNA named *MUC12* (Genbank accession number AF147790) which localises to chromosome band 7q22 (8). Conceptual protein translation of this cDNA reveals two EGF-like domains, a transmembrane region and a cytoplasmic tail with considerable homology to the rodent mucins. The 3′ sequence of *MUC12* is distinct from that of *MUC3* suggesting the presence of a cluster of mucin genes on human chromosome band 7q22. Our preliminary analysis of the published partial hMUC3 cDNA sequence revealed several possible open reading frames which included a potential second EGF-like motif, as well as a putative transmembrane region and cytoplasmic tail. To determine the true nature of the carboxyl terminus of hMUC3, we employed RT-PCR and sequencing of hMUC3 transcripts from colon, small intestine and 10 colorectal cancer cell lines. We have identified two discrete PCR products, the largest of which upon sequencing confirmed that *MUC3* encodes a transmembrane mucin, distinct from MUC12, and 260 amino acids longer than originally proposed. Thirty-eight percent overall identity of the conceptual protein translation of this new hMUC3 sequence with rat and mouse Muc3 suggests it is closely related to the rodent mucins.

MATERIALS AND METHODS

RNA extraction. Total RNA was isolated from cell lines and tissues listed below by the method of Chomczynski and Sacchi (9). Concentration and purity was determined by spectrophotometry at 260 and 280nm. The integrity of the RNA was assessed by denaturing agarose gel electrophoresis.

Touchdown RT-PCR. RT-PCR was performed on total RNA isolated from two normal colonic mucosa samples, one small intestine sample and 10 colorectal cancer cell lines, Caco-2, LIM1215, LIM1899, HCT116, SW116, LoVo, LS 174T, KM12SM, LISP-1 and SW620. First strand cDNA synthesis was accomplished using 1 μg of total RNA. Tissue and cell line RNA was initially screened for *MUC3* expression by performing PCR amplification of cDNA in a total volume of 25 μl containing 0.5 μl of the first strand cDNA synthesis reaction products, 2.5 μl 10x PCR buffer, 0.2 mM dNTPs, 1.5 mM $MgCl_2$, 20 pmol each of the forward and reverse primers, 5% DMSO and 1.25 U AmpliTaq gold (Perkin Elmer, Norwalk, CT). Based upon the published (7) sequence of the partial hMUC3 cDNA (Genbank accession no. AF007194), gene-specific forward and reverse primers (MUC3F1; 5′-TGTGACAATGGTGGCACCTGG-3′ and MUC3R1; 5′-GGGATTTGGGGGAACAGTCTC-3′) were designed to produce a predicted PCR product of 1126 bp.

PCR amplification conditions began with an initial denaturation step of 94°C for 10 min followed by a "touchdown" procedure (10). This involved 2 cycles each at annealing temperatures decreasing at 1° intervals from 68°C to 64°C (30 s), with denaturation at 94°C (45 s) and extension at 72°C (1 min). Then followed by 30 cycles of 94°C (45 s), annealing at 63°C (30 s) and extension at 72°C (1 min) were performed. PCR products were electrophoresed on 1% agarose, 1x TBE gels and photographed. For sequencing purposes, PCR amplification was also performed using high fidelity Expand polymerase (Boehringer Mannheim, Roche Diagnostics, Mannheim, Germany).

Cloning and sequencing. Amplified products from normal colonic and small intestinal mucosa, and from Caco-2 and LoVo cell lines were purified, cloned into pGEM-T (Promega Corporation, Madison, WI) and sequenced. cDNAs were sequenced in a cycle sequencing reaction with 2.5 pmol of primer and 4 μl of BigDye reaction mix (DNA Cycle Sequencing Kits, Perkin-Elmer) in a total volume of 10 μl. Cycling reactions were as follows: 25 cycles of denaturation at 96°C (30 s), primer annealing at 50°C (15 s) and extension at 60°C (4 min). Unincorporated nucleotides were removed by isopropanol precipitation. Products were analysed on a Model 377A automated DNA sequencer (PE Applied Biosystems). Sequences were analysed by multiple sequence similarity searches using BLAST algorithms (11) accessed through the National Centre of Biotechnology Information (NCBI;http://www.ncbi.nlm.nih.gov).

Computer analysis of protein sequences. Examination of protein sequences for potential structural domains was performed using ExPASy sequence analysis programs accessed through http:/www.expasy.ch/tools/.

RESULTS

Identification by RT-PCR of 1247 bp, 1131 bp and 1073 bp PCR products. In PCR reactions with RNA from normal colon, small intestine and 8 of 10 colorectal cancer cell lines (Caco-2, LIM1899, LIM1215, LS174T, HCT116, SW116, KM12SM, and LISP-1), an abundant PCR product of 1247 bp was observed (band identified by upper arrow in Fig. 1), in contrast to the predicted 1126 bp PCR product expected from the previously published hMUC3 sequence. The SW620 cell line did not express *MUC3*. Sequence analysis of this PCR product generated from normal small intestinal, and colonic and Caco-2 RNA, confirmed its sequence was amplified as a result of specific priming of the forward and reverse hMUC3 cDNA primers. The DNA sequence and conceptual protein translation of this product (Genbank accession no. AF143371) is shown in Fig. 2 and differs from that previously reported in several respects. Firstly, an additional previously unrecognized exon is found in the present hMUC3 sequence (from nucleotides 182–297 inclusive, Fig. 2). This sequence is almost identical to nucleotides 1950–

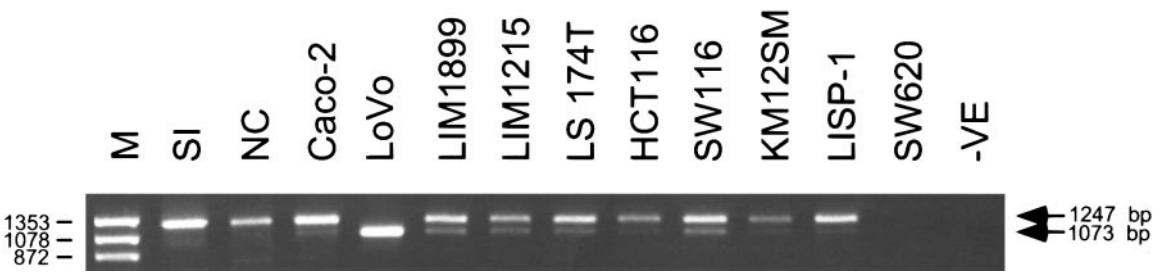

FIG. 1. RT-PCR showing amplification of hMUC3 mRNA transcripts from normal colonic mucosa (NC), small intestine (SI), and 10 colorectal cancer cell lines. M denotes molecular size markers.

2063 of Genbank accession no. AF007196, described by Gum *et al.* (5) as intronic sequence, except for two additional nucleotide insertions at positions 183 and 184 (Fig. 2). Secondly, there are 5 single nucleotide deletions in the original sequence (at positions 3026, 3254, 3269, 3479 and 3538 of Genbank accession no. AF007194) corresponding to nucleotides 589, 818, 834, 1045 and 1105 respectively of the new hMUC3 cDNA sequence shown in Fig. 2. Thirdly, the hMUC3 cDNA presented here contains 2 nucleotide substitutions (GC rather than CG) at positions 855 and 856 (Fig. 2).

Computer analysis revealed conceptual protein translation of this cDNA to contain two cysteine-rich EGF-like domains (amino acid positions 1–51 and 195–261 in Fig. 2) separated by an N-glycosylated domain and a coiled-coil region (amino acids 138–169 in Fig. 2), followed by a transmembrane domain (amino acids 269–291 in Fig. 2), containing 23 hydrophobic or uncharged amino acids, and a cytoplasmic tail of 75 amino acids at the carboxyl terminus (amino acids 292–366 in Fig. 2). A YVAL motif in the cytoplasmic tail (at amino acid positions 342–345 in Fig. 2) is similar to specific motifs recognized by SH-2 domain-containing proteins (12).

Interestingly, sequencing of one of eight clones derived from an extracted band containing the 1247 bp PCR product from small intestinal RNA, revealed an 1131 bp product (Genbank accession no. AF143373) almost identical to the previously described MUC3 sequence except for the 5 single nucleotide insertions reported above. The conceptual translation of this splice form (generated by deleting the sequence between the first two arrow heads in Fig. 2) is as previously described (7) and due to a shift in the open reading frame, results in a product with a different carboxyl terminus to the predominant isoform described above.

A smaller, less abundant PCR product of 1073 bp (Genbank accession no. AF143372) was also readily identifiable in 9 of 10 colorectal cancer cell lines (band identified by lower arrow in Fig. 1). The LoVo cell line was unusual in that it expressed this product almost exclusively. In repeated PCR reactions of the same cDNAs, this product was not always detected in PCR reactions from normal colonic cDNA and faint bands were only rarely seen in reactions with small intestinal cDNA. Sequence analysis of this PCR product generated from LoVo and normal colonic RNA confirmed it was as a result of specific priming of the forward and reverse hMUC3 cDNA primers. A section of sequence, from nucleotides 740–914, was found to be spliced out of this transcript resulting in removal (between the third and fourth arrow heads in Fig. 2) of the latter part of the second EGF-like domain and the entire transmembrane region from the conceptual translation product.

Amino acid sequence alignment of the C-terminal human hMUC3, rMuc3 and mMuc3 is shown in Fig. 3. Conservation of the cysteine spacing between the hMUC3 and rodent Muc3 proteins is observed in both EGF-like domains. The hMUC3 C-terminal amino acid sequence was found to have 38% overall identity with rat and mouse Muc3.

DISCUSSION

Previously, hMUC3 has been described as containing a single EGF-like domain and lacking a transmembrane domain. Using RT-PCR, we have established that the carboxy terminus of hMUC3 contains two EGF domains, a hydrophobic region consistent with a transmembrane domain, and a long cytoplasmic tail. The discrepancies between the C-terminal amino acid sequence we have described and that described previously (7) are due to an additional exon and five single base pair insertions. The previously described sequence was assembled from several clones with little overlapping sequence (a genomic clone named GM3-1; a cDNA clone named clone20 and a 3′ RACE PCR product; both of the latter being generated from small intestine). Much of the previously reported 3′ sequence of hMUC3 relies on the accuracy of a single clone. Sequence analysis has revealed we have derived an identical 1247 bp PCR product from three different sources of RNA using high stringency RT-PCR. The amplification of the alternative splice form of *MUC3*

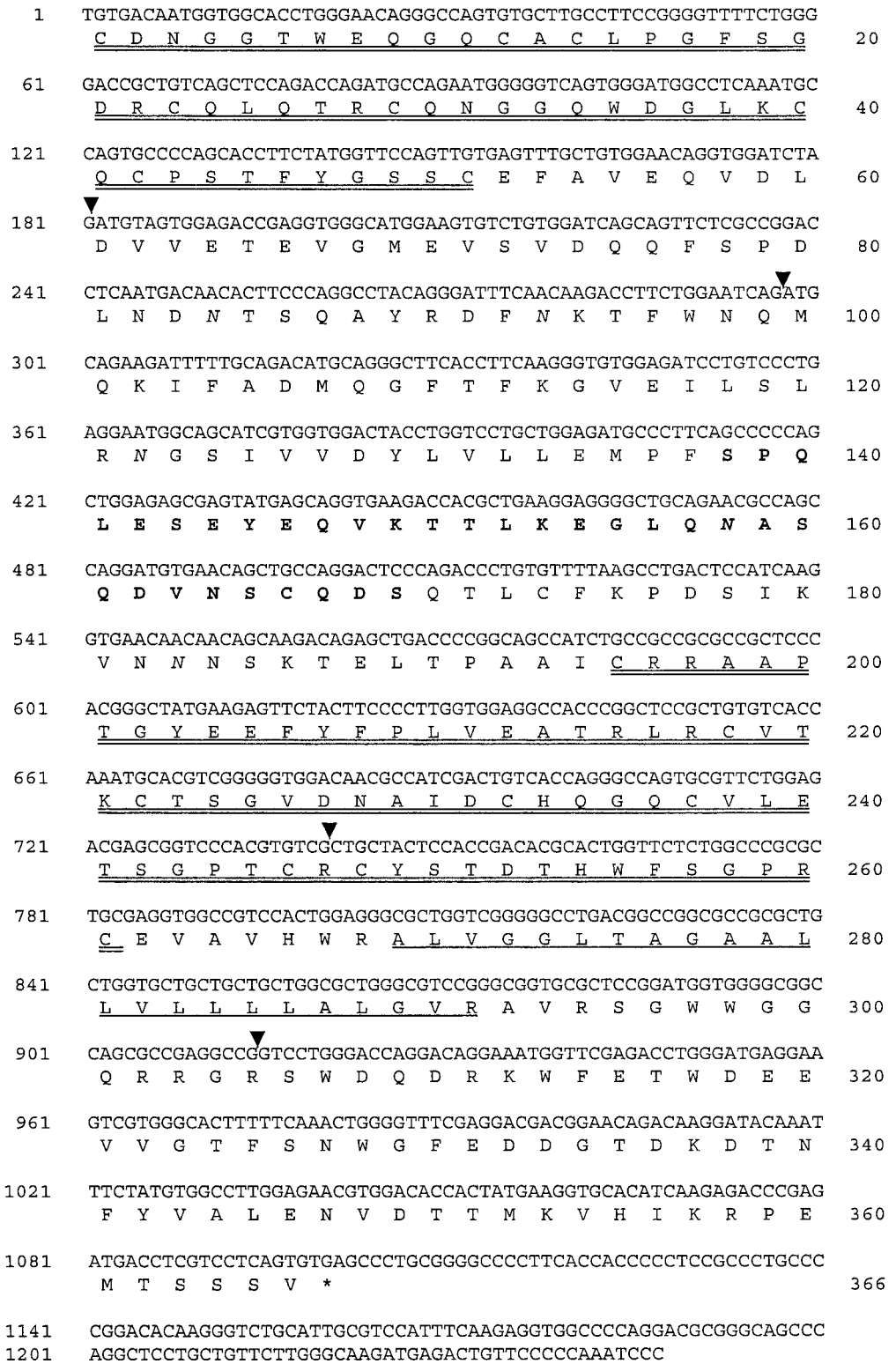

FIG. 2. DNA sequence and predicted amino acid sequence of hMUC3. Numbering of nucleotides is given on the left and amino acids on the right. Five potential N-glycosylation sites are shown in italic, the transmembrane region is singly underlined and two cysteine-rich EGF-like domains are double underlined. A potential coiled-coil region is in bold. The alternatively spliced regions are marked by arrows. The stop codon is denoted by an asterisk.

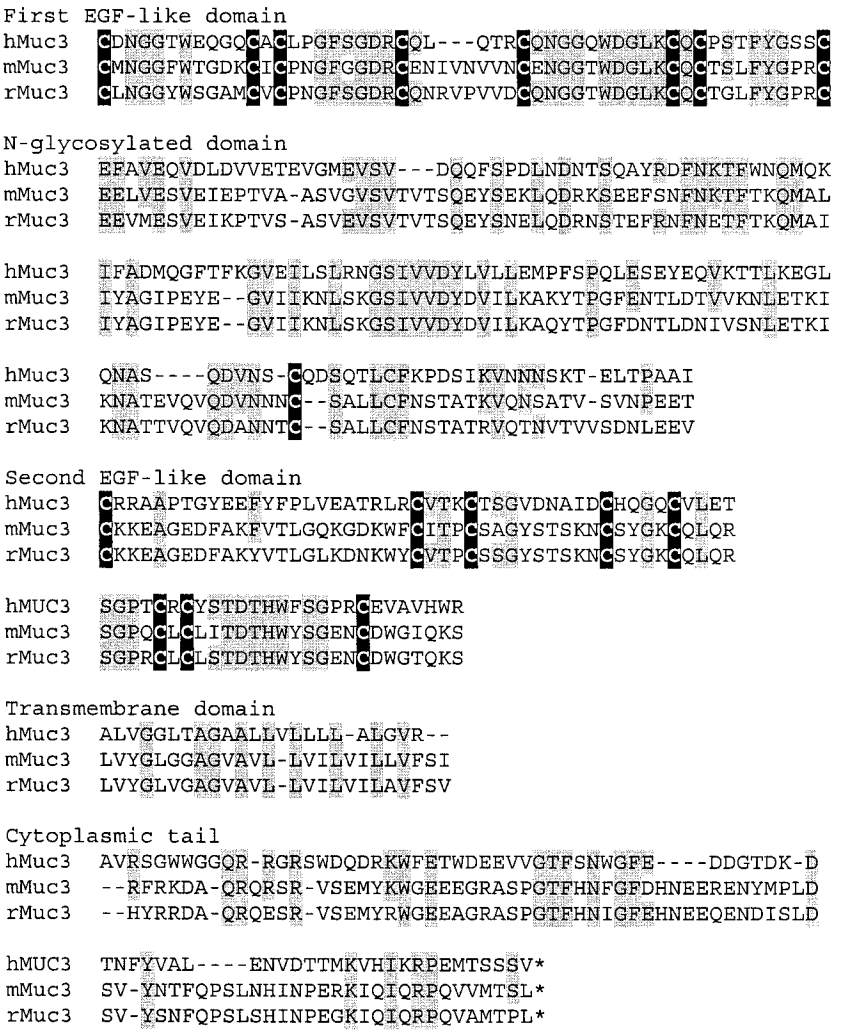

FIG. 3. Amino acid sequence alignment of the carboxyl termini of hMUC3 (amino acids 1–366) (Fig. 2), rMuc3 (16, 17) (amino acids 356–447 and 1–379, respectively), mMuc3 (18) (amino acids 637–1015) grouped according to functional domains. Shading demonstrates identity of hMUC3 with the rodent proteins. Hyphens indicate gaps inserted to optimize the alignment.

lacking a transmembrane domain (1073 bp), was not consistently expressed in the colonic tissue-derived RNA sample, and not seen in PCR reactions generated from small intestinal RNA, suggesting the newly described *MUC3* transcript encoding a transmembrane mucin represents the major form of hMUC3 produced in the intestine. An 1131 bp PCR product representing the previously described *MUC3* carboxyl terminus (7) was isolated from small intestinal RNA. This product was not readily visible under the present stringent PCR conditions and suggests this transcript is relatively rare.

MUC3 thus appears to be a transmembrane mucin and as such is only the third human membrane-anchored epithelial mucin to be described to date, along with MUC1 and MUC4. MUC1 has been shown to be involved in cell signaling via multiple tyrosine phosphorylation sites on its highly conserved cytoplasmic tail (13). At its carboxyl terminus, MUC3 possesses a cytoplasmic tail containing a YVAL sequence which is similar to motifs recognized by SH-2 domain-containing proteins (12), suggesting that MUC3, like MUC1, could be involved in signal transduction. Furthermore an extracellular potential coiled-coil domain and two EGF-like domains suggest roles in protein-protein interactions and ligand binding. We have identified two alternative splice forms of *MUC3,* their conceptual protein translations suggesting that they both could be secreted as they both lack a transmembrane domain. The secreted isoforms of MUC3 may function as protective mucins, perhaps as a co-constituent with gel-forming mucins in mucus, or may act at the apical cell surface as a ligand for other cell surface molecules such as EGF-like growth factors. Interestingly, the

MUC1 gene also demonstrates differential splicing and similarly has secreted and transmembrane isoforms (14, 15). MUC1/REP encodes a tandem repeat domain, followed by a transmembrane region and a 72 amino acid cyoplasmic tail, whereas the MUC1/SEC variant lacks a transmembrane region and is secreted from the cell. Another isoform, MUC1/Y, contains the transmembrane region but lacks the tandem repeat domain and is expressed by various human epithelial tumors, whilst not being detectable in adjacent normal tissue.

The carboxyl terminus of hMUC3 shows areas of high homology to the equivalent regions of both rat and mouse Muc3 proteins (16, 17, 18). Extensive conservation of the two EGF-like domains (61% and 42% respectively) suggest they have been functionally important throughout evolution. Given the very high degree of conservation between the rodent proteins (81% amino acid identity), it is clear that rMuc3 and mMuc3 are orthologues. In contrast, the present hMUC3 conceptual protein has only 38% identity overall with the rodent proteins. Despite conservation of parts of the carboxyl terminus of hMUC3 with rMuc3 and mMuc3, it is likely that hMUC3 represents a closely related protein family member rather than the human orthologue of the rodent proteins.

Following this reappraisal of the hMUC3 gene, our cloning of MUC12 and the recent cloning of MUC4, it appears that there is a distinct subfamily of epithelial mucins with a conserved C-terminal domain structure. Data concerning functions of this mucin subfamily are restricted to studies of the rat MUC4 orthologue (ASGP1/2), or the rat sialomucin complex. Interestingly, the ASGP2 isoform which contains EGF-like domains but lacks a mucin domain, has been shown in transfection studies to bind and activate the c-erbB-2 growth factor receptor, both in the presence and absence of the c-erbB-2 ligand, resulting in increased mitogenesis (6). In fact, the EGF-like domain in ASGP2 shows homology with the c-erbB-2 ligands (heregulins), and the first EGF-like domain in hMUC3 shows conservation of cysteine residues and limited homology to a number of EGF receptor-binding growth factors such as TGFα, amphiregulin and betacellulin. In contrast to ASGP2 which has only a short cytoplasmic tail, MUC3 and MUC12 have large cytoplasmic tails consistent with a role in signal transduction.

Mucins have been implicated in the pathology of carcinomas as well as several non-malignant diseases such as cystic fibrosis and inflammatory bowel disease. Specifically, *MUC3* has been reported to be downregulated in colorectal cancers (19, 20); although we show here that *MUC3* is expressed in 9/10 colorectal cancer cell lines. A recent report has also described evidence for linkage between inflammatory bowel disease and markers on chromosome 7q22 (21). *MUC3* has therefore been proposed as a candidate susceptibility gene for this disease and preliminary evidence suggests a possible relationship between rare *MUC3* alleles and inflammatory bowel disease (22).

The present study demonstrates that the *MUC3* gene encodes a human transmembrane mucin, only the third to be described to date. We have also identified two minor alternative splice forms of *MUC3*, their conceptual protein translations suggesting that they are secreted. The finding that *MUC3* encodes a transmembrane molecule presents a new paradigm for the structure of this mucin and the manner in which it may function.

ACKNOWLEDGMENTS

This research was supported by a grant from the NH&MRC, Grant 9936503. The authors thank Dr. Nick Hayward for helpful discussions.

REFERENCES

1. Seregni, E., Botti, C., Massaron, S., Lombardo, C., Capobianco, A., Bogni, A., and Bombardieri E. (1997) *Tumori* **83,** 625–632.
2. Gendler, S. J., Spicer, A. P., Lalani, E. N., Duhig, T., Peat N., Burchell, J., Pemberton, L., Boshell, M., and Taylor-Papadimitriou, J. (1991) *Am. Rev. Respir. Dis.* **144,** S42–S47.
3. Sheng, S., Wu, K., Carraway, K. L., and Fregien, N. (1992) *J. Biol. Chem.* **267,** 16341–16346.
4. Moniaux, N., Nollet, S., Porchet, N., Degand, P., Laine, A., and Aubert, J. P. (1999) *Biochem. J.* **338,** 325–333.
5. Campbell, I. D., and Bork, P. (1993) *Curr. Opin. Struct. Biol.* **3,** 385–392.
6. Mcneer, R. R., Price-Schiavi, S. A., Komatsu, M., Fregien, N., Carraway, C. A. C. and Carraway, K. L. (1998) *Front. Bioscience* **2,** 449–459.
7. Gum, J.R., Ho, J. J. L., Pratt, W. S., Hicks, J. W., Hill, A. S., Vinall, L. E., Roberton, A. M., Swallow, D. M., and Kim, Y. S. (1997) *J. Biol. Chem.* **272,** 26678–26686.
8. Williams, S. J., McGuckin, M. A., Gotley, D. J., Eyre, H., Sutherland, G., and Antalis, T. M. (1998) *Cancer Res.*, in press.
9. Chomczynski, P., and Sacchi, N. (1987) *Anal Biochem.* **162,** 156–159.
10. Don, R. H., Cox, P. T., Wainwright, B. J., Baker, K., and Mattick, J. S. (1991) *N.A.R.* **19,** 4008.
11. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) *J. Mol. Biol.* **215,** 403–410.
12. Songyang, Z., Shoelson, S. E., McGlade, J., Olivier, P., Pawson, T., Bustello, X. R., Barbacid, M., Sabe, H., Hanafusa, H., Yi, T., Ren, R., Baltimore, D., Ratnofsky, S., Feldman, R. A. and Cantley, L. C. (1994) *Mol. Cell Biol.* **14,** 2777–2785.
13. Zrihan-Licht, S., Baruch, A., Elroy-Stein, O., Keydar, I., and Wreschner, D. H. (1994) *FEBS Lett.* **356,** 130–136.
14. Wreschner, D. H., Hareuvini, M., Tsarfaty, I., Smorodinsky, N., Horev, J., Zaretsky, J., Kotkes, P., Weiss, M., Lathe, R., Dion, A. S., and Keydar, I. (1990) *Eur. J. Biochem.* **189,** 463–473.
15. Zrihan-Licht, S., Vos, H. L., Baruch, A., Elroy-Stein, O., Sagiv, D., Keydar, I., Hilkens, J., and Wreschner, D. H. (1994) *Eur. J. Cancer* **224,** 787–795.
16. Gum, J. R., Hicks, J. W., Lagace, R. L., Byrd, J. C., Toribara,

N. W., Siddiki, B., Fearney, F. J., Lamport, D. T., and Kim, Y. S. (1991) *J. Biol. Chem.* **266,** 22733–22738.

17. Khatri, I. A., Forstner, G. G., and Forstner, J. F. (1997) *Biochim. Biophys. Acta* **1326,** 7–11.
18. Shekels, L. L., Hunninghake, D. A., Tisdale, A. S., Gipson, I. K., Kieliszewski, M., Kozak, C. A, and Ho, S. B. (1998) *Biochem. J.* **330,** 1301–1308.
19. Ogata, S., Uehara, H., Chen, A., and Itzkowitz, S. H. (1992) *Cancer Res.* **52,** 5971–5978.
20. Chang, S., Dohrman, A. F., Basbaum, C. B., Ho, S. B., Tsuda, T., Toribara, N. W., Gum, J. R., and Kim, Y. S. (1994) *Gastroenterology* **107,** 28–26.
21. Satsangi, J., Parkes, M., Louis, E., Hashimoto, L., Kato, N., Welsh, K., Terwilliger, J. D., Lathrop, G. M., Bell, J. I., and Jewell, D. P. (1996) *Nature Genet.* **14,** 199–202,
22. Kyo, K., Parkes, M., Takei, Y., Nishimori, H., Vyas, P., Satsangi, J., Simmons, J., Nagawa, H., Baba, S., Jewell, D., Muto, T., Lathrop, M., and Nakamura Y. (1999) *Hum. Mol. Gen.* **8,** 307–311.