Glycoconjugate Journal (1994) **11**:172–179

GLYCOPINIONS MINI-REVIEW

Mucin genes: structure, expression and regulation

MUKESH VERMA and EUGENE A. DAVIDSON*

Department of Biochemistry, Georgetown University Medical Center, 3900 Reservoir Road, NW, Washington, DC 20007, USA

Received 13 May 1994, revised 9 June 1994

Keywords: cystic fibrosis; epithelial cells; gene regulation; glycoprotein; immortalization; mucin; papilloma virus; vectors; gene regulation

1. Introduction: mucin gene family

Mucus is defined as a viscous secretion of one or more glycoproteins (mucins) secreted by epithelial cells of normal and cancerous tissues. Mucins isolated from the gastrointestinal, respiratory, or urinary tracts exhibit different properties (for example, amino acid composition) [1]. Mucin has been isolated and characterized from the submaxillary gland of bovine and porcine species, skin of frog, mammary gland of mouse and human, intestine of rat and human, trachea of dog, human, rat and saliva of human and from other sources [2, 3].

Mucins are defined by their characteristic *O*-glycosylated domains. Typically, the domains contain a (semi)repetitive protein backbone with a particularly high content of Thr and Ser residues interspersed by Pro residues. This reflects the importance of local secondary structure for *O*-glycosylation by the polypeptide GalNAc transferase. The clustered O-linked oligosaccharides result in a stiff, extended conformation of the glycoprotein. Due to the presence of charge (sulfation of carbohydrate moieties and sialic acid residues), mucin can be considered as a highly charged thread-like polymer. As a consequence, during exocytosis the mucin matrix undergoes an ion-triggered transition from a condensed to an expanded hydrated phase, caused by the mutual repulsion of the anionic charges. This massive swelling is the biophysical basis for many general properties typical of mucin, particularly their high viscosity. These molecules have a strong tendency to bind to various proteins and peptides and their purification is difficult. On the basis of their sensitivity to proteases, mucins contain two different types of domains: highly glycosylated regions (rich in Ser and Thr) and naked hydrophobic regions that are exclusively devoid of carbohydrates. Due to the high carbohydrate content of mucins, the elucidation of the primary structure of their peptide moieties, or apomucin, has been extremely difficult by conventional protein sequencing and only a few stretches of mucin peptide have been sequenced (reviewed in [2]). These difficulties encountered using conventional procedures have prompted several laboratories to clone apomucin cDNAs from different sources. To date a number of mucin genes from different species have been isolated and a nomenclature has been assigned to these genes.

According to the recent nomenclature, human mucin genes are referred to as *MUC1* (mammary gland), *MUC2* (intestinal, basic), *MUC3* (intestinal, neutral), *MUC4* (tracheo bronchial), *MUC5* (tracheo-bronchial), *MUC6* (stomach), and *MUC7* (salivary). Mucin genes from other species are: *RIM* (rat intestinal); *M-MUC1* (mouse homologue of *MUC1*); *FIM* (frog integumentary); *CTM* (canine tracheo-bronchial); *BSM* (bovine submaxillary gland); and *PSM* (porcine submaxillary mucin). In this review, the above names will be used throughout.

Mucins are clinically significant molecules. Blood group antigens were identified on mucins from secretory positive individuals, and tissue specific patterns of sialylation and sulfation were detected [4, 5]. The branching of oligosaccharides was also found to have tissue specificity, giving a characteristic mucin oligosaccharide pattern for each animal mucosa [6].

A. General features – site of synthesis, protective function, clinical significance, mucin as a ligand for Pseudomonas *and* Staphylococcus

The common features of mucin synthesized in different tissues are their high sugar content, the presence of five amino acids (Ala, Gly, Pro, Thr and Ser) in predominant amounts, amino acid repeats, very few aromatic amino acids, polydisperse message, and a cysteine-rich domain. Since

* To whom correspondence should be addressed.

0282-0080 © 1994 Chapman & Hall

they provide protection to epithelia, the composition of mucin is tightly regulated. The selected product should not be too thin so that it cannot form a protective layer while at the same time, it should not be too thick to hinder the movement of cilia (where relevant) of the epithelium. How this regulation occurs in the cell is not completely understood. In diseases such as chronic bronchitis, asthma, cystic fibrosis (CF), and in some cancers, mucin is apparently over-expressed and abnormally glycosylated. Mucin molecules may provide the site of attachment for microorganisms such as *Pseudomonas* and *Staphylococcus*. According to current theory, sialic acid residues serve as a ligand in this process.

B. Heterogeneity – at the mRNA and protein level

The mucin isolated from gastrointestinal tract, respiratory tract, and urinary tract exhibits different amino acid composition. When the primary structure of mucin was determined by cloning of the cDNAs encoding these genes, heterogeneity was observed at the nucleotide level as well. The size of the message for mucin ranges from 0.5 to 18 kb which also suggests the existence of heterogeneity at the RNA level. In some instances, dense bands of the mRNA could be observed, for example, a 2.5 kb band in the case of MUC7 [see 1, 7]. Polymorphism for MUC1, MUC2, and FIM has been reported [1, 8, 9].

C. Tissue specificity – lung mucin gene is different than other mucins

Mammalian mucosal epithelia are generally coated with mucus. The specific properties of this material vary among the different tissue sources (e.g. salivary glands, respiratory tract, gastrointestinal tract, and cervix) but common features are observed. These reflect both physical aspects such as lubrication or protection, and chemical composition. In diseases like cystic fibrosis, chronic bronchitis, asthma, cancer, and during viral infection, mucin is secreted in higher amounts and its glycosylation is aberrant [10–12]. It has been reported that the mucin gene is up-regulated in CF where *Staphylococcus aureus* and *Pseudomonas aeruginosa* infections are common [3, 14]. The respiratory mucin has an additional function of moving particulate materials in an upward direction by ciliary movement. Since the entire system is responsive to a variety of stimuli, close regulation of synthesis and secretion is clearly necessary. To date, the regulation of biosynthesis of the tracheo bronchial mucin (TBM) has not been studied. In general, the tracheal mucin is rich in Thr and Pro whereas salivary gland mucin is rich in thr and ser.

D. Alternative splicing

It is not yet completely understood why mucin gene mRNA appears polydisperse on Northern blots. In the case of the pancreatic mucin, it was demonstrated that different forms of mucin were synthesized due to alternative splicing [15]. Since information regarding the genomic organization (the size and the number of the introns and exons) of mucin is generally lacking, such experiments have not been performed for most types. However, the information available about the sequences of the cDNAs encoding different mucin genes will be useful in determining whether polydispersity is due to alternative splicing. It is quite possible that the epithelial cells, which are the major source of mucin production, contain cell-specific splicing factors. To establish a protocol for *in vitro* splicing, we have started characterizing the components of the splicing machinery of the canine tracheal epithelial cells [16].

E. Repeated sequences

In spite of the presence of tandem amino acid repeats in most of the mucins (exceptions are the BSM and CTM), there is no homology in the repeat sequences (see Table 1). In most of the examples, Ser and Thr are the essential component of the repeats, but amino acids such as Pro and His are also present. Being rich in Thr and Ser, these repeats provide potential sites for *O*-glycosylation. The number of amino acids in the tandem repeats of human mucins are smaller than in PSM. The tandem repeats of the human acidic mucin (MUC2) and neutral (MUC3) intestinal mucins are markedly different from other mammalian mucins and from each other. The longest repeat (169 amino acids) is present in MUC6. Probably these repeats provide protection for the molecule.

F. Cysteine rich domain

In the canine tracheal mucin sequence, 29 Cys residues are present; 16 of those are in the carboxyl terminal region

Table 1. Repeats in mucins.

Mucin	*Repeating sequence*
MUC1	PDTRPAPGSTAPPAHGVTSA
MUC2	PTTTPITTTTTVTPTPTPTGTQT
MUC3	HSTPSFTSSITTTETTS
MUC4	TSSVSTGHATSLPVTD
MUC5	TTSTTSAP
MUC6	SPFSSTGPMTATSFQTTTTYPTPSHPQTTLPTH VPPFSTSLVTPSTGTVIPTHAOMATSASIHS TPTGTIPPPTTLKATGSTHTAPPMTPTTSGTS QAHSSFSTAKTSTSLHSHTSSTHHPEVTPTSTT TITPNPTSTGTSTPVAHTTSATSSELPTPFTHH SPPTGS
MUC7	CRPKLPPSPNKPPKFPNPHQPPK
BSM	No repeats
CTM	No repeats
FIM	VPTTPETTT
PSM	GAGPGTTASSVGVTETARPSVAGSGTTGTVS GASGSTGSSSGSPGATGASIGQPETSRISVAGSS GAPAVSSGASQAAGTS
RIM	TTTPDV

of the protein (starting from nt 2649 to nt 3382). A Cys-rich domain in the COOH-terminal region was first reported in BSM [17], and subsequently in PSM [18], RIM [19], FIM [20] and MUC2 [1]. The pattern of the distribution of Cys in all the above cases was different. For example, in BSM, all 30 Cys residues are present in the carboxyl terminal region with none present in the first 329 amino acid residues (out of 563 amino acids). In RIM, the distribution of Cys was throughout the molecule and comprised 10% of the total amino acids. It will be interesting to know whether the Cys rich domain has some functional significance. Eckhardt *et al.* [18] suggested that this domain has secondary structures that permit the formation of a globular structure. A globular structure would also be likely if the half-cysteine residues form many intramolecular disulfide bonds.

Table 2. Human mucin genes and their chromosomal location.

Mucin gene	*Chromsomal location*	*Tissue Type*
MUC1	1q21	Mammary gland
MUC2	11p15.5	Intestinal
MUC3	7	Intestinal
MUC4	3	Tracheal
MUC5	11p15	Tracheal
MUC6	11p15.4–11p15.5	Gastric
MUC7	4	Salivary
CF	11p13–11pTer	Tracheal
CTM	11p15	Tracheal

2. Gene structure (at the cDNA and genomic levels)

To understand the primary structure of different mucins it is desirable to delineate the nucleotide sequence of the genes encoding them. Using antibodies (against deglycosylated mucin) and synthetic oligonucleotides (based on the partial peptide sequence, available in some of the cases) as probes, cDNA libraries were screened and mucin genes identified. Thus far, seven human mucin genes *MUC1* to *MUC7*) have been identified. The cDNA for mucin from other species has also been characterized, for example, BSM, CTM, FIM, PSM. The genomic structure is available for *MUC1* and partly for *MUC2* and *CTM*. The chromosomal location of all human mucin genes has been assigned [1, 2] and it appears that chromosome 11 is the most common locus for mucin genes. In Table 2 is shown the chromosomal location for all mucin genes known to date (the assignment of the *CTM* gene is done by hybridization of human somatic hybrids with a CTM cDNA probe, Verma, Banerji, Murthy, and Davidson, unpublished).

A. Full length vs partial cDNA sequences

The full length sequences for cDNAs of BSM [17], CTM [21], MUC1 [8] and MUC7 [7] are available. Since the information about other mucin genes is based on partial cDNA sequences is it not known how long the full coding sequences are. Moreover, the chromosomal location of all mucin genes (other than human mucin genes) is also not available to determine whether partial cDNA sequences are part of a single gene. MUC2 cDNA has been characterized. It contains a single peptide characteristic for the secretory protein [1]. The transcription initiation site for *MUC2* has also been determined. Gum *et al.* [1] have suggested that the full coding sequence is over 15000 bases in the most common allelic form encoding a protein containing more than 5100 residues. *MUC7* is the smallest human mucin gene which codes for 377 amino acids [7]. The first 20 *N*-terminal residues are very hydrophobic and probably comprise the leader peptide. The region coding for the secreted protein can be divided into three distinct domains: unique 5′- and 3′-translated regions containing 4 and 1 potential *N*-glycosylation sites, respectively, and a central region of six tandem repeats of 23 amino acid residues. The mRNA is about 2.5 kb long and its expression appears to be species-, tissue-, and cell-specific.

The nucleotide (nt) sequence of the CTM showed some interesting features. The complete sequence determined for 3.7 kb of cDNA contained an open reading frame with coding capacity for 1118 amino acids. The length of the 3′UTR (untranslated region) was 388 bases. Two translation initiation sites (ATG sequence with Kozak consensus sequence) were located at 31 and 182 nt. The translation termination codon was located at nt 3385. One poly-adenylation signal (AAATAAAAA at nt 3742) and a polyA stretch were also observed. The size of the mucin gene is 40 kb and Southern hybridization experiments indicate that in the canine genome a single *TBM* gene exists. When primers were chosen from the cDNA sequences which do not show homology in the Thr and Pro rich motifs, and used for PCR amplification, single bands of the expected sizes were obtained. It suggests that the cDNA which was isolated and sequenced is not a product of alternative splicing. Larger size fragment(s) would have been obtained if this was due to alternative splicing. Primer extension analysis suggests multiple sites of initiation of transcription. The canine mucin cDNA derived amino acid sequence showed the characteristics of a typical mucin: it is rich in Thr, Ser, Pro, Ala, and Gly and poor in aromatic amino acids. The distribution of the above amino acids is almost uniform throughout. Although tandem repeats of amino acids were not observed in the canine mucin sequence, motifs TPTPTPTG or TTTTTTM/V were found. No similarity was observed among repeats from the many examples of mucins which suggests the heterogeneity of the protein in different systems. BSM also does not contain any repeats [17]. The existence of repeats is a characteristic feature of several mucins, although their functional significance is not yet understood.

B. Homology among all known mucin gene sequences (at the nucleotide and amino acid levels)

Striking homology is seen only in the cys-rich domains of *BSM*, *PSM* and *FIM* [18, 20]. In all other regions, homology (partial) is seen only in the apparent glycosylation regions. Mucin genes, *MUC2* and *FIM*, exhibit some homology (in the D domain) to the prepro-von Willebrand factor. In *MUC2* four such domains have been reported [1]. The prepro-von Willebrand factor is a 741 residue protein that has been implicated in the disulfide linkage oligomerization of the factor into secondary vacuoles. The physiological significance of these homologous sequences is not known. Gambus *et al.* [22] demonstrated detection of the *MUC2* apomucin tandem repeat with a mouse monoclonal antibody (mab) (LDQ10). This mab also shows strong reactivity with colorectal and stomach cancer and weaker reactivity with pancreas, breast and bladder cancers. They suggested that LDQ10 detects a peptide epitope of *MUC2* that becomes cryptic on glycosylation. Altered synthesis of the *MUC2* apomucin takes place in a variety of carcinomas [22]. Antisera against human intestinal mucin react with the mucin of the upper airways [23]. Pemberton *et al.* [24] have characterized antibodies to the cytoplasmic domain of the *MUC1* gene which show conservation through mammals. We have also seen CTM-like sequences in a variety of mammals [25]. This may be due to the presence of motifs TPTPTP and TTTTT in the CTM. Perini *et al.* [23] showed antigenic similarities between bronchial and intestinal mucins.

C. Potential N-*glycosylation sites*

The nucleotide sequences of mucins show the presence of potential *N*-glycosylation sites. In some cases, the *N*-glycosylation sites have been identified chemically (such as in RIM, see [19]). Three potential *N*-glycosylation sites were located at amino acid positions 565, 939, and 1077 in CTM. In other mucins, a varying number of *N*-glycosylation sites has been reported, e.g. MUC2 (two sites; [26–28]), MUC4 (two sites; [15]), MUC1 (one site; [8, 29, 30]), PSM (seven sites; [18]), RIM (13 sites; [19, 28]), M-MUC1 (one site; [31]). Since *N*-glycosylation occurs co-translationally, it is possible that the *N*-glycans stabilize the nascent apomucin prior to *O*-glycosylation or they play a role in the intracellular targeting of this macromolecule. On the other hand, the *N*-glycans may be important in the functioning of the mature mucin or may have no role (or may not be present). Further studies are needed to define the role of *N*-linked carbohydrate chains in mucin.

D. Secondary structure of mucins

Based on computer analysis, the secondary structure of some mucin proteins has been predicted. The predicted secondary structure of CTM is presented in Fig. 1. It does not have any transmembrane region, but, helices and pleated sheets can be seen. The Pro-rich regions probably provide folding to the molecule. Eckhardt *et al.* [18] analysed the secondary structure of the PSM. The polypeptide from residues 502–920 appeared to contain primarily turns with only a few stretches of sequence that can form helices or pleated sheets. In contrast, the carboxyl-terminal 240 residues in the polypeptide appear to contain both helices and pleated sheets suggesting that this domain has a more globular structure than the remainder of the mucin [18].

E. Structure of the promoter region and regulatory elements

The *MUC1* promoter has been characterized extensively and some information is available about the *MUC2* and the *CTM* promoter regions [1, 32]; Verma and Davidson, unpublished observations). In *MUC1*, 803 bp of the 5′ end upstream region are known. *MUC1* contains 7 exons which vary in size from 4 kb to 7 kb depending on the number of repeats in exon 2. Expression of *MUC1* was obtained from a genomic clone in an Epstein-Barr virus based vector, after transfection into a human epithelial cell line, indicating the presence of effective regulatory sequences in this clone [32]. Some recent experiments done with canine tracheal immortalized cells (CT1) suggest that the mucin gene expression is regulated by signal transduction; cytosolic cAMP levels can affect mucin gene expression (this induction requires *de novo* RNA and protein synthesis) (Verma, Sanadi, and Davidson, unpublished). The nucleotide sequence of the *CTM* gene promoter contains a cAMP response element [33–35]. In the future, detailed experiments can be designed to investigate the possible mechanisms (along with the signal transduction pathway) involved in the over-expression of mucin genes in lung disease.

The nucleotide sequence in the potential promoter regions of the CTM indicated the presence of the regulatory elements recognized by some of the well characterized transcription factors. The functionality of the promoter was tested by transient expression of the *CAT* gene and clones were found to contain the promoter for the *CTM* gene; fine mapping of the promoter region is yet to be completed. To determine whether these clones contain the binding sites for regulatory proteins, these clones were further digested into smaller fragments and tested for binding to the nuclear proteins of canine tracheal cells (CT1). EMSA (electrophoretic mobility shift assay) and Southwestern blot analysis were performed [36]; results suggest that the regulatory sequences of the *CTM* gene contain binding sites for some nuclear proteins (Verma and Davidson, unpublished). It is not yet known whether these proteins are functionally important. The specificity of binding can be tested by oligonucleotide competition assay, *in vitro* mutagenesis and Southwestern analysis. In our laboratory, immortalized cells from a cystic fibrosis patient (CFT1) and a normal human subject (HBE1E6E7) are available. Whether CF-specific nuclear proteins exist which regulate mucin gene expression can be tested by using the same DNA fragments (which show binding to the tracheal nuclear cells)

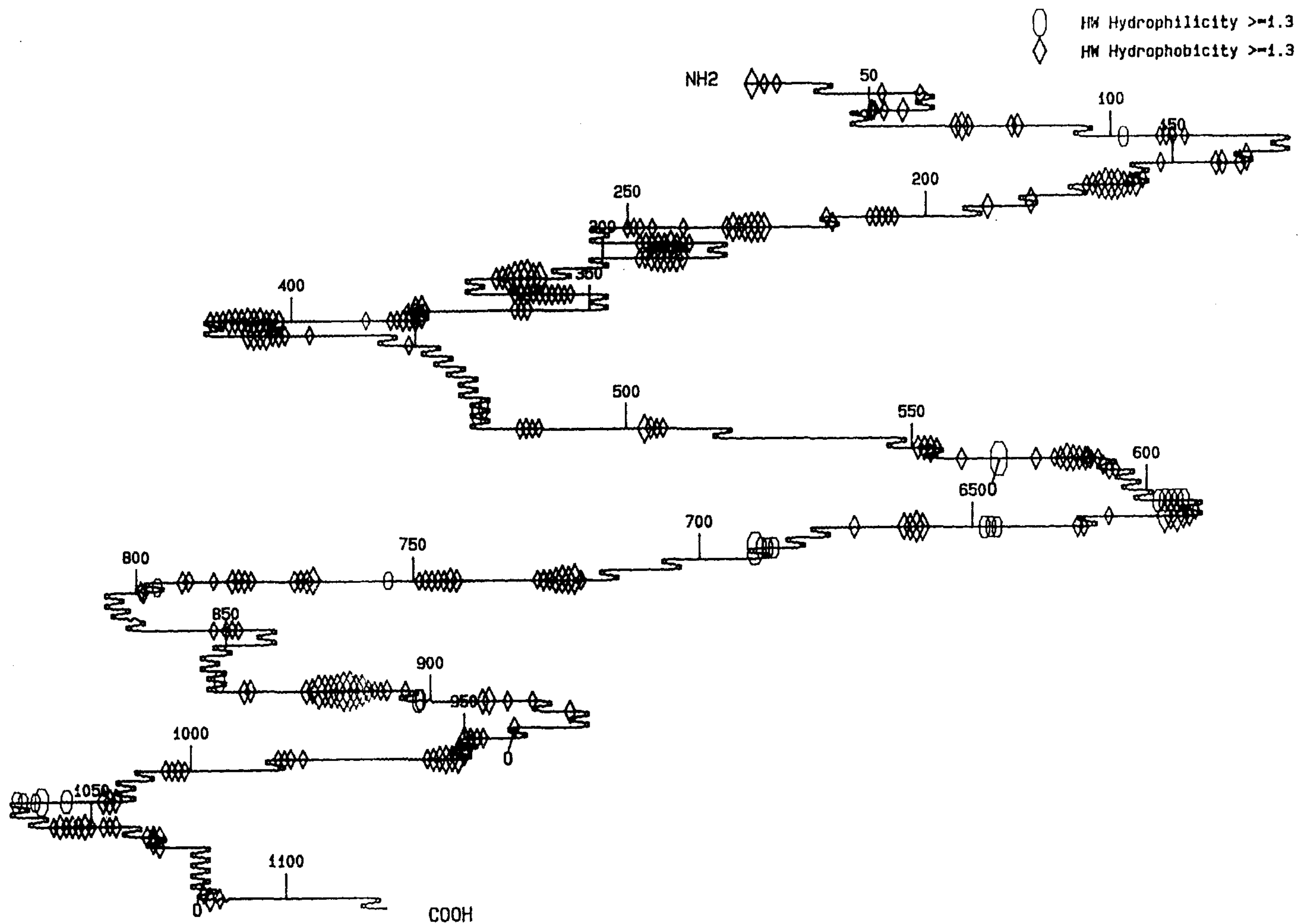

Figure 1. Secondary structure of CTM. The structure is based on Chou-Fasman model (generated by Genetic Computer Group software).

and using nuclear extracts from non-CF cells (HBE1E6E7). A difference in pattern in the EMSA analysis of HBE1E6E7 cell nuclear extract with those from CFT1 cells will suggest the presence of CF-specific nuclear factor(s). DNA fragments showing binding for proteins (factors) can be characterized by deletion mapping. Finally, the proteins binding to these sequences can be functionally characterized. Such studies have not been performed in other mucins.

3. Biosynthesis of mucin

The steps involved in the biosynthesis of rat gastric mucin have been studied in detail. In rat gastric mucin it was demonstrated that oligomerization is necessary for the functioning of the proteins, and *N*-glycosylation precedes the oligomerization [37]. Mucin probably is *N*-glycosylation co-translationally in the rough endoplasmic reticulum (RER). The peptide is released in the lumen of the RER where oligomerization starts. Tunicamycin inhibits the *N*-glycosylation step but does not affect the oligomerization of the mucin peptide. It was observed that: (i) disulfide bond formation plays a significant role in the oligomerization; and (ii) improperly oligomerized mucin cannot be processed further. After oligomerization, this protein is translocated to the *cis*-Golgi where *O*-glycosylation is initiated. In later Golgi compartments the glycoprotein undergoes sulfation, fucosylation, and sialylation (utilizing specific transferases). The above pathway was characterized using inhibitors of the general post-translational modification pathway. The synthesized mucin was detected by immunoprecipitation and the oligomers identified by gel electrophoresis [38]. The same strategy can be adopted to study mucin synthesis in other systems as well.

The regulation of biosynthesis of the mucins is virtually unknown. Since we have determined the nucleotide sequence of the canine cDNA (which can detect human message), developed an immortalized tracheobronchial epithelial cell line (CT1) system, raised antibodies against TBM, and acquired human epithelial cell lines (HBE1E6E7 and CFT1) which express the mucin gene, the biosynthesis of lung mucin can be followed in CF and compared with normal cells.

Mucin synthesized by primary culture cells has been characterized from canine [39], Chinese hamster [40] and guinea pig trachea [41]. The main problem with these systems is that the cell survive only for a few weeks and thus do not provide a good model system to study the regulation of mucin gene expression. To define the steps involved in the biosynthesis of mucin, well characterized inhibitors were used and intermediates were isolated and characterized.

4. Regulation of mucin gene expression

This field has not been explored very thoroughly. McCool *et al.* [42] studied the synthesis of mucin in T84 human colonic adenocarcinoma cell line and its release in response to different secretagogues. Carbachol (10 mM), prostaglandin E1 (PGE1) (1 mM) and vasoactive intestinal polypeptide (VIP) (0.1 mM) all stimulated mucin release, but histamine (1 mM) had no effect. Whereas VIP is reported to stimulate chloride secretion more strongly than carbachol, it was less effective that carbachol in stimulating mucin secretion. Phorbol 12-myristate 13-acetate (PMA) (0.1–10 mM) also stimulated mucin release strongly, implicating a responsive protein kinase C-pathway. Additive secretory responses were obtained with combined stimulation by VIP (10 nM–1 mM) and carbachol (1 mM). Responses to stimulation with A23187 (1–10 mM) together with PMA (10 nM–10 mM) suggest that cytosolic Ca^{2+} concentration is a modulator of PMA activity. In most other examples, the primary cultures of epithelial cells have been isolated and utilized to study regulation [39, 40].

A. Cell line model – immortalization by HPV18

To study regulation of mucin synthesis in homogeneous population of cells, we immortalized canine tracheal epithelial cells using a vector encoding the human papillomavirus (type 18) *E6* and *E7* genes as has been previously established by Yankaskas *et al.* [43]. The *E6* and *E7* genes are essential and sufficient for the immortalization of human genital keratinocytes as well as human tracheal epithelium [43, 44]. Primary epithelial cells from dog trachea were transfected with a vector containing HPV 18 genes *E6* and *E7*. The resultant cells (CT1) were cloned and maintained in selective medium supplemented with growth factors and hormones. These cells synthesize a mucin glycoprotein which exhibits properties similar to native CTM. Yankaskas *et al.* [43] immortalized human tracheal epithelial cells from normal subjects and cystic fibrosis patients using the papilloma virus vector. They demonstrated that these cells retain a well differentiated phenotype but did not study the mucin expression in these cells.

As yet there has been no suitable model (except the one described above) to study regulation of mucus secretion by tracheal gland cells since *in vitro* propagation of epithelial cells generally results in cessation of differentiated function. Additionally, normal epithelial cells in culture undergo senescence after a short time, thus limiting their usefulness for the study of differentiated functions. The immortalized cells isolated in our laboratory provide a model system to examine the regulation of synthesis of mucin in normal and disease state.

B. Regulation at the pre-mRNA (splicing) synthesis level

Polydispersity of the mRNA is an inherent feature of the mucin synthesizing cells but generally not seen for other genes. Analysis by Northern blot hybridization of RNA from mucus-rich tissue and an adenocarcinoma cell line has revealed multiple transcripts. Polydispersity of mucin mRNA has been observed even when hybridization with probes for housekeeping genes has demonstrated the integrity of the RNA samples. These results may reflect rapid turnover, instability, or incomplete or alternative splicing of this mRNA (see review by Smith *et al.* [45]. The latter process has been demonstrated in MUC1 (4.1 kb allele) [46]. In fish polysialoglycoprotein genes, an alternative possibility was observed that multiple mRNA species are transcribed by multiple related genes which contain diverged numbers of exact repeats [47]. To date there is no concrete evidence for multiple closely related human mucin protein genes. Gerard *et al.* [15] have suggested that polydispersity increases in pathological samples (especially from cystic fibrosis patients). We have demonstrated that polydispersity is less when 3′ UTR (untranslated region) cDNA is used as a probe as opposed to the coding region of the cDNA [21] because the sequence of the 3′ UTR is unique.

C. Regulation at the transcriptional level

Based on the results of the expression of the mucin gene in response to the treatment with different secretagogues, we infer that the induction in the mRNA levels for CTM in CT1 cells requires *de novo* RNA and protein synthesis. Nuclear run-on assay has indicated that the regulation of CTM occurs at the transcriptional level. Whether other steps of regulation are also crucial has not been tested so far (in the case of the CTM). Carcinogens such as benzo-α-pyrene (BAP) also enhance mucin gene expression in HBE1E6E7 cells (Verma and Davidson, unpublished).

D. Regulation at the translational level

Very little has been reported about the regulation of mucin gene expression at the translation and post-translation levels. Peat *et al.* [48] have developed a transgenic mouse strain expressing the human *MUC1* gene product in a tissue-specific manner. The TG4 strain was established using a 40 kb fragment containing 4.5 kb of 5′ and 27 kb of 3′ flanking sequence. The TG18 strain was developed using a 10.6 kb Sac II fragment from the 40 kb fragment (this fragment contained 1.6 kb of 5′ sequence and 1.9 kb of 3′ flanking sequence). Both strains showed tissue specificity of

expression of the *MUC1* gene, which was very similar to the profile of expression seen in human tissues. They also established that the distribution of the SM-3 epitope of PEM (polymorphic epithelial mucin, now called MUC1) in the tissues of the transgenic mice is similar to that seen in humans. Thus, these studies provided the basis for the development of a preclinical model for the evaluation of PEM-based antigens and of antibodies directed to PEM in cancer therapy. Boshell *et al.* [49] have demonstrated the product of the human *MUC1* gene secreted by mouse cells transfected with the full length cDNA. Lesuffleur *et al.* [50] demonstrated differential expression of MUC1 to MUC5 in relation to growth and differentiation of different mucus-secreting HT-29 cell subpopulations.

E. Regulation in disease states

In cancer, and lung diseases such as cystic fibrosis, chronic bronchitis, and asthma, the mucin gene is apparently over-expressed and the protein is aberrantly glycosylated. New insight gained into the mucin gene structure and the organization of the transcriptional regulatory sequences may lead to a better understanding of the presently poorly understood mechanisms of expression of mucin gene in different diseases.

F. Role of second messengers at different steps of mucin synthesis

Data from primary cultures of tracheal epithelial cells isolated from dog and Chinese hamster indicate that the induction of mucin synthesis by these cells is mediated by second messengers such as cAMP. Recently, Manna *et al.* [51] reported the effect of retinoic acid on mucin gene expression in rat airways *in vitro*. Their results demonstrated that retinoic acid stimulated the synthesis of mucin in primary cultures although it is not clear whether retinoic acid-receptor complex directly regulates the mucin gene or induces a cascade of regulatory proteins, one or several of which regulate the mucin gene. Virmani *et al.* [39] demonstrated that cAMP and its analogues enhanced mucin synthesis and secretion in primary cultures of canine tracheal epithelial cells; this secretion stopped when atropine was included in the medium.

5. Future aspects/conclusion: gene therapy (in diseases where the mucin gene is over-expressed)

Mucin gene over-expression may be controlled by genetic therapy to treat the disease. Antisense oligodeoxynucleotides (ODNs) can be designed to complement the splice-junctions, ribosome binding site, and translation start region (covering the ATG codon present in the cDNA of the mucin genes), and used to investigate whether they prevent the expression of these mucin genes. In cancer and infectious diseases, it has been demonstrated that oligonucleotides can act as antisense complements to target sequences of mRNAs to selectively regulate gene expression. Chemically modified (thiophosphate) analogues that are nuclease-resistant can be utilized in such studies. Wagner *et al.* [52] have synthesized phosphorothioate ODNs in which the U and C residues are modified at the 5 position with a propyne group. This modification so improves affinity that inhibition of gene expression could be seen at nanomolar concentrations when the ODNs were microinjected into cells. Cellular uptake of oligonucleotides, either as such or as conjugates, has been well established. [53–55].

Use of antisense oligonucleotides and ribozymes (in diseases such as cancer, cystic fibrosis, chronic bronchitis, asthma)

Ribozymes can be designed (based on the mucin cDNA nucleotide sequence) and tested to prevent mucin gene transcription [56, 57]. This approach has been adapted to design so-called hammerhead ribozymes that can be produced inside the cell by the introduction of viral or plasmid vectors, and could then bind to an antisense sequence and hydrolyse a specific target RNA sequence (demonstrated in the case of HIV, [57] and references therein). A potential difficulty of this approach is that ribozymes may get degraded by ribonucleases *in vivo*.

Designer mucins

Knowledge about mucin structure and regulation may be applied in future in the following areas (designer mucins): the production of synthetic tears in dry eye disease, strengthening of mucin gels in stomach ulcer and inflammatory bowel diseases and gels suitable for transport on mucociliary mucosae in respiratory and genital track ailments.

Conclusion

Mucus has inherent protective and lubricative properties for epithelial surfaces in many biological systems and is overproduced in chronic obstructive pulmonary diseases. Despite a vast number of studies on mucus and its components, fundamental questions such as the protective mechanism of mucus, functional roles of specific components, and identification of structural features that determine functions, have not been clearly explained. Current information indicates the existence of a family of mucin proteins, some of which may exhibit cell and/or tissue specificity. Elucidation of the expression of mucin proteins may provide insight into the role of mucins in health and disease. If airway mucin genes are transcriptionally regulated (in fact studies with the CTM suggest this), initiation of transcription of individual mucin core protein gene may be a primary event leading to goblet and mucous cell metaplasia and hyperplasia in the airways [2, 58–60]. Characterization of the promoter region will be useful in designing modulators which affect mucin gene expression and response to environmental stimuli, inflammatory mediators, bacterial or viral pathogens.

References

1. Gum JR, Hicks JW, Toribara NW, Siddiki B, Kim Y (1994) *J Biol Chem* **269**:2440–6.
2. Rose M (1992) *Am J Physiol* **7**:L413–29.
3. Gum JR (1992) *Am J Resp Cell Mol Biol* **7**:557–64.
4. Carlstedt I, Sheehan JK, Cornfield AP, Gallagher JT (1985) *Essays Biochem* **20**:40–76.
5. Feizi T, Childs RA (1987) *Biochem J* **245**:1–11.
6. Cornfield T (1992) *Glycoconj J* **9**:217–21.
7. Bobek LA, Tsai H, Biesbrock AR, Levine MJ (1993) *J Biol Chem* **268**:20563–69.
8. Gendler SJ, Lancaster CA, Taylor-Papadimitriou J, Duhig T, Peat N, Burchell J, Pemberton L, Lalani E, Wilson D (1990) *J Biol Chem* **265**:15286–93.
9. Hauser F, Hoffman W (1992) *J Biol Chem* **267**:24620–4.
10. Jany B, Gallup M, Tsuda T, Basbaum C (1991a) *Biochem Biophys Res Comm* **181**:1–8.
11. Jany B, Gallup M, Yan PS, Gum J, Kim Y, Basbaum C (1991b) *J. Clinical Invest* **87**:77–82.
12. Bhavanandan VP (1991) *Glycobiology* **1**:493–503.
13. Carnoy C, Ramphal R, Scharfman A, LoGuidice JM, Hourdret N, Klein A, Galabert C, Lamblin G, Roussel P (1993) *Am J Resp Cell Mol Biol* **9**:323–34.
14. Schwab UE, Wold AE, Carson JL, Leigh MW, Cheng P, Gilligan PH, Boat TF (1993) *Am J Resp Dis* **148**:365–9.
15. Gerard C, Eddy RL, Shows TB (1990) *J Clin Invest* **86**:1921–7.
16. Verma M, Davidson EA (1994) *Cancer Biochem Biophys* (in press).
17. Bhargawa AK, Woitach JT, Davidson EA, Bhavanandan VP (1990) *Proc Natl Acad Sci USA* **87**:6798–802.
18. Eckhardt AE, Timpte CS, Abernethy JL, Zhao, Hill R (1991) *J Biol Chem* **256**:9678–86.
19. Xu G, Huan L. Khatri IA, Wang D, Bennick A, Fahim REF, Forstner GG, Forstner JF (1992) *J Biol Chem* **267**:5401–7.
20. Probst JC, Gertze E, Hoffman W (1990) *Biochemistry* **29**:6240–4.
21. Verma M, Davidson EA (1993) *Proc Natl Acad Sci USA* **90**:7144–8.
22. Gambus G, Bolos CD, Andreu D, Franci C, Egea G, Real FX (1992) *Gastroenterology* **104**:93–102.
23. Perini JM, Marianne T, Lafitte JJ, Lamblin G, Roussel P, Mazzuca N (1989) *J Histochem Cytochem* **37**:869–75.
24. Pemberton L, Taylor-Papadimitriou J, Gendler SJ (1992) *Biochem Biophys Res Comm* **185**:167–75.
25. Verma M, Madhu M, Marrota C, Lakshmi CV, Davidson EA (1994) *Cancer Biochem Biophys* (in press)
26. Gum JR, Byrd JC, Hicks JW, Torbiara NW, Lamport DTA, Kim YS (1989) *J Biol Chem* **264**:6480–7.
27. Gum JR, Hicks JW, Swallow DM, Lagace RL, Byrd JC, Lamport DTA, Kim YS (1990) *Biochem Biophys Res Comm* **171**:407–15.
28. Gum JR, Hicks JW, Lagace RE, Byrd JC, Toribara NW, Siddiki B, Fearney FJ, Lamport DTA, Kim YS (1992) *J Biol Chem* **266**:22733–8.
29. Gendler SJ, Burchell JM, Duhig T, Lamport D, White R, Parker M, Taylor-Papadimitriou J (1987) *Proc Natl Acad Sci USA* **84**:6060–4.
30. Gendler S, Taylor-Papadimitriou J, Duhig T, Rothbard J, Burchell J (1988) *J Biol Chem* **263**:12820–3.
31. Spicer AP, Parry G, Patton S, Gendler S (1991) *J Biol Chem* **266**:15099–109.
32. Lancaster CA, Peat N, Duhig T, Wilson D, Taylor-Papadimitrou J, Gendler S (1990) *Biochem Biophys Res Comm* **173**:1019–29.
33. Penotti FE (1990) *J Mol Biol* **213**:37–52.
34. Murre C, McCaw PS, Baltimore D (1989) *Cell* **56**:777–83.
35. Pani L, Overdier DG, Porchella A, Qian X, Lai E, Costa RH (1992) *Mol Cell Biol* **12**:3723–32.
36. West M, Mikovitz J, Princlear G, Liu Y, Ruscetti FW, Kung H, Raziuddin (1992) *J Biol Chem* **267**:24948–52.
37. Strous GJ, Dekker J (1992) *Critical Rev Biochem Mol Biol* **27**:57–92.
38. Dekker J, Strous G (1990) *J Biol Chem* **265**:18116–22.
39. Virmani AK, Naziruddin B, Desai VC, Lowry JP, Graves DC, Sachdev GP (1992) *In Vitro Cell Dev Biol* **28A**:120–7.
40. Christensen TG, Breuer R, Haddad C, Niles RM (1993) *Am J Resp Cell Mol Biol* **9**:287–94.
41. Rahmoune H, Rounding HP, McDonald-Gibson WJ, Lambin G, Hall RL, Roussel P (1991) *Am J Resp Cell Mol Biol* **4**:156–65.
42. McCool DJ, Marcon MA, Forstner JF, Forstner GG (1990) *Biochem J* **267**:491–500.
43. Yankaskas JR, Haizlip JE, Conard M, Koval D, Lazarowski E, Paradiso AM, Rinehart CA, Sarkadi B, Schlegel R, Boucher RC (1993) *Am J Physiol* **264**:C1219–30.
44. Villa LL, Schlegel R (1991) *Virology* **181**:374–7.
45. Smith CWJ, Patton JG, Nadal-Ginnard B (1989) *Ann Rev Gen* **23**:527–77.
46. Lightenberg MJL, Rudolp SA, Genissen AMC, Hilkens J (1990) *J Biol Chem* **265**:5573–8.
47. Sorimachi HY, Emori Y, Kawaski H, Kitajim K, Innoue S, Suzuki K, Inoue Y (1988) *J Biol Chem* **263**:17678–84.
48. Peat N, Gendler SJ, Lalanai EN, Duhig T, Taylor-Papadimitriou J (1992) *Cancer Res* **52**:1954–60.
49. Boshell M, Lalani EN, Pemberton L, Burchell J, Gendler S, Taylor-Papadimitriou J (1992) *Biochem Biophys Res Comm* **185**:1–8.
50. Lesuffleur T, Porchet N, Aubert JP, Swallow D, Gum JR, Kim YS, Real FX, Zweibaum A (1993) *J Cell Soc* **106**:771–83.
51. Manna B, Lund M, Ashbaugh P, Kaufman B, Bhattacharya SN (1994) *Biochem J* **297**:309–13.
52. Wagner RW, Matteucci MD, Lewis JG, Gutierrez AJ, Moulds C, Froehler BC (1993) *Science* **260**:1510–13.
53. Moxham CM, Ho, Y, Malbon CC (1993) *Science* **260**:991–5.
54. Cohen JC (1991) *Pharmac Ther* **52**:211–25.
55. Crooke RM (1991) *Anti-cancer Drug Design* **6**:609–46.
56. Cech TR (1988) *JAMA* **25**:3030–4.
57. Mulligan RC (1993) *Science* **260**:926–32.
58. Basbum CB, Finkbeiner WE (1989) In *Lung Cell Biology* (Massaro D, ed.) pp. 37–79. Marcel Dekker, New York.
59. Basbaum C, Gallup M, Gum J, Kim Y, Jany B (1990a) *Biorheology* **27**:485–9.
60. Basbaum CB, Jany B, Finkebeiner WE (1990b) *Ann Rev Physiol* **52**:97–113.