[Frontiers in Bioscience 6, d1192-1206, October 1, 2001]

STRUCTURAL ORGANIZATION AND CLASSIFICATION OF THE HUMAN MUCIN GENES

Nicolas Moniaux[1, 2], Fabienne Escande[2], Nicole Porchet[2], Jean-Pierre Aubert[2] and Surinder K. Batra[1]

[1]Department of Biochemistry and Molecular Biology, The Eppley Institute for Research in Cancer and Allied Diseases, University of Nebraska Medical Center, Omaha, NE 68198-4525, USA; [2]Unité 377 INSERM, Place de Verdun, 59045 Lille Cedex and Laboratoire de Biochimie et de Biologie Moléculaire de l'Hôpital C. Huriez, 59037 Lille Cedex, France

TABLE OF CONTENTS

1. ABSTRACT

The cells of living organisms in contact with the external environment are constantly attacked by different kinds of substances such as micro-organisms, toxins, and pollutants. With evolution, defense mechanisms, such as the secretion of mucus has been developed. Mucins are the main components of mucus. They are synthesized and secreted by specialized cells of the epithelium and in some case, by non mucin-secreting cells. Little was known about the structure of mucins until a decade ago. This is principally due to heavy glycosylation of mucins, which complicated their analysis. With the application of molecular biological methods, structures of the mucin core peptides (apomucins) are beginning to be elucidated. A total of eleven human mucin (MUC) genes have been identified and numbered in chronological order of their description: *MUC1-4*, *MUC5AC*, *MUC5B*, *MUC6-8*, and *MUC11-12*. Of these, the complete cDNA sequence are published only for six mucins *MUC1*, *MUC2*, *MUC4*, *MUC5B*, *MUC5AC*, and *MUC7*. Human mucin genes, in general, show three common features: I) a nucleotide tandem repeat domain; II) a predicted peptide domain containing a high percentage of serines and threonines; III) complex RNA expression. The tandem repeats in mucins make up the majority of the backbone. Related to their structure, mucins can be classified in three distinct sub-families: gel-forming, soluble, and membrane-bound. Each member from one family possesses common characteristics and probably specific functions. For a long time, they were thought to have the unique function of protecting and lubricating the epithelial surfaces. The study of the mucins structure as well as the relationship between structure and function show that mucins also possess other important functions, such as growth, direct implication in the fetal development, the epithelial renewal and differentiation, the epithelial integrity, carcinogenesis, and metastasis. This review presents the actual knowledge on the mucins

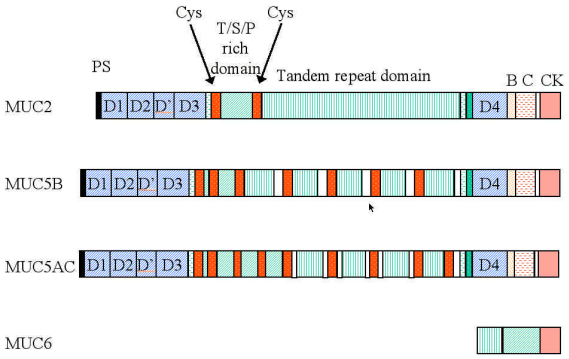

Figure 1. Schematic representation of the deduced amino acid sequences from the mucin genes *MUC2*, *MUC5B*, *MUC5AC*, and *MUC6* localized on the chromosome 11 in p15.5. The domains D1, D2, D3, D', D4, B, C, and CK (cystin knot) show a high level of similarity with the respective domains of the pre pro-von Willebrand factor. The T/S/P domains are the domains rich in serine, threonine, and proline amino acid residues. Cys are domains rich in cystein amino acid residues, and PS is the peptide signal.

structure and the best-characterized function related to their structure.

2. INTRODUCTION

Mucins are O-glycoproteins with a high molecular weight and are produced by secretory epithelial cells for the lubrication and protection of ducts and lumen. Historically, purified mucin has been identified by its amino and carbohydrate composition that consists of a high percentage of serine, threonine, proline, alanine, glycine, and a large proportion of O-linked oligosaccharides (up to 80% of the total mass). The definition of mucins was unclear for a long period. Several families of proteins, membrane-bound or secreted by distinct cellular types with adhesion properties in common, were also called mucins. Thus, some authors make the epithelial mucins out from the leucocytes mucins and the endothelial mucins (1).

The epithelial mucins share a number of common points, primarily in their coding sequence, with repetitions organized in tandem that code for peptides very rich in serine, threonine, and proline residues. The repetitive domain, which contains many putative O-glycosylation sites, for all mucins is in the central position. For the mucins whose genomic organizations are known, this domain is encoded by a unique exon. The size of this exon varies from 2.2 kb for *MUC7* (2) to 21 kb for *MUC4* (3).

When the repetitive domain is longer than several kilobases, it is characterized by an inter-individual VNTR (variable number of tandem repeats) polymorphism. The VNTR polymorphism is caused by the instability of the number of repetitions from generation to generation (4-7). This VNTR polymorphism can be detected at the genomic level, the protein level (8) and the RNA level (9). Southern blot techniques also reveal the presence of a mutational polymorphism for the repetitive domain of the mucin genes (4,5,7,10).

A second typical feature of mucins is the presentation in Northern blots of a polydisperse signal from low to very high molecular weight (11-16). The polydisperse pattern has been considered an original and inherent feature of mucins, and for a long period was unexplained. However, it has been demonstrated that mucin mRNAs are not polydisperse but rather very large (9), among the largest reported for eukaryotes. Debailleul et al showed that the polydisperse signals observed for mucin mRNAs are inherent to degradation, and the preparation of mucin RNAs without extraction artifacts require an improved method to analyze large mRNAs (9).

Because of their complex structure, mucins are difficult to study by classical biochemical procedures. With the application of recombinant technology, structures of the mucin core peptides (called apomucins) are being elucidated. Eleven human mucin (MUC) genes have been identified and designated as *MUC1-4*, *MUC5AC MUC5B*, *MUC6-8* and *MUC11-12* (2,12,14,16-23). Based on recently deduced amino acid sequences, mucins are now categorized in three distinct families: gel forming (MUC2, MUC5AC, MUC5B and MUC6), soluble (MUC7), and membrane-bound (MUC1, MUC3, MUC4 and MUC12). MUC8 and MUC11 remain unclassified on this review.

3. HUMAN MUCIN GENES

3.1. Gel forming mucins

The gel forming family of mucins is composed of MUC2, MUC5AC, MUC5B, and MUC6. Their genes are clustered in a complex of 400 kb very rich in CpG islands on chromosome 11 in region p15.5 (24). The cluster is localized between HRAS and IGF2. The deduced amino acid sequences of these genes are organized in domains with similar structural schemes. They exhibit in their distal part a high level of similarity with the pro-von Willebrand factor (Figure 1) (25,26). The four genes are organized in a complex in which the distribution of the restriction sites (24) with a great deal of symmetry and repetition seems to demonstrate the existence of many events of duplication. Computational and phylogenetic analyses have permitted the development of an evolutionary history of the four human mucin genes from an ancestor gene common to the human von Willebrand factor gene (27,28).

3.1.1. MUC2

MUC2 was first identified and described by Gum et al in 1989 (13). The three first cDNA clones isolated, SMUC40, 41, and 42, have been cloned after the screening of a small intestine cDNA library with antisera prepared against the deglycosylated protein backbone of human colon cancer xenograft mucin. The three clones are composed of sequences repeated in tandem with a repetition unit of 69 bp. The full-length cDNA sequence (Figure 1), with its largest allele measuring 15,720 bp was characterized after a new cDNA library screening with these probes and RACE-PCR (Rapid Amplification of cDNA Ends-polymerase chain reaction) procedures (18,29).

The central domain of MUC2 is composed of two highly repetitive sequences. The first, in the central position, is characterized by the perfect repetition of one motif of 23 amino acids. This domain shows in Southern blot a VNTR polymorphism with the number of repetitions varying from 51 to 115 (6). The second, located upstream, is composed of an irregular sequence repeated in tandem with a unit of 347 amino acids. These two sequences are rich in amino acid residues of serine, threonine, and proline. Study of mucins isolated from nude mouse xenografts of the LS174T colonic adenocarcinoma cell line by gel filtration and CsCl density gradient centrifugation showed that 78 % of the threonine residues are O-glycosylated. MUC2 may possess up to 1,000 oligosaccharidic chains (30). Two sequences rich in cysteine residues flank the N-terminal domain made up of the repeat of 347 residues. These domains are called Cys domains. MUC2 also possesses five D domains, so called because of their homology with the D domains of the von Willebrand factor. The D1, D2, D', and D3 domains are localized in its N-terminal part, whereas the D4 domain is localized in the C-terminal position. Downstream of the D4 domain, three other sequences show similarity with domains of the von Willebrand factor, one domain C, one domain B, and one domain CK (Cystin Knot). The CK domain is also found in other secreted proteins (31) such as the NDP (Norries Disease Protein). Sequence pattern searches and three-dimensional modeling suggest that the CK domain of the NDP has a tertiary structure similar to that of the transforming growth factor beta (TGFbeta) (32).

The initiation point for *MUC2* gene transcription lies within a 7000-base GC-rich region. Like many genes exhibiting tissue-specific expression, it contains the TATA element located 25 bases from its initiation (33). Computer analysis revealed the presence of a CACCC motif between bases -91 and -73. This CACCC box appears to be important for *MUC2* gene transcription. Sp1 is the most abundant binding factor for this motif. In addition, this element seems able to bind other factors. Likely candidates for additional factors that are able to bind this element include Sp2, Sp3, and Sp4. These are zinc-finger proteins with sequence similarity to Sp1 (34,35). Two other regions have also been identified as important for the expression of the *MUC2* gene; one localized between bases -228 and -171 that may confer cell-type specificity, and the second a nuclear factor-kappaB site located between bases -1452 and -1441 that participates in the induction of *MUC2* by *Pseudomonas* (34,36). Elements required for small intestine specific expression are also located between bases -2864 and +17 of the *MUC2* 5' flanking sequence (37).

3.1.2. MUC5B

A human tracheobronchial lambda gt 11 cDNA library was screened using an antiserum prepared against the deglycosylated protein backbone of human tracheobronchial mucins. Two cDNAs, designated JER28 and JER57, were obtained (12). These clones consisted of an imperfect repetition of one unit of 87 bp. Using these probes, a full-length cDNA of 16986 bp has been isolated and characterized (Figure 1) (38-43). The repetitive domain of *MUC5B* encodes a large exon in the central position. Composed of 3570 amino acid residues, the MUC5B central domain is organized by three distinct sequences that alternate. Nineteen subdomains can be individualized (41). Seven subdomains of 108 amino acid residues, Cys1 through Cys7 with 10 residues of cysteine, have a structure similar to the Cys domains of MUC2. Five subdomains are composed of the imperfect repetition of 29 residues. Four of these five subdomains are flanked in one of their extremities by the same unique sequence. Finally, three sub-domains possess a unique sequence rich in serine, threonine, and proline residues. The central domain of MUC5B is composed of four super-repeats of 528 amino acid residues. The 5'-terminus extremity of its cDNA, 4023 bp long, is constituted of 30 exons. The deduced amino acid sequence codes for four D domains, D1, D2, D', and D3, similar to those of MUC2 and von Willebrand factor (39,43). Its 3'-terminus extremity, 2.9 kb long, possesses 18 exons. The deduced amino acid sequences also code for the domain D4, B, C, and CK domains similar to the von Willebrand factor (40,42). The peptide deduced from the sequence of the 48 exons of MUC5B consists of a 5,662-amino acid polypeptide with a Mr approximately 600 kDa.

The first data regarding the sequence of the 5'-flanking region and the promoter activity of the human mucin gene *MUC5B* have been published recently (44). The 5'-flanking region upstream of the transcription start site revealed the presence of a TATA box-like sequence (TACATAA) located between bases -32 and -26. Near to the TATA box-like sequence, are found potential binding sites for c-Myc/N-Myc/Max transcription factors (consensus CACGTG), as well as Sp1 and numerous GC and CACCC boxes. Further upstream, putative binding sites were found for AP-1, NF-kappaB, cAMP-response-element-binding protein (CREB), glucocorticoid response element (GRE), and a silencer called CIIS1. Moreover, throughout the 5'-flanking sequence upstream of the TATA box, numerous binding sites are present for factors involved in intestinal [e.g., hepatocyte nuclear factor (HNF)-1, HNF-3, and gut-enriched Küppel factor (GKLF)] or respiratory [HNF-3, thyroid transcription factor (TTF)-1] cell differentiation. In the first intron, a high number of GC boxes, Sp1 binding sites and CACCC are present. The central part of the intron is clustered by CACCC boxes, followed by an array of eight GA repeated in tandem containing the consensus sequence GGGGAGGGGCT, each separated by 8 to 10 bp. Other putative sites for the transcription factors have been shown such as three Sp1 binding sites, two activator protein (AP)-2 sites, one NF-kapaB, one GATA-1 and one Adh-1 site. Sp1 site showed direct involvement in the regulation of *MUC5B*, in the promoter sequence as well as in the intron 1. Moreover, another factor, called NF1-MUC5B nuclear factor, has already been demonstrated to bind into the intron 1 (45).

3.1.3. MUC5AC

The structural organization of MUC5AC is now completely known (Fgure 1). Two clones, JER47 and JER58, were isolated, in parallel with the work on *MUC5B*, by screening of a human tracheobronchial lambda gt 11 cDNA library using antiserum prepared against the deglycosylated protein backbone of human

tracheobronchial mucins (46). These clones are composed of a sequence repeated in tandem of 24 bp, flanked in the case of JER47 by a sequence of 330 bp that codes for domains rich in cysteine amino acid residues. These Cys domains are similar to those of MUC2 and MUC5B. The clone NP3a was isolated from a human nasal polyp cDNA library using two unique nucleotide probes for human tracheobronchial mucin glycoprotein (TBM) generated via PCR with degenerate primers deduced from the TBM:TR-3A tryptic peptide sequence (47). The clone L31 was isolated from an HT29-MTX (methotrexate) expression library using a polyclonal serum specific for normal gastric mucosa (48). The two clones code for the C-terminus extremity of MUC5AC. They possess the domains similar to those of the von Willebrand factor D4, C, B, and CK. The isolation and characterization of a genomic cosmid clone, designated ELO9, spanning the 3'-region of *MUC5AC* and the 5'-region of *MUC5B*, shows that *MUC5AC* and *MUC5B* have the same transcriptional orientation. Moreover, comparative molecular analysis of the entire sequence of the 3'-region from *MUC5AC* and *MUC5B* points to a remarkable similarity in the size and the distribution of exons (18), and in the type of splice sites (49). An oligonucleotide based on the sequence isolated by successive CsCl-gradient ultracentrifugation in the presence of guanidinium hydrochloride from human gastric mucin was used to screen a human stomach lambda ZAPII cDNA library. Several clones have been isolated, the largest one, HGM-1 encoded 850 amino acid residues (50). A second screening of the human gastric cDNA library with the previously identified *MUC5AC* sequences (51) and a RACE-PCR procedure (52) better characterized the full-length sequence of the 5'-terminus region of MUC5AC. The region, 1,858 amino acid residues long, is composed of the D1, D2, D', and D3 domain in a way comparable to MUC2 and MUC5B. The size of *MUC5AC* cDNA is estimated to be about 16.6 kb long, therefore encoding a 5525 amino acid long peptide.

Recently, the central domain of MUC5AC has been characterized (53). It presents a structural organization similar to those of MUC5B. It is composed of 17 major domains. Nine code for cysteine-rich domains (Cys1 to Cys9) and exhibit high sequence similarity to the cysteine-rich domains described in the central region of MUC2 and MUC5B. Domains Cys1 to Cys5 are interspersed by domains enriched with serine, threonine and proline residues. Domains Cys5 to Cys9 are interspersed by four domains (TR1 to TR4) composed of various numbers of MUC5AC-type repeats.

Much less information is available regarding *MUC5AC* promotor. To date, only the computer analysis of its AUG upstream sequence is known. The upstream sequence contains a TATA box located between bases -23 and -29, as well as a nuclear factor -kapaB, Sp1, GRE, AP-2, and a CACCC box (52). MUC5AC has been shown to be up-regulated by *Pseudomonas aeruginosa*, with a 15 to 20-fold induction of transcription activity in epithelial cells stably transfected with *MUC5AC*-luciferase reporter constructs and exposed to *Pseudomonas aeruginosa*. Several responsive elements for *Pseudomonas aeruginosa* have been identified in the 4 kb DNA fragment immediately upstream of the *MUC5AC* transcription site. Li et al. (54) showed that *P. aeruginosa* lipopolysaccharide activated the SRC-ras-MEK-pp90rsk signaling pathway that leads to the activation of NF-kappaB.

3.1.4. MUC6

Of the four mucins genes clustered on chromosome 11p15, MUC6 is the least characterized (Figure 1). The first cDNA clone of *MUC6* was originally isolated from a human gastric cDNA library (16) The cDNA sequence is characterized by a tandem repeat region whose individual repeat unit is 507 base pairs (169 amino acids) long. A combination of genomic, cDNA, and PCR techniques was used to isolate the carboxyl-terminal end of MUC6 (55). The 3'-unique sequence contained 1083 base pairs of coding sequence (361 amino acids) followed by 632 base pairs of the 3'-untranslated region. The coding sequence consists of two distinct regions; the first is 270 amino acids long (62% serine, threonine, and proline with no cysteine residues), and the second containing the carboxy-terminal 91 amino acids (with 12% Cysteine). This domain has approximately 25% amino acid similarity to the CK domain of the human mucins MUC2, 5AC, and 5B and the von Willebrand factor.

3.1.5. Mucus network

Not only is the general structural organization of the gel forming mucins similar to those of the von Willebrand factor, both types of proteins share common properties, such as the formation of inter-molecular disulfide bridges. Indeed, the hemostatic functions of the human von Willebrand factor depend on the normal assembly of disulfide-linked multimers from approximately 250 kDa subunits. Subunits initially form dimers through disulfide bonds of their CK domains. Dimers then form multimers through disulfide bonds of their D domains at the N-terminus of each subunit (56,57). In a similar way, mucin monomers form dimers with their CK domains and then, with the D domains from their N-terminal part, form multimers responsible for the tri-dimensional mucus network (58,59). The events of disulfide-linked dimerization and the N-glycosylation of the mucin monomers are achieved in the rough reticulum endoplasmic (rER), before the O-glycosylation and the sulfation. Each of the precursors of MUC2, MUC5AC, MUC5B, and MUC6 form a single species of disulfide-linked homo-dimers (59).

Some animal mucins, like the porcine submaxillary mucin (PSM) (60-62), the bovine submaxillary mucins (BSM) 1 and 2 (63-65), and the frog integumentary mucins (FIM-B.1) (66,67), also possess in their distal parts the D (D1, D2, D', D3, and D4), C, and CK domains. In a similar way, the animal mucins form homo-dimer and then multimerize via these domains (58,68-70).

MUC2, MUC5AC, and MUC5B possess in their central domain several domains rich in cysteine amino acid residues, the Cys domains. The numbers of the Cys domains differ from one mucin to another. The Cys domain could be implicated in inter-molecular disulfide formation.

Figure 2. Schematic representation of the mucin MUC7. Structural organizarion of MUC7. PS: peptide signal, T/S/P: domain rich in serine, threonine, and proline amino acid residues.

As the composition of gel forming mucin differs from tissue to tissue, according to the mucins present, the rheologic properties of the mucus could also be different and specific. This hypothesis has not been corroborated and further investigations are needed.

3.2. Soluble mucin, MUC7

3.2.1. Genomic organization

Screening a human submandibular gland cDNA library with a rabbit antibody, anti apo-MG2, isolated the first cDNA clone of *MUC7* (71). MG2 is a low-molecular-mass mucin population (150 to 200 kDa) secreted by the submandibular gland and the sublingual salivary glands (72). The full-length cDNA was isolated and characterized after screening the same library with the previous probe (73). Compared to the other human mucins, MUC7 has a very simple architectural organization (Figure 2) (a gene of 10 kb long and only 3 exons). The cDNA sequence of *MUC7* encodes a 39 kDa protein of 377 amino acid residues. MUC7 reveals five distinct domains (74). The central domain of the protein is constituted of five or six (depending on the allele) perfect tandem repeats, each comprising 23 residues. A histatin-like domain with a leucine-zipper segment, followed by a moderately glycosylated domain, constitute the N-terminus part, a heavily glycosylated domain, and a second leucine-zipper segment for the C-terminus extremity. The distal regions of MUC7 do not exhibit any cysteine rich domain, only two cysteine residues are present toward its N-terminus part.

Little information is available regarding the regulatory sequences of *MUC7*. A TATAA box is present at -24 to -19 and a CAAT sequence at -83 to -79 (2). Several other regulatory elements have also been found including the AP-1 element, the GRE, and the cAMP reponse element.

3.2.2. Antimicrobial agent

Although MUC7 is a low molecular weight mucin with a simple structure, it has a very important function as an antimicrobial agent in the oral cavity. Previous studies have reported that MUC7 in salivary secretions could interact with a variety of microorganisms, such as oral *Streptococci* (75-77) *Pseudomonas aeruginosa*, and *Staphylococcus aureus* (78), *Actinobacillus actinomycetemcomitans* (79), and *Eikenella corrodens* (80). It has also been reported to agglutinate the AIDS virus HIV-1 (81,82) and to inhibit HIV infection (83). The two cysteine residues located in the N-terminal region of MUC7 seem to be directly implicated in these activities (77). Moreover, MUC7 has an anti-candidal activity (84) via its histatin-like domain. The histatins are a family of small histidine-rich peptides found in parotid and submandibular secretions (85-87), and the candidacidal properties of these are well known (88,89).

3.3. Membrane-bound mucins

The membrane-bound or membrane-associated mucins family is composed of MUC1 (19,90), MUC3 (MUC3A and MUC3B)(14,91), MUC4 (3,92), and MUC12 (23). Until recently, MUC1 was the only identified (19,90,93) member of this group and is considered as a mucin-like molecule. At least three members of the subfamily are clustered on chromosome 7 in the region of q22 (14,23,91). They seem to share a common evolutionary history and may come from a common ancestor gene.

The membrane-associated mucins share several properties, as to be expressed by distinct cellular types, epithelial or not. They can be expressed in four distinct forms; membrane-anchored, soluble (proteolytic cleavage of the membrane-bound form), secreted (alternative splicing variants), and lacking the main feature that characterize all mucins, the tandem repeat array (alternative splicing variants) (94-97).

For MUC1 and the rat Muc4, the biosynthesis has been shown to follow a specific and uncommon course. Indeed, they reach the apical cell surface in an incompletely glycosylated state and additional oligosaccharides are added to the glycoproteins in a second process involving recycling (98,99). The function of this biosynthetic process is still unknown. It has been proposed that in each form, the membrane-associated mucins may play distinct roles.

3.3.1. MUC1

MUC1 is known by several names, the most common being PEM (Polymorphic Epithelial Mucin), episialin, DUPAN-2, DF3, HMFG (human milk fat globule), EMA (epithelial membrane antigen), CD227, and MUC1 (100). It has been isolated from various tissue samples including human mammary epithelial cells (90,93), ovarian cells (101), and pancreatic cells (19). Although in all these tissues, MUC1 apomucin appears to be identical, each tissue expresses distinct glycoforms with a molecular weight varying from 250 to 500 kDa in the mammary glands (102) or up to 1000 kDa in the pancreas (103).

Like the other mucins, MUC1 is organized structurally in domains (Figure 3), with a central domain made up of a sequence repeated in tandem with a perfect unit of repetition of 20 amino acid residues (104). This domain presents a VNTR polymorphism, varying in size from 400 to 2400 residues. The sequences located on both sides of the central domain are composed of the same unit of repetition with an imperfection that increases with the distance to the center of the protein. The N-terminal extremity is composed of the leader sequence. The deduced amino acids of the remaining part code for three domains: a unique extracellular sequence, a transmembrane sequence, and a cytoplasmic tail. The cDNA consists of seven exons; exon 1 encodes the leader peptide, exon 2 the central domain, and exons 6 and 7 respectively the transmembrane sequence and the cytoplasmic tail.

Anchored in the membrane with its full O-glycan

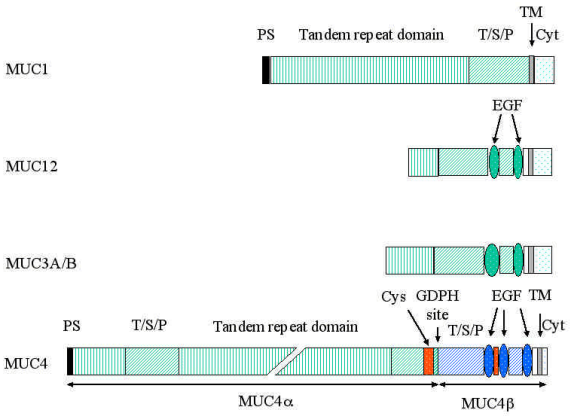

Figure 3. Schematic representation of the membrane-bound mucins. Structural organization of the membrane-bound mucin MUC1, MUC3A/B, MUC4, and MUC12. PS: peptide signal, T/S/P: domain rich in serine, threonine, and proline amino acid residues, Cys: cysteine rich domain, GDPH site: GDPH proteolitic cleavage site, EGF: EGF-like domain, TM: transmembrane domain, Cyt: cytoplasmic tail.

moiety, MUC1 presents a large extended conformation (105,106). The negative charges carried by the glycan moiety extend the protein to a size that is predicted to be around 500 nm for its largest allele (106). This conformation provides anti-adhesive properties to MUC1 (107), properties directly implicated in the morphogenesis of the epithelial tissues (108,109), as well as in tumor progression or metastasis by disturbing the cell-cell and/or cell-matrix interactions. However, perturbations of its glycosylation relevant to numerous pathologic situations (110) create new glycosidic epitopes (sialyl Lewisx and sialyl Lewisa), ligands for the P- and E-selectins (111), and ICAM-1 (112). In this case, MUC1 presents adhesive properties.

MUC1 interacts directly with the beta-catenin via the SXXXXXSSL motif in its cytoplasmic tail (113). The beta-catenin is a protein that has important functions in the formation of the junction cell-cell by interaction with the E-cadherin (114,115). The beta-catenin also binds APC (adenomatous polyposis coli) (114-116). APC is an essential partner in the signal pathway Wingless/Wnt-1 (117). This intracellular pathway is directly implicated in the development of the brain and the axis specification (118). The activation of this pathway (Wnt-1 expression) results in the accumulation of beta-catenin free in the cytoplasm via an inhibition of the GSK3beta (119,120). Whatever the partner that binds the beta-catenin, the complexes are mutually exclusive (121). APC overexpression reduces the level of free cytoplasmic beta-catenin, and thus reduces the complexes beta-catenin/E-cadherin along with reducing the intercellular adherence (122,123). The formation of these complexes is regulated by the phosphorylation of the cytoplamic tail of each partner by the GSK3beta (124). After phosphorylation, the beta-catenin is degraded. The GSK3beta is also able to phosphorylate the beta-catenin binding site on the MUC1 cytoplasmic tail (125). The more the cytoplasmic tail of MUC1 is phosphorylated, the less MUC1 interacts with the beta-catenin (126). The relative levels of MUC1, E-cadherin, beta-catenin, GSK3beta, and APC seem to be very critical to maintaining the epithelium integrity.

However, when the cytoplasmic tail of MUC1 is glycosylated, it is able to interact with other partners like MUC1/SEC and MUC1/Y (127,128). MUC1/SEC (97,129) and MUC1/Y are alternative splice variants of the MUC1 gene (130).

MUC1/SEC is an isoform of MUC1, resulting from an alternative splicing event occurring in the 3'-extremity of the tandem repeat array. MUC1/SEC is co-linear with the gene. It encodes an open reading frame containing only 160 amino acid residues downstream from the tandem repeat array (97). MUC1/SEC does not possess the transmembrane sequence and the cytoplasmic tail of MUC1. MUC1 and MUC1/SEC do possess 149 amino acid residues in common in their C-terminal extremity, although, MUC1/SEC has a unique sequence of 11 residues that has been used to generate specific monoclonal antibodies (131). With these antibodies, the existence of the MUC1/SEC protein as a secreted form has been shown by breast cancer cells as well as by body fluids obtained from breast cancer patients (131).

MUC1/Y is also an alternative splice variant from the MUC1 gene, with a 1.2 kb full-length cDNA. It is characterized by the deletion of the central domain (encoded completely by exon 2), corresponding to the highly O-glycosylated repetitive sequence (130). Using a specific antibody (6E6/2), MUC1/Y showed expression in various epithelial tumors, such as breast and ovarian cancers, but it is undetectable in the adjacent normal tissue (132,133). MUC1/Y seems to act as a membrane receptor that undergoes tyrosine and serine phosphorylation and to activate the signaling cascade via GRB2 (128). MUC1/Y has been shown to enhance tumor initiation and progression in vivo (133). Recently, MUC1/SEC has been identified as a cognate binding protein of MUC1/Y (127). The interaction of MUC1/SEC with MUC1/Y induces MUC1/Y phosphorylation and changes the cell morphology. Other splice variants deleted for the tandem repeat array are also identified for MUC1, such as MUC1/X (133) or MUC1/Z (134), but no real function are yet known for these variants.

MUC1 is an O-glycoprotein that is expressed at a basal level in most epithelial cells (135). Computer analysis of the 5'-sequence upstream of the initiation point reveals the presence of 104 Cys elements in the sense strand and 67 in the antisense (33,35,136,137). Ubiquitous Cys elements are found, such as TATA box, CCAAT-, E-, and GC-boxes. Other elements found are boxes that bind AP1, AP2, AP3, AP4, CTF/NF1, ER, PR, Sp1, STAT1, STAT3, STAT5, and YY1. Different boxes for tissue-specific regulation are also found including elements regulating the transcription in mammary epithelial cells (MAF, MGF, MP4, RME, PMR, SpA, and WAP), elements specific for transcription in hemapoetic cells (BKLF/TEFII, GATA1, v-Myb, c-Myb, and MZF1), elements responsible for the

transcription in immunospecific cells (AML-1, Gfi-1, Ikaros, IL-6 RE, LyF1, NF-GMCSF, NF-mu-E1, NF-Y, Pu-box, SRY, TCF-1, TdT Inr, XBP1, and W-element), the elements that are responsible for the transcription in hepatocytes (ARP-1, HNF-5, LF-A1, and H-APF1), elements that control the transcription in muscle cells (Myo D, Nkx-2.5, and SEF1), and elements specific for viral promoters (JCV, LBP-1, LVC, PEA1, PEA3, PV-E2, T-ag SV40, retroviral TATA-box, TEF-1, TEF-2, and TFII-ML-Inr2) (138). Interestingly, a large proportion of these cys elements are clustered or have overlapping sequences that suggest a very precise mechanism of regulation. Moreover, the 5' and 3' regions of the MUC1 promoter possess independent promoter activities. The TATA box and the GC boxes located at -30/-25 and -90/-137, respectively, could govern formation of the initiation complex (ITC), in the 3'-end regulated transcription, while two initiator elements, TFII-1-ML-Inr2 (-661/-653) and TdT Inr (-634/-627), might control the ITC formation in the 5'-end regulated transcription (137). These different ITCs could be involved in the transcription of specific MUC1 isoforms.

3.3.2. Cluster of mucins located in 7q22

3.3.2.1. MUC3

The sequences of MUC3, A and B, are partially known (Figure 3). MUC3 was initially reported after the screening of a small intestinal lambda gt11 cDNA library using antibodies raised against deglycosylated small intestinal mucins (14). Two partial cDNA clones, SIB124 and SIB139, were isolated and sequenced. They encoded 17 amino acid residues repeated in tandem. With the use of this sequence as a probe, two other coding sequences repeated in tandem have also been described for *MUC3*; one of 1125 bp (139), and one of 177 bp (140). The organization of the central repetitive domain of *MUC3* remains unclear. It presents a VNTR polymorphism with a 51 bp repetitive sequence (139). MUC3 has been located on chromosome 7 in the region q22. The smallest fragment recognized by the 51 bp tandem repeat has a size of 200 kb (139). *MUC3* appears to be a huge gene.

RT-PCR and cDNA library screening procedures have been used to characterize the 3'-terminus region of MUC3. It appears that the *MUC3* gene, by an alternative splicing mechanism, encodes a family of proteins that can be membrane-bound or secreted (96,141). To date, four distinct splice variants have been described for *MUC3*. One has a 3'-terminus extremity composed of 10 exons that code for the membrane bound form of MUC3 (141). The exon 2 and 7 code for two EGF-like domains, exon 8 codes for the transmembrane domain and the exons 9, 10, and 11 code for the cytoplasmic tail. The three other splice variants code for the secreted forms of MUC3, in which distinct domains, such as the second EGF-like domain, are deleted (96,141).

The 5'-extremity of the central domain of *MUC3* as well as the 5'-terminal sequences is still unknown. The organization of the central domain, which appears to be complex, might explain why the sequence of *MUC3* is still uncompleted. Its rat homologue (142) as well as its mouse homologue (143) have been identified but also in these cases, only a partial sequence of the central domain and 3'-terminal extremity are known. With their human homologue, they code for a membrane-bound mucin that contains two EGF-like domains. No splice variant has been described.

Recently, based on nucleotide changes observed in its sequence from one single individual, the existence of a second *MUC3* gene carried by the same 200 kb DNA fragment has been proposed (91). Both *MUC3* genes, now called *MUC3A* and *MUC3B*, show a unique exonic sequence ranging from 94 to 100% identity and 95% similarity for the intronic sequences. To date, nothing is known about the functions of MUC3 (A and B).

3.3.2.2. MUC12

MUC12 is a recently described mucin. The first cDNA fragment of *MUC12*, dd29, has been identified using a differential display procedure using colorectal cancer and normal colon samples (23). The dd29 clone encodes a 28-amino acid residue degenerated tandem repeat that presents 71% similarity with that of MUC11. Using cDNA library screening and RT-PCR techniques, the 3'-terminus of MUC12 has been characterized. It presents the same structure as MUC3, with two EGF-like domains, one transmembrane sequence, and a cytoplasmic tail. MUC12 and MUC3 C-termini present 34% and 38% homology, respectively, with the rodent protein known as rMuc3. The three molecules also share the same domain organization distinct from that of MUC4 (Figure 3). It is difficult to actually describe rMuc3 as the rodent homologue of MUC3 or MUC12. Even if MUC12 seems to be closely related to MUC3, it presents a distinct pattern of expression (23).

The recent descriptions of MUC12 as well as the putative existence of two MUC3 (MUC3A and MUC3B) open the field of investigation for the comprehension of the functions of the membrane-bound mucins.

3.3.3. MUC4

MUC4 was initially identified after the screening of a lambda gt11 cDNA library constructed from human tracheo-bronchial mucosa with a polyclonal antiserum raised against deglycosylated glycopeptides from human bronchial mucins (15). One cDNA fragment has been isolated and named JER64. It contains 48 bp tandem repeat sequence. The corresponding gene has been localized with JER64 used as a probe on the chromosome 3 in the region q29 (4,15). Using this probe, *MUC4* has been shown to exhibit a VNTR polymorphism in its tandem repeat array. Alleles observed vary between 7 and 19 kb after digestion by EcoRI/PstI endonuclease and correspond to a variation in the number of repetitions ranging from 145 to 395 units (3,4). By RT-PCR RACE-PCR experiments and cDNA library screening, the full-length sequence of MUC4 has been characterized (Figure 3) (3,92). Its deduced amino acid sequence consists of a 27-residue peptide signal followed by three imperfect repetitions of a motif varying from 126 to 130 residues and by a unique sequence of 554 residues. The central domain is composed of a perfect repetition of 16 residues. The C-terminal region can be divided into 12 domains (CT1 to CT12). It possesses two

domains rich in N-glycosylation sites, three EGF-like domains, a transmembrane sequence and, a cytoplasmic tail (3,92,144). A GDPH cleavage site is found between the domains CT4 and CT5. The MUC4 precursor, a 930 kDa apomucin, provides the MUC4alpha and MUC4beta subunits. MUC4 is predicted to be a membrane-associated 2.12 mm long mucin, in which MUC4alpha is the mucin type-associated subunit and MUC4beta is the growth factor-like subunit.

To date, 24 distinct cDNAs, transcript from *MUC4* gene, have been identified (94,95,144). They result from a complex alternative splicing mechanism, mainly of the 3'-terminal region, but also for two of them, by an alternative splicing of the central repetitive domain. The 24 isoforms are called sv0 to sv21-MUC4, MUC4/Y and MUC4/X. These 24 distinct cDNAs encode 19 different forms of MUC4: 5 membrane-bound, 12 secreted, and 2 growth factor-like membrane-bound forms without the tandem repeat domain. Several of these splice variants encode the same protein. Even though no quantitative expression studies have been performed, it appears that the full-length MUC4 form, also called sv0-MUC4, is the main isoform expressed by the different tissue samples studied (95). MUC4/X and MUC4/Y forms are expressed by cancer tissue samples including lung and pancreas. It is not yet known whether their expression is related to the carcinogenesis as it is the case for MUC1/Y.

No precise functions have been attributed to MUC4 products, but some functions are known for the rat homologue of MUC4, rMuc4, or SMC (sialo mucin complex) (92,145,146). Even if the *rMuc4* shares the same structural organization as *MUC4*, no splice variants have yet been characterized. rMuc4 is characterized under two distinct forms, a membrane-bound and a soluble form. The soluble form results from the proteolityc cleavage of the membrane-associated form (147). As is the case for MUC1, anchored in the membrane with its full O-glycan moiety, rMuc4 presents a large extended conformation that provides anti-adhesive properties (148,149). These properties are implicated in the cell-matrix and cell-cell interactions, in tumor progression or metastasis (150), as well as in protection again the natural killer cells (151). rMuc4 also appears to be able to interact with the proto-oncogene ErbB2 (152), but the effect of this interaction is unclear.

3.4. Uunclassified mucins

The unclassified mucins are partially known mucins for which only a short cDNA sequence of the repetitive sequence has been characterized.

3.4.1. MUC8

Polyclonal antibodies against deglycosylated human tracheo-bronchial mucin were used to select immunoreactive clones from a Uni-ZAP cDNA expression library prepared from normal human tracheal mRNA (153). One positive clone, designated pAM1, revealed a partial 941 bp cDNA that encoded a 313-amino acid polypeptide. It consisted of imperfect 41-nucleotide tandem repeats that encoded a unique polypeptide with two types of consensus repeats. The corresponding gene was mapped to chromosome 12 in the region q24.3. Using the RACE-PCR procedure, the 3'-terminus region has been cloned (22). It contains only a very short coding unique sequence followed by a stop codon and a 458 bp of 3'-untranslated region. Due to the lack of longer carboxy coding sequence, MUC8 is still considered an unclassified mucin.

3.4.2. MUC11

Like *MUC12*, the first cDNA fragment of *MUC11*, called dd34, has been identified using a differential display procedure of colorectal cancer and normal colon tissue samples (23). This cDNA fragment is the only sequence known to date for *MUC11*. It is 2.8 kb long and encodes a degenerated sequence repeat in tandem of 28 amino acid residues. The repetitive sequence of MUC11 shares 71% similarity with that of MUC12. Because of its localization, chromosome 7 in q22, *MUC11*, like *MUC12*, *MUC3A*, and *MUC3B* may encode a membrane-bound mucin.

4. PERSPECTIVES

This article presents for the first time a complete review of current knowledge of structural organization of all discovered human mucin genes.

Mucin, or not mucin, that is the question?

New molecules, discovered all along the years, share with mucins some structural properties, like secreted by epithelial cells, possess a mucin-like domain (name given to a domain rich in serine, threonine and proline amino acid residues), and an EGF-repeat growth-factor like subdomain. Even if these molecules do not possess the classical mucin structure, some of them received a MUC designation, for instance, the recently discovered MUC13 (154). MUC13 is a low molecular weight membrane-bound protein (54 kDa) that possesses a mucin like small domain (two repetitions rich in serine, threonine, and proline residues) and three EGF-like domains. Because of its structural organization, MUC13 can be defined as a "mini-mucin". It is possible that the growth factor functions of MUC13, devoid of mucin repeat, are more important than the classical mucin functions. It has maintained the tandem repeat domain for its simpler expression. MUC13 appears to be a mucin for which the growth factor functions are more important than the mucin functions and so have raised the evolution of the molecule.

In summary, mucins appear to be a complex family with three distinct structures (gel-forming, soluble, and membrane-bound,) and so potentially three distinct patterns of functions. Historically, mucins were defined as the main component of the mucus, with the function of protecting and lubricating all epithelium. Mucins were considered to be the first immunological system. Recent developments regarding the relation structure-function of the mucins carry implications for the mucins. They may be directly implicated in the development as well as in the integrity of the epitheliums. Dysregulation of their expression seems to play a key role in tumoral and metastatic progression. Now that the structure of the mucin genes are being known, research can be directed toward the comprehension of these complex mechanisms.

5. ACKNOLEDGMENTS

This work was supported by the grants from the National Institutes of Health (P5O CA72712 and RO1 CA78590). Ms. Kristi L.W. Berger, communications specialist and editor, Eppley Institute, and Mr. Erik Moore, University of Nebraska Medical Center, are acknowledged for their editorial assistances.

6. REFERENCES

1. Shimizu Y. & S. Shaw: Cell adhesion. Mucins in the mainstream. *Nature* 366, 630-631 (1993)

2. Bobek L. A., J. Liu, S. N. Sait, T. B. Shows, Y. A. Bobek & M. J. Levine: Structure and chromosomal localization of the human salivary mucin gene, MUC7. *Genomics* 31, 277-282 (1996)

3. Nollet S., N. Moniaux, J. Maury, D. Petitprez, P. Degand, A. Laine, N. Porchet & J. P. Aubert: Human mucin gene MUC4: organization of its 5'-region and polymorphism of its central tandem repeat array. *Biochem J* 332, 739-748 (1998)

4. Gross M. S., V. Guyonnet-Duperat, N. Porchet, A. Bernheim, J. P. Aubert & V. C. Nguyen: Mucin 4 (MUC4) gene: regional assignment (3q29) and RFLP analysis. *Ann Genet* 35, 21-26 (1992)

5. Pigny P., W. S. Pratt, A. Laine, A. Leclercq, D. M. Swallow, V. C. Nguyen, J. P. Aubert & N. Porchet: The MUC5AC gene: RFLP analysis with the Jer58 probe. *Hum Genet* 96, 367-368 (1995)

6. Toribara N. W., J. R. J. Gum, P. J. Culhane, R. E. Lagace, J. W. Hicks, G. M. Petersen & Y. S. Kim: MUC-2 human small intestinal mucin gene structure. Repeated arrays and polymorphism. *J Clin Invest* 88, 1005-1013 (1991)

7. Vinall L. E., A. S. Hill, P. Pigny, W. S. Pratt, N. Toribara, J. R. Gum, Y. S. Kim, N. Porchet, J. P. Aubert & D. M. Swallow: Variable number tandem repeat polymorphism of the mucin genes located in the complex on 11p15.5. *Hum Genet* 102, 357-366 (1998)

8. Swallow D. M., S. Gendler, B. Griffiths, A. Kearney, S. Povey, D. Sheer, R. W. Palmer & J. Taylor-Papadimitriou: The hypervariable gene locus PUM, which codes for the tumour associated epithelial mucins, is located on chromosome 1, within the region 1q21-24. *Ann Hum Genet* 51, 289-294 (1987)

9. Debailleul V., A. Laine, G. Huet, P. Mathon, M. C. d'Hooghe, J. P. Aubert & N. Porchet: Human mucin genes MUC2, MUC3, MUC4, MUC5AC, MUC5B, and MUC6 express stable and extremely large mRNAs and exhibit a variable length polymorphism. An improved method to analyze large mRNAs. *J Biol Chem* 273, 881-890 (1998)

10. Gum J. R. J.Mucin genes and the proteins they encode: structure, diversity, and regulation. *Am J Respir Cell Mol Biol* 7, 557-564 (1992)

11. Crepin M., N. Porchet, J. P. Aubert & P. Degand: Diversity of the peptide moiety of human airway mucins. *Biorheology* 27, 471-484 (1990)

12. Dufosse J., N. Porchet, J. P. Audie, D. Guyonnet, V, A. Laine, I. Van-Seuningen, S. Marrakchi, P. Degand & J. P. Aubert: Degenerate 87-base-pair tandem repeats create hydrophilic/hydrophobic alternating domains in human mucin peptides mapped to 11p15. *Biochem J* 293, 329-337 (1993)

13. Gum J. R., J. C. Byrd, J. W. Hicks, N. W. Toribara, D. T. Lamport & Y. S. Kim: Molecular cloning of human intestinal mucin cDNAs. Sequence analysis and evidence for genetic polymorphism. *J Biol C*hem 264, 6480-6487 (1989)

14. Gum J. R., J. W. Hicks, D. M. Swallow, R. L. Lagace, J. C. Byrd, D. T. Lamport, B. Siddiki & Y. S. Kim: Molecular cloning of cDNAs derived from a novel human intestinal mucin gene. *Biochem Biophys Res Commun* 171, 407-415 (1990)

15. Porchet N., J. Dufosse, J. P. Audie, V. G. Duperat, J. M. Perini, V. C. Nguyen, P. Degand & J. P. Aubert: Structural features of the core proteins of human airway mucins ascertained by cDNA cloning. *Am Rev Respir D*is 144, S15-S18 (1991)

16. Toribara N. W., A. M. Roberton, S. B. Ho, W. L. Kuo, E. Gum, J. W. Hicks, J. R. J. Gum, J. C. Byrd, B. Siddiki & Y. S. Kim: Human gastric mucin. Identification of a unique species by expression cloning. *J Biol Chem* 268, 5879-5885 (1993)

17. Aubert J. P., N. Porchet, M. Crepin, M. Duterque-Coquillaud, G. Vergnes, M. Mazzuca, B. Debuire, D. Petitprez & P. Degand: Evidence for different human tracheobronchial mucin peptides deduced from nucleotide cDNA sequences. *Am J Respir Cell Mol Biol* 5, 178-185 (1991)

18. Gum J. R. J., J. W. Hicks, N. W. Toribara., B. Siddiki & Y. S. Kim: Molecular cloning of human intestinal mucin (MUC2) cDNA. Identification of the amino terminus and overall sequence similarity to prepro-von Willebrand factor. *J Biol C*hem 269, 2440-2446 (1994)

19. Lan M. S., S. K. Batra, W. N. Qi, R. S. Metzgar & M. A. Hollingsworth: Cloning and sequencing of a human pancreatic tumor mucin cDNA. *J Biol Chem* 265, 15294-15299 (1990)

20. Lapensee L., Y. Paquette & G. Bleau: Allelic polymorphism and chromosomal localization of the human oviductin gene (MUC9). *Fertil Steril* 68, 702-708 (1997)

21. Porchet N., V. C. Nguyen, J. Dufosse, J. P. Audie, V. Guyonnet-Duperat, M. S. Gross, C. Denis, P. Degand, A. Bernheim & J. P. Aubert: Molecular cloning and chromosomal localization of a novel human tracheo-bronchial mucin cDNA containing tandemly repeated sequences of 48 base pairs. *Biochem Biophys Res Commun* 175, 414-422 (1991)

22. Shankar V., P. Pichan, R. L. J. Eddy, V. Tonk, N. Nowak, S. N. Sait, T. B. Shows, R. E. Schultz, G. Gotway, R. C. Elkins, M. S. Gilmore & G. P. Sachdev: Chromosomal localization of a human mucin gene (MUC8) and cloning of the cDNA corresponding to the carboxy terminus. *Am J Respir Cell Mol Biol* 16, 232-241 (1997)

23. Williams S. J., M. A. McGuckin, D. C. Gotley, H. J. Eyre, G. R. Sutherland & T. M. Antalis: Two novel mucin genes down-regulated in colorectal cancer identified by differential display. *Cancer Res* 59, 4083-4089 (1999)

24. Pigny P., V. Guyonnet-Duperat, A. S. Hill, W. S. Pratt, S. Galiegue-Zouitina, M. C. d'Hooge, A. Laine, I. Van-Seuningen, P. Degand, J. R. Gum, Y. S. Kim, D. M. Swallow, J. P. Aubert & N. Porchet: Human mucin genes assigned to 11p15.5: identification and organization of a cluster of genes. *Genomics* 38, 340-352 (1996)

25. Mayadas T. N. & D. D. Wagner: von Willebrand factor biosynthesis and processing. *Ann N Y Acad Sci* 614, 153-166 (1991)

26. Mayadas T. N. & D. D. Wagner: Vicinal cysteines in the prosequence play a role in von Willebrand factor multimer assembly. *Proc Natl Acad Sci U S A* 89, 3531-3535 (1992)

27. Desseyn J. L., M. P. Buisine, N. Porchet, J. P. Aubert, P. Degand & A. Laine: Evolutionary history of the 11p15 human mucin gene family. *J Mol Evol* 46, 102-106 (1998)

28. Desseyn J. L., J. P. Aubert, N. Porchet & A. Laine: Evolution of the large secreted gel-forming mucins. *Mol Biol Evol* 17, 1175-1184 (2000)

29. Gum J. R. J., J. W. Hicks, N. W. Toribara, E. M. Rothe, R. E. Lagace & Y. S. Kim: The human MUC2 intestinal mucin has cysteine-rich subdomains located both upstream and downstream of its central repetitive region. J Biol Chem 267, 21375-21383 (1992)

30. Byrd J. C., J. Nardelli, B. Siddiqui & Y. S. Kim: Isolation and characterization of colon cancer mucin from xenografts of LS174T cells. *Cancer Res* 48, 6678-6685 (1988)

31. Sun P. D. & D. R. Davies: The cystine-knot growth-factor superfamily. *Annu Rev Biophys Biomol Struct* 24, 269-291 (1995)

32. Meitinger T., A. Meindl, P. Bork, B. Rost, C. Sander, M. Haasemann & J. Murken: Molecular modelling of the Norrie disease protein predicts a cystine knot growth factor tertiary structure. *Nat Genet* 5, 376-380 (1993)

33. Kovarik A., N. Peat, D. Wilson, S. J. Gendler & J. Taylor-Papadimitriou: Analysis of the tissue-specific promoter of the MUC1 gene. *J Biol Chem* 268, 9917-9926 (1993)

34. Gum J. R., J. W. Hicks & Y. S. Kim: Identification and characterization of the MUC2 (human intestinal mucin) gene 5'-flanking region: promoter activity in cultured cells. *Biochem J* 325, 259-267 (1997)

35. Kovarik A., P. J. Lu, N. Peat, J. Morris & J. Taylor-Papadimitriou: Two GC boxes (Sp1 sites) are involved in regulation of the activity of the epithelium-specific MUC1 promoter. *J Biol Chem* 271, 18140-18147 (1996)

36. Li J. D., A. F. Dohrman, M. Gallup, S. Miyata, J. R. Gum, Y. S. Kim, J. A. Nadel, A. Prince & C. B. Basbaum: Transcriptional activation of mucin by Pseudomonas aeruginosa lipopolysaccharide in the pathogenesis of cystic fibrosis lung disease. *Proc Natl Acad Sci U S A* 94, 967-972 (1997)

37. Gum J. R. J., J. W. Hicks, A. M. Gillespie, E. J. Carlson, L. Komuves, S. Karnik, J. C. Hong, C. J. Epstein & Y. S. Kim: Goblet cell-specific expression mediated by the MUC2 mucin gene promoter in the intestine of transgenic mice. *Am J Physiol* 276, G666-G676 (1999)

38. Troxler R. F., G. D. Offner, F. Zhang, I. Iontcheva & F. G. Oppenheim: Molecular cloning of a novel high molecular weight mucin (MG1) from human sublingual gland. *Biochem Biophys Res Com*mun 217, 1112-1119 (1995)

39. Offner G. D., D. P. Nunes, A. C. Keates, N. H. Afdhal & R. F. Troxler: The amino-terminal sequence of MUC5B contains conserved multifunctional D domains: implications for tissue-specific mucin functions. *Biochem Biophys Res Commun* 251, 350-355 (1998)

40. Keates A. C., D. P. Nunes, N. H. Afdhal, R. F. Troxler & G. D. Offner: Molecular cloning of a major human gall bladder mucin: complete C-terminal sequence and genomic organization of MUC5B. *Biochem J* 324, 295-303 (1997)

41. Desseyn J. L., V. Guyonnet-Duperat, N. Porchet, J. P. Aubert & A. Laine: Human mucin gene MUC5B, the 10.7-kb large central exon encodes various alternate subdomains resulting in a super-repeat. Structural evidence for a 11p15.5 gene family. *J Biol Chem* 272, 3168-3178 (1997)

42. Desseyn J. L., J. P. Aubert, S. Van, I, N. Porchet & A. Laine: Genomic organization of the 3' region of the human mucin gene MUC5B. *J Biol Chem* 272, 16873-16883 (1997)

43. Desseyn J. L., M. P. Buisine, N. Porchet, J. P. Aubert & A. Laine: Genomic organization of the human mucin gene

MUC5B. cDNA and genomic sequences upstream of the large central exon. *J Biol Chem* 273, 30157-30164 (1998)

44. Van-Seuningen I., M. Perrais, P. Pigny, N. Porchet & J. P. Aubert: Sequence of the 5'-flanking region and promoter activity of the human mucin gene MUC5B in differant phenotypes of colon cancer cells. *Biochem J* 348, 675-686 (2000)

45. Pigny P., S. Van, I, J. L. Desseyn, S. Nollet, N. Porchet, A. Laine & J. P. Aubert: Identification of a 42-kDa nuclear factor (NF1-MUC5B) from HT-29 MTX cells that binds to the 3' region of human mucin gene MUC5B. *Biochem Biophys Res Commun* 220, 186-191 (1996)

46. Guyonnet D., V, J. P. Audie, V. Debailleul, A. Laine, M. P. Buisine, S. Galiegue-Zouitina, P. Pigny, P. Degand, J. P. Aubert & N. Porchet: Characterization of the human mucin gene MUC5AC: a consensus cysteine-rich domain for 11p15 mucin genes? *Biochem J* 305, 211-219 (1995)

47. Meezaman D., P. Charles, E. Daskal, M. H. Polymeropoulos, B. M. Martin & M. C. Rose: Cloning and analysis of cDNA encoding a major airway glycoprotein, human tracheobronchial mucin (MUC5). *J Biol Chem* 269, 12932-12939 (1994)

48. Lesuffleur T., F. Roche, A. S. Hill, M. Lacasa, M. Fox, D. M. Swallow, A. Zweibaum & F. X. Real: Characterization of a mucin cDNA clone isolated from HT-29 mucus-secreting cells. The 3' end of MUC5AC*?* *J Biol Chem* 270, 13665-13673 (1995)

49. Buisine M. P., J. L. Desseyn, N. Porchet, P. Degand, A. Laine & J. P. Aubert: Genomic organization of the 3'-region of the human MUC5AC mucin gene: additional evidence for a common ancestral gene for the 11p15.5 mucin gene family. *Biochem J* 332, 729-738 (1998)

50. Klomp L. W., L. Van Rens & G. J. Strous: Cloning and analysis of human gastric mucin cDNA reveals two types of conserved cysteine-rich domains. *Biochem J* 308, 831-838 (1995)

51. van de Bovenkamp J. H., C. M. Hau, G. J. Strous, H. A. Buller, J. Dekker & A. W. Einerhand: Molecular cloning of human gastric mucin MUC5AC reveals conserved cysteine-rich D-domains and a putative leucine zipper motif. *Biochem Biophys Res Commun* 245, 853-859 (1998)

52. Li D., M. Gallup, N. Fan, D. E. Szymkowski & C. B. Basbaum: Cloning of the amino-terminal and 5'-flanking region of the human MUC5AC mucin gene and transcriptional up-regulation by bacterial exoproducts. *J Biol Chem* 273, 6812-6820 (1998)

53. Escande, F., Aubert, J. P., Porchet, N., and Buisine, M. P. Human mucin gene MUC5AC: organization of its 5'- and central repetitive regions. *Biochem J* (in press) (2001)

54. Li J. D., W. Feng, M. Gallup, J. H. Kim, J. Gum, Y. Kim & C. Basbaum: Activation of NF-kappaB via a Src-dependent Ras-MAPK-pp90rsk pathway is required for Pseudomonas aeruginosa-induced mucin overproduction in epithelial cells. *Proc Natl Acad Sci U S A* 95, 5718-5723 (1998)

55. Toribara N. W., S. B. Ho, E. Gum, J. R. J. Gum, P. Lau & Y. S. Kim: The carboxyl-terminal sequence of the human secretory mucin, MUC6. Analysis Of the primary amino acid sequence. *J Biol Chem* 272, 16398-16403 (1997)

56. Dong Z., R. S. Thoma, D. L. Crimmins, D. W. McCourt, E. A. Tuley & J. E. Sadler: Disulfide bonds required to assemble functional von Willebrand factor multimers. *J Biol Chem* 269, 6753-6758 (1994)

57. Voorberg J., R. Fontijn, J. Calafat, H. Janssen, J. A. van Mourik & H. Pannekoek: Assembly and routing of von Willebrand factor variants: the requirements for disulfide-linked dimerization reside within the carboxy-terminal 151 amino acids. *J Cell Biol* 113, 195-205 (1991)

58. Perez-Vilar J. & R. L. Hill: The carboxyl-terminal 90 residues of porcine submaxillary mucin are sufficient for forming disulfide-bonded dimers. *J Biol Chem* 273, 6982-6988 (1998)

59. Van Klinken B. J., A. W. Einerhand, H. A. Buller & J. Dekker: The oligomerization of a family of four genetically clustered human gastrointestinal mucins. *Glycobiology* 8, 67-75 (1998)

60. Eckhardt A. E., C. S. Timpte, J. L. Abernethy, A. Toumadje, W. C. J. Johnson & R. L. Hill: Structural properties of porcine submaxillary gland apomucin. *J Biol Chem* 262, 11339-11344 (1987)

61. Eckhardt A. E., C. S. Timpte, J. L. Abernethy, Y. Zhao & R. L. Hill: Porcine submaxillary mucin contains a cystine-rich, carboxyl-terminal domain in addition to a highly repetitive, glycosylated domain. *J Biol Chem* 266, 9678-9686 (1991)

62. Eckhardt A. E., C. S. Timpte, A. W. DeLuca & R. L. Hill: The complete cDNA sequence and structural polymorphism of the polypeptide chain of porcine submaxillary mucin. *J Biol Chem* 272, 33204-33210 (1997)

63. Jiang W., J. T. Woitach, D. Gupta & V. P. Bhavanandan: Sequence of a second gene encoding bovine submaxillary mucin: implication for mucin heterogeneity and cloning. *Biochem Biophys Res Commun* 251, 550-556 (1998)

64. Jiang W., J. T. Woitach, R. L. Keil & V. P. Bhavanandan: Bovine submaxillary mucin contains multiple domains and tandemly repeated non-identical sequences. *Biochem J* 331, 193-199 (1998)

65. Jiang W., D. Gupta, D. Gallagher, S. Davis & V. P. Bhavanandan: The central domain of bovine submaxillary

mucin consists of over 50 tandem repeats of 329 amino acids. Chromosomal localization of the BSM1 gene and relations to ovine and porcine counterparts. *Eur J Biochem* 267, 2208-2217 (2000)

66. Hoffmann W. & F. Hauser: Biosynthesis of frog skin mucins: cysteine-rich shuffled modules, polydispersities and genetic polymorphism. *Comp Biochem Physiol* [B] 105, 465-472 (1993)

67. Hoffmann W. & W. Joba: Biosynthesis and molecular architecture of gel-forming mucins: implications from an amphibian model system. *Biochem Soc Trans* 23, 805-810 (1995)

68. Perez-Vilar J., A. E. Eckhardt, A. DeLuca & R. L. Hill: Porcine submaxillary mucin forms disulfide-linked multimers through its amino-terminal D-domains. *J Biol Chem* 273, 14442-14449 (1998)

69. Perez-Vilar J. & R. L. Hill: Identification of the half-cystine residues in porcine submaxillary mucin critical for multimerization through the D-domains. Roles of the CGLCG motif in the D1- and D3-domains. *J Biol Chem* 273, 34527-34534 (1998)

70. Perez-Vilar J. & R. L. Hill: The structure and assembly of secreted mucins. *J Biol Chem* 274, 31751-31754 (1999)

71. Reddy M. S., L. A. Bobek, G. G. Haraszthy, A. R. Biesbrock & M. J. Levine: Structural features of the low-molecular-mass human salivary mucin. *Biochem J* 287, 639-643 (1992)

72. Cohen R. E., A. Aguirre, M. E. Neiders, M. J. Levine, P. C. Jones, M. S. Reddy & J. G. Haar: Immunochemistry and immunogenicity of low molecular weight human salivary mucin. *Arch Oral Biol* 36, 347-356 (1991)

73. Bobek L. A., H. Tsai, A. R. Biesbrock & M. J. Levine: Molecular cloning, sequence, and specificity of expression of the gene encoding the low molecular weight human salivary mucin (MUC7). *J Biol Chem* 268, 20563-20569 (1993)

74. Gururaja T. L., N. Ramasubbu, P. Venugopalan, M. S. Reddy, K. Ramalingam & M. J. Levine: Structural features of the human salivary mucin, MUC7. *Glycoconj J* 15, 457-467 (1998)

75. Levine M. J., M. C. Herzberg, M. S. Levine, S. A. Ellison, M. W. Stinson, H. C. Li & T. van Dyke: Specificity of salivary-bacterial interactions: role of terminal sialic acid residues in the interaction of salivary glycoproteins with Streptococcus sanguis and Streptococcus mutans. *Infect Immun* 19, 107-115 (1978)

76. Murray P. A., A. Prakobphol, T. Lee, C. I. Hoover & S. J. Fisher: Adherence of oral streptococci to salivary glycoproteins. *Infect Immun* 60, 31-38 (1992)

77. Liu B., S. Rayment, F. G. Oppenheim & R. F. Troxler: Isolation of human salivary mucin MG2 by a novel method and characterization of its interactions with oral bacteria. *Arch Biochem Biophys* 364, 286-293 (1999)

78. Biesbrock A. R., M. S. Reddy & M. J. Levine: Interaction of a salivary mucin-secretory immunoglobulin A complex with mucosal pathogens. *Infect Immun* 59, 3492-3497 (1991)

79. Groenink J., A. J. Ligtenberg, E. C. Veerman, J. G. Bolscher & A. A. Nieuw: Interaction of the salivary low-molecular-weight mucin (MG2) with Actinobacillus actinomycetemcomitans. *Antonie Van Leeuwenhoek* 70, 79-87 (1996)

80. Ebisu S., H. Fukuhara & H. Okada: Purification and characterization of Eikenella corrodens aggregating factor from submandibular-sublingual saliva. *J Periodontal Res* 23, 328-333 (1988)

81. Bergey E. J., M. I. Cho, M. L. Hammarskjold, D. Rekosh, M. J. Levine, B. M. Blumberg & L. G. Epstein: Aggregation of human immunodeficiency virus type 1 by human salivary secretions. *Crit Rev Oral Biol Med* 4, 467-474 (1993)

82. Bergey E. J., M. I. Cho, B. M. Blumberg, M. L. Hammarskjold, D. Rekosh, L. G. Epstein & M. J. Levine: Interaction of HIV-1 and human salivary mucins. *J Acquir Immune Defic Syndr* 7, 995-1002 (1994)

83. Nagashunmugam T., D. Malamud, C. Davis, W. R. Abrams & H. M. Friedman: Human submandibular saliva inhibits human immunodeficiency virus type 1 infection by displacing envelope glycoprotein gp120 from the virus. *J Infect Dis* 178, 1635-1641 (1998)

84. Liu B., S. A. Rayment, C. Gyurko, F. G. Oppenheim, G. D. Offner & R. F. Troxler: The recombinant N-terminal region of human salivary mucin MG2 (MUC7) contains a binding domain for oral Streptococci and exhibits candidacidal activity. Biochem J 345 Pt 3, 557-564 (2000)

85. Pollock J. J., L. Denepitiya, B. J. MacKay & V. J. Iacono: Fungistatic and fungicidal activity of human parotid salivary histidine-rich polypeptides on Candida albicans. *Infect Immun* 44, 702-707 (1984)

86. Oppenheim F. G., T. Xu, F. M. McMillian, S. M. Levitz, R. D. Diamond, G. D. Offner & R. F. Troxler: Histatins, a novel family of histidine-rich proteins in human parotid secretion. Isolation, characterization, primary structure, and fungistatic effects on Candida albicans. *J Biol Chem* 263, 7472-7477 (1988)

87. Troxler R. F., G. D. Offner, T. Xu, J. C. Vanderspek & F. G. Oppenheim: Structural relationship between human salivary histatins. *J Dent Res* 69, 2-6 (1990)

88. Xu T., S. M. Levitz, R. D. Diamond & F. G. Oppenheim: Anticandidal activity of major human salivary histatins. *Infect Immun* 59, 2549-2554 (1991)

89. Tsai H. & L. A. Bobek: Human salivary histatins: promising anti-fungal therapeutic agents. *Crit Rev Oral Biol Med* 9, 480-497 (1998)

90. Gendler S. J., C. A. Lancaster, J. Taylor-Papadimitriou, T. Duhig, N. Peat, J. Burchell, L. Pemberton, E. N. Lalani & D. Wilson: Molecular cloning and expression of human tumor-associated polymorphic epithelial mucin. *J Biol Chem* 265, 15286-15293 (1990)

91. Pratt W. S., S. Crawley, J. Hicks, J. Ho, M. Nash, Y. S. Kim, J. R. Gum & D. M. Swallow: Multiple transcripts of MUC3: evidence for two genes, MUC3A and MUC3B. *Biochem Biophys Res Commun* 275, 916-923 (2000)

92. Moniaux N., S. Nollet, N. Porchet, P. Degand, A. Laine & J. P. Aubert: Complete sequence of the human mucin MUC4: a putative cell membrane-associated mucin. *Biochem J* 338, 325-333 (1999)

93. Gendler S. J., J. M. Burchell, T. Duhig, D. Lamport, R. White, M. Parker & J. Taylor-Papadimitriou: Cloning of partial cDNA encoding differentiation and tumor-associated mucin glycoproteins expressed by human mammary epithelium. *Proc Natl Acad Sci U S A* 84, 6060-6064 (1987)

94. Choudhury A., N. Moniaux, J. P. Winpenny, M. A. Hollingsworth, J. P. Aubert & S. K. Batra: Human MUC4 mucin cDNA and its variants in pancreatic carcinoma. *J Biochem* 128, 233-243 (2000)

95. Moniaux N., F. Escande, S. K. Batra, N. Porchet, A. Laine & J. P. Aubert: Alternative splicing generates a family of putative secreted and membrane-associated MUC4 mucins. *Eur J Biochem* 267, 4536-4544 (2000)

96. Williams S. J., D. J. Munster, R. J. Quin, D. C. Gotley & M. A. McGuckin: The MUC3 gene encodes a transmembrane mucin and is alternatively spliced. *Biochem Biophys Res Commun* 261, 83-89 (1999)

97. Wreschner D. H., M. Hareuveni, I. Tsarfaty, N. Smorodinsky, J. Horev, J. Zaretsky, P. Kotkes, M. Weiss, R. Lathe & A. Dion: Human epithelial tumor antigen cDNA sequences. Differential splicing may generate multiple protein forms. *Eur J Biochem* 189, 463-473 (1990)

98. Pimental R. A., J. Julian, S. J. Gendler & D. D. Carson: Synthesis and intracellular trafficking of Muc-1 and mucins by polarized mouse uterine epithelial cells. *J Biol Chem* 271, 28128-28137 (1996)

99. Litvinov S. V. & J. Hilkens: The epithelial sialomucin, episialin, is sialylated during recycling. *J Biol Chem* 268, 21364-21371 (1993)

100. Gendler S. J. & A. P. Spicer: Epithelial mucin genes. *Annu Rev Physiol* 57, 607-634 (1995)

101. Stern L., M. Palatsides, T. de Kretser & M. Ford: Expression of the tumor-associated mucin MUC1 in an ovarian tumor cell line. *Int J Cancer* 50, 783-790 (1992)

102. Shimizu M. & K. Yamauchi: Isolation and characterization of mucin-like glycoprotein in human milk fat globule membrane. *J Biochem* 91, 515-524 (1982)

103. Lan M. S., R. C. J. Bast, M. I. Colnaghi, R. C. Knapp, D. Colcher, J. Schlom & R. S. Metzgar: Co-expression of human cancer-associated epitopes on mucin molecules. *Int J Cancer* 39, 68-72 (1987)

104. Gendler S., J. Taylor-Papadimitriou, T. Duhig, J. Rothbard & J. Burchell: A highly immunogenic region of a human polymorphic epithelial mucin expressed by carcinomas is made up of tandem repeats. *J Biol Chem* 263, 12820-12823 (1988)

105. Fontenot J. D., N. Tjandra, D. Bu, C. Ho, R. C. Montelaro & O. J. Finn: Biophysical characterization of one-, two-, and three-tandem repeats of human mucin (muc-1) protein core. *Cancer Res* 53, 5386-5394 (1993)

106. Jentoft N.Why are proteins O-glycosylated*? Trends Biochem Sci* 15, 291-294 (1990)

107. Ligtenberg M. J., F. Buijs, H. L. Vos & J. Hilkens: Suppression of cellular aggregation by high levels of episialin. *Cancer Res* 52, 2318-2324 (1992)

108. Braga V. M., L. F. Pemberton, T. Duhig & S. J. Gendler: Spatial and temporal expression of an epithelial mucin, Muc-1, during mouse development. *Development* 115, 427-437 (1992)

109. Chambers J. A., M. A. Hollingsworth, A. E. Trezise & A. Harris: Developmental expression of mucin genes MUC1 and MUC2. *J Cell Sci* 107 (Pt 2), 413-424 (1994)

110. Reis C. A., L. David, M. Seixas, J. Burchell & M. Sobrinho-Simoes: Expression of fully and under-glycosylated forms of MUC1 mucin in gastric carcinoma. *Int J Cancer* 79, 402-410 (1998)

111. Majuri M. L., P. Mattila & R. Renkonen: Recombinant E-selectin-protein mediates tumor cell adhesion via sialyl-Le(a) and sialyl-Le(x*). Biochem Biophys Res Commun* 182, 1376-1382 (1992)

112. Regimbald L. H., L. M. Pilarski, B. M. Longenecker, M. A. Reddish, G. Zimmermann & J. C. Hugh: The breast mucin MUCI as a novel adhesion ligand for endothelial intercellular adhesion molecule 1 in breast cancer. *Cancer Res* 56, 4244-4249 (1996)

113. Yamamoto M., A. Bharti, Y. Li & D. Kufe: Interaction of the DF3/MUC1 breast carcinoma-associated

antigen and beta-catenin in cell adhesion. *J Biol Chem* 272, 12492-12494 (1997)

114. Hulsken J., W. Birchmeier & J. Behrens: E-cadherin and APC compete for the interaction with beta-catenin and the cytoskeleton. *J Cell Biol* 127, 2061-2069 (1994)

115. Hulsken J., J. Behrens & W. Birchmeier: Tumor-suppressor gene products in cell contacts: the cadherin-APC-armadillo connection. *Curr Opin Cell Biol* 6, 711-716 (1994)

116. Munemitsu S., I. Albert, B. Souza, B. Rubinfeld & P. Polakis: Regulation of intracellular beta-catenin levels by the adenomatous polyposis coli (APC) tumor-suppressor protein. *Proc Natl Acad Sci U S A* 92, 3046-3050 (1995)

117. Peifer M.Regulating cell proliferation: as easy as APC. *Science* 272, 974-975 (1996)

118. Bhat R. V., J. M. Baraban, R. C. Johnson, B. A. Eipper & R. E. Mains: High levels of expression of the tumor suppressor gene APC during development of the rat central nervous system. *J Neurosci* 14, 3059-3071 (1994)

119. Hinck L., I. S. Nathke, J. Papkoff & W. J. Nelson: Beta-catenin: a common target for the regulation of cell adhesion by Wnt-1 and Src signaling pathways. *Trends Biochem Sci* 19, 538-542 (1994)

120. Hinck L., W. J. Nelson & J. Papkoff: Wnt-1 modulates cell-cell adhesion in mammalian cells by stabilizing beta-catenin binding to the cell adhesion protein cadherin. *J Cell Biol* 124, 729-741 (1994)

121. Rubinfeld B., B. Souza, I. Albert, S. Munemitsu & P. Polakis: The APC protein and E-cadherin form similar but independent complexes with alpha-catenin, beta-catenin, and plakoglobin. *J Biol Chem* 270, 5549-5555 (1995)

122. Peifer M.Cancer, catenins, and cuticle pattern: a complex connection. *Science* 262, 1667-1668 (1993)

123. Burchill S. A.The tumour suppressor APC gene product is associated with cell adhesion. Bioessays 16, 225-227 (1994)

124. Rubinfeld B., I. Albert, E. Porfiri, C. Fiol, S. Munemitsu & P. Polakis: Binding of GSK3beta to the APC-beta-catenin complex and regulation of complex assembly. *Science* 272, 1023-1026 (1996)

125. Li Y., A. Bharti, D. Chen, J. Gong & D. Kufe: Interaction of glycogen synthase kinase 3beta with the DF3/MUC1 carcinoma-associated antigen and beta-catenin. *Mol Cell Biol* 18, 7216-7224 (1998)

126. Quin R. J. & M. A. McGuckin: Phosphorylation of the cytoplasmic domain of the MUC1 mucin correlates with changes in cell-cell adhesion. *Int J Cancer* 87, 499-506 (2000)

127. Baruch A., M. Hartmann, M. Yoeli, Y. Adereth, S. Greenstein, Y. Stadler, Y. Skornik, J. Zaretsky, N. I. Smorodinsky, I. Keydar & D. H. Wreschner: The breast cancer-associated MUC1 gene generates both a receptor and its cognate binding protein. *Cancer Res* 59, 1552-1561 (1999)

128. Zrihan-Licht S., A. Baruch, O. Elroy-Stein, I. Keydar & D. H. Wreschner: Tyrosine phosphorylation of the MUC1 breast cancer membrane proteins. Cytokine receptor-like molecules. *FEBS L*ett 356, 130-136 (1994)

129. Ligtenberg M. J., H. L. Vos, A. M. Gennissen & J. Hilkens: Episialin, a carcinoma-associated mucin, is generated by a polymorphic gene encoding splice variants with alternative amino termini. *J Biol Chem* 265, 5573-5578 (1990)

130. Zrihan-Licht S., H. L. Vos, A. Baruch, O. Elroy-Stein, D. Sagiv, I. Keydar, J. Hilkens & D. H. Wreschner: Characterization and molecular cloning of a novel MUC1 protein, devoid of tandem repeats, expressed in human breast cancer tissue. *Eur J Biochem* 224, 787-795 (1994)

131. Smorodinsky N., M. Weiss, M. L. Hartmann, A. Baruch, E. Harness, M. Yaakobovitz, I. Keydar & D. H. Wreschner: Detection of a secreted MUC1/SEC protein by MUC1 isoform specific monoclonal antibodies. *Biochem Biophys Res Commun* 228, 115-121 (1996)

132. Hartman M., A. Baruch, I. Ron, Y. Aderet, M. Yoeli, O. Sagi-Assif, S. Greenstein, Y. Stadler, M. Weiss, E. Harness, M. Yaakubovits, I. Keydar, N. I. Smorodinsky & D. H. Wreschner: MUC1 isoform specific monoclonal antibody 6E6/2 detects preferential expression of the novel MUC1/Y protein in breast and ovarian cancer. *Int J Cancer* 82, 256-267 (1999)

133. Baruch A., M. Hartmann, S. Zrihan-Licht, S. Greenstein, M. Burstein, I. Keydar, M. Weiss, N. Smorodinsky & D. H. Wreschner: Preferential expression of novel MUC1 tumor antigen isoforms in human epithelial tumors and their tumor-potentiating function. *Int J Cancer* 71, 741-749 (1997)

134. Oosterkamp H. M., L. Scheiner, M. C. Stefanova, K. O. Lloyd & C. L. Finstad: Comparison of MUC-1 mucin expression in epithelial and non-epithelial cancer cell lines and demonstration of a new short variant form (MUC-1/Z). *Int J Cancer* 72, 87-94 (1997)

135. Patton S., S. J. Gendler & A. P. Spicer: The epithelial mucin, MUC1, of milk, mammary gland and other tissues. *Biochim Biophys Acta* 1241, 407-423 (1995)

136. Abe M. & D. Kufe: Characterization of cis-acting elements regulating transcription of the human DF3 breast carcinoma-associated antigen (MUC1) gene. *Proc Natl Acad Sci U S A* 90, 282-286 (1993)

137. Zaretsky J. Z., R. Sarid, Y. Aylon, L. A. Mittelman, D.

H. Wreschner & I. Keydar: Analysis of the promoter of the MUC1 gene overexpressed in breast cancer. *FEBS Lett* 461, 189-195 (1999)

138. Weis L.& D. Reinberg: Accurate positioning of RNA polymerase II on a natural TATA-less promoter is independent of TATA-binding-protein-associated factors and initiator-binding proteins. *Mol Cell Biol* 17, 2973-2984 (1997)

139. Gum J. R. J., J. J. L. Ho, W. S. Pratt, J. W. Hicks, A. S. Hill, L. E. Vinall, A. M. Roberton, D. M. Swallow & Y. S. Kim: MUC3 human intestinal mucin. Analysis of gene structure, the carboxyl terminus, and a novel upstream repetitive region. *J Biol Chem* 272, 26678-26686 (1997)

140. Van Klinken B. J., T. C. Van Dijken, E. Oussoren, H. A. Buller, J. Dekker & A. W. Einerhand: Molecular cloning of human MUC3 cDNA reveals a novel 59 amino acid tandem repeat region. *Biochem Biophys Res Commun* 238, 143-148 (1997)

141. Crawley S. C., J. R. J. Gum, J. W. Hicks, W. S. Pratt, J. P. Aubert, D. M. Swallow & Y. S. Kim: Genomic organization and structure of the 3' region of human MUC3: alternative splicing predicts membrane-bound and soluble forms of the mucin. *Biochem Biophys Res Commun* 263, 728-736 (1999)

142. Khatri I. A., G. G. Forstner & J. F. Forstner: The carboxyl-terminal sequence of rat intestinal mucin RMuc3 contains a putative transmembrane region and two EGF-like motifs. *Biochim Biophys Acta* 1326, 7-11 (1997)

143. Shekels L. L., D. A. Hunninghake, A. S. Tisdale, I. K. Gipson, M. Kieliszewski, C. A. Kozak & S. B. Ho: Cloning and characterization of mouse intestinal MUC3 mucin: 3' sequence contains epidermal-growth-factor-like domains. *Biochem J* 330 (Pt 3), 1301-1308 (1998)

144. Choudhury A., N. Moniaux, J. Ringel, J. King, E. Moore, J. P. Aubert, and & S. K. Batra: Alternate splicing at the 3'-end of the human pancreatic tumor-associated mucin MUC4 cDNA. *Teratogenesis, Carcinogenesis, and Mutagenesis* 21, 83-96 (2001)

145. Sheng Z., K. Wu, K. L. Carraway & N. Fregien: Molecular cloning of the transmembrane component of the 13762 mammary adenocarcinoma sialomucin complex. A new member of the epidermal growth factor superfamily. *J Biol Chem* 267, 16341-16346 (1992)

146. Wu K., N. Fregien & K. L. Carraway: Molecular cloning and sequencing of the mucin subunit of a heterodimeric, bifunctional cell surface glycoprotein complex of ascites rat mammary adenocarcinoma cells. *J Biol Chem* 269, 11950-11955 (1994)

147. Rossi E. A., R. R. McNeer, S. A. Price-Schiavi, J. M. Van den Brande, M. Komatsu, J. F. Thompson, C. A. Carraway, N. L. Fregien & K. L. Carraway: Sialomucin complex, a heterodimeric glycoprotein complex. Expression as a soluble, secretable form in lactating mammary gland and colon. *J Biol Chem* 271, 33476-33485 (1996)

148. Carraway K. L., N. Fregien & C. A. Carraway: Tumor sialomucin complexes as tumor antigens and modulators of cellular interactions and proliferation. *J Cell Sci* 103, 299-307 (1992)

149. Komatsu M., C. A. Carraway, N. L. Fregien & K. L. Carraway: Reversible disruption of cell-matrix and cell-cell interactions by overexpression of sialomucin complex. *J Biol Chem* 272, 33245-33254 (1997)

150. Steck P. A., S. M. North & G. L. Nicolson: Purification and partial characterization of a tumour-metastasis-associated high-Mr glycoprotein from rat 13762NF mammary adenocarcinoma cells. *Biochem J* 242, 779-787 (1987)

151. Moriarty J., C. M. Skelly, S. Bharathan, C. E. Moody & A. P. Sherblom: Sialomucin and lytic susceptibility of rat mammary tumor ascites cells. *Cancer Res* 50, 6800-6805 (1990)

152. Carraway K. L., E. A. Rossi, M. Komatsu, S. A. Price-Schiavi, D. Huang, P. M. Guy, M. E. Carvajal, N. Fregien & C. A. Carraway: An intramembrane modulator of the ErbB2 receptor tyrosine kinase that potentiates neuregulin signaling. *J Biol Chem* 274, 5263-5266 (1999)

153. Shankar V., M. S. Gilmore, R. C. Elkins & G. P. Sachdev: A novel human airway mucin cDNA encodes a protein with unique tandem-repeat organization. *Biochem J* 300, 295-298 (1994)

154. Williams S.J., D. H. Wreschner, M. Tran, H. J. Eyre, G. R. Sutherland, & M. A. McGuckin: MUC13 - a novel human cell surface mucin expressed by epithelial and hemopoietic cells. *J Biol Chem* (in press) (2001)

Abbreviations: Mr, relative molecular weight; Kb, kilo base pair; bp, base pair; EGF, epidermal growth factor; TGF, transforming growth factor; VNTR, variable number of tandem repeat; kDa, kilo Dalton; PCR, polymerase chain reaction; RACE-PCR, rapid amplification of cDNA ends; RT-PCR, reverse transcription-polymerase chain reaction; UTR, untranslated region; nm, nanometer; mm, micrometer

Key words: Mucin, Relation Structure-Function, EGF-Like Domaine, TGFβ-Like Domain, Secreted, Membrane-Bound, Cystin Knot Domain, Review

Send correspondence to: Surinder K. Batra, Ph.D., Department of Biochemistry and Molecular Biology, Eppley Institute for Research in Cancer and Allied Diseases, University of Nebraska Medical Center, 984525 Nebraska Medical Center, Omaha, NE 68198-4525, USA, Tel:402-559-5455, Fax:402-559-6650, E-mail: sbatra@unmc.edu