Structure and function of epithelial mucins

Irina Gout

Lviv Medical Institute
35a Nekrasova str., Lviv, Ukraine

Institute of Ophthalmology
11 Both str., London, UK

Mucins are the structural components of the epithelial mucosa that protect the respiratory, gastrointestinal and reproductive tracts from the hostile environments, including microorganisms, toxins and abrasives. Mucins constitute a group of high molecular weight (> 200 kDa), polydisperse and highly glycosylated proteins which are present on the surface of most epithelial tissues. Our understanding of the structure and function of mucins has advanced significantly in the last decade. This progress was mainly associated with the isolation of the cDNA clones, encoding a family of epithelial mucins. To date, this family includes eight mucin genes (MUC1—MUC8) and more await to be discovered. Based on sequence analysis and studies of subcellular localisation, epithelial mucins could be divided into two classes: membrane-associated (MUC1) and secretory (MUC2-8). This review is focused on our current knowledge of the structure of products of mucin genes and their function in normal tissues and in disease. The regulation of the expression of mucin genes, posttranslational modifications and alterations in secretion and processing will also be discussed.

Mucins as an integral part of mucus. Most mammalian epithelia that are in contact with hostile environments need protection, which is provided by secretion of a mucus gel. Specialised epithelial cells, such as goblet and non-goblet cells, secrete mucus which adheres to the epithelial surface and serves to hydrate and protect the underlying epithelium and also to trap foreign substances for their removal [1, 2]. Mucus can be characterised as a slimy, viscous secretion that contains water, minerals, immunoglobulins, and mucins. The specific properties of mucus vary among the different tissues, but there are some common features which include both physical aspects such as lubrication, protection and chemical composition. Mucins make up a large part of the mucus and contribute to its viscosity, elasticity, and protective functions.

Structural and functional diversity of mucin genes. Mucins constitute a heterogeneous group of high molecular weight (> 200 kDa) glycoproteins which are present on the surface of most epithelial cells [1, 3, 4]. Biochemical characterisation of mucins, purified from different tissues, indicated that these molecules are heavily glycosylated and, in some cases, contain more than 75 % of carbohydrates by weight [5]. On the basis of their sensitivity to proteolytic enzymes, the existence of so called «naked regions» that are devoid of carbohydrates was demonstrated. It was also found that sugar groups are mainly linked to the protein backbone via O-glycosidic bonds, formed between N-acetylgalactosamine and serine or threonine residues [3].

In general, mucins are difficult to analyse by conventional biochemical and biophysical methods, due to their large size and high degree of glycosylation. However, cDNA cloning and sequencing of mucin genes unfolded considerable insight into the primary structure and, subsequently, the function of these molecules.

The first mucin gene, called MUC1, was identified in 1987 [6]. To date, eight different mucin genes have been cloned (MUC1—MUC8) and more await to be discovered [7—16]. Due to the high level of heterogeneity and presence of numerous carbohydrate chains on mucin polypeptide backbones, they have proven to be difficult for protein sequencing analysis. As a result of that, most mucin genes were cloned by screening of expression libraries with polyclonal and monoclonal antibodies, raised against purified and chemically deglycosylated mucins. No
sequence homology was revealed between mucin genes from the analysis of corresponding cDNA clones, but common features of the internal organisation of the polypeptide backbone were uncovered [3, 4].

First of all, a major portion of each mucin gene was found to consist of tandem repeats, which are specific for each type of mucin. Secondly, serine and threonine residues, located within the tandem repeats, create potential glycosylation sites. Thirdly, the high degree of heterogeneity can be explained by a variable number of tandem repeats and different glycosylation patterns. Finally, the alternating pattern of hydrophylic and hydrophobic domains was found.

Tandem repeat sequences. Tandem repeats of mucin genes are not homologous to each other in sequence and differ significantly in length. For example, the MUC5AC gene product contains the shortest repeat structure of 8 amino acids, while the MUC6 gene product has the longest repeat unit of 169 amino acids [11, 12]. A summary of the mucin tandem repeats and their amino acid composition is presented in Table 1.

Among these differences, mucin tandem repeats have one common feature—a high percentage of three amino acids: serine, threonine and proline. These three amino acids together constitute 20—55 % of mucin tandem repeats. The serine and threonine residues of tandem repeats are potential glycosylation sites, through which hundreds of oligosaccharide chains may attach to a single core protein.

It has been proposed that proline residues in tandem repeats may play a structural role, since this amino acid is known to be involved in the formation of the α-turn left-handed helixes, which exhibit rigid rod-like structures [17]. The presence of proline residues near serine and threonine, particularly at position −1 and +3, was shown to be important for the initiation of O-glycosylation of mucins [17].

Electron microscopy revealed native mucin molecules as long filamentous structures, varying in length from 200 to >1000 nm [18]. Profound glycosylation and presence of a large number of proline residues in tandem repeats may determine the filamentous conformation of mucin molecules, observed by electron microscopy. Significant variability in the size of mucin molecules can be explained by the variable number of tandem repeats and polymerisation of subunits.

The arrangement of the mucin subunits in the mucus is also controversial. Several structural models have been proposed, including end-to-end attachment of mucin subunits by disulphide bridges and a flexible-thread model with or without the linking peptide [1, 19].

Glycosylation of mucins. Glycosylation of mucins take place in the cis-Golgi compartments and is subsequently followed by sialylation and sulphation [5]. Attached oligosaccharide chains differ in length: from 1 to 20 sugars per chain. It has also been demonstrated that the number and composition of associated oligosaccharides varies between different mucins which may explain the inherent heterogeneity of these molecules. Interestingly, the same mucins contain different oligosaccharide chains in different tissues and their composition is altered in diseases [20—22].

Sequence analysis of mucin genes show the presence of potential N-glycosylation sites. Xu et al. [23] identified N-glycosylation of rat intestinal mucin (RIM) by chemical analysis and suggested that this modification may play a role in the intracellular

<table>
<thead>
<tr>
<th>Mucins</th>
<th>Tandem repeat sequence</th>
<th>Length (aa)</th>
<th>Number of repeats</th>
</tr>
</thead>
<tbody>
<tr>
<td>MUC1</td>
<td>PDPAPAPGSTAPAHGVSTA</td>
<td>20</td>
<td>4—7</td>
</tr>
<tr>
<td>MUC2</td>
<td>PTTPTPTTTTTTPTPTPTQF</td>
<td>23</td>
<td>8</td>
</tr>
<tr>
<td>MUC3</td>
<td>HSTPSPTSSTTTGTTS</td>
<td>17</td>
<td>13</td>
</tr>
<tr>
<td>MUC4</td>
<td>TSSASTGHTTPLPTTD</td>
<td>16</td>
<td>10</td>
</tr>
<tr>
<td>MUC5AC</td>
<td>TSTTSAP</td>
<td>8</td>
<td>Not known</td>
</tr>
<tr>
<td>MUC5B</td>
<td>SSTPTAHHTLT1TTATATATGSTATP</td>
<td>29</td>
<td>Not known</td>
</tr>
<tr>
<td>MUC6</td>
<td>A very long repeat (169 aa)</td>
<td>169</td>
<td>Not known</td>
</tr>
<tr>
<td>MUC7</td>
<td>CRRPK1PPSPVKKPPFPNPQPQPK</td>
<td>23</td>
<td>Not known</td>
</tr>
<tr>
<td>MUC8</td>
<td>Two repeats (13 and 41 aa long)</td>
<td>13/41</td>
<td>Not known</td>
</tr>
</tbody>
</table>
targeting of mucin. In another study, N-glycosylation was shown to be important for the oligomerisation of gastric mucin [24]. However, more studies are necessary to clarify the importance of this posttranslational modification in mucin function.

Based on the composition of attached groups, mucins can be divided into three main classes: sialomucins (acidic mucins), sulfomucins (very acidic mucins) and neutral mucins, containing mainly hexoses. These different subclasses of mucins can be easily distinguished from each other by special histochemical analysis.

Membrane-associated and secretory mucins. Based on the structure and cellular localisation, epithelial mucins can be subdivided into membrane-associated (MUC1) and secretory (MUC2—MUC8), as shown in Fig. 1.

MUC1 is the only cloned membrane-associated mucin which consists of a large highly glycosylated extracellular domain, a membrane spanning domain, and a short cytoplasmic domain [6]. It has also been demonstrated that MUC1 is not able to form oligomeric complexes, which are specific for secretory mucins [23]. The extracellular domain extends well above the apical cell membrane and this can be important for specific function of this mucin. Membrane association of the MUC1, observed by immunohistochemical analysis, could be explained by the presence of a hydrophobic membrane-spanning domain. The function of the cytoplasmic domain is not known, but it has been suggested that this domain may provide a link with the actin cytoskeleton [26].

MUC1 mucin is expressed in nearly all epithelial tissues. Attention has recently focused on the aberrant expression of MUC1 by carcinomas of different origins, which makes it an important marker in cancer [27—29].

Secretory mucins constitute the viscous mucus of the tracheobronchial, gastrointestinal, and reproductive tracts and typically form extremely large oligomers through linkage of protein monomers via disulfide bridges. Complete cDNA sequences are available only for two secretory mucins: MUC2 and MUC7 [7, 9]. The cDNAs for MUC5A, B, and C were originally cloned from human tracheobronchial mucosa cDNA libraries [8, 11]. Chromosomal mapping has demonstrated that MUC5A and MUC5C are part of the same gene [30] and therefore were termed as MUC5AC. The presence of specific regions at the N-terminus, called leader sequences, determines their extracellular localisation. Besides the tandem repeat sequences, secretory mucins (MUC2, MUC5AC) were shown to contain a cysteine-rich domains [7, 11]. These domains are known to be important in mediating protein-protein interactions via the formation of disulfide bridges. The identification of cystein-rich domains in secretory type of mucins may explain the polymerisation of mucin molecules into gel. It is interesting that MUC7 lacks cystein-rich domain and is not able to form oligomers. This secretory mucin was found to be expressed only in submandibular salivary gland, but not all epithelial tissues have been analysed [9]. Cloning and characterisation of the full length cDNAs for other secretory

---

![Classification of mucins, based on the structure and tissue expression](image)
mucin genes (MUC3, MUC4, MUC6, and MUC8) is essential for the elucidation of their function.

Schematic structures of the transmembrane (MUC1) and secretory mucins (MUC2) are shown in Fig. 2.

**Mucin-like molecules.** Mucin-like proteins have been identified in non-epithelial tissues, such as endothelium and lymphocytes. Endothelial and leukocyte type of mucins are membrane associated and heavily glycosylated proteins, involved in lymphocyte migration via specific binding to selectins [3]. It is suggested that mucin-like proteins play an important role in inflammation via the initiation of the adhesion cascade and lymphocyte activation [31].

**Expression of mucin genes and its regulation.** Initial analysis of mucin gene expression by immunohistological methods was not very productive, because of the lack of specific antibodies directed against these

---

**Fig. 2.** Schematic structures of the membrane-bound (A) and secretory (B) mucins.

- N-linked oligosaccharides
- O-linked oligosaccharides
- S-S - disulfide bonds
proteins. Polyclonal and monoclonal antibodies raised against native mucins recognised mainly carbohydrate epitopes, which are difficult to classify and not just expressed on mucins. The use of chemically deglycosylated and recombinant (unglycosylated) mucins as antigens allowed generating of a panel of specific and high affinity antibodies [32].

The availability of specific antibodies and also cDNA probes enabled significant analysis of mucin gene expression in mammalian epithelia. Different techniques have been employed to fulfil this aim, including Northern and Western blot analysis, immunohistochemistry, RT-PCR and in situ hybridisation.

Numerous studies provided evidence that mucins are expressed in an extremely tissue-specific manner. Mucin transcripts and proteins have only been detected in epithelial tissues [6, 7, 9, 33, 34, 50—56]. Differential expression of mucin genes in various epithelial tissues has also been observed. For example, MUC1 and MUC5B exhibit a widespread pattern of expression and are detected in all or almost all epithelial tissue. On the other hand, MUC7 has so far found to be expressed only in submandibular salivary glands [9]. A summary of mucin gene expression in different human epithelial tissues is presented in Table 2.

Aberrant overexpression of mucin genes has been also detected in patients with cystic fibrosis, asthma and chronic bronchitis [21, 35]. Abnormal expression of mucin, was demonstrated in skin fibroblasts, derived from a patient with long-standing scleroderma. Histological examination of a biopsy specimen from the affected skin indicated significant accumulation of mucin between collagen bundles [36].

The understanding the mechanisms of tissue- and cell type-specific expression of mucin genes remains a topic of intense research. Developmental expression of mucin genes was shown to correlate well with the differentiation of epithelial cells in mammary glands, gastrointestinal and respiratory tracts [37, 38].

Many endogenous and exogenous factors, including growth factors and hormones, pathogenic organisms and chemicals, were found to influence expression of mucin genes. For example, expression of MUC1 in mammary glands was found to be

<table>
<thead>
<tr>
<th>Tissue</th>
<th>MUC1</th>
<th>MUC2</th>
<th>MUC3</th>
<th>MUC4</th>
<th>MUC5AC</th>
<th>MUC5B</th>
<th>MUC6</th>
<th>MUC7</th>
<th>MUC8</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bronchus</td>
<td>++</td>
<td>—</td>
<td>—</td>
<td>++</td>
<td>+</td>
<td>+</td>
<td>±</td>
<td>ND</td>
<td>+</td>
<td>[20, 33, 52, 53]</td>
</tr>
<tr>
<td>Trachea</td>
<td>+</td>
<td>+</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>±</td>
<td>±</td>
<td>ND</td>
<td>+</td>
<td>[20, 33, 52]</td>
</tr>
<tr>
<td>Lung</td>
<td>+</td>
<td>+</td>
<td>—</td>
<td>+</td>
<td>+</td>
<td>±</td>
<td>±</td>
<td>ND</td>
<td>+</td>
<td>[20, 33, 52]</td>
</tr>
<tr>
<td>Salivary gland</td>
<td>+</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>+</td>
<td>—</td>
<td>+</td>
<td>ND</td>
<td>+</td>
<td>[9, 20, 33, 53]</td>
</tr>
<tr>
<td>Esophagus</td>
<td>+</td>
<td>—</td>
<td>ND</td>
<td>+</td>
<td>—</td>
<td>—</td>
<td>ND</td>
<td>ND</td>
<td>+</td>
<td>[13, 20, 52, 53]</td>
</tr>
<tr>
<td>Stomach</td>
<td>+</td>
<td>—</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>—</td>
<td>—</td>
<td>ND</td>
<td>+</td>
<td>[9, 20, 33, 53]</td>
</tr>
<tr>
<td>Small intestine</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>ND</td>
<td>—</td>
<td>[13, 20, 33, 52, 53]</td>
</tr>
<tr>
<td>Colon</td>
<td>+</td>
<td>—</td>
<td>+</td>
<td>+</td>
<td>±</td>
<td>+</td>
<td>±</td>
<td>ND</td>
<td>ND</td>
<td>[13, 20, 33, 52]</td>
</tr>
<tr>
<td>Pancreas</td>
<td>+</td>
<td>±</td>
<td>—</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>[20, 53, 54]</td>
</tr>
<tr>
<td>Gall bladder</td>
<td>+</td>
<td>±</td>
<td>+</td>
<td>ND</td>
<td>±</td>
<td>+</td>
<td>±</td>
<td>ND</td>
<td>ND</td>
<td>[13, 20, 33, 53]</td>
</tr>
<tr>
<td>Mammary gland</td>
<td>+</td>
<td>—</td>
<td>ND</td>
<td>—</td>
<td>—</td>
<td>±</td>
<td>±</td>
<td>ND</td>
<td>ND</td>
<td>[20, 52, 53]</td>
</tr>
<tr>
<td>Uterus</td>
<td>+</td>
<td>—</td>
<td>ND</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>ND</td>
<td>ND</td>
<td>[20, 51—53]</td>
</tr>
<tr>
<td>Cervix</td>
<td>+</td>
<td>±</td>
<td>—</td>
<td>ND</td>
<td>+</td>
<td>+</td>
<td>±</td>
<td>ND</td>
<td>ND</td>
<td>[33, 52, 53]</td>
</tr>
<tr>
<td>Ovary</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>+</td>
<td>ND</td>
<td>ND</td>
<td>—</td>
<td>ND</td>
<td>—</td>
<td>[20, 53]</td>
</tr>
<tr>
<td>Prostate</td>
<td>+</td>
<td>±</td>
<td>±</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>[20]</td>
</tr>
<tr>
<td>Conjunctiva</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>+</td>
<td>+</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>[55, 56]</td>
</tr>
</tbody>
</table>

Level of expression: (+++) — high, (+++) — medium, (+) — low, (−) — not expressed, (ND) — not determined
stimulated by insulin, prolactin and hydrocortisone [39]. In another study, it was shown that EGF can regulate differentiation of tracheal epithelial cells and subsequently stimulate expression of mucins [40]. Interestingly, an increased expression of EGF and its functional analogue TGFα has been demonstrated in response to the inflammatory process or tissue injury, which also induce mucin expression. On the other hand, vitamin A, which is known as an important factor in the regulation of airway epithelial cell function, exhibits an opposite effect on mucin expression. Both time- and dose-dependent downregulation of MUC2 expression by vitamin A has recently been reported in tracheobronchial epithelial cells [41]. Taken together, these data indicate that the expression of mucin genes can be regulated in both a stimulatory and an inhibitory fashion.

Differential expression of mucin genes, induced by various stimuli, has also been reported. Estrogen and dexamethasone were found to increase MUC4 and MUC5 expression in human epithelial cell lines 4 fold, but not MUC1 [42]. In the same study, it was shown that progesterone does not exhibit a stimulatory effect. It was also demonstrated that differential expression of mucin genes depends on the growth and differentiation of various subpopulations of mucus-secreting epithelial cell lines [43].

In a number of studies, it has been shown that viral or bacterial infections lead to a dramatic increase in mucin expression [44]. Irritation of mucous membranes by chemicals also produces a stimulatory effect on the expression of mucin genes [45]. The molecular mechanism of this stimulation remains unclear. Recently, it was demonstrated that inflammation-associated mediators stimulate mucin secretion in tracheal epithelial cells via a mechanism involving intracellular production of nitric oxide, which serves as a signalling messenger [46].

Taken together these data indicate that expression of mucin genes is very complex and could be regulated at different levels, including transcription, alternative splicing and translation (schematic representation in Fig. 3).

**Regulation at the level of transcription.** Induced expression of different mucins in response to treatment with various stimuli, was shown to require *de novo* RNA and protein synthesis, indicating transcriptional regulation of this process. Chromosomal mapping and sequence analysis of genomic clones has opened the field of transcriptional regulation of mucin genes. Promoter sequences, identified within the MUC1 and MUC5B genes, greatly facilitated affinity purification and characterisation of transcription factors, involved in the regulation of mucin genes [47, 48]. A nuclear protein, called NF1-MUC5B, was found to modulate the transcriptional activity of the mucin MUC5B gene [47]. Further characterisation of the new promoter regions in other mucin genes will create the platform for the discovery of common and specific transcription factors, involved in transcriptional regulation of mucin genes.

**Heterogeneity of mucin mRNA.** Polydispersity of the messenger RNA, which is a characteristic of most mucins, with the exception of MUC1, may also contribute to the regulation of their expression pattern [49]. Many factors can contribute to the heterogeneity of mucin messages. Firstly, rapid turnover of mucin RNA, confirmed experimentally by using the pulse-chase technique, is one of the most likely explanations of the observed polydispersity. Several destabilising sequences have been found within the 3' untranslated region of mucin RNAs, which may at least in part explain their instability. It is known that nuclear and cytoplasmic proteins can specifically bind to these sequences and induce a rapid degradation of mucin RNA.

Secondly, an alternative splicing of mucin mRNAs has also been suggested as a contributing factor. However sequence analysis of many mucin cDNA clones did not provide sufficient evidence in favour of this explanation. Alternatively, susceptibility to RNAase degradation and physical instability of very large mucin messages (well over 10,000 bases), may be an essential factor in generating an artificial ladder of degraded mRNAs [10]. This hypothesis
could be ruled out by data indicating that polydispersity of mucin mRNAs is also observed in the RNA samples, which showed no sign of degradation. Taken together, these results suggest that rapid turnover of mucin mRNAs seems the most likely explanation of polydispersity. Regulation of mucin mRNA stability by intracellular factors is going to be the subject of intense investigation in the near future.

**Regulation at the translational level.** There are some data indicating that restriction of mucin expression to mucous cells is regulated at the level of translation [37]. This conclusion was based on the analysis of MUC2 and MUC3 expression in human bronchus, where mRNAs specific for these mucins were found both in mucin-expressing and mucin-non-expressing cells. So far, there are no data concerning the regulation of mucin gene expression at the post-translational level.

**Conclusion.** The last decade has been very productive and exciting in the study of mucin genes and their products. Isolation and sequence analysis of cDNA clones encoding eight mucin genes, provided insights into the structure of their products. Based on these studies, common features of the internal organisation of the polypeptide backbone were uncovered within the family of epithelial mucins. The availability of cDNA clones has opened new areas of research on mucins. Immunohistochemical and *in situ* hybridisation experiments with specific antibodies or cDNA probes provided the information on tissue distribution and subcellular localisation of the different mucin gene products. Expression studies of mucins *in vitro*, using cDNAs and cell lines, have also been initiated and should generate data which will enable better understanding of their role in physiological and pathological processes. However, much work remains to be done in elucidating the regulation of mucin gene expression. In order to examine this at the molecular level, the sequences of the mucin gene promoters have to be determined. The next few years will witness new insights into the research on mucins, especially with the generation of transgenic mice in which a particular mucin gene will be inactivated or over-expressed in epithelial tissues.

**Acknowledgments.** I am grateful to Prof. G. Semenova for support and encouragement. I thank Dr. I. T. Gout for the critical reading of the manuscript.

**I. T. Gout**

**Структура і функції епітеліальних мухин

**Резюме**

Мухини — це структурні компоненти епітеліальних шарів мухин, які захищують респіраторний, травний та репродуктивні імунні та епітеліальні процеси. Мухини складають групу високомолекулярних (> 200 кДа), полідисперсних та сильноглікозійованих білків, що знаходяться на поверхні більшості епітеліальних тканин. На сьогодні це сімейство включає 8 мухинових генів (MUC1—MUC8) і охоплює 200-сімейство мухинових генів та їхні функції в нормальних тканинах і при патологіях. Обговорюються також регуляція експресії мухинових генів, посттрансляційні модифікації та зміни в секреції i процесингу.

**REFERENCES**

cloning of human intestinal mucin (MUC2) cDNA // J. Biol. Chem.—1994.—269.—N.4.—P. 2440—2446.
43. Lalani E. N. Expression of the gene encoding for a human mucin in mouse mammary tumour cells can effect their tumorigenicity // J. Biol. Chem.—1991.—266.—P. 15420—15426.


53. De-Bolos C., Garrido M., Real F. X. MUC6 apomucin shows a distinct normal tissue distribution that correlates with Lewis antigen expression in the human stomach // Ibid.—1995.—109, N 3.—P. 723—734.


Received 14.01.98