Genome-Wide Search and Identification of a Novel Gel-Forming Mucin *MUC19/Muc19* in Glandular Tissues

Yin Chen, Yu Hua Zhao, Tejas Baba Kalaslavadi, Edward Hamati, Keith Nehrke, Anh Dao Le, David K. Ann, and Reen Wu

Center for Comparative Respiratory Biology and Medicine and Division of Pulmonary and Critical Care Medicine, University of California, Davis, California; Center for Oral Biology, University of Rochester, Rochester, New York; Department of Oral and Maxillofacial Surgery, Drew University of Medicine and Science, Los Angeles; and Department of Molecular Pharmacology and Toxicology, University of Southern California, Los Angeles, California

Gel-forming mucins are major contributors to the viscoelastic properties of mucus secretion. Currently, four gel-forming mucin genes have been identified: *MUC2, MUC5AC, MUC5B*, and *MUC6*. All these genes have five major cysteine-rich domains (four von Willebrand factor [vWF] C or D domains and one Cystine-knot [CT] domain) as their distinctive features, in contrast to other non–gel-forming type of mucins. The CT domain is believed to be involved in the initial mucin dimer formation and have very succinct relationship between different gel-forming mucins across different species. Because of gene duplication and evolutional modification, it is very likely that other gel-forming mucin genes exist. To search for new gel-forming mucin candidate genes, a "Hidden Markov Model"(HMM) was built from the common features of the CT domains of those gel-forming mucins. By using this model to screen all protein databases as well as the six-frame translated expression sequence tag and translated human genomic databases, we identified a locus located at the peri-centromere region of human chromosome 12 and the corresponding homologous region of mouse chromosome 15. We cloned the 3′ end of this gene and its mouse homolog. We found one vWF C domain, one CT domain, and various mucin-like threonine/serine-rich repeats. Phylogenetic analysis indicated the close relationship between this gene and the submaxillary mucin from porcine and bovine. A polydispersed signal was observed on the Northern blot, which indicates very large mRNA size. Further analysis of the upstream genomic sequences generated from human and mouse genome projects revealed three additional vWF D domains and many mucin-like threonine/serine-rich repeats. The expression of this gene is restricted to the mucous cells of various glandular tissues, including sublingual gland, submandibular gland, and submucosal gland of the trachea. Based on the chronological convention, we have given the name *MUC19* to the human ortholog and *Muc19* to the mouse.

(*Received in original form March 21, 2003 and in revised form July 2, 2003*)

Address correspondence to: Reen Wu, Ph.D., Center for Comparative Respiratory Biology and Medicine, Surge 1 Annex, Room 1121, University of California at Davis, One Shields Ave., Davis, CA 95616. E-mail: rwu@ucdavis.edu

Human *MUC19*: GenBank Accession Number AY236870 2. Mouse *Muc19*: GenBank Accession Number AY193891.

Abbreviations: bovine submaxillary mucins, BSM; Cystine-knot domain, CT; glyceraldehyde-phosphate dehydrogenase, GAPDH; Hidden Markov Model, HMM; expression sequence tag, EST; polymerase chain reaction, PCR; porcine submaxillary mucins, PSM; reverse transcriptase, RT; saline sodium citrate, SSC; von Willebrand factor C domain, VWC; von Willebrand factor D domain, VWD; von Willebrand factor, vWF.

Am. J. Respir. Cell Mol. Biol. Vol. 30, pp. 155–165, 2004
Originally Published in Press as DOI: 10.1165/rcmb.2003-0103OC on July 25, 2003
Internet address: www.atsjournals.org

Mucus is a viscoelastic gel-like substance that covers the mammalian epithelial surface of various tissues. The main functions of mucus include lubricating and protecting of epithelia from environmental insults. The viscous and elastic properties of mucus gel are generally attributable to the physical properties and structural features of mucin glycoproteins, specifically gel-forming mucins. *MUC*2, *MUC*5AC, *MUC*5B, *MUC*6 define this mucin subgroup and they are believed to have evolved from one common ancestor with von Willebrand factor (vWF) (1). Bovine and porcine submaxillary mucins (BSM, PSM) also belong to this subgroup (1). All of these gel-forming mucins have very large size (15–40 kb cDNA); they also share a similar structure and substantial sequence homology in the conserved regions. The cDNA sequences of those mucins have multiple "cysteine-rich" vWF C (VWC) and vWF D (VWD) domains in the flanking region of the mucin-like threonine/serine-rich repeats and Cystine knot (CT) domains in their C-terminal regions (1, 2). Both the cysteine number and their positions are extremely conserved in those domains, which play an essential role in forming disulfide-linked dimers (3–5) and multimers (1, 6, 7). No such domains are found in other non–gel-forming type of mucins. Their large size and the capability of forming multimers support the notion that these mucins have played a pivotal role in forming the mucus gel. Indeed, those gel-forming mucins have been proven to be major components of the mucus secretion of various organs (8–11).

In addition to the gel-forming mucins, fifteen other human mucin genes have been cloned and named as *MUC*1, 3–4 and 7–18 (http://www.ncbi.nlm.nih.gov/LocusLink/). Generally speaking, individual mucins are named because that they have so-called "threonine/serine-rich mucin repeats," and they share no apparent sequence similarities as a big group (12). Among those mucins, some are membrane-tethered (*MUC*1, 3, 4, 11, 12) and some are very small (*MUC*7, 9, 10) (12). The contribution of those mucins to the biophysical and biochemical properties of mucus gel is not entirely clear.

For many years, the total number of mucin genes has remained a mystery. Currently, four human gel-forming mucin genes have been identified: *MUC*2, *MUC*5AC, *MUC*5B, and *MUC*6. New gel-forming mucin may also exist due to gene duplication, chromosomal exchange, or other genetic alterations. Current progress in DNA sequencing has led to the creation of many sequence databases that are useful resources for the discovery of new proteins. Now that the human genome project has been completed, potential gene candidates can be predicted from their genomic sequence. In addition, another useful database is dbEST (NCBI expression sequence tag [EST] database http://www.ncbi.nlm.nih.gov/dbEST/). dbEST contains the cloned cDNA sequences by the reverse transcription of mRNA samples from various tissues, and has been widely used for the study of gene expression.

One general approach to discover new members of a gene family is to search the nucleotide databases for similar sequences of this gene family by BLAST program (http://www.ncbi.nlm.nih.gov/BLAST/). However, many gene families, such as gel-forming mucins, don't have the overall sequence similarities; rather, they only share some conserved "motifs" such as the CT domain. This difficulty can be overcome by searching the database using sequence profiles rather than merely the sequence *per se*. There are many methods for constructing sequence profile from a multiple sequence alignment; the resulting profile represents the mathematical summary of the specific features of these sequences extracted from those known members of a given gene family. Searching the database by using a sequence profile is like looking for the general "features" of those genes rather than just similar DNA sequences (13, 14). "Hidden Markov Model" (HMM) is one of the most powerful tools in this regard (15, 16).

Using this HMM-based searching method, Schultz and coworkers (17) have discovered more than 1,000 new putative human small GTPase proteins. Combined with EST database search and BLAST search on genomic sequence, Wittenberger and colleagues (13) have uncovered new members of the G-protein–coupled receptor superfamily. Therefore, this HMM-based search approach will be more robust and specific than the BLAST program. In this report, we have used this approach to identify *MUC19/Muc19*, as a novel glandular tissue–specific gel-forming mucin gene.

Materials and Methods

Screening the Novel Gel-Forming Mucin Genes

As shown in Figure 1, we collected all the known gel-forming mucin genes, including those of human and other animal species, from the NCBI database. We chose the 3′ end sequences because of the concern that some genes, such as *MUC6,* only have 3′ end sequences. Moreover, most of the EST sequences were generated from the 3′ end. All sequences were selected and processed with Blast2 (NCBI software program). Only the most representative sequences were preserved. These genes were then aligned by the ClustalW program (18). A gel-forming mucin gene-specific HMM was built based on the alignment data by using HMMER2.2 software from Sean Eddy's Lab Home Page of Washington University at St. Louis, MO (http://hmmer.wustl.edu/). NCBI human and mouse EST databases were downloaded to an in-house Linux computer. All those sequences were six-frame translated, then they were screened using the "gel-forming Hidden Markov Model" by HMMSEARCH in the HMMER2.2 software package. Initially, a default cutoff value (< 1) was used. All hits were then used to search the NCBI nr database to find out if those ESTs corresponded to the known genes by using an in-house search program. By visual inspection, we found that there was a large gap in the scores among all those hits. All the known nonmucin genes have a score much smaller than 0.01. Thus, a second cutoff value (< 0.01) was used to filter the results. The same method was also used to search the human and mouse genomic databases from NCBI again. The only difference in this search was that all the genomic sequences were first translated by GENESCAN program (19) before the search.

3′ and 5′-RACE

The RACE kit (Roche Diagnostics Corporation, Indianapolis, IN) was used to synthesize the first-strand cDNA from total RNA (3 μg) isolated from human and mouse salivary gland tissues. All the procedures followed the manufacturer's instructions. Briefly, Oligo-dT anchor primer or antisense gene-specific primer corresponding to different regions of *MUC19/Muc19* message were used to initiate first-strand cDNA synthesis. For the 3′-RACE, PCR was performed by 5′ gene specific primers and 3′ oligo d(T) anchor primer. For 5′-RACE, a 3′ tailing with oligo d(A) with terminal deoxynucleotidyl transferase was performed on the first-strand cDNA, then a PCR was performed using the nested gene-specific primer and the 5′ oligo d(T) anchor primer. The PCR products were subcloned into the TA vector (Invitrogen, Carlsbad, CA) for cloning and DNA sequencing. All primer sequences used in this study are listed on the Table 1.

Reverse Transcriptase–Polymerase Chain Reaction Amplification

cDNA was synthesized from total RNA (3 μg) by RT with oligo d(T) primer. The resulting single-strand cDNA was used as a template for polymerase chain reaction (PCR) amplification by *MUC19/Muc19* gene-specific primers (Table 1). PCR products were TA cloned and sequenced.

Phylogenetic Analysis

All non-repetitive 3′ end peptide sequences from gel-forming mucins of different species were aligned using ClustalW program (18). The alignment was edited and the tree was built by Jalview program (Michele Clamp [michele@ebi.ac.uk]).

Genomic Structure and Localization

The chromosomal location of *MUC19/Muc19* was determined by BLAST search of NCBI genomic databases (http://www.ncbi.nih.gov) and Blat search of UCSC human genome draft (University of California, Santa Cruz, CA). The whole genomic structure was deduced from the comparison between the porcine submaxillary gland mucin (GenBank: AAC62527) and the human/mouse genomic sequences in the *MUC19/Muc19* locus by locally installed Genewise program (Ewan Birney [http://www.sanger.ac.uk/Software/Wise2]).

RNA Isolation and Northern Blot Hybridization

RNA was isolated from human and mouse tissues by a single-step acid guanidinium thiocyanate phenol-chloroform extraction method (20). For Northern blot hybridization, equal amounts of total RNA (20 μg/lane) were subjected to electrophoresis on a 1.2% agarose gel in the presence of 2.2 M formaldehyde and then transblotted onto Nytran membranes. The RNA was cross-linked to membrane by an ultraviolet Stratalinker 2,400 (Stratagene, La Jolla, CA). The clones corresponding to the 3′ end sequence of *MUC19/Muc19* were labeled with ^{32}P-dCTP by ready-to-go kit (Amersham Biosciences Corp., Piscataway, NJ). After hybridization, all the blots were

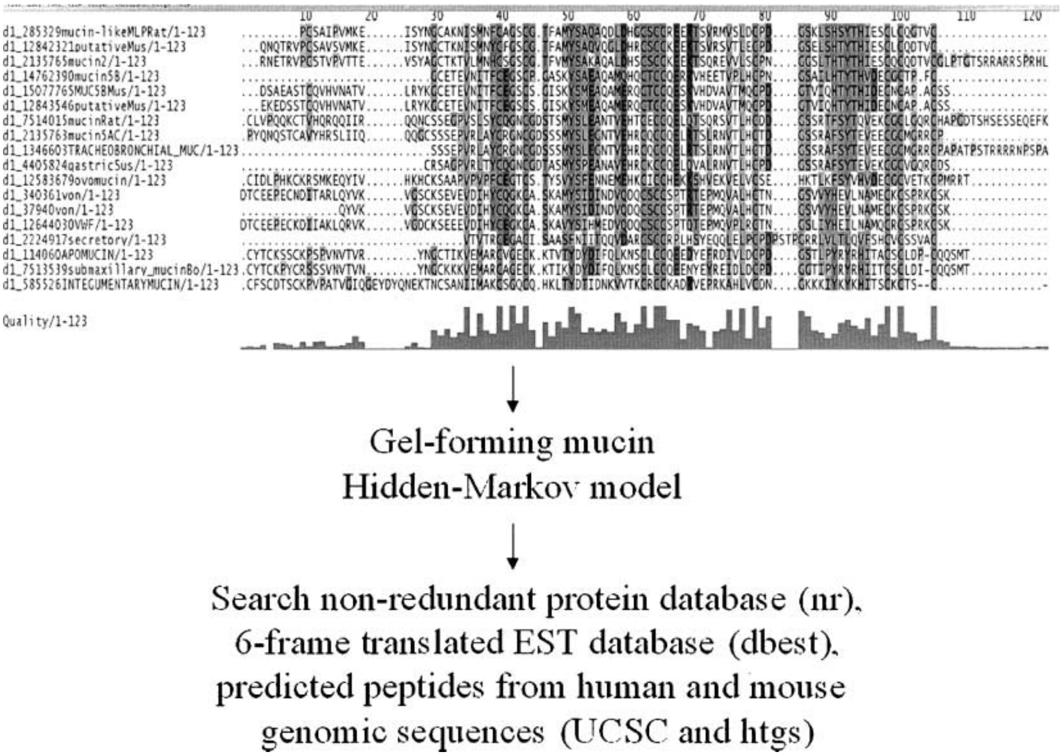

Quality/1-123
Gel-forming mucin
Hidden-Markov model
Search non-redundant protein database (nr),
6-frame translated EST database (dbest),
predicted peptides from human and mouse
genomic sequences (UCSC and htgs)

Figure 1. The strategy of creating "HMM" for gel forming mucin. Currently known gel-forming mucin genes from all species were collected and their C-terminal sequences were aligned and used for the creation of HMM for genome-wide search. The "gel-forming mucin HMM" was then used to screen various sequence databases, as described in MATERIALS AND METHODS.

exposed overnight to the phosphor screen and read by the STORM system (Molecular Dynamics, Sunnyvale, CA) (21). The integrity of the RNA sample was verified by visualization of ribosomal 18S and 28S bands in the ethidium bromide–stained gel.

Expression Analysis by Quantitative Reverse Transcriptase-PCR

SYBR Green based quantitative reverse transcriptase (RT)-PCR approach was performed to characterize the message expression of *MUC19/Muc19* and the control gene, glyceraldehyde-phosphate dehydrogenase (GAPDH), in various human and mouse tissues. Mouse cDNA samples were generated in house from various tissue RNAs by reverse transcriptase using Oligo d(T) anchor primer. Human cDNA samples (human multiple tissue cDNA panel I, cat #K1420–1 and human multiple tissue cDNA panel II, cat # K1421–1) from various human tissues were purchased from BD Biosciences Clontech (Palo Alto, CA). Gene-specific primers were designed according to the cDNA region. Each PCR reaction contained 10 μM primers for a total volume of 50 μl PCR reaction solution. SYBR Green PCR kit (Applied Biosystems, Foster City, CA) was used according to manufacturer's instructions. Real-time PCR data were obtained using GeneAmp 5,700 Sequence Detection System (Applied Biosystems). The normalized expression values of *MUC19/Muc19* were obtained by dividing the corresponding

TABLE 1
Primers used for RT-PCR and 3′/5′ RACE

Genes		Primer Sequences	
MUC19	hmuc19_1878	F	5′-GAGTTCAGATGGCAAAATGCACA-3′
	hmuc19_2021	R	5′-TGCCATCAGGACAGTCAAGTACA-3′
	hmuc19_1110	F	5′-GATCAACGGGAGTCACCAGC-3′
	hmuc19_1431	R	5′-ACTGGAGCTGGTGGAAGTG-3′
	hmuc19_1333	F	5′-ACCACAAGTATCCCAGCCAG-3′
	hmuc19_1426	R	5′-AGCTGGTGGAAGTGAGGCTA-3′
Muc19	mmuc19_1392	F	5′-GATTATGCGATTGGTTCATCCT-3′
	mmuc19_1740	R	5′-GTGCAATGTCCCTGAACTCATA-3′
	mmuc19_1378	F	5′-TATTTAACAATACCGATTATG-3′
	mmuc19_1443	R	5′-AGGAGAGGCATGGGTTGCTTG-3′
Oligo d(T) anchor primer		R	5′-GACCACGCGTATCGATGTCGACTTTTTTTTTTTTTTTTV-3′
GAPDH		R	5′- TGAAGGTCGGAGTCAACGGATTTGGT -3′
		F	5′- CATGTGGGCCATGAGGTCCACCAC -3′

expression values of GAPDH. To improve the readability of the data, all final expression values (listed in Table 2) were further multiplied by a factor (10,000). The tissue with the value that is less than 0.01 has undetectable *MUC19/Muc19* expression using this PCR method.

In Situ Hybridization

Glass slide sections from various tissue blocks were hybridized in the hybridization solution using biotin-labeled antisense or sense probes synthesized by *in vitro* transcription of *MUC19/Muc19* clones. *In situ* hybridization was performed as per the manufacturer's protocol (Roche Diagnostics Corp., Indianapolis, IN) and modified as described before (21). Briefly, slide sections were treated with 10 μg/ml Proteinase K in 50 mM Tris-Cl (pH 8.0) and 50 mM ethylenediamenetetraacetic acid for 15 min at 37°C, rinsed twice in 0.2× saline sodium citrate (SSC) thereafter, and then postfixed in 4% paraformaldehyde/phosphate-buffered saline for 20 min. Slides were treated twice for 5 min each time with 0.1 M triethanolamine (pH 8.0) and blocked by 0.25% acetic anhydride in 0.1 M triethanolamine. The sections were then dehydrated through the ethanol series. For each section, 0.5 pmol biotin-labeled oligonucleotide probe in 50 μl of hybridization buffer was applied. The hybridization buffer contained 2× SSC, 1× Denhardt's solution, 10% dextran sulfate, 50 mM phosphate buffer (pH 7.0), 50 mM dithiothreitol, 250 μg/ml yeast tRNA, 100 μg/ml poly A, and 500 μg/ml salmon sperm DNA. The section was hybridized at 45°C overnight in a humidified chamber. After hybridization, the section was washed twice for 15 min each time at 37°C with 2× SSC, twice for 15 min each time with 1× SSC, and twice for 15 min each time with 0.25× SSC. After the wash, the slide was reacted with anti-biotin primary antibody conjugated with alkaline phosphatase. After several washes, the reacted probes in the slide were color-developed with the Biotin Nucleic Acid Detection kit from Roche Diagnostics Corp.

TABLE 2
Tissue-specific MUC19/Muc19 *expression in various human and mouse tissues*

Tissues	*MUC19* Expression (Human)	*Muc19* Expression (Mouse)
Brain	< 0.01*	< 0.01
Colon	< 0.01	< 0.01
Heart	< 0.01	< 0.01
Kidney	< 0.01	< 0.01
Leukocytes	< 0.01	ND
Liver	< 0.01	< 0.01
Lung	< 0.01	< 0.01
Pancreas	< 0.01	< 0.01
Parotid gland	ND	< 0.01
Prostate	< 0.01	< 0.01
Skeletal muscle	< 0.01	ND
Small intestine	< 0.01	< 0.01
Spleen	< 0.01	ND
Sublingual gland	ND	891.23
Submandibular gland	482.03	283.84
Thymus	< 0.01	< 0.01
Trachea	12.35	70.38

Definition of abbreviation: ND, not determined.

* The normalized expression values of *MUC19/Muc19* were obtained by dividing the corresponding expression values of GAPDH. In order to improve the readability of the data, all final expression values were further multiplied by a factor (10,000). The tissue with the value that is < 0.01 has undetectable *MUC19/Muc19* expression using this PCR method.

Results

Developing HMM for the Genome-Wide Search of New Gel-Forming Mucin Genes

To conduct a genome-wide search for new gel-forming mucin genes, a specific HMM was developed based on the sequence alignment of all known gel-forming mucins (Figure 1). To enhance the representation of this model, sequences from species other than human and mouse were also included in the alignment. Using this finalized model, a comprehensive search of the human and mouse EST databases revealed many hits with high scores, especially those from the mouse EST databases. Most of these hits were parts of the known gel-forming mucin genes (*MUC2/Muc2*, *MUC5AC/Muc5AC*, etc.). Because of significant high score of these hits in the mouse EST database, we decided to focus on the mouse gene. After processing those results by an in-house program, 24 mouse ESTs that did not match any known mouse mucin gene were obtained from the search. These ESTs were in fact generated from the same gene. The translated product of this new gene has a bona fide gel-forming mucin like CT domain (Figure 2Ab).

Molecular Cloning and Sequence Characterization of the 3′ End of Novel Gel-Forming Mucin Gene, *Muc19*

We then performed 5′/3′-RACE using the primers deduced from the potential coding region of this new gene. The total mouse salivary gland RNA was used because all the ESTs from this new gene were obtained from mouse salivary gland library SG2. For 5′-RACE, we used mmuc19_1740 as gene specific primer; for 3′-RACE, we used mmuc19_1392. Sequences of the primers are listed in the Table 1. By these methods, we were able to obtain two cDNA clones (1.897 kb and 2.023 kb) that were generated by different polyadenylation sites (Figure 2Aa). The longer transcript has the same ORF as the shorter one, but has longer 3′ UTR. The sequence has been deposited into GenBank under the accession number AY193891. The deduced peptide sequence has significantly high threonine and serine content (35.9%) and several mucin-like threonine/serine-repeats (Figure 2Ac). It also has the signature motifs of gel-forming mucin: VWC and CT domains (Figure 2Ab). Because mucins are named numerically in chronological order, we therefore named this new mouse mucin gene as *Muc19*. By comparing the *Muc19* sequence with the UCSC and NCBI human genome sequence database, the cloned 3′ end sequence of *Muc19* was found to reside at chromosome 15 (Figure 3A) and consists of 9 exons (Figure 2Aa).

Identification of the Human *MUC19* Locus by Searching the Translated Genomic Database with "Gel-Forming Mucin HMM"

In contrast to mouse EST database search, the human *MUC19* was not found in the human EST library. After looking through the current human EST libraries, we realized this problem might be due to the lack of the human salivary gland library in the human EST database. To overcome this obstacle, we performed the screening using the translated human genomic databases deduced from the publicly available GenBank database. By using this approach, we were

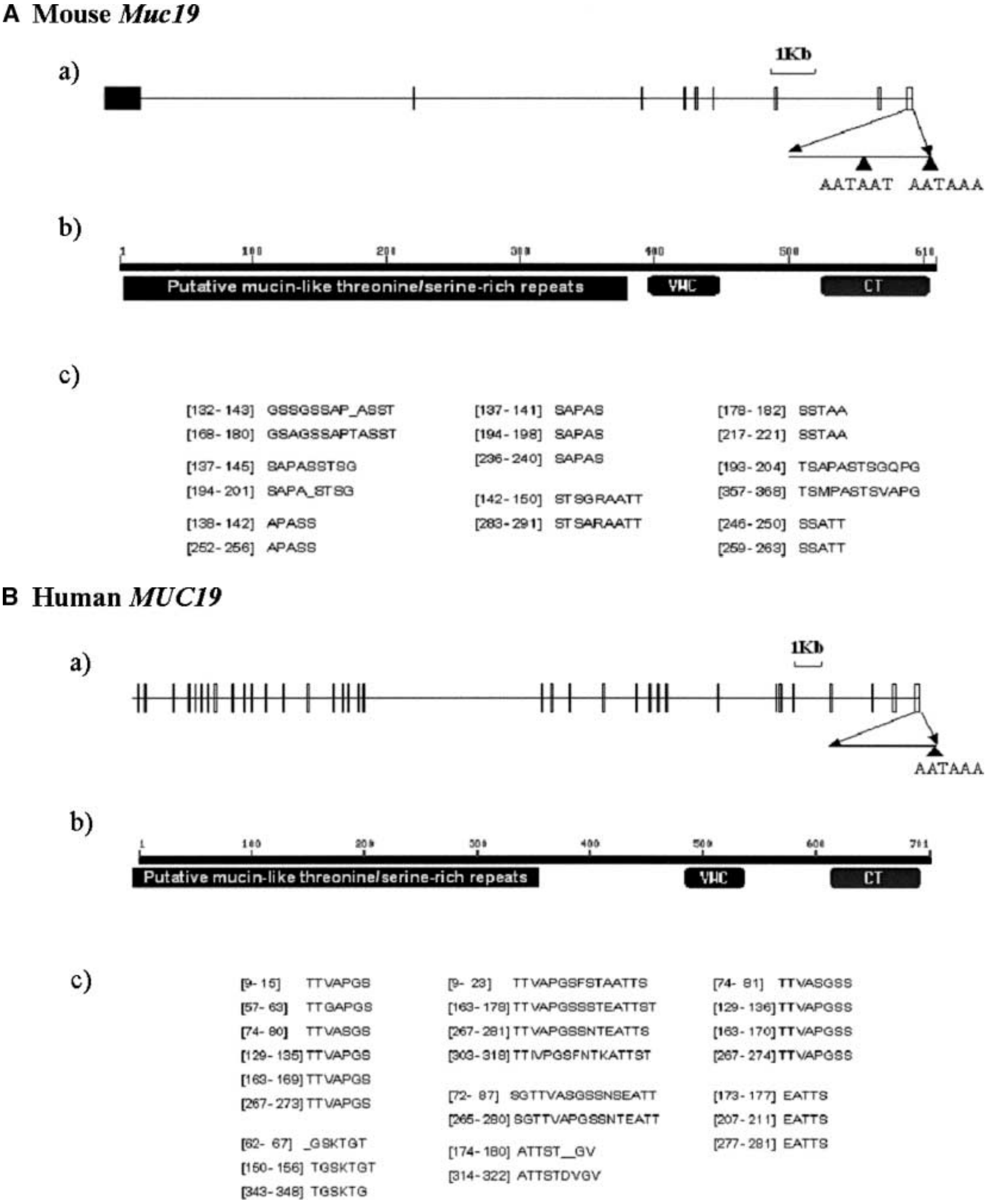


Figure 2. The gene structure and feature of 3′ end of human and mouse *Muc19* gene. (*A*) Mouse *Muc19* gene. (*a*) Gene structure of the 3′ end of *Muc19*. It has 9 exons, and the first one is incomplete. Two polyadenylation sites are detected (AATAAT and AATAAA), but both transcripts have the same open reading frame. (*b*) The deduced peptide sequence of the C-terminus of mouse Muc19 gene product. Both putative mucin repeats, VWC, and CT domains are indicated in the figure. (*c*) The putative mucin-like threonine/serine-rich repeats identified in the primary amino acid sequence. *Numbers in parentheses* indicate the amino acid position of these repeats. (*B*) Genetic structure and deduced amino acid sequence of the 3′end of human *MUC19*. (*a*) The genetic structure of the 3′ end of human *MUC19* gene with 35 exons and one polyadenylation site (AATAAA). (*b*) The deduced peptide sequence of the 3′ end of human *MUC19* gene, including the presence of both VWC and CT domains. (*c*) The mucin-like threonine/serine-rich repeats in the sequence of *MUC19* are indicated, and their positions in this region are indicated in the *parentheses*.

able to identify the putative human *MUC19* locus in chromosome 12 (Figure 3B).

We also screened the translated mouse genomic databases and found the mouse *Muc19* locus at chromosome 15 (Figure 3B), which further confirmed the sensitivity and accuracy of our screening method. Interestingly, this portion of mouse chromosome 15 seems to be the homologous region to the human chromosome 12.

Notably, we were unable to identify any candidates other than *MUC19/Muc19* by this search on both human and mouse genomes.

Molecular Cloning and Sequencing of the 3′ End of Human *MUC19*

To clone the human *MUC19*, we designed various primers corresponding to the deduced cDNA region of the *MUC19* locus. RT-PCR and 5′/3′-RACE were used to amplify *MUC19* cDNA from human salivary gland RNA. For 5′-RACE, we used primer hmuc19_2021; for 3′-RACE, we used hmuc19_1878. We further used the hmuc19_1110/hmuc19_2021 primer pair to confirm the sequence by RT-PCR. All the primer sequences are listed in Table 1. Using these approaches, we cloned and sequenced a 2.23-kb

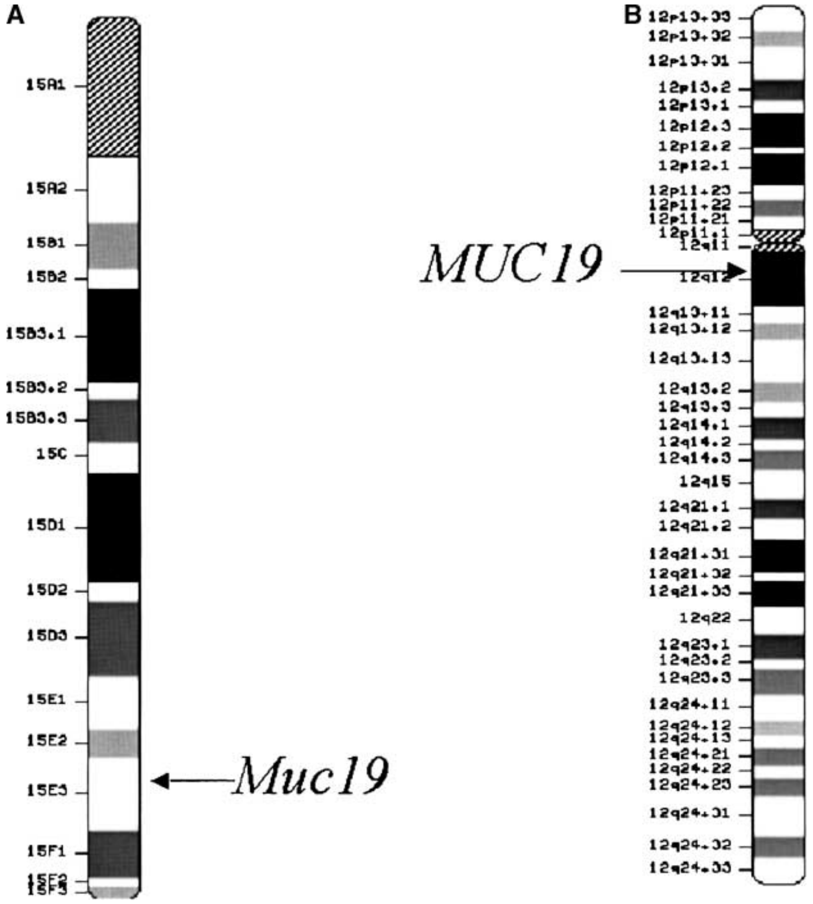


Figure 3. Chromosomal locations of *Muc19* and *MUC19* genes in mouse and human chromosomes. (*A*) Mouse *Muc19* at chromosome 15E3. (*B*) Human *MUC19* at chromosome 12q12.

MUC19 cDNA fragment (Figure 2B) (GenBank accession: AY236870). The deduced peptide sequence also has very high threonine and serine content (31.4%) and many mucin-like threonine/serine repeats (Figure 2Bc). VWC and CT domains were also identified in the sequence (Figure 2Bb). Interestingly, we did not find an alternative polyadenylation site in the human sample (Figure 2Ba). The cloning of *MUC19* from human salivary gland tissue demonstrates the similarity of the expression between human and mouse clones in terms of tissue specificity.

Phylogenetic and Sequence Analysis of Gel-Forming Mucin Genes

Phylogenetic analysis of various gel-forming mucin genes and vWF from different species indicates that *MUC19* belongs to the PSM/BSM cluster (Figure 4), which is consistent with the detection of *MUC19/Muc19* transcripts in the human and mouse salivary glands. The sequence alignment also demonstrates the numerous similarities between *MUC19/Muc19* and PSM (Figure 5). It also appears that *MUC19* is much more similar to PSM than *Muc19* (Figure 5). Notably, the similarities among those three sequences are particularly high within the last 250AA of the C-terminus, where the CT domain resides. CT domains have been shown to play a crucial role in dimer formation (4, 5). It appears that the mucin repeat regions are very diversified even among the homologs in different species, which is also true for other gel-forming mucins. The only common feature of those mucin repeats is that they are all threonine/serine-rich and contain potential sites for O-glycosylation.

The Predicted Gene Structure Upstream of the Cloned 3′ End of *MUC19/Muc19*

The genomic sequences from both the human *MUC19* and mouse *Muc19* locus allow us to deduce the genomic structure and protein motifs. Most importantly, those sequences have been shown to be very similar to PSM. We then tried to predict the gene structure upstream of the cloned *MUC19/Muc19* sequences by comparing their genomic sequence with PSM peptide sequence using Genewise program (Ewan Birney [http://www.sanger.ac.uk/Software/Wise2]). The benefit of this prediction program is that it uses sequence homology in addition to sequence statistics to facilitate the exon prediction. Thus it is more accurate than the conventional exon prediction method like GENESCAN that is solely dependent on sequence statistics (19). As shown in Figures 6A and 6B, both peptide sequences deduced from the genomic sequences share similar structural domains with other gel-forming mucin genes: 5′-VWD-VWD-VWD-mucin repeats-VWC-CT-3′. Both genes seem to have a very large central region containing most of the serine/threonine-rich repeats, which is reminiscent of the large central exon of *MUC5B* gene (22). Those structural features are very similar to PSM (Figure 6C). As we ex-

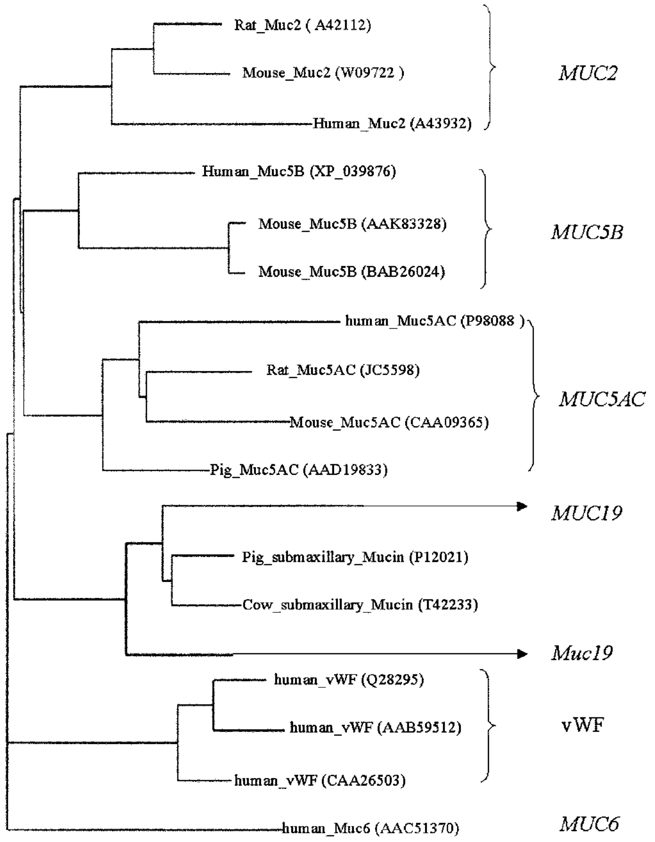


Figure 4. Phylogenetic tree analysis of all the gel-forming mucin genes and vWF from different species. Sequences used for the construction of this phylogentic tree were from GenBank with an accession number of: A43932 for human *MUC2*; W09722 *for mouse *Muc2*; A42112 for rat *Muc2*; A57534 and P98088 for human *MUC5AC*; JC5598 for rat *Muc5ac*; AAD19833 for Pig *Muc5ac*; XP_039876 for human *MUC5B*; AAK83328 and BAB26024 for mouse *Muc5b*; AAC51370 for human *MUC6*; AAB59512, Q28295, and CAA26503 for human *vWF*; P12021 for pig *PSM*; and T42233 for bovine *BSM* mucin. This analysis suggests that both human *MUC19* and mouse *Muc19* belong to PSM/BSM cluster. *W09722 is a nucleotide sequence. In this phylogenetic tree construction, the deduced peptide sequence from this accession number was used.

pected, the predicted peptide sequences from *MUC19* and *Muc19* were very similar with the PSM sequence (Figure 7). Highly homologous sequences were found at both the C terminus and putative N terminus of the peptide sequences of *MUC19/Muc19*, while no significant homology was seen in the central repetitive regions (Figure 7). Both *MUC19* and *Muc19* are very large genes. Human *MUC19* has more than 180kb of genomic sequence with a deduced peptide sequence larger than 7,000 amino acids, whereas mouse *Muc19* has ~ 80 kb of genomic sequences with ~ 3,000 amino acids. The smaller size of mouse *Muc19* might result from more gaps and much lower quality of the mouse genomic sequences available in the current database. We expect that the genomic size of mouse *Muc19* is probably similar to human *MUC19* when the mouse genomic project is complete.

Characterization of the Expression of *MUC19/Muc19* *In Vitro* and *In Vivo*

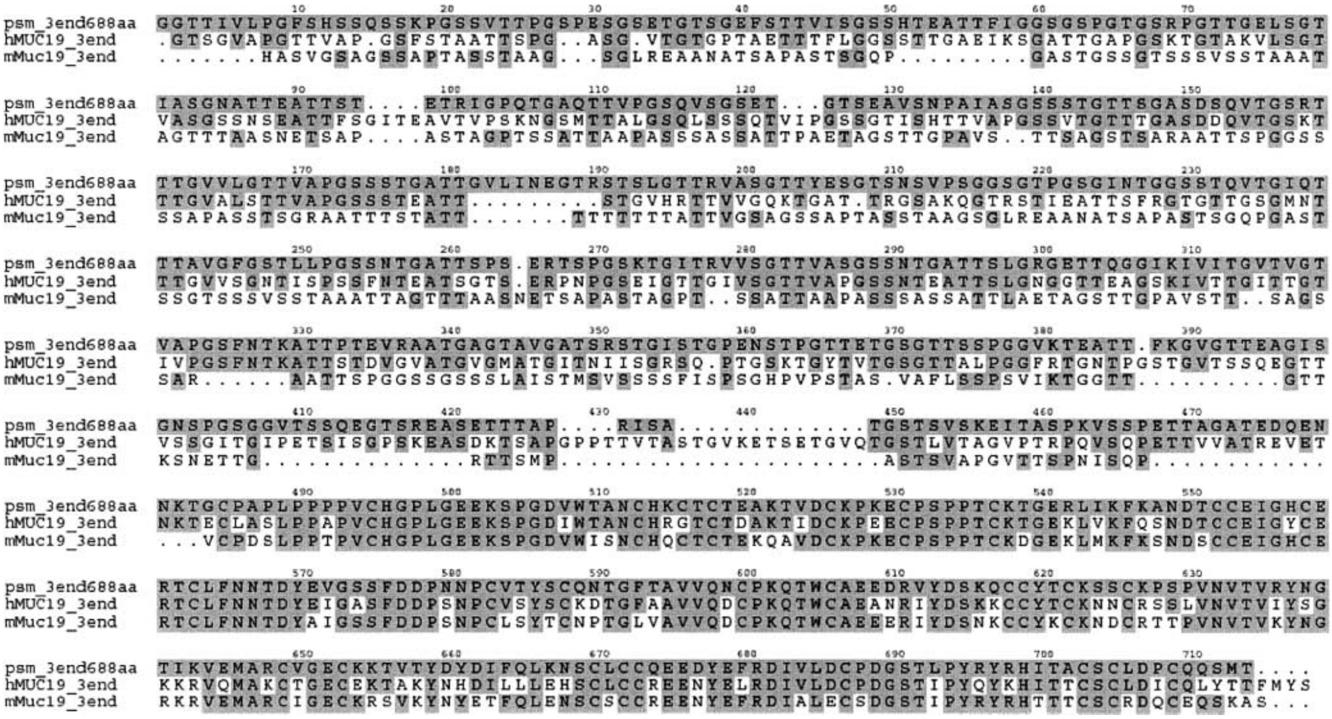
To further examine the expression of *MUC19/Muc19* in various tissues, both Northern blot and RT-PCR approaches were used to screen the mouse and human multitissue panels. Like other gel forming mucin gene messages, the Northern blot revealed a polydispersed feature of *MUC19/Muc19* messages in salivary gland and tracheal tissues (Figure 8A). In mouse, *Muc19* is mainly expressed in the two major salivary glands, sublingual and submandibular, and to a much lesser extent in trachea (Figure 8A). *Muc19* is expressed at a higher level in the sublingual gland than that in the submandibular gland, and it is undetectable in the parotid gland (Figure 8A). This result is consistent with the distribution of mucous cell population in those glands. In these three major salivary glands, the sublingual gland contains mostly the mucous cell type, the submandibular gland contains a mixture of mucous and serous cell types, and the parotid gland cells are mostly the serous cell type. In human tissues, we also detected similar polydispersed signals from trachea and submandibular gland RNA samples (Figure 8B). To increase the sensitivity and the coverage of this tissue distribution study, we further used the quantitative RT-PCR method to screen additional human and mouse tissue samples. In the screening, primers hmuc19_1333/hmuc19_1426 were used for human, and primers mmuc19_1378/mmuc19_1443 were used for mouse. As summarized in Table 2, *MUC19/Muc19* expression is very restricted and cannot be detected by RT-PCR in various nonglandular tissues.

We used *in situ* hybridization to further examine the specific cell types that express *MUC19/Muc19* messages. *MUC19* transcripts were detected in the mucous cells of the submandibular gland and submucosal gland of the trachea from human (Figure 9). A similar positive hybridization of mouse *Muc19* probe was seen in mouse tissue sections from the sublingual gland and tracheal submucosal gland (Figure 10). Notably, there is no hybridization signal in most serous cells of these glands. The strict cell type specificity of *MUC19/Muc19* may explain why low levels of these transcripts in the tracheal RNA sample in which most of the RNA species are generated from the nonglandular portion.

Discussion

The current explosion of sequence data from the genome project and EST project of different species make it much easier to identify new gene family members. In addition to the simple sequence similarity search, pattern-based search methods have proven to be more robust (13, 17). In this study, we successfully used the HMM-based approach to identify a novel gel-forming mucin gene, *MUC19/Muc19*, which are specifically expressed in various glandular tissues.

In contrast to conventional biological discovery, the bioinformatic discovery approach requires a precise mathematical definition of the specific feature of the gene family of interest. Our initial attempt to define the "mucin-like threonine/serine-rich repeats" for discovering new mucin genes was a complete failure. This was partly due to the



```
                          10        20        30        40        50        60        70
psm_3end688aa  GGTTIVLPGFSHSSQSSKPGSSVTTPGSPESGSETGTSGEFSTTVISGSSHTEATTFIGGSGSPGTGSRPGTTGELSGT
hMUC19_3end    .GTSGVAPGTTVAP.GSFSTAATTSPG..ASG.VTGTGPTAETTTFLGGSSTTGAEIKSGATTGAPGSKTGTAKVLSGT
mMuc19_3end    .......HASVGSAGSSAPTASSTAAG...SGLREAANATSAPASTSGQP.........GASTGSSGTSSSVSSTAAAT

                          90       100       110       120       130       140       150
psm_3end688aa  IASGNATTEATTST....ETRIGPQTGAQTTVPGSQVSGSET...GTSEAVSNPAIASGSSSTGTTSGASDSQVTGSRT
hMUC19_3end    VASGSSNSEATTFSGITEAVTVPSKNGSMTTALGSQLSSSQTVIPGSSGTISHTTVAPGSSVTGTTTGASDDQVTGSKT
mMuc19_3end    AGTTTAASNETSAP....ASTAGPTSSATTAAPASSSASSATTPAETAGSTTGPAVS..TTSAGSTSARAATTSPGGSS

                         170       180       190       200       210       220       230
psm_3end688aa  TTGVVLGTTVAPGSSSTGATTGVLINEGTRSTSLGTTRVASGTTYESGTSNSVPSGGSGTPGSGINTGGSSTQVTGIQT
hMUC19_3end    TTGVALSTTVAPGSSSTEATT.........STGVHRTTVVGQKTGAT.TRGSAKQGTRSTIEATTSFRGTGTTGSGMNT
mMuc19_3end    SSAPASSTSGRAATTTSTATT.......TTTTTTTATTVGSAGSSAPTASSTAAGSGLREAANATSAPASTSGQPGAST

                         250       260       270       280       290       300       310
psm_3end688aa  TTAVGFGSTLLPGSSNTGATTSPS.ERTSPGSKTGITRVVSGTTVASGSSNTGATTSLGRGETTQGGIKIVITGVTVGT
hMUC19_3end    TTGVVSGNTISPSSFNTEATSGTS.ERPNPGSEIGTTGIVSGTTVAPGSSNTEATTSLGNGGTTEAGSKIVTTGITTGT
mMuc19_3end    SSGTSSSVSSTAAATTAGTTTAASNETSAPASTAGPT..SSATTAAPASSSASSATTLAETAGSTTGPAVSTT..SAGS

                         330       340       350       360       370       380       390
psm_3end688aa  VAPGSFNTKATTPTEVRAATGAGTAVGATSRSTGISTGPENSTPGTTETGSGTTSSPGGVKTEATT.FKGVGTTEAGIS
hMUC19_3end    IVPGSFNTKATTSTDVGVATGVGMATGITNIISGRSQ.PTGSKTGYTVTGSGTTALPGGFRTGNTPGSTGVTSSQEGTT
mMuc19_3end    SAR......AATTSPGGSSGSSSLAISTMSVSSSSFISPSGHPVPSTAS.VAFLSSPSVIKTGGTT..........GTT

                         410       420       430       440       450       460       470
psm_3end688aa  GNSPGSGGVTSSQEGTSREASETTTAP....RISA.............TGSTSVSKEITASPKVSSPETTAGATEDQEN
hMUC19_3end    VSSGITGIPETSISGPSKEASDKTSAPGPPTTVTASTGVKETSETGVQTGSTLVTAGVPTRPQVSQPETTVVATREVET
mMuc19_3end    KSNETTG..............RTTSMP......................ASTSVAPGVTTSPNISQP............

                         490       500       510       520       530       540       550
psm_3end688aa  NKTGCPAPLPPPPVCHGPLGEEKSPGDVWTANCHKCTCTEAKTVDCKPKECPSPPTCKTGERLIKFKANDTCCEIGHCE
hMUC19_3end    NKTECLASLPPAPVCHGPLGEEKSPGDIWTANCHRGTCTDAKTIDCKPEECPSPPTCKTGEKLVKFQSNDTCCEIGYCE
mMuc19_3end    ...VCPDSLPPTPVCHGPLGEEKSPGDVWISNCHQCTCTEKQAVDCKPKECPSPPTCKDGEKLMKFKSNDSCCEIGHCE

                         570       580       590       600       610       620       630
psm_3end688aa  RTCLFNNTDYEVGSSFDDPNNPCVTYSCQNTGFTAVVQNCPKQTWCAEEDRVYDSKQCCYTCKSSCKPSPVNVTVRYNG
hMUC19_3end    RTCLFNNTDYEIGASFDDPSNPCVSYSCKDTGFAAVVQDCPKQTWCAEANRIYDSKKCCYTCKNNCRSSLVNVTVIYSG
mMuc19_3end    RTCLFNNTDYAIGSSFDDPSNPCLSYTCNPTGLVAVVQDCPKQTWCAEEERIYDSNKCCYKCKNDCRTTPVNVTVKYNG

                         650       660       670       680       690       700       710
psm_3end688aa  TIKVEMARCVGECKKTVTYDYDIFQLKNSCLCCQEEDYEFRDIVLDCPDGSTLPYRYRHITACSCLDPCQQSMT....
hMUC19_3end    KKRVQMAKCTGECEKTAKYNHDILLLEHSCLCCREENYELRDIVLDCPDGSTIPYQYKHITTCSCLDICQLYTTFMYS
mMuc19_3end    RKRVEMARCIGECKRSVKYNYETFQLENSCSCCREENYEFRDIALECSDGSTIPYRYRHTTTCSCRDQCEQSKAS...
```

Figure 5. Comparisons of the C-terminal peptide sequences of *MUC19* and mouse *Muc19* with PSM. Multiple sequence alignment of C-terminal peptide sequences of *MUC19* (hMUC19_3end), *Muc19* (mMuc19_3end), and PSM (psm_3end 688aa). The identical amino acids are marked by *shading*.

heterogeneous nature of the mucin genes; some of them are named quite arbitrarily. *MUC7*, for example, was named as mucin only because of the presence of four mucin-like serine/threonine-rich repeats (23). As a matter of fact, many immunoglobulin genes have more mucin repeats than *MUC7*. It seems that the conventional mucin definition is too loose to distinguish the real mucins from other mucin-like genes. Thus, we tried to define the mucin genes based on additional features of their peptide sequences. We found that a mucin subgroup called "gel-forming mucin" (1, 12) was much easier to be defined. All of these gel-forming mucin genes share similar conserved motifs and structures.

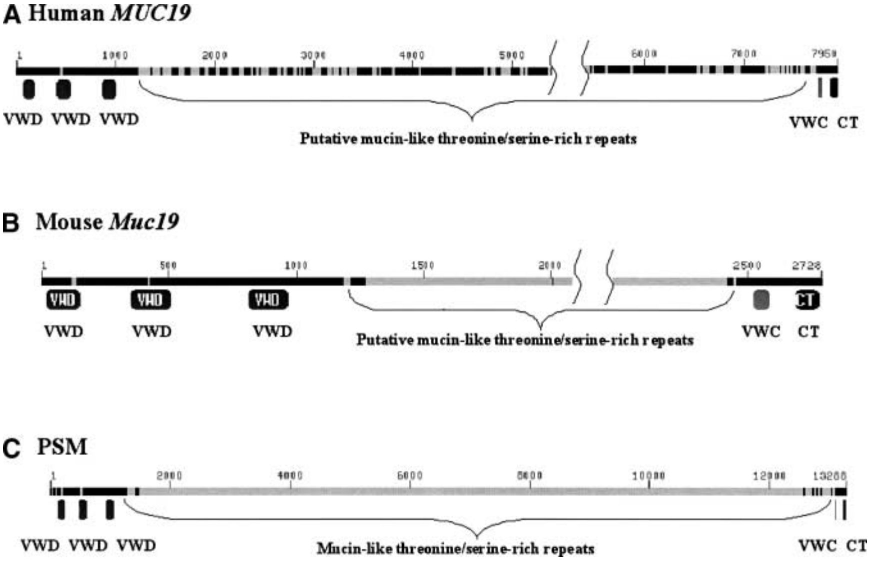


Figure 6. The predicted full-length peptide sequences of *MUC19* and *Muc19*. (*A*) A predicted peptide sequence of human *MUC19*. The cloned sequence has also been combined together. VWD, VWC, and CT domains as well as mucin threonine/serine-rich repeats are indicated in the figure. The *S-shaped lines* across the sequence indicate the gap of the predicted sequence, and this is due to either the existing gap in the genomic sequence or the sequence homology is too low for the prediction to proceed. (*B*) A predicted peptide sequence of mouse *Muc19*. The cloned sequence has also been combined together. VWD, VWC, and CT domains as well as mucin-like threonine/serine-rich repeats are indicated in the figure. The *S-shaped lines* across the sequence indicate the gap of the predicted sequence, and this is due to either the existing gap in the genomic sequence or the sequence homology is too low for the prediction to proceed. (*C*) PSM sequence (GenBank: AAC 62527) is included for comparison.

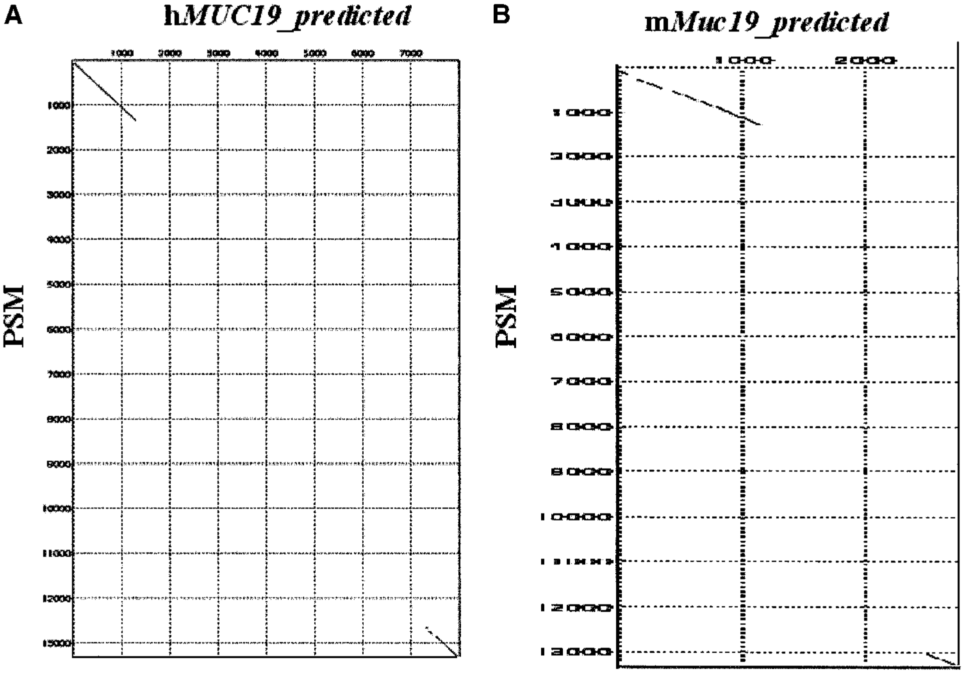


Figure 7. Comparison of the whole predicted peptide sequences of *MUC19* and *Muc19* with PSM. (*A*) Dotplot comparison of the predicted peptide sequences between *MUC19* (h*MUC19*_predicted) and PSM. Sequences with > 70% similarity are indicated in the figure. (*B*) Dotplot comparison of the predicted peptide sequences between *Muc19* (m*Muc19*_predicted) and PSM. Sequences with > 70% similarity are indicated in the figure.

Most notably, they have been suggested to be the determining factor for the viscoelastic properties of mucus secretion and mucus gel formation in various organs. We therefore defined the "gel-forming mucin-specific HMM" based on specific features at the 3′ ends of known gel-forming mucin genes in various species. After screening the ESTs databases, we found this "gel-forming mucin-specific HMM" to be very specific and discriminating. The approach identified all previously known gel-forming mucin genes of various species without missing any. No other hits had a high enough score to be considered except *MUC19/Muc19*. That was also true when translated human and mouse genomic databases were included for the screening.

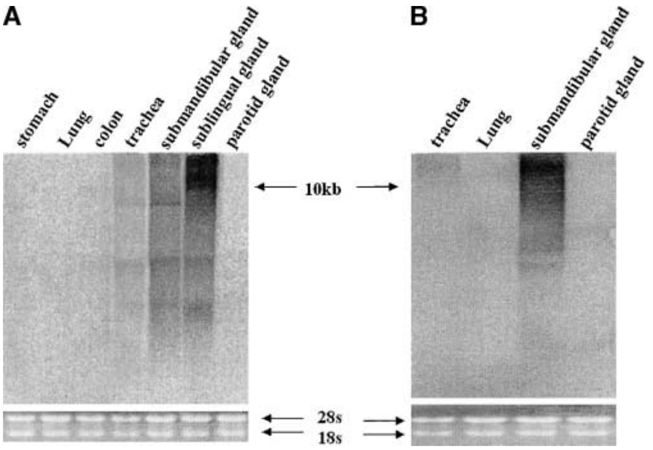


Figure 8. Northern blot analysis of *MUC19/Muc19* expression in various human and mouse tissues. (*A*) Tissue-specific expression of mouse *Muc19*. *Lane 1:* stomach; *lane 2:* lung; *lane 3:* colon; *lane 4:* trachea; *lane 5:* submandibular gland; *lane 6:* sublingual gland; *lane 7:* parotid gland. (*B*) Tissue-specific expression of human *MUC19*. *Lane 1:* trachea; *lane 2:* lung; *lane 3:* submandibular gland; *lane 4:* parotid gland.

The newly identified *MUC19/Muc19* gene has the gel-forming mucin feature with a structure significantly similar to the porcine and bovine submaxillary mucins. It has been suggested that all the known gel-forming mucin genes are evolved from one common ancestor with vWF by gene duplication events (1). Structurally, *MUC19/Muc19* are also very similar to vWF as well as other gel-forming mucin genes. Interestingly, human *MUC19* resides in chromosome 12q12, which is close to the location of vWF (12p13). In the phylogenetic tree, *MUC19* is much closer to the *MUC2/MUC5AC/MUC5B* than *MUC6*, although *MUC6* is also located in the 11p15 (24). We suspect that *MUC19* shares a similar ancestor with the other gel-forming mucins and branched out evolutionarily later than *MUC6*.

The most striking feature of *MUC19/Muc19* is their size. Of the known sequences, *MUC5B* is the largest human gel-forming mucin, consisting of ~ 5,000 amino acids (21, 22, 25). However, newly identified human MUC19 has more than 7,000 amino acids based on the known sequence. Considering that there are gaps in the sequence and its porcine counterpart has 13,288 amino acids (26), MUC19 must be the largest gel-forming mucin protein ever identified. Because of their huge size, *MUC19/Muc19* may play a significant role in the regulation of the viscosity of mucus secretions. Such a role may be critical not only to the normally protective function of mucus, but also to its pathologic nature in diseases when mucus secretion is too thick and viscous to be cleared. Further studies of *MUC19/Muc19* expression in various airway and glandular diseases could help to elucidate the contributing role played by MUC19/Muc19 in mucus secretion.

Similar to their porcine/bovine counterparts, *MUC19/*

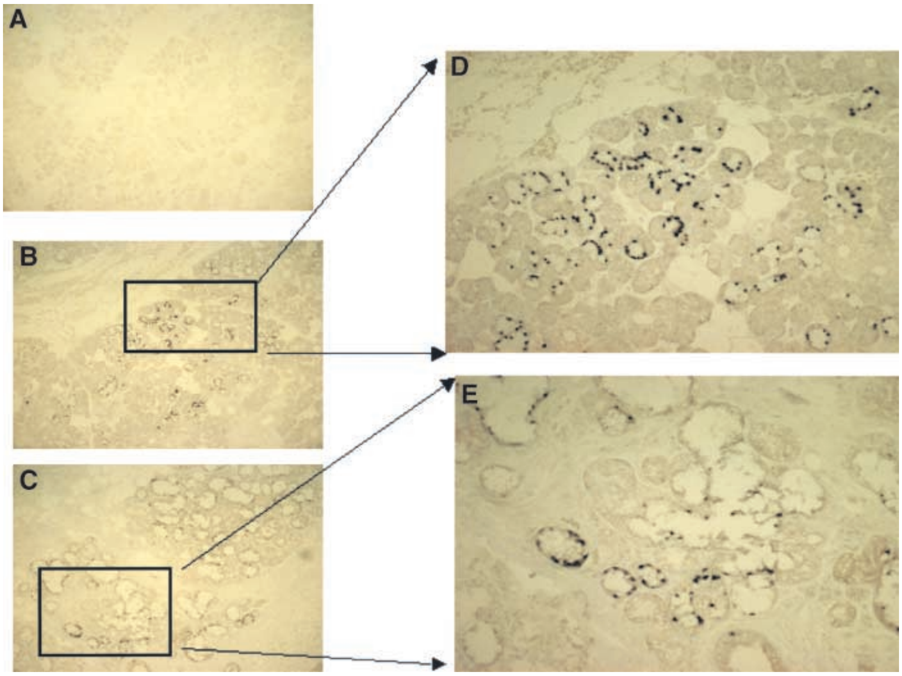


Figure 9. *In situ* hybridization of *MUC19* in human salivary gland and tracheal tissue sections. (*A*) Control, sense probe. (*B*) Submandibular gland section with antisense probe. (*C*) Tracheal section with submucosal gland region. *D* and *E* are enlarged pictures of *B* and *C*, respectively.

Muc19 is expressed mainly in the major salivary glands, including both the sublingual and submandibular glands. This then raises the question: what is the major mucin component in the saliva? Previous study has suggested that MUC5B protein is the major mucin component in the high molecular weight portion of salivary mucus based on the comparison of the known mucin species in the saliva as well as in RNA samples from salivary gland (27, 28). However, a recent paper indicates that concentrated solutions of salivary MUC5B protein alone cannot replicate the gel-forming properties of saliva (29), which suggests the presence of additional mucin(s) in mediating mucus gel formation. In this study, we have demonstrated that *MUC19/Muc19* transcripts are present in the major salivary glands at a high level. Because its large size, this new mucin may be one of the major components contributing to the viscosity of salivary mucus.

We have also demonstrated the expression of *MUC19/Muc19* in the mucous cells of airway submucosal glands. Submucosal gland is one of the major sources for the airway mucus secretion. Until now, MUC5B protein is the only gel-forming mucin identified in the mucous cells of human

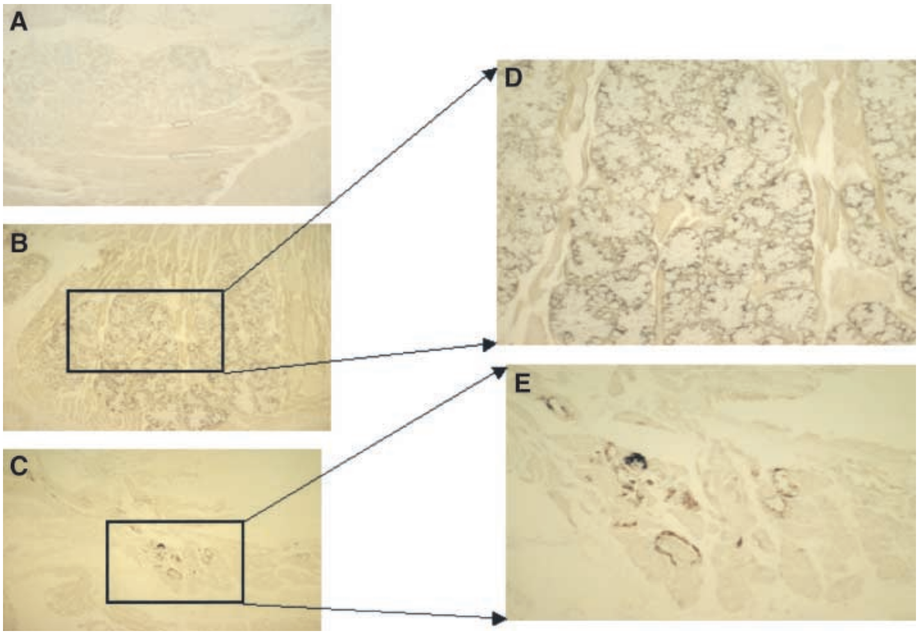


Figure 10. *In situ* hybridization of *Muc19* in mouse salivary gland and tracheal tissue sections. (*A*) Control, sense probe. (*B*) Sublingual gland section with antisense probe. (*C*) Tracheal section with submucosal gland region. *D* and *E* are enlarged pictures of *B* and *C*, respectively.

airway submucosal glands (21). Both MUC5B and MUC5AC mucin proteins have been identified in human airway secretions from normal and patients with various chronic diseases (8, 9, 30–32). It is very possible that MUC19/Muc19 protein also contributes to airway mucus secretion. Because of its huge size, MUC19/Muc19 mucin may be essential to determining the viscoelasticity of the airway mucus secretion. In the chronic airway diseases such as asthma and COPD, the presence of unusually high level of MUC19/Muc19 mucin could be detrimental to the morbidity and mortality of these diseases by increasing the tenacious nature of mucus plugs in airways.

In summary, we have identified a novel gland-specific gel-forming mucin gene, *MUC19/Muc19* by using HMM-based genome-wide search approach. Molecular cloning and sequence information suggest that this mucin gene is probably the largest gel-forming mucin gene ever identified, and it has all the features of the known gel-forming mucins. Expression analyses, based on Northern blot and *in situ* hybridization, demonstrate that *MUC19/Muc19* is mainly expressed in the mucous cells of various glands, including the major salivary glands (sublingual and submandibular glands), and the submucosal gland of large airways. Further studies of the expression and the biochemical properties of this novel mucin gene in various mucus secretions will be essential to understanding the function and the regulation of this newly found mucin in the normal and disease.

Acknowledgments: This manuscript is supported in part by NIH grants (HL35635, ES06230, ES09701, AI50496, ES04699, and ES05707), the California Tobacco-Related Disease Research Program (10RT-0262), and Cystic Fibrosis Research, Inc. (Grant #03–006j).

References

1. Desseyn, J. L., J. P. Aubert, N. Porchet, and A. Laine. 2000. Evolution of the large secreted gel-forming mucins. *Mol. Biol. Evol.* 17:1175–1184.
2. Offner, G. D., D. P. Nunes, A. C. Keates, N. H. Afdhal, and R. F. Troxler. 1998. The amino-terminal sequence of MUC5B contains conserved multifunctional D domains: implications for tissue-specific mucin functions. *Biochem. Biophys. Res. Commun.* 251:350–355.
3. Bell, S. L., I. A. Khatri, G. Xu, and J. F. Forstner. 1998. Evidence that a peptide corresponding to the rat Muc2 C-terminus undergoes disulphide-mediated dimerization. *Eur. J. Biochem.* 253:123–131.
4. Perez-Vilar, J., A. E. Eckhardt, and R. L. Hill. 1996. Porcine submaxillary mucin forms disulfide-bonded dimers between its carboxyl-terminal domains. *J. Biol. Chem.* 271:9845–9850.
5. Perez-Vilar, J., and R. L. Hill. 1998. The carboxyl-terminal 90 residues of porcine submaxillary mucin are sufficient for forming disulfide-bonded dimers. *J. Biol. Chem.* 273:6982–6988.
6. Perez-Vilar, J., A. E. Eckhardt, A. DeLuca, and R. L. Hill. 1998. Porcine submaxillary mucin forms disulfide-linked multimers through its amino-terminal D-domains. *J. Biol. Chem.* 273:14442–14449.
7. Perez-Vilar, J., and R. L. Hill. 1998. 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.
8. Thornton, D. J., I. Carlstedt, M. Howard, P. L. Devine, M. R. Price, and J. K. Sheehan. 1996. Respiratory mucins: identification of core proteins and glycoforms. *Biochem. J.* 316:967–975.
9. Thornton, D. J., M. Howard, N. Khan, and J. K. Sheehan. 1997. Identification of two glycoforms of the MUC5B mucin in human respiratory mucus: evidence for a cysteine-rich sequence repeated within the molecule. *J. Biol. Chem.* 272:9561–9566.
10. Tytgat, K. M., H. A. Buller, F. J. Opdam, Y. S. Kim, A. W. Einerhand, and J. Dekker. 1994. Biosynthesis of human colonic mucin: Muc2 is the prominent secretory mucin. *Gastroenterology* 107:1352–1363.
11. de Bolos, C., F. X. Real, and A. Lopez-Ferrer. 2001. Regulation of mucin and glycoconjugate expression: from normal epithelium to gastric tumors. *Front. Biosci.* 6:D1256–D1263.
12. Moniaux, N., F. Escande, N. Porchet, J. P. Aubert, and S. K. Batra. 2001. Structural organization and classification of the human mucin genes. *Front. Biosci.* 6:D1192–D1206.
13. Wittenberger, T., H. C. Schaller, and S. Hellebrand. 2001. An expressed sequence tag (EST) data mining strategy succeeding in the discovery of new G-protein coupled receptors. *J. Mol. Biol.* 307:799–813.
14. Bateman, A., E. Birney, R. Durbin, S. R. Eddy, R. D. Finn, and E. L. Sonnhammer. 1999. Pfam 3.1: 1313 multiple alignments and profile HMMs match the majority of proteins. *Nucleic Acids Res.* 27:260–262.
15. Sonnhammer, E. L., S. R. Eddy, E. Birney, A. Bateman, and R. Durbin. 1998. Pfam: multiple sequence alignments and HMM-profiles of protein domains. *Nucleic Acids Res.* 26:320–322.
16. Hofmann, K. 2000. Sensitive protein comparisons with profiles and hidden Markov models. *Brief. Bioinform.* 1:167–178.
17. Schultz, J., T. Doerks, C. P. Ponting, R. R. Copley, and P. Bork. 2000. More than 1,000 putative new human signalling proteins revealed by EST data mining. *Nat. Genet.* 25:201–204.
18. Thompson, J. D., D. G. Higgins, and T. J. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22:4673–4680.
19. Burge, C. B., and S. Karlin. 1998. Finding the genes in genomic DNA. *Curr. Opin. Struct. Biol.* 8:346–354.
20. Chomczynski, P., and N. Sacchi. 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* 162:156–159.
21. Chen, Y., Y. H. Zhao, Y. P. Di, and R. Wu. 2001. Characterization of human mucin 5B gene expression in airway epithelium and the genomic clone of the amino-terminal and 5′-flanking region. *Am. J. Respir. Cell Mol. Biol.* 25:542–553.
22. Desseyn, J. L., V. Guyonnet-Dupérat, N. Porchet, J. P. Aubert, and A. Laine. 1997. 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.
23. Bobek, L. A., J. Liu, S. N. Sait, T. B. Shows, Y. A. Bobek, and M. J. Levine. 1996. Structure and chromosomal localization of the human salivary mucin gene, MUC7. *Genomics* 31:277–282.
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, and N. Porchet. 1996. Human mucin genes assigned to 11p15.5: identification and organization of a cluster of genes. *Genomics* 38:340–352.
25. Desseyn, J. L., J. P. Aubert, I. Van Seuningen, N. Porchet, and A. Laine. 1997. Genomic organization of the 3′ region of the human mucin gene MUC5B. *J. Biol. Chem.* 272:16873–16883.
26. Eckhardt, A. E., C. S. Timpte, A. W. DeLuca, and R. L. Hill. 1997. The complete cDNA sequence and structural polymorphism of the polypeptide chain of porcine submaxillary mucin. *J. Biol. Chem.* 272:33204–33210.
27. Nielsen, P. A., E. P. Bennett, H. H. Wandall, M. H. Therkildsen, J. Hannibal, and H. Clausen. 1997. Identification of a major human high molecular weight salivary mucin (MG1) as tracheobronchial mucin MUC5B. *Glycobiology* 7:413–419.
28. Thornton, D. J., N. Khan, R. Mehrotra, M. Howard, E. Veerman, N. H. Packer, and J. K. Sheehan. 1999. Salivary mucin MG1 is comprised almost entirely of different glycosylated forms of the MUC5B gene product. *Glycobiology* 9:293–302.
29. Raynal, B. D., T. E. Hardingham, D. J. Thornton, and J. K. Sheehan. 2002. Concentrated solutions of salivary MUC5B mucin do not replicate the gel-forming properties of saliva. *Biochem. J.* 362:289–296.
30. Sheehan, J. K., C. Brazeau, S. Kutay, H. Pigeon, S. Kirkham, M. Howard, and D. J. Thornton. 2000. Physical characterization of the MUC5AC mucin: a highly oligomeric glycoprotein whether isolated from cell culture or in vivo from respiratory mucous secretions. *Biochem. J.* 347:37–44.
31. Sheehan, J. K., M. Howard, P. S. Richardson, T. Longwill, and D. J. Thornton. 1999. Physical characterization of a low-charge glycoform of the MUC5B mucin comprising the gel-phase of an asthmatic respiratory mucous plug. *Biochem. J.* 338:507–513.
32. Thornton, D. J., T. Gray, P. Nettesheim, M. Howard, J. S. Koo, and J. K. Sheehan. 2000. Characterization of mucins from cultured normal human tracheobronchial epithelial cells. *Am. J. Physiol. Lung Cell. Mol. Physiol.* 278:L1118–L1128.