Degenerate 87-base-pair tandem repeats create hydrophilic/hydrophobic alternating domains in human mucin peptides mapped to 11p15

José DUFOSSE,* Nicole PORCHET,*† Jean-Pierre AUDIE,* Véronique GUYONNET DUPERAT,* Anne LAINE,* Isabelle VAN-SEUNINGEN,* Slaheddine MARRAKCHI,† Pierre DEGAND,*† and Jean-Pierre AUBERT*†‡

*Unité 16 INSERM, Place de Verdun, 59045 Lille Cédex, and †Laboratoire de Biochimie de l'Hôpital C. Huriez, CHRU, 59037 Lille Cédex, France

A human tracheobronchial λ gt 11 cDNA library was screened using antiserum prepared against the deglycosylated protein backbone of human tracheobronchial mucins. Two cDNAs, designated JER 28 and 57, obtained from this immunoscreening, were used to isolate two other cDNA clones, JUL 7 and JUL 10, from a human tracheobronchial λ gt 10 cDNA library. These four clones (561, 1830, 1631 and 991 bp), which mapped to chromosome 11p15, were all found to contain degenerate 87-base-pair tandem repeats which encode non-repetitive peptides.

Numerous deletions or insertions in an otherwise virtually perfect 87-base-pair tandem repeat create many shifts in reading frame which completely destroy the repetitive peptide structure. The peptide is composed of alternate hydrophobic and hydrophilic domains which probably differ in the extent to which they are glycosylated. The mRNAs are expressed both in the respiratory and in the digestive tracts. These human mucin probes may be important in assessing the abnormal mucins associated with inflammatory diseases or carcinoma from human mucosae.

INTRODUCTION

Mucins are secretory glycoproteins which constitute the viscous gel that covers mucosal epithelia of respiratory, gastrointestinal and reproductive tracts. The protective and lubricative properties of mucus are due to the viscous and viscoelastic properties of mucins, the major glycoprotein components of mucus. Mucins are characteristically heavily glycosylated (70–80% of their molecular mass) via the hydroxy groups on serine and threonine, which are major constituents of their protein backbone (for a recent review, see [1]). The composition of these molecules and their high level of heterogeneity made the elucidation of their primary peptide structures extremely difficult. Until 1989, no more than 20 or 30 consecutive amino acids were identified by Edman degradation [2].

Several authors have, however, proposed structural models: Allen's 'windmill' model in which four highly glycosylated subunits were linked together by a link peptide [3]; a flexible-thread model without link peptide, presented by Carlstedt [4] and Slayter [5], in which subunits were attached end-to-end by disulphide bridges. These structural models have been constructed to explain the diverse physico-chemical properties of these molecules: abrupt loss of viscosity of mucus gel observed after contact with reducing agents, which cleave disulphide bonds between cysteine residues [6], possible existence of a central link protein [7] and sensitivity to different proteolytic enzymes [8]. Although the arrangement and the size of the mucin subunits are controversial, most authors agree that the peptide cores contain heavily glycosylated domains (resistant to enzymic attack and rich in hydroxy amino acid residues) and also poorly glycosylated regions (so called 'naked regions') mainly consisting of hydrophobic amino acids such as leucine, isoleucine and valine and which are sensitive to proteolytic enzymes.

The isolation of mucin cDNA clones has introduced a new view of the structure of mucins: a repetitive model involving numerous tandem repeats of a short peptide sequence comprising

a large part of the coding region. At present, all the published peptide sequences deduced from mucin cDNA clones (porcine submaxillary-gland mucin [9], breast- or pancreatic-tumour 'mucin-like' glycoproteins [10–12], human intestinal mucins [13,14] and human tracheobronchial mucins [15,16]) contain this repetitive peptide structure.

Preliminary information concerning two other short clones has been published elsewhere [16].

Here we report for the first time the complete peptide sequences deduced from four longer human tracheobronchial cDNA clones which clearly demonstrate a new type of peptide organization composed of two distinct alternating domains.

EXPERIMENTAL

RNA purification and Northern-blot analysis

Total RNA from different human mucosae was purified using the guanidine isothiocyanate/CsCl method [17]. Polyadenylated [poly(A)⁺] mRNAs were selected on two cycles of oligo(dT)–cellulose affinity chromatography using standard procedures.

RNA was fractionated through 0.9% agarose gel containing glyoxal/dimethyl sulphoxide. Electrophoresis buffer (10 mM sodium phosphate, pH 7.0) was recirculated. Transfer to reinforced cellulose nitrate membrane BA-S83 (Schleicher and Schuell) was achieved by capillary blotting. rRNAs (28S and 18S) and a mixture of RNAs from Bethesda Research Laboratories were both used as molecular-mass markers after membrane staining with Methylene Blue.

Probe hybridization was performed at 42 °C for 16 h with a mixture comprising 50% formamide, 3 × SSC (1 × SSC is 0.15 M NaCl/0.015 M sodium citrate) 10 µg/ml sheared salmon sperm DNA, 1 × Denhardt's solution (0.1 g Ficoll, 0.1 g polyvinylpyrrolidone and 0.1 g BSA in 500 ml of water), Hepes buffer (50 mM, pH 7.0), 12% dextran sulphate and a minimum of 10⁶ c.p.m. of radiolabelled probe/ml. The membranes were washed

Abbreviations used: poly(A)⁺, polyadenylated; 1 × SSC, 0.15 M NaCl/0.015 M sodium citrate; 1 × SSPE, 0.15 M NaCl/0.01 M NaH₂PO₄/0.001 M EDTA; T_m, 'melting' temperature; TFMS, trifluoromethane sulphonic acid.

‡ To whom correspondence should be addressed.

[illegible][illegible]

The asterisk (*) indicates a shift in the reading frame.

Three short oligonucleotides (16 or 17 bp) were hybridized to Northern blots under the following conditions. Prehybridization was carried out at 10 °C below the T_m for at least 6 h (the T_m of each oligonucleotide was for 56–58 °C) in 6 × SSPE (1 × SSPE is 0.1 g Ficoll, 0.1 g of polyvinylpyrrolidone and 0.1 g of BSA in 500 ml of water)/10 × Denhardt's solution/0.05 % SDS buffer. Hybridization was at the same temperature for 15 h in 6 × SSPE/10 × Denhardt's solution/0.5 % SDS/0.1 mg/ml Salmon sperm DNA/10 % dextran sulphate, using 1 pmol/ml 5'-end-radio-labeled probe ($[\gamma\text{-}^{32}\text{P}]\text{ATP}$). Post-hybridization washings were first carried out in 4 × SSPE/0.1 % SDS at 5 °C below the T_m of each oligonucleotide. To guarantee the signal specificity, more stringent temperature conditions were then used, namely a 5 min washing at 1 °C below the T_m .

Human bronchial sputum was collected from a single patient with blood group O suffering from chronic bronchitis. Pronase-resistant glycopeptides from mucins were separated by anion-exchange chromatography according to the acidic content of their carbohydrate chains. All details concerning preparation, purification and biochemical analysis of mucin glycopeptides were described in [16,18,19]. Five fractions, corresponding to neutral, acidic or very acidic glycopeptides, were selected and then were deglycosylated with trifluoromethanesulphonic acid (TFMS) in the presence of phenylalanine (as scavenger) and Triton X-100. The precise protocols adapted for each fraction were described previously [16,20]. Two New Zealand White rabbits were immunized with an antigenic mixture containing 100 μ g of each deglycosylated fraction (protocol of immunization was as in [16]).

cDNAs were prepared from normal human tracheobronchial mucosa poly(A)⁺ RNAs using the cDNA Synthesis System Plus manufactured by Amersham Corp. Synthesis of the first cDNA strand was done using oligo(dT)₁₂₋₁₈ as primer for reverse transcription, and the second cDNA strand was synthesized by using RNAase H, DNA polymerase I and T4 DNA polymerase.

Sequencing of cDNAs

The cDNA clones were sequenced by Sanger's method [21] using Sequenase T4 DNA polymerase under conditions recommended

Figure 2 Deduced amino acid sequences and amino acid compositions of (a) JUL 7, (b) JUL 10, (c) JER 57 and (d) JER 28

by the supplier (United States Biochemical Corp). The cDNA inserts were sequenced either by directional subcloning or by using exonuclease III-deleted clones. The length of deletions were estimated by gel electrophoresis. In each case sequences were determined on both strands.

Three oligonucleotide probes were chosen in particular domains of clones corresponding to degenerate domains:

NAU 23: 5'GGT GGC CGT GGG GCC A 3'

(this sequence was chosen to be able to hybridize to 568–585 nucleotides from JUL 7)

NAU 24: 5'T CGC AAT GGG GCG CAG T 3'

[this sequence was complementary to nucleotide sequence of JUL 10 (positions 115–131)]

NAU 24: 5' AC GGG TTG TGG GTG TGG 3'

[this sequence was complementary to nucleotide sequence of JER 57 (positions 1344–1360)]

Southern-blot analysis and chromosome mapping

Human genomic DNA was extracted from lymphocytes by SDS/proteinase K digestion and phenol/chloroform extraction. DNA was digested by the indicated enzymes and was fractionated

Figure 3 Hydropathy curves for JER 57, JUL 7, JUL 10 and JER 28

Black rectangles represent the potentially heavily glycosylated domains which are hydrophilic domains. Diagonally hatched rectangles indicate the potentially naked domains which are hydrophobic (closely spaced) or very hydrophobic (widely spaced) domains. The abscissa shows the numbers of amino acids.

by 1 %-agarose-gel electrophoresis. Southern blotting was performed on Nylon N⁺ membranes (Amersham Corp). Probes were labelled with [³²P]dCTP using a commercially available random-primed labelling kit (Boehringer Mannheim).

DNA samples of 25 human × rodent hybrids were digested with *Bam*HI, electrophoresed and transferred as described above. The complete procedure, including the source and description of hybrid cells, chromosome studies and *in situ* hybridization, was described previously [22,23].

Hydrophilic/hydrophobic index

Hydropathy curves were computed using the method described by Kyte and Doolittle [24].

RESULTS

Isolation of tracheobronchial mucin cDNAs: Immunological characterization

The same polyclonal rabbit antiserum which was previously used for the isolation of the tandem repeat probes from MUC-4 (the gene located on chromosome 3q29) [15,25] and MUC-5 (the gene located on chromosome 11p15) [16,20] was used in the present study. This immune serum was directed against a mixture of TFMS-deglycosylated mucin peptides. The extent of deglycosylation varied from 83 to 96 % [16]. The immune serum reacts strongly with the β -galactosidase fusion proteins produced by JER 57 and JER 28. As was previously done for MUC-4 and MUC-5C cDNA probes [20], purified antibodies were prepared from the immune serum by specific adsorption on these two fusion proteins. Purified antibodies from JER 57 and those from JER 28 were able to label mucous cells from human

submaxillary glands and both goblet cells and mucous cells from human bronchial (J. P. Audié, N. Porchet, A. Janin and J. P. Aubert, unpublished work) and tracheal [20] mucosae. Peroxidase labelling was only situated in the rough endoplasmic reticulum of the goblet cells and of the mucous cells. No serous cells were labelled. The secretion granules containing glycosylated mucins were not labelled, owing to the immune-serum specificity for mucin peptide cores.

Of the 20 positive clones identified in our original screening, two appeared to have distinctive structural features: they cross-hybridized with each other, but not with other clones, this constituting the third family of our sequence classification [16,20]. Furthermore, the fusion proteins were also shown to bind an anti-(respiratory mucin) monoclonal antibody which has been isolated independently using a quite different apomucin preparation [26]. The larger of these two clones, JER 57, was used as a radiolabelled probe to screen the λ gt 10 cDNA library, and two other clones, JUL 7 and JUL 10, were selected among other positive clones because of their large size. We then analysed the details of their structures.

The complete sequence of each clone was determined and shown on Figure 1. The insert sizes were 561 bp for JER 28, 1830 bp for JER 57, 1631 bp for JUL 7 and 991 bp for JUL 10.

The deduced amino acid sequences and the percentages of each amino acid are shown in Figure 2.

Comparison of the four deduced peptide sequences

The following common features were noticed: (i) hydroxy amino acids represent 34–41 % of the total amino acids, with a Ser/Thr ratio of 1:2; (ii) proline constitutes 11–18 % and histidine 10 % of the total amino acids; (iii) cysteine is present at a constant percentage of 1–1.5 %; these residues are dispersed along the

Table 1 Hydroxy amino acid content of each peptide domain from JER 57, JUL 7, JUL 10 and JER 28

According to hydropathy curves (Figure 3), each peptide sequence of each clone was cut into peptide domains. Arrows on Figure 3 indicate the bounds of each domain. The percentages of serine + threonine were calculated.

Clone	Domains (amino acid positions)	Hydrophilic (H) or hydrophobic (h)	Ser + Thr (%)
JER 57	1-99	H	25.2
	100-152	h	22.6
	153-284	H	52.9
	285-348	h	26.5
	349-478	Very h	10.6
	479-532	H	53.7
	533-543	h	36.3
JUL 7	544-610	H	53.7
	1-29	h	0.0
	30-354	H	49.4
	355-382	h	7.1
	383-437	Very h	10.9
JUL 10	438-544	H	50.4
	1-40	H	50.0
	41-62	h	4.5
	63-90	H	42.7
	91-135	h	26.6
	136-181	H	54.2
	182-204	h	4.3
	205-242	Very h	12.8
JER 28	243-313	H	40.8
	314-330	h	23.4
	1-95	H	45.1
	96-156	h	23.3
	157-187	H	48.3

Figure 4 Southern-blot analysis

Human genomic DNA blots were analysed by using JER 57 as a probe. DNA prepared from the lymphocytes of one healthy donor (J.P.A.) was digested with *EcoRI* (lane 1), *BamHI* (lane 2), *EcoRI-BamHI* (lane 3), *HindIII* (lane 4), *XbaI* (lane 5) or *PstI* (lane 6). A 20 µg sample of digested DNA was loaded per lane. The values to the left and right of the gels are sizes in kb.

sequences, except in the case of the N-terminal regions of JER 57 and JUL 7, where they are grouped; (iv) tryptophan, one of the most hydrophobic amino acids, is present at about 1% in three of the four cDNA clones, and at 2.3% in JER 57.

Figure 5 Nucleotide sequence analysis

The nucleotide sequences of the four cDNA clones were determined, and the nucleotide sequence representing the most frequent 87-base-pair tandem repeat was deduced. The Figure shows its sequence.

Besides common features in their amino acid composition, and the fact that no tandemly repeated sequences are observed, the four peptides show a similar internal organization in which an alternation of domains can be seen.

Figure 3 shows the hydrophilic/hydrophobic index of peptide stretches and demonstrates that each sequence can be cut into hydrophilic and hydrophobic zones which alternate somewhat. For example, JUL 7 contains two large hydrophilic domains and two short hydrophobic domains. The JUL 10 clone, in contrast, shows shorter hydrophilic/hydrophobic zones. In the JER 57 clone, the three hydrophobic domains exhibit very variable size (from 10 to 193 amino acids). As regards JER 28, the shortest clone, only a few domains are detectable, two hydrophilic and one hydrophobic (Figure 3).

A consideration of the percentage of serine and threonine of each domain for each clone (Table 1) indicates that the hydrophilic domains are potentially heavily glycosylated, with 47.3% of serine and threonine on average (range 25.4–54.2). The hydrophobic domains are probably less heavily glycosylated, with 15.7% of serine and threonine on average (range 0–33.3).

Hydrophilic domains vary in size from 20 to 357 amino acid residues, although domains located close to N- or C-terminal part of the clone might be longer on the corresponding full-length mRNA. Hydrophobic or 'naked' regions vary from 33 to 203 amino acids.

A third point worth noting is the striking partition of some residues like proline and histidine. In the four clones, polyproline structures comprising three to six consecutive proline residues are only found in hydrophobic domains. The same observation can be made for histidine residues, which are also mainly found in the hydrophobic domains, in some cases as polyhistidine structures of two to four consecutive residues. All the cysteine residues of each clone are located in hydrophobic domains. The presence of five cysteine residues in the N-terminal hydrophilic domains of JER 57 is an exception to this rule.

Nucleotide sequence analysis of mucin cDNAs

(1) Southern-blot analysis, chromosomal localization and *in situ* hybridization

Human genomic DNA from the lymphocytes of nine healthy volunteers was digested with *EcoRI*, *BamHI*, *HindIII*, *XbaI* or *PstI*. In each instance the four cDNA probes hybridized to the same restriction fragments (Figure 4): a single band > 30 kb was seen with *EcoRI*, digestion with *BamHI* revealed three fragments at 1.8, 1.3 and 0.3 kb, a single 25 kb fragment was seen with *HindIII*, a single broad band > 30 kb with *XbaI* and then four fragments were cut by *PstI*, the highest at 1.4 kb, a broad band at 1–1.2 kb and two small bands at 0.9 and 0.7 kb.

In order to map these four cDNAs, hybridization to a panel of 25 independent human × rodent somatic-cell hybrids was performed. No cross-hybridization with hamster or mouse DNA was observed. There was a complete correlation between the

Table 2 Degeneracy into tandem-repeat nucleotide sequences

Each tandem repeat of each clone was compared with the virtually 87-base-pair most frequent tandem repeat. All nucleotide insertions or deletions are listed.

Clone	Nucleotide position	Number of nucleotides inserted (i) or deleted (d)	Reading frameshift	Amino acid position	Peptide hydrophilic or hydrophobic (h) alternation
JER 57	295	1 i	+	99	H → h
	456	1 d	+	152	h → H
	516	9 d	0	172	0
	699	19 d	+	233	0
	701	1 i	+	234	0
	852	1 d	+	284	H → h
	872	3 d	0	291	0
	1043	1 i	+	348	h → more h
	1098	13 i	+	366	h → more h
	1155	27 d	0	385	0
	1258	16 d	+	420	0
	1345	15 i	0	449	0
	1437	4 d	+	479	h → H
	1594	2 d	+	532	H → h
	1627	4 d	+	543	h → H
	1638	3 i	0	546	0
	1767	9 d	0	589	0
JUL 7	87	29 d	+	29	h → H
	182	21 d	0	61	0
	371	9 d	0	124	0
	832	8 d	+	278	0
	932	10 d	+	311	0
	1061	1 d	+	354	H → h
	1145	16 d	+	382	h → more h
	1223	15 i	0	408	0
	1311	4 d	+	437	h → H
JUL 10	118	1 i	+	40	H → h
	186	1 d	+	62	h → H
	269	2 d	+	90	H → h
	405	1 d	+	135	h → H
	441	9 d	0	147	0
	541	1 d	+	181	H → h
	610	16 d	+	204	h → more h
	726	1 d	+	242	h → H
	773	4 d	+	258	0
	823	1 i	+	275	0
	937	1 i	+	313	H → h
JER 28	115	3 d	0	39	0
	285	1 d	+	95	H → h
	314	15 d	0	105	0
	394	1 d	+	132	0
	468	16 d	+	156	h → H

presence of *Bam*HI hybridization bands and the presence of human chromosome 11 in all hybrids. So the four sequences were localized on chromosome 11. Complete analysis, including *in situ* hybridization, was carried out with the JER 57 probe, which was assigned to 11p15. This work has been described in detail elsewhere [23].

(2) Comparison of the nucleotide sequences

The four nucleotide sequences showed a high level of similarity (Figure 1) [from 71% (JUL 10 compared with JER 57) to 96% (JER 28 compared with JER 57)].

In spite of the fact that no repetitive peptide sequence could be deduced from the nucleotide sequences, we observed that some nucleotides were constantly repeated in the sequence, for example three consecutive triplets, CCC–CTC–CTC. Using these obser-

ventions, we were able to show that the sequences of the four clones were based upon a succession of imperfect tandem repeats (Figure 1). Compiling the 55 complete tandem repeat sequences, the most frequent is an 87-base-pair structure, its sequence is shown on Figure 5. However, only 22 units have the perfect 87-base-pair sequence, 27 are shorter because of nucleotide deletions and six are longer because of nucleotide insertions. The shortest consists of 25 nucleotides, the longest of 102 nucleotides.

By analysing the clones in this way, one can make the following observations about each clone: JER 28 consists of seven tandem repeats. Four nucleotide deletions were detected (Figure 1d). The first one (following nucleotide position 115) results in the deletion of a threonine residue. The three other deletions introduce a shift in reading frame that results in the substitution of many serine and threonine residues by proline and histidine residues.

In the three other clones, most of nucleotide insertions or

Figure 6 RNA blot analysis

Poly(A)⁺ RNA samples (2 µg) from human mucosae were analysed as described in the Experimental section. In (a), bronchial mucosa was probed with NAU 24 (lane 1), NAU 28 (lane 2) and 89 bp *EcoRI*–*PstI* fragment from JER 57 (lane 3). The autoradiographic exposure time was 15 h. In (b), tracheal mucosa (lane 1), bronchial mucosa (lane 2), gastric mucosae (fundus in lane 3 and antrum in lane 4), colonic mucosae (ascending colon in lane 5, descending colon in lane 6 and transverse colon in lane 7) were probed with JER 57. Hybridization signals were obtained after autoradiographic exposure time(s) of 3 h for respiratory mucosae samples and 24 h for digestive-mucosae samples. In (c), bronchial mucosa was probed with β -actin. Sizes are given in kb, and the positions of rRNAs are given (28 S and 18 S). SAP, sample application point.

deletions introduce a shift (*) in the reading frame (Figure 1). Deletions are more frequent modifications than insertions. Variable numbers of nucleotides are deleted (from 1 to 62) or inserted (from 1 to 15). Most modifications in the nucleotide tandem repeat create shifts in reading frame: out of 42 modifications, 30 shifts in reading frame occur. Four shifts in the reading frame nullify each other, in two cases because of a deletion immediately followed by an insertion which restores the reading frame. Comparison of the bounds of each domain on hydropathic curves of the proteins and the nucleotide sequence reveals that, out of 26 shifts in the reading frames (arrows on hydropathic curves), 21 coincide with inversion in hydropathic curves (Table 2).

The first 204 nucleotides of JER 57 do not exhibit any sequence similarity with the 87-base-pair tandem repeat and may be considered as a very long inserted structure, but because this structure coincides with the N-terminal part of the clone, it may be incomplete.

RNA blot analysis

A 2 µg sample of poly(A)⁺ RNA from different human mucosae [tracheal, bronchial, gastric (antrum and fundus) and colonic (ascending colon, transverse colon, descending colon)] were subjected to Northern-blot analysis. The same hybridization signal was obtained on each blot with the four cDNA probes. In all cases, the message was uniformly polydisperse in size (up 20 kb to 1 kb). A very strong signal after only 2 h of autoradiography was observed for tracheal or bronchial mucosae. A longer exposure time was required (from 24 to 48 h) for the digestive mucosae tested. The signal was resistant to very stringent washing conditions (1 × SSC/0.1 % SDS, 65 °C).

Several other human tissues (heart, brain, placenta, lung, liver, skeletal muscle, kidney and pancreas) were also tested in the same way, using a Northern blot obtained from Clontech Laboratories. No hybridization signal could be observed, even after 1 week's exposure (results not shown).

An *EcoRI*–*PstI* fragment from JER 57, which comprises the

first 89 nucleotides, was also used to probe the bronchial-mucosa Northern blot and produced the same hybridization signal as the complete probe (Figure 6).

Three short oligonucleotides chosen to be able to recognize a perfect 87-base-pair tandem repeat (NAU 23) or inserted domains of degenerate tandem repeats (NAU 24 and NAU 28) produce the same polydisperse hybridization signal as the corresponding entire probes (Figure 6).

'Homology search'

Computer searches for similarities of the deduced amino acid sequences with proteins registered in the University of Geneva Protein (Swiss-Prot) data bank revealed no statistically significant sequence similarity with any protein, including MUC 1 [10,12], MUC 2 and 3 [13,14], MUC 4 and 5 [15,16] and porcine [9] or bovine [32] submaxillary mucins.

DISCUSSION

Here we report the complete analysis of four mucin cDNAs which all map to chromosome 11p15 and which constitute a new mucin peptide family. The special feature of these peptides is the complete lack of tandem-repeat organization, in spite of the fact that nucleotide sequence contains numerous internal repeats. The peptide structure in each case appears as an irregular alternation of potentially heavily glycosylated domains and potentially naked domains. The 20 clones originally isolated from a λ gt 11 cDNA library can be classified into four families: JER 64 (called 'MUC 4' in human gene mapping nomenclature) was mapped to 3q29 [15,25]; the three other clone families all mapped to 11p15 [27]. Complete analysis of one of them is presented here. This family comprises four clones: JER 28 and JER 57 (from the λ gt 11 cDNA library) and JUL 7 and JUL 10 (from the λ gt 10 cDNA library). There are six lines of evidence that these four clones encode human mucins.

(1) The immune serum used for their isolation was raised against human tracheobronchial apo-mucins. It is unlikely that the immune serum contains significant levels of antibodies directed against other tracheobronchial proteins, because of the extensive Pronase digestion and drastic TFMS treatment which are used in the preparation of the apo-mucins.

(2) Purified antibodies prepared from fusion proteins recognized only mucous cells and goblet cells from tracheal and bronchial epithelia [20].

(3) Northern-blot analysis showed that the mRNA detected by the four cDNA clones was large and very polydisperse and characteristic of the patterns seen with other mucin probes. It was expressed in various mucosae, but not in the other tissues tested.

(4) The amino acid composition and, in particular, the high percentage of hydroxy amino acids in the four deduced peptide sequences is characteristic of mucins.

(5) The probes map to chromosome 11p15, which is the same region to which two other mucin cDNA families map: intestinal cDNAs encoding for 23-amino-acid tandem-repeat peptides and products of the gene MUC 2 [13], and tracheobronchial cDNAs encoding for 8-amino-acid tandem-repeat peptides [23].

(6) Nucleotide sequences of the four cDNAs presented here show the typical internal tandem-repeat organization described for all animal or human mucin structures.

However, the peptide sequences deduced from cDNAs of this mucin family contrast with others previously described by a complete lack of tandem repeats. This is due to numerous

deletions or insertions in a repetitive nucleotide structure which introduce shifts in reading frame.

A 89 bp DNA probe and 16 or 17 bp oligonucleotides which all only recognized inserted regions from degenerate tandem repeats gave on RNA blots the same polydisperse signal as the entire four cDNA probes.

This clearly indicates that these nucleotide structures are also present in bronchial-mucosa RNA from another individual and not only in our cDNA library.

This phenomenon raises several questions: why is there such degeneracy? Is it related to biological functions within the mucosa? Is it related to the position of these cDNA stretches within the corresponding full-length mRNA, as was previously observed for MUC 1 [28]? In this case the full-length protein may be composed of a repetitive domain with non-repetitive domains located in the N- and C-terminal regions. If this is the case, the full-length mRNA must be very long.

The existence of alternating hydrophilic and hydrophobic domains is perfectly consistent with structural studies on mucins: Sheehan and Carlstedt [29] have shown, by electron microscopy, that antibodies raised against naked domains bind to cervical-mucin molecules in an irregular manner. The same observation has been made for gastric, bronchial and salivary mucins [30]. Furthermore, Shankar et al. [31] have provided evidence for the presence of hydrophobic zones in human airway mucins by the use of a fluorescent hydrophobic probe.

Thus, for the first time, the nucleotide sequences of this mucin family are consistent with a classical structural model of mucin proteins.

It is noteworthy that these new mucin sequences map to the same region of chromosome 11 as the (two) other classes of mucin clones we have characterized and also the MUC 2 gene. The lack of sequence identity with MUC 2 and MUC 5C and difference in tissue expression indicates that none of the tracheo-bronchial mucin clones are part of MUC 2 and suggest that at least one more gene (MUC 5B) is located in this region. However, it is not yet clear whether all the classes of tracheobronchial mucin clones which map to 11p15 are products of a single or several genes, or how large the region is. We hope that restriction mapping of large-size DNA fragments using pulse field gel electrophoresis will help to answer these questions.

This work was supported by the Association pour la Recherche sur le Cancer. We thank Danièle Petitprez and Pascal Mathon for their excellent technical assistance, Dr. Dallas Swallow for helpful discussions and reading the manuscript, and Dr. Nguyen Van Cong for chromosomal localization.

REFERENCES

- 1 Strous, G. J. and Dekker, J. (1992) *Crit. Rev. Biochem. Mol.* **27**, 57–92
- 2 Rose, M. C., Kaufman, B. and Martin, B. (1989) *J. Biol. Chem.* **264**, 8193–8199
- 3 Allen, A. (1981) in *Structure and Function of Gastrointestinal Mucus* (Johnson, L. R., ed.) pp. 617–639, Raven Press, New York
- 4 Carlstedt, I., Lindgren, H., Sheehan, J. K., Ulmsen, U. and Wingerup, L. (1983) *Biochem. J.* **213**, 427–435
- 5 Slayter, H. S., Lamblin, G., Le Treut, A., Galabert, C., Houdret, N., Degand, P. and Roussel, P. (1984) *Eur. J. Biochem.* **142**, 209–218
- 6 Allen, A. and Snary, D. (1972) *Gut* **13**, 666–672
- 7 Slomiany, A., Okazaki, K., Tamura, S. and Slomiany, B. L. (1991) *Arch. Biochem. Biophys.* **286**, 383–391
- 8 Scawen, M. and Allen, A. (1977) *Biochem. J.* **163**, 363–368
- 9 Timpl, C. S., Eckhardt, A. E., Abernethy, J. L. and Hill, R. L. (1988) *J. Biol. Chem.* **263**, 1081–1088
- 10 Gendler, S., Taylor-Papadimitriou, J., Duhig, T., Rothbard, J. and Burchell, J. (1988) *J. Biol. Chem.* **263**, 12820–12823
- 11 Siddiqui, J., Abe, M., Hayes, D., Shani, E., Yunis, E. and Kufe, D. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 2320–2323
- 12 Lan, M. S., Batra, S. K., Qi, W. N., Mezgar, R. S. and Hollingsworth, M. A. (1990) *J. Biol. Chem.* **265**, 15294–15299
- 13 Gum, J. R., Byrd, J. C., Hicks, J. W., Toribara, N. W., Lamport, D. T. A. and Kim, Y. S. (1989) *J. Biol. Chem.* **264**, 6480–6487
- 14 Gum, J. R., Hicks, J. W., Swallow, D. M., Lagace, R. L., Byrd, J. C., Lamport, D. T. A., Siddiqui, B. and Kim, Y. S. (1990) *Biochem. Biophys. Res. Commun.* **171**, 407–415
- 15 Porchet, N., Nguyen, V. C., Dufosse, J., Audie, J. P., Guyonnet-Duperat, V., Gross, M. S., Denis, C., Degand, P., Bernheim, A. and Aubert, J. P. (1991) *Biochem. Biophys. Res. Commun.* **175**, 414–422
- 16 Aubert, J. P., Porchet, N., Crepin, M., Duterque-Coquilaud, M., Vergnes, G., Mazzuca, M., Debuire, B., Petitprez, D. and Degand, P. (1991) *Am. J. Respir. Cell. Mol. Biol.* **5**, 178–185
- 17 Glisin, V., Orkvenjakov, R. and Byus, C. (1974) *Biochemistry* **13**, 2633–2637
- 18 Degand, P., Roussel, P. and Lamblin, G. (1973) *Biochem. Biophys. Acta* **320**, 318–330
- 19 Lamblin, G., Degand, P., Roussel, P., Havez, R., Hartemann, E. and Filliat, M. (1972) *Clin. Chim. Acta* **36**, 329–340
- 20 Crepin, M., Porchet, N., Aubert, J. P. and Degand, P. (1990) *Biorheology* **27**, 471–484
- 21 Sanger, F., Nicklen, S. and Colson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463–5467
- 22 Nguyen, V. C., Weil, D., Finaz, C., Cohen-Haguenaue, R., Gross, M. S., Jegou-Foubert, C., Tand, M. F., Cochet, C., Grouchy, J. and Frezal, J. (1986) *Ann. Génét. (Paris)* **29**, 20–26
- 23 Nguyen, V. C., Aubert, J. P., Gross, M. S., Porchet, N., Degand, P. and Frezal, J. (1990) *Hum. Genet.* **86**, 167–172
- 24 Kyte, J. and Doolittle, R. F. (1982) *J. Mol. Biol.* **157**, 105–132
- 25 Gross, M. S., Guyonnet-Duperat, V., Porchet, N., Bernheim, A., Aubert, J. P. and Nguyen, V. C. (1992) *Ann. Génét. (Paris)* **35**, 21–26
- 26 Perini, J. M., Vandamme-Cubadda, N., Aubert, J. P., Porchet, N., Mazucca, M., Lamblin, G., Hercovics, A. and Roussel, P. (1991) *Eur. J. Biochem.* **196**, 321–328
- 27 Porchet, N., Dufosse, J., Audie, J. P., Guyonnet-Duperat, V., Perini, J. M., Nguyen, V. C., Degand, P. and Aubert, J. P. (1991) *Am. Rev. Respir. Dis.* **144**, 15–18
- 28 Ligtenberg, M. J. L., Vos, H. L., Gennissen, M. C. and Hilken, J. (1990) *J. Biol. Chem.* **265**, 5573–5578
- 29 Sheehan, J. K. and Carlstedt, I. (1990) *Biochem. J.* **265**, 169–178
- 30 Sheehan, J. K., Boot-Handford, R. P., Chantler, E., Carlstedt, I. and Thornton, D. J. (1991) *Biochem. J.* **274**, 293–296
- 31 Shankar, V., Naziruddin, B., Reyes De La Rocha, S. and Sachdev, G. P. (1990) *Biochemistry* **29**, 5856–5864
- 32 Bhargava, A. K., Weitach, H. T., Davidson, E. A., Bhavanandan, V. P. (1990) *Proc. Natl. Acad. Sci. U.S.A.* **87**, 6798–6802