Genomic organization of the 3'-region of the human *MUC5AC* mucin gene: additional evidence for a common ancestral gene for the 11p15.5 mucin gene family

Marie-Pierre BUISINE*†‡, Jean-Luc DESSEYN*, Nicole PORCHET*†‡, Pierre DEGAND*†‡, Anne LAINE* and Jean-Pierre AUBERT*†¹

*INSERM Unité 377, Place de Verdun, 59045 Lille Cedex, France, †Laboratoire de Biochimie et Biologie Moléculaire, Hôpital C. Huriez, Centre Hospitalier Régional et Universitaire, 59037 Lille Cedex, France, and ‡Faculté de Médecine, Université de Lille II, 59045 Lille Cedex, France

The human mucin gene *MUC5AC* is mapped clustered with *MUC2*, *MUC5B* and *MUC6* on chromosome 11p15.5. We report here the isolation and characterization of a genomic cosmid clone, designated ELO9, spanning the 3'-region of *MUC5AC* and the 5'-region of *MUC5B*, allowing us to conclude that *MUC5AC* and *MUC5B* have the same transcriptional orientation. We determined the genomic organization and the entire

sequence of the 3'-region of *MUC5AC*. The comparative molecular analysis of *MUC5AC* and *MUC5B* points to a remarkable similarity in the size and the distribution of exons, and in the type of splice sites, supporting the notion that *MUC5AC* and *MUC5B* have evolved from a single common ancestral gene. The derivation of the four genes of the 11p15.5 mucin gene family from a single ancestral gene is discussed.

INTRODUCTION

The epithelial surface of the respiratory, gastrointestinal and reproductive tracts is coated with mucus which is secreted by specialized epithelial cells, e.g. goblet cells and submucosal gland cells. Mucus secretions provide important protective and lubricative functions varying among the tissues. Most of the properties of mucus have been attributed to mucins.

To date, nine human mucin genes (*MUC1*, *MUC2*, *MUC3*, *MUC4*, *MUC5AC*, *MUC5B*, *MUC6*, *MUC7*, *MUC8*) have been identified (for reviews, see [1,2]). Additional partial cDNAs have also been described, including NP3a which has been reported to be part of *MUC5* [3], and L31 and HGM-1 which have been reported to be part of *MUC5AC* [4,5]. Four of the nine mucin genes, *MUC2*, *MUC5AC*, *MUC5B*, and *MUC6*, have been mapped to chromosome 11p15.5 [6–8]. Recently the gene order has been determined to be *MUC6-MUC2-MUC5AC-MUC5B* from the telomere to the centromere [9]. All mucin genes share common features, including tandemly repeated sequences flanked by non-repeat regions. Tandem repeat units vary in length from eight amino acid residues in *MUC5AC* [10] to 169 amino acid residues in *MUC6* [8]. They encode peptides rich in threonine and serine which support the numerous O-glycan chains. Cysteine-rich domains have been reported in the N- and C-terminal regions of *MUC2*, the C-terminal region of *MUC5B*, the C-terminal region of *MUC6*, in NP3a, L31 and HGM-1 [3–5,11,13,14]. The C-terminal regions of *MUC2* and *MUC5B*, NP3a and L31 exhibit striking sequence similarities with the D4, B, C and CK domains of the human von Willebrand factor (vWF) [3,4,10,11,13,15]. Recently, the C-terminal region of *MUC6* has been reported to contain only the CK-vWF-like domain [14]. Other cysteine-rich domains, designated cysteine-rich subdomains, have been reported in the central repetitive domains of *MUC5AC* and *MUC5B* [10,12].

Qualitative and quantitative alterations in the expression of the *MUC5AC* gene have been reported in both preneoplastic and

neoplastic lesions. We have shown that *MUC5AC* is expressed in rectosigmoid villous adenomas but is absent from normal intestine and colon cancers [16]. The expression level of *MUC5AC* in rectosigmoid villous adenomas is correlated to the degree of dysplasia [16]. Moreover, *MUC5AC* is expressed in embryonic and foetal intestine [16a]. Likewise, *MUC5AC* mRNAs are detectable in pancreatic cancers but not in normal pancreas [17].

Although *MUC5AC* and *MUC5B* have been shown by physical mapping and expression pattern to be distinct mucin genes [10], confusion has been introduced in the nomenclature with the cloning of a new cDNA NP3a that has been designated as *MUC5* [3].

In the present study, we report the organization and the complete nucleotide sequence of the 3'-region of the human *MUC5AC* gene and provide evidence that the previously described cDNA clones L31 [4] and NP3a [3] correspond to the 3'-end of *MUC5AC*. We compared the 3'-regions of *MUC5AC*, *MUC5B*, *MUC2* and *MUC6*. The comparative molecular analysis of *MUC5AC* and *MUC5B* revealed striking common features, supporting the notion that they have evolved from a single common ancestral gene. The derivation of the four genes of the 11p15.5 mucin gene family from a single ancestral gene is discussed.

EXPERIMENTAL

Oligonucleotide primers

Oligonucleotide primers used in amplification and sequence experiments were synthesized by Eurogentec (Liège, Belgium).

Reverse transcription-PCR (RT-PCR) and amplification

Human tracheal poly(A⁺) RNA (Clontech) (0.4 µg) was reverse transcribed using the first-strand cDNA synthesis kit (Clontech)

Abbreviations used: vWF, von Willebrand factor; RT-PCR, reverse transcription-PCR.

¹ To whom correspondence should be addressed. (e-mail jpa@lille.inserm.fr).

The nucleotide sequences reported in this paper have been submitted to the EMBL Data Bank with accession numbers AJ001402 and AJ001403.

Figure 1 Partial restriction map of ELO9 and sequencing strategy for the 3'-region of *MUC5AC*

(A) ELO9 and the *MUC5B* genomic clones BEN1 and BEN2 are indicated. The orientation of the DNA fragments is 5' → 3' from the left to the right. The thicker line represents the central exon of *MUC5B*. B, *Bam*HI; Bg, *Bgl*II; Bss, *Bss*HII; H, *Hind*III; E, *Eco*RI; K, *Kpn*I; N, *Not*I; S, *Sma*I; SI, *Sac*I; SII, *Sac*II; X, *Xho*I. All inserts are flanked by *Not*I sites of the cosmids indicated by N*. (B) Sequence strategy for the 3'-region of *MUC5AC*. Restriction fragments and inserts generated by amplification are shown. They were entirely sequenced on both strands. Horizontal small open arrows indicate the location and direction of the two primers NAU171 and NAU118. Direct sequencing was also performed; horizontal arrows indicate the length and direction of individual sequencing reactions. Filled rectangles indicate the location of specific primers and are not to scale.

and random primers according to the manufacturer's instructions. Amplifications were performed in a Perkin-Elmer Thermal Cycler 480. PCR amplification reactions were conducted in 50 μ l reaction volumes containing 5 μ l of 10 $\times$ buffer (100 mM Tris/HCl/15 mM MgCl₂/500 mM KCl, pH 8.3), 4 μ l of 10 mM deoxynucleoside triphosphates, 1–8 μ l of first-strand tracheal cDNA, 10 pmol of each primer and 2 units of *Taq* DNA polymerase (Boehringer-Mannheim). After overlaying with 50 μ l of mineral oil, the mixture was denatured at 94 °C for 4 min followed by 30 cycles at 94 °C for 45 s, 58–60 °C for 45 s and 72 °C for 2 min. The final elongation step was extended for an additional 15 min. The amplified products were purified using Preps DNA purification resin (Promega) and cloned into pCR2.1 vector with the TA cloning kit (Invitrogen) according to the manufacturer's instructions. Clones were sequenced as described below. The first RT-PCR reaction was conducted on tracheal cDNA using sense primer NAU171 (5'-ACCACCAGCACA-ACCTCTGCT-3') located in tandemly repeated sequence of *MUC5AC* (EMBL Data Bank numbers Z34277, Z34278 [10]) with antisense primer NAU118 (5'-GATGCGCGAGACCA-CGATGCC-3'), the sequence of which was deduced from the NP3a cDNA (nt 1211–1231) (EMBL Data Bank number U06711 [3]) and from the L31 cDNA (nt 942–962) (EMBL Data Bank number Z48314), which has been reported to be a putative 3'-end for the *MUC5AC* gene [4]. A 1.3-kb product was obtained and

subcloned. The corresponding cDNA clone was designated TH171-118. Oligonucleotide sequences for further RT-PCR experiments were deduced from the cosmid clone sequence (see below).

Library screening

A human placenta genomic DNA library in cosmid vector pWE15 (Stratagene) was screened using the cDNA probe TH171-118 generated by RT-PCR as described previously [12]. One positive clone designated ELO9 was isolated and characterized.

Characterization of the genomic organization of the cosmid clone

The restriction mapping of ELO9 was conducted as described previously [12]. Some restriction fragments were subcloned into the pBluescript KS(+) vector (Stratagene). Other fragments were generated by PCR using 20 ng of cosmid as template and subcloned into pCR2.1 vector (Invitrogen). Clones were sequenced as described below.

Amplification on genomic DNA

PCR amplification reactions were conducted as described above

using 500 ng of genomic DNA as template. Oligonucleotide sequences were deduced from the cDNA sequence.

Sequencing of DNA fragments and data analyses

Clones were sequenced on both strands using the Sequenase 2.0 kit (U.S. Biochemical Corp.) according to the protocol indicated by the manufacturer, using [α - 35 S]dATP (Amersham) and either universal primers or specific internal oligonucleotides. Direct sequencing was also performed on the cosmid clone using 1.5 μ g of DNA as described previously [12]. Sequences were analysed with PC/Gene software (IntelliGenetics Inc.).

RESULTS

Cloning and characterization of the *MUC5AC*-specific cDNA probe TH171-118

Since *MUC5AC* is highly expressed in trachea and has been initially isolated from a human tracheobronchial cDNA library [18], RT-PCR reactions were performed with human tracheal RNA. The first amplification was conducted using sense primer NAU171 with antisense primer NAU118 (see the Experimental section). A 1.3-kb cDNA product was obtained and designated TH171-118. The nucleotide sequence is available from EMBL Data Bank under accession number AJ001402 (nt 1–1353). TH171-118 contains four *MUC5AC*-type tandem repeats, two degenerate repeats of eight amino acids followed by a unique cysteine-rich sequence. The unique sequence was identical, except for one nucleotide substitution, with clone L31 (nt 1–962) isolated by Lesuffleur et al. [4] from an HT29-MTX expression library using a polyclonal serum specific for normal gastric mucosa, and was highly similar, except for three nucleotide deletions, three nucleotide inversions and four nucleotide substitutions, to clone NP3a (nt 8–1231) isolated by Meerzaman et al. [3] from a human nasal polyp cDNA library using a probe for human tracheobronchial mucin glycoprotein (TR-3A tryptic peptide).

Isolation and characterization of the genomic DNA clone ELO9

The *MUC5AC* specific cDNA probe TH171-118 was used to screen a cosmid genomic DNA library. Only one positive clone containing an insert of approx. 40 kb and designated ELO9 was isolated and characterized. The partial restriction map of ELO9 is shown in Figure 1(A). ELO9 contains a CpG island located at 1.6 kb from the 3'-end of the insert. Restriction-enzyme-digested DNA of ELO9 hybridized with several probes, including the TH171-118 *MUC5AC*-specific probe and probes corresponding to unique regions of the *MUC5B* genomic clone BEN1, e.g. the 1.3-kb *NotI*–*EcoRI* fragment located in the 5'-end of BEN1 and the 0.5-kb *BssHII*–*NotI* fragment located within the I8 CpG island reported by Desseyn et al. [12] (see their location in Figure 1A). ELO9 overlaps the 5'-region of the *MUC5B* genomic clone BEN1 on approx. 25 kb, as ascertained by restriction-map analysis and Southern-blot hybridization. In fact, the CpG island in ELO9 corresponds to the I10 CpG island on the macrocartography performed in the 11p15.5 region [9]. The 1.3-kb *NotI*–*EcoRI* fragment from BEN1 hybridized to the 12-kb *NotI*–*EcoRI* fragment located in the 5'-end of ELO9. The 0.5-kb *BssHII*–*NotI* fragment from BEN1 hybridized to the 0.5-kb *BssHII*–*NotI* fragment located upstream of the CpG island within ELO9 (Figure 1A). Our results indicate that ELO9 spans the 3'-region of *MUC5AC* and the 5'-region of *MUC5B*.

Thus, *MUC5AC* and *MUC5B* have the same transcriptional orientation.

Sequence data and analysis of the 3'-region of the *MUC5AC* tracheal cDNA

To further determine the genomic organization of the 3'-region of *MUC5AC*, specific oligonucleotides (Figure 1B) were synthesized and used to clone the corresponding tracheal cDNA portion by RT-PCR. The sequences of the specific oligonucleotides were deduced from ELO9, NP3a [3] and L31 [4]. The sequences obtained were compiled with that of clone TH171-118 yielding a sequence of 3681 bp, with a continuous reading frame of 3243 bp and a 3'-untranslated region of 438 bp. The open reading frame encoded a 1227-amino-acid peptide containing five major domains: (i) a domain composed of four *MUC5AC*-type repeats of eight amino acid residues, (ii) a unique conserved domain with an atypical repeat, (iii) a first cysteine-rich domain (10 cysteine residues/110 amino acids), (iv) a domain of 50 amino acid residues rich in serine, threonine, alanine, proline and valine containing four degenerate SSSVA tandem repeats, and (v) a unique sequence of 874 amino acid residues rich in cysteine residues (85 cysteine residues, 10% of sequence). The first three domains are likely to correspond to a part of the last super-repeat of the central exon of *MUC5AC*, as defined for *MUC5B* [12]. The first cysteine-rich subdomain contains the TR-3A tryptic peptide (WFDVDFPSPGPHGGDKETYNNTI) which was initially isolated from human lung mucus [19]. The TR-3A tryptic peptide was found in clone JER47, which was isolated from a tracheobronchial cDNA library [18] and belonged to *MUC5AC* [10]. The last two domains correspond to the so-called C-terminal region of the tracheal *MUC5AC* mucin. The last domain contains the MUC11p15-type and the A3uD4 subdomains described by Desseyn et al. [13], and the D4-, B-, C- and CK-vWF-like subdomains. The cDNA has a CATAAA polyadenylation signal [20] 413 bp downstream from the stop codon.

Comparative sequence analyses of the 3'-region of mucin gene cDNAs and deduced peptides

Comparative sequence analyses revealed that the 3'-region of the *MUC5AC* tracheal cDNA was nearly identical (99.7% identity) to L31 [4], except for seven nucleotide substitutions and one nucleotide inversion. The 3'-region of the *MUC5AC* tracheal cDNA also displayed a high percentage identity (98.7%) with NP3a at the nucleotide level [3], but there was more divergences at the deduced peptide level resulting from nucleotide substitutions and inversions and, above all, from nucleotide insertions and deletions introducing shifts in the reading frame. Moreover, an additional C in the *MUC5AC* tracheal cDNA compared with L31 and NP3a resulted in an early stop codon, and thus in a shorter C-terminal region than L31 and NP3a. The first in-frame stop codon (TGA), located at position nt 3244, is closer to that of *MUC2* [11] and *MUC5B* [13] (in black boxes in Figure 2). As described in L31 [4], the deduced amino acid sequence of the 3'-region of the *MUC5AC* tracheal cDNA contains 12 typical potential N-glycosylation sites (Asn-Xaa-Thr/Ser) (underlined in Figure 2).

MUC5AC differed from *MUC2* [11], *MUC5B* [12,13] and *MUC6* [14], since the last super-repeat of the central exon of *MUC5AC* ended with a cysteine-rich domain immediately downstream of the *MUC5AC*-tandem repeat. Figure 2 shows, however, the striking sequence similarities between the 3'-regions of *MUC5AC*, *MUC5B*, *MUC2* and *MUC6* at the nucleotide and

MUC5AC	TTC	TTC	TTC	AAC	GTG	GCT	- - -	GAC	CGG	CTC	TAC	CTT	GCA	A	TCC	ACC	669
MUC5B	TTC	TTC	TTC	AAG	GCA	TTT	- - -	GGA	CAG	TTT	TTC	TTC	TTC	GGA	GGA	GTC	121
MUC2	CTT	ATC	TTC	TTC	GTG	CTG	AAC	GAC	ACC	TAC	TAC	GCA	CCC	GGA	GGA	GAG	459
	Cys	Phe	Cys		Val			Asp			Tyr		Pro	Gly	Gln		
MUC5AC	ATA	TAC	CGC	CAC	AGA	GAC	CTC	GCT	GGC	CAT	TGC	TAT	- - -	TAT	GCC	CTG	714
MUC5B	ATC	TAC	AAT	AAG	ACA	GAC	CGA	GCT	GGC	- - -	TGC	CAT	- - -	TAT	GCA	GTG	16
MUC2	GTG	TAC	AAT	GGC	ACA	- - -	TAC	GGA	GAC	ACC	TAT	TAT	TTC	TAT	GTC	AAC	501
	Ile	Tyr	Asn		Thr	Asp		Ala	Gly		Cys	Tyr	Phe	Tyr	Ala		
MUC5AC	TGT	AGC	CAG	GAC	TGC	CAA	GTG	GTG	AGA	GGG	GTT	GAC	AGT	GAC	TGT	CCG	762
MUC5B	TGT	AAT	CAG	CAC	TGT	GAC	ATT	GAC	CGC	TTC	- - -	CAG	GGC	GCC	TGT	CCC	211
MUC2	TGT	TCA	CTG	AGC	TGT	ACG	TTG	GAG	- - -	TTC	TAT	AAC	TGG	TCC	TGT	CCA	546
	Cys	Ser	Gln		Cys			Arg	Phe						Cys	Pro	
MUC5AC	TCC	ACC	- - -	ACG	CTG	CCT	CCT	GCC	CCA	GCC	- - -	ACG	TCC	CCT	TCA	ATA	804
MUC5B	TCC	ACC	TCC	CCA	CTG	CCT	GTG	TCC	TCC	GCC	- - -	CTG	TCC	- - -	TCA	- - -	250
MUC2	TCC	ACC	- - -	CCC	TCC	CCA	ACA	- - -	CCC	ACC	CCC	- - -	TCC	AAG	TCA	ACG	585
	Ser	Thr		Pro		Pro			Pro	Ala	Pro		Ser		Ser		
MUC5AC	- - -	TCC	ACC	- - -	TCC	GAG	CCC	- - -	GTC	ACT	- - -	- - -	GAG	CTG	- - -	- - -	831
MUC5B	CCC	TCC	- - -	CCT	GCC	- - -	CCT	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	265
MUC2	CCC	- - -	ACC	CCT	TCC	AAG	CCA	TCC	ACC	CCC	CCC	TCC	AAG	CCG	ACG	CCC	630
	Pro	Ser	Thr	Pro	Ser		Pro		Thr								
MUC5AC	GGA	- - -	- - -	- - -	- - -	- - -	TGC	CCA	- - -	AAT	GCG	GTT	CCC	CCC	AGA	AAG	861
MUC5B	GGA	- - -	- - -	- - -	- - -	- - -	TGT	- - -	GAC	AAT	GCC	ATC	CCT	CTC	CGA	CA	295
MUC2	GGA	ACC	AAG	CCC	CCC	GAG	TGT	CCA	GAC	- - -	TTC	GAT	CCT	CCC	CGA	G	675
	Gly						Cys	Pro	Asp	Asn			Pro	Pro	Arg	Gln	
MUC5AC	AAA	GGT	GAG	ACC	GCC	TGG	ACA	CCC	- - -	AAC	TCC	TCC	GAG	GCC	ACC	TGT	906
MUC5B	GAG	AAT	GAG	ACC	- - -	TGG	ACC	CTG	GAG	AAC	TCC	ACG	GAG	GCC	ACC	TGT	340
MUC2	GAG	AAC	GAG	ACT	TGG	TGG	- - -	CTG	TCC	GAC	TCC	TTC	ATG	GCC	ACC	TGT	720
		Asn	Glu	Thr		Trp	Thr	Leu		Asn	Cys			Ala	Thr	Cys	
MUC5AC	GAG	GGC	AAC	AAC	- - -	GTC	- - -	ATC	TCC	CTG	- - -	AGC	CCG	CGC	ACG	TGT	945
MUC5B	GAG	GGT	AAC	AAC	CGT	GTC	- - -	- - -	CTG	CTG	CTG	GAG	CCA	AAG	- - -	- - -	376
MUC2	AAG	TAC	AAC	AAC	ACG	GTC	GAG	ATC	Val	AAG	GTC	GAG	- - -	- - -	- - -	TGT	759
		Gly	Asn	Asn		Val		Ile	Val	Leu			Pro			Cys	
MUC5AC	- - -	CCG	AGG	GTG	GAG	AAG	CCC	ACT	TGT	GCC	AAC	GGC	TAC	CCG	GCT	GTG	990
MUC5B	- - -	CCT	- - -	GTG	GCC	AAC	GTC	ACC	TGT	GTG	AAC	AAG	CAC	CTG	CCC	ATC	418
MUC2	GAG	CCG	CCG	- - -	CCC	ATG	CCC	ACC	TGT	TCC	Asn	Gly	CTC	CAA	CCC	GTG	804
		Pro		Val			Pro	Thr	Cys						Pro	Val	
MUC5AC	AAG	GTG	GCT	GAC	- - -	CAA	GAT	GGC	TGC	TGC	CAT	CAC	TAC	CAG	TGC	CAG	1035
MUC5B	AAG	GTG	TGC	GAC	CCC	AGC	CAG	CCC	TGT	GAC	TTC	CAC	TAT	GAG	TGC	GAG	466
MUC2	CGC	GTG	GAG	GAC	CCC	- - -	GAC	GGC	TGT	TGC	TGG	CAC	TAT	GAG	TGC	GAC	849
	Lys	Val		Asp	Pro		Asp	Gly	Cys	Cys		His	Tyr	Glu	Cys		
MUC5AC	TGT	GTG	TGC	AGC	GGC	TGG	GGT	GAC	CCC	CAC	TAC	ATC	ACC	TTC	GAC	GGC	1083
MUC5B	TGT	ATC	TGC	AGC	ATG	TGG	GGC	GGC	TCC	CAC	TAT	TCC	ACC	TTC	GAC	GGC	514
MUC2	TGT	TAC	TGC	ACG	GTC	TGG	GGC	GAC	CCC	CAC	TAT	GTC	ACC	TTC	GAC	GGA	897
	Cys		Cys	Ser	Gly	Trp	Gly	Asp	Pro	His	Tyr		Thr	Phe	Asp	Gly	
MUC5AC	ACC	TAC	TAC	ACC	TTC	CTG	GAC	AAC	TGT	ACG	TAC	GTG	CTG	GTG	CAG	CAG	1131
MUC5B	ACC	TCT	TAC	ACC	TTC	CGG	GGC	AAC	TGT	ACC	TAT	GTG	CTC	ATG	AGA	GAG	562
MUC2	CTC	TAC	TAC	AGC	TAC	CAG	GGC	AAC	TGT	ACC	TAT	GTG	CTC	GTG	GAG	GAG	945
	Thr	Tyr	Tyr	Thr	Phe		Gly	Asn	Cys	Thr	Tyr	Val	Leu	Val		Glu	
MUC5AC	ATT	GTG	CCC	GTG	TAT	GGC	CAC	TTC	CGC	GTG	CTC	- - -	GTC	GAC	AAC	- - -	1173
MUC5B	ATC	CAT	GCA	CGC	TTT	GGG	AAT	CTC	AGC	- - -	CTC	- - -	GTC	GAC	AAC	CAC	607
MUC2	ATC	AGC	CCC	TCC	GTG	GAC	AAC	TTC	GGA	GTT	TAC	TAC	ATC	GAC	AAC	- - -	987
	Ile		Pro			Gly	Asn	Phe		Val	Leu	Tyr		Asp	Asn		
MUC5AC	TAC	TTC	TGC	GGT	GCC	- - -	- - -	GAG	GAC	GGG	CTC	TCC	TCC	TGC	CCG	AGG	1212
MUC5B	TAC	- - -	TGC	ACG	GCC	TCT	- - -	GCC	GCT	GCC	CTC	CCG	CCG	TGC	CCC	CGC	652
MUC2	TAC	CAC	TGC	GAT	CCC	- - -	- - -	AAC	GAC	AAG	GTG	TCC	TCC	TGC	CCC	CGC	1026
	Tyr		Cys						Asp				Ser	Cys	Pro	Arg	
MUC5AC	TCC	ATC	- - -	ATC	CTG	GAG	TAC	CAC	CAG	- - -	GAC	CGC	GTG	GTG	CTG	ACC	1254
MUC5B	GCC	CTC	AGC	ATC	CAC	- - -	TAC	AAG	TCC	ATG	GAT	ATC	GTG	GTG	CTC	ACT	694
MUC2	ACC	CTC	- - -	ATC	GTG	- - -	CGC	CAC	GAG	ACC	CAG	GAG	GTG	GTG	CTC	ATC	1065
		Leu		Ile			Tyr	His			Asp		Val		Leu	Thr	
MUC5AC	CGC	AAG	CCA	- - -	GTG	CAC	GGG	GTG	ATG	ACA	AAC	GAG	- - -	ATC	ATC	- - -	1293
MUC5B	GTG	- - -	ACC	ATG	GTG	CAT	GGG	- - -	- - -	AAG	GAG	GAG	GGC	GTG	ATC	CTG	733
MUC2	- - -	AAG	ACC	- - -	GTG	CAT	- - -	ATG	ATG	CCC	ATG	- - -	CAG	GTG	CAG	GTG	1101
		Lys	Thr		Val	His	Gly		Met			Glu			Ile		
MUC5AC	TTC	AAC	AAC	AAG	- - -	GTG	GTG	AGC	CCC	GGC	TTC	CGG	AAA	AAC	- - -	GGC	1335
MUC5B	TTC	GAC	CAA	ATT	CCG	GTG	GTG	AGC	AGC	GGT	TTC	AGC	AAA	AAC	- - -	GGC	775
MUC2	CAG	GTG	AAC	AGG	CAG	GCG	GTG	- - -	G A	CTG	CCC	TAC	AAG	AAG	TAC	GGG	1146
	Pro		Asn			Val	Val	Ser		Gly	Phe		Lys	Asn		Gly	
MUC5AC	ATC	GTG	- - -	- - -	GTG	TCC	CGC	ATC	GGC	GTG	AAG	- - -	ATG	TAC	GCG	ACC	1374
MUC5B	- - -	GTG	CTT	- - -	GTG	TCT	GTG	CTG	GGG	ACC	ACC	ACC	ATG	CTG	GTG	GAC	817
MUC2	- - -	- - -	CTG	GAG	GTG	TAC	CAG	TCT	GGC	ATC	AAC	TAC	GTG	- - -	GTG	GAC	1185
		Val	Leu		Val	Ser		Gly					Met		Val	Asp	
MUC5AC	ATC	CCG	GAG	CTG	GGA	GTG	CAG	GTG	ATG	TTC	TCC	GGC	CTC	ATC	TTC	TCC	1422
MUC5B	ATC	CCT	GCC	CTG	GGC	GTG	ACG	GTG	ACC	TTC	AAAT	GGC	CAA	GTC	TTC	CAG	865
MUC2	ATC	CCC	GAG	CTG	GGT	GTG	CTC	GTG	TCC	TAC	AAAT	GGC	CTG	TCC	TTC	TCC	1233
	Ile	Pro	Glu	Leu	Gly	Val		Val		Phe	Asn	Gly	Leu		Phe	Ser	
MUC5AC	GTG	GAG	GTG	CCC	TAC	AGC	AAG	TTT	GCC	AAC	AAC	ACC	GAG	GGC	CAG	TGC	1470
MUC5B	GCC	CGG	CTG	CCC	TAC	AGC	CTC	TTT	CAC	AAC	AAC	ACC	GAG	GGC	CAG	TGC	913
MUC2	GTG	AGG	CTG	CCC	TAC	CAC	CGG	TTT	GCC	AAC	AAC	ACC	AAG	GGC	CAG	Cys	1281
	Val	Arg	Leu	Pro	Tyr	Ser		Phe		Asn	Asn	Thr	Glu	Gly	Asn		

Figure 2 For legend see page 735.

<i>MUC5AC</i>						GAC	AGG	AAG				CGC	ACG		AGG		1518
<i>MUC5B</i>						AAC	CAG	ACC				CTC	CAG		GAC		961
<i>MUC2</i>						AAC	ACC	TCC				ATT	CTG		AGC		1329
<i>MUC5AC</i>		GTG				- - -		TCC				GGC	CTC	- - -			1560
<i>MUC5B</i>		ACT				AGT		AAG				AAG	CTC	- - -			1003
<i>MUC2</i>		ATC				AAC		- - -				GCT	GAC	CAG			1371
<i>MUC5AC</i>		AGC			GAC	CAG		GCC		CAC	CGG			CCG			1608
<i>MUC5B</i>		CCC			AGA	AAG		GGC		TGG	GCC			GGC			1051
<i>MUC2</i>		AAC			TCC	AAG		CAC		- - -	- - -			AGC			1413
<i>MUC5AC</i>			GTG	GGG		- - -	ACC		GTT	- - -			ACC		GTG		1650
<i>MUC5B</i>			GTC	AGG			GCC		GTT	- - -			AGC		CCC		1093
<i>MUC2</i>			AAG	CGC			GTC		GTT	CCC			GGC	GTT	AAA		1458
<i>MUC5AC</i>			ACA	GTT	GGG	TCT	ACC	ACC	GTG	GGG		ACC	ACA	CCG	CCT	GCT	1698
<i>MUC5B</i>			- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -		ACC	- - -	- - -	- - -	- - -	1102
<i>MUC2</i>			- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -		CAC	AAG	GAC	- - -	- - -	1473
<i>MUC5AC</i>	CCG		CTG				ATC						- - -				1743
<i>MUC5B</i>	- - -		CCA										- - -				1144
<i>MUC2</i>	- - -		ACC										AGC				1515
<i>MUC5AC</i>		GAG	CCG			ACT	GTG	ATC			CTG	CTG			GAG		1791
<i>MUC5B</i>		GAG	GAG			AAC	CTT				CTG	CTG			AAC		1192
<i>MUC2</i>		GCC	CAG			GCA					CTG	CTG			GAT		1563
<i>MUC5AC</i>		GTG			CGG		CAC		ACG	- - -	GAC		GAT		GTG		1836
<i>MUC5B</i>		ATC			CAG		AGG	- - -	GCC	CGC		GAG		GTG	CCC		1234
<i>MUC2</i>		GTG			AGC		TTC		CCG	GCC	- - -		AGC		GAG		1608
<i>MUC5AC</i>	TCC				CTG						- - -		TCC	CAC	GAC		1881
<i>MUC5B</i>	CAG				GCT						CGC		- - -	CGG	GGA		1279
<i>MUC2</i>	GCC	AGT			GCC						- - -		CAG	CAG	AAC		1653
<i>MUC5AC</i>		ATC			AGA		CGG	ACC			- - -	ATG				- - -	1923
<i>MUC5B</i>		AGT			CAG		GCA	ACT			- - -	CTG			- - -	CTC	1321
<i>MUC2</i>		CTC			CGG		CAC	ACG			- - -	GCC				GTG	1695
<i>MUC5AC</i>				GCC	GAC		GTG							AGC	AAC	- - -	1968
<i>MUC5B</i>				CCC	ACC									ATA	CAG	- - -	1366
<i>MUC2</i>				TCT	CAC									GCA	GAA	GAG	1743
<i>MUC5AC</i>		TCC			TAC	GGG		GAC		GCC		CTC		GCT	CTC	CGG	2016
<i>MUC5B</i>		GCC			- - -	- - -			- - -	- - -		- - -			CAG	AGC	1396
<i>MUC2</i>		- - -			- - -	- - -				TCC		TCC			- - -	AAC	1779
<i>MUC5AC</i>	GAG	GCC		CCC	- - -	ATC	ACC									ATG	2061
<i>MUC5B</i>	CCA	CAG		GAG	GGG	ATG	GCG									CAG	1444
<i>MUC2</i>	ACA	GTG		- - -	- - -	GTG	- - -									ACC	1818
<i>MUC5AC</i>	ACC				ACC	AGT	GCC	- - -	CAG				CCC			GCC	2103
<i>MUC5B</i>	ATC				GCA	CAC	ATG		- - -				CAG			- - -	1483
<i>MUC2</i>	ATG	Ile	Phe	Asn	TAC	GCT	CCT		TTT	GAT			AAG			- - -	1857
<i>MUC5AC</i>		CCC			CTG				GAG		GAG		GTG				2151
<i>MUC5B</i>		- - -			GTG				TTT		TTT		AAA				1528
<i>MUC2</i>		- - -			Val	Gly	Pro	Asp	Gly	Val	Pro		AGA	GAG			1902
<i>MUC5AC</i>	ACC	GTG		ATG				GAG		ACG		GAG		GCC	ACG	TGG	2199
<i>MUC5B</i>	CGG	TGG		AGC				TCC				GAC			TCA	GTG	1576
<i>MUC2</i>	CAC	TTC		TTC				AAC				CTG			GGA	AGT	1950
<i>MUC5AC</i>	ACG	CTG	ACC		CAG			CTC		CCG	- - -	CTG	- - -	- - -		- - -	2235
<i>MUC5B</i>	TCG	GTG	CAG		AGA			CCC		GAC	GCC		GGT	CAG		- - -	1621
<i>MUC2</i>	GGC	ATC	ATC		CAA			AGG		AGC	- - -		- - -	AAG		GTT	1992
<i>MUC5AC</i>		GCC		CCC	CTG					CCT			GCA	GCC	CAC	CAG	2283
<i>MUC5B</i>		CCG		AAC	CGT					ACC			AGG	CCC	CGG	CGG	1669
<i>MUC2</i>		ACC		GTG	GAA					CTC			AGG	GTC	AAC	CCT	2040
<i>MUC5AC</i>		GGC	CAG				CAG	TAC	AGC							CGC	2331
<i>MUC5B</i>		AAC	CCC				GAG									ACC	1717
<i>MUC2</i>		GAC	ACC				ATT									CTG	2088
<i>MUC5AC</i>			GCG	- - -	- - -		- - -		GCC			GAG		GCC		GCG	2370
<i>MUC5B</i>			CAG	AGC	CTG		- - -		- - -			CCT		CAG		TCC	1759
<i>MUC2</i>			GAG	AAG	- - -		TCC		- - -			CTG		TTC		GTC	2130

Figure 2 For legend see page 735.

MUC5AC	ATC	CCG	ACC	TAC	CAG	GAG	GGG	GCC	TGC	TGC	CCA	GTC	- - -	CAC	AAC	TGC	2415
MUC5B	ATC	TGC	ACC	CAG	GAG	GAG	GGC	GAC	TGC	TGT	CCC	ACC	TTC	- - -	CGC	TGC	1804
MUC2	AAG	AGC	ATG	ATG	GTG	GCT	GGA	AGG	TGC	TGT	CCC	- - -	TTC	TAC	TGG	TGT	2175
	Ile		Thr			Glu	Gly		Cys	Cys	Pro		Phe			Cys	
MUC5AC	ATG	TGG	ACA	- - -	GTG	TGC	AGC	ATC	AAC	GGG	- - -	ACC	CTG	TAC	CAG	CCC	2457
MUC5B	ATG	CCT	CAG	- - -	CTG	TGT	TGC	TAC	AAT	GGC	- - -	ACC	TTC	TAC	GGG	GTT	1846
MUC2	GAG	TCC	AAG	GGG	GTG	TGT	GTT	CAC	- - -	GGG	AAT	GCT	GAG	TAC	CAG	CCC	2220
					Val	Cys			Asn	Gly		Thr		Tyr	Gln	Pro	
MUC5AC	GGT	GCC	GTG	GTG	- - -	TCC	TCC	AGC	CTG	- - -	TGC	GAA	ACC	TGC	AGG	TGT	2499
MUC5B	GGT	GCA	ACC	TTC	CCA	GGC	GCC	- - -	CTT	CCC	TGC	CAC	ATG	TGT	ACC	TGC	1891
MUC2	GGT	TCT	CCA	GTT	TAT	TCC	TCC	AAG	- - -	- - -	TGC	CAG	GAC	TGC	GTG	TGC	2262
	Gly	Ala		Val		Ser	Ser		Leu		Cys			Cys		Cys	
MUC5AC	GAG	CTG	CCG	GGT	GGC	CCC	- - -	CCA	TGC	GAC	GCG	TTC	GTG	GTG	AGC	TGT	2544
MUC5B	CTC	TCT	- - -	GGG	GAC	- - -	- - -	ACC	CAG	GAC	GCA	ACG	GTG	- - -	CAC	TGT	1927
MUC2	ACG	GAC	AAG	GTG	GAC	AAC	AAC	Thr	CTG	Asp	CTC	AAC	Val	ATC	GCC	TGC	2307
				Gly	Asp											Cys	
MUC5AC	- - -	GAG	ACC	CAG	ATC	TGC	AAC	- - -	ACA	CAC	TGC	CCT	GTG	GGC	TTC	GAG	2586
MUC5B	CAG	GAG	- - -	GAT	GCC	TGC	AAC	AAT	ACT	ACC	TGT	CCC	CAG	GGC	TTT	GAG	1972
MUC2	ACC	- - -	CAC	GTG	CCC	TGC	AAC	- - -	ACC	TCC	TGT	AGC	CTT	GGC	TTT	GAA	2349
		Glu				Cys	Asn		Thr		Cys	Pro	Gly	Phe		Glu	
MUC5AC	TAC	CAG	GAG	CAG	- - -	AGC	GGG	CAG	TGC	TGT	GGG	ACC	TGT	GTG	CAG	GTC	2631
MUC5B	TAC	AAG	AGA	GTG	- - -	- - -	GGG	CAG	TGC	TGT	GGG	ACC	TGT	GTG	CAG	ACC	2017
MUC2	CTC	ATG	GAG	- - -	GCC	CCC	GGG	GAG	TGC	TGT	AAG	AAG	TGT	GAA	CAG	ACG	2394
	Tyr		Glu		Ala		Gly	Gln	Cys	Cys	Gly		Cys	Val	Gln	Thr	
MUC5AC	GCC	TGT	GTG	ACC	AAC	ACC	AGC	AAG	AGC	CCC	GCC	- - -	CAG	CAC	CTC	TTC	2673
MUC5B	GCC	TGT	CTC	ACG	- - -	- - -	- - -	- - -	- - -	CCC	GAT	GGC	CAG	CAC	GTC	- - -	2047
MUC2	CAC	TGT	ATC	ATC	- - -	- - -	- - -	AAA	CAG	CCC	GAC	AAC	CAG	CAC	GTC	ATC	2433
	la	Cys		Thr				Lys		Pro	Asp		Gln	His	Val		
MUC5AC	- - -	TAC	CCT	GGC	GAG	ACC	TGG	- - -	TCA	GAC	GCA	GGG	AAC	CAC	TGT	- - -	2712
MUC5B	- - -	CAG	CTG	AAT	GAA	ACC	TGG	GTG	AAC	AGC	CAT	GTG	GAC	AAC	TGC	ACC	2092
MUC2	CTG	AAG	CCC	GGG	GAC	TTC	- - -	AAG	AGC	GAC	CCG	AAG	Asn	Asn	Cys	Thr	2478
			Pro	Gly	Glu	Thr	Trp			Asp							
MUC5AC	GTG	ACC	CAC	CAG	TGT	GAG	AAG	- - -	CAC	- - -	CAG	GAT	GGG	CTC	GTG	- - -	2751
MUC5B	GTG	- - -	TAC	CTC	TGT	GAG	GCT	GAG	GGT	- - -	- - -	- - -	GGG	- - -	GTG	CAT	2125
MUC2	- - -	TTC	TTC	AGC	TGC	GTG	AAG	ATC	CAC	AAC	CAG	- - -	Gly	CTC	ATC	TCC	2517
		Val			Cys	Glu	Lys		His		Gln			Leu	Val		
MUC5AC	GTG	GTG	ACC	ACG	AAG	AAG	GCG	- - -	TGC	CCC	- - -	- - -	- - -	CCG	CTC	AGC	2787
MUC5B	TTC	CTG	ACC	CCA	CAG	CCT	GCA	TCC	TGC	CCA	- - -	- - -	GAT	GTG	TCC	AGC	2167
MUC2	- - -	GTG	TCC	- - -	AAC	ATC	- - -	ACC	TGC	CCC	AAC	TTC	GAT	GCC	AGC	ATT	2559
		Val	Thr				Ala		Cys	Pro			Asp		Ser		
MUC5AC	TGT	TCT	CTG	GAT	GAG	GCC	CGC	ATG	AGC	AAG	GAC	GGC	TGC	TGC	CGC	TTC	2835
MUC5B	TGT	- - -	AGG	GGG	AGC	- - -	- - -	CTC	AGG	AAA	ACC	GGC	TGC	TGC	TAC	TCC	2206
MUC2	TGT	ATC	CCG	GGC	TCC	ATC	ACA	TTC	ATG	CCC	AAT	GGA	TGC	TGC	AAG	ACC	2607
	Cys		Gly	Ser						Lys		Gly	Cys	Cys			
MUC5AC	TGC	CCG	CCG	CCC	CCG	CCC	CCG	TAC	CAG	AAC	CAG	TGC	ACC	- - -	TGT	GCT	2880
MUC5B	TGT	GAG	- - -	- - -	- - -	- - -	- - -	- - -	- - -	AAC	CAG	GAC	- - -	TCC	TGT	CAC	2227
MUC2	TGC	ACC	CCT	- - -	GTG	CCC	- - -	- - -	CGC	AAT	GAG	GTC	ACC	AGG	TGC	TCC	2643
MUC6	Cys	Pro				Pro				Asn	Glu	Thr			Cys	AGT	816
MUC5AC	- - -	GTG	- - -	TAC	CAT	AGG	AGC	CTG	ATC	ATC	CAG	CAG	CAG	GGC	TGC	AGC	2922
MUC5B	- - -	GTC	- - -	CGC	ATC	AAC	ACG	ACC	ATC	CTG	TGG	CAC	CAG	GGC	TGC	GAG	2269
MUC2	ACC	GTC	- - -	CCC	GTC	ACC	ACG	GAG	ATC	GTT	TCC	TAC	GCC	GGC	TGC	- - -	2682
MUC6	- - -	GTG	CGG	GAG	CAG	CAG	GAG	GAG	ATC	ACG	TTC	AAG	Gln	Gly	Cys	ATG	858
		Val				Thr	Glu	Ile									
MUC5AC	TCC	TCC	GAG	CCC	GTG	CGC	CTG	- - -	GCT	TAC	TGC	CGG	GGG	AAC	TGT	- - -	2964
MUC5B	ACC	- - -	GAG	- - -	GTG	AAC	- - -	ATC	ACC	TTC	TGC	GAG	GGG	TCC	TGC	CCC	2308
MUC2	ACC	AAG	- - -	ACC	GTG	- - -	CTC	ATG	AAT	CAT	TGC	TCC	GGG	TCC	TGC	- - -	2721
MUC6	GCG	AAC			Val	ACG		GTA	ACC	CGC	TGC	GAG	GGC	GCC	TGC	ATT	897
	Thr		Glu				Leu		Thr		Cys	Glu	Gly	Ser	Cys		
MUC5AC	GGG	GAC	AGC	TCT	TCC	ATG	TAC	TCC	TGC	GAG	GCC	AAC	ACG	GTG	GAG	CAC	3012
MUC5B	GGA	GCG	- - -	TCC	TCC	AAG	TAC	TCC	TGC	GCA	GCC	CAG	GCC	ATG	CAG	CAC	2350
MUC2	GGG	- - -	ACA	TTT	GTC	ATG	TAC	TCC	TGC	TCC	GCC	CAG	GCC	CTG	GAG	CAC	2766
MUC6	TCC	GCT	- - -	- - -	GCC	ATG	TTC	AAC	ATC	ATC	ACC	CAG	CAG	GTG	GAT	GCC	939
	Gly				Ser	Met	Tyr	Ser		Glu	Ala	Ala	Gln	Val	Asp	His	
MUC5AC	AGG	TGC	CAG	TGC	TGC	CAG	GAG	CTG	CGG	ACC	TCC	CTG	- - -	AGG	- - -	- - -	3051
MUC5B	CAG	TGC	ACC	TGC	TGC	CAG	GAG	AGG	CGG	GTC	CAC	GAG	- - -	- - -	- - -	GAG	2389
MUC2	AGC	TGC	TCC	TGC	TGC	AAG	GAG	GAG	CGG	ACC	AGC	GAG	- - -	CGT	- - -	GAG	2808
MUC6	CGC	TGC	AGC	TGC	TGC	CGC	CCC	CTC	CAC	TCC	TAT	GAG	Gln	CAG	CTG	GAG	987
	Arg	Cys		Cys	Cys	Gln	Glu	Leu	Ala	Thr		Glu				Glu	
MUC5AC	AAT	GTG	ACC	CTG	CAC	TGC	ACC	GAC	GGC	- - -	TCC	AGC	- - -	CGG	GCC	TTC	3093
MUC5B	ACG	GTG	CCC	TTC	CAC	TGC	CCT	AAC	GGC	- - -	TCA	GCC	ATC	CTG	- - -	CAC	2431
MUC2	- - -	GTG	GTG	CTG	AGC	TGC	CCC	AAT	GGC	GCC	TCC	- - -	- - -	CTG	ACA	CAC	2847
MUC6	- - -	- - -	- - -	CTG	CCC	TGC	CCC	GAT	CCC	AGC	ACC	GCC	GCC	CGG	CGG	CTC	1026
		Val		Leu	His	Cys	Pro	Asn/Asp	Gly		Ser		Leu/Ala		His		
MUC5AC	AGC	TAC	ACC	- - -	GAG	GTG	- - -	GAA	GAG	TGC	GGC	TGC	ATG	GGC	CGG	CGG	3135
MUC5B	ACC	TAC	ACC	- - -	CAC	GTG	- - -	GAT	GAG	TGT	GGC	TGC	ACG	CCC	TTC	- - -	2470
MUC2	ACC	TAC	ACC	- - -	CAC	- - -	ATC	GAG	AGC	TGC	CAG	TGC	CAG	GAC	ATC	GTC	2889
MUC6	GTA	CTC	ACC	CTG	CAG	GTG	TTC	AGC	CAC	TGC	GTC	TGC	AGC	TCT	GTG	GCC	1074
	Thr	Tyr	Thr		His	Val		Glu	Glu	Cys	Gly						
MUC5AC	TGC	CCT	GCG	CCG	- - -	GGC	GAC	ACC	CAG	CAC	TCC	GAG	GAG	GGC	GAA	CCC	3180
MUC5B	TGT	GTC	- - -	CCT	GCG	- - -	CCC	ATG	GCT	CCC	CCA	CAC	- - -	ACC	CGT	GGC	2509
MUC2	TGT	GGG	CTC	CCC	ACC	GGC	- - -	ACC	- - -	- - -	TCC	CAG	CGG	GGC	CGC	GCC	2928
MUC6	TGT	GGA	GAC	TAG	CAG	TCC	CTG	TGC	TTC	TGC	GCT	GAA	ACT	GAG	TGA	GAA	1134
	Cys	Ser		Pro		Glu		Thr			Ser		Ala	Arg			

Figure 2 For legend see opposite.

<i>MUC5AC</i>	G A G	C C C	A G C	C A G	G A G	G C A	G A G	A G T	G G G	A G C	T G G	G A G	A G A	G G C	G T C	C C A	3228
<i>MUC5B</i>	T T C	C C G	G C C	C A G	G A G	G C C	A C T	G C T	G T C	T G A	G A A	C G T	T C T	G C C	T C C	A T G	2563
<i>MUC2</i>	T C C	C C T	A G G	C A T	- - -	C T G	G G G	A G C	G G G	T G A	G C G	G G G	T G G	G C A	C A G	C C C	2979
<i>MUC6</i>	C A G	G A A A A	C A C	C A C	A G C	C C C	T T C	C G C	T T G	C A G	C A G	C T G	C T T	T C C	T G T	C A C	1197
		Pro		Gln	Glu	Ala		Ser	Gly	Stop							
<i>MUC5AC</i>	G T G	T C C	C C C	A T G	C A C	T G A	C C A	G C A	C T G	C C G	C C C	T C C	T G A	C C T	C C A	A G G	3286
<i>MUC5B</i>	C T C	T G T	C C A	C C T	G G A	G C C	A G A	T G T	G C A	T T G	T C T	C T G	A T C	A A A	A C C	T T G	2626
<i>MUC2</i>	C T G	C C C	T G A	C A G	C T T	T A C	C T C	C C C	G G A	C C C	T C T	C T G	A G C	T C C	T A G	C T C	3042
<i>MUC6</i>	C T G	G C C	C C A	A G T	G C C	C T G	G C C	G T G	C T C	C C T	G G G	G C A	C G G	T T G	A G A	G G G	1260
<i>MUC5AC</i>	A T A	T G T	C C T	C T G	A G C	T C G	G C T	T T C	C A A	G G C	C A G	T G G	A A C	T T G	T G C	C C T	3349
<i>MUC5B</i>	G A G	C C C	C C G	C C T	G T G	T G T	G G C	A C C	C G C	G C T	C C G	T G C	T C C	T G C	C C A	C C C	2689
<i>MUC2</i>	C A G	A T A	T T A	T T G	T C T	G A G	T C T	T T G	T T C	A G T	C C T	T G C	T T T	C C A	A A T	<u>A A T A A A</u>	3105
<i>MUC6</i>	A G G	G G C	T C A	G A C	T A C	C A C	T C T	G C A	G A C	C C C	T G A	G C C	A G A	G A G	<u>A G G G A C</u>	T G A	1323
<i>MUC5AC</i>	G C T	T T G	A A C	A C A	C A C	T G T	C C A	G C C	C C G	C T T	T T C	T T G	T G G	A G G	T G T	G G G	3412
<i>MUC5B</i>	A A C	G G G	C C C	C A G	A A G	G G T	A G G	G G G	C C A	G C A	G A G	A C C	T T T	C G G	A G G	G C C	2752
<i>MUC2</i>	T G C	A															3109
<i>MUC6</i>	T G G	T C A	C G G	A C C	T C C	A G G	C A C	A C A	G G G	C A C	T C C	G A C	C A C	C C C	T G C	C A C	1386
<i>MUC5AC</i>	C C T	G G A	G G A	G G G	C C C	T T A	C C C	A C C	C G C	C T G	C A G	C C A	C C T	C T C	A G G	A C C	3475
<i>MUC5B</i>	T A C	C C T	G G G	A G A	G C C	T G T	G G C	C C A	C C T	T T G	C C C	T T G	C C C	T C C	T G A	T G T	2815
<i>MUC6</i>	C C C	A G C	C C C	T G A	A C T	T G G	C C C	A G C	C C T	G C T	T G G	C C C	A G A	A C C	T G C	A G A	1449
<i>MUC5AC</i>	G C C	G A G	C T C	C T C	T G G	C C A	T G C	A T C	C A G	C C T	T G C	T G T	T T C	T G G	G A C	G T G	3538
<i>MUC5B</i>	G G A	A C A	A A C	T A A	G C A	T G T	G C G	G G C	C T A	T G T	G T C	C C T	T G C	A C G	G C C	G G A	2878
<i>MUC6</i>	C A G	G C G	C T C	G A C	C A G	A C C	C A T	C A G	G G G	C G A	G A G	A G G	G C A	C G G	A A A	A C C	1512
<i>MUC5AC</i>	C T C	A G G	A A T	G A C	G C T	T G G	A C A	T G G	T G A	T C A	G C T	G C C	T G G	T G G	C T G	C A G	3601
<i>MUC5B</i>	G A T	T C A	G C T	T G C	C A C	G T C	C G C	G C G	T G G	G C A	G A C	A G C	T G T	T C A	G G C	A A G	2941
<i>MUC6</i>	A A G	A G G	C C C	A G T	C A G	C C A	C A G	C A C	A G A	A A G	A G A	G A G	A T C	C C C	A G A	G T C	1575
<i>MUC5AC</i>	T C C	T A C	C T C	A G C	C C T	C A G	C C T	G C G	C T C	C C C	T C C	T C A	G T A	C A C	G G C	A A T	3664
<i>MUC5B</i>	G C C	A G G	A A G	C T G	C G A	C A G	G C A	A G C	G G C	C G C	C T G	T C C	A T G	C C T	G C T	G C A	3004
<i>MUC6</i>	G T G	G C A	G C A	G G G	C A G	G C A	G C C	C G C	C C C	G C T	C C C	A G C	A G G	C A G	A A G	G C C	1638
<i>MUC5AC</i>	C A C	T T G	A G C	A T T	T T G	C A											3681
<i>MUC5B</i>	G A G	G T C	G C A	A C G	G C C	A G G	T C A	G A G	A G G	G G T	C A G	C A T	C C C	A A G	C C C	C C T	3067
<i>MUC6</i>	A C A	C C C	A T C	C C C	A G C	A G C	C T G	T C C	T T G	G G A	G G G	G C G	T A C	C C G	G T C	A G A	1701
<i>MUC5B</i>	C C A	G T T	T T G	C A A A A	<u>T A A A A</u>	C C C	T G A	G C A	T T G	A G T A							3100
<i>MUC6</i>	C C G	G T T	C T T	G T C	A A G	G C A											1719

Figure 2 Sequence similarity between the 3'-regions of the *MUC5AC*, *MUC5B*, *MUC2* and *MUC6* cDNAs

The nucleotide sequences of the 3'-regions of the *MUC5AC*, *MUC5B* (EMBL Data Bank accession number Y09788) [13], *MUC2* (EMBL Data Bank accession number M94132) [11] and *MUC6* (EMBL Data Bank accession number U97698) [14] cDNAs were aligned to maximize sequence similarity and minimize gaps. Gaps are indicated by dashes. The numbers on the right refer to nucleotides. Nucleotide triplets coding for identical amino acids are enclosed in open boxes. Codons identical in more than two genes are enclosed in bold open boxes. Triplets coding for cysteine residues are shaded. Derived amino acids identical in at least two of the four genes are given below the nucleotide sequence. Derived amino acids identical in more than two genes are given in bold letters. Cysteine residues are underlined twice. Typical potential N-glycosylation sites are underlined once. The polyadenylation signals are in italics and underlined. The stop codons are indicated in black boxes. The location of introns in *MUC5AC*, *MUC5B* [13], and *MUC6* [14] is shown by filled arrowheads. The putative location of some of the introns in *MUC2* is shown by open arrowheads. Introns are named by capital letters from A to R.

deduced peptide level. Nucleotide triplets coding for nearly all of the cysteine residues (in shaded boxes and underlined twice in Figure 2) and several other amino acids (in open boxes in Figure 2), including some of the potential N-glycosylation sites, were present at the same positions in the coding region of *MUC5AC*, *MUC5B* and *MUC2*. These are also conserved in the D4, B, C and CK domains of vWF [15]. Moreover, the cysteine-rich domain of the C-terminal region of *MUC6* exhibits sequence similarities with the CK domain of vWF and the corresponding domains in *MUC2*, *MUC5AC* and *MUC5B*. Nucleotide triplets coding for the 11 cysteine residues and several other amino acids within the CK-vWF-like domain were located at the same positions (Figure 2). In contrast, there was no significant sequence similarity among any of the 3'-untranslated regions of *MUC5AC*, *MUC5B*, *MUC2* and *MUC6*. The *MUC5AC* tracheal cDNA did not have a typical AATAAA polyadenylation signal as did *MUC5B*, *MUC2* and *MUC6*, the polyadenylation signal in this case being CATAAA [20].

Sequence strategy for the 3'-region of *MUC5AC*

The sequence strategy is given in Figure 1(B). *Xho*I-*Xho*I and *Sma*I-*Sma*I restriction fragments derived from ELO9 and inserts

generated by amplification using ELO9 as template and specific primers were subcloned and sequenced using universal and internal primers. Direct sequencing on ELO9 was also performed (horizontal arrows in Figure 1B). In order to obtain the complete genomic nucleotide sequence encoding the 3'-portion of *MUC5AC* toward the putative central exon, an amplification procedure on human genomic DNA was employed. Oligonucleotide sequences were deduced from the TH171-118 cDNA clone and chosen by nucleotide comparison with the C-terminal region of *MUC2* and *MUC5B* to avoid (i) cross-hybridization and (ii) hybridization of the oligonucleotides to putative intron sites. Because of the sequencing problems encountered because of the high proportion of G+C, most of the inserts were sequenced several times on both strands.

Sequence data and analysis of the 3'-region of *MUC5AC*

Comparison of the nucleotide sequence of ELO9 with that of the cDNA allowed identification of 15 introns. Three additional introns were identified by comparison of the nucleotide sequence of the genomic clones generated by amplification with the nucleotide sequence of the cDNA clone TH171-118. The 18 introns designated A to R were entirely sequenced. The complete sequence of the 3'-region of *MUC5AC* encompassed 8290 bp.

Table 1 Comparison of the characteristics of the exon–intron junctions of the 3′-regions of the *MUC5AC*, *MUC5B* and *MUC6* genes

Capital letters indicate exons and small letters indicate introns. Dots represent identity with the *MUC5AC* sequence. The positions of the introns in *MUC5B* was reported by Desseyn et al. [13], and the positions of the introns in *MUC6* was reported by Toribara et al. [14]. The sequences of the two introns of *MUC6* (designated 1 and 2) were not available from the EMBL Data Bank.

Gene	Protein Domain	Exon		5′-Splice donor	Intron			3′-Splice acceptor
		No.	Size (bp)		Name	Size (bp)	Class	
<i>MUC5AC</i>	central	0	?	CTGCAG <u>gt</u> tcgtgagtggttcc	A	257	1	cccttccctcttacc <u>ag</u> GATCCA
<i>MUC5B</i>			10,713	.GC.C. . . g a . . . c a . . . g g a		679	1	. t g t . g t t c t . c GGAAG
<i>MUC6</i>			?	.AC.C.	1	~2500	1	
<i>MUC5AC</i>	MUC11p15	1	200	AGAAAG <u>gt</u> aacccccctacttc	B	502	0	tctgtgcttctgtgc <u>ag</u> AAAGGT
<i>MUC5B</i>			182	C.GC. . . . gg g c c t g c .		283	0	. g . t c c . . . g . c c GTGAA .
<i>MUC5AC</i>	A3uD4	2	175	GCCAGT <u>gt</u> gagtgaggagcgccg	C	447	1	cctgggggcccctgggc <u>ag</u> GTGTGT
<i>MUC5B</i>			172	. . G t g g g t . . g t .		1118	1	a t c t c t . t . c c C A . C .
<i>MUC5AC</i>	D4-like	3	251	AACGAG <u>gt</u> ggggggcgccgcccg	D	126	0	acgccctgctgtcc <u>ag</u> ATCATC
<i>MUC5B</i>			260	GG.CT. a . t c . . a g . . t g c		338	0	c . t g g a . t . c c C . G
<i>MUC5AC</i>	D4-like	4	184	AGTGC <u>gt</u> gagggccacagggc	E	340	1	ccacgctctgtccc <u>ag</u> GCACTT
<i>MUC5B</i>			187	 t g g g . g . c . g		159	1	a . g . c . c a c t C .
<i>MUC5AC</i>	D4-like	5	268	GAGCAA <u>gt</u> gagacttgggtgc	F	539	2	ccatccctcccaac <u>ag</u> GGTCTT
<i>MUC5B</i>			226	 C t . c . c c c . c g		114	2	. g c . t t g c
<i>MUC5AC</i>	D4-like	6	179	TGTGCC <u>gt</u> gagtgccaccact	G	333	1	tccccctcccaac <u>ag</u> CATTCA
<i>MUC5B</i>			176	 G g g g g . g c c		597	1	. c t . t g . g
<i>MUC5AC</i>	D4-like	7	88	CCTCG <u>gt</u> agggcaccctccct	H	141	2	ctatgctctcttac <u>ag</u> GCTCTC
<i>MUC5B</i>			70	. T C T A a . t . . . a g g g a t g		443	2	. . t g t . a t c . . . g
<i>MUC5AC</i>	D4-like	8	104	GCCCCA <u>gt</u> acgtgcccaggcc	I	156	1	tctctctctcttcc <u>ag</u> GGTGTC
<i>MUC5B</i>			101	C.TG.C. a . c t . . . g c c a . .		463	1	c t g t . t . . t
<i>MUC5AC</i>	D4-like	9	32	GTGAAG <u>gt</u> gagtggaaggcat	J	558	0	tccccctctcttgc <u>ag</u> GTGGGC
<i>MUC5B</i>			32	AAATTT c t c c a c c		258	0	c t t g a . c . a . . . c
<i>MUC5AC</i>	D4+B-like	10	172	GCTGC <u>gt</u> aagccctttgtctg	K	263	1	ctcccgcccccgcc <u>ag</u> CCTGCA
<i>MUC5B</i>			181	TG a . g c . g c a g a		389	1	t c . a t c . t
<i>MUC5AC</i>	B+C-like	11	102	ACTGCA <u>gt</u> gagtgggcctggac	L	263	1	tcttctctgtgggc <u>ag</u> GCTGGA
<i>MUC5B</i>			105	G g g c t . g g		126	1	. . c c t . t c c t c
<i>MUC5AC</i>	C-like	12	38	TACCAG <u>gt</u> aagagccacggag	M	567	0	cctggctctctccc <u>ag</u> CCCGGC
<i>MUC5B</i>			38	 G G a c a . c . .		574	0	t . c c a . c . c t t g
<i>MUC5AC</i>	C-like	13	123	CCTGTG <u>gt</u> gagcgctcccacc	N	90	0	ccatccctcccgcc <u>ag</u> GGCTTC
<i>MUC5B</i>			120	. . C C A a c . c a g g . a		696	0	. t c a t g t c a
<i>MUC5AC</i>	C-like	14	99	TTCTAC <u>gt</u> gagtagtggtctgc	O	185	0	ctggcctccctccc <u>ag</u> CCCGGC
<i>MUC5B</i>			87	G C . G t a a c a g c a g g c a		120	0	. c t t g . a t t t c a g
<i>MUC5AC</i>	-	15	120	TCTCTG <u>gt</u> gaggtccaggatc	P	204	0	ccctctgtcggcac <u>ag</u> GATGAG
<i>MUC5B</i>			123	. G C A G t . t g . t g . a g g .		264	0	t t . g g c c . c c a c
<i>MUC5AC</i>	-	16	70	AGAAC <u>gt</u> gagtagccagcct	Q	89	1	tgtgttccctctcc <u>ag</u> AGTCGA
<i>MUC5B</i>			43	. . G . G G a g g a a g c . a c		232	1	c t . t c c t . c t c
<i>MUC6</i>						1		G . G T C T
<i>MUC5AC</i>	CK	17	115	ATGTAC <u>gt</u> gacctgggcagcag	R	467	2	agccccctgccccac <u>ag</u> GTACTC
<i>MUC5B</i>			103	G . C C . A a g t c . t .		656	2	. c . . t g c . t
<i>MUC6</i>		1	103	TGCC.G	2	273	2	C . T . A A
<i>MUC5AC</i>	CK+3′-UTR	18	700					
<i>MUC5B</i>			781					
<i>MUC6</i>		2	811					

The locations of the introns are shown in Figure 2 (filled arrowheads). Exon and intron sizes, splice junction types and sequences of the exon–intron boundaries are given in Table 1. The sizes of the 18 introns ranged from 89 to 567 bp. All the intron–exon junctional sequences were in agreement with the consensus GT–AG motifs described for splice sites in eukaryotic genes [21]. Introns A, C, E, G, I, K, L and Q are of phase class 1 (they interrupt the codon between the first and the second nucleotide), introns F, H and R are of phase class 2 (they interrupt the codon between the second and the third nucleotide) and introns B, D, J, M, N, O and P are of phase class 0 (they lie between two codons). The *Bam*HI–restriction site located at nucleotide 661 in the sequence of the 3′-region of the *MUC5AC* tracheal cDNA is not present in the genomic DNA sequence, since it overlaps the central exon (designated exon 0) and exon 1, and is disrupted by intron A (Figure 2, Table 1). Several GC box binding sites were found in introns. Intron F contains five GC

boxes. Two potential polyadenylation signals CATAAA were found 413 bp and 469 bp downstream from the stop codon.

Comparative sequence analysis of the 3′-region of *MUC5AC*

Comparison of the genomic organization of the 3′-regions of *vWF*, *MUC5B* and *MUC5AC* revealed additional striking similarities. The location of the introns and the type of intron–exon junctions were identical in *MUC5AC* and *MUC5B* (Figure 2, Table 1). As previously reported, several introns are also conserved in *vWF* [13]. Introns B, C, E, G, J, K, L, M and N in *MUC5AC* and *MUC5B* were located at the same positions as introns 33, 34, 36, 37, 38, 40/41, 45, 46 and 43/47 respectively in *vWF*. Moreover, the two introns in the 3′-end of *MUC6* are present at the same positions as introns Q and R in *MUC5AC* and *MUC5B* (Figure 2). Interestingly, introns A and Q in *MUC5AC* and *MUC5B* and the first intron in *MUC6* are of the

Figure 3 Comparison of the genomic organization of the 3'-regions of *MUC5AC* and *MUC5B*

Exons are represented by boxes. They are numbered consecutively above the boxes with 0 for the central exons (grey boxes). Black boxes indicate 3'-untranslated segments. Introns are indicated by elbowed lines. They are named below the lines by capital letters. Lengths of the exons and introns are shown to scale.

same phase class, i.e. phase 1 (Figure 2, Table 1). As shown in Figure 3 and Table 1, the major length differences between *MUC5AC* and *MUC5B* are in the introns. This results in a shorter 3'-region for *MUC5AC* when compared with *MUC5B*. There was no significant sequence similarity among any of the introns of *MUC5AC* and *MUC5B*. More particularly, characteristics that have been described in the introns of the 3'-region of *MUC5B*, such as tandemly repeated sequences in introns A, C, G and P [13] were not found in the 3'-region of *MUC5AC*. Nevertheless, sequence similarities were found in exon-intron boundaries (Table 1) and in short stretches distributed within introns.

DISCUSSION

In the present study, we report the organization and the complete nucleotide sequence of the 3'-region of the human *MUC5AC* gene. The comparative molecular analysis between *MUC5AC* and *MUC5B* revealed striking common features, supporting the notion that they have evolved from a single common ancestral gene.

The nucleotide sequence of the *MUC5AC* tracheal cDNA presented here shows a very high level of identity to the clone NP3a, which was isolated from a human nasal polyp cDNA library [3] and to the clone L31, which was isolated from an HT29-MTX expression library [4]. In contrast, the *MUC5AC* tracheal cDNA and L31 show a low level of identity to NP3a at the deduced peptide level because of nucleotide insertions and deletions. Lesuffleur et al. [4] suggested that L31 and NP3a might correspond to different 3'-ends of *MUC5AC*, divergences resulting from alternate splicing. But it is more likely that differences are due to problems encountered with premature termination and band compression in sequencing because of the high G + C content. A tracheobronchial cDNA clone, designated TH51, was reported to be another candidate for the 3'-end of *MUC5AC* [10]. In this clone containing a poly(A) tail, the tandem repeat domain is interrupted by the stop codon. Nevertheless, we were unable to reproduce this cDNA using RT-PCR and it is likely that TH51 is the result of a rearrangement into clones from the cDNA library. Some of the sequence differences between the *MUC5AC* tracheal cDNA, L31 and NP3a may also

be attributable to punctual polymorphic differences between individuals [22].

We have isolated a cosmid clone ELO9 that contains the main part of the 3'-region of *MUC5AC* and overlaps on approx. 25 kb the 5'-region of the *MUC5B* genomic clone BEN1 reported by Desseyn et al. [12].

The 3'-region of *MUC5AC* contains 18 exons, the last one containing the 3'-untranslated region. Two potential polyadenylation signals, CATAAA, were found downstream from the stop codon. It will be of interest to study the putative implication of these two signals in various tissues. Exons of *MUC5AC* and *MUC5B* are quite similar in length. Furthermore, as previously shown for the C-terminal regions of *MUC5B*, *MUC2*, *L31* and *vWF* [13], nearly all of the cysteine residues are located in same positions, and several other amino acids are also conserved. The sequence of the 3'-region of *MUC6* differs from that of *MUC2*, *MUC5AC* and *MUC5B*. It contains two regions, the first one coding for a serine-, threonine- and proline-rich domain with no cysteine residues, the second one coding for a cysteine-rich domain (11 cysteine residues, 12%) [14]. The cysteine-rich domain of *MUC6* exhibits sequence similarities with the CK domain of *vWF* and the corresponding domains in *MUC2*, *MUC5AC* and *MUC5B*. The disparity in the lengths of the 3'-regions of *MUC5AC* and *MUC5B* is primarily due to the size of their introns, which otherwise are located in identical positions. The remarkable conservation in the distribution, the size and the sequence of the exons coding for the 3'-part of these two genes, and in the location and phase class of introns are strongly suggestive of an evolutionary pressure from a common ancestor. Moreover, this suggests that after *MUC5AC* and *MUC5B* had been duplicated from a common ancestor, no recombination events between the exons were tolerated. In contrast, the 3'-region of *MUC6* contains only two exons coding for a CK-*vWF*-like domain, and two introns, which otherwise are present at the same position as the two last introns Q and R in *MUC5AC* and *MUC5B*. Interestingly, introns A and Q in *MUC5AC* and *MUC5B* and the first intron in *MUC6* are of same phase class, not disrupting the reading frame. Thus although the D4, B and C domains are lacking in *MUC6*, the CK domain is conserved. Present data reinforce the previously reported hypothetical diagram showing evolution of the 11p15.5 mucin genes *MUC6*, *MUC2*, *MUC5AC* and *MUC5B* from a single

ancestral gene [23]. It is likely that exon shuffling mechanisms have played an important role in generating the common ancestor, which may have given rise by successive duplications to the different members of the 11p15.5 mucin gene family. The exon-intron distribution remains to be elucidated for *MUC2*. Our alignment of the nucleotide sequences of *MUC5AC*, *MUC5B* and *MUC2* cDNAs (Figure 2) will enable one to choose oligonucleotides to amplify and determine the exact location and size of the introns in the 3'-region of *MUC2*. We can, from now on, predict from Figure 2 the exact location of introns A, C, E, I, K, M, the probable location of introns B, F, H, Q, R and the approximate location of the other introns (open arrowheads in Figure 2). Contrasting with the coding region, there is little sequence similarity among any of the 3'-untranslated regions of *MUC2*, *MUC5AC*, *MUC5B* and *MUC6*. Previous studies with closely related proteins have shown that the 3'-end non-coding regions diverge more rapidly than the coding regions, or at least as rapidly [24].

In some introns of the *MUC5B* gene, unique tandemly repeated sequences have been identified [13]. Intron G is of particular interest, since it contains nine copies of a perfect 59-bp repeat, each repeat containing one GC box binding site [13]. A nuclear factor designated NF1-MUC5B extracted from the mucus-secreting subclone HT29-MTX selected by adaptation to methotrexate from parental HT29 cells, and different from Sp1, has been shown to bind this site [25]. No tandemly repeated sequences exist in the introns of *MUC5AC*. More particularly, no tandemly repeated sequence containing GC box binding sites were found in intron G. However, five GC boxes are present in intron F, and it will be of interest to investigate their functional significance. Therefore, if some of these regions are functional, it would lead to different regulation of *MUC5AC* and *MUC5B* apomucin production.

The order of the mucin genes on chromosome 11p15.5 has been determined recently to be *TelMUC6-MUC2-MUC5AC-MUC5B^{cen}* [9]. We have isolated a cosmid ELO9 spanning the C-terminal region of *MUC5AC* and the 5'-region of *MUC5B*. This confirms that *MUC5AC* is close to *MUC5B*, and demonstrates that *MUC5AC* and *MUC5B* have the same transcriptional orientation. Furthermore, Velcich et al. [26] have recently shown that a mucin gene, designated *MUCX*, is transcribed in opposite orientation to that of *MUC2*. *MUCX* is likely to be *MUC6* since (1) *MUCX* is located upstream of *MUC2* [26], as *MUC6* is [9], (2) the 5'-region of *MUCX* and the entire *MUC2* locus are encompassed in a 85-kb fragment [26], and the distance between the tandem repeat domains of *MUC6* and *MUC2* has been determined to be of approx. 100 kb [9] and (3) *MUCX* and *MUC6* are expressed with a similar pattern [26]. If *MUCX* is demonstrated to be *MUC6*, thus the transcriptional orientation of the four mucin genes within the *MUC* cluster on 11p15.5 can be deduced as follows: *MUC2*, *MUC5AC* and *MUC5B* are transcribed in same orientation, i.e. from telomere to centromere, and *MUC6* is transcribed in opposite orientation. Nevertheless, further investigations are necessary to ensure that *MUCX* corresponds to *MUC6*.

This work was supported by the Comité du Nord de la Ligue contre le Cancer and the Association de Recherche contre le Cancer. We thank D. Petitprez, P. Mathon and V. Mortelet for excellent technical assistance.

REFERENCES

- Gendler, S. J. and Spicer, A. P. (1995) *Annu. Rev. Physiol.* **57**, 607–634
- Shankar, V., Pichan, P., Eddy, R. L., Tonk, V., Nowak, N., Sait, S. N. J., Shows, T. B., Schultz, R. E., Gotway, G., Elkins, R. C. et al. (1997) *Am. J. Respir. Cell Mol. Biol.* **16**, 232–241
- Meerzaman, D., Charles, P., Daskal, E., Polymeropoulos, M. H., Martin, B. M. and Rose, M. C. (1994) *J. Biol. Chem.* **269**, 12932–12939
- Lesuffleur, T., Roche, F., Hill, A. S., Lacasa, M., Fox, M., Swallow, D. M., Zweibaum, A. and Real, F. X. (1995) *J. Biol. Chem.* **270**, 13665–13673
- Klomp, L. W. J., Van Rens, L. and Strous, G. J. (1995) *Biochem. J.* **308**, 831–838
- Griffiths, B., Matthews, D. J., West, L., Attwood, J., Povey, S., Swallow, D. M., Gum, J. R. and Kim, Y. S. (1990) *Ann. Hum. Genet.* **54**, 277–285
- Nguyen, V. C., Aubert, J.-P., Gross, M. S., Porchet, N., Degand, P. and Freizal, J. (1990) *Hum. Genet.* **86**, 167–172
- Toribara, N. W., Robertson, A. M., Ho, S. B., Kuo, W. L., Hicks, J. W., Gum, J. R., Byrd, J. C., Siddiki, B. and Kim, Y. S. (1993) *J. Biol. Chem.* **268**, 5879–5885
- Pigny, P., Guyonnet Dupérat, V., Hill, A. S., Pratt, W. S., Galiègue-Zouitina, S., Collyn D'Hooghe, M., Laine, A., Van Seuningen, I., Degand, P., Gum, J. R., et al. (1996) *Genomics* **38**, 340–352
- Guyonnet Dupérat, V., Audié, J.-P., Debailleul, V., Laine, A., Buisine, M.-P., Zouitina, S., Pigny, P., Degand, P., Aubert, J.-P. and Porchet, N. (1995) *Biochem. J.* **305**, 211–219
- Gum, J. R., Hicks, J. W., Toribara, N. W., Rothe, E.-M., Lagace, R. E. and Kim, Y. S. (1992) *J. Biol. Chem.* **267**, 21375–21383
- Desseyn, J.-L., Guyonnet Dupérat, V., Porchet, N., Aubert, J.-P. and Laine, A. (1997) *J. Biol. Chem.* **272**, 3168–3178
- Desseyn, J.-L., Aubert, J.-P., Van Seuningen, I., Porchet, N. and Laine, A. (1997) *J. Biol. Chem.* **272**, 16873–16883
- Toribara, N. W., Ho, S. B., Gum, E., Gum, J. R., Lau, P. and Kim, Y. S. (1997) *J. Biol. Chem.* **272**, 16398–16403
- Mancuso, D. J., Tuley, E. A., Westfield, L. A., Worrall, N. K., Shelton-Inloes, B. B., Sorace, J. M., Alevy, Y. G. and Sadler, J. E. (1989) *J. Biol. Chem.* **264**, 19514–19527
- Buisine, M.-P., Janin, A., Maunoury, V., Audié, J.-P., Delescaut, M.-P., Copin, M.-C., Colombel, J.-F., Degand, P., Aubert, J.-P. and Porchet, N. (1996) *Gastroenterology* **110**, 84–91
- Buisine, M.-P., Devisme, L., Savidge, T. C., Gespach, C., Gosselin, B., Porchet, N. and Aubert, J.-P. (1998) *Gut*, in the press
- Balagué, C., Audié, J.-P., Porchet, N. and Real, F. X. (1995) *Gastroenterology* **109**, 953–964
- Aubert, J.-P., Porchet, N., Crépin, M., Duterré-Coquillaud, M., Vergnes, G., Mazzuca, M., Debuire, B., Petitprez, D. and Degand, P. (1991) *Am. J. Respir. Cell Mol. Biol.* **5**, 178–185
- Rose, M. C., Kaufman, B. and Martin, B. M. (1989) *J. Biol. Chem.* **264**, 8193–8199
- Wickens, M. and Stephenson, P. (1984) *Science* **226**, 1045–1051
- Mount, S. M. (1982) *Nucleic Acids Res.* **10**, 459–472
- Pigny, P., Pratt, W. S., Laine, A., Leclercq, A., Swallow, D. M., Nguyen, V. C., Aubert, J.-P. and Porchet, N. (1995) *Hum. Genet.* **96**, 367–368
- Desseyn, J.-L., Buisine, M.-P., Porchet, N., Aubert, J.-P. and Laine, A. (1998) *J. Mol. Evol.* **46**, 102–106
- Hennighausen, L. G., Sippel, A. E., Hobbs, A. A. and Rosen, J. M. (1982) *Nucleic Acids Res.* **10**, 3733–3744
- Pigny, P., Van Seuningen, I., Desseyn, J.-L., Nollet, S., Porchet, N., Laine, A. and Aubert, J.-P. (1996) *Biochem. Biophys. Res. Commun.* **220**, 186–191
- Velcich, A., Palumbo, L., Selleri, L., Evans, G. and Augenlicht, L. (1997) *J. Biol. Chem.* **272**, 7968–7976