J. Biochem. 128, 233–243 (2000)

Human *MUC4* Mucin cDNA and Its Variants in Pancreatic Carcinoma[1]

Amit Choudhury,[*] Nicolas Moniaux,[*,‡] John P. Winpenny,[†,§] Michael A. Hollingsworth,[†] Jean-Pierre Aubert,[‡] and Surinder K. Batra[*,2]

[*]Department of Biochemistry and Molecular Biology, [†]Eppley Institute for Research in Cancer and Allied Diseases, University of Nebraska Medical Center, Omaha, NE 68198-4525, USA; [‡]Unité 377 INSERM, Place de Verdun, 59045 Lille Cedex and Laboratoire de Biochimie et de Biologie Moléculaire de l'Hôpital C. Huriez, 59037 Lille Cedex, France; and [§]School of Sciences, University of Sunderland, Darwin Building, Wharncliffe Street, Sunderland, SR1 3SD, UK

Received April 21, 2000; accepted May 24, 2000

The human *MUC4* gene is not expressed in normal pancreas; however, its dysregulation results in high levels of expression in pancreatic tumors. To investigate the tumor-associated expression, *MUC4* cDNA was cloned from a human pancreatic tumor cell line cDNA expression library using a polyclonal antibody raised against human deglycosylated mucin and RT-PCR. Pancreatic *MUC4* cDNA shows differences in 12 amino acid residues in the non-tandem repeat coding region with no structural rearrangement as compared with tracheal *MUC4*. The full-length *MUC4* cDNA includes a leader sequence, a serine and threonine rich non-tandem repeat region, a central large tandem repeat domain containing 48 bp repetitive units, regions rich in potential *N*-glycosylation sites, two cysteine-rich domains, EGF-like domains, and a transmembrane domain. We also report the presence of a new EGF-like domain in *MUC4* cDNA, located in the cysteine-rich region upstream from the first EGF-like domain. Four distinct splice events were identified in the region downstream of the central tandem repeat domain that generate three new *MUC4* cDNA sequences (sv4, sv9, and sv10). The deduced amino acid sequences of two of these variants lack the transmembrane domain. Furthermore, two unique forms of *MUC4* (*MUC4/Y* and *MUC4/X*) generated as a result of alternative splicing lack the salient feature of mucins, the tandem repeat domain. A high degree of polymorphism in the central tandem repeat region of *MUC4* was observed in various pancreatic adenocarcinoma cell lines, with allele sizes ranging from 23.5 to 10.0 kb. *MUC4* mRNA expression was higher in differentiated cell lines, with no detectable expression in poorly differentiated pancreatic tumor cell lines.

Key words: MUC4 mucin, pancreatic carcinoma, polymorphism, variants.

Mucins are important molecules produced by secretory epithelial cells for the lubrication and protection of ducts and lumen within the human body. Mucins contain a high percentage of serine, threonine, proline, alanine, and glycine residues, and are heavily glycosylated (*O*-linked) (*1*, *2*). Mucins from different organ sites generally show similar, but not identical, amino acid compositions. However, mucins from different organs usually have distinct physical and biochemical properties, including oligosaccharide structure and composition, sulfation, molecular weight, and viscoelastic properties (*3*). The common structural feature among all mucins is the presence of tandemly repeated amino acid regions rich in serine and threonine. In general, the *O*-glycosylated tandem repeat regions are not conserved between species in terms of either size or amino-acid sequence; however, flanking amino-terminal and carboxyl-terminal regions typically show some similarity, presumably reflecting functional relatedness.

Currently, ten human mucin genes have been identified and designated *MUC1-4, MUC5B, MUC5AC, MUC6-7, MUC11,* and *MUC12* (*4–13*), with the complete deduced primary amino acid sequences of *MUC1, MUC2, MUC4, MUC5B,* and *MUC7*. Mucins have been classified into two main categories, secreted (gel-forming and non-gel-forming) and membrane-bound. The membrane-associated mucin family comprises *MUC1, MUC3, MUC4,* and *MUC12*.

MUC1 and rat SMC (sialomucin complex, human *MUC4* homologue) are well characterized and exist as both soluble and membrane-associated forms (*14*). MUC1 mucin is expressed by all human glandular epithelial cells and plays a distinct role in adhesion modulation and intracellular signalling (*1*). Alternate splicing generates a secreted form MUC1/SEC and a membrane-bound form lacking the tandem repeat domain, MUC1/Y (*15*). In rat *Muc4* two EGF-like domains (EGF1 and EGF2) are present. The EGF1 domain has been demonstrated to associate with the c-erbB-2 growth factor receptor (*16*), indicating the possibility of growth factor receptor modulation by membrane-anchored mucins containing EGF-like domains. Alternative splice forms (secreted and membrane-associated) of another

[1] This work was supported by a grant from the National Institutes of Health CA 78590.

[2] To whom correspondence should be addressed: Phone: +1-402-559-5455, Fax: +1-402-559-6650, E-mail: sbatra@unmc.edu

Abbreviations: SMC, sialomucin complex; ASGP, ascites sialoglycoprotein; RT-PCR, reverse transcriptase polymerase chain reaction; EGF, epidermal growth factor; VNTR, variable number of tandem repeats; UTR, untranslated region; hEGF, human EGF.

membrane-bound mucin, MUC3, and more recently, MUC4, have been documented (*17*, *18*).

The apomucins *MUC1, MUC5B,* and *MUC6* are expressed in a normal human pancreas; *MUC1* and *MUC5B* are overexpressed whereas *MUC2* and *MUC5AC* are expressed only in human pancreatic tumors (*19–21*). In pancreatic tumor cell lines, high levels of *MUC1*, extremely low levels of *MUC2* mRNA, and no detectable *MUC3* mRNA have been reported. *MUC4* is expressed at high levels in human pancreatic tumors and cell lines, with undetectable levels in normal pancreas (*19*, *22*). So far, no MUC4-protein expression data are available, as MUC4-specific monoclonal antibodies are not available. Furthermore, the *MUC4* RNA signal on Northern blots appears as a polydisperse pattern due to the degradation of the RNA during sample preparation (*23*).

Cancer of the pancreas represents a useful model for study, because this tumor type frequently contains mucin-producing cells, while corresponding normal tissues do not. We previously cloned and sequenced a full-length cDNA for a pancreatic tumor mucin *MUC1*, on which the DU-PAN-2 antigen was expressed (*4*, *24*). High levels of *MUC4* mRNA expression by pancreatic tumor cells suggest the possibility that mucin carbohydrate epitopes defined by antibodies such as DU-PAN-2 may be expressed on a second mucin core protein produced by pancreatic tumor cells.

In this paper, we describe the cloning, sequencing, and identification of an additional EGF-like domain and putative variant forms of human pancreatic tumor–associated mucin *MUC4*. We also investigated the level of expression of the *MUC4* transcripts by Northern blot analysis and the occurrence of polymorphism in the *MUC4* gene by Southern blot analysis in a large panel of pancreatic tumor cell lines.

MATERIALS AND METHODS

Preparation and Screening of a MUC4 cDNA Library—cDNA libraries from HPAF and SUIT-2 pancreatic tumor cell lines were prepared as described earlier (*4*). The λ-gt11 HPAF cDNA expression library was screened using polyclonal rabbit antiserum against human deglycosylated mucins (*25*). Positive plaques were detected with a commercially available alkaline phosphatase-conjugated goat anti-rabbit antibody (Promega, Madison, WI). The screening conditions recommended by the manufacturer were followed except that the primary and secondary antibodies were incubated with the filters in phosphate-buffered saline (PBS) plus 5% nonfat dry milk for 1 h each at 25°C, followed by three washes in PBS after both incubations. Specific reactivity was confirmed by three subsequent subcloning and screening cycles. Seven positive clones were initially selected by antibody screening of 3×10^5 plaques. Among the isolated clones, one showed >98% similarity to the published *MUC4* cDNA sequence and was named DUM4.6. Using this DNA insert, 12 additional plaques were selected from cDNA libraries prepared from HPAF and another pancreatic tumor cell line, SUIT-2. Nitrocellulose membranes were used to obtain plaque lifts. These membranes were pre-hybridized and hybridization was performed with 2.5×10^5 cpm/membrane overnight at 42°C. Inserts of positive phages were subcloned into pBluescript KS+ vector.

Derivation of MUC4 Genomic Clones and Their Sequencing—Genomic clone for *MUC4* were obtained by screening a P1 human foreskin fibroblast genomic DNA library and a bacterial artificial chromosome (BAC) library (Genome Systems, St. Louis, MO) with a *MUC4* cDNA probe (M4S4.9) isolated from the HPAF cDNA library. The clones were digested with *Eco*RI, *Bam*HI, *Pst*I, and *Eco*RI + *Pst*I, and the resulting fragments were analyzed by Southern blot analysis for hybridization with specific MUC4 probes. Fragments were either subcloned first in phage vector λ-DASH II and subsequently cloned in pGEM-11Zf or cloned directly into pCR2.1 vector. These cloned DNA fragments were either partially or fully sequenced. Sequence information was analyzed using Genetics Computer Group software (Madison, WI).

Southern Blot Analysis—Genomic DNA from human pancreatic tumor cell lines was digested with *Eco*RI and *Pst*I restriction endonucleases. Fragments were separated by electrophoresis in 0.8% agarose gels and transferred to nylon membranes. The blots were hybridized with *MUC4* tandem repeat cDNA probe.

Isolation of RNA—Total cellular RNA from various tumor cell lines was isolated by the guanidine isothiocyanate and cesium chloride (CsCl) cushion ultracentrifugation method (*26*). Cells were washed twice with ice-cold PBS (pH 7.4) and lysed with a solution containing 4 M guanidine isothiocyanate, 0.05 M sodium acetate, and 250 mM 2-mercaptoethanol. Total RNA was recovered *via* sedimentation through a 5.7 M CsCl, 0.025 M sodium acetate cushion in a Beckman SW40Ti rotor centrifuged at 32,000 rpm for 18 h. RNA pellets were resuspended in 0.3 M sodium acetate and precipitated with ethanol (*26*, *27*).

Oligonucleotide Primers—The primers designed from HPAF *MUC4* cDNA, submitted to the EMBL data bank under accession number AJ276359, are as follows: sense, P1 (1–21); antisense, P2 (2899-2919); sense, P3 (3156–3176); antisense, P4 (6678–6699); sense, P5 (2650–2670); antisense, P6 (3006–3024); sense, P7 (3036–3059); antisense, P8 (3470–3491); sense, P9 (33–53); sense P10 (58–79); antisense P11 (5838–5858). The positions and directions of these primers are shown in Fig. 2.

RT-PCR Cloning of MUC4—Reverse transcription reactions using Superscript II RNnase H^- reverse transcriptase and reagents (Gibco BRL, MD) according to the manufacturer's instructions. Total RNA (1–2 μg) from an HPAF pancreatic tumor cell line was used to prepare a single strand cDNA using oligo dT primer (Pharmacia Biotech. Piscataway, NJ). The 5′-end and the tandem repeat flanking regions were amplified using sense primers P1, P5, and P7 and antisense primers P2, P6, and P8, respectively (Fig. 2), by classical RT-PCR. The reaction was performed in a programmable thermal controller (PTC-100, MJ Research, Watertown, MA). PCR parameters were 94°C for 4 min, followed by 30 cycles at 94°C for 45 s, annealing at 60°C for 45 s, and extension at 72°C for 1 min, followed by a final elongation at 72°C for 15 min. PCR was performed using 2.5 units of *Taq* DNA polymerase.

The 3′-end was amplified using the Expand™ Long Template PCR System (Boehringer Mannheim, GmbH, Germany) with the sense primer P3 and the antisense primer P4 (Fig. 2). The PCR amplification reaction mixture consisted of 5 μl of cDNA, 10 mM sodium dNTPs, 0.4 pmol of each primer, 5 μl of 10× Expand™ Long Template PCR

buffer 3, 0.75 mM $MgCl_2$, and 2.5 units of enzyme mixture in a final volume of 50 μl. The reaction was performed in a programmable thermal controller. The parameters for the reaction were 94°C for 2 min, followed by 30 cycles at 94°C for 30 s, annealing at 60°C for 45 s, and elongation at 71°C for 4 min, during which the elongation time in the last 20 cycles was extended for 40 s for each new cycle, followed by a final elongation at 71°C for 15 min. The amplified products were separated electrophoretically in a 1% agarose gel and visualized with ethidium bromide. RT-PCR products were subcloned into the Original TA Cloning Kit (Invitrogen, CA) without further purification. At least 20 clones and splice events only were sequenced for each amplified PCR product.

Sequence Analyses—Sequences were determined on both strands by automatic sequencing using internal primers with an ABI Prism model 377 XL automatic sequencer. Analyses of nucleic acid and protein sequences were performed with GCG software. The nucleotide sequences reported in this paper were submitted to the EMBL Data Bank with the accession numbers AJ276359, AJ400633, and AJ277412.

RESULTS

MUC4 cDNA Isolation and Sequencing—Approximately

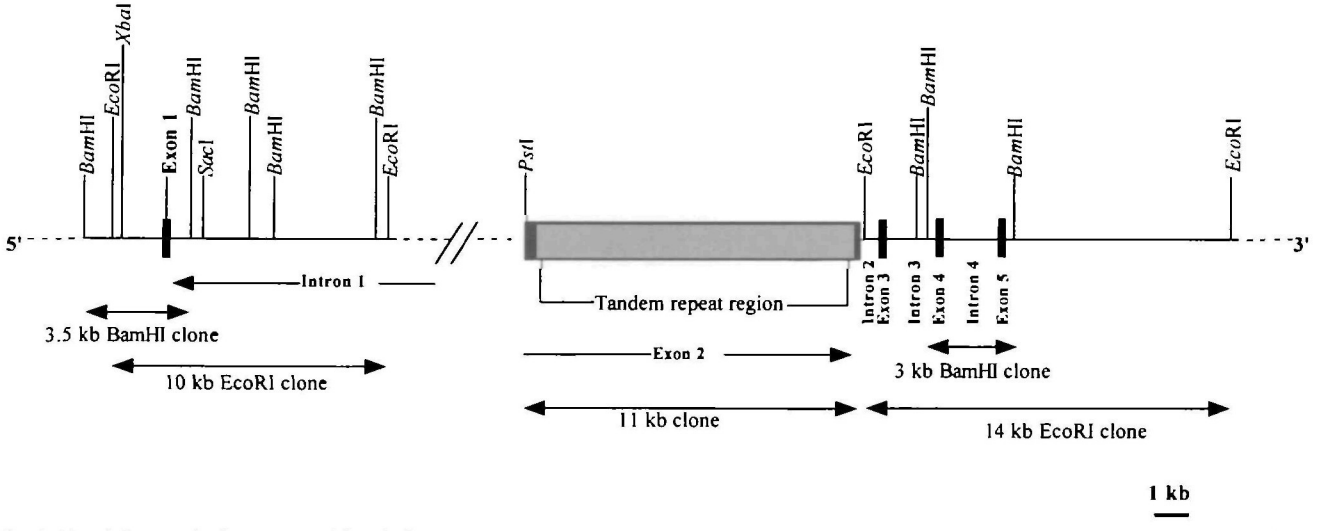

Fig. 1. **Partial restriction map of DNA fragments (3.5 kb *Bam*HI, 10 kb *Eco*RI, 11 kb, 3 kb *Bam*HI, 14 kb *Eco*RI) isolated from genomic clones (P1 and BAC).** The solid bold lines and blocks perpendicular to the thin horizontal lines (intronic sequence) represent exonic sequences.

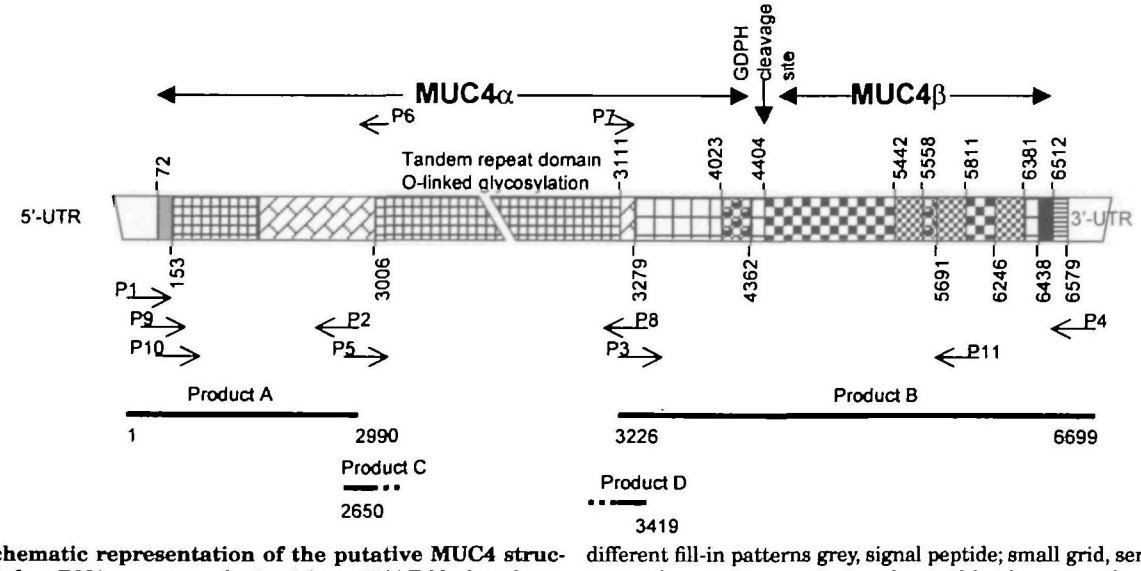

Fig. 2. **Schematic representation of the putative MUC4 structure and the cDNA sequence derived from HPAF.** Nucleotide positions are indicated by the numbers. The location of primers P1–P11 is indicated by the line with arrows. The coding regions are represented by boxes with fill-in patterns whereas the non-coding regions are indicated as empty boxes. The various domains are indicated with different fill-in patterns grey, signal peptide; small grid, serine, threonine-rich repetitive sequences; diagonal brick, serine, threonine-rich non-repetitive sequences; large grid, unique sequence; sphere, cysteine-rich domain; large checker board, potential *N*-linked glycosylation sites; small checker board, EGF-like domain; black, potential transmembrane sequence; horizontal lines, cytoplasmic tail.

500,000 recombinants were screened with rabbit polyclonal antiserum against deglycosylated mucins and twenty plaques were isolated. The DNA was purified from the phages, subcloned, and sequenced. Out of 11 clones, two showed homology with our earlier published *MUC1* sequence (*4*). One clone (DUM4.6; accession number AF177925) showed similarity to the published *MUC4* cDNA sequence (*7*).

The DUM4.6 clone was used for further screening of HPAF and SUIT-2 cDNA libraries. The clones obtained were DUM3.4, DUM4.9, M4H24.5, M4H1.2, M4H4.2, M4H6.1, M4H3.1, M4S1.1, M4S9.3, M4S2.1, M4S4.9, and M4JC5.1 with insert sizes of 0.5, 0.7, 0.9, 1.6, 4.0, 1.4, 0.4, 2.4, 5.4, 1.2, 4.1, 4.4, and 5.9 kb, respectively. All the cDNA clones showed a *MUC4* tandem repeat sequence at their 5′-ends and unique sequences at the 3′-ends. One of the

TABLE I. **Summary differences between the cDNA sequences of HPAF and tracheal MUC4.**

	Differences between pancreatic and tracheal cDNA sequences						
	Position of altered nucleotide			Nucleotide bases	At codon level	At AA level	Position of altered AA
Pancreatic (AJ276359)	1413	1802	Tracheal (AJ000281)	C for T	CTA for TTA	Leu for Leu	448
	1656	2045		A for G	AGG for GGG	Arg for Gly	529
	1838	2227		T for C	TAT for TAC	Try for Tyr	589
	1846	2235		T for C	ATT for ACT	Ile for Thr	592
	2008	2397		G for A	CGA for CAA	Arg for Gln	646
	2190	2579		A for G	AAT for GAT	Asn for Asp	707
	2232	2621		G for A	GCA for ACA	Ala for Thr	721
	2826	3215		G for A	GGA for AGA	Gly for Arg	919
	2969	3358		G for A	GGG for GGA	Gly for Gly	966
	3848	738	Tracheal (AJ010901)	G for C	AAG for AAC	Lys for Asn	1259
	3883	773		T for A	TTT for TAT	Phe for Tyr	1271
	3984	874		G for A	GGG for AGG	Gly for Arg	1305
	4741	1631		C for T	GCC for GTC	Ala for Val	1557
	4745	1635		C for G	TCC for TCG	Ser for Ser	1558
	4803	1693		T for G	TCC for GCC	Ser for Ala	1578
	6182	3072		T for C	ATT for ATC	Ile for Ile	2037
	6187	3077		G for A	GGG for GAG	Gly for Glu	2039
	6350	3240		G for A	GAG for CAA	Gln for Gln	2093

AA, amino acid.

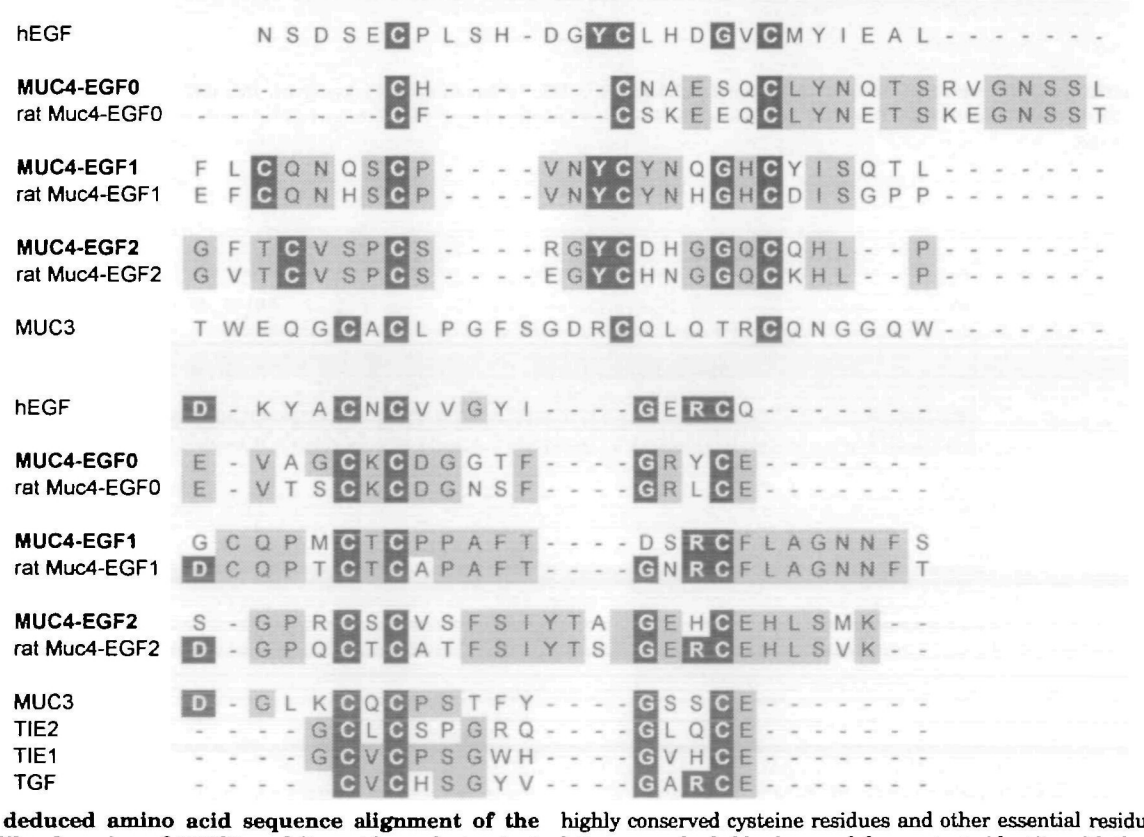

Fig. 3. **The deduced amino acid sequence alignment of the three EGF-like domains of MUC4 and its rat homologue (rat Muc4) with hEGF and the EGF-like motifs present in several proteins.** Bold type white letters on a dark background represent highly conserved cysteine residues and other essential residues, and letters on a shaded background demonstrate identity with the rat homologue and other proteins. Hyphens indicate gaps inserted to optimize the alignment.

clones, M4S4.9, contained introns as well as part of a sequence corresponding to exon 3 in the published *MUC4* cDNA sequence from trachea (*28*). Specific primers could not be designed to sequence within the tandem repeat sequence, as they would bind to multiple sites. The tandem repeat region of the M4S4.9 clone was further confirmed by enzymatic digestion with the *Dde*I enzyme that cuts once within each tandem repeat unit. A complete digestion of M4S4.9 into 48 bp fragments along with a non-tandem repeat region was detected (data not shown).

Characterization of Genomic Fragments—Five DNA fragments were subcloned from the genomic clones digested with *Eco*RI, *Bam*HI, and *Eco*RI+*Pst*I, hybridizing to *MUC4* cDNA probes. Four of these DNA fragments were isolated from P1 and one from BAC human genomic libraries. The alignment and partial restriction maps of these subclones are shown in Fig. 1. Sequence analysis of the DNA fragments subcloned from the genomic clones revealed exonic and intronic regions. The exonic regions (exon 1–5) shared 100% sequence homology with the recently published tracheal *MUC4* cDNA sequence (*28*, *29*). The splice donor and acceptor sequences for intron 1–4 followed convention, conforming to the gt-ag rule (data not shown). The subclones, 10 kb *Eco*RI and 3.5 kb *Bam*HI, showed overlapping sequences with 5′ UTR and the 82 bp common signal sequence (exon 1). The sequence analysis of 14 kb *Eco*RI revealed the presence of at least three *MUC4* exons, *i.e.*, exons 3–5 (*28*). The 3 kb *Bam*HI clone overlapped with 14 kb *Eco*RI and had exons 4 and 5 in common. The 11 kb *Pst*I +*Eco*RI fragment from the BAC clone consisted of an intact tandem repeat domain of exon 2. The entire tandem repeat domain could not be sequenced because of the repetitive nature of the sequence. However, digestion with the *Dde*I yielded 48 bp fragments and 96 bp fragments that corresponded to one or two tandem repeat units.

RT-PCR Amplification and Generation of Full-Length Coding Sequence of MUC4 from HPAF Cell Line—To obtain the full length cDNA sequence encoding *MUC4*, primers were designed in the exonic regions of the genomic sequences as well as in the 3′-UTR sequence (AJ010901). The sense primer P1 was designed from sequence information of the two DNA fragments subcloned from the genomic clone spanning the 5′-UTR (Fig. 2). An antisense primer P2 was designed in the non–tandem repeat exonic sequence obtained from the DNA fragment subcloned from the BAC clone, which contains a 274 bp non–repetitive coding sequence located upstream of the tandem repeat region (Fig. 1). Using classical RT-PCR with total RNA from the HPAF cell line, a 2991 bp fragment was amplified, corresponding to exon 2 or the 5′-end of the *MUC4* cDNA (Figs. 2 and 6A). For the second set of primers, a sense primer (P3) was designed in the exonic sequence (*i.e.*, in exon 3) of the 14 kb *Eco*RI subclone (Fig. 2). An antisense primer (P4) was designed in the 3′-UTR from the sequence information obtained from the recently published tracheal *MUC4* cDNA sequence (*28*). Using expand long RT-PCR on the total RNA isolated from the HPAF cell line, a 3,474 bp fragment was amplified (Product B; Figs. 2 and 6A).

Sequences from the DNA fragments subcloned from the genomic clones and the RT-PCR amplified products were compared, and the full-length *MUC4* cDNA sequence was deduced (Fig. 2). The HPAF *MUC4* showed structural motifs similar to those found in the tracheal MUC4. The differences observed between the HPAF *MUC4* cDNA and tracheal *MUC4* sequence are summarized in Table I. Six amino acid variations were observed in the deduced *MUC4* amino-acid sequence (submitted to the EMBL data bank

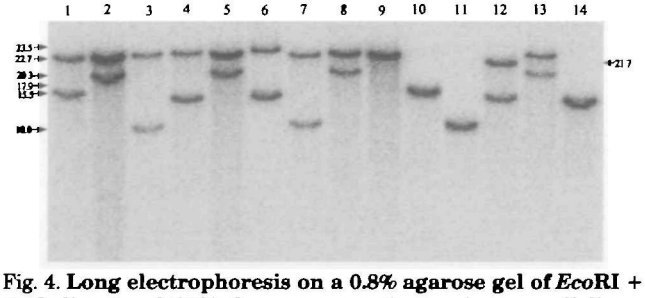

Fig. 4. **Long electrophoresis on a 0.8% agarose gel of *Eco*RI + *Pst*I digests of DNA from pancreatic carcinoma cell lines.** BxPc3 (1), Panc89 (2), AsPc-1 (3), QGP1 (4), Capan1 (5), Capan2 (6), Panc-1 (7), T3M4 (8), HPAF (9), Colo357 (10), MiaPaCa (11), S2CP9 (12), HPAC (13), Hs766T (14). The membrane was probed with a ^{32}P-labeled *MUC4* tandem repeat cDNA probe.

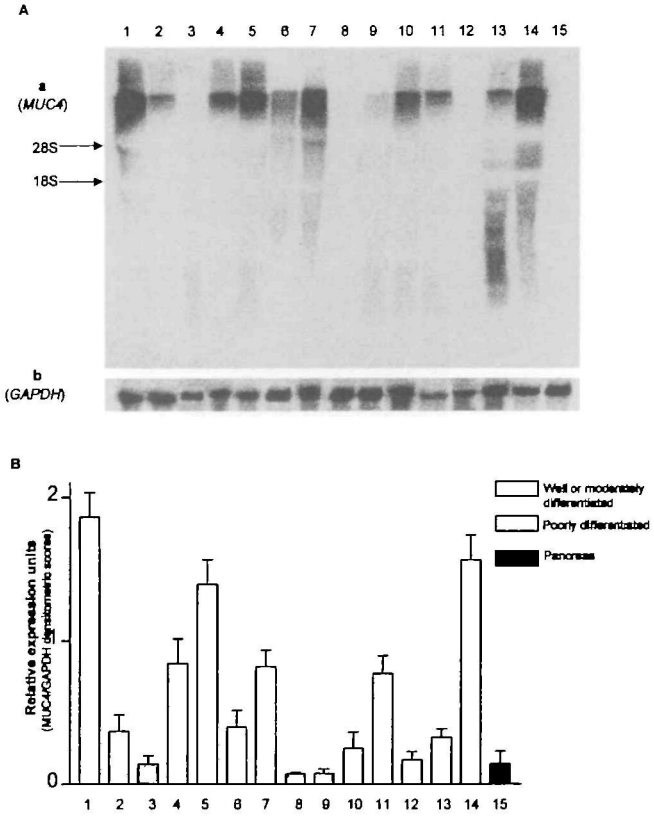

Fig. 5. **A: Northern blot of total cellular RNA (20 μg) separated in a 1% agarose/formaldehyde gel.** Pancreatic carcinoma cell lines HPAF (1), QGP1 (2), HCG-25 (3), Colo357 (4), Capan1 (5), Capan2 (6), HPAC (7), Panc-1 (8), MiaPaCa (9), T3M4 (10), S2CP9 (11), AsPc-1 (12), Panc-89 (13), BxPc-3 (14) and normal human pancreas (15). a, probed with ^{32}P-labeled *MUC4* TR cDNA probe. b, the same membrane shown in (a) probed with a GAPDH cDNA probe. **B: Densitometric values (±SE) for the *MUC4*, normalized to *GAPDH* band intensity in three different experiments were determined using the Molecular Dynamics Image Quant software program.**

under accession number AJ276359): Arg, Ile, Arg, Asn, Ala, and Gly at positions 529, 592, 646, 707, 721, and 919, respectively, at the NH_2-terminus. Among the few differences found at the COOH-terminus were Lys, Phe, Gly, Ala, Ser, and Gly at positions 1259, 1271, 1305, 1557, 1578, and 2039, respectively. Among these, the G for A change was detected in 7 of 18 differences at the nucleotide level (Table I). We also observed a deletion of A after the 43rd bp in the 5′-UTR.

Analysis of the coding sequence derived from pancreatic *MUC4* cDNA revealed the presence of another EGF-like domain (EGF0; 5442–5558) apart from the two previously described (EGF1 and EGF2) (*28*). We have also identified a similar EGF-like domain in the ASGP-2 subunit of rat Muc4 (*30*) and, therefore, have named it rat Muc4-EGF0. The similarity between the EGF0 domain of human MUC4 and rat Muc4 is 64%. Hence, the putative *MUC4*β subunit consists of an EGF0 domain followed by a short cysteine-rich domain (5559–5690; Fig. 2.), and includes two more EGF-like domains (EGF1 and EGF2). The spacing of the cysteines in the EGF0 domains of MUC4 and rat Muc4, and to a lesser extent the overall sequence, are similar to that in EGF-like structural motifs present in several other proteins (Fig. 3).

Polymorphism Studies—To investigate polymorphism in the tandem repeat domain of the MUC4 gene, which contains 48 bp repetitive units, purified DNA from pancreatic tumor cell lines was digested with *Eco*RI and *Pst*I enzymes, because these sites flank the tandem repeat region (Fig. 1). With the MUC4 tandem repeat cDNA probe, Southern blot analysis revealed two distinct alleles in 10 out of 14 cell lines; BxPc-3, Panc89, AsPc-1, QGP-1, Capan-1, Capan-2, Panc-1, T3M4, S2CP9, and HPAC. Four cell lines (HPAF, MiaPaCa, Hs766T, and Colo357) showed a single band and appeared to be homozygous or hemizygous for the *MUC4* gene. The size of these alleles varied from 23.5–10 kb. Seven different allele sizes were noticed (Fig. 4), with the 22.7 kb allele being the most common form. No *MUC4* gene amplification was observed among these tumor cell lines. The length of the *Pst*I/*Eco*RI fragment containing the tandem repeat region in the BAC clone corresponds to the 10 kb allele.

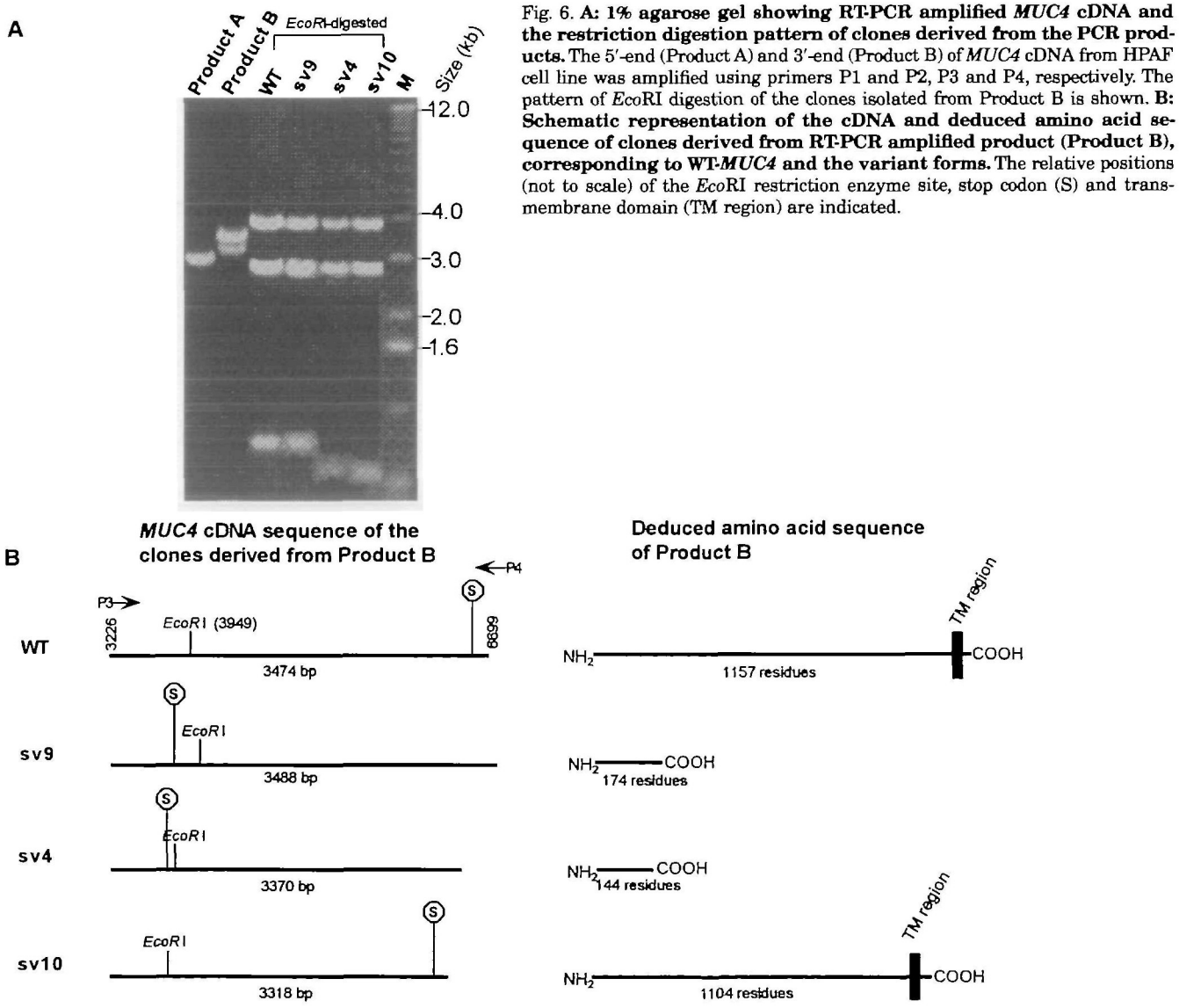

Fig. 6. **A: 1% agarose gel showing RT-PCR amplified *MUC4* cDNA and the restriction digestion pattern of clones derived from the PCR products.** The 5′-end (Product A) and 3′-end (Product B) of *MUC4* cDNA from HPAF cell line was amplified using primers P1 and P2, P3 and P4, respectively. The pattern of *Eco*RI digestion of the clones isolated from Product B is shown. **B: Schematic representation of the cDNA and deduced amino acid sequence of clones derived from RT-PCR amplified product (Product B), corresponding to WT-*MUC4* and the variant forms.** The relative positions (not to scale) of the *Eco*RI restriction enzyme site, stop codon (S) and transmembrane domain (TM region) are indicated.

MUC4 Expression in Pancreatic Tumor Cell Lines—Steady state expression levels of *MUC4* mRNA were investigated in 14 pancreatic adenocarcinoma cell lines with different grades of differentiation and in normal pancreas. Upon hybridization with a tandem repeat *MUC4* cDNA probe, a high molecular weight band (>10 kb) was detected on the Northern blots (Fig. 5). Several cell lines, including HPAF, HPAC, Capan1, Capan2, Colo357, T3M4, QGP1, S2CP9, BxPc-3, and Panc 89, showed moderate to high levels of *MUC4* RNA expression. A low level of expression was observed in AsPc-1, and cell lines HCG-25, Panc1, and MiaPaCa showed no detectable mRNA for the *MUC4* gene. No expression of *MUC4* RNA was detected in the normal human pancreas.

MUC4 Splice Variants—Amplification of the 3′-end of the *MUC4* cDNA isolated from HPAF cells using different sets of primers revealed some deletions from the published tracheal sequence. Recently, five secreted forms (sv1, sv2, sv3–5, sv6, and sv7) and two membrane-associated forms (sv0 and sv8) of *MUC4* in human testis were reported (*18*). These variant forms were identified by RT-PCR amplification of the 3′-end of MUC4 cDNA using polyA$^+$ RNA. We further investigated the existence of alternate splice forms in pancreatic tumor cells (HPAF) and have used similar nomenclature for the variants. Using primers (P3 and P4; Fig. 2) used in our previous report (*18*), the 3′-end of *MUC4* cDNA (3226–6699, 3,474 bp; AJ276359) was amplified by an expanded long RT-PCR procedure using total RNA from the HPAF cell line. Instead of a single band, diffuse doublet bands in the range of 3.2 to 3.5 kb were detected (Fig. 6A). The PCR product was subcloned directly into pCR2.1 vector, and clones of different sizes were isolated and sequenced. Restriction digestion of these clones with *Eco*RI generated an RFLP pattern that is different for each variant (Fig. 6A).

Twelve of the 20 clones had sequences similar to our pancreatic *MUC4* isolated from HPAF cells termed wild type *MUC4* cDNA (WT-*MUC4*), with differences in 9 bp from tracheo-bronchial *MUC4* cDNA as shown in Table I. Eight of the 20 clones screened had deletions in the WT-*MUC4* sequence. The four splice events found are summarized in Table II, and are responsible for generating three putative variant forms of *MUC4* cDNA (Fig. 6B). We detected one insertion (I1) and three deletion (D1–3) events and the splice junctions of two of these events are shown in Fig. 7. The genomic organization after exon 5 is not known, therefore, we cannot define the splice junction for events D2 and D3 (Table II). Because of these splice events (I1 and D1–3), the reading frames of the translated 3′-end of *MUC4* for some of these variants (sv9-MUC4 and sv4-MUC4) are altered, which may eventually lead to protein variants of MUC4 (Table II, Fig. 6). The sv10-MUC4 showed an in-frame deletion of residue 156 in the region of the unique sequence.

From our polymorphism studies, we expect that the size of exon 2 in HPAF will be approximately 25 kb, of which 22

TABLE II. **Summary of events detected in the 3′-end amplified RT-PCR product (Product B) of *MUC4* cDNA from HPAF cell.**

Form	Frequency	*Eco*RI digested product size (bp)	Splice event(s)			Product		Position of SC in the deduced AA sequence	Reading frame	Putative form
			Type	Length (bp)	Position in WT-*MUC4*	size (bp)	Accession number			
WT	12/20	724+2750	ND	—	—	3474	AJ276359	1158	In-frame	Membrane bound
sv9	2/20	738+2750	I1	14	Between 3285-3286	3488	AJ400633	175	Modified	secreted
sv4	2/20	648+2722	D1	76	3420-3495	3370	AJ242544	145	Modified	secreted
			D2	28	3842-3869					
sv10	4/20	568+2750	D3	156	3584-3739	3318	AJ277412	1105	In-frame	Membrane bound

ND, not detected; I, insertion; D, deletion; WT, wild-type *MUC4*; sv, splice variant; SC, stop codon; AA, amino acid; bp, nucleotide base pair.

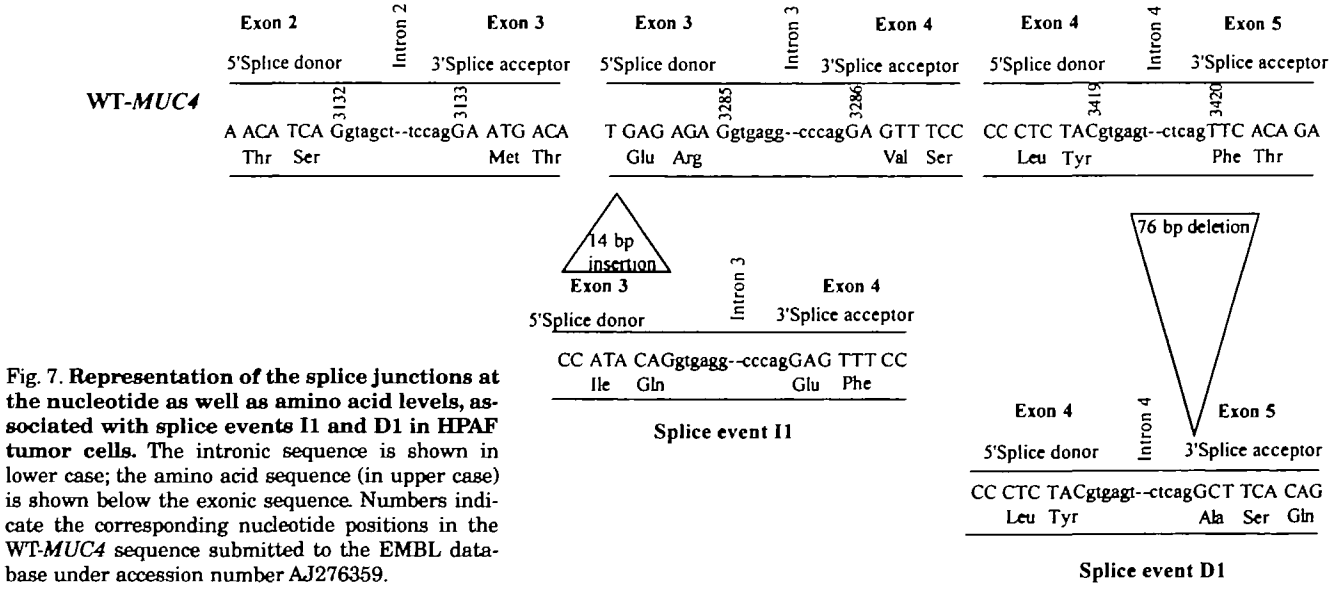

Fig. 7. **Representation of the splice junctions at the nucleotide as well as amino acid levels, associated with splice events I1 and D1 in HPAF tumor cells.** The intronic sequence is shown in lower case; the amino acid sequence (in upper case) is shown below the exonic sequence. Numbers indicate the corresponding nucleotide positions in the WT-*MUC4* sequence submitted to the EMBL database under accession number AJ276359.

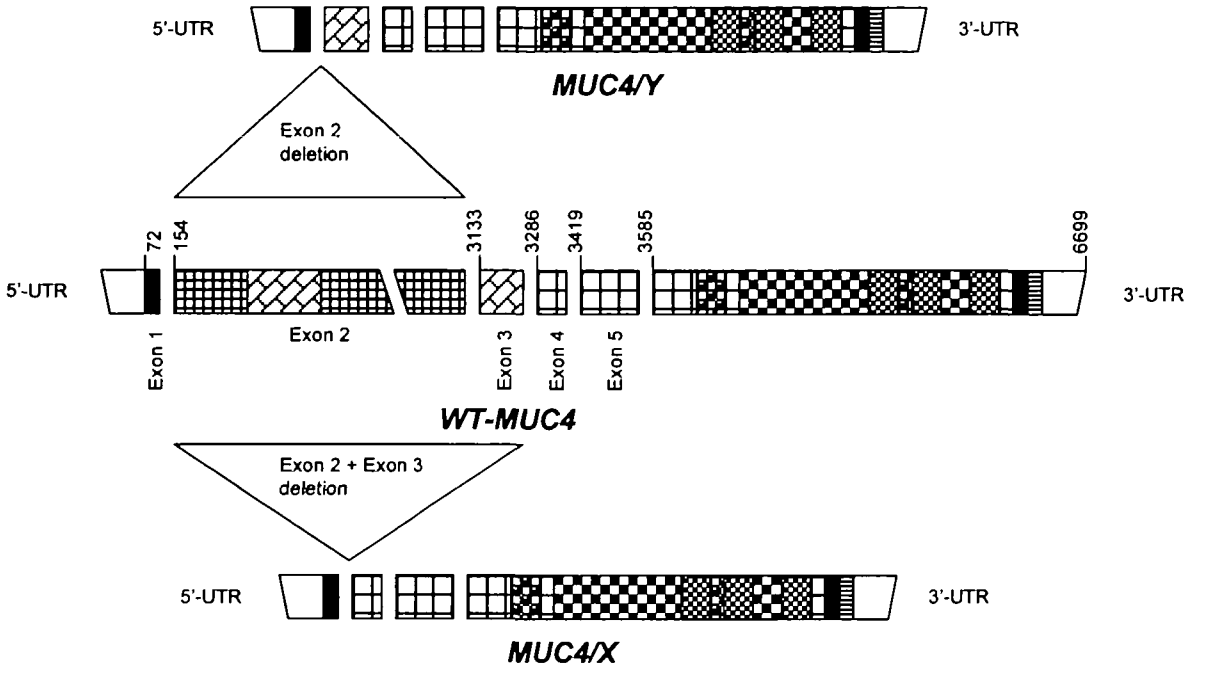

Fig. 8. **Schematic representation of the splice events that generate *MUC4/Y* and *MUC4/X*.** The boxes with fill-in patterns represent the coding sequences; empty boxes denote non-coding regions. The boxes with fill-in patterns separated from each other represent exonic sequences for which the genomic organization is known. The various domains are indicated with different fill-in patterns: grey, signal peptide; small grid, serine, threonine-rich repetitive sequences; diagonal brick, serine, threonine-rich non-repetitive sequences; large grid, unique sequence; sphere, cysteine-rich domain; large checker board, potential *N*-linked glycosylation sites; small checker board, EGF-like domain; black, potential transmembrane sequence; horizontal lines, cytoplasmic tail.

kb is the tandem repeat array. It was therefore impossible to amplify the 5'-end along with the tandem repeat region for each variant and subclone the complete cDNA corresponding to each splice variant. RT-PCR using total RNA from HPAF cells was performed to investigate the presence of potential alternative splice events in the non-tandem repeat regions, junctions between exon 1 and exon 2, exon 2 and exon 3, as well as within exons 2 and 3 (Product A, C, D; Fig. 2). No alternative splice events were detected in these regions after the screening of 25 different subclones from each PCR product (Product A, C, and D) cloned in the pCR2.1 vector. Therefore, we assume that, unlike the 3'-end of the *MUC4* cDNA (Product B, Fig. 2), the 5'-end and the central tandem repeat domain of each splice variant are similar to those reported in the WT-*MUC4* sequence (deposited in GenBank under accession number AJ276359) as described in Fig. 2.

We also identified two new variants of *MUC4* lacking the tandem repeat domain in HPAF tumor cells. RT-PCR amplification using primer P4 and P9 (Fig. 2) on total RNA isolated from HPAF cells revealed undetectable levels of the product. Therefore, re-amplification was done using a set of nested primers (P10 and P11, Fig. 2). Two PCR-products of 2821 bp and 2668 bp were detected. These products were cloned, sequenced and compared with WT-*MUC4*. Both these clones showed an in-frame deletion of exon 2 and exon 2-3, respectively (Fig. 8). We have reported similar variants in human lung tumors (*18*) named *MUC4/Y* (exon 2 deleted) and *MUC4/X* (exon 2-3 deleted). The remaining downstream sequence of the *MUC4* cDNA was also amplified using the first PCR product (obtained with primers P4 and P9) as a template and no splice event was found in this region.

DISCUSSION

The *MUC4* mucin is considered to be a homologue of the rat sialomucin complex (SMC) due to their similar structural organizations. SMC is a heterodimeric glycoprotein complex composed of mucin (ascites sialoglycoprotein 1, ASGP-1) and transmembrane ASGP-2 subunits (*30*, *31*). Like SMC, MUC4 may also exist as two subunits: a mucin type known as MUC4α and a growth factor–like transmembrane subunit, MUC4β. *MUC4* is expressed in numerous normal tissues including stomach, ovary, salivary gland, colon, lung, trachea, uterus, and prostate (*32–35*). Furthermore, there is an aberrant expression of *MUC4* in pancreatic tumors and in pancreatic tumor cell lines (*2*, *19*, *22*). However, there is no detectable expression in normal pancreas. It is, therefore, important to determine the *MUC4* sequence in pancreatic tumor cells.

The data presented here reveal differences in the sequence of the full-length *MUC4* cDNA derived from a HPAF tumor cell line compared with the previously published human tracheobronchial mucin *MUC4* sequence (*28*, *29*). However, some of these variations at the amino-acid level are also observed in other pancreatic tumor cell lines such as S2CP9, BxPc-3 and QGP1 (data not shown). Furthermore, these differences do not alter the putative motifs and/or domain structure of this molecule as described in

the previous reports (*28*, *29*). At present, we do not know whether the polymorphism observed at the amino acid level is specific to the pancreatic adenocarcinoma, because we have not compared our sequence with the MUC4 sequence derived from other normal human tissues. The alignment of different genomic clones with the published tracheal *MUC4* cDNA (AJ000281 and AJ010901) sequence and the partial genomic map provides information about exon positions, intronic sequences and the translation start site. The majority of the cDNA clones isolated in the present investigation shared similar tandem repeat sequences at the 5′ end. However, none of the clones had sequences in common at the 3′ end. Comparing these sequences with the genomic clones revealed that the majority of the clones have intronic sequences at the 3′ end. These intronic sequences in the cDNA clones might be derived from the pre-mRNA pools of the cDNA library, a situation frequently encountered in mucin cDNA cloning. Comparing the HPAF cDNA sequence with the tracheal cDNA sequence revealed differences in the translated portion of the cDNA.

We have identified another EGF-like domain (MUC4-EGF0) in the previously reported cysteine-rich domain upstream of the EGF1 domain (*28*). A similar EGF-like motif is present in the cysteine-rich domain of MUC3 (*36*) and shares 46% similarity with the motif present in MUC4 and varying degree of similarity with several other proteins such as TIE1, TIE2, EGF, and TGF. TIE1 and TIE2 represent tyrosine protein kinase receptor precursors, and EGF, and TGF, epidermal growth factor precursor and TGFα growth factor precursor (*37–40*). Among the membrane-bound mucins, MUC1 contains no EGF-like domains (*4*), MUC3 has two EGF-like domains (*41*) and MUC4 has three EGF-like domains. However, the functions of the EGF-like domain have not been demonstrated. The EGF1-domain in ASGP-2, the membrane-associated subunit of rat Muc4, is believed to interact as well as activate the tyrosine kinase $p185^{neu}$ (*16*).

A relatively higher level of *MUC4* expression in well or moderately differentiated tumor cell lines was observed as compared to poorly differentiated tumor cell lines. In normal pancreas, the level of *MUC4* is undetectable. However, the variable levels of expression within similar differentiation grades are likely to be the consequence of the different tumor cell cultures examined containing heterogeneous mixtures of epithelial cells. We performed RT-PCR analysis on the same RNA preparation using primers designed in the non-tandem repeat coding region, and have obtained a similar expression pattern (data not shown).

The whole genomic tandem repeats domain hybridizing with the tandem repeat cDNA probe was defined by a *Pst*I/*Eco*RI band of approximately 11 kb derived from a BAC clone. Digestion with *Dde*I allowed us to establish that the *Pst*I/*Eco*RI fragment consists entirely of 48 or 96 bp. Thus, the tandem repeats in the BAC clone seem to contain 208 uninterrupted individual 48 bp units. We observed potential allelic variation in Southern blots of DNA isolated from 14 pancreatic tumor cell lines, digested with *Eco*RI+*Pst*I, and probed with a pancreatic mucin *MUC4* tandem repeat cDNA probe with no signs of gene amplification. These results suggest that there is a high degree of VNTR polymorphism in the *MUC4* gene among various pancreatic tumor cell lines. Similar, VNTR polymorphism in the *MUC4* gene has been demonstrated for healthy normal individuals (*7*, *42*). In pancreatic tumor cell lines, the 22.7 kb allele was more common than the 19 kb allele that has been reported to be prevalent in the normal Caucasian population (*29*). Allelotyping of the *MUC4* gene in a large number of pancreatic tumor samples may prove to be a useful diagnostic marker.

The existence of alternate forms of mucin molecules is known for membrane-associated mucins such as *MUC1*, *MUC3*, and rat *Muc4*. Rat *Muc4* is reported to exist in both the soluble and membrane-bound forms (*43*). MUC1 is also known to exit as MUC1/SEC, which is generated by the proteolytic cleavage of the membrane-bound form, and MUC1/Y, which lacks the complete tandem repeat domain (*15*). The presence of both the secreted and membrane-associated forms has been documented for another membrane-bound mucin *MUC3*, and two putative secreted forms of *MUC3* have been reported (*17*). However, the presence of such forms of MUC3 at the protein level has not been demonstrated.

In the present study, we have identified five putative variant forms of *MUC4* in HPAF tumor cells. The sv9-MUC4 and sv4-MUC4 variants are the secreted forms and lack the Cys-rich domains, all three EGF-like domains, and the transmembrane (TM) domain. We previously reported a similar variant (sv4-MUC4) in human testis (*18*). The sv9-MUC4 codes for the 174 residue COOH-terminus, a similar variant was reported in our earlier publication (sv2-MUC4) (*18*); however, the splice events responsible for generating these variants are different. The third variant form, sv10-MUC4, is similar to the WT-MUC4, but is shorter in length with a 156 bp truncation in the unique sequence of the *MUC4*β subunit along with intact *MUC4*β. Recently, we found that sv10-MUC4 (a putative membrane-bound variant) is detectable in 37% of pancreatic tumor cell lines but not in any normal human tissues examined including trachea, testis, and lung (*18*, *44*).

The variants *MUC4/Y* and *MUC4/X* lack exon 2 or exon 2-3, respectively. Hence, MUC4/Y and MUC4/X are *N*-glycosylated, heterodimeric non-mucin type growth factor–like molecules. Currently, we are investigating the expression of each of these variant forms among different grades of pancreatic tumors and cell lines and in normal pancreatic tissues. A similar variant, MUC1/Y, is preferentially expressed in breast cancer tissues, as compared to MUC1 isoforms (*15*). MUC1/Y has been suggested to be involved in the oncogenic process as its expression correlates with the growth and proliferation of breast cancer (*15*). It has also been demonstrated that another MUC1 variant form MUC1/SEC acts as a cognate binding protein for MUC1/Y, causing phosphorylation of MUC1/Y and resulting in altered cell morphology (*15*). It is therefore believed that alternative splicing controls the biological effects elicited by the interaction of these two isoforms by regulating the relative levels of the cognate binding protein and receptor. The existence of such variant secreted forms is not known for rat MUC4. However, a soluble form of rat MUC4 has been identified in several rat tissues, resulting from the proteolytic cleavage of the membrane form (*43*). Similar post-translational proteolytic processes might exist in membrane-bound WT-*MUC4* and its variants (sv10-MUC4, MUC4/Y, and MUC4/X).

We do not know the significance of the co-existence of various splice variants in pancreatic adenocarcinoma. We

could not confirm the RT-PCR amplified variant forms of *MUC4* at the level of transcription by Northern blot or RNase protection assay because the RNA is very large (28 kb) and extremely difficult to purify without degradation. Previously, a typical feature of mucins was thought to be the polydisperse signal pattern on Northern blots; but later this was attributed to the degradation of the RNA as a preparation artifact. So far, there is no MUC4-specific monoclonal antibody available with which to study these variants at the protein level.

In conclusion, the present study demonstrates the variation in the amino-acid sequences of pancreatic tumor *versus* normal tracheal MUC4. Genomic hybridization in a panel of pancreatic tumor cell lines with a tandem repeat cDNA probe revealed varying allele sizes for the *MUC4* gene. The level of *MUC4* expression increases with the differentiation grades of tumor cell lines. We have identified variant forms of *MUC4* cDNA, with the conceptual protein translations suggesting the existence of three main mucin types: membrane-associated, secreted, and a membrane-bound form lacking the tandem repeat domain. The presence of various alternate forms of MUC4 in one cell type presents a new paradigm for the structure and the biological roles of these molecules.

The authors would like to acknowledge the technical support of Erik Moore and the Molecular Biology Core Facility, UNMC, for oligonucleotide synthesis and DNA sequencing. JPW was funded by a Wellcome Trust (UK) Advanced Training Fellowship (Grant No. 047002/Z/95). We would also like to acknowledge Takashi Iwamura for generating the SUIT-2 pancreatic tumor cell line.

REFERENCES

1. Gendler, S.J. and Spicer, A.P. (1995) Epithelial mucin genes. *Annu. Rev. Physiol.* **57**, 607–634
2. Kim, Y.S., Gum, J.R., Jr., Crawley, S.C., Deng, G., and Ho, J.J. (1999) Mucin gene and antigen expression in biliopancreatic carcinogenesis. *Ann. Oncol.* **10**, 51–55
3. Rose, M.C. (1988) Epithelial mucous glycoproteins and cystic fibrosis. *Horm. Metab. Res.* **20**, 601–608
4. Lan, M.S., Batra, S.K., Qi, W.N., Metzgar, R.S., and Hollingsworth, M.A. (1990) Cloning and sequencing of a human pancreatic tumor mucin cDNA. *J. Biol. Chem.* **265**, 15294–15299
5. Gum, J.R., Jr., Hicks, J.W., Toribara, N.W., Siddiki, B., and Kim, Y.S. (1994) Molecular cloning of human intestinal mucin (MUC2) cDNA. Identification of the amino terminus and overall sequence similarity to prepro-von Willebrand factor. *J. Biol. Chem.* **269**, 2440–2446
6. Gum, J.R., Hicks, J.W., Swallow, D.M., Lagace, R.L., Byrd, J.C., Lamport, D.T., Siddiki, B., and Kim, Y.S. (1990) Molecular cloning of cDNAs derived from a novel human intestinal mucin gene. *Biochem. Biophys. Res. Commun.* **171**, 407–415
7. 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) Molecular cloning and chromosomal localization of a novel human tracheobronchial mucin cDNA containing tandemly repeated sequences of 48 base pairs. *Biochem. Biophys. Res. Commun.* **175**, 414–422
8. Dufosse, J., Porchet, N., Audie, J.P., Guyonnet Duperat, V., Laine, A., Van-Seuningen, I., Marrakchi, S., Degand, P., and Aubert, J.P. (1993) Degenerate 87-base-pair tandem repeats create hydrophilic/hydrophobic alternating domains in human mucin peptides mapped to 11p15. *Biochem. J.* **293**, 329–337
9. Aubert, J.P., Porchet, N., Crepin, M., Duterque-Coquillaud, M., Vergnes, G., Mazzuca, M., Debuire, B., Petitprez, D., and Degand, P. (1991) Evidence for different human tracheobronchial mucin peptides deduced from nucleotide cDNA sequences. *Am. J. Respir. Cell. Mol. Biol.* **5**, 178–185
10. Toribara, N.W., Roberton, A.M., Ho, S.B., Kuo, W.L., Gum, E., Hicks, J.W., Gum, J.R., Jr., Byrd, J.C., Siddiki, B., and Kim, Y.S. (1993) Human gastric mucin. Identification of a unique species by expression cloning. *J. Biol. Chem.* **268**, 5879–5885
11. Bobek, L.A., Liu, J., Sait, S.N., Shows, T.B., Bobek, Y.A., and Levine, M.J. (1996) Structure and chromosomal localization of the human salivary mucin gene, MUC7. *Genomics* **31**, 277–282
12. Shankar, V., Pichan, P., Eddy, R.L., Jr., Tonk, V., Nowak, N., Sait, S. N., Shows, T. B., Schultz, R.E., Gotway, G., Elkins, R.C., Gilmore, M.S., and Sachdev, G.P. (1997) Chromosomal localization of a human mucin gene (MUC8) and cloning of the cDNA corresponding to the carboxy terminus. *Am. J. Respir. Cell. Mol. Biol.* **16**, 232–241
13. Williams, S.J., McGuckin, M.A., Gotley, D.C., Eyre, H.J., Sutherland, G.R., and Antalis, T.M. (1999) Two novel mucin genes down-regulated in colorectal cancer identified by differential display. *Cancer Res.* **59**, 4083–4089
14. McNeer, R.R., Price-Schiavi, S., Komatsu, M., Fregien, N.R., and Carraway, K.L. (1997) Sialomucin complex in tumors and tissues. *Front. Biosci.* **2**, d449–459
15. Baruch, A., Hartmann, M., Yoeli, M., Adereth, Y., Greenstein, S., Stadler, Y., Skornik, Y., Zaretsky, J., Smorodinsky, N.I., Keydar, I., and Wreschner, D.H. (1999) The breast cancer-associated MUC1 gene generates both a receptor and its cognate binding protein. *Cancer Res.* **59**, 1552–1561
16. Carraway, K.L., 3rd, Rossi, E.A., Komatsu, M., Price-Schiavi, S. A., Huang, D., Guy, P.M., Carvajal, M.E., Fregien, N., Carraway, C.A., and Carraway, K.L. (1999) An intramembrane modulator of the ErbB2 receptor tyrosine kinase that potentiates neuregulin signaling. *J. Biol. Chem.* **274**, 5263–5266
17. Crawley, S.C., Gum, J.R., Jr., Hicks, J.W., Pratt, W.S., Aubert, J.P., Swallow, D.M., and Kim, Y.S. (1999) Genomic organization and structure of the 3′ region of human MUC3: alternative splicing predicts membrane-bound and soluble forms of the mucin. *Biochem. Biophys. Res. Commun.* **263**, 728–736
18. Moniaux, N., Escande, F., Batra, S.K., Porchet, N., Laine, A., and Aubert, J.-P. (2000) Alternate splicing generates a family of both secreted and membrane-associated *MUC4* apomucins. *Eur. J. Biochem.* **267**, 4536–4544
19. Balague, C., Gambus, G., Carrato, C., Porchet, N., Aubert, J.P., Kim, Y.S., and Real, F.X. (1994) Altered expression of MUC2, MUC4, and MUC5 mucin genes in pancreas tissues and cancer cell lines. *Gastroenterology* **106**, 1054–1061
20. Balague, C., Audie, J.P., Porchet, N., and Real, F.X. (1995) *In situ* hybridization shows distinct patterns of mucin gene expression in normal, benign, and malignant pancreas tissues. *Gastroenterology* **109**, 953–964
21. Reid, C.J., Hyde, K., Ho, S.B., and Harris, A. (1997) Cystic fibrosis of the pancreas: involvement of MUC6 mucin in obstruction of pancreatic ducts. *Mol Med* **3**, 403–411
22. Hollingsworth, M.A., Strawhecker, J.M., Caffrey, T.C., and Mack, D.R. (1994) Expression of MUC1, MUC2, MUC3 and MUC4 mucin mRNAs in human pancreatic and intestinal tumor cell lines. *Int. J. Cancer* **57**, 198–203
23. Debailleul, V., Laine, A., Huet, G., Mathon, P., d'Hooghe, M.C., Aubert, J.P., and Porchet, N. (1998) Human mucin genes *MUC2*, *MUC3*, *MUC4*, *MUC5AC*, *MUC5B*, and *MUC6* express stable and extremely large mRNAs and exhibit a variable length polymorphism. An improved method to analyze large mRNAs. *J. Biol. Chem.* **273**, 881–890
24. Batra, S.K., Kern, H.F., Worlock, A.J., Metzgar, R.S., and Hollingsworth, M.A. (1991) Transfection of the human Muc 1 mucin gene into a poorly differentiated human pancreatic tumor cell line, Panc1: integration, expression and ultrastructural changes. *J. Cell Sci.* **100**, 841–849
25. Bhattacharyya, S.N., Enriquez, J.I., Sr., and Manna, B. (1990) Deglycosylation of neutral and acidic human colonic mucin. *Inflammation* **14**, 93–107
26. Batra, S.K., Metzgar, R.S., and Hollingsworth, M.A. (1991) Molecular cloning and sequence analysis of the human ribosomal protein S16. *J. Biol. Chem.* **266**, 6830–6833

27. Batra, S.K., Metzgar, R.S., and Hollingsworth, M.A. (1991) Isolation and characterization of a complementary DNA (PD-1) differentially expressed by human pancreatic ductal cell tumors. *Cell Growth Differ.* **2**, 385–390
28. Moniaux, N., Nollet, S., Porchet, N., Degand, P., Laine, A., and Aubert, J.P. (1999) Complete sequence of the human mucin MUC4: a putative cell membrane-associated mucin. *Biochem. J.* **338**, 325–333
29. Nollet, S., Moniaux, N., Maury, J., Petitprez, D., Degand, P., Laine, A., Porchet, N., and Aubert, J.P. (1998) Human mucin gene MUC4: organization of its 5′-region and polymorphism of its central tandem repeat array. *Biochem. J.* **332**, 739–748
30. Sheng, Z., Wu, K., Carraway, K.L., and Fregien, N. (1992) Molecular cloning of the transmembrane component of the 13762 mammary adenocarcinoma sialomucin complex. A new member of the epidermal growth factor superfamily. *J. Biol. Chem.* **267**, 16341–16346
31. Sherblom, A.P. and Carraway, K.L. (1980) A complex of two cell surface glycoproteins from ascites mammary adenocarcinoma cells. *J. Biol. Chem.* **255**, 12051–12059
32. Audie, J.P., Janin, A., Porchet, N., Copin, M.C., Gosselin, B., and Aubert, J.P. (1993) Expression of human mucin genes in respiratory, digestive, and reproductive tracts ascertained by *in situ* hybridization. *J. Histochem. Cytochem.* **41**, 1479–1485
33. Buisine, M.P., Devisme, L., Copin, M.C., Durand-Reville, M., Gosselin, B., Aubert, J.P., and Porchet, N. (1999) Developmental mucin gene expression in the human respiratory tract. *Am. J. Respir. Cell. Mol. Biol.* **20**, 209–218
34. Gipson, I.K., Spurr-Michaud, S., Moccia, R., Zhan, Q., Toribara, N., Ho, S.B., Gargiulo, A.R., and Hill, J.A., 3rd (1999) MUC4 and MUC5B transcripts are the prevalent mucin messenger ribonucleic acids of the human endocervix. *Biol. Reprod.* **60**, 58–64
35. Reid, C.J., and Harris, A. (1998) Developmental expression of mucin genes in the human gastrointestinal system. *Gut* **42**, 220–226
36. Gum, J.R., Jr., Ho, J.J.L., Pratt, W.S., Hicks, J.W., Hill, A.S., Vinall, L.E., Roberton, A.M., Swallow, D.M., and Kim, Y.S. (1997) MUC3 human intestinal mucin. Analysis of gene structure, the carboxyl terminus, and a novel upstream repetitive region. *J. Biol. Chem.* **272**, 26678–26686
37. Sato, T.N., Qin, Y., Kozak, C.A., and Audus, K.L. (1993) Tie-1 and tie-2 define another class of putative receptor tyrosine kinase genes expressed in early embryonic vascular system [published erratum appears in *Proc. Natl. Acad. Sci. USA* 1993 Dec 15;90(24):12056]. *Proc. Natl. Acad. Sci. USA* **90**, 9355–9358
38. Partanen, J., Armstrong, E., Makela, T.P., Korhonen, J., Sandberg, M., Renkonen, R., Knuutila, S., Huebner, K., and Alitalo, K. (1992) A novel endothelial cell surface receptor tyrosine kinase with extracellular epidermal growth factor homology domains. *Mol Cell Biol.* **12**, 1698–1707
39. Bell, G.I., Fong, N.M., Stempien, M.M., Wormsted, M.A., Caput, D., Ku, L.L., Urdea, M.S., Rall, L.B., and Sanchez-Pescador, R. (1986) Human epidermal growth factor precursor: cDNA sequence, expression *in vitro* and gene organization. *Nucleic Acids Res.* **14**, 8427–8446
40. Derynck, R., Roberts, A.B., Winkler, M.E., Chen, E.Y., and Goeddel, D.V. (1984) Human transforming growth factor-alpha: precursor structure and expression in *E. coli*. *Cell* **38**, 287–297
41. Crawley, S.C., Gum, J.R., Jr., Hicks, J.W., Pratt, W.S., Aubert, J.P., Swallow, D.M., and Kim, Y.S. (1999) Genomic organization and structure of the 3′ region of human MUC3: Alternative splicing predicts membrane-bound and soluble forms of the mucin. *Biochem. Biophys. Res. Commun.* **263**, 728–736
42. Gross, M.S., Guyonnet-Duperat, V., Porchet, N., Bernheim, A., Aubert, J.P., and Nguyen, V.C. (1992) Mucin 4 (MUC4) gene: regional assignment (3q29) and RFLP analysis. *Ann. Genet.* **35**, 21–26
43. Rossi, E.A., McNeer, R.R., Price-Schiavi, S.A., Van den Brande, J.M., Komatsu, M., Thompson, J.F., Carraway, C.A., Fregien, N.L., and Carraway, K.L. (1996) Sialomucin complex, a heterodimeric glycoprotein complex. Expression as a soluble, secretable form in lactating mammary gland and colon. *J. Biol. Chem.* **271**, 33476–33485
44. Choudhury, A., Moniaux, N., Ringel, J., King, J., Moore, E., Aubert, J.-P., and Batra, S.K. (2000) Alternate splicing at the 3′-end of the human pancreatic tumor-associated mucin MUC4 cDNA. *Teratogenesis, Carcinogenesis, and Mutagenesis*, in press