Genomic Organization of the Human Mucin Gene MUC5B

cDNA AND GENOMIC SEQUENCES UPSTREAM OF THE LARGE CENTRAL EXON*  

(Received for publication, April 30, 1998, and in revised form, September 1, 1998)  

Jean-Luc Desseyn‡§, Marie-Pierre Buisine‡¶, Nicole Porchet‡§, Jean-Pierre Aubert‡¶, and Anne Laine‡¶  

From the ‡Unité 377 INSERM, Laboratoire de Recherche Gérard Biserd, Place de Verdun, 59045 Lille Cedex and §Laboratoire de Biochimie et de Biologie Moleculaire de l'Hôtel C. Huriez, CHRU, 59037 Lille Cedex, France  

The complete structure of the DNA encoding the polypeptide chain of human mucin MUC5B has been determined. In this paper, we report the full-length cDNA (3886 bp) and genomic (15,143 bp) sequences upstream of the unusually large central exon of the human mucin gene MUC5B. This region, composed of 29 exons, encodes 1283 amino acid residues. Exon sizes vary from 44 to 262 bp, and intron sizes range from 87 to 1703 bp. We determined the 5’-end of MUC5B by performing rapid amplification of cDNA ends-polymerase chain reaction experiments leading to the same length of the amplified product and by using primer extension experiments. A putative translation start site was found at position 44 to 262 bp, and intron sizes range from 87 to 1703 bp. We determined the 5’-end of MUC5B by performing rapid amplification of cDNA ends-polymerase chain reaction experiments leading to the same length of the amplified product and by using primer extension experiments. A putative translation start site was found at position 44 to 262 bp, and intron sizes range from 87 to 1703 bp. We determined the 5’-end of MUC5B by performing rapid amplification of cDNA ends-polymerase chain reaction experiments leading to the same length of the amplified product and by using primer extension experiments. A putative translation start site was found at position 44 to 262 bp, and intron sizes range from 87 to 1703 bp.}

Mucins are the major components of the mucus, the viscoelastic fluid that protects and lubricates the epithelial surfaces of the respiratory, digestive, and reproductive tracts. Mucins constitute a family of high molecular mass, polydisperse, highly glycosylated proteins. They are synthesized and secreted by specialized epithelial cells. Alterations of the biosynthesis of mucins affecting the protein core and/or the carbohydrate chains linked to the peptide have been observed in numerous pathological situations (1–4). At least eight human mucin genes have been identified (for review, see Gendler and Spicer (5) and Ref. 6). Four of these mucin genes, MUC6, MUC2, MUC5AC, and MUC5B, have been mapped to 11p15.5 on a single band of 400 kilobases (7). Only the genomic structure of the two shortest human mucins, MUC1 (8, 9) and MUC7 (10, 11), have been reported. The central domains of the large apomucins MUC2, RMuc2, MUC5AC, FIM-B.1, and PSM, which contain tandem repeats rich in Ser, Thr, Gly, and Ala, are flanked at both ends by unique cysteine-rich domains (12–21). These domains have significant amino acid sequence similarity to one another and to those in human pro-von Willebrand Factor (pro-vWF) (22).

To date, full-length cDNAs have been published for MUC2 (12–14, 23, 24), FIM-B.1 (20), PSM (21), and SMC (25). Much less information is available regarding the exon-intron structure of these large mucins and only the first three introns of MUC2 (26), and the exon-intron organizations downstream of the central part of MUC5AC, MUC5B, and MUC6 have been determined (27–29).

The MUC5B gene is expressed mainly in bronchus glands and also in submaxillary glands, endocervix, gall bladder, and pancreas (30–34). We have previously subcloned and sequenced the unusually large central exon of 10,713 bp of MUC5B (35) and the 3’ region (28). The limited data known until recently concerning the putative region upstream the tandem repeat array of MUC5AC (18) suggested that the amino-terminal region of MUC5AC could also be similar to that containing several D-domains of pro-vWF as in MUC2, PSM, and FIM-B.1 (12–14, 20, 21). This has been confirmed very recently (19, 36). Hypothesizing that such a similarity also exists in the amino-terminal part of MUC5B, we performed RT-PCR using degenerate primers chosen in sequences that show a high degree of similarity between the known sequences in MUC genes and vWF gene. In this study, we present the genomic organization and the complete sequence (15,143 bp) of the region upstream of the central exon of MUC5B and the cDNA sequence encoding the amino-terminal region (1283 aa) of the apomucin. The comparison of the deduced amino acid sequence shows that MUC5B and MUC5AC have higher similarities.

EXPERIMENTAL PROCEDURES

Synthesis of Oligonucleotide Primers—Oligonucleotide primers used in PCR, RACE-PCR, RT-PCR, and sequencing experiments were synthesized by Eurogentec (Liège, Belgium). Their positions and sequences are indicated in Table I.  

5’-End Amplification of cDNAs—The 5’-AmpliFINDER RACE kit (CLONTECH) was used to synthesize first-strand cDNA from human trachea poly(A)+ RNA (1 μg) obtained from CLONTECH using NAU65 or NAU75 as first primer, followed by ligation of the 5’-anchor adapter.

* This work was supported by “le Comité du Nord de la Ligue Nationale contre le Cancer” and “l’Association de Recherche contre le Cancer.” The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ Supported by a fellowship from the “Ministère de l’Education Supérieure et de la Recherche.” Present address: Dept. of Pharmacology, P.O. Box 357750, University of Washington, Seattle, WA 98195.

¶ To whom correspondence should be addressed: INSERM U-377, 59045 Lille Cedex, France. Tel.: 33-3-20-29-88-59; Fax: 33-3-20-53-85-62; E-mail: laine@lille.inserm.fr.
The PCR was then performed using the nested primer NAU66 or NAU76 respectively and the 5'-anchor primer. Nested PCRs (RACE66 and RACE76 in Fig. 1) involving a second or a third round amplification were carried out with 1 μl of the reaction mixture obtained from each previous round of PCR as template. Another RACE-PCR (RACE367, Fig. 1) experiment was performed using NAU433 and then NAU367 as nested primer. These specific primers were designed from the sequences of progressively amplified products using RT-PCR. The last 5'-RACE-PCR (RACE324, Fig. 1) was performed using the 5'-/3'-RACE-PCR kit (Boehringer Mannheim) and the primer NAU180 and purified using the High Pure PCR Product Purification Kit (Boehringer Mannheim). Terminal transferase was used to add a homopolymeric A-tail to the 3'-end of the cDNA, and tailed cDNA was then amplified by PCR using NAU324 and the oligo(dT)-anchor primer. The PCR products were subcloned into pGEM T-vector (Promega) or PCR2–1 vector (Invitrogen).

RT-PCR Amplification—Total RNA (1 μg) of human gall bladder was extracted as described previously (37). Single-stranded cDNA was performed using the First-strand cDNA Synthesis Kit (CLONTECH) and random hexamers. The 4-fold degenerate primer NAU105 (Table I) was based on seven amino acids, CGLCGNF, found in vWF domain D3 (where the first C residue is Cys-993), MUC2 (where the first C residue

---

### Table I

| Primer designation | Primer sequence (5' to 3') | Position | Orientation |
|--------------------|----------------------------|----------|-------------|
| NAU65              | CTTCCTGATGATGGTGCCGTT       | 3790–3810 | AS          |
| NAU66              | GTCGTCGGCAGATGCGGATCAAGC   | 3767–3788 | AS          |
| NAU75              | CTGCAGCTGACCCGGCAGGTGTATG  | 3522–3540 | AS          |
| NAU91              | CAGTCGGGACCTCGCTGGACGCTG   | 3462–3486 | AS          |
| NAU76              | TGGGCTCTSTGYGGGAGTAYT       | 3482–3498 | AS          |
| NAU101             | TGGGCTCTSTGYGGGAGTAYT       | 529–547   | S           |
| NAU101             | TGGGCTCTSTGYGGGAGTAYT       | 1594–1612 | S           |
| NAU101             | TGGGCTCTSTGYGGGAGTAYT       | 2983–3001 | S           |
| NAU105             | TGGGCTCTSTGYGGGAGTAYT       | 2998–3001 | AS          |
| NAU111             | CACAGAGGCCCAAGTTGTCTG      | 521–541   | S           |
| NAU524             | AGGTCGTCGACACCCGGCCATTTG   | 133–157   | AS          |
| NAU546             | CTCAGGCTGCGGACGAGCTGAGCC   | 93–116    | AS          |
| NAU547             | TGCCAGGGCCCACTAGGAGGCCG    | 24–44     | S           |
| NAU366             | CTGGAGCTGCGCAATG GCCAGGCC | 496–519   | S           |
| NAU367             | AGAAGACGCCGCTCCAAGATTGT    | 989–1012  | AS          |
| NAU433             | GCCAGAGTTGCGTGAAGTCTACAG   | 1030–1053 | AS          |

* The underlined EcoRI site has been added. Y, C or T; R, A or G; S, C or G.

* Strand orientation: sense (S), antisens (AS).
Amino-terminal Region of MUC5B

is Cys-980), and HGM1 (where the first C residue is Cys-28). The 32-fold degenerate primer NAU101 (Table I) was based on seven amino acids, CGLCG(N/D)(F/Y), which occur in pro-vWF domains D1 (where the first C residue is Cys-521), and MUC2 (where the first C residue is Cys-166). Both were utilized as sense primers. For the third RT-PCR, we used the primer NAU366 (see Table I) of which the sequence has been determined from the genomic sequence of the 1.3-kilobase SmaI-SmaI fragment hybridizing with NAU101 (in the left part of Fig. 2A). NAU76, NAU111, and NAU114 were used as antisense primers (Table I) for the PCR. The first-strand cDNA was amplified with the three pairs of primers NAU105/NAU76, NAU101/NAU111, and NAU366/NAU114 (Fig. 1) on a DNA Thermal Cycler 480 (Perkin-Elmer). PCR parameters were 94°C for 2 min, followed by 30 cycles at 94°C for 30 s, 60°C for 1 min and 72°C for 2 min, followed by a final extension at 72°C for 15 min. After electrophoresis, the amplified products were purified (Wizard PCR Prep DNA Purification System, Promega) and subcloned into the pMOSBlue T-vector (Amersham). We confirmed the sequences of all these cDNA fragments using the Titan™One tube RT-PCR kit (Boehringer Mannheim) on human trachea poly(A)+ RNA (CLONTECH) using various couples of specific primers. The double-stranded plasmid inserts were sequenced manually using [γ-32P]dATP (Amersham) and Sequenase 2.0 (U. S. Biochemical Corp.) according to the protocol indicated by the manufacturer. Universal primers or a series of specific oligonucleotides (Table I) were used. The clones were sequenced on both strands several times. Sequence analyses were performed with help of the PC/GENE Software.

**Primer Extension**—Primer extension was performed using two different primers end-labeled with [γ-32P]dATP, NAU346 and NAU324, which correspond to cDNA positions +93→+116 and +133→+157, respectively. Total human tracheobronchial RNA and human poly(A)+ salivary mRNA (CLONTECH) were used. The primers were annealed to RNA, which was then extended by reverse transcription. The product was analyzed on a 6% polyacrylamide gel (Sequagel-6™, National Diagnostics) alongside a sequencing ladder, i.e., the sequencing reaction performed on M13 vector with the −40 primer using [γ-32P]dATP (Amersham) as indicated above.

**Restriction Mapping of the Cosmid BEN1**—Cosmid BEN1 DNA was analyzed by restriction digestion and Southern blot as described previously (28, 35) to identify useful DNA fragments and especially those hybridizing with the primers NAU105 and NAU101.

**Characterization of Genomic DNA, Sequencing, and Sequence Analysis**—We performed DNA sequencing directly on the cosmid BEN1 as described by Desseyn et al. (35) with the primers used for PCR. Sequencing was also performed as described above after subcloning into pBluescript II KS+ vector (Stratagene). The genomic sequence of the 5’-region of MUC5B reported in this paper has been submitted to the EBI data bank with accession number AJ004862.

## RESULTS

**Strategy for Cloning the cDNA of the Amino-terminal Region of MUC5B**—In a previous work we performed a 5’ RACE-PCR experiment and produced the RACE57 clone (985 bp), which allowed us to identify the 5’-end of the central exon and the intron just upstream of this exon (35). Our strategy is summarized in Fig. 1. Two overlapping clones were obtained using the RACE protocol: RACE66 and RACE76 of 600 and 200 bp, respectively. Then, using the RT-PCR with the degenerate oligonucleotides and specific primers, we obtained three overlapping fragments, 105/76 (516 bp), 101/111 (1425 bp), and 366/114 (1156 bp). To elongate the unknown nucleotide sequence encoding the amino-terminal portion of MUC5B, two other RACE-PCR experiments were performed using NAU433 and then NAU367 as nested primer (RACE367) and NAU180 and NAU324 as nested primer (RACE324). RT-PCR using various oligonucleotides were performed to confirm the sequences (91/57, 253/367, 253/286, 366/334, Fig. 1).

**Characterization of the Genomic Fragments of the 5’-End Region of MUC5B**—We first determined the partial restriction map of the genomic clone BEN1. It is shown in Fig. 2A, together with the overlapping fragments, which were subcloned into pBluescript II KS+ vector to facilitate sequence analysis. The fragment, 375-bp HindIII-KnaI (in the right part of Fig. 2A), contains the 5’-end of the central exon (35) indicated with a bold line.

**Transcription Initiation Site of MUC5B**—Repeated RACE-PCR experiments to generate larger products using total RNA from trachea or gall bladder and the primers NAU324 and NAU346 yielded PCR products with identical sequence that were 157 and 116 bp, respectively, ending at the same nucleotide. Primer extension experiments were carried out using two primers, oligonucleotides NAU346 located in exon 2 and NAU324 located in exon 3. The location of transcription initiation was determined by comparison of the migration of the extension product with the sequencing ladder. After electrophoresis, one band for each extended primer was revealed (Fig. 3, for NAU346), supporting the view suggested by the RACE-PCR results that the start site is approximately 95 bp upstream of NAU346. The transcription start site, further designed +1, was designated to be 36 nucleotides upstream of the translation initiation codon ATG which is embedded in a Kozak consensus sequence (38). The 5’-untranslated DNA sequence is CCGTGGGACCGAGCTGGGGGAAATGCGAGGGCCACACCCG(TAG).

**Sequencing Data and Genomic Organization of MUC5B**—The 5’-region of human MUC5B encompasses 15,143 bp from the start site. Comparison between the genomic sequence and the cDNA sequence allowed us to determine the complete genomic organization of the region upstream of the central large exon of MUC5B. The 5’-region of MUC5B, depicted in Fig. 2B, is composed of 29 exons ranging in size from 44 to 262 bp (Table II). The first exon contains an UTR region of 36 bp, the translation start site, and the first three amino acids. As typical of secretory proteins, the sequence of MUC5B starts with a hydrophobic sequence, which is not indicated by the computer as a typical signal sequence. The 5’-region of MUC5B contains 29 introns ranging in size from 87 to 1703 bp. Table II summarizes the length and class of each intron.

The sequences at the exon/intron boundaries are highly conserved with respect to canonical acceptor/donor site (ag/gt) (39) except for the intron 8 that started with a gc instead of gt dinucleotide (Table II). This intron is flanked on either side by the perfect hexanucleotide sequence CAGGCA (boxed in Table II). Intron 9 contains a 455-bp motif found twice and intron 24 contains almost six irregular tandemly repeated motifs of 15 bp ((C/T)GGTTGGGCA(A/G)/A(T)/(G/T)/(C/A/G)). These repeated sequences were not found identical to any registered nucleotide sequences of the EBI/GenBank™.

**Sequencing Data and Amino Acid Analyses**—The predicted translation product deduced from the 3850-nucleotide open
FIG. 4. Comparison of MUC5B amino-terminal protein with pro-vWF, MUC2, RMuc2, FIM-B.1, PSM, and MUC5AC. Dashes indicate gaps introduced in the sequence for alignment purposes. Cysteine residues are shaded. N-potential glycosylation sites are indicated by bold italic letters and shaded. Black down-pointing arrowheads indicate the positions of the introns of human vWF and their class and are named according to Mancuso et al. (22). Gray up-pointing arrowheads indicate the positions of MUC5B introns and are marked with their number and their class according to the Table II. The three introns reported for MUC2 gene (26) are indicated by gray down-pointing arrowheads. Highly conserved amino acids are bold boxed. The highly conserved dibasic sequence is indicated by two asterisks.
reading frame encodes a 1283-amino acid peptide (Fig. 4) rich in cysteine (9.3%) and proline (6.5%). This region is relatively poor in serine and threonine (6.9 and 7.4%, respectively). Ten potential N-glycosylation sites (Asn-X-Ser/Thr; X, any amino acid except Pro) were found (shaded and in italic in Fig. 4) in the amino-terminal sequence of MUC5B.

**Fig. 4—continued**
vWF especially allowed us to dissect this region in six subdomains (Fig. 5). The subdomain, we called 5B, contains 36 amino acid residues. The site for proteolytic cleavage of the signal peptide has not been identified. The predicted NH2-terminal region of MUC5B contains four D-domains, a 348-aa D1-domain, a 368-aa D2-domain, a 101-aa D9-domain, and a 364-aa D3-domain (Figs. 3 and 5). These four D-domains are similar to those of pro-vWF (GenBankTM accession number X04385) (22), MUC2 (GenBankTM accession number L21998) (14), MUC5AC (GenBankTM accession numbers X81649 and AF015521) (18, 19), RMuc2 (GenBank TM accession number Y08296) (20) and PSM (GenBankTM accession number AF005273) (21). We distinguished a short additional domain of 66 aa located just downstream of the D3-domain and upstream of the central part of MUC5B. We called it “MUC11p15-type.” This domain is somewhat different from the A1 domain of vWF.

**DISCUSSION**

The studies reported here present the amino-terminal sequence of MUC5B and the exon-intron organization of this gene, which is now completely known from the transcription start site to the end of the gene. At +37 is an ATG codon. The sequence surrounding this ATG matched well with the Kozak consensus sequence for predicted eukaryotic translation initiation sites (38). The results obtained using repeated RACE-PCR experiments and primer extension carried out on RNA samples from trachea and gall bladder confirmed that we cloned the complete 5′-region cDNA of MUC5B. The 29 introns have an average length of 388 bp, this is in good agreement with previous surveys (40). Out of the 5′-splicing donor sites, there was one site in intron 8) containing gc instead of the more commonly used gt. A gc dinucleotide instead of gt has been detected in a number of genes (41, 42), and in vitro and in vivo studies of the effects of this alteration on mRNA processing have shown that it allowed normal splicing.

Comparison of the vWF gene with MUC5B shows that their intron positions are highly conserved (Fig. 4). The 14 introns, 3, 5–7, 9, 14, 16, 17, 20, 21, 23–25, and 27, of MUC5B show conserved positions and identical phase classes (Fig. 4). Nucleotide sequences of the introns of the two genes do not share any similarity showing that intron positions are much more highly conserved than intron lengths and sequences. The sequence overlapping the first four exons of MUC2 has been recently published (GenBankTM accession number U67167) (26). Our comparison between the genomic organizations of MUC2 and MUC5B shows that their intron positions and classes, downstream of the signal peptide, are highly conserved (Fig. 4), but intron lengths and sequences differ. Moreover, the tandemly repeated sequence of 38 bp (GGGTGGGTAGAGGCCCTCAGGCATGGGC(A/T)CGGGC(A/G)GGT) in intron 2 of MUC2 was not found in the corresponding intron of MUC5B (intron 3). The amino-terminal region of MUC5B shares high levels of sequence similarity with the NH2 termini of pro-vWF, MUC2, RMuc2, MUC5AC, PSM, and FIM-B.1. As shown in Fig. 4, nearly all the cysteine residues are conserved in the seven sequences compared. Out of the 424 highly conserved amino acid residues, the two more conserved are Cys (106 residues, i.e. 25%), which is important in tertiary structure, and Gly (54 residues, i.e. 13%), which may have a role in secondary structure. Interestingly, MUC5AC and MUC5B are more similar to each other than MUC5AC or MUC5B to the other peptides. These observations are consistent with (i) the evolutionary model we proposed before (43) for the three human mucin genes, MUC2, MUC5B, and MUC5AC, illustrating that they may have evolved from a common ancestor, and (ii) the fact that the genes MUC2, and its homologue RMuc2, MUC5AC, and MUC5B define a subclass of mucins characterized by the presence of subdomains called Cys subdomains (35, 43) that interrupt several times their heavily glycosylated regions.

**FIG. 4—continued**

Amino-terminal Region of MUC5B
FIG. 5. Schematic representation of the genomic organization and peptide structure of MUC5B. Boxes indicate subdomains; the seven Cys subdomains (108 aa, 10 Cys) are numbered from 1 to 7. Horizontal lines represent the 5' -UTR (36 bp) and the 3' -UTR (564 bp). Vertical arrow indicates the AATAAA signal. The 47 introns are represented by vertical lines, and some of them are numbered. The two black boxes represent the two different MUC11p15 domains (flanking the central part), found in MUC2 and MUC5AC and each coded by one exon in MUC5B gene.

TABLE II

Characteristics of the exon-intron junctions of the 5' region of MUC5B

Uppercase letters indicate exons, lowercase letters indicate introns.

| Protein domain | EXON N° | Size (bp) | 5' Splice donor | INTRON N° | Size (bp) | Class | 3' Splice acceptor |
|----------------|---------|-----------|-----------------|-----------|-----------|-------|-------------------|
| 5'UTR-5B      | 1       | 46        | ATG GCA         | 1         | 464       | 1     | cag ctc ccc cag   |
| D1+D1-like    | 2       | 72        | TAG ACC          | 2         | 338       | 1     | ccc gct ccc cag   |
| D1-like       | 3       | 262       | GCA GCA         | 3         | 154       | 2     | ccc gct ctc cag   |
| D1-like       | 4       | 115       | CCG TGG          | 4         | 162       | 0     | ccc ccc cct cag   |
| D1-like       | 5       | 91        | CCA CCA         | 5         | 274       | 1     | gtt tt cc cag     |
| D1-like       | 6       | 107       | AGC GAG         | 6         | 786       | 1     | cc cct gct cag    |
| D1-like       | 7       | 202       | TCT GGC         | 7         | 403       | 1     | gct ctt cag       |
| D1-like       | 8       | 126       | CCA GAG         | 8         | 394       | 1     | cag cgc ccc cag   |
| D1-like       | 9       | 118       | CCT GAG         | 9         | 196       | 2     | cag ctc ccc cag   |
| D1+D2-like    | 10      | 139       | CCA GAG         | 10        | 345       | 1     | gtt ctc cag       |
| D2-like       | 11      | 111       | GCA GCA         | 11        | 337       | 0     | cgc ctc cag       |
| D2-like       | 12      | 70        | CGA GAG         | 12        | 419       | 1     | ccc cct cag       |
| D2-like       | 13      | 138       | TGT GCG         | 13        | 438       | 1     | cca ctc cag       |
| D2-like       | 14      | 162       | AGA GAG         | 14        | 290       | 1     | cag ctc cag       |
| D2-like       | 15      | 95        | CAT GAG         | 15        | 98        | 0     | cct ccc cag       |
| D2-like       | 16      | 127       | TCT GGC         | 16        | 242       | 1     | ccc ccc cag       |
| D2-like       | 17      | 256       | CTT GGC         | 17        | 523       | 2     | gtc gtc cag       |
| D3-like       | 18      | 56        | GCA GCA         | 18        | 347       | 1     | cag ctc cag       |
| D3-like       | 19      | 101       | CCT GGC         | 19        | 223       | 0     | cgg ctc cag       |
| D3-like       | 20      | 152       | CAC GAG         | 20        | 393       | 2     | ttc ggc cag       |
| D3+D3-like    | 21      | 139       | CCC CAG         | 21        | 98        | 0     | ttc ctc cag       |
| D3-like       | 22      | 111       | GTT GAG         | 22        | 297       | 0     | ttc ctc cag       |
| D3-like       | 23      | 177       | TCA GAG         | 23        | 357       | 0     | ttc ctc cag       |
| D3-like       | 24      | 240       | TCC GAG         | 24        | 1703      | 0    | tgc ggc cag       |
| D3-like       | 25      | 157       | CCG TGG         | 25        | 412       | 1     | ccc ctc cag       |
| D3-like       | 26      | 129       | TGG AAG         | 26        | 231       | 1     | ccc ctc cag       |
| D3-like       | 27      | 145       | GAC TGG         | 27        | 100       | 2     | tcc ccc cag       |
| D3-like       | 28      | 44        | TGG AAG         | 28        | 87        | 1     | cgc ctc cag       |
| MUC11p15 type | 29      | 198       | CGA AAG         | 29        | 468       | 1     | ccc ctc cag       |
| Central       | 30      | 10713     | CGC CAG         | 30        |           |       | ccc ccc ccc cag   |

Consensus (39):

```
C    A
G g a t
g a t
A    G
```

![Schematic diagram](image-url)
Amino-terminal Region of MUC5B

We thank Karine Rousseau, Michel Crépin, and Pascal Mathon for their excellent technical assistance.

REFERENCES

1. Kaliner, M., Shelhamer, J. H., Borson, B., Nadel, J., Patow, C., and Marow, Z. (1986) Am. Rev. Respir. Dis. 134, 612–621
2. Ho, S. B., Roberton, A. M., Shekels, L. L., Lyftogt, C. T., Niehans, G. A., and Biesbrock, A. R. (1993) J. Biol. Chem. 268, 2440–2446
3. Verma, M. (1994) Cancer Biochem. Biophys. 1019–1029
4. Buisine, M. P., Janin, A., Maunoury, V., Audie, J. P., Delescaut, M. P., Copin, M. C., Colombel, J. F., Degenand, P., Aubert, J. P., and Porchet, N. (1996) Gastroenterology 100, 84–91
5. Gendler, S. J., and Spicer, A. F. (1995) Annu. Rev. Physiol. 57, 607–634
6. Shankar, V., Pichan, P., Eddy, R. L., Tonk, V., Nowak, N., Sait, S. N. J., Shows, T. B., Schultz, R. E., Getway, G., Elkins, R. C., Gilmore, M. S., and Sachdev, S. G. (1997) Am. J. Respir. Cell Mol. Biol. 16, 232–241
7. Pigay, P., Guyonnet Dupérat, V., Hill, A. S., Pratt, W. S., Galiegue-Zouitina, S., Callin D’Hooghe, M., Laine, A., Van Sequenning, I., Gum, J. R., Kim, Y. S., Swallow, D. M., Aubert, J. P., and Porchet, N. (1996) Genomics 38, 340–352
8. Lan, M. S., Batra, S. B., Qi, W. N., Metzgar, R. S., and Hollingsworth, M. A. (1990) J. Biol. Chem. 265, 15294–15299
9. Lancaster, C. A., Peat, N., Duhig, T., Wilson, D., Taylor-Papadimitriou, J., and Gendler, S. J. (1996) Biochem. Biophys. Res. Commun. 173, 1019–1029
10. Bobek, L. A., Tsai, H., Biesbrock, A. R., and Levine, M. J. (1993) J. Biol. Chem. 268, 20363–20369
11. Bobek, L. A., Liu, J., Sait, S. N. J., Shows, T. B., Bobek, Y. A., and Levine, M. J. (1996) Genomics 31, 277–282
12. Gum, J. R., Jr., Byrd, J. C., Hicks, J. W., Toribara, N. W., Lamport, D. T. A., and Kim, Y. S. (1989) J. Biol. Chem. 264, 6480–6487
13. Gum, J. R., Jr., Hicks, J. W., Toribara, N. W., Rothe, E. M., Laine, A., and Kim, Y. S. (1992) J. Biol. Chem. 267, 21375–21383
14. Gum, J. R., Jr., Hicks, J. W., Toribara, N. W., Siddiki, B., and Kim, Y. S. (1994) J. Biol. Chem. 269, 2440–2446
15. Ohmori, H., Dohrman, A. F., Gallup, M., Tsuda, T., Kai, H., Gum, J. R., Jr., Kim, Y. S., and Basbaum, C. B. (1994) J. Biol. Chem. 269, 17833–17840
16. Xu, G., Huan, L.-J., Khatri, I. A., Wang, D., Bennick, A., Fahim, R. E. F., and Basbaum, C. B. (1995) J. Biol. Chem. 270, 13665–13673
17. Lesaffre, T., Roches, F., Hill, A. S., Lacasa, M., Fox, M., Swallow, D. M., Zweibaum, A., and Real, F. X. (1995) J. Biol. Chem. 270, 3401–3407
18. Klomp, L. W. J., Van Rens, L., and Strous, G. J. (1995) Biochim. Biophys. Acta 538, 771–813
19. Li, D., Gallup, M., Fan, N., Szymkowski, D. E., and Basbaum, C. B. (1998) J. Biol. Chem. 273, 6812–6820
20. Jobs, W., and Hoffmann, W. (1997) J. Biol. Chem. 272, 1805–1810
21. Eckhardt, A. E., Timpte, C. S., Desseyn, J.-L., Aubert, J. P., and Porchet, N. (1997) J. Biol. Chem. 272, 33204–33210
22. Mancuso, D. J., Tuley, E. A., Westfield, L. A., Worrall, N. K., Shelton-Inloes, B. B., Toribara, J. S., Aleye, Y. G., and Sadler, J. E. (1999) J. Biol. Chem. 264, 19514–19527
23. Gum, J. R., Hicks, J. W., Swallow, D. M., Lague, R. E., Byrd, J. C., Lamport, D. T. A., Siddiki, B., and Kim, Y. S. (1990) Biochem. Biophys. Res. Commun. 171, 407–415
24. Toribara, N. W., Gum, J. R., Culhane, P. J., Lague, R. E., Hicks, J. W., Petersen, G. M., and Kim, Y. S. (1991) J. Clin. Invest. 88, 1005–1013
25. Wu, K., Fregien, N., and Carraway, K. L. (1994) J. Biol. Chem. 269, 11950–11955
26. Gum, J. R., Hicks, J. W., and Kim, Y. S. (1997) Biochem. J. 325, 259–267
27. Buisine, M. P., Desseyn, J. L., Porchet, N., Degenand, P., Laine, A., and Aubert, J. P. (1998) Biochimie. J. 262, 729–738
28. Desseyn, J. L., Aubert, J. P., Van Vanseunening, I., Porchet, N., and Laine, A. (1997) J. Biol. Chem. 272, 16873–16883
29. Toribara, N. W., Ho, S. B., Gum, E. J., Gum, J. R., Lau, P., and Kim, Y. S. (1999) Proc. Natl. Acad. Sci. U. S. A. 96, 88–92
30. Audie, J. P., Janin, A., Porchet, N., Copin, M. C., Gosselin, B., and Aubert, J. P. (1993) J. Histochem. Cytochem. 41, 1479–1485
31. Audie, J. P., Tetaert, D., Pigay, P., Buisine, M. P., Janin, A., Aubert, J. P., Porchet, N., and Boerema, A. (1990) Hum. Reprod. 10, 94–102
32. Balague, C., Gambus, G., Carrato, C., Porchet, N., Aubert, J. P., Kim, Y. S., and Real, F. X. (1994) Gastroenterology 106, 1056–1061
33. Campion, J. P., Porchet, N., Aubert, J. P., L’Helgoualc’h, A., and Clément, B. (1995) Hepatol. 21, 223–231
34. Balague, C., Audie, J. P., Porchet, N., and Real, F. X. (1995) Gastroenterology 109, 953–964
35. Desseyn, J.-L., Guyonnet-Dupérat, V., Porchet, N., Aubert, J.-P., and Laine, A. (1997) J. Biol. Chem. 272, 3168–3178
36. van de Bovenkamp, J. H., Hau, C. M., Strous, G. J., Butler, H. A., Dekker, J., and Einerhand, A. W. (1998) Biochem. Biophys. Res. Commun. 253, 853–859
37. Glisin, V., Orkenjakover, B., and Byus, C. (1974) Biochemistry 13, 2633–2637
38. Kozak, M. (1991) J. Biol. Chem. 266, 19867–19870
39. Mount, S. M. (1982) Nucleic Acids Res. 10, 459–472
40. Hawkins, J. D. (1988) Nucleic Acids Res. 16, 9893–9905
41. Jackson, I. J. (1991) Nucleic Acids Res. 19, 3795–3798
42. Bao, Y., Dawson, T. L., and Chen, Y. T. (1996) Genomics 38, 155–165
43. Desseyn, J. L., Buisine, M. P., Porchet, N., Aubert, J. P., Degenand, P., and Laine, A. (1998) J. Mol. Evol. 46, 102–106
44. Mayadas, T. N., and Wagner, D. J. (1992) Proc. Natl. Acad. Sci. U. S. A. 89, 3531–3535
45. Perez-Vilar, J., Eckhardt, A. E., DeLuca, A., and Hill, R. L. (1998) J. Biol. Chem. 273, 14442–14449