Molecular cloning of a major human gall bladder mucin: complete C-terminal sequence and genomic organization of MUC5B

Andrew C. KEATES*‡, David P. NUNES*, Nezam H. AFDHAL*, Robert F. TROXLER† and Gwynneth D. OFFNER*§

*Section of Gastroenterology, Department of Medicine and †Department of Biochemistry, Boston University School of Medicine and Boston City Hospital, Boston, MA 02118, U.S.A.

Gall bladder mucin has been shown to play a central role in the pathogenesis of cholesterol gallstone disease. While cloning and sequencing studies have provided a wealth of information on the structure of other gastrointestinal and respiratory mucins, nothing is known about the primary structure of human gall bladder mucin. In this study, we show that the tracheobronchial mucin MUC5B is a major mucin gene product expressed in the gall bladder. Antibodies directed against deglycosylated human gall bladder mucin were used to screen a gall bladder cDNA expression library, and most of the isolated clones contained repetitive sequences nearly identical with those in the tandem repeat region of MUC5B. An additional clone (hGBM2-3) contained an open reading frame coding for a 389 residue cysteine-rich sequence. The arrangement of cysteine residues in this sequence was very similar to that in the C-terminal regions of MUC2, MUC5AC and human von Willebrand factor. This cysteine-rich sequence was connected to a series of degenerate MUC5B tandem repeats in a 7.5 kb *Hin*cII genomic DNA fragment. This fragment, with ten exons and nine introns, contained MUC5B repeats in exon 1 and a 469 residue cysteine-rich sequence in exons 2–10 that provided a 152 nucleotide overlap with cDNA clone hGBM2-3. Interestingly, the exon–intron junctions in the MUC5B genomic fragment occurred at positions equivalent to those in the D4 domain of human von Willebrand factor, suggesting that these proteins evolved from a common evolutionary ancestor through addition or deletion of exons encoding functional domains.

INTRODUCTION

Mucous glycoproteins or mucins are the principal protein component of the mucus gel that lines epithelial surfaces in the gastrointestinal, respiratory and genitourinary tracts [1]. The viscoelastic and lubricative properties of mucins are important in protection of these surfaces against physical and chemical injury, as well as against desiccation and bacterial assault [2–4]. In addition to these protective functions, several lines of evidence have shown that gall bladder mucin plays an integral role in the pathogenesis of cholesterol gallstone disease. First, hypersecretion of gall bladder mucin precedes gallstone formation in cholesterol-fed prairie dogs [5] and inhibition of mucin secretion with aspirin prevents stone formation [6]. Secondly, purified human gall bladder mucin accelerates the nucleation of cholesterol crystals *in vitro* in a time- and concentration-dependent manner [7], and thirdly, the presence of mucin within cholesterol gallstones has been demonstrated by electron-microscopic and biochemical techniques [8,9].

In the past decade, a vast amount of structural information has been obtained on human mucins and on mucins from other vertebrates (reviewed in [2–4,10,11]). At present, at least nine human mucin gene products have been identified and the complete nucleotide sequences of MUC1 [12–15], MUC2 [16–19] and MUC7 [20] have been reported. In addition, partial sequences of MUC3 [21], MUC4 [22], MUC5 (now referred to as MUC5AC; [23–27]), MUC5B [28], MUC6 [29], a novel tracheobronchial mucin (possibly MUC8; [30]) and a sublingual-gland mucin [31] have been described. All of these proteins contain tandem repeating sequences, rich in serine, threonine and proline, which are thought to comprise an extended array in the central region of the polypeptide backbone. In addition, cysteine-rich domains have been identified in the N- and C-terminal regions of MUC2 [18,19] and the C-terminal region of MUC5AC [24,26]. These cysteine-rich regions are believed to participate in the formation of intermolecular disulphide bonds linking mucin monomers to form dimers and higher-order oligomers.

Despite the considerable progress that has been made towards elucidating the structure of other mucins, the only gall bladder mucin that has been characterized to date at both a biochemical and a structural level is bovine gall bladder mucin. Previous work from this laboratory has shown that bovine gall bladder mucin contains two distinct functional domains [32]. One domain is densely glycosylated and resistant to digestion with proteolytic enzymes. The second domain is poorly glycosylated, susceptible to proteolytic cleavage, and has been shown to facilitate binding of bilirubin [33] and biliary lipids [34]. Recently, we have shown that these two domains contain distinct tandem repeating structural units [35]. The glycosylated domain comprises 20 amino acid serine- and threonine-rich tandem repeats, whereas the non-glycosylated domain comprises 127 amino acid cysteine-rich repeats with striking similarity to the scavenger receptor cysteine-rich domains found in a number of receptor and ligand-binding proteins [35].

In the present investigation, we have used both protein sequencing and molecular biological techniques to identify MUC5B as a major human gall bladder mucin. In addition, we describe the complete nucleotide and deduced amino acid sequence of the cysteine-rich C-terminal region of MUC5B and show that it, like the corresponding regions in MUC2 [18,19] and

Abbreviation used: pfu, plaque-forming units.

‡ Present address: Section of Gastroenterology, Department of Medicine, Beth Israel Hospital and Harvard Medical School, Boston, MA 02215, U.S.A.
§ To whom correspondence should be addressed.

The nucleotide sequences presented in this paper are deposited in GenBank with the following accession numbers: U78550, U78551, U78552, U78553 and U78554.

MUC5AC [24,26], comprises domains with striking sequence similarity to the D4, C1 and extreme C-terminal domains of human von Willebrand factor [36–38].

EXPERIMENTAL

Isolation of human gall bladder mucin

Mucosal scrapings from human gall bladders obtained at cholecystectomy were added to four volumes of ice-cold 6 M guanidine hydrochloride/50 mM Tris/HCl (pH 7.5)/5 mM EDTA, and gently dispersed using a Potter–Elvehjem homogenizer. The homogenate was stirred for 72 h at 4 °C to solubilize the mucin, centrifuged for 30 min at 30000 ***g***, and the supernatant subjected to size-exclusion chromatography on Sepharose CL-4B in 4 M guanidine hydrochloride containing 50 mM Tris/HCl (pH 7.5)/5 mM EDTA. Material eluted in the column void volume was concentrated by ultrafiltration using an XM-300 membrane (Amicon, Bedford, MA, U.S.A.). Solid CsCl was added to a density of 1.45 g/ml and the sample was subjected to equilibrium density gradient centrifugation for 75 h at 150000 ***g***. Gradient fractions were dialysed against distilled water, and those containing periodic acid/Schiff reagent-positive mucin were run on a second CsCl density gradient using the conditions described above, except that the concentration of guanidine hydrochloride was reduced to 1 M. Mucin-containing fractions were dialysed against distilled water, freeze-dried and stored at −20 °C.

Deglycosylation of human gall bladder mucin

Freeze-dried mucin (45 mg) was deglycosylated by treatment with anhydrous hydrogen fluoride as described previously [39]. Deglycosylated mucin was dissolved in 8 M urea, dialysed exhaustively against distilled water, freeze-dried and stored at −20 °C. Antibodies were prepared against both native and deglycosylated human gall bladder mucin as described [39].

Isolation and sequencing of gall bladder mucin peptides

Deglycosylated mucin (660 μg) was digested with chymotrypsin (Boehringer Mannheim, Indianapolis, IN, U.S.A.) in 2 M urea/0.1 M Tris/HCl (pH 7.8)/10 mM $CaCl_2$ at an enzyme-to-substrate ratio of 1:100 (w/w) for 24 h at 25 °C. Peptides were fractionated by reversed-phase HPLC using a Vydac C_{18} column (4.6 mm × 150 mm) developed with a 90 min linear gradient from 100 % solvent A [0.1 % (v/v) trifluoroacetic acid in water] to 100 % solvent B [0.1 % (v/v) trifluoroacetic acid in acetonitrile/water (8:1, v/v)]. Column eluate was monitored at 229 nm and 1 ml fractions were collected. Selected peptides were further purified by rechromatography under isocratic conditions at a concentration of solvent B that was 16 % less than that at which the peptide eluted originally. The amino acid sequences of purified peptides were determined on an ABI 470A gas-phase sequencer. Peptide sequences were compared with those in the PIR database of GenBank.

Human gall bladder cDNA library construction and screening

RNA was isolated from normal human gall bladder tissue [40] and affinity purified using the PolyATract system (Promega, Madison, WI, U.S.A.). A random-primed human gall bladder cDNA library in Lambda Zap II (Stratagene, La Jolla, CA, U.S.A.) was prepared according to the manufacturer's protocols, except that random hexamers (Pharmacia, Piscataway, NJ, U.S.A.) were used to prime first-strand cDNA synthesis. Approximately 600000 plaque-forming units (pfu) were plated on *Escherichia coli* SURE at a density of 37000 pfu/150 mm Petri plate. After incubation at 42 °C for 3.5 h, plates were overlaid with nitrocellulose filters soaked in 10 mM isopropyl β-D-thiogalactopyranoside and incubated at 37 °C for a further 3 h. After blocking, filters were incubated with a 1:500 dilution of the anti-deglycosylated human gall bladder mucin antiserum, which had been pretreated with an *E. coli* lysate. Filters were then incubated with a 1:7500 dilution of alkaline phosphatase-conjugated goat anti-rabbit IgG (Promega) and colour was developed with 5-bromo-4-chloro-3-indolyl phosphate/Nitro Blue Tetrazolium. Positive clones were replated and rescreened until plaque purified.

DNA sequencing

Phagemid DNA was isolated from clones that cross-reacted most intensely with the anti-deglycosylated gall bladder mucin antibody and was sequenced with universal primers using the dideoxy method [41] with Sequenase v. 2.0 (Amersham, Chicago, IL, U.S.A.). The complete sequence of one clone (hGBM4-1; see the Results section) was determined from unidirectional deletions prepared using a commercially available exonuclease III system (Erase-a-Base, Promega). Sequences of other cDNA and genomic clones were determined either using nested deletions or using specific oligonucleotide primers. The sequences of exons in genomic clones were confirmed from the sequences of cDNAs obtained by reverse transcriptase PCR from human gall bladder RNA. PCR products were cloned into pCRScript (Stratagene, La Jolla, CA, U.S.A.) and sequenced on an ABI model 373A automated sequencer using universal and specific primers. The nucleotide sequences of all cDNA clones and of coding regions in genomic clones were determined from sequencing both DNA strands. Nucleotide and deduced amino acid sequences were compared with those in GenBank and the PIR database.

Northern, Southern and dot hybridization

RNA from human gall bladder isolated as described above and RNA from human trachea, small intestine and stomach (Clontech) was electrophoresed on 1 % (w/v) agarose denaturing gels and transferred to Hybond N+ membranes (Amersham). Restriction digests of phage lambda DNA were electrophoresed on 0.6–0.8 % agarose gels and blotted on to Hybond N+ membranes. Phagemid DNA (approx. 100 ng) was heat denatured and applied to Hybond N+ membranes for dot hybridization analysis. Northern, Southern and dot blots were hybridized with random-primer-labelled [42] probes at 42 °C in a solution containing 25 mM potassium phosphate, pH 7.4, 5 × SSC (1 × SSC = 0.15 M NaCl/0.015 M sodium citrate), 5 × Denhardt's [1 × Denhardt's = 0.02 % (w/v) Ficoll 40/0.02 % (w/v) polyvinylpyrrolidone/0.02 % (w/v) BSA], 100 μg/ml denatured salmon sperm DNA, 0.1 % (w/v) SDS, 50 % (v/v) formamide and 10 % (w/v) dextran sulphate. Final washes were performed in 0.25 × SSC at 42 °C.

Human genomic library screening

A commercially available human genomic library in Lambda Fix (Stratagene) was plated on *E. coli* LE392 at a density of 37000 pfu/150 mm Petri plate. Plaque filters were hybridized with random-primer-labelled gall bladder mucin cDNA probes (hGBM2-3 and hGBM4-1, see below) under the conditions used for Northern hybridization. Lambda DNA was isolated from a single positive clone and digested with several restriction enzymes. Southern blots of the digests were probed with either hGBM2-3 or hGBM4-1 and hybridizing fragments were subcloned into pBluescript (Stratagene) and sequenced as described above.

RESULTS

Sequencing of human gall bladder mucin chymotryptic peptides

Human gall bladder mucin purified by size-exclusion chromatography and equilibrium density gradient centrifugation was free of contaminating proteins and glycoproteins as judged by SDS/PAGE and staining with either silver or periodic acid/Schiff reagent (results not shown). Digestion of deglycosylated mucin with chymotrypsin followed by fractionation of the digest by reversed-phase HPLC yielded eight peptides for which sequences were determined (Table 1). The chymotryptic peptides contained several features of note. First, serine, threonine and proline make up 55–75% of the total amino acids, which tentatively placed these peptides in the tandem repeat domain of gall bladder mucin. Secondly, the peptide sequences obtained were relatively short, ranging in length from 6 to 13 amino acid residues. In most cases, the C-terminal residue was not an amino acid expected from chymotryptic cleavage and it is possible that some of these peptides may have arisen by cleavage of the polypeptide backbone during treatment with anhydrous hydrogen fluoride. Finally, a search of the PIR database revealed that none of these peptides had sequences that were contained in any known mucin or other protein.

Table 1 Amino acid sequences of peptides isolated from deglycosylated human gall bladder mucin

Peptide	Sequence
1	Thr Ala Thr Ala Thr Pro
2	Lys Val Pro Pro Pro Pro Ser Thr
3	Ser Thr Pro Pro Gly Thr Thr Leu
4	Thr Thr Lys Val Pro Thr Thr Thr
5	Thr Ala Ile Pro Ser Ser
6	Lys Leu Pro Pro Pro Ser Ala Ala Pro Glu Ser Val
7	Lys Leu Pro Gly Pro Pro Ala Thr Pro Ser
8	Ser Leu Pro Pro Leu Ser Ser Leu Ile Ser Gly

Cloning and sequencing of human gall bladder mucin cDNAs

To identify the gall bladder mucin from which the chymotryptic peptides were derived, a human gall bladder cDNA library was screened with a polyclonal antiserum directed against deglycosylated gall bladder mucin. Immunopositive plaques (37) were identified from a screening of approx. 600000 pfu. Initially, seven of the most immunoreactive clones were plaque purified for further characterization. Sequence analysis of the inserts in these clones using universal primers indicated that all contained repetitive sequences that were strikingly similar to the degenerate 87 nucleotide tandem repeats in the tracheobronchial mucin MUC5B [28].

The insert in the largest of the seven strongly immunoreactive cDNA clones, named hGBM4-1, was sequenced completely and found to contain an open reading frame coding for 328 amino acids comprising eleven complete MUC5B-type tandem repeats and one partial repeat (Figure 1). The eleven complete tandem repeats ranged from 72 to 87 bp in length and were nearly identical with the previously described MUC5B tandem repeat

```
                                                GCA AAG CTG ACC ACA ACA GCC ACT ACG ACT GAG TCC ACT GGA TCC ACG GCC ACC CCG  57
                                                Ala Lys Leu Thr Thr Thr Ala Thr Thr Thr Glu Ser Thr Gly Ser Thr Ala Thr Pro  19

TCC TCC ACC CCA GGG ACC ACC TGG ATC CTC ACA GAG CCG AGC ACT ACA GCC ACC GTG ACG GTG CCC ACC GGA TCC ACG GCC ACC GCC 144
Ser Ser Thr Pro Gly Thr Thr Trp Ile Leu Thr Glu Pro Ser Thr Thr Ala Thr Val Thr Val Pro Thr Gly Ser Ser Ala Thr Ala  48

TCC TCC ACC CAG GCA ACT GCT GGC ACC CCA CAT GTG ACG ACC ACG GCC ACG ACA CCC ACA GTC ATC AGC TGG AAA GCC ACT CCC     228
Ser Ser Thr Gln Ala Thr Ala Gly Thr Pro His Val Thr Thr Thr Ala Thr Thr Pro Thr Val Ile Ser Trp Lys Ala Thr Pro      76

TTC TCC AGT CCA GGG ACT GCA ACC GCC CTT CCA GCA CTG AGA AGC ACA GCC ACC ACA CCC ACA GCT ACC AGC TTT ACA GCC ATC CCC 315
Phe Ser Ser Pro Gly Thr Ala Thr Ala Leu Pro Ala Leu Arg Ser Thr Ala Thr Thr Pro Thr Ala Thr Ser Phe Thr Ala Ile Pro 105

TCC TCC TCC CTG GGC ACC ACC TGG ACC CGC CTA TCA CAG ACC ACC ACA CCC ACG GCC ACC ATG TCC ACA GCC ACA CCC             393
Ser Ser Ser Leu Gly Thr Thr Trp Thr Arg Leu Ser Gln Thr Thr Thr Pro Thr Ala Thr Met Ser Thr Ala Thr Pro             131

TCC TCC ACT CCA GAG ACT GCC CAC ACC TCC ACA GTG CTT ACC ACT ACG GCC ACC ACA ACC AGG GCC ACC GGC TCT GTG GCC ACC CCC 480
Ser Ser Thr Pro Glu Thr Ala His Thr Ser Thr Val Leu Thr Thr Thr Ala Thr Thr Thr Arg Ala Thr Gly Ser Val Ala Thr Pro 160

TCC TCC ACC CCA GGA ACA GCT CAC ACT ACC AAA GTG CCG ACT ACC ACA ACC ACG GGC TTC ACA GTC ACC CCC                   552
Ser Ser Thr Pro Gly Thr Ala His Thr Thr Lys Val Pro Thr Thr Thr Thr Thr Gly Phe Thr Val Thr Pro                   184

TCC TCC AGC CCA GGG ACG GCA CGC ACG CCT CCA GTG TGG ATC AGC ACA ACC ACC ACA CCC ACA ACC AGT GGC TCC ACG GTG ACC CCC 639
Ser Ser Ser Pro Gly Thr Ala Arg Thr Pro Pro Val Trp Ile Ser Thr Thr Thr Thr Pro Thr Thr Ser Gly Ser Thr Val Thr Pro 213

TCC TCC ATC CCG GGG ACC ACC CAC ACC CCC ACA GTG CTG ACC ACC ACC ACC ACA ACT GTG GCC ACT GGT TCT ATG GCA ACA CCC     723
Ser Ser Ile Pro Gly Thr Thr His Thr Pro Thr Val Leu Thr Thr Thr Thr Thr Val Ala Thr Gly Ser Met Ala Thr Pro      241

TCC TCT AGC ACA CAG ACC AGT GGT ACT CCC CCA TCA CTG ATC ACC ACG GCC ACT ACG ATC ACG GCC ACC GGC TCC ACC ACC AAC CCC 810
Ser Ser Ser Thr Gln Thr Ser Gly Thr Pro Pro Ser Leu Ile Thr Thr Ala Thr Thr Ile Thr Ala Thr Gly Ser Thr Thr Asn Pro 270

TCC TCA ACT CCA GGG ACA ACA CCT ATC CCC CCA GTG CTG ACC ACC ACC GCC ACC ACA CCT GCA GCC ACC AGC AGC ACA GTG ACT CCC 897
Ser Ser Thr Pro Gly Thr Thr Pro Ile Pro Pro Val Leu Thr Thr Thr Ala Thr Thr Pro Ala Ala Thr Ser Ser Thr Val Thr Pro 299

TCC TCT GCC CTA GGG ACC ACC CAC ACA CCC CCA GTG CCG AAC ACC ACG GCC ACC ACA CAC GGG CGA TCC CTG TCC CCC AGC AGT CCC 984
Ser Ser Ala Leu Gly Thr Thr His Thr Pro Pro Val Pro Asn Thr Thr Ala Thr Thr His Gly Arg Ser Leu Ser Pro Ser Ser Pro 328

MUC5B Repeat Consensus Sequences:
                                        Thr
(1; This work) Ser Ser Thr Pro Gly Thr Ala His Thr Pro Pro Val Leu Thr Thr Thr Ala Thr Thr Thr Ala Thr Gly Ser Thr Ala Thr Pro

(2; Ref 28)    Ser Ser Thr Pro Gly Thr Ala His Thr Leu Thr Val Leu Thr Thr Thr Ala Thr Thr Pro Thr Ala Thr Gly Ser Thr Ala Thr Pro
```

Figure 1 Nucleotide and deduced amino acid sequence of clone hGBM4-1 containing MUC5B-like tandem repeats

The consensus sequence generated from the individual repeats in this clone is given in (1) at the bottom of the Figure and the consensus sequence of the MUC5B tandem repeats reported by Dufosse et al. [28] is given in (2). Identical amino acids are indicated with bold-faced type. Sequences identical with those of peptides 4 and 5 isolated from deglycosylated human gall bladder mucin (Table 1) are doubly underlined. The singly underlined sequence is identical with peptide 1 (Table 1) with one mismatch.

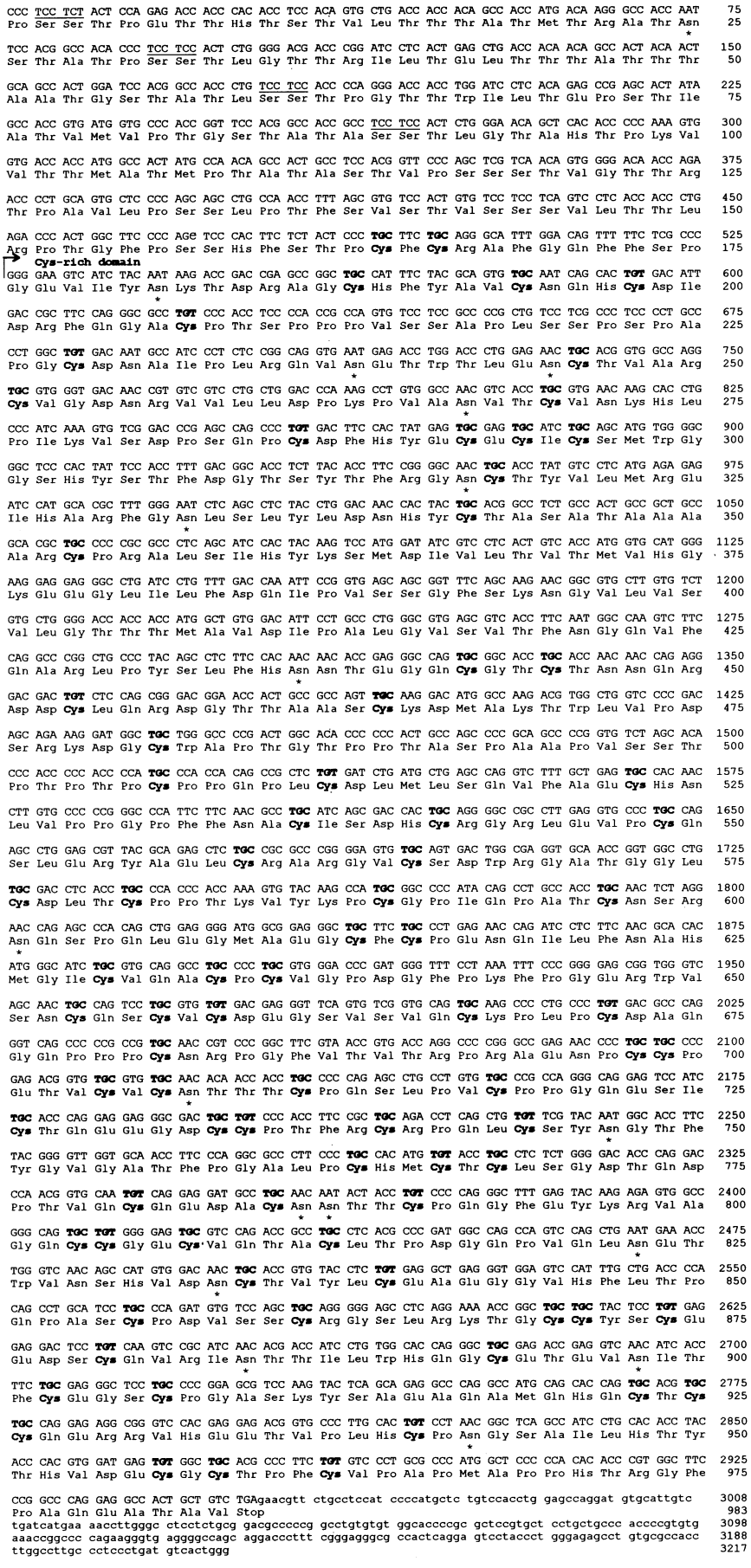

Figure 2 For legend see facing page.

sequences [28]. The latter sequences were found to contain numerous shifts in reading frame resulting from insertions and deletions [28]. However, no shifts in reading frame were evident in any of the MUC5B-type tandem repeats sequenced in this study.

The consensus sequence derived from the tandem repeats in hGBM4-1 is 90% identical with the consensus sequence of the MUC5B repeats (Figure 1). Since the tandem repeats in MUC5B are highly degenerate and the tandem repeat domain is quite large, it seems likely that hGBM4-1 encodes MUC5B repeats that come from a different part of the tandem repeat array from those previously reported. However, it is also possible that some of the sequence differences noted are the result of MUC5B gene polymorphisms in the individual samples from which RNA was obtained.

The sequences of chymotryptic peptides 4 and 5 (Table 1) are contained within hGBM4-1 (double underline in Figure 1). The sequence of chymotryptic peptide 1 is identical with residues 82–87 in hGBM4-1 with a single amino acid substitution (single underline in Figure 1). The similarity of the sequences of peptides 1, 4 and 5 to the remaining peptide sequences (Table 1) suggests that the latter peptides may occur in a different region of the MUC5B tandem repeat array.

Since all seven of the most immunoreactive clones isolated from the human gall bladder cDNA library contained MUC5B repeats, it seemed likely that some of the less immunoreactive clones might contain cDNAs encoding other regions of this mucin. In order to identify recombinants containing such sequences, DNA inserts from the remaining 30 immunopositive clones were screened by dot-blot hybridization using hGBM4-1 as probe. DNA from 19 clones hybridized to this probe, indicating the presence of MUC5B repeats, and these clones were not studied further. The remaining eleven clones were partially sequenced with universal primers as described above. The deduced sequence of one of these clones, hGBM2-3, was enriched with respect to cysteine and this clone was sequenced completely using both exonuclease III-generated nested deletions and specific oligonucleotide primers.

The insert in hGBM2-3 contains 1565 bp, an open reading frame encoding 389 amino acids, followed by a TGA stop codon and 266 bp of 3′-untranslated region (Figure 2: nucleotides 1780–3217). The coding region in this insert contained no tandem repeats, 51 cysteine residues (13.1 mol%) and ten potential N-glycosylation sites (marked with asterisks in Figure 2). Analysis of the nucleotide and deduced amino acid sequences in GenBank revealed that hGBM2-3 is unique, but displays a significant degree of similarity with the C-terminal domains of MUC2 [18], MUC5AC [24,26] and human von Willebrand factor [36–38]. These results, coupled with the large number of positive clones that encoded MUC5B repeats (70%), suggested that the insert in hGBM2-3 was likely to represent a portion of the C-terminal domain of MUC5B.

Northern blot analysis

To test the hypothesis that hGBM2-3 may encode the C-terminal region of MUC5B, cDNA inserts from hGBM4-1 (encoding MUC5B tandem repeats) and hGBM2-3 were used to probe

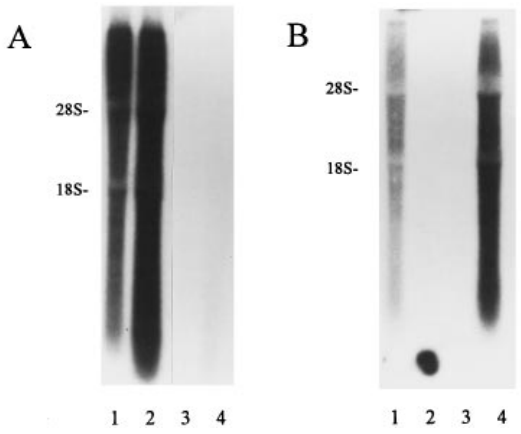

Figure 3 Northern blot analysis of human RNAs probed with the insert in clone hGBM4-1 containing only tandem repeats (A) and the insert in clone hGBM2-3 containing the cysteine-rich C-terminal region (B) of human gall bladder mucin

Hybridization and wash conditions are as described in the text. (**A**) RNA from: lane 1, gall bladder; lane 2, trachea; lane 3, stomach; lane 4, small intestine. (**B**) RNA from: lane 1, gall bladder; lane 2, stomach; lane 3, small intestine; lane 4, trachea. The positions of the 28 S and 18 S ribosomal subunits are marked.

Northern blots to determine the tissue distribution of mRNAs hybridizing to each clone. The tandem repeat probe hybridized only to RNA from gall bladder and trachea (Figure 3A). When the insert in hGBM2-3 (cysteine-rich C-terminal domain) was used to probe a second identical blot, hybridization was again seen only with RNA from gall bladder and trachea (Figure 3B). The tissue distribution of hybridizing transcripts was therefore consistent with the premise that the inserts in hGBM4-1 and hGBM2-3 encode different portions of the same mucin. A polydisperse hybridization pattern from greater than 9 kb to 1 kb was seen with both probes and this pattern is typical of that observed with other mucin mRNAs, although its basis is not known. Rehybridization of the blots with a cDNA probe for glyceraldehyde-3-phosphate dehydrogenase gave discrete bands in each lane (results not shown), but this would not exclude degradation due to shearing of the very large mucin mRNAs.

Analysis of genomic DNA fragments

In order to obtain further sequence information on the region 5′ to that contained within clone hGBM2-3 (cysteine-rich C-terminal domain), differential screening of the gall bladder cDNA library was carried out using the inserts in hGBM4-1 and hGBM2-3. In three separate screenings, no clones were identified that hybridized to both probes. Therefore a human genomic DNA library was screened using the same differential hybridization procedure. One positive clone, designated hGBM G1-4, was identified from screening approx. 375000 pfu. DNA was isolated from this clone and digested with several restriction endonucleases. These analyses showed that the size of the insert in the genomic clone was approx. 18 kb. When duplicate Southern blots of the restriction digests were hybridized with either

Figure 2 Nucleotide and deduced amino acid sequence of the C-terminal region of human gall bladder mucin

The sequence is a composite of the sequence of genomic clone hGBM G1-4 (nucleotides 1–1932) and the sequence of cDNA clone hGBM2-3 (nucleotides 1780–3217). The first two codons of each of four tandem repeats are underlined, the cysteine-rich domain of human gall bladder mucin is marked with an arrow, cysteine codons and cysteine residues are shown in bold-faced type and potential N-glycosylation sites are marked with asterisks.

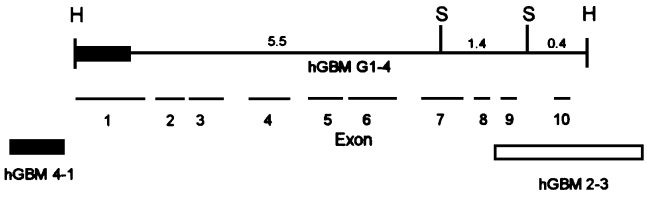

Figure 4 Schematic representation of cDNA clones hGBM4-1 and hGBM2-3 and the *Hin*cII fragment of genomic clone hGBM G1-4 used to determine the nucleotide sequence of the C-terminal region of MUC5B

Clone hGBM4-1 and the portion of the *Hin*cII fragment containing MUC5B tandem repeats are represented as filled boxes. Clone hGBM2-3 is represented as an open box. Exons are numbered provisionally, and the sizes of exons 1–10 are not represented to scale. H, *Hin*cII; S, *Sac*I.

hGBM4-1 or hGBM2-3, a 7.5 kb *Hin*cII fragment (designated hGBM G1-4/7.5) was identified that was recognized by both probes. Digestion of the 7.5 kb *Hin*cII fragment with *Sac*I followed by Southern blot analysis identified three restriction subfragments: a 5.5 kb fragment that hybridized only to the hGBM4-1 repeat probe, a 1.4 kb fragment that hybridized only to the hGBM2-3 cysteine-rich probe and a 0.4 kb fragment that hybridized to neither probe (Figure 4). Each of the *Sac*I restriction fragments was subcloned into pBluescript and sequenced using both exonuclease III-generated subclones and specific oligonucleotide primers.

Sequence analysis of these restriction fragments showed that the 5.5 kb *Hin*cII–*Sac*I fragment contained open reading frames coding for six exons (designated exons 1–6) and a portion of a seventh exon. The 1.4 kb fragment contained open reading frames coding for the remainder of exon 7 and exons 8 and 9. The 0.4 kb fragment contained an open reading frame coding for exon 10 (Figure 4). It should be noted that exons are numbered provisionally, since the entire genomic structure of this mucin is not yet known. The partial sequence of exon 1 contained MUC5B tandem repeats at both its 5′ and 3′ ends. Those at the 3′ end are shown in Figure 2 (nucleotides 1–525). The sequences of these repeats were similar to, but not identical with, any of the degenerate repeats in the insert in hGBM4-1. Exons 2–10 encoded a cysteine-rich non-repeating region. The sequence of this region, shown in Figure 2 (nucleotides 526–1932), provides a 152 bp overlap with the 5′ end of the insert in hGBM2-3 (Figure 2, nucleotides 1780–3217). Thus the sequence of the exons 1–10 in hGBM G1-4/7.5 directly connects MUC5B tandem repeats (exon 1) with the cysteine-rich C-terminal domain in the cDNA clone hGBM2-3 described above.

The complete C-terminal sequence of MUC5B downstream of the tandem repeats (nucleotides 526–2949, Figure 2) codes for 807 amino acids, 81 of which are cysteine residues. Analysis of the deduced amino acid sequence of the C-terminal region of MUC5B in Genbank revealed that the positions of the cysteine residues were nearly identical with those in the C-terminal regions of MUC2 [18], MUC5AC [24,26] and the D4, C1 and C-terminal domain of human von Willebrand factor [36–38]. In addition, the sequence of the extreme C-terminal region of MUC5B (amino acids 738–983) is similar to the corresponding regions of porcine submaxillary mucin [43], bovine submaxillary mucin [44] and frog integumentary mucin B.1 [45], and identical with that of a recently described human salivary mucin [31].

An alignment of the deduced amino acid sequences of the entire C-terminal regions of MUC5B, MUC2 [18,19] and MUC5AC [24,26], and the D4, C1 and C-terminal domains of

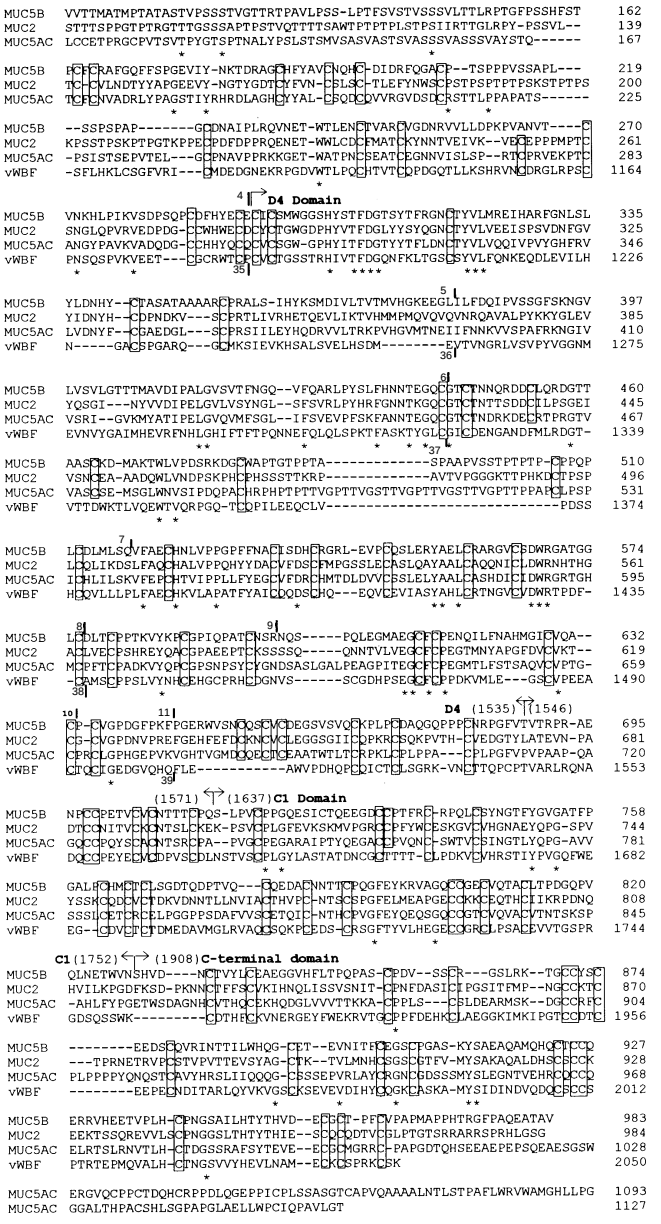

Figure 5 Comparison of the deduced amino acid sequences of the C-terminal regions of MUC5B (this report), MUC2 [18], MUC5AC [24,26] and human von Willebrand factor (vWBF; [36–38]) with gaps introduced to maximize sequence similarity

Cysteine residues identical in at least three of the four sequences are enclosed in boxes and other amino acids that are identical in all four sequences are marked with asterisks. The sequence of human vWBF has been divided into several discontinuous domains indicated in bold type, whereas the other three sequences are presented continuously. The positions of the exon–intron junctions in the D4 domains of MUC5B and vWBF are indicated with a vertical line and exon number.

von Willebrand factor is presented in Figure 5. Overall, this region of MUC5B displays 33.2% identity with both MUC2 and MUC5AC and 26.9% identity with von Willebrand factor. Despite this low overall degree of similarity, all of the cysteine residues in the C-terminal domain of MUC5B are present at the same position in MUC2 and MUC5AC. Of the 71 cysteines in the region of MUC5B, which can be aligned with the D4, C1 and

Table 2 Exon–intron structure of MUC5B and human von Willebrand factor (vWBF) genes

Exons in MUC5B are numbered provisionally according to their position in the sequence of the genomic clone hGBM G1-4/7.5 and exons in vWBF are numbered as in [49]. Exon and intron sizes are given exactly except where shown by the symbol (∼).

Protein	Exon	5′ Splice site	3′ Splice site	Exon size (kb)	Domain	Intron size (kb)	Type
MUC5B	1		TTC TCG CCC G/gtgag	∼ 1.8	–	∼ 0.9	1
MUC5B	2	cacag/GG GAA GTC	CTC CGG CAG/gtggg	0.181	–	0.285	0
MUC5B	3	cccag/GTG AAT GAG	GAG TGC GAG T/gtgag	0.172	–	1.118	1
vWBF	34	tgcag/CCC GGG GAC	ACC TGC CCC T/gtgag	0.178	–	∼ 16.4	1
MUC5B	4	cgcag/GC ATC TGC	GAG GGC CTG/gtgag	0.260	D	0.344	0
vWBF	35	tccag/GC GTG TGC	GAC ATG GAG/gtgag	0.221	D4	1.4	0
MUC5B	5	cccag/ATC CTG TTT	GGC CAG TGC G/gtgag	0.187	D	0.165	1
vWBF	36	tgcag/GTG ACG GTG	GGT CTG TGT G/gtgac	0.193	D4	0.211	1
MUC5B	6	cccag/GC ACC TGC	CTG AGC CA/gtgag	0.226	D	0.114	2
MUC5B	7	cacag/G GTC TTT	GGC CTG TGC G/gtgag	0.176	D	0.210	1
vWBF	37	tctag/GG ATC TGT	GAT TTC TGT G/gtgag	0.342	D4	∼ 2.0	1
MUC5B	8	cacag/AC CTC ACC	AAC TCT AG/gtaag	0.071	D	0.444	2
MUC5B	9	ggcag/G AAC CAG	CAG GCC TGC C/gtaag	0.101	D	0.469	1
MUC5B	10	tccag/CC TGC GTG	CCT AAA TTT/gtgag	0.032	D	0.195	0
vWBF	38	tacag/CT ATG TCA	CAG CAC CAG/gtagg	0.200	D4	∼ 6.5	0

C-terminal domains of von Willebrand factor, the positions of 64 cysteines are conserved in both proteins.

Analysis of MUC5B genomic structure

As described above, exons 2–10 of MUC5B encode 469 amino acids of the C-terminal cysteine-rich region immediately following the MUC5B tandem repeat array. Analysis of the exon/intron boundaries in this region reveals three type 0, five type 1 and two type 2 splice junctions (Table 2). The sequences of the 5′ and 3′ splice junctions in each intron conform to the 'GT-AG' rule and with previously established consensus sequences [46]. Seven of the ten 5′ splice junctions are specified by the sequence GTGAGT (Table 2).

Exons 4–10 encode a region of MUC5B with extensive sequence similarity to the D4 domain of human von Willebrand factor that is encoded in exons 35–39 of the gene for this protein [48]. When the nucleotide sequences of the D4 domains in the two genes are compared, a striking coincidence in both the position of the exon–intron boundaries and the splice junction type is observed. Exon 4 in MUC5B is approximately the same length as exon 35 in von Willebrand factor. Both are followed by a type 0 exon–intron junction (Table 2) and these exons code for the same regions in the two proteins (Figure 5). Similarly, exon 5 in MUC5B and exon 36 in von Willebrand factor are of comparable length, are followed by a type 1 exon–intron junction (Table 2) and occur at identical positions in the two proteins (Figure 5). Exon 37 in von Willebrand factor is split into exons 6 and 7 in MUC5B, whereas exon 38 in von Willebrand factor is split into exons 8, 9 and 10 in MUC5B (Table 2). As shown in Table 2, exon 6 in MUC5B is preceded by a type 1 junction and exon 7 is followed by a type 1 junction, analogous to the type 1 junctions that flank exon 37 in von Willebrand factor. Exon 8 in MUC5B is preceded by a type 1 exon–intron junction and exon 10 is followed by a type 0 junction; exon 38 in von Willebrand factor is preceded by a type 1 junction and followed by a type 0 junction. Gene structure analysis revealed that the similarity between the positions of exon–intron junctions in the D4 domains in MUC5B and von Willebrand factor is greater than that between the exon–intron junctions in the D1, D2, D3 and D4 domains of von Willebrand factor itself.

DISCUSSION

The structure of gall bladder mucin is of considerable interest because of its key role in the pathophysiology of cholesterol gallstone disease. In this paper, we describe the first nucleotide sequences of clones isolated from a human gall bladder cDNA library. These data identified a major mucin in human gall bladder that is likely to be the tracheobronchial mucin MUC5B based on the following observations: (a) transcripts for both the gall bladder mucin and MUC5B have an identical tissue distribution, (b) the consensus sequences of the tandem repeats in the gall bladder mucin and MUC5B are 90% identical at the amino acid level and (c) Southern analysis of human genomic DNA probed with clone hGBM4-1 (Figure 1) revealed exactly the same pattern of hybridizing bands (results not shown) as that seen using a MUC5B probe [28]. While the above strongly suggest that the mucin described in this report is MUC5B, it cannot be ruled out that human gall bladder expresses a closely related gene product.

It has now become clear that a given mucin gene is expressed in more than one human tissue and that frequently tissues express more than one mucin gene [2–4]. In earlier studies, almost all of the known human mucin genes have been shown to be expressed at some level in the human gall bladder epithelium [29,47,49,50]. Since these studies have been conducted in numerous laboratories using different techniques and have examined both normal and inflamed gall bladder tissue, it is difficult to conclude which of these genes encodes the predominant gall bladder mucin.

In the present investigation, several lines of evidence indicate that the mucin identified as MUC5B is a major mucin gene product in the gall bladder. First, cDNA clones encoding MUC5B were isolated from a human gall bladder cDNA expression library using an antiserum raised against deglycosylated human gall bladder mucin. Of the 37 clones that were initially isolated from a screening of 600000 pfu, 26 were shown to contain MUC5B repeats by either direct sequencing or dot

hybridization. The large number of MUC5B clones isolated from the cDNA library is suggestive of a highly expressed gene product. Since the polyclonal antiserum used to screen the library was raised against purified mucin obtained from gall bladder mucosal scrapings, this antibody preparation would be expected to recognize all of the mucins present in human gall bladder epithelium. Secondly, the sequences of two chymotryptic peptides isolated from deglycosylated gall bladder mucin were contained within the deduced amino acid sequence of clone hGBM4-1 encoding MUC5B tandem repeats, and the sequences of the other peptides suggest that they are derived from regions of the degenerate tandem repeat array not yet sequenced. Since none of the chymotryptic-peptide sequences were contained in the tandem repeats of other known mucins, the primary sequence data suggest that the gene for MUC5B is the most highly expressed in human gall bladder epithelium. Thirdly, Northern blot analyses showed that a MUC5B tandem repeat probe hybridized strongly to human gall bladder RNA, consistent with a high level of expression in human gall bladder epithelium.

The complete nucleotide sequence of the cysteine-rich C-terminal region of MUC5B was determined from overlapping cDNA and genomic clones. The deduced amino acid sequence is similar to the cysteine-rich C-terminal regions of MUC2 [18] and MUC5AC [24,26], and the D4, C1 and C-terminal domains of human von Willebrand factor [36–38]. As might be expected, the highest degree of overall sequence similarity (33.2%) was observed between MUC5B and either MUC2 or MUC5AC. The positions of all of the cysteine residues in the C-terminal region of MUC5B were conserved in the other two mucins. Furthermore, the cysteine-containing sequences, GQCGTCTN and EGCFCPE (marked with asterisks in Figure 5), which have been previously shown to occur in both MUC2 and MUC5AC, are also contained in the C-terminal region of MUC5B. It seems likely that both the overall conservation in the position of cysteine residues and the occurrence of the conserved sequences above are indicative of structural features that are required for the disulphide-linked polymerization of mucin monomers. Comparison of the sequences of the individual structural domains in MUC5B, MUC2, MUC5AC and von Willebrand factor shows that the D4 domains in each of the mucins are approx. 38% identical with each other, whereas the D4 domain in any of the mucins is only approx. 26% identical with the D4 domain in von Willebrand factor. The higher degree of sequence similarity among the D4 domains in the three mucins may suggest that this domain has evolved to perform a 'mucin-specific' function distinct from that in von Willebrand factor. In contrast, the similarity between the extreme C-terminal domains of MUC5B, MUC2 and MUC5AC ranges from 23.2 to 31.8%, and the similarity between the extreme C-terminal domain of the three mucins and von Willebrand factor ranges from 24.6 to 31.6%, suggesting that this domain has a common function in all four proteins. In von Willebrand factor, this domain appears to be the only structural requirement for the C-terminal-to-C-terminal dimerization of protein monomers because deletion mutants lacking this region are unable to dimerize and mutants containing only the C-terminal 151 amino acids are fully capable of dimerization [51]. Since the cysteine-rich extreme C-terminal domain has been found in several animal mucins [43–45] and is present in MUC5B (the present investigation), MUC2 [18] and MUC5AC [24,26], this domain is likely to play a critical role in polymerization of monomers during the secretory process. Although the precise location of the cysteine residues involved in the polymerization of von Willebrand factor have not been established, one of the cysteines required for multimerization has been localized to the extreme C-terminal domain of the polypeptide chain by protein sequencing studies [52].

In addition to the sequence similarity noted at the amino acid level between the D4 domains of MUC5B and von Willebrand factor, the gene structure of these regions also appear to be related.

Comparison of the sequences of the two genes revealed a striking coincidence in the positions of exon–intron boundaries and splice junction types (Table 2; Figure 5). For example, the D4 domain begins with exon 4 in MUC5B and exon 35 in von Willebrand factor (Figure 5). These exons are of similar size and are followed by a type 0 splice junction. Exon 5 in MUC5B also corresponds to exon 36 in the D4 domain of von Willebrand factor. However, an additional exon is inserted into the MUC5B gene such that exons 6 and 7 together comprise the region encoded in exon 37 in von Willebrand factor. Similarly, exon 38 in von Willebrand factor codes for a region of the D4 domain encoded by exons 8, 9 and 10 in MUC5B. It remains to be determined whether the genomic structure of MUC5B (containing additional exons not found in the von Willebrand gene) is unique or is a general feature of mucin genes located on chromosome 11p15.

Although the origin of introns is still the subject of debate, accumulating evidence suggests that the vast majority of introns were inserted into existing genes late in eukaryotic evolution [53]. Introns appear to have played a significant role in the evolution of eukaryotic genomes by promoting exon shuffling between genes [54]. In particular, shuffling of exons containing various protein modules has been important in the evolution of cell surface and extracellular proteins [55]. A common feature of many of these modules is the presence of type 1 introns at their 5′ and 3′ ends [56]. Interestingly, in MUC5B, the tandem repeat array (exon 1) and the cysteine-rich domain (beginning with exon 2) are separated by a type 1 intron, suggesting that these two distinct regions of the protein may have been assembled via exon shuffling. This may indicate that the mechanism for the evolution of mucin genes with distinct tandem repeat arrays may have involved the insertion and duplication of exons encoding repeats into a primordial gene for a mucin-like molecule.

The biochemical mechanism by which gall bladder mucin promotes gallstone formation is unknown, but previous studies have shown that the non- or poorly glycosylated regions of the molecule are essential for this process. Human gall bladder mucin contains numerous low-affinity binding sites for hydrophobic ligands, and binding of cholesterol and phosphatidylcholine to these sites could be abolished by proteolysis [7]. In addition, treatment of bovine gall bladder mucin with reducing agents increased the number of available ligand-binding sites, suggesting that cysteine-containing non-glycosylated portions of the molecule are the regions that promote cholesterol crystal nucleation [57]. In the present investigation, we have shown that MUC5B is a major human gall bladder mucin and have identified structural features in the C-terminal region of MUC5B that are fully consistent with earlier experimental observations characterizing the functional domains of gall bladder mucin. These are: (a) the C-terminal domain of MUC5B is not heavily glycosylated and therefore is susceptible to digestion with proteolytic enzymes, (b) the C-terminal domain of MUC5B contains stretches of hydrophobic amino acids that might serve to bind ligands such as cholesterol and other biliary lipids, and (c) the C-terminal domain of MUC5B is enriched with respect to cysteine. Future studies will identify the exact structural features of MUC5B that govern its interaction with biliary lipids leading to stone formation.

Supported by Grants DK44619, DK02327 and DE11691 from the National Institutes of Health.

REFERENCES

1 Neutra, M. R. and Forstner, J. F. (1987) in Physiology of the Gastrointestinal Tract (Johnson, L. R., ed.), 2nd edn., pp. 975–1009, Raven Press, New York
2 Gum, J. R. (1992) Am. J. Respir. Cell Mol. Biol. **7**, 557–564
3 Strous, G. J. and Dekker, J. (1992) Crit. Rev. Biochem. Mol. Biol. **27**, 57–92
4 Gendler, S. J. and Spicer, A. P. (1995) Annu. Rev. Physiol. **57**, 607–634
5 Lee, S. P., LaMont, J. T. and Carey, M. C. (1981) J. Clin. Invest. **67**, 1712–1723
6 Lee, S. P., LaMont, J. T. and Carey, M. C. (1981) Science **211**, 1429–1432
7 Smith, B. F. (1987) J. Lipid Res. **28**, 1088–1097
8 Bills, P. M. and Lewis, D. L. (1975) Gut **16**, 630–637
9 Been, J. M., Bills, P. M. and Lewis, D. L. (1979) Gastroenterology **76**, 548–555
10 Verma, M. and Davidson, E. A. (1994) Glycoconjugate J. **11**, 172–179
11 Kim, Y. S. and Gum, J. R. (1995) Gastroenterology **109**, 999–1001
12 Siddiqui, J., Abe, M., Hayes, D., Shani, E., Yunis, E. and Kufe, D. (1988) Proc. Natl. Acad. Sci. U.S.A. **85**, 2320–2323
13 Gendler, S. J., Lancaster, C. A., Taylor-Papadimitriou, J., Duhig, T., Peat, N., Burchell, J., Pemberton, L., Lalani, E. and Wilson, D. (1990) J. Biol. Chem. **265**, 15286–15293
14 Lan, M. S., Batra, S. K., Qi, W., Metzgar, R. S. and Hollingsworth, M. A. (1990) J. Biol. Chem. **268**, 15294–15299
15 Wreschner, D. H., Hareuveni, M., Tsarfaty, H., Smorodinsky, N., Horev, J., Zaretsky, J., Kotkes, P., Weiss, M., Lathe, R., Dion, A. and Keydar, I. (1990) Eur. J. Biochem. **189**, 463–473
16 Gum, Jr., J. R., Byrd, J. C., Hicks, J. W., Toribara, N. W., Lamport, D. T. A. and Kim, Y. S. (1989) J. Biol. Chem. **264**, 6480–6487
17 Toribara, N. W., Gum, J. R., Culhane, P. J., Lagace, R. E., Hicks, J. W., Petersen, G. M. and Kim, Y. S. (1991) J. Clin. Invest. **88**, 1005–1013
18 Gum, Jr., 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
19 Gum, Jr., J. R., Hicks, J. W., Toribara, N. W., Siddiki, B. and Kim, Y. S. (1994) J. Biol. Chem. **269**, 2440–2446
20 Bobek, L. A., Tsai, H., Biesbrock, A. R. and Levine, M. J. (1993) J. Biol. Chem. **268**, 20563–20569
21 Gum, J. R., Hicks, J. W., Swallow, D. M., Lagace, R. L., Byrd, J. C., Lamport, D. T. A., Siddiki, B. and Kim, Y. S. (1990) Biochem. Biophys. Res. Commun. **171**, 407–415
22 Porchet, N., Van Cong, N., Dufosse, J., Audie, J. P., Guyonnet-Duperat, V., Gross, M. S., Denis, C., Degand, P., Bernheim, A. and Aubert, J. P. (1991) Biochem. Biophys. Res. Commun. **175**, 414–422
23 Aubert, J. P., Porchet, N., Crepin, M., Duterque-Coquillaud, M., Verges, G., Mazzuca, M., Debuire, B., Petitprez, D. and Degand, P. (1991) Am. J. Respir. Mol. Biol. **5**, 178–185
24 Meerzaman, D., Charles, P., Daskal, E., Polymeropoulos, M. H., Martin, B. M. and Rose, M. C. (1994) J. Biol. Chem. **269**, 12932–12939
25 Guyonnet Duperat, V., Audie, J.-P., Debailleul, V., Laine, A., Buisine, M.-P., Galiegue-Zoutina, S., Pigny, P., Degand, P., Aubert, J.-P. and Porchet, N. (1995) Biochem. J. **305**, 211–219
26 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**, 13665–13678
27 Ho, S. B., Roberton, A. M., Shekels, L. L., Lyftogt, C. T., Niehans, G. A. and Toribara, N. W. (1995) Gastroenterology **109**, 735–747
28 Dufosse, J., Porchet, N., Audie, J. P., Guyonnet-Duperat, V., Laine, A., Van-Seuningen, I., Marrakchi, S., Degand, P. and Aubert, J. P. (1993) Biochem. J. **293**, 329–337
29 Toribara, N. W., Roberton, A. M., Ho, S. B., Kuo, W.-L., Gum, E., Hicks, J. W., Gum, J. R., Byrd, J. C., Siddiki, B. and Kim, Y. S. (1993) J. Biol. Chem. **268**, 5879–5885
30 Shankar, V., Gilmore, M. S., Elkins, R. C. and Sachdev, G. P. (1994) Biochem. J. **300**, 295–298
31 Troxler, R. F., Offner, G. D., Zhang, F., Iontcheva, I. and Oppenheim, F. G. (1995) Biochem. Biophys. Res. Commun. **217**, 1112–1119
32 Afdhal, N. H., Offner, G. D., Murrey, F. E., Troxler, R. F. and Smith, B. F. (1990) Gastroenterology **98**, 1631–1641
33 Smith, B. F. and LaMont, J. T. (1983) Gastroenterology **85**, 707–712
34 Smith, B. F. and LaMont, J. T. (1984) J. Biol. Chem. **259**, 12170–12177
35 Nunes, D. P., Keates, A. C., Afdhal, N. H. and Offner, G. D. (1995) Biochem. J. **310**, 41–48
36 Shelton-Inloes, B. B., Titani, K. and Sadler, J. E. (1986) Biochemistry **25**, 3164–3171
37 Titani, K., Kuman, S., Takio, K., Ericsson, L. H., Wade, R. D., Ashida, K., Walsh, K. A., Chopek, M. W., Sadler, J. E. and Fujikawa, K. (1986) Biochemistry **25**, 3171–3184
38 Shelton-Inloes, B. B., Broze, G. J., Miletich, J. P. and Sadler, J. E. (1987) Biochem. Biophys. Res. Commun. **144**, 657–665
39 Afdhal, N. H., Offner, G. D. and Smith, B. F. (1990) Gastroenterology **99**, 1493–1501
40 Chomczynski, P. and Sacchi, N. (1987) Anal. Biochem. **162**, 156–159
41 Sanger, F., Nicklen, S. and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U.S.A. **74**, 5463–5467
42 Feinberg, A. P. and Vogelstein, B. (1984) Anal. Biochem. **137**, 266–267
43 Eckhardt, A. E., Timpte, C. S., Abernethy, J. L., Zhao, Y. and Hill, R. L. (1991) J. Biol. Chem. **266**, 9678–9686
44 Bhargava, A. K., Woitach, J. T., Davidson, E. A. and Bhavanandan, V. P. (1990) Proc. Natl. Acad. Sci. U.S.A. **87**, 6798–6802
45 Hoffman, W. (1988) J. Biol. Chem. **263**, 7686–7690
46 Breathnach, A. and Chambon, P. (1981) Annu. Rev. Biochem. **50**, 349–383
47 Carrato, C., Balague, C., DeBolos, C., Gonzalez, E., Gambus, G., Planas, J., Perini, J. M., Andreu, D. and Real, F. X. (1994) Gastroenterology **107**, 160–172
48 Mancuso, D. J., Tuley, E. A., Westfield, L. A., Worrall, N. K., Shelton-Inloes, B. B., Sorace, J. M., Alery, Y. G. and Sadler, J. E. (1989) J. Biol. Chem. **264**, 19514–19527
49 Baeckstrom, D., Karlsson, N. and Hansson, G. (1994) J. Biol. Chem. **269**, 14430–14437
50 Campion, J.-P., Porchet, N., Aubert, J.-P., L'Helgoualc'h, A. and Clement, B. (1995) Hepatology **21**, 223–231
51 Voorberg, J., Fontijn, R., Calafat, J., Janssen, H., vanMourik, J. A. and Pannekoek, H. (1991) J. Cell Biol. **113**, 195–205
52 Marti, T., Rosselet, S. J., Titani, K. and Walsh, K. A. (1987) Biochemistry **26**, 8099–8109
53 Palmer, J. D. and Langdon, J. M. (1991) Curr. Opin. Genet. Dev. **1**, 470–477
54 Rogers, J. H. (1990) FEBS Lett. **268**, 339–343
55 Baron, M., Norman, D. G. and Campbell, I. D. (1991) Trends Biochem. Sci. **16**, 13–17
56 Patthy, L. (1987) FEBS Lett. **214**, 1–7
57 Smith, B. F. and LaMont, J. T. (1984) J. Biol. Chem. **259**, 12170–12177

Received 9 September 1996/12 December 1996; accepted 6 January 1997