JCI The Journal of Clinical Investigation

The core polypeptide of cystic fibrosis tracheal mucin contains a tandem repeat structure. Evidence for a common mucin in airway and gastrointestinal tissue.

C Gerard, … , R L Eddy Jr, T B Shows

J Clin Invest. 1990;86(6):1921-1927. https://doi.org/10.1172/JCI114925.

Research Article

A cystic fibrosis trachea cDNA library was constructed and probed with a synthetic oligonucleotide containing a consensus sequence recently identified in human intestinal mucin. One of the isolated clones, AMN-22, has been characterized extensively. The cDNA sequence of this 884-bp fragment was determined, and revealed a tandem repeat structure rich in threonine and proline residues. The repeating sequence of AMN-22 was similar but not identical to that determined for gut mucin. When examined by Northern analysis, the mRNA hybridizing to AMN-22 is extremely polydisperse in cystic fibrosis (CF) trachea, with apparent message length varying from approximately 2 kb to greater than 10 kb. A similar pattern was observed, with less abundant message, in CF bronchiectatic lung parenchyma. The lung cDNA hybridized to a similarly polydisperse message in ulcerative colitis colon RNA, but did not hybridize to control RNA from U937 lymphoma cells or stomach RNA. Pedigree analysis of restriction digests of genomic DNA revealed a pattern indicating a single polymorphic locus for the mucin gene expressed in the lung and the intestine. Southern analyses of human:mouse somatic cell hybrid cell lines allow a chromosomal localization for the mucin gene to human chromosome II, within the region 11p13-11pTer. Taken together, these data demonstrate that a polymorphic gene encodes a mucin core polypeptide expressed in both lung and intestine.

The Core Polypeptide of Cystic Fibrosis Tracheal Mucin Contains a Tandem Repeat Structure

Evidence for a Common Mucin in Airway and Gastrointestinal Tissue

Craig Gerard,* Roger L. Eddy, Jr.,‡ and Thomas B. Shows‡

**Ina Sue Perlmutter Cystic Fibrosis Research Laboratory, Childrens' Hospital, the Thorndike Laboratory of Harvard Medical School, and the Department of Pediatrics, Harvard Medical School, Boston, Massachusetts 02115; and ‡Department of Human Genetics, Roswell Park Memorial Institute, Buffalo, New York 14263*

Abstract

A cystic fibrosis trachea cDNA library was constructed and probed with a synthetic oligonucleotide containing a consensus sequence recently identified in human intestinal mucin. One of the isolated clones, AMN-22, has been characterized extensively. The cDNA sequence of this 884-bp fragment was determined, and revealed a tandem repeat structure rich in threonine and proline residues. The repeating sequence of AMN-22 was similar but not identical to that determined for gut mucin.

When examined by Northern analysis, the mRNA hybridizing to AMN-22 is extremely polydisperse in cystic fibrosis (CF) trachea, with apparent message length varying from ~ 2 kb to > 10 kb. A similar pattern was observed, with less abundant message, in CF bronchiectatic lung parenchyma. The lung cDNA hybridized to a similarly polydisperse message in ulcerative colitis colon RNA, but did not hybridize to control RNA from U937 lymphoma cells or stomach RNA.

Pedigree analysis of restriction digests of genomic DNA revealed a pattern indicating a single polymorphic locus for the mucin gene expressed in the lung and the intestine. Southern analyses of human:mouse somatic cell hybrid cell lines allow a chromosomal localization for the mucin gene to human chromosome 11, within the region 11p13-11pTer.

Taken together, these data demonstrate that a polymorphic gene encodes a mucin core polypeptide expressed in both lung and intestine. (*J. Clin. Invest.* 1990. 86:1921–1927.) Key words: molecular cloning • mucins • chromosome 11 • cystic fibrosis • gene mapping

Introduction

Mucus obtained from the lumen of airways is a complex product made up of secretions produced by a number of different epithelial cells, as well as variable amounts of nucleic acids, lipids, and cell debris. The bulk of the secreted product derives from submucosal glands, goblet cells, and clara cells, and the basic composition of mucus is 95% water, 1% protein, 0.9% carbohydrate, and 0.8% lipid (1). The protein component contains variable amounts of protein transudated from serum, serous gland–derived lactoferrin, lysozyme, and locally produced immunoglobulins of various classes. The unique rheological properties of mucus, however, are determined by the presence of heavily glycosylated macromolecules known as mucus glycoproteins or mucins.

Over the past decade, a number of laboratories have initiated characterization of the airway mucin from normal, asthmatic, chronic bronchitic, and cystic fibrosis sputa, as well as mucins from the submaxillary gland, intestine, and reproductive tract (2–7). Much debate arises from a lack of consensus about methods of purification and criteria for purity of mucins. The size of the mucin family is unknown. Generally, there is agreement (reviewed in reference 7) that the core polypeptides of mucins are enriched in the hydroxy amino acids threonine and serine, and that up to 80% of the apparent molecular weight of the mucin is *O*-linked carbohydrate moieties. Typical molecular weights for mucin macromolecules range from 10^5 to 10^6 D. Because these chemical features thwart analysis by classical protein chemical techniques, the size of the mucin monomer is unknown at present as is the number of unique gene products termed "mucins".

Recently, partial cDNAs for mucins from the submaxillary gland and intestine and breast have been cloned using antisera recognizing the deglycosylated core protein (8–10). In each instance, a tandem repeat structure was observed, although each mucin possessed a unique primary structure. The tandem repeat motif may be explained in terms of the function of the polypeptide as a scaffold for the carbohydrate moieties, which impart the biological properties to the macromolecule.

A comparison of the amino analysis of human intestinal mucin (11) with those determined for human tracheal mucins (4–6) reveals a pattern not shared with mucin isolated from submaxillary gland (8, 12). Both lung and intestinal mucins have a predominance of threonine over serine residues, with abundant proline content. Submaxillary mucin, distinctly, is enriched in serine relative to threonine. Although lung and intestinal mucins have always been assumed to be unique molecules, because of the similarities in amino acid composition between lung and intestinal mucins, we constructed and screened a cDNA library from cystic fibrosis trachea using a synthetic oligonucleotide based upon the intestinal mucin tandem repeat. We identified an abundant population of clones hybridizing with the intestinal probe. The analysis of a typical clone from this library is presented, along with evidence which demonstrates that a single polymorphic allele located on chromosome 11 constitutes a mucin gene expressed in both the lung and gastrointestinal tract.

Address correspondence to Craig Gerard, Ph.D., M.D., Enders Research Building, 300 Longwood Ave., Boston, MA 02115.

Received for publication 22 January 1990 and in revised form 5 July 1990.

J. Clin. Invest.

0021-9738/90/12/1921/07 $2.00
Volume 86, December 1990, 1921–1927

Methods

Construction of cystic fibrosis trachea cDNA libraries. Cystic fibrosis (CF)[1] trachea was removed at autopsy within 4 h postmortem, washed in sterile saline, frozen in liquid nitrogen, and stored at −70°C until needed. Samples of ~ 1–2 cm length (3–5 g) were pulverized in liquid nitrogen using a precooled mortar and pestle. RNA was prepared using the guanidine-CsCl gradient method (13) modified for tracheal tissues as follows. The powdered tracheal tissue was suspended in 10–20 ml. 4 M guanidinium thiocyanate containing 0.1 M sodium acetate, pH 5.2, 5 mM EDTA, and 1% 2-mercaptoethanol, and homogenized using a Polytron at full power for 60 s. The homogenate was made 0.5% in sodium lauryl sarcosinate (Sarkosyl, Sigma Chemical Co., St. Louis, MO) and was centrifuged at 2,000 *g* for 10 min. Solid CsCl was added (1 g/2.5 ml homogenate) and dissolved before layering ~ 3.5 ml of the solution over a 1.2-ml cushion of 5.7 M CsCl in 0.1 M sodium acetate, pH 5.2, 5 mM EDTA. The gradient was centrifuged at 35,000 *g* using a model SW 50.1 rotor (Beckman Instruments, Inc., Fullerton, CA) for ~ 18 h at 20°C. The supernatant was aspirated, and the tube was cut at the level of the CsCl cushion. The pellet was washed with 80% ethanol, dissolved in 10 mM Tris buffer containing 10 mM EDTA and 0.5% SDS, pH 7.6, and was extracted with phenol until the material at the interface was negligible. The extracted RNA was precipitated with sodium acetate and ethanol, and redissolved in diethylpyrocarbonate-treated water and stored at −70°C. The yield of RNA was typically 1 mg per 2 cM trachea.

Polyadenylated RNA was purified using oligo-*dT* cellulose (14), and cDNA was synthesized using oligo-*dT* primers with avian myeloblastosis virus reverse transcriptase (Boehringer Mannheim Biochemicals). Second strand synthesis followed the method of Gubler and Hoffman, (15) and double stranded cDNA was blunt ended with T4 polymerase (Boeringer Mannheim Biochemicals, Indianapolis, IN). ~ 2 μg CF tracheal cDNA was suspended in 30 μl ligase buffer, and double stranded oligonucleotide adapters were ligated as described by Aruffo and Seed (16). The adapted cDNA was precipitated with ammonium acetate and ethanol, and residual adapters were removed by passing the cDNA over a 1.2-ml column of Sephacryl S-400 which had been preequilibrated in 10 mM Tris buffer containing 1 mM EDTA and 0.1 M NaCl. The partially size-fractionated cDNA was next sized using potassium acetate gradients (16), and material < 500 bp was discarded. The sized cDNA was ligated into the COS cell expression vector pCDM8, which was the generous gift of Dr. Brian Seed, Harvard Medical School. The plasmid library was transformed into MC1061p3, and the library plated onto nitrocellulose filters. The master filters were replica plated, and the masters frozen at −20°C after soaking in 15% glycerol in LB broth. The library was found to contain 5×10^5 independent recombinants by miniprep analysis.

Library screening. An antisense 30-mer oligonucleotide corresponding to the sequence TTTTVTPTPT (5′-tgt tgg ggt tgg ggt cac cgt ggt ggt ggt-3′) (9) was synthesized and labeled with polynucleotide kinase (New England Biolabs, Burlington, MA) and 32P-ATP. Replica filters of the original library were grown for 12 h at 37°C, and the colonies lysed by placing the filters colony side up for 10 min on a sheet of Whatman paper saturated with 0.5 M NaOH–1.5 M NaCl. The filters were neutralized in a similar fashion using 0.5 M Tris-1.5 M NaCl (pH 7.0), air dried, and then baked 2 h in a vacuum oven at 80°C. Baked replicas were prehybridized for several hours in ~ 50 ml 6× standard sodium citrate (SSC), 0.1% SDS, 0.2% Denhardt's solution, and 50 μg/ml each tRNA and boiled salmon sperm DNA. ~ 10^7 dpm labeled oligonucleotide probe was then added and incubated at 42°C overnight. Filters were washed three times for 10 min each in 500 ml 2× SSC–0.1% SDS at room temperature, then for 1 h at 68°C with 0.5× SSC–0.1% SDS and vigouous agitation. Dried replicates were exposed for 48–72 h with Kodak X-Omat film and an intensifying screen at −70°C.

DNA sequencing. Clones hybridizing with the synthetic oligonucleotide probe were isolated and plasmid purified using CsCl gradients (17). Plasmids were screened by sequencing both strands from polylinker primers using the dideoxynucleotide technique with a double-stranded sequencing protocol (Sequenase Version II) provided by United States Biochemical Corp., Cleveland, OH. Clone AMN-22 was excised from pCDM8 and subcloned in both orientations into pBluescript SK−. Nested deletion clones (18) were prepared using exonuclease III and S1 nuclease (Promega Biotec, Madison, WI). Extensive sequence analysis of these subclones from both orientations established the sequence of AMN-22.

RNA and genomic DNA hybridization. 5–20 μg of total RNA prepared by the guanidine-CsCl method (13) was denatured with formaldehyde and formamide, and electrophoresed through 1.2% agarose gel containing formaldehyde (17). CF trachea RNA was obtained from an autopsy specimen as previously described. CF bronchiectatic lobe RNA was obtained from tissue frozen at −70°C immediately at time of lobectomy. U937 lymphoma cell RNA was prepared from freshly cultured cells. Ulcerative colitis colon RNA and human stomach RNA were obtained from specimens obtained within 30 min of resection, and were obtained as purified RNA as the generous gift of Dr. Chris Stevens, Beth Israel Hospital and Harvard Medical School. Duplicate lanes were stained with ethidium bromide to mark position of 18 and 28 S RNA. RNA was transferred to GeneScreen Plus membranes by blotting with 10× SSC. Genomic DNA was prepared from the peripheral blood of a family with no history of CF, laboratory personel, or patients with documented CF. 7.5-μg samples of genomic DNA were digested to completion with restriction enzyme (Sau 3a, Eco RI, Hind III, Hph1,) in a 50-μl reaction mixture at 37°C for 18 h. Samples were electrophoresed through a 1% agarose gel, and blotted as described above. Transferred nucleic acids were prehybridized at 42°C in 50% formamide, 1% SDS, 1 M NaCl, and 10% dextran sulphate (Pharmacia Fine Chemicals, Piscataway, NJ) for 2–18 h and were then probed with ~ 5×10^5 dpm/ml of clone AMN-22 which had been labeled with 32P-dCTP using the random primer method (Boehringer Mannheim kit). Blots were washed with 2 × 100 ml 2× SSC at room temperature, 2 × 200 ml 2× SSC–1% SDS at 65°C for 30 min, and finally with 2 × 100 ml 0.1× SSC at room temperature. Autoradiographs were prepared using Kodak X-Omat film and an intensifying screen at −70°C for 2–24 h.

Results

Isolation of cystic fibrosis tracheal mucin cDNA clones. High-quality RNA was isolated from CF tracheal tissue, despite the potential autolysis and presence of a chronic exudate in the specimen. When screened with an oligonucleotide designed after a portion of the consensus sequence recently described for human intestinal mucin (9), 47 strongly positive duplicate clones were identified, with many other clones less strongly hybridizing. 10 clones were initially selected and colony purified. These clones contained inserts ranging from 0.75 to 1.5 kb. Each clone was screened by sequencing both strands from either polylinker region. Eight of the 10 clones contained similar or identical sequence at both terminii, and thus appeared to contain a repeating structure. Two clones contained repeat units at their 5′ end and polyadenylated tracts at the 3′-terminus, but the sequences of these clones differed at the 3′ region. They are being analyzed separately at this time. Of the remaining eight clones, one of these, AMN-22, was selected for further characterization. Restriction analysis revealed that the repeating unit contained sites for the restriction enzymes Hph I and Taq IIa. When clone AMN-22 was digested with Hph I, the insert was completely digested, leaving only vector-derived fragments when analyzed by agarose gel electrophoresis. This

1. *Abbreviations used in this paper:* CF, cystic fibrosis; CFTR, CF transmembrane conductance regulator.

result (not shown) indicated that the entire cDNA consisted of repeating units.

Sequence analysis of AMN-22. Because a repeating structure was anticipated from the restriction analysis, the only sequencing strategy available required the production of nested deletions. AMN-22 was subcloned into pBluscript phagemid at the Xba I site in both orientations relative to the T7 primer. Nested deletions were prepared using the Kpn I and Bam HI sites 5′ to the cDNA. In this fashion, the T7 primer was used for all dideoxy sequencing reactions. The sequence of AMN-22 is presented in Fig. 1.

RNA analysis. RNA isolated from CF trachea, CF lung parenchyma (bronchiectatic lobe), lymphoma cell line U937, ulcerative colitis colon, and human stomach was probed with AMN-22 cDNA. As shown in Fig. 2 *A*, a polydisperse message

```
          10                  20                  30                  40
A ACC CCA ACC CCA ACC CCC ACT GGC ACA CAG ACC CCA ACC CCG ACA
  Thr Pro Thr Pro Thr Pro Thr Gly Thr Gln Thr Pro Thr Pro Thr

      50                  60                  70                  80                  90
CCC ATC ACC ACC ACC ACC ACT ATG GTG ACC CCA ACC CCA ACA ATC
Pro Ile Thr Thr Thr Thr Thr MET Val Thr Pro Thr Pro Thr Ile

          100                 110                 120                 130
ACC AGC ACA CAG ACC CCA ACC CCG ACA CCC ATC ACC ACC ACT ACG
Thr Ser Thr Gln Thr Pro Thr Pro Thr Pro Ile Thr Thr Thr Thr

      140                 150                 160                 170                 180
GTG ACC CCA ACC CCA ACC CCC ACC AGC ACA CAG AGA ACA ACA CCG
Val Thr Pro Thr Pro Thr Pro Thr Ser Thr Gln Arg Thr Thr Pro

          190                 200                 210                 220
ACA TCC ATC ACC ACC ACC ACC ACG GTG ACC CCA ACC CCA ACA CCC
Thr Ser Ile Thr Thr Thr Thr Thr Val Thr Pro Thr Pro Thr Pro

      230                 240                 250                 260                 270
ATC ACC ACC ACC ACC ACG GTG ACC CCA ACC CCA ACA CCC ACC GGC
Ile Thr Thr Thr Thr Thr Val Thr Pro Thr Pro Thr Pro Thr Gly

          280                 290                 300                 310
ACA CAG ACC CCA ACA ACG ACA CCC ATC AGC ACC ACC ACC ACG GTG
Thr Gln Thr Pro Thr Thr Thr Pro Ile Ser Thr Thr Thr Thr Val

      320                 330                 340                 350                 360
ACC CCA ACC CCA ACA CCC ACT GGA ACA CAG ACC CTA ACC CCA ACA
Thr Pro Thr Pro Thr Pro Thr Gly Thr Gln Thr Leu Thr Pro Thr

          370                 380                 390                 400
CCC ATC ACC ACC ACC ACT ACG GTG ACC CCA ACC CCT ACA CCC ACC
Pro Ile Thr Thr Thr Thr Thr Val Thr Pro Thr Pro Thr Pro Thr

      410                 420                 430                 440                 450
GGC ACA CAG ACC CCA ACA TCG ACA CCC ATC ACC ACC ACC ACT ACG
Gly Thr Gln Thr Pro Thr Ser Thr Pro Ile Thr Thr Thr Thr Thr

          460                 470                 480                 490
GTG ACC CCA ACA CCA ACA CCC ACC GGG ACA CAG ACC CCA ACC ACG
Val Thr Pro Thr Pro Thr Pro Thr Gly Thr Gln Thr Pro Thr Thr

      500                 510                 520                 530                 540
ACA CCC ATC ACC ACC ACC ACT ACG GTG ACC CCA ACC CCA ACA CCC
Thr Pro Ile Thr Thr Thr Thr Thr Val Thr Pro Thr Pro Thr Pro

          550                 560                 570                 580
ACC GGC ACA AGA TAC CCA ACT CCG ACA CCA ATC ACC ACC ACC ACT
Thr Gly Thr Arg Tyr Pro Thr Pro Thr Pro Ile Thr Thr Thr Thr

      590                 600                 610                 620                 630
ACG GTG ACC CCA ACC CCA ACA CCC ACC AGC ACA CAG AGT ACA ACA
Thr Val Thr Pro Thr Pro Thr Pro Thr Ser Thr Gln Ser Thr Thr

          640                 650                 660                 670
CCA ACA CCC ATC ACC ACC ACC AAT ACG GTA ACC CCA ACC CCA ACC
Pro Thr Pro Ile Thr Thr Thr Asn Thr Val Thr Pro Thr Pro Thr

      680                 690                 700                 710                 720
CCC ACT GGC ACA CAG ACC CCA ACC CCG ACA CCC ATC ACC ACC ACC
Pro Thr Gly Thr Gln Thr Pro Thr Pro Thr Pro Ile Thr Thr Thr

          730                 740                 750                 760
ACC ACT ATG GTG ACC CCA ACC CCA ACA ATC ACC AGC ACA CAG ACC
Thr Thr MET Val Thr Pro Thr Pro Thr Ile Thr Ser Thr Gln Thr

      770                 780                 790                 800                 810
CCA ACC CCG ACA CCC ATC ACC ACC ACT ACG GTG ACC CCA ACC CCA
Pro Thr Pro Thr Pro Ile Thr Thr Thr Thr Val Thr Pro Thr Pro

          820                 830                 840                 850
ACA CCC ACC AGC ACA CAG AGA ACA ACA CCG ACA TCC ATC ACC ACC
Thr Pro Thr Ser Thr Gln Arg Thr Thr Pro Thr Ser Ile Thr Thr

      860                 870                 880
ACC ACC ACG GTG ACC CCA ACC CCA ACA C
Thr Thr Thr Val Thr Pro Thr Pro Thr
```

Figure 1. cDNA sequence of human CF tracheal mucin fragment AMN-22.

length was observed for CF tracheal RNA with the bulk intensity > 2 kb. A fine band at ~ 15 kb is observed, which may represent the primary transcript. When comparing the CF trachea with CF lung parenchyma, it is clear that the message is predominant and abundant in the trachea (compare 4- and 18-h exposures). That the polydispersity does not represent message degradation is shown by the sharp 18- and 28-S RNA bands (inset). Additionally, unrelated cDNA probes display sharp bands with CF trachea and lobe RNA (data not shown, see Discussion). Two additional RNA samples expected to contain a mucin message, as well as a negative control, were analyzed. As shown in Fig. 2 *B*, an extremely strong signal (12-h exposure) was seen in the ulcerative colitis colon RNA, whereas stomach and U937 RNA did not hybridize to AMN-22 cDNA.

Genomic DNA analysis. Genomic DNA prepared from lymphocytes of four healthy volunteers or four patients with CF was digested with HphI, Sau 3a, Hind III, or Eco RI. In each instance, a single band > 12 kb was seen with Eco RI, and digestion of genomic DNA with Hph 1 revealed no fragments hybridizing with AMN-22 cDNA (data not shown). Analysis of five individuals (three CF, two non-CF) whose DNA was digested with Hind III showed a single band at > 12 kb in four cases; one individual showed two fragments. To ascertain the nature of the latter polymorphism, Hind III–restricted genomic DNA from the parents of the control subject was probed. As shown in Fig. 3 *A*, one parent had the single Hind III fragment at > 12 kb, whereas both the smaller fragment at 12 kb, and the common > 12-kb band were present in the other parent. To clarify this observation, the kindred was next probed after digestion of genomic DNA with Sau 3A. As shown in Fig. 3 *B*, a clear pattern was obtained, which indi-

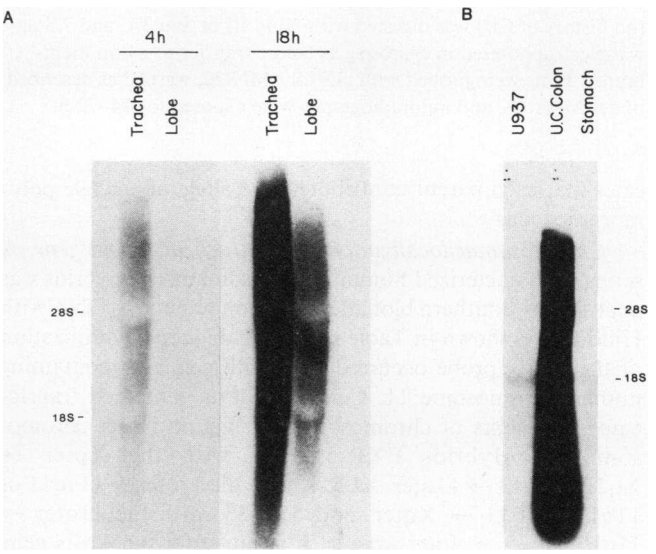

Figure 2. RNA analyses. 10 μg of total RNA from the indicated tissues was subjected to denaturing agarose gel electrophoresis, and blotted to nylon membranes. Blots were probed with [^{32}P]-labeled cDNA AMN-22, and washed as described in Methods. (*A*) Autoradiograph of RNA hybridizing with AMN-22 from CF trachea and lobe exposed for 4 or 8 h. The inset shows the ethidium stained ribosomal RNA of the CF trachea sample. (*B*) Autoradiograph of RNA probed with AMN-22 from U937 lymphoma, ulcerative colitis colon, or stomach resected for carcinoma (tumor free tissue), exposed for 12 h.

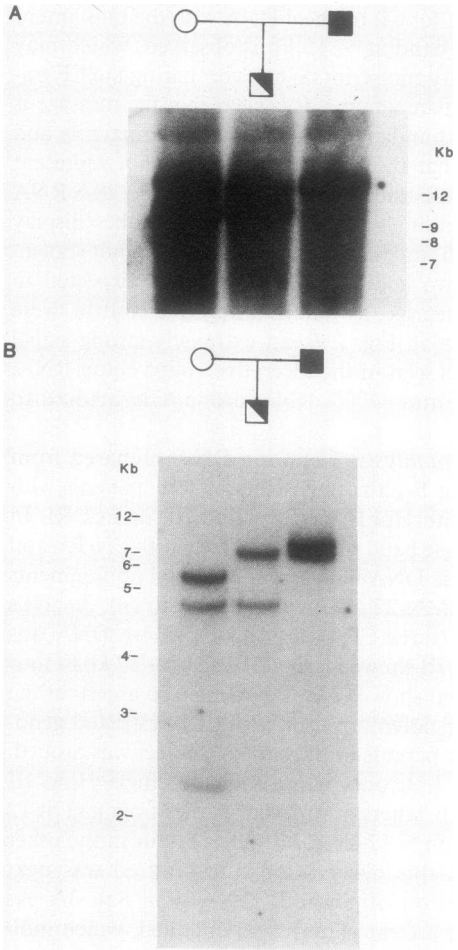

Figure 3. Genomic DNA analyses. Genomic DNA from a kindred (no history of CF) was digested with Hind III or Sau 3A, and 7.5 μg was electrophoresed in agarose gels before transfer to nylon membranes. Blots were probed with cDNA AMN-22, washed as described under Methods, and autoradiographs were exposed for 24–72 h.

cates that each parent contributed one allele of a single polymorphic locus.

Chromosomal localization of the tracheal mucin gene. A series of characterized human:mouse somatic cell hybrids was analyzed by Southern blot after digestion of nuclear DNA with Hind III. As shown in Table I, an unambiguous hybridization of the AMN probe occurred only with cell lines containing human chromosome 11. Certain hybrids contained translocated fragments of chromosome 11, but no intact chromosome 11. Hybrids EXR-5CSAZ, with the Xpter → Xq22::11q13 → 11qter, XER-7, with the 11qter → (11q11 or 11q12)::Xq11 → Xqter, and 55R-33, with the 11qter → 11q13::4q25 → 4qter were each negative for the AMN gene scoring. These hybrids localized the AMN gene to the p13 → pter region of human chromosome 11.

Discussion

The core polypeptide of cystic fibrosis tracheal mucin contains tandem repeats. Using an oligonucleotide based on the repeat structure present in a partial cDNA for human intestinal mucin, a relatively abundant cDNA population was identified in a primary plasmid library prepared from cystic fibrosis trachea. Analysis to date of 10 clones purified from an initial 47 positives revealed that eight of 10 contained only repeating structures, and two clones contained poly-A+ tails. These latter clones contained the repeat structure contiguous with different sequences within the 3′ region. At present, it is not clear whether they represent alternative splicing of the heteronuclear RNA, are the products of different mucin alleles, or represent a cloning artifact. Further analysis of these partial cDNAs is clearly required, and will ultimately require analysis of genomic clones because of the presence of alu family repeat sequences in the 3′ ends of the clones. Of the eight clones containing a repeating structure, AMN-22 was selected because of its relatively small size (884 bp). The sequence of AMN-22 is remarkably similar, but not identical, to the sequences determined for the small intestine mucin (9). As shown in Fig. 4, the repeating unit in AMN-22 is irregular, and varies between 66 and 72 nucleotides, except the segment beginning at nucleotide 226, where an apparent deletion of ~ 27 nucleotides is inferred. In the case of SMUC 40, 41 and 42 partial cDNAs from the small intestine, the tandem repeat unit was exactly 69 nucleotides in the 14 repeats observed in three separate partial clones (9).

In some cases, identity between the intestine and CF trachea clone occurs. For example, the sequence for the tracheal mucin between nucleotides 314 and 436 is identical to that of SMUC 40 between nucleotides 42 and 166. However, the flanking repeats are not identical. As can be seen from comparison of the consensus sequences for small intestine and CF tracheal mucins, a discrepancy occurs in two positions. The inference of these data is that a single polymorphic allele might encode a mucin core polypeptide in the gut and the lung. These data are supported by the RNA and genomic DNA analyses presented below. However, because the tracheal cDNA was derived from a CF patient, the assumption is made that the disease is not related to the core polypeptide of the mucins. The molecular cloning of the CF transmembrane conductance regulator (CFTR, 19) has identified the primary defect in CF. Whether or not changes in the core protein of mucin occur as an epiphenomenon of the disease (for example altered processing, alternative splicing, or regulation) requires analysis of additional clones from non-CF individuals.

RNA analysis. The RNA obtained from autopsy CF trachea was of high quality, which was surprising given the chronic airway bacterial infection and the time elapsing before the tissue was frozen in liquid nitrogen, ~ 4 h. As shown in the inset to Fig. 2, ribosomal RNA bands were sharp. Additionally, we have cloned a ~ 3-kb cDNA for neutral metalloendopeptidase (E.C. 3.4.24.11) from the library used in this study; this cDNA hybridizes to a prominent 5.2-kb message which is undergraded in the CF tracheal RNA (H. Iijima and C. Gerard, unpublished data). Thus the polydisperse message length of CF tracheal mucin does not appear to represent degraded message. The mechanism by which this occurs may be a consequence of alternative splicing involving a repetitive motif, where the donor and acceptor splice junctions themselves may be repetitive.

The mucins purified from lung and gut have long been known to be polydisperse with respect to molecular weight, and this has been attributed to variations in the *O*-linked oligosaccharide length, as well as proteolysis and shearing of the molecule during isolation (7). Interestingly, Gum and col-

Table I. Segregation of AMN Probe with Human Chromosomes in Hind III Digested Human-Mouse Cell Hybrid DNA

DNA No. Hybrid	*AMN*	Human chromosomes																							Translocations
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	X	
845 38L-33	+	−	−	+	−	−	−	+	−	−	−	+	−	−	−	−	−	+	−	−	−	+	−	−	
1246 55R-33	−	+	+	+	+	+	+	+	+	+	−	t	+	+	+	−	+	+	−	−	−	+	+	+	11/4
660 ATR-13	−	+	+	+	+	+	+	+	+	+	+	−	+	+	+	+	+	+	+	+	−	−	+	t	5/X
233 DUA-3BSAGA	−	−	+	−	−	−	−	+	+	−	−	−	−	+	+	−	−	+	−	−	−	−	−	−	
197 DUA-5BSAGA	+	−	−	+	−	+	−	−	−	−	−	+	−	−	+	−	−	+	+	−	−	+	−	−	
859-DUA-6	−	−	+	−	+	+	−	−	−	−	−	−	−	−	−	+	−	−	+	+	−	−	−	+	
186 DUM-13	+	+	+	+	−	+	+	+	−	−	+	+	+	−	+	t	+	+	+	+	+	+	+	t	X/15, 15/X
952 EXR-5CSAZ	−	+	+	+	+	+	+	+	+	−	+	t	+	+	+	+	−	+	+	+	+	+	+	+	X/11
1185 GAR-1	−	−	−	+	−	+	−	−	+	−	+	−	+	+	−	+	+	−	−	−	+	−	−	+	
389 JSR-2	−	−	−	+	+	−	−	+	−	−	−	−	−	+	+	−	−	−	−	−	−	−	−	+	
402 JSR-14	−	−	+	+	+	+	+	−	−	−	−	−	+	+	−	−	−	+	−	−	+	+	−	+	
44 JSR-17S	+	+	+	+	−	+	−	t	+	+	+	+	+	+	+	+	+	+	+	−	+	+	+	−	7/9
830 KER-3	+	−	−	−	−	−	−	−	−	−	−	+	−	+	+	−	+	−	−	−	−	−	−	+	
1146 NSL-9	−	−	−	−	−	+	−	−	+	t	+	−	+	+	+	+	+	+	+	−	+	+	+	−	17/9
192 NSL-16	−	−	−	+	+	+	−	+	+	t	+	−	+	−	+	+	+	+	+	−	+	+	−	−	17/9
42 REW-11	+	−	−	−	+	−	−	+	−	−	−	+	+	+	−	−	+	−	−	−	+	+	+	+	
184 REX-11BSAgB	−	−	−	+	−	−	−	−	−	−	+	−	−	−	+	+	−	−	+	−	−	−	−	−	
254 REX-11BSHF	−	−	−	+	−	−	−	−	−	−	+	−	−	−	+	−	−	−	+	−	−	−	t	t	22/X
394 REX-26	+	+	+	+	+	−	−	+	+	+	+	+	+	−	+	+	+	+	+	+	−	+	t	t	22/X
1162 RSR-3	+	−	−	−	+	−	−	+	−	−	+	+	−	−	+	−	+	+	−	−	−	+	−	+	
390 SIR-11	−	−	−	−	−	−	−	+	−	−	−	−	−	+	−	−	−	−	−	−	−	+	+	+	
643 TSL-1	+	−	+	+	+	−	−	−	−	−	+	+	−	+	−	−	+	+	+	−	+	+	−	−	
644 TSL-2	−	−	+	t	−	+	+	−	+	−	+	−	+	−	+	+	−	t	+	−	+	+	−	+	17/3
395 VTL-6	+	−	+	−	−	−	+	+	+	−	+	+	−	−	−	+	−	+	−	+	+	+	+	−	
407 VTL-17	+	−	−	−	−	+	−	+	−	−	+	+	−	+	+	−	−	+	−	−	+	+	−	−	
212 WIL-2	−	−	−	−	−	−	−	−	+	−	+	−	+	−	−	+	−	+	−	−	−	+	−	+	
425 WIL-6	+	−	+	−	+	+	+	+	+	−	+	+	−	−	+	−	−	+	−	+	+	+	−	+	
424-WIL-8X	+	−	−	+	+	+	−	+	+	−	+	+	+	−	+	−	−	+	+	+	+	+	−	+	
347 WIL-14	−	+	−	+	−	+	−	+	+	−	+	−	+	−	+	+	−	+	−	−	−	−	−	+	
25 WIL-15	+	−	+	+	+	−	+	+	−	−	+	+	+	+	+	+	−	+	+	−	+	+	−	+	
640 XER-7	−	+	+	+	+	+	+	+	+	+	+	t	+	+	+	+	−	−	+	+	−	+	−	+	11/X
534 XOL-6	+	t	−	−	−	+	+	+	−	−	+	+	+	−	+	−	−	+	−	+	+	−	+	t	1/X
555 XOL-13	+	t	−	−	+	+	−	+	−	−	+	+	+	−	−	−	−	+	−	−	+	−	−	+	X/1
1107 XOL-21	+	−	−	+	−	−	−	t	+	+	+	+	+	−	+	−	−	+	+	−	+	−	−	+	ISO7p
Corcordant No. of	(+/+)	3	7	9	8	8	5	12	6	3	13	17	9	6	12	4	7	15	8	6	12	13	5	8	
hybrids	(−/−)	12	9	5	9	6	11	8	6	12	6	14	6	7	5	6	12	7	8	13	11	8	11	4	
Discordant No. of	(+/−)	12	10	8	9	9	12	3	11	14	4	0	8	11	5	12	10	2	9	11	5	4	11	6	
hybrids	(−/+)	5	8	11	8	11	6	9	11	3	11	0	11	10	12	11	5	9	9	4	6	9	5	11	
% Discordancy		53	53	58	50	59	53	37	65	53	44	0	56	62	50	70	44	33	53	44	32	38	50	59	

This table is compiled from 34 cell hybrids involving 17 unrelated human cell lines and four mouse cell lines. The hybrids were characterized by karyotypic analysis and by mapped enzyme markers (23–25). A *t* in the table indicates a chromosome translation with no intact chromosome present (see Translocation; 22–24). The DNA probe was hybridized to Southern blots containing Hind III digested DNA from the human-mouse hybrids. The scoring was determined by the presence (+) or absence (−) of human bands in the hybrids on the blots. The scoring was compared to the presence or absence of human chromosomes in each hybrid. A 0% discordancy indicates a matched segregation of the probe with a chromosome and is the basis for chromosome assignment.

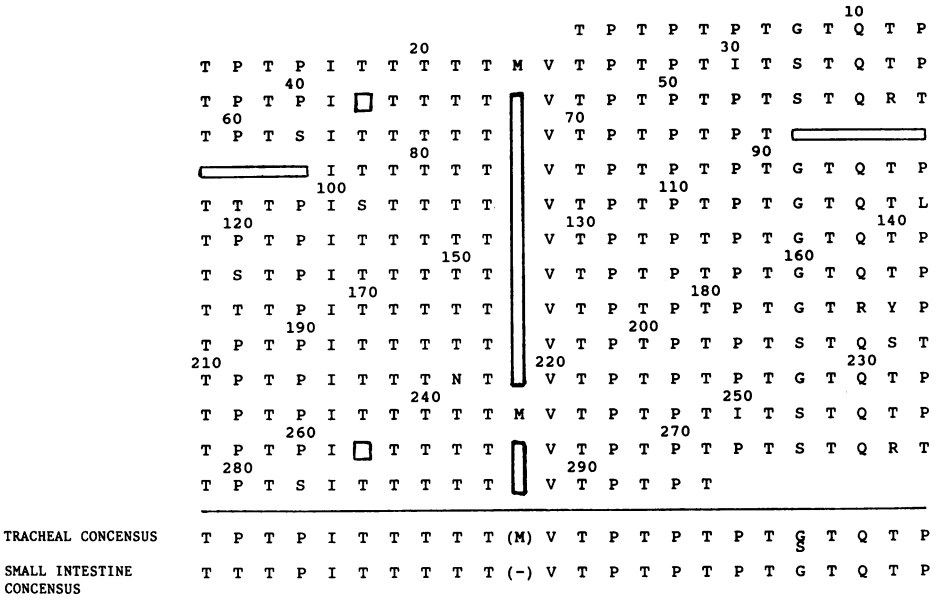

Figure 4. Tandem repeating structure for CF tracheal mucin. Protein sequence deduced for the core protein for the cDNA fragment AMN-22 is aligned to optimize the repeat structure. Deletions have been inserted to preserve sequence identities. The sequence for the tracheal and intestinal mucin fragments (9) are compared.

leagues saw a variable result on RNA analysis with intestinal mucin probes. Four colon cell lines and normal small bowel and colon appear to hybridize to a much less polydisperse message when compared with patient S.J. colon and tumor RNA (9). We were not able to obtain normal human tracheal RNA for the present study, but the possibility exists that pathological mucus secretion such as is seen in CF or ulcerative colitis may be associated with a more polydisperse mucin message.

The tracheal probe hybridized to an abundant message with tracheal RNA but a less abundant signal was seen with an identical quantity of RNA from CF bronchiectatic lobe. This would correlate pathologically with the decreased number of mucin-producing cells in peripheral airways. No staining was seen with RNA from the lymphoma cell line U937. As would be predicted from the high degree of similarity between the AMN-22 and SMUC-41, the CF tracheal clone recognizes an extremely abundant and polydisperse message in ulcerative colitis colon RNA. Surprisingly, stomach RNA did not hybridize with AMN-22. This result does not appear to be due to degradation of the RNA, as 18 and 285 ribosomal RNA bands was undegraded. Additionally, this RNA preparation has been subsequently used for the molecular cloning of the neurokonin A receptor, an extremely rare RNA message in stomach tissue (19a).

Evidence for a single polymorphic gene encoding tracheal and intestinal mucins. Given the similarities between the partial cDNAs for the intestinal and CF tracheal mucin, Southern analyses were performed. We found a limited polymorphism in comparing restriction fragments from several unrelated individuals using Hind III, Sau 3a, and Eco RI. Such patterns are sometimes interpretable in a pedigree analysis. An individual who had a polymorphism revealed with Hind III was studied in this fashion. The data in Fig. 3 *A* suggested that each parent contributed a different single allele. To confirm this, digestion with second enzyme was performed. Gum and colleagues have previously reported the unusual result that the gut clone hybridized to relatively large genomic fragments after Sau 3A digestion (9). (Theoretically, this four-base cutter should cleave once every 256 bases, assuming a random assortment of GATC sequences). When the kindred was probed after Sau3A digestion of genomic DNA, an apparent independent segregation of Sau3A restriction fragment length polymorphisms was seen. These data indicate that a single length–polymorphic allele encodes a mucin core protein expressed in both the lung and the gut. This finding suggests an interesting possibility that the mucin core protein secreted may be a mosaic of the polymorphic allele, or perhaps only a single allele is expressed in tissue. Whether or not there are clinical correlates to particular alleles in disorders associated with mucus secretion is under investigation.

To provide additional definitive data demonstrating a single gene hybridizing with clone AMN-22, we determined the chromosome assignment for the putative unique locus. As shown in Table I, an unambiguous localization was made to chromosome 11p13 → pter. This data is consistent with a proposed localization of 11p15 for the MUC2 gene from small intestine (20). Thus, by sequence analysis of cDNA clones, Northern analysis, Southern analysis, and chromosome localization, we demonstrate that the same polymorphic gene encodes both a tracheal and intestinal mucin core polypeptide. No previous data has suggested that the core proteins for a lung and intestinal mucins are products of the same gene.

Relationship of tracheal mucin to cystic fibrosis. The recent molecular cloning of the CFTR molecule (19) does not shed immediate light on the relationship of this molecule to the mucus inspissation seen in CF. One hypothesis which may relate the primary defect to the mucins is as follows. Goblet cells secrete mucins into exocrine ducts and lumena of all tissue affected in CF, including the gall bladder, testes, pancreas, lung, and bowel. After biosynthesis, the mucins are packed into secretory granules in a condensed form in the goblet cell. As Donnan polyelectrolytes (21), the mucins swell in hydrated environments as a function of pH and ionic strength. Conversely, the mucins condensed in secretory vesicles would require a high concentration of cation to neutralize their anionic properties, and exocytosis itself may depend on swelling of the granule by ion currents which disturb the Don-

nan equilibrium of the vesicle (21). The exocytic granule may thus require the proper functioning of CFTR, perhaps the granule membrane itself contains CFTR. If the Donnan equilibrium is not maintained within the CF exocytic granule a tonic secretion of granules might result. The altered salt and water content of the extracellular lumena might then contribute to inspissation of oversecreted mucin, leading to the obstruction classically associated with CF pathology. A similar epiphenomenon has been proposed for the increased sulfation in CF epithelial glycoconjugates, where abnormal sulfate ion transport resulting from linkage to the chloride flux allows for increased availability of sulfate for mucin modification (22). These epiphenomenon might together create the niche which favors colonization of the airways with bacterial organisms.

Acknowledgments

This study could not have been carried out without the valuable advice and support of a number of individuals, including Drs. Norma Gerard, Lynne Reid, K. R. Bhaskar, Hideya Iijima, Chris Stevens, and Jeffrey Drazen. The research assistance of Christine Squassoni is greatly appreciated.

This research was supported by National Institutes of Health grant HL 41277. Dr. Gerard is an RJR Nabisco Faculty Scholar at the Harvard Medical School.

References

1. Hayez, R., and P. Roussel. 1976. Bronchial mucus: physical and biochemical features. *In* Bronchial Asthama: Mechanisms and Therapeutics. E. B. Weiss and M. S. Segal, editors. Little, Brown and Co., Boston. 409–422.

2. Creeth, J. M., K. R. Bhaskar, J. R. Horton, I. Das, M. T. Lopez-Vidriero, and L. M. Reid. 1977. The separation and characterization of bronchial glycoproteins by density-gradient methods. *Biochem. J.* 167:557–569.

3. Feldhoff, P. A., V. P. Bhavandan, and E. A. Davidson. 1979. Purification, properties and analysis of human asthmatic bronchial mucin. *Biochemistry.* 18:2430–2436.

4. Woodward, H., B. Horsey, V. P. Bhavandan, and E. A. Davidson. 1982. Isolation, purification and properties of respiratory mucus glycoproteins. *Biochemistry.* 21:694–700.

5. Chase, K. V., M. Flux, and G. P. Sachdev. 1985. Comparison of physiochemical properties of purified mucus glycoproteins isolated from respiratory secretions of cystic fibrosis and asthmatic patients. *Biochemistry.* 24:7334–7341.

6. Rose, M. C., W. S. Lynn, and B. Kauffman. 1979. Resolution of the major components of human lung mucosal gel and their capabilities for reaggregation and gel formation. *Biochemistry.* 18:4030–4037.

7. Carlstedt, I. P., J. K. Sheehan, A. P. Corfield, and J. T. Gallagher. 1985. Mucus glycoproteins: a gel of a problem. *Essays Biochem.* 20:40–68.

8. Eckhardt, A. E., C. S. Timpte, J. L. Abernathy, A. Toumadje, W. C. Johnson, Jr., and R. L. Hill. 1987. Porcine submaxillary gland apomucin contains tandemly repeated, identical sequences of 81 residues. *J. Biol. Chem.* 262:11339–11344.

9. Gum, J. R., J. C. Byrd, J. W. Hicks, N. W. Toribara, D. T. A. Lamport, and Y. S. Kims. 1989. Molecular cloning of human intestinal mucins cDNAs. *J. Biol. Chem.* 264:6480–6487.

10. Gendler, S., J. Papadimitriou-Taylor, T. Duhig, J. Rothbard, and J. Burchell. 1988. A highly immunogenic region of a human polymorphic epithelial mucin expressed by carcinomas is made up of tandem repeats. *J. Biol. Chem.* 263:12820–12823.

11. Wesley, A., M. Mantle, D. Man, R. Qureshi, G. Forstner, and J. Forstner. 1985. Neutral and acidic species of human intestinal mucin: evidence for different core peptides. *J. Biol. Chem.* 260:7955–7959.

12. Hill, H. D., Jr., M. Schwyzer, H. M. Steinman, and R. L. Hill. 1977. Ovine submaxillary mucin. Primary structure and substrates of UPD-N-acetylgalactosaminyl: mucin transferase. *J. Biol. Chem.* 252:3799–4010.

13. Chirgwin, J. M., A. E. Przbyla, R. J. Macdonald, and W. J. Rutter. 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. *Biochemistry.* 18:5294–5299.

14. Aviv, H., and P. Leder. 1972. Purification of biologically active globin messenger RNA by chromatography on oliogothymidylic acid-cellulose. *Proc. Natl. Acad. Sci. USA.* 69:1408–1412.

15. Gubler, U., and B. J. Hoffman. 1983. A simple and very efficient method for generating cDNA libraries. *Gene (Amst.).* 25:263–269.

16. Arrufo, A., and B. Seed. 1987. Molecular cloning of CD28 cDNA by a high-efficiency COS cell expression system. *Proc. Natl. Acad. Sci. USA.* 84:8573–8577.

17. Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular Cloning. A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

18. Henikoff, S. 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. *Gene (Amst.).* 28:351–355.

19. Riordan, J. R., J. M. Rommens, K. Bat-Sheva, N. Alon, R. Rozmahel, Z. Grzelczak, J. Zielenski, S. Lok, N. Plasvic, J.-L. Chou, M. L. Drumm, M. C. Ianuzzi, F. S. Collins, and L.-C. Tsui. 1989. Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. *Science (Wash. DC).* 245:1066–1072.

19a. Gerard, N. P., and C. Gerard. 1990. The human neurokinin A (substance K) receptor: molecular cloning of the gene, chromosome localization, and isolation of cDNA from tracheal and gastric tissues. *J. Biol. Chem.* In press.

20. Griffiths, B., J. Gum, L. F. West, S. Povey, D. M. Swallow, and Y. S. Kim. 1989. Mapping of the gene coding for intestinal mucin to chromosome 11p15. *Cytogenet. Cell Genet.* 51:1008. (Abstr.)

21. Tam, P. Y., and P. Verdugo. 1981. Control of mucus hydration as a Donnan equilibrium process. *Nature (Lond.).* 292:340–342.

22. Cheng, P.-W., T. F. Boat, K. Cranfill, J. R. Yankaskas, and R. C. Boucher. 1989. Increased sulfation of glycoconjugates by cultured epithelial cells from patients with cystic fibrosis. *J. Clin. Invest.* 84:68–72.

23. Shows, T. B., A. Y. Sakaguchi, and S. L. Naylor. 1982. Advances in Human Genetics. Vol. 12. H. Harris and K. Hirschhorn, editors. Plenum Publishing Corp., New York. 341–452.

24. Shows, T., R. Eddy, L. Haley, M. Byers, M. Henry, T. Fujita, H. Matsui, and T. Taniguchi. 1984. Interleukin 2 (IL2) is assigned to human chromosome 4. *Somatic Cell Mol. Genet.* 10:315–318.

25. Shows, T. B., J. A. Brown, L. L. Haley, M. G. Byers, R. L. Eddy, E. S. Cooper, and A. P. Goggin. 1978. Human beta glucuronidase: assignment of the structural gene to chromosome 7 using somatic cell hybrids. *Cytogenet. Cell Genet.* 21:99–104.