BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS **219,** 140–145 (1996)
ARTICLE NO. 0195

Human CFTR Gene Sequences in Regions Flanking Exon 10: A Simple Repeat Sequence Polymorphism in Intron 9

Zhidong Xu* and Dieter C. Gruenert*,†,[1]

**Cardiovascular Research Institute, Gene Therapy Core Center; and †Departments of Laboratory Medicine and Stomatology, University of California San Francisco, San Francisco, California 94143-0911*

Received December 13, 1995

A 2,908-bp segment of genomic DNA containing exon 10 and flanking intron regions of the human cystic fibrosis transmembrane conductance regulator gene was sequenced. A 30-bp sequence discrepancy and three missing nucleotides were detected when compared to a previously published 831-bp sequence. In the 30-bp region of sequence discrepancy, only a primer based on the new sequence information presented in this study gave products from polymerase chain reaction amplification of cellular DNA and a plasmid DNA encompassing the exon 10 region of CFTR. A 4-bp (TAAA) simple repeat sequence was also identified in intron 9 region. This repeat is dimorphic with nine $(TAAA)_9$ or eleven $(TAAA)_{11}$ copies on different chromosomes. Eleven repeats were exclusively associated with chromosomes carrying the δF508 mutation. Both 9 and 11 repeats were detected in non-δF508 chromosomes.

Cystic fibrosis (CF) is a severe autosomal recessive disorder affecting approximately 1 in 2500 live births in the Caucasian population (1). The gene responsible for CF, the cystic fibrosis transmembrane conductance regulator (CFTR), was recently identified (2,3,4). Genomic sequences of exons and adjacent introns of CFTR have been subsequently determined (5). This information has greatly facilitated mutation detection following polymerase chain reaction (PCR) amplification of individual CFTR exons (5). In addition, the availability of the CFTR gene has made it possible to design gene therapy strategies for CF. The exon 10 region of the CFTR gene contains many clinically important mutations (6) and it is important to have correct DNA sequence information to maintain the integrity of mutation detection and for development of gene therapy strategies.

Enhancement of mutation analysis can be facilitated by identification of polymorphisms associated with specific mutations. A simple repeat sequence (TAAA) has been previously identified and linked to specific genes (7,8,9). Although this repeat is polymorphic when comparing different genes, polymorphisms between different alleles of the same gene have not been described.

Attempts to amplify DNA comprising exon 10 and adjacent introns by PCR with a primer based on the 5′ end of the published sequence data (5) were unsuccessful using either a cloned DNA fragment found in plasmid T6/20 (3,4) or human cellular DNA. Plasmid T6/20 was previously isolated from a cDNA library. Although the sequence flanking exon 10 has not been previously verified, it has been assumed that the flanking regions were segments of CFTR introns 9 and 10.

This study reports the complete DNA sequence (2908-bp) of the *EcoRI* insert from plasmid T6/20 DNA which is identical to the human genomic DNA segment encompassing this region. A 30-bp sequence discrepancy and three missing nucleotides were detected when sequence information generated in this study was compared to the published sequence. A 4-bp (TAAA) simple repeat sequence was also detected in intron 9 region, which was polymorphic on δF508 or non-δF508 chromosomes.

[1] To whom reprint requests should be addressed at 203 Surge, Box 0911, Cardiovascular Research Institute, Gene Therapy Core Center, University of California, San Francisco, CA 94143-0911. Fax: (415) 476-9749.

0006-291X/96 $18.00

MATERIALS AND METHODS

Cell and Nucleic Acid Isolation

The T6/20 plasmid was obtained from American Type Cell Culture (Rockville, MD). The plasmid DNA was prepared by the standard CsCl ultracentrifugation and characterized by restriction enzyme digestion (10).

The 16HBE14o$^-$ is a wild-type human bronchial epithelial cell line (11) and σCFTE29o$^-$ is a human trachea epithelial cell line homozygous for the δF508 mutation (12). Both cell lines are transformed by an origin of replication defective SV40 containing plasmid. Cells were grown in Eagle's minimal essential medium (MEM) supplemented with 10% fetal bovine serum and antibiotics at 37°C under an atmosphere 5% in CO_2. Genomic DNA was isolated from cell cultures using a salting-out procedure as described (13).

Polymerase Chain Reaction

PCR was carried out in 50 μl volumes containing 0.1 μg human DNA/or 1 ng of plasmid DNA, 10 mM Tris-HCl, pH 8.4, 500 mM KCl, 2 mM $MgCl_2$, 0.001% (w/v) gelatin, 20 pmole of each primer, and 1 unit of *AmpliTaq* DNA polymerase (Perkin-Elmer). After denaturation at 94°C for 5 min, PCR was performed for 30 cycles of 94°C/30 sec, denaturation; 69°C/30 sec, annealing; and 72°C/30 sec-2 min, extension; followed by a final extension at 72°C for 5 min. Negative controls without addition of DNA templates were included in parallel during PCR experiments to ensure there was no exogenous contamination. PCR products were analyzed on 1% agarose gels stained with ethidium bromide.

DNA Sequence Analysis

The entire sequence of the *EcoRI* CFTR insert in the plasmid T6/20 was determined using Prism DyeDeoxy Terminator Cycle Sequencing kit (Perking-Elmer, Norwalk, CT). Sequencing reactions were prepared according to manufacturer's instructions except that an annealing temperature of 62°C was used. The dye labeled sequencing reactions were purified by passage through Centri-sep columns (Princeton Separations, Adelphia, NJ), and analyzed by electrophoresis using Applied Biosystems Model 373A automated DNA sequencer on 6% polyacrylamide denaturing gels.

Three overlapping genomic segments were amplified from the 16HBE14o$^-$ DNA by PCR using primers CF46/CF37 (617-bp), CF33/CF48 (1436-bp), and CF34/CF36 (2436-bp), respectively. PCR products were separated on 1% SeaPlaque GTG low melting point agarose gels (FMC, Rockland, ME) in 1X TAE buffer and the corresponding bands were excised and purified by GeneClean II kit (Bio101, Vista, CA). The purified genomic segments were then directly sequenced using γ-^{32}P-ATP end labeled primers and Sequenase version 2 (USB, Cleveland, OH).

RESULTS

PCR amplification of genomic CFTR DNA fragments comprising exon 10 and adjacent intron sequences employed oligonucleotide primers based on the 831-bp sequence published previously (5). One primer, CF2, was derived from 5′ end of the published sequence (Table 1). This primer, in conjunction with other downstream primers, failed to generate detectable PCR products from either genomic or T6/20 plasmid DNA after a thorough testing for optimal PCR conditions. Plasmid T6/20 was originally constructed from a cDNA library (3,4) and was thought to contain the CFTR gene exon 10 flanked by intron 9 and intron 10 sequences. To elucidate potential sequence discrepancies, the T6/20 plasmid and genomic DNA segments were sequenced directly from PCR amplification products.

DNA sequencing reactions were initially performed with primers based on the exon 10 region and M13 sequences in the cloning site of the plasmid. Additional primers based on newly generated sequences were synthesized for further sequencing reactions (Table 1). The relative position and orientation of the primers are shown in Fig. 1. To ensure the accuracy of the nucleotide sequence, both strands of the insert were sequenced. The overlapping sequence alignments were analyzed using the Gene Works software (Inteligenetics, Inc., Mountain View, CA).

The complete 2908-bp sequence of the *Eco RI* DNA insert in plasmid T6/20 is presented in Fig. 2. The data includes the previously published 831-bp sequence (5). When the sequence data were aligned and compared, a 30-bp nucleotide (nt) sequence discrepancy (nt 757-786) was detected at the 5′ end of the published sequence. In addition, three bases at nt position 818, 1548, and 1578, respectively, were found absent in the published sequence.

To verify that the sequence of the CFTR insert in plasmid T6/20 was in fact, genomic in origin,

TABLE 1
Oligonucleotides Used for Sequencing and Amplification by PCR

CF2(+)	ACTGTAGCTGTACTACCTTCCATC
CF32(−)	GACAATTACAATACAGTGTGACAAG
CF33(+)	GGGTTCATTTGATCACAATAAATGC
CF34(+)	CAGCTTTTCTTAATAAAGCAATCAG
CF35(+)	TGCAATTCTTTGATGCAGAGGCAA
CF36(−)	AGAATGAGAGACCCACAGTACTAAA
CF37(−)	ATTCAGGTAATATTGTTCCCATGAG
CF38(+)	TATTGACAGTATACTCCAAATAGTG
CF39(+)	TAACCTTTCCCATTCTTCCTCCA
CF40(+)	TCTACTTTGTAGGATTTCTGTGAAG
CF41(−)	ATTCTCTGCTGGCAGATCAATGC
CF42(+)	TGACAGAGTGAGACTCTGTCTC
CF43(−)	AGCAGAGGAAGAAAAAGCACTGAT
CF46(+)	GAATTCCAGCCAGACGTGATGG
CF48(−)	GAATTCCCTCTTCTCCTTTTTTCA
CF49(+)	ATGGTGGTTTGATTTCCTCAAGTC
C16B(+)	GTTTTCCTGGATTATGCCTGGCAC
C16D(−)	GTTGGCATGCTTTGATGACGCTTC

Note: All oligonucleotides are presented from 5′ → 3′. Primers C16B and C16D have been previously described (2). Primers were synthesized and redissolved in double deionized H_2O without further purification. (+) = sense primers; (−) = antisense primers.

DNA from a human bronchial epithelial cell line, 16HBE14o$^-$ (11) was sequenced directly from PCR amplification products. Three overlapping genomic segments of DNA of 617-bp, 1,436-bp, and 2,436-bp in size, were sequenced. Manual sequence analysis showed that the genomic sequence is identical to that of the plasmid T6/20 insert (Fig. 2). The last 30-bp from the 5′ end of the published sequence (nt 757-786 in Fig. 2) was also not present in the genomic DNA.

PCR was performed with primers based on the DNA sequence generated in this study. Primer CF39 was used to replace primer CF2 (Table 1). The expected 485-bp fragment was observed using primer CF39 and C16D in samples from 16HBE14o$^-$, ΣCFTE29o$^-$, and the plasmid T6/20 (Fig. 3).Similarly, amplification with primers CF40/CF41, localized outside previously published sequence, gave the expected 1057-bp fragment from the three DNA samples (Fig. 3). The identity of the PCR products was further verified by membrane transfer and hybridization with an internal oligonucleotide, C16B, which is localized in exon 10 of the CFTR gene (data not shown).

Upon inspection of the intron regions, a 4-bp $(TAAA)_n$ tandem sequence repeat was detected in intron 9 from nt 155-190 (908-bp upstream from exon 10) (Fig. 2). This repeat contained 9 or 11 copies upon screening with two flanking primers, CF42 and CF43 (Table 1). Initial analysis (Table 2) showed that 11 repeats $(TAAA)_{11}$ were linked to δF508 chromosomes from homozygous

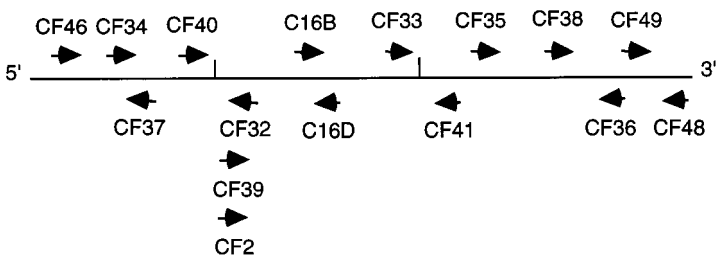

FIG. 1. Schematic diagram of primer locations for PCR amplification and sequencing reaction. Vertical lines represent the region of published DNA sequence (not to scale).

```
GAATTCCAGC CAGACGTGAT GGCGGGTGCC CGTAGTCCCA GCTACTCGGG AGGCTGAGGC          60
AGGAGAATGG CGTGAACCCA GGAGGCAGAA CTTGCAGTGA GCCGAGATCG CGCCACTGCA         120
CTCTAGCCTG GGTGACAGAG TGAGACTCTG TCTCTAAATA AATAAATAAA TAAATAAATA         180
AATAAATAAA ATCAGTGCTT TTTCTTCCTC TGCTACCTCC TTTCCTTCTA CTCAGTTTTA         240
GTCAGTAGTA TTATCTTTTT TCAGATTTAT CTTTGTATTG TTAAATCTGC TTATGCTTCT         300
ATTACTTTAT TTATTAGCTT TAAATGATAC CTTTTGACTT TCAGCTTTTC TTAATAAAGC         360
AATCAGCAAA TTTCCTTTAC ACTCCACACT TATACCCCAT TTCCTTTGTT TGTTTATTTG         420
GTTTTTACTT CTAACTTTTC TTATTGTCAG GACATATAAC ATATTTAAAC TTTGTTTTTC         480
AACTCGAATT CTGCCATTAG TTTTAATTTT TGTTCACAGT TATATAAATC TTTGTTCACT         540
GATAGTCCTT TTGTACTATC ATCTCTTAAA TGACTTTATA CTCCAAGAAA GGCTCATGGG         600
AACAATATTA CCTGAATATG TCTCTATTAC TTAATCTGTA CCTAATAATA TGAAGGTAAT         660
CTACTTTGTA GGATTTCTGT GAAGATTAAA TAAATTAATA TAGTTAAAGC ACATAGAACA         720
GCACTCGACA CAGAGTGAGC ACTTGGCAAC TGTTAGCTGT TACTAACCTT TCCCATTCTT         780
                                          cactgtagct gtactacctt ccat
CCTCCAAACC TATTCCAACT ATCTGAATCA TGTGCCCCTT CTCTGTGAAC CTCTATCATA         840
ctcctc.... .......... .......... .......-.. .......... ..........
                                                                          //

.......................................... AAACGTCCTC AATGGTTATT        1560
.......................................... ........-.. ..........
TATATGGCAT GCATATAAGT GATATGTGGT ATCTTTTTAA AAGATACCAC AAAATATGCA        1620
.......... ........-.. ..........
TCTTTAAAAA TATACTCCAA AAATTATTAA GATTATTTTA ATAATTTTAA TAATACTATA        1680
GCCTAATGGA ATGAGCATTG ATCTGCCAGC AGAGAATTAG AGGGGTAAAA TTGTGAAGAT        1740
ATTGTATCCC TGGCTTTGAA CAAATACCAT ATAACTTCTA GTGACTGCAA TTCTTTGATG        1800
CAGAGGCAAA ATGAAGATGA TGTCATTACT CATTTCACAA CAATATTGGA GAATGAGCTA        1860
ATTATCTGAA AATTACATGA AGTATTCCAA GAGAAACCAG TATATGGATC TTGTGCTGTT        1920
CACTATGTAA ATTGTGTGAT GGTGGGTTCA GTAGTTATTG CTGTAAATGT TAGGGCAGGG        1980
CAATATGTTA CTATGAAGTT TATTGACAGT ATACTCCAAA TAGTGTTTGT GATTCAAAAG        2040
CAATATCTTT GATAGTTGGC ATTTGCAATT CCTTTATATA ATCTTTTATG AAAAAAATTG        2100
CAGAGAAAGT AAAATGTAGC TTAAAATACA GTATCCAAAA AAATGGAAAA GGGCAAACCG        2160
TGGATTAGAT AGAAATGGCA ATTCTTATAA AAAGGGTTGC ATGCTTACAT GAATGGCTTT        2220
CCATGTATAT ACTCAGTCAT TCAACAGTTT TTTTTTTAGA GCCCCATCCT TATTTTTTAT        2280
ACACTTTGAG AGCATAATGA AAAGAAAAGC TACCTGCAAA AGTTTTGGAC TTACCTCAAA        2340
GAGGATATAC TACATTCCTC AAAAGGCCTT CTTCCAGGAA TAGTATTTCA TAACCTGGAG        2400
GTTGGAAAAA TCTGGATTAG TTACAAAAAA ATCTGAGTGT TTCTAGCGGA CACAGATATT        2460
TGTCTAGGAG GGGACTAGGT TGTAGCAGTG GTAGTGCCTT ACAAGATAAA TCATGGGCTT        2520
TATTTACTTA CGAGTGGAAA AGTTGCGGAA GGTGCCTTAC AGACTTTTTT TTTGCGTTAA        2580
GTATGTGTTT TCCCATAGGA ATTAATTTAT AAATGGTGGT TTGATTTCCT CAAGTCAACC        2640
TTTAAAAGTA TATTTAGCCA AAATATAGCT TAAATATATT ACTAGTAATA AATTTAGTAC        2700
TGTGGGTCTC TCATTCTCAA AATGAGCATT TACTAATTTC TGAACACTGT GCTAGGTCCT        2760
GGGAATACCA AATTGAATAA GACATAGTCT ATTTTTCTGA AGGGTTTATA GCAGAGTCCC        2820
CTGTGTTAAT AATGAAGGAG TGTGTGGTAT GTGAATCATA TATCAATAGG GTTGTTAAAA        2880
ATAATGAAAA AAGGAGAAGA GGGAATTC                                           2908
```

FIG. 2. Nucleotide sequence spanning exon 10 and adjacent intron regions of the human CFTR gene. The capital letters are the sequence data generated from this study. Lowercase letters and dots represent the published data (5). The $(TAAA)_n$ repeat is underlined from nucleotide (nt) 155–190. The 30-bp sequence discrepancy is indicated by underlined lowercase letters (nt 757–786) at the 5′ end of the published data. Single nt deletions (nt 818, 1548, 1578, respectively) are indicated by dashes. The region from nt 841–1541 (//) is not included since it is identical to data already published. The GenBank Accession # is L49160.

patients (n = 20). Of 10 samples with one δF508 and one non-δF508 chromosome, 12 chromosomes were associated with 11-repeats and 8 chromosomes with 9-repeats. Analysis of 15 non-δF508 DNA samples revealed that 9 chromosomes were linked with 11-repeats and another 21 chromosomes were associated with 9-repeats (Table 2).

DISCUSSION

Genomic CFTR sequences of exons and flanking introns were previously reported (5) and have been extensively used for mutation detection and prenatal diagnosis. In the present study, the

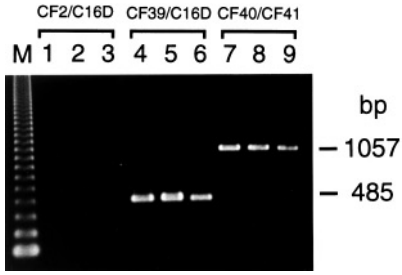

FIG. 3. PCR analysis of DNA fragments generated from the CFTR gene exon 10 and adjacent introns. PCR primers used in each sample are indicated on the top. Lanes 1,4,7, plasmid T6/20; lanes 2,5,8, normal human bronchial epithelia cell 16HBE14o$^-$ DNA; lanes 3,7,9, human trachea epithelia cell with homozygous δF508 mutation σCFTE29o$^-$ DNA. M, 123 bp molecular weight ladders. No PCR products are produced with primers CF2 and C16D as shown in lanes 1–3. The expected PCR fragment sizes in bp are indicated on the right.

known genomic human CFTR sequence, comprising exon 10 and adjacent introns, was expanded from 831-bp to 2908-bp. By sequencing plasmid T6/20 and genomic DNA, plasmid T6/20 was verified to contain a genomic CFTR insert.

The new sequence revealed a region containing a 30-bp sequence discrepancy as well as 3 single nucleotide deletions when compared to the previously reported sequence. The cause of these sequence variations in the earlier report are not known, but it is possible that they could be the result of base compression or other complications often associated with sequencing reactions.

The 4-bp sequence repeats $(TAAA)_n$ detected in intron 9 represent a microsatellite unit not previously reported in the CFTR gene. Several other simple repeat sequences have been described to be polymorphic in this gene (5). The most common simple repeat sequences are dinucleotides, such as $(CA/TG)_n$ and $(TA)_n$. These dinucleotide repeats were reported to be highly polymorphic for carrier detection and prenatal diagnosis of CF (14,15,16). In addition, a GATT tandem repeat was found to exist in dimorphic forms of either 6 or 7 units, with 6 repeats, $(GATT)_6$, always linked with the δF508 mutation (17).

The $(TAAA)_n$ repeats detected in this study appear to be dimorphic, either 9 repeats $(TAAA)_9$ or 11 repeats $(TAAA)_{11}$. Initial PCR screening showed that the $(TAAA)_{11}$ repeat was linked to ΔF508 chromosomes tested (n = 30) (Table 2). Of the 40 non-ΔF508 chromosomes, $(TAAA)_{11}$ was detected in 11 (2 + 9) chromosomes, and $(TAAA)_9$ was detected in 29 (8 + 21) chromosomes.

Further studies are required to determine whether or not the $(TAAA)_9$ repeat is exclusively associated with the wild-type CFTR gene or if the $(TAAA)_{11}$ repeat is only linked to chromosomes with CFTR mutations. The TAAA repeats have been reported in other genes, such as annexin III gene (7), the adenosine deaminase gene (8,9), and *Alu* repetitive elements (8). However, the repeats are polymorphic between genes, and have not been associated with specific mutations. While the functional role of the repeat elements remains to be defined for CFTR, the polymorphisms of TAAA units may be useful in genotying and characterizing the evolution of the δF508 mutation.

TABLE 2
Polymorphic Analysis of TAAA Repeats in CFTR Gene

DNA Sample Number	Chromosome Copy Number	Genotype for δF508 Mutation[a]	$(TAAA)_{11}$	$(TAAA)_9$
10	20	ΔF508/ΔF508	20	0
10	20	ΔF508/Non-ΔF508	10/2[b]	8
15	30	Non-ΔF508/Non-ΔF508	9	21

Note: [a] Genotypic status was determined for the presence of the ΔF508 mutation while mutational status for non-ΔF508 chromosomes was unknown. [b] Ten were detected from ΔF508 chromosomes and two from non-ΔF508 chromosomes.

ACKNOWLEDGMENTS

The authors wish to thank Dr. Farid Chehab and Jeff Wall for providing human lymphocyte DNA samples. This work was supported by NIH Grants DK46002 and DK47766 and a grant from the Cystic Fibrosis Foundation.

REFERENCES

1. Boat, T. F., Welsh, M. J., and Beaudet, A. L. (1989) *in* Cystic Fibrosis (Scriver, C. L., Beaudet, A. L. Sly, W. S., and Valle, D., Eds.), pp. 2649–2680, McGraw–Hill, New York.
2. Kerem, B., Rommens, J. M., Buchanan, J. A., Markiewicz, D., Cox, T. K., Chakravarti, A., and Tsui, L.-C. (1989) *Science* **245,** 1073–1080.
3. Riordan, J. R., Rommens, J. M., Kerem, B., Alon, N., Rozmahel, R., Grzelczak, Z., Zielenski, J., Lok, S., Plavsic, N., Chou, J.-L., Drumm, M., Iannuzzi, M. C., Collins, S. F., and Tsui, L.-C. (1989) *Science* **245,** 1066–1073.
4. Rommens, J. M., Iannuzzi, M. C., Kerem, B., Drumm, M. L., Melmer, G., Dean, M., Rozmahel, R., Cole, J., Kennedy, D., Hidaka, N., Zsiga, M., Buchwald, M., Riordan, J., Tsui, L.-C., and Collins, F. S. (1989) *Science* **245,** 1059–1065.
5. Zielenski(a), J., Rozmahel, R., Bozon, D., Kerem, B., Grzelczak, Z., Riordan, J. R., Rommens, J., and Tsui, L.-C. (1991) *Genomics* **10,** 214–228.
6. Tsui, L.-C. (1992) *Trends. Genet.* **8,** 392–398.
7. Tait, J. F., Smith, C., Xu, L., and Cookson, B. T. (1993) *Genomics* **18,** 79–86.
8. Economou, E. P., Bergen, A. W., Warren, A. C., and Atonarakis, S. E. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 2951–2954.
9. Chen, E. H., Tartaglia, A. P., and Mitchell, B. S. (1993) *Am. J. Hum. Genet* **53,** 889–893.
10. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *in* Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
11. Cozens, A. L., Yezzi, M. J., Kunzelmann, K., Ohrui, T., Chin, L., Eng, K., Finkbeiner, W. E., Widdicombe, J. H., and Gruenert, D. C. (1994) *Am. J. Respir. Cell. Mol. Biol.* **10,** 38–47.
12. Kunzelmann, K., Schwiebert, E., Kuo, W.-L., Stanton, B. A., and Gruenert, D. C. (1993) *Am. J. Respir. Cell. Mol. Biol.* **8,** 522–529.
13. Miller, S. A., Dykes, D. D., and Polesky, H. F. (1988) *Nucleic. Acid. Res.* **16,** 1215.
14. Zielenski(b), J., Markiewicz, D., Rininsland, F., Rommens, J., and Tsui, L.-C. (1991) *Am. J. Hum. Genet.* **49,** 1256–1262.
15. Mornet, E., Chateau, C., Simon-Bouy, B., Boue, J., Zielenski, J., Tsui, L. C., and Boue, A. (1992) *Hum. Genet.* **88,** 479–481.
16. Morral, N., Girbau, E., Zielenski, J., Nunes, V., Casals, T., Tsui, L.-C., and Estivill, X. (1992) *Hum. Genet.* **88,** 356.
17. Chehab, F. F., Johnson, J., Louie, E., Goossens, M., Kawasaki, E., and Erlich, H. (1991) *Am. J. Hum. Genet.* **48,** 223–226.