Identification of the Cystic Fibrosis Gene: Cloning and Characterization of Complementary DNA

John R. Riordan, Johanna M. Rommens, Bat-sheva Kerem, Noa Alon, Richard Rozmahel, Zbyszko Grzelczak, Julian Zielenski, Si Lok, Natasa Plavsic, Jia-Ling Chou, Mitchell L. Drumm, Michael C. Iannuzzi, Francis S. Collins, Lap-Chee Tsui

Overlapping complementary DNA clones were isolated from epithelial cell libraries with a genomic DNA segment containing a portion of the putative cystic fibrosis (CF) locus, which is on chromosome 7. Transcripts, approximately 6500 nucleotides in size, were detectable in the tissues affected in patients with CF. The predicted protein consists of two similar motifs, each with (i) a domain having properties consistent with membrane association and (ii) a domain believed to be involved in ATP (adenosine triphosphate) binding. A deletion of three base pairs that results in the omission of a phenylalanine residue at the center of the first predicted nucleotide-binding domain was detected in CF patients.

Cystic Fibrosis (CF) is an autosomal recessive genetic disorder affecting a number of organs, including the lung airways, pancreas, and sweat glands (1). Abnormally high electrical potential differences have been detected across the epithelial surfaces of the CF respiratory tract, including the trachea and nasal polyps, as well as across the walls of CF sweat gland secretory coils and reabsorptive ducts (2). The basic defect has been associated with decreased chloride ion conductance across the apical membrane of the epithelial cells (3). That the defect also appeared to persist in cultured cells derived from several epithelial tissues suggested that the CF gene is expressed in these cells (4). More recently, patch clamp studies showed that this defect is probably due to a failure of an outwardly rectifying anion channel to respond to phosphorylation by cyclic AMP–dependent protein kinase (protein kinase A) or protein kinase C (5). Although progress has been made in the isolation of polypeptide components of an epithelial chloride channel that mediates conductance (6), their relation to the kinase-activated pathway and CF has yet to be established, and the basic biochemical defect in CF remains unknown.

Molecular cloning experiments have permitted the isolation of a large, contiguous segment of DNA spanning at least four transcribed sequences from a region thought to contain the CF locus (7). These sequences were initially identified on the basis of their ability to detect conserved sequences in other animal species by DNA hybridization and were subsequently characterized by RNA hybridization experiments, cDNA isolation, and direct DNA sequence analysis (7). Three of the transcribed regions were excluded from being the CF locus by earlier genetic or DNA sequence analyses (7, 8). The fourth one, as shown by genetic analysis (9) and DNA sequencing analysis presented below, corresponds to a portion of the CF gene locus.

Isolation of cDNA clones. Two DNA segments (E4.3 and H1.6) that detected cross-species hybridization signals (7) were used as probes to screen cDNA libraries made from several tissues and cell types (10). After screening seven different libraries, one single clone (10-1) was isolated with H1.6 from a cDNA library made from the cultured epithelial cells of the sweat glands of an unaffected (non-CF) individual (10).

DNA sequencing showed that 10-1 contained an insert of 920 base pairs (bp) in size and one potential, long open reading frame (ORF). Since one end of the sequence shared perfect sequence identity with H1.6, it was concluded that the cDNA clone was probably derived from this region. The DNA sequence in common was, however, only 113 bp long (Figs. 1 and 2). This sequence in fact corresponded to the first axon of the putative CF gene. The short sequence overlap thus explained the weak hybridization signals in library screening and our inability to detect transcripts in RNA gel-blot analysis. In addition, the orientation of the transcription unit was tentatively established on the basis of alignment of the genomic DNA sequence with the presumptive ORF of 10-1.

Since the corresponding transcript was estimated to be about 6500 nucleotides in length by RNA gel-blot hybridization experiments, further cDNA library screening was required in order to clone the remainder of the coding region. As a result of several successive screenings with cDNA libraries generated from the colon carcinoma cell line T84, normal and CF sweat gland cells, pancreas,
and adult lungs, 18 additional clones were isolated (Fig. 1). DNA sequence analysis revealed that none of these cDNA clones corresponded to the length of the observed transcript, but it was possible to derive a consensus sequence based on overlapping regions. Further cDNA clones corresponding to the 5' and 3' ends of the transcript were derived from 5' and 3' primer-extension experiments (Fig. 1). Together, these clones span about 6.1 kb and contain an ORF capable of encoding a protein of 1480 amino acids (Fig. 2).

It was unusual that most of the cDNA clones isolated here contained sequence insertions at various locations (Fig. 1). While many of these extra sequences corresponded to intron regions reverse-transcribed during the construction of the cDNA, as revealed on alignment with genomic DNA sequences, the identities of several others were uncertain because they did not align with sequences at the corresponding exon-intron junctions, namely, the sequences at the 5' ends of clones 13a and T16-1 and at the 5' and 3' ends of T11, and the insertions between exons 3 and 4 in 13a and between exons 10 and 11 in T16-4.5 (legend to Fig. 1). More puzzling were the sequences corresponding to the reverse complement of exon 6 at the 5' end of 11a and the insertion of a segment of a bacterial transposon in clone C16-1; none of these could be explained by mRNA processing errors.

In that the number of recombinant cDNA clones for the putative CF gene detected in the library screening was much less than would have been expected from the abundance of transcripts estimated from RNA hybridization experiments, it seemed probable that the clones that contained aberrant structures were preferentially retained while the proper clones were lost during propagation. Consistent with this interpretation, poor growth was observed for most of our recombinant clones isolated, regardless of the vector used.

**RNA analysis.** To visualize the transcript of the putative CF gene, we used RNA gel-blot hybridization with the 10-1 cDNA as the probe (Fig. 3). The analysis revealed a prominent band, about 6.5 kb in size, in T84 cells. Identical results were obtained with other cDNA clones as probes. Similar, strong hybridization signals were also detected in pancreas and primary cultures of cells from nasal polyps, suggesting that the mature mRNA of the putative CF gene is about 6.5 kb. Minor hybridization signals, probably representing degradation products, were detected at the lower size ranges, but they varied between different experiments. On the basis of the hybridization band intensity and comparison with those detected for other transcripts under identical experimental conditions, it was estimated that the putative CF gene transcripts constituted about 0.01 percent of total mRNA in T84 cells.

Additional tissues were analyzed by RNA gel-blot hybridization in an attempt to correlate the expression pattern of the putative CF gene and the pathology of CF. Transcripts, all of identical size, were found in lung, colon, sweat glands (cultured epithelial cells), placenta, liver, and parotid gland, but the signal in these tissues was generally weaker than that detected in the pancreas and nasal polyps (Fig. 3). Intensity varied among different preparations; for example, hybridization in kidney was not detectable in the preparation shown in Fig. 3 but was clearly discernible subsequently. Transcripts were not detected in the brain or adrenal gland, nor in skin fibroblast and lymphoblast cell lines.

Thus, expression of the putative CF gene appeared to occur in many of the tissues examined, with higher levels in those tissues severely affected in CF. While this epithelial tissue–specific expression pattern is in good agreement with the disease pathology, no significant difference was detected in the amount or size of transcripts from CF and control tissues (Fig. 3), consistent with the assumption that CF mutations are subtle changes at the nucleotide level.

**Characterization of cDNA clones.** As indicated above, a contig-
The open triangle indicates the position at which 3 bp are deleted in CF.

4561 GATACAAGGCTTTAGAGAGCAGCATAAATGT-ACATGGGACATTTGCTCATGGAATTGG
4581 AAAACAAGGATGAATTAAGTTTTTTTTTAAAAAAGAAACATTTGGTAAGGGGAATTGAGG

were performed to ensure the authenticity of the consensus sequencing overlapping cDNA clones. Since most of the cDNA clones were presumably coding region of the CF locus could be deduced from sequence are shown above [nucleotides 4561 to 6129 plus the poly(A) tail] numbers on the right amino acid residue positions. The first base position corresponds to the first nucleotide in the 5' extension clone PA3-5, which is one nucleotide longer than TB2-7 (12). The 3' end and the noncoding DNA sequencing was performed by the dideoxy chain termination method over the alignment clone PA3-5, which is one nucleotide longer than TB2-7 (12). The 3' end and the noncoding DNA sequencing was performed by the dideoxy chain termination method Eisenberg et al. (35) are enclosed in boxes. Amino acids comprising putative ATP-binding folds are underlined. Possible sites of phosphorylation (27) by protein kinases A or C are indicated by open and closed circles, respectively. The open triangle indicates the position at which 3 bp are deleted in CF. Abbreviations for the amino acid residues are: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

The 5' terminus of the transcript was determined by primer extension (11). A modified polymerase chain reaction, anchored PCR (12), was also used to facilitate cloning of the 5' end sequences. Two independent 5' extension clones, one from pancreas and the other from T84 RNA, were characterized by DNA sequencing and restriction enzyme analysis. The positions of exon junctions are indicated by vertical lines. Potential membrane-spanning segments ascertained with the use of the algorithm of Eisenberg et al. (35) are enclosed in boxes. Amino acids comprising putative ATP-binding folds are underlined. Possible sites of phosphorylation (27) by protein kinases A or C are indicated by open and closed circles, respectively. The open triangle indicates the position at which 3 bp are deleted in CF. Abbreviations for the amino acid residues are: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

The cDNA was further characterized in gel hybridization experiments with genomic DNA. Five to six different restriction fragments could be detected with the 10-1 cDNA in Eco RI- or Hind III-digested total human DNA and a similar number of fragments with several other cDNA clones, suggesting the presence of multiple exons for the putative CF gene. The hybridization studies also identified the cDNA clones with unprocessed intron sequences when they showed preferential hybridization to a smaller subset of genomic DNA fragments with relatively greater intensities. For the confirmed cDNA clones, their corresponding genomic DNA segments were isolated (7) and the exons and exon-intron boundaries were sequenced. In all, 24 exons were identified (Fig. 2). Physical mapping experiments (7) showed that the gene locus spanned about 250 kb.

The 5' terminus of the transcript was determined by primer extension (11). A modified polymerase chain reaction, anchored PCR (12), was also used to facilitate cloning of the 5' end sequences. Two independent 5' extension clones, one from pancreas and the other from T84 RNA, were characterized by DNA sequencing and differed by only 1 base in length, thus establishing the most probable initiation site for the transcript (Fig. 2). Since the initial cDNA clones did not contain a poly(A)+ tail indicative of the end of a mRNA, anchored PCR was also applied to the 3' end of the transcript (12). The results derived from the use of several different 3'-extending oligonucleotides were consistent with the interpreta-
tion that the end of the transcript was about 1.2 kb downstream of the Hind III site at nucleotide position 5027 (Fig. 2).

The complete cDNA sequence spans 6129 base pairs excluding the poly(A)+ tail at the end of the 3' untranslated region and it contains an ORF capable of encoding a polypeptide of 1480 amino acids (Fig. 2). An ATG (AUG) triplet is present at the beginning of this ORF (base position 133–135). Since the nucleotide sequence surrounding this codon (5'-AGACCAUGCA-3') has the proposed features of the consensus sequence (CC)ACCAUGG(G) of a eukaryotic translation initiation site (13), with a highly conserved A at the −3 position, it is highly probable that this AUG corresponds to the first methionine codon for the putative polypeptide.

Detection of mutation. A comparison between the cDNA sequences derived from CF and unaffected (N) individuals was next conducted. Two clones, C16-1 and C1-1/5, were derived from a CF patient. A deletion (AF508) was detected in both CF clones. This deletion (AF508) was detected in both CF clones. The full cDNA sequence was a 3-bp deletion (Fig. 4), which would result in the loss of a phenylalanine residue (position 508) in the predicted CF polypeptide. This deletion (AF508) was detected in both CF clones. To exclude the possibility that this difference was due to a cloning artifact, sequence-specific oligonucleotides were used to screen DNA samples from CF families. Specific hybridization could be observed for each oligonucleotide probe with genomic DNA amplified by PCR, confirming the presence of corresponding genomic DNA sequences (9). Furthermore, the oligonucleotide specific for the 3-bp deletion hybridized to 68 percent of chromosomes carrying a CF mutation but not to any of the normal chromosomes (0/198), an indication that a silent sequence polymorphism was unlikely.

Sequence differences were found elsewhere among different cDNA clones probably representing sequence polymorphisms or cDNA cloning artifacts (14).

Predicted protein structure. Analysis of the sequence of the overlapping cDNA clones (Fig. 2) predicted a polypeptide of 1480 amino acids with a molecular mass of 168,138 daltons. The most characteristic feature of the predicted protein is the presence of two repeated motifs, each of which consists of a domain capable of spanning the membrane several times and sequences resembling consensus nucleotide (ATP)-binding folds (NBF's) (15) (Figs. 5 and 6). These characteristics are remarkably similar to those of the mammalian multidrug resistance P-glycoprotein (16) and a number of other membrane-associated proteins (as discussed below), suggesting that the predicted CF gene product is likely to be involved in the transport of substances (ions) across the membrane and is probably a member of a membrane protein superfamily (17). For the convenience of future discussion and to avoid confusion with the previously named CF protein and CF factor (18), we will call the protein Cystic Fibrosis Transmembrane Conductance Regulator (CFTR).

**Fig. 5.** Hydropathy profile and predicted secondary structures of the CFTR. (A) The mean hydropathy index determined according to Kyte and Doolittle (19) of nine-residue peptides is plotted against the amino acid sequence. (B) The corresponding positions of features of secondary structure predicted according to Garnier et al. (19). C, coil; T, turn; S, sheet; H, helix.

**Fig. 6.** Alignment of the three most conserved segments of the amino acid sequences (single letter code) of the extended NBF's of CFTR with comparable regions of other proteins. These three segments consist of residues 433 to 473, 488 to 513, and 542 to 584 of the amino-terminal (N) half and 1219 to 1259, 1277 to 1302, and 1340 to 1382 of the carboxyl-terminal (C) half of CFTR. The heavy overlining points out the regions of greatest similarity. The star indicates the position corresponding to the conserved feature of the consensus sequence (CC)ACCAUGG(G) of a eukaryotic translation initiation site (13), with a highly conserved A at the −3 position, it is highly probable that this AUG corresponds to the first methionine codon for the putative polypeptide. The heavy overlining points out the regions of greatest similarity. The star indicates the position corresponding to the conserved feature of the consensus sequence (CC)ACCAUGG(G) of a eukaryotic translation initiation site (13), with a highly conserved A at the −3 position, it is highly probable that this AUG corresponds to the first methionine codon for the putative polypeptide. The corresponding positions of features of secondary structure predicted according to Garnier et al. (19). C, coil; T, turn; S, sheet; H, helix.
putative CF gene product the cystic fibrosis transmembrane conductance regulator (CFTR).

Each of the predicted membrane-associated regions of CFTR consists of six hydrophobic segments capable of spanning a lipid bilayer (19), which are followed by a large hydrophilic region containing the NBF's (Fig. 5). On the basis of sequence alignment with other nucleotide-binding proteins, each of the putative NBF's in CFTR comprises at least 150 residues (Fig. 6). The single residue deletion (ΔF_{508}) detected in most of the CF patients is in the first NBF, between the two most highly conserved segments within this sequence. The amino acid sequence identity between the region surrounding the ΔF_{508} mutation and the corresponding regions of several other proteins suggests that this region is of functional importance (Fig. 6). A hydrophobic amino acid, usually one with an aromatic side chain, is present in most of these proteins at the position corresponding to Phe^{508} of CFTR.

Despite the overall symmetry in the two-motif structure of the protein and the sequence conservation of the NBF's, sequence identity between the two motifs of the predicted CFTR protein is modest. The strongest identity is between sequences at the carboxyl ends of the NBF's. Of the 66 residues aligned within these regions, 27 percent are identical and 11 percent are functionally similar. The overall weak internal sequence identity is in contrast to the much higher degree (>70 percent) in P-glycoprotein for which a sequence duplication hypothesis has been proposed (16). The lack of conservation in the relative positions of the exon-intron boundaries in the CF gene also argues against recent exon duplication as a mechanism in the evolution of this gene (Fig. 2).

Since there is apparently no signal-peptide sequence at the amino terminus of CFTR (Fig. 7), the highly charged hydrophilic segment preceding the first transmembrane sequence is probably oriented in the cytoplasm. Each of the two sets of hydrophobic helices are expected to form three traversing loops across the membrane and little of the sequence of the entire protein is expected to be exposed to the exterior surface, except the region between transmembrane segments 7 and 8. It is of interest that the latter region contains two potential sites for N-linked glycosylation (20).

A highly charged cytoplasmic domain can be identified in the middle of the predicted CFTR polypeptide, linking the two halves of the protein. This domain, named the R domain, is operationally defined by a single large exon in which 69 of the 241 amino acids are polar residues arranged in alternating clusters of positive and negative charges. In addition, four of the ten sites at which there are consensus sequences for phosphorylation by protein kinase A and seven of the potential substrate sites for protein kinase C found in CFTR are located in this exon (21).

Properties of CFTR could be further derived from comparison to other membrane-associated proteins (Fig. 6). In addition to the overall structural similarity with P-glycoproteins, each of the two predicted motifs in CFTR shows resemblance to the single motif structure of hemolysin B of Escherichia coli (22) and the product of the White gene of Drosophila (23). These proteins are involved in the transport of the lytic peptide of the hemolysin system and of eye pigment molecules, respectively. The vitamin B12 transport system of E. coli, BtuD (24), and MbpX (25), which is a liverwort chloroplast gene product whose function is unknown, also have a similar structural motif. Further, CFTR shares structural similarity with several of the periplasmic solute transport systems of Gram-negative bacteria, where the transmembrane region and the ATP-binding folds are contained in separate proteins that function in concert with a third substrate-binding polypeptide (26).

The overall structural arrangement of the transmembrane domains in CFTR is similar to several cation channel proteins (27) and some cation-translocating adenosine triphosphatases (ATPases) (28) as well as the recently described adenylyl cyclase of bovine brain (29). Short regions of sequence identity have also been detected between the putative transmembrane regions of CFTR and other membrane-spanning proteins (30). In addition, a sequence of 18 amino acids situated approximately 50 residues from the carboxyl terminus of CFTR shows some identity with the ras serine-threonine kinase proto-oncogene product of Xenopus laevis (31).

Finally, a sequence identity (10 of 13 amino acid residues) has been noted between the hydrophilic segment (position 701 to 713) within the highly charged R domain of CFTR and a region immediately preceding the first transmembrane loop of the sodium channels in both rat brain and eel (32). This feature of CFTR is not shared with the topologically closely related P-glycoprotein; the 241-amino acid linking peptide is apparently the major difference between the two proteins.

Relevance to the CF anion transport defect. In view of the genetic data of Kerem et al. (9) and the tissue specificity and predicted properties of the CFTR protein, it is reasonable to conclude that CFTR is directly responsible for CF. It remains unclear, however, how CFTR is involved in the regulation of ion conductance across the apical membrane of epithelial cells.

It is possible that CFTR serves as an ion channel itself. For example, 10 of the 12 putative transmembrane regions contain one or more amino acids with charged side chains (Fig. 7), a property similar to that of the brain sodium channel and the γ-aminobutyric acid (GABA) receptor chloride channel subunits, where charged residues are present in four of the six, and three of the four, respective membrane-associated domains per subunit or repeat unit (32, 33). The amphipathic nature of these transmembrane segments is believed to contribute to the channel-forming capacity of these molecules. In contrast, the closely related P-glycoprotein, which is
not believed to conduct ions, has only two charged residues in all 12 transmembrane domains. Alternatively, CFTR may not be an ion channel but instead may serve to regulate ion channel activities. In support of the latter possibility, none of the recently purified polypeptides (from trachea and kidney) that are capable of reconstituting chloride channels in lipid membranes (6) appear to be CFTR, judged on the basis of molecular mass. In any case, the presence of ATP-binding domains in CFTR suggests that ATP hydrolysis is directly involved and required for the transport function. The high density of phosphorylation sites for protein kinases A and C and the clusters of charged residues in the R domain may both serve to regulate this activity. The deletion of Phe508 in the NBF may prevent proper binding of ATP or the conformational change required for normal CFTR activity, consequently resulting in the observed insensitivity to activation by protein kinase A–or protein kinase C–mediated phosphorylation of the CF apical chloride conductance pathway (5). Since the predicted structure of CFTR contains several conserved domains and belongs to a family of proteins, most of which function as parts of multicomponent molecular systems (15), the CFTR protein may also participate in epithelial cell functions not related to ion transport.

To understand the basic defect in CF, it is necessary to determine the precise role of Phe508 in the regulation of ion transport and to understand the mechanism that leads to the pathophysiology of the disease. With the CF gene (that is, the cDNA) now isolated, it should be possible to elucidate the control of ion transport pathways in epithelial cells in general. Knowledge gained from study of the CF gene product (CFTR), both the normal and mutant forms, will provide a molecular basis for the development of improved means of treatment of the disease.

REFERENCES AND NOTES


10. The cDNA libraries from cultured epithelial cells were prepared as follows: sweat gland cells derived from a non-CF individual and from a CF patient were grown to first passage as described [G. Collie, M. Buchwald, P. Harper, J. Riordan, In Vivo Cell. Dev. Biol. 21, 592 (1985)]. The presence in these cells of an outwardly rectifying Cl– channel was confirmed [T. T. Bachmann, T. J. Jensen, J. R. Riordan, J. W. Hanrahan, J. Membrane Biol., in press], but the CF cells were insensitive to activation by cyclic AMP [T. J. Jensen, J. W. Hanrahan, A. A. Bacharova, M. Buchwald, J. Riordan, J. Biol. Chem. 263, 6508 (1988)]. Polyanucleotides were isolated from [J. M. Chitwood, A. E. Prebyla, R. L. MacDonald, W. J. Rutter, Biochemistry 18, 5294 (1979)]; H. Aiv and P. Leeder, Proc. Natl. Acad. Sci. U.S.A. 69, 1408 (1972) and used as template for the synthesis of cDNA according to U. Gubler and B. Hoffman [Genet 25, 263 (1983)]. After methylation of internal Eco RI sites, ends were made flush with T4 DNA polymerase, and phosphorylated Eco RI linkers were added to the cDNA. After digestion with Eco RI and removal of excess linkers, the cDNA products were ligated into the Eco RI site of a ZAP vector (Stratagene, San Diego, CA). The same procedures were used to construct a library from RNA isolated from preconfluent cultures of the CF apical chloride conductance pathway (5). Since the predicted structure of CFTR contains several conserved domains and belongs to a family of proteins, most of which function as parts of multicomponent molecular systems (15), the CFTR protein may also participate in epithelial cell functions not related to ion transport.

11. To understand the basic defect in CF, it is necessary to determine the precise role of Phe508 in the regulation of ion transport and to understand the mechanism that leads to the pathophysiology of the disease. With the CF gene (that is, the cDNA) now isolated, it should be possible to elucidate the control of ion transport pathways in epithelial cells in general. Knowledge gained from study of the CF gene product (CFTR), both the normal and mutant forms, will provide a molecular basis for the development of improved means of treatment of the disease.
Identification of the Cystic Fibrosis Gene: Genetic Analysis

BAT-SHEVA KEREM, JOHANNA M. ROMMENS, JANET A. BUCHANAN, DANUTA MARKIEWICZ, TARA K. COX, ARAVINDA CHAKRAVARTI, MANUEL BUCHWALD, LAP-CHEE TSAU

Approximately 70 percent of the mutations in cystic fibrosis patients correspond to a specific deletion of three base pairs, which results in the loss of a phenylalanine residue at amino acid position 508 of the putative product of the cystic fibrosis gene. Extended haplotype data based on DNA markers closely linked to the putative disease gene locus suggest that the remainder of the cystic fibrosis mutant gene pool consists of multiple, different mutations. A small set of these latter mutant alleles (about 8 percent) may confer residual pancreatic exocrine function in a subgroup of patients who are pancreatic sufficient. The ability to detect mutations in the cystic fibrosis gene at the DNA level has important implications for genetic diagnosis.

Although the frequency of cystic fibrosis (CF) is not uniformly high among all Caucasian populations, a consensus estimate is that it occurs once in 2000 live births (1). On the basis of the autosomal recessive mode of inheritance for this disease, a mutant allele frequency of 0.022 may be derived. Several different mechanisms, including high mutation rate (2), heterozygote advantage (3), genetic drift (4), multiple loci (5), and reproductive compensation (6), have been proposed in attempts to explain the high incidence and, indirectly, the nature of the CF mutations. Although some of these hypotheses could not be further addressed because of the lack of knowledge about the basic defect in CF, several important observations have been made during the past few years through genetic analysis of the families of affected individuals (7–20).

Extensive linkage analysis provides evidence for the existence of a single CF locus on human chromosome 7 (region q31) (7–10, 21). The detection of allelic and haplotype association between the CF...