Absence of mutations in the interspecies conserved regions of the CFTR promoter region in cystic fibrosis (CF) and CF related patients

Claudine Verlingue, Sandrine Vuillaumier, Bernard Mercier, Grald Le Gac, Jacques Elion, Claude Frec, Erick Denamur

Abstract

This study was aimed at testing if a 5.2 kb untranslated region on both sides of the first CFTR exon, shown to contain regulatory elements, could carry mutations responsible for cystic fibrosis (CF) or CF related phenotypes. Selection of the DNA segments studied within this region was based upon the identification of conserved sequences throughout evolution (phylogenetic footprints, PFs). Comparison of the CFTR sequences in eight species representing four orders of mammals (man, gibbon, rhesus monkey, squirrel, monkey, rabbit, cow, rat, and mouse) identified four clusters of PFs within the 3.9 kb of DNA sequence upstream from the initiation codon, as well as two nearby PFs at +1 kb within intron 1. Six DNA segments containing PFs were scanned for mutations by denaturing gradient gel electrophoresis (DGGE) in patients with CF (n=29), congenital bilateral absence of the vas deferens (n=143), or disseminated bronchiectasis (n=33), for whom only one or no mutations had been identified despite extensive DGGE analysis of the 27 CFTR exons and exon/intron boundaries. Only one polymorphism (-966 T→G) was identified with a frequency of 2.2% and no other sequence variations were found. This study reinforces the idea that the promoter region in the CFTR is not frequently mutated.

(J Med Genet 1998;35:137-140)

Keywords: CFTR mutations; phylogenetic footprints; CFTR promoter

Cystic fibrosis (CF) is one of the most common genetic diseases in white populations, affecting 1 in 2500 neonates.¹ While the $\Delta F508$ mutation is the major molecular defect, accounting for 20 to 90% of mutations depending on the population,²⁻⁴ over 600 sequence alterations within the coding sequences and the exon/intron junctions have been reported so far (Cystic Fibrosis Genetic Analysis Consortium, personal communication). However, extensive and efficient screening procedures covering the whole CFTR (cystic fibrosis transmembrane conductance regulator) gene (cDNA and splicing sites) have been unsuccessful in identifying mutations in some CF patients. Furthermore, other phenotypes, such as congenital bilateral absence of

the vas deferens (CBAVD) or disseminated bronchiectasis (DB), have been described as being linked to specific mild mutations of the CFTR gene.⁵⁻⁶ In β thalassaemia, the paradigm of most monogenic diseases, some mutations are located in the promoter region of the β globin gene or far upstream in the locus control region.⁷ Thus, one would expect some CF mutations to be present within CFTR regulatory regions. The CFTR gene is tightly regulated and mainly expressed in epithelial cells. The molecular mechanisms leading to this regulation are largely unknown but some recent investigations have identified potential elements involved in transcriptional regulation.⁸⁻¹⁴

In this study, we have screened for mutations the sequences that are conserved between mammals in 5.2 kb of untranslated DNA on both sides of the first CFTR exon. The region we studied spans over 3.9 kb of upstream sequences, including the CFTR promoter, down to 1.3 kb in the first intron. It has been shown previously to contain potential regulatory elements.^{8-12, 14} The CFTR cDNA is highly conserved in a wide range of vertebrate species¹⁵ and we postulated that homologous sequences between the promoter regions of various mammals correspond to sequences under strong positive selective pressure because of a specific functional role. Search for sequences that have been conserved throughout evolution in promoter regions has been performed in other systems and referred to as phylogenetic footprinting. A good correlation was found between phylogenetic footprints (PFs) and binding sites for nuclear proteins in the β globin gene cluster.¹⁶ By using this phylogenetic approach, we selected six DNA segments, containing 20 PFs altogether, which we screened for possible mutations in patients with various CF and CF related phenotypes (CBAVD and DB).

Patients and methods

PATIENTS

Three groups of patients were included in this study. The first was a cohort of 29 typical CF patients for whom only one CFTR mutation had been characterised, the other allele remaining unidentified even though the CFTR gene was extensively analysed by denaturing gradient gel electrophoresis (DGGE scanning) of the 27 exons and of the intron/exon boundaries. The second group was a cohort of 143 CBAVD patients. Among them, 42 were

ETSBO-CHU-UBO,
46 rue Flix Le Dantec,
BP 454, 29275 Brest
Cedex, France
C Verlingue
B Mercier
G Le Gac
C Frec

Biochimie
Gntique/INSERM
U458, Hpital Robert
Debr, 48 Boulevard
Srurier, 75019 Paris,
France
S Vuillaumier
J Elion
E Denamur

Correspondence to:
Dr Frec

Received 28 May 1997
Revised version accepted for
publication 13 August 1997

identified carriers of one mutant CFTR allele including the 5T variant; for the remaining 101 patients, no CF mutation was found by DGGE scanning. The third group was a cohort of 33 patients with bronchiectasis and a borderline sweat test; 11 of them were carriers of one CFTR mutation.

DNA SEQUENCES

A 5.2 kb region containing the promoter region of the CFTR gene and part of the first intron (3.9 kb upstream and 1.3 kb downstream from the first exon) was analysed for nucleotide sequence similarities between human, non-human primate, cow, rabbit, and rodent sequences. Sequence comparisons were performed using the CLUSTAL V multiple alignment program.¹⁷ The human promoter region sequence was obtained from GenBank (accession number M58478) and the intron 1 sequence was completed by sequencing the pDAN1 plasmid clone (a kind gift from Dr A Harris, Institute of Molecular Medicine, Oxford, UK) using the primer DEN2R (5'-GAAGTTATATCCAACCTGAC-3'). The mouse sequence was kindly provided by Dr R Rozmahel and Dr L C Tsui (Hospital for Sick Children, Toronto, Canada). The remaining sequences have been published previously¹⁵ and are available under the following GenBank accession numbers: X95930 (gibbon), X95929 (rhesus monkey), X95928 (squirrel monkey), X95931 (rabbit), X95926 (cow), and X95927 and Y10147 (rat). For all the compared sequences, nucleotide numbering was related to the human ATG initiation codon.

DGGE

Oligonucleotides for amplification of the interspecies conserved regions are: 3855-5': 5'-GGTTGTAGGGAAGAGGGGTT-3'; 3720-3' GC: 5'-CGCCCGCCGCGCCCGCCCGCGCCG CCGTCCCGCCGCCCCCGCCCGGGC CACAGTGATTACCTAGGAT-3'; 3650-5' GC: 5'-CGCCCGCCGCGCCCGCCCGCG CCGCGTCCCGCCCGCCCGCCCGACTGT GGCCCACTGT TGAAGAGCTG-3'; 3281-3': 5'-ATATATGAAAGATTATAAGTGATA ACC-3'; 1440-5': 5'-CATCATGAAATGCA CATTATTATTG-3'; 1200-3' GC: 5'-CGCCC GCGCGCCCCGCGCCCGTCCCGCCG CCCCCGCCCCCTTTTGTGTAGGATTGC CTGTGTCTC-3'; 1200-5' GC: 5'-CGCCCG CCGCGCCCCGCGCCCGTCCCGCCCC CGCCCCGCCTTCTCAGAAGGAAGGCG CCTACGC-3'; 881-3': 5'-GTAAGAGGAG ATAATGCTTTG-3'; 240-5' GC: 5'-CGCC CGCCGCGCCCCGCGCCCGTCCCGCC GCCCCGCCCCGCAATTTGGGGCCG GACCAG-3'; 60-3': 5'-AGACCTACTACT CTGGGTG-3'; 990-5': 5'-GCTCATATAT GTGTGTAGGG-3'; 1180-3' GC: 5'-CGC CCGCCGCGCCCCGCGCCCGTCCCGC CGCCCCGCCCCCTAACATCAACACAAA AAGAT-3'.

Conditions for the DGGE analyses have been described elsewhere.^{4 5 18 19} Amplification products were run on 6.5% polyacrylamide gels containing a linear gradient from 20% to 70% of denaturant, where 100% denaturant

Table 1 Sequences and positions of the phylogenetic footprints (PFs) within a region extending from 2.5 kb upstream from the ATG initiation codon to 1.3 kb downstream from the first CFTR exon

	Sequence 5' to 3'	Position*
PF +1	AGGAGA	+1110 to +1116
PF +2	AGAAGA	+1138 to +1143
PF -1	AAAAGGG	-97 to -91
PF -2	TGACATCAC	-119 to -111
PF -3	TGGGTGGGGG	-143 to -134
PF -4	GGTGTG	-160 to -155
PF -5	AACCTGGGC	-203 to -195
PF -6†	AAGGAGCG	-295 to -288
PF -7	AGTTTA	-1012 to -1007
PF -8	TTTGCT	-1053 to -1047
PF -9	GCTTCA	-1098 to -1093

*Nucleotide numbering refers to the human ATG initiation codon.

†PF -6 was not analysed.

equals 7 mol/l urea with 40% formamide (v/v). Electrophoresis was performed at 75 V for seven to nine hours, depending upon the melting properties of the amplified sequence. PCR fragments displaying altered behaviour in the gel, indicative of the presence of a mutation or polymorphism in this particular region of the gene, were subsequently sequenced.

DNA SEQUENCING

Single stranded DNA was obtained in an asymmetrical PCR reaction. Each single stranded product was concentrated using a Centricon 100 (Amicon, Danver, Mass, USA) and sequenced by the dideoxynucleotide chain termination method.

Results

SELECTION OF THE DNA SEGMENTS STUDIED

CFTR sequences of eight species representing four orders of mammals (man, gibbon, rhesus monkey, squirrel monkey, rabbit, cow, rat, and mouse) were compared over a region spanning from 2.5 kb upstream from the human ATG initiation codon to 1.3 kb downstream from the first exon. A phylogenetic footprint (PF) is defined as a non-coding sequence motif that shows 100% conservation in several species over a region of ≥ 6 contiguous base pairs.¹⁶ Our analysis allowed the identification of 11 PFs (table 1, fig 1). Two of them (PFs +1 and +2) are located at +1 kb in the first intron and are separated by only 21 bp. Five PFs (-1 to -5) are clustered in a 116 bp region, 95 bp upstream from the ATG. This region has been defined as the minimal promoter, both in human and mouse, by transfection assays with reporter genes.⁸⁻¹¹ PF -6 is located at position -288. Three PFs (-7 to -9) cluster in a 100 bp region 1.3 kb upstream from the ATG. Lastly, between nucleotides (nts) -1246 and -1236, we observed a region in which eight out of 11 nts are conserved between all the species studied. However, since this region does not contain 6 contiguous bp, it cannot be considered as a PF, and we will refer to it as a "region of similitude" (RS).

Comparison was extended further upstream for the three species for which sequence data are available, that is, human, rat, and mouse (fig 2). No similitude was observed between the

Thus, four clusters of PFs and one RS are observed within the 3.9 kb of DNA sequence upstream from the initiation codon of CFTR. In addition, two nearby PFs are located at +1 kb in intron 1 (fig 1). The length of these PFs varies from 6 to 12 nts (table 1, fig 2). From these data, six DNA segments ranging from 125 to 370 bp were selected to be studied by DGGE: segment A (nts -3855 to -3720), segment B (nts -3650 to -3281), segment C (nts -1440 to -1200), segment D (nts -1200 to -881), segment E (nts -240 to -60), and segment F (nts +990 to 1180) (fig 1). PF -6 was not analysed.

Altogether, 366 chromosomes were studied by DGGE analysis of the six DNA segments described above. We observed no variations in the pattern of migration of any of the corresponding PCR products except for fragment D where a T→G nucleotide change was identified at position -966 from the ATG initiation codon. This change does not fall into a PF and has not been reported so far. Familial segregation clearly indicates that this nucleotide change is a polymorphism and it has been observed in the homozygous state in some parents of affected patients. Moreover, in two families, the father was a carrier of the -966 T→G and the variant was not transmitted to his affected CF child. The overall frequency of

this polymorphism was around 2.2% in the sample studied.

Discussion

To look for mutations within the CFTR promoter region, we used a targeted approach based on evolutionary clues. Recently, phylogenetic footprinting has been successful in identifying a cAMP responsive element (CRE) and a 12-O-tetra decanoylphorbol-13 acetate (TPA) responsive element (TRE) within the CFTR promoter. These two elements are located at -0.1 and -1.3 kb relative to the ATG, respectively (that is, in the PF -7/-9 and PF -1/-5 clusters). In vitro analysis of protein/DNA interactions and ex vivo transfection assays have confirmed the functional relevance of these two conserved sequences in the transcriptional regulation of CFTR.^{12 14} Once these potential regulatory elements were identified, we used a GC clamp DGGE assay to scan these regions for possible mutations in patients with various CF related phenotypes. Based upon our previous results, this DGGE approach should be the most effective in identifying sequence variations in the region studied, even if this region is particularly GC rich. Except for clear polymorphisms found in CF and non-CF subjects, the assignment of a base change in a non-coding sequence as a molecular defect is difficult to assess²⁰ (CF Genetic Analysis Consortium, personal communication). Actually, we have been surprised by the very low level of polymorphism exhibited by the sequences we tested. Indeed, only one polymorphism (-966 T→G) was identified and we did not find any other sequence variations. We have looked at the distribution of the -966 T→G variant in CF, CBAVD, or disseminated bronchiectasis patients and the percentage of -966 T→G carriers is the same in these three groups, so the chance that this change could be a mild mutation, like, for example, the intron 8 5T variant in CBAVD patients, is very low.

This may be a sign that these regions containing PF are under selective pressure even for sequences that surround the PFs. Among the 600 mutations reported so far, only a very few nucleotide changes have been reported in the promoter region²⁰ (CF Genetic Analysis Consortium, personal communication). In this study, DNA of the CF or CBAVD patients had been previously thoroughly studied for mutations throughout the whole coding sequence of the CFTR gene. We have previously shown that DGGE scanning can detect up to 100% of the mutations in a PCR fragment.^{21 22} The fact that we have been unable to find any abnormalities reinforces the idea that the promoter region in the CFTR is not frequently mutated. However, our approach could have missed some important sites involved in the regulation of the human CFTR gene because, firstly, the consensus regulatory sequences are usually relatively short and they may vary by one or two nucleotides between species and, secondly,

species specific regulatory elements seem to exist.¹² Conversely, no functional studies have been performed for regions encompassing PFs -10 to -20 and PFs +1 and +2 and this prevents definite conclusions being drawn concerning their role in CFTR regulation. Still, we think that this approach should certainly be pursued and extended, in particular to the CFTR regulatory element newly described as the 181 +10 kb element.¹³

We thank Pierre Lehn and Ann Harris for fruitful discussion and encouragement, and Bernard Kaltenboeck for computer analysis of the nucleotide sequences. This work was supported in part by the Association Française de Lutte contre la Mucoviscidose (AFLM) and the Contrat de Recherche INSERM No 4U007B.

- 1 Welsh MJ, Tsui LC, Boat TF, Beaudet AL. Cystic fibrosis. In: Scriver CR, Beaudet AL, Sly WS, Valle D, eds. *The metabolic and molecular bases of inherited diseases*. New York: McGraw-Hill, 1995:3799-876.
- 2 The Cystic Fibrosis Genetic Analysis Consortium. Population variation of common cystic fibrosis mutations. *Hum Mutat* 1994;4:167-77.
- 3 Zielenski J, Tsui LC. Cystic fibrosis: genotypic and phenotypic variations. *Annu Rev Genet* 1995;29:777-807.
- 4 Messaoud T, Verlingue C, Denamur E, et al. Distribution of CFTR mutations in cystic fibrosis patients of Tunisian origin: identification of two novel mutations. *Eur J Hum Genet* 1996;4:20-4.
- 5 Chillon M, Casals T, Mercier B, et al. Mutations in the cystic fibrosis gene in patients with congenital absence of the vas deferens. *N Engl J Med* 1995;332:1475-80.
- 6 Pignatti PF, Bombieri C, Marigo C, Benetazzo M, Luisetti M. Increased incidence of cystic fibrosis gene mutations in adults with disseminated bronchiectasis. *Hum Mol Genet* 1994;4:635-40.
- 7 Crossley M, Orkin SH. Regulation of the β -globin locus. *Curr Opin Genet Dev* 1993;3:232-7.
- 8 Yoshimura K, Nakamura H, Trapnell BC, et al. The cystic fibrosis gene has a "housekeeping"-type promoter and is expressed at low levels in cells of epithelial origin. *J Biol Chem* 1991;266:9140-4.
- 9 Chou JL, Rozmahel R, Tsui LC. Characterization of the promoter region of the cystic fibrosis transmembrane conductance regulator gene. *J Biol Chem* 1991;266:24471-6.
- 10 Koh J, Sferri TJ, Collins FS. Characterization of the cystic fibrosis transmembrane conductance regulator promoter region. *J Biol Chem* 1993;268:15912-21.
- 11 Denamur E, Chehab FF. Analysis of the mouse and rat promoter regions. *Hum Mol Genet* 1994;3:1089-94.
- 12 Vuillaumier S, Dixmeras I, Messai H, et al. Cross-species characterisation of the CFTR promoter region reveals multiple levels of regulation. *Biochem J* (in press).
- 13 Smith AN, Barth ML, McDowell TL, et al. A regulatory element in intron 1 of the cystic fibrosis transmembrane conductance regulator gene. *J Biol Chem* 1996;271:9947-54.
- 14 Matthews RP, McKnight GS. Characterization of the cAMP response element of the cystic fibrosis transmembrane conductance regulator gene promoter. *J Biol Chem* 1996;271:31869-77.
- 15 Vuillaumier S, Kaltenboeck B, Lecointre G, Lehn P, Denamur E. Phylogenetic analysis of cystic fibrosis transmembrane conductance regulator in mammalian species argues for the development of a rabbit model for cystic fibrosis. *Mol Biol Evol* 1997;14:372-80.
- 16 Gumucio DL, Heilstedt-Williamson H, Gray TA, et al. Phylogenetic footprinting reveals a nuclear protein which binds to silencer sequences in the human γ and ϵ globin genes. *Mol Cell Biol* 1992;12:4919-29.
- 17 Higgins DG, Sharp PM. Clustal: a package for performing multiple sequence alignments on a microcomputer. *Gene* 1988;73:237-44.
- 18 Lerman LS, Silverstein K. Computational simulation of DNA melting and its application to denaturing gradient gel electrophoresis. *Methods Enzymol* 1987;155:482-501.
- 19 Myers RM, Maniatis T, Lerman LS. Detection and localization of single base changes by denaturing gradient gel electrophoresis. *Methods Enzymol* 1987;155:501-27.
- 20 Bienvenu T, Lacroque V, Raymondjean M, et al. Three novel sequence variations in the 5' upstream region of the cystic fibrosis transmembrane conductance regulator (CFTR) gene: two polymorphisms and one putative molecular defect. *Hum Genet* 1995;95:698-702.
- 21 Férec C, Audrézet MP, Mercier B, et al. Detection of over 98% cystic fibrosis mutations in a Celtic population. *Nat Genet* 1992;1:188-91.
- 22 Macek M Jr, Mercier B, Mackova A, et al. Sensitivity of the denaturing gradient gel electrophoresis technique in detection of known mutations and novel Asian mutations in the CFTR gene. *Hum Mutat* 1997;9:136-47.