Cross-species characterization of the promoter region of the cystic fibrosis transmembrane conductance regulator gene reveals multiple levels of regulation

Sandrine VUILLAUMIER*, Isabelle DIXMERAS*, Habib MESSAÏ*, Claudine LAPOUMÉROULIE*, Dominique LALLEMAND†, Jean GEKAS*, Farid F. CHEHAB‡, Christine PERRET§, Jacques ELION* and Erick DENAMUR*[1]

*INSERM U 458, Hôpital Robert Debré, 48 boulevard Sérurier, 75019 Paris, France, †Unité des Virus Oncogènes, UA 1149 du CNRS, Département des Biotechnologies, Institut Pasteur, 25 rue du Dr. Roux, 75724 Paris cedex 15, France, ‡Department of Laboratory Medicine, University of California, San Franscico, CA 94143-0134, U.S.A. and §INSERM U 129, 24 rue du Faubourg Saint Jacques, 75014 Paris, France

The cystic fibrosis transmembrane conductance regulator (*CFTR*) gene is highly conserved within vertebrate species. Its pattern of expression *in vivo* seems to be tightly regulated both developmentally and in a tissue-specific manner, but shows differences with species. To identify transcriptional regulatory elements in the *CFTR* promoter region, we have used a combined approach based both on the analysis of the chromatin structure *in vivo* in rat tissues and on evolutionary clues (i.e. phylogenetic footprinting). In *CFTR*-expressing tissues, 15 DNase I-hypersensitive sites were identified within a 36 kb region encompassing exon 1. Eleven of them are clustered in a 3.5 kb region that exhibits eleven phylogenetic footprints observed when comparing sequences from eight mammalian species representing four orders (Primates, Artiodactylia, Lagomorpha and Rodentia). Comparison of the two sets of data allows the identification of two types of regulatory elements. Some are conserved between species, such as a non-consensus cAMP response element (CRE) and a PMA-responsive element (TRE) located respectively at positions −0.1 and −1.3 kb relative to ATG. Some are species-specific elements such as a 300 bp purine·pyrimidine (Pu·Py) stretch that is present only in rodents. Analysis of protein/DNA interactions *in vitro* with rat tissue protein extracts on the conserved elements revealed that the TRE site binds a specific heterodimeric complex composed of Fra-2, Jun D and a protein immunologically related to Jun/CRE-binding protein in the duodenum, whereas the CRE-like site binds ATF-1 ubiquitously. Functional analysis in Caco-2 cells showed that the CRE-like site supports a high basal transcriptional activity but is not able by itself to induce a response to cAMP, whereas the TRE site acts as a weak transactivator stimulated by PMA. Lastly, we found that the rodent-specific Pu·Py stretch confers nuclease S1 hypersensitivity under conditions of acidic pH and supercoiling. This indicates a non-B DNA conformation and thus reinforces the biological significance of non-random Pu·Py strand asymmetry in the regulation of transcription. Thus the tight transcriptional regulation of *CFTR* expression involves the combination of multiple regulatory elements that act in the chromatin environment *in vivo*. Some of them are conserved throughout evolution, such as the CRE-like element, which is clearly involved in the basal level of transcription; others are species-specific.

INTRODUCTION

Regulation of expression of the cystic fibrosis transmembrane conductance regulator (*CFTR*) gene, which is altered in cystic fibrosis [1], is complex. Both CFTR mRNA and protein are highly conserved within a wide range of vertebrate species representing evolutionary distances of up to 420 million years. They show patterns of expression *in vivo* that seem to be tightly regulated both developmentally [2,3] and in a tissue-specific manner [4,5], but with differences depending on the species. Expression of *CFTR* is mainly restricted to epithelial cells in the lung, intestine, pancreas, gall bladder, kidney, salivary and sweat glands, testies and uterus. In each of these tissues only a subpopulation of specified cells is involved in *CFTR* expression. For example, in the human respiratory tract, the predominant site of epxression is a subpopulation of cells in submucosal glands [6]. In the rat intestine a decreasing gradient of expression of *CFTR* is observed both on the crypt–villus and the proximal–distal axes [5]. Furthermore *CFTR* expression is regulated in rat both during the cycle of the seminiferous epithelium and the oestrous cycle [7]. In human cell lines it is modulated by cAMP [8], PMA [9,10] and divalent cations [11].

DNA sequence analysis of the *CFTR* promoter region in human [12] and rodents [13] has revealed a high GC content, no TATA box and putative Sp1- and AP-1-binding sites. In human, multiple transcription start sites including a major one have been described, but the precise location of these sites is not consistent from one study to another [12,14,15]. Experiments with reporter genes and transient transfection assays in *CFTR*-expressing and non-expressing cell lines in both human [12,14,15] and mouse [13] have identified the basal *CFTR* promoter as a 250 bp fragment upstream of the ATG translation start codon driving a weak expression. However, no sequences that might account for tissue specificity have yet been found and published data about the precise dissection of the promoter region are not consistent. Promising data have been obtained from DNase I-hypersensitive site (DHS) analysis as it explores chromatin structure *in vivo* and thus is more relevant to a physiological process. In human cell lines, several groups have identified DHSs 5′ to *CFTR* or within the first intron, and some of them show a correlation with *CFTR*

Abbreviations used: ATF, activating transcription factor; *CAT*, chloramphenicol acetyltransferase gene; *CFTR*, cystic fibrosis transmembrane conductance regulator gene; CRE, cAMP response element; CREB, CRE-binding protein; DHS, DNase I-hypersensitive site; PF, phylogenetic footprint; Pu·Py, purine·pyrimidine; TRE, PMA-responsive element PMR; Pu·Py mirror repeat element.

[1] To whom correspondence should be addressed.

expression [15–18]. Using this approach, a regulatory element has been localized at 181 + 10 kb within the human first intron [18]. These studies have, however, been essentially performed on long-term cell lines, which might not adequately reflect the regulation of *CFTR* expression *in vivo*. Discrepancies in the methylation status of CpG sites in the mouse and human *CFTR* promoters have been demonstrated between long-term cell lines and tissues, and are attributed to methylation *de novo* during the process of immortalization of the cell lines [19].

Because of these somewhat confusing data, we decided to investigate the bases of the transcriptional regulation of *CFTR* by a combined approach including DNase I hypersensitivity assays in tissues and phylogenetic footprinting. Most DHSs are found only in chromatin of cells in which the associated gene is being expressed and are related to gene expression [20]. Phylogenetic reconstructions of the *CFTR* promoter region and cDNA sequences among mammalian species have yielded two deeply separated groups (man, non-human primates, cow, and rabbit; and rodents) correlated with the pattern of *CFTR* expression [21]. Nevertheless we chose to perform the DNase I hypersensitivity assays on nuclei isolated from rat tissues for the following reasons. First, rat tissues are easily available; secondly, the observed differences between rodents and other mammalian species in *CFTR* regulation could themselves be interesting; and thirdly, the characterization of the rat promoter will permit, in a second step, the production of transgenic mice with rat promoter sequences linked to a reporter gene. This should eliminate artifacts that could be generated in transgenic experiments when a promoter sequence belonging to the human group is studied within a rodent background. Phylogenetic footprinting is an evolutionary approach that postulates that, although species-specific patterns of *CFTR* expression exist, the basic mechanisms of *CFTR* transcriptional regulation are conserved throughout the mammalian phylogeny. Phylogenetic footprints are defined as non-coding sequence motifs that show 100 % conservation in several species over a region of six or more contiguous base pairs. A high correlation was found between the presence of a phylogenetic footprint and the binding of nuclear proteins in the β-globin gene cluster [22]. With this combined approach, four regions corresponding to conserved DNA sequences and exhibiting tissue-specific DNase I hypersensitivity in the chromatin structure were identified within the *CFTR* promoter region. Two of them, at -1.3 and -0.1 kb relative to the initiation codon, were characterized and correspond to a PMA-responsive element (TRE) and a non-consensus cAMP responsive element (CRE) respectively. These DNA motifs are the ultimate targets for transcriptional control of the two major signal transduction pathways that use diacylglycerol and inositol trisphosphate, and cAMP, as secondary messengers respectively. TRE sequences are usually binding sites for the heterodimeric transcription factor AP-1, whereas CRE sequences are recognized by the members of a large family known as the CRE-binding (CREB)/activating transcription factor (ATF) proteins. Both groups of proteins belong to the class of leucine zipper (bZip) transcription factors [23–25]. Additional DHSs have been mapped in regions that are not conserved between species and could correspond to elements involved in *CFTR* regulation in the rodents specifically. Further analysis of two of these sites suggests a physiological significance of regions of non-random purine·pyrimidine (Pu·Py) strand asymmetry in the regulation of *CFTR* expression.

MATERIALS AND METHODS

Analysis of DHSs

The search for DHSs within chromatin isolated from rat tissues (duodenal mucosa, liver, kidney, lung) was realized with the indirect end-labelling technique of Wu [26]. Nuclei isolation, DNase I digestion, DNA isolation and Southern blots were performed as previously described [27]. The probes were generated by PCR and labelled by random priming with [α-^{32}P]dCTP to a specific radioactivity of at least 10^8 c.p.m./μg. To design primers, nucleotide sequences were obtained by direct sequencing of two overlapping phage clones isolated from a rat genomic library constructed in the EMBL3 SP6/T7 vector (Clontech, Palo Alto, CA, U.S.A.) [21]. As positive controls, the presence of known tissue-specific DHSs (duodenum- and liver-specific) in the *CaBP9k* gene [27] was checked on all the blots tested. For an accurate estimation of the molecular mass of hybridized DNA fragments, DNA size markers [λ-DNA cleaved with *Eco*RI and *Hin*dIII or SPP1 DNA cleaved with *Eco*RI (Boehringer, Mannheim, Germany)] were loaded with each DNA sample in the same well and the blots were re-hybridized with ^{32}P-labelled λ and SPP1 DNA species as above.

Computer analysis of the DNA sequences

Multiple nucleotide sequence alignments of a 3.5 kb region encompassing *CFTR* exon 1 (2.3 kb upstream of the ATG initiation start codon and 1.2 kb of the first intron) in human, gibbon, cynomolgus, squirrel monkey, cow, rabbit, mouse and rat [21] were performed with the Clustal V program [28]. Search for transcription factor recognition sequences was performed on the rat sequence with the SIGMUC program [29].

DNase I footprinting *in vitro*

Nuclear extract preparation from rat tissues and cell lines was performed as described by Lambert et al. [30]. DNA fragments used as probes were generated by PCR under the previously described conditions [13], subcloned in the PCR® II vector (InVitrogen, SanDiego, CA, U.S.A.) and end-labelled on either strand with the Klenow fragment of DNA polymerase 1 and [α-^{32}P]dATP and dGTP. Footprinting experiments were performed as follows. Nuclear proteins (40 μg) were preincubated for 15 min on ice in 16 μl of binding buffer containing 250 ng of poly(dI-dC)·poly(dI-dC) and 50000 c.p.m. of the labelled probe. Samples were subjected to limited DNase I digestion, and the DNA was purified and resolved by denaturing 6 % (w/v) PAGE and autoradiography. The chemical G+A and C+T sequencing reactions were performed by the Maxam–Gilbert method.

Electrophoretic mobility-shift assays

Complementary oligonucleotides were 5′ end-labelled with T4 polynucleotide kinase and [γ-^{32}P]ATP. End-labelled double-stranded oligonucleotide (10000 c.p.m.) was incubated for 10 min at 4 °C in 20 μl of storage buffer with 5–10 μg of nuclear extracts, 1 μg of poly(dI-dC)·poly(dI-dC), in the presence or absence of double-stranded competitor oligonucleotide. Samples were loaded on a 6 % (w/v) polyacrylamide gel, electrophoresed and analysed by autoradiography. In some experiments, antibodies raised against known transcription factors were pre-incubated for 1 h at 4 °C with the proteins before the incubation with the labelled probe to supershift or abolish the protein–DNA complex. In these cases, samples were loaded on 4 % (w/v) polyacrylamide gels. Tested sequences were as follows: GS9, TTGCTTCAGTGACTCAGAGTCCG; GS9/1M, TTGCTgggGTGACTCAGAGTCCG; GS9/2M, TTGCTTCAGTGACgggGAGTCCG; GS9/1+2M, TTGCTgggGTGACgggGAGTCCG; GS1, TGCAAATGACATCACCTCAGGTCTGAGTAAAAGGGACGAGC; GS1M, TGCAAAcatCAcCACCTCA-

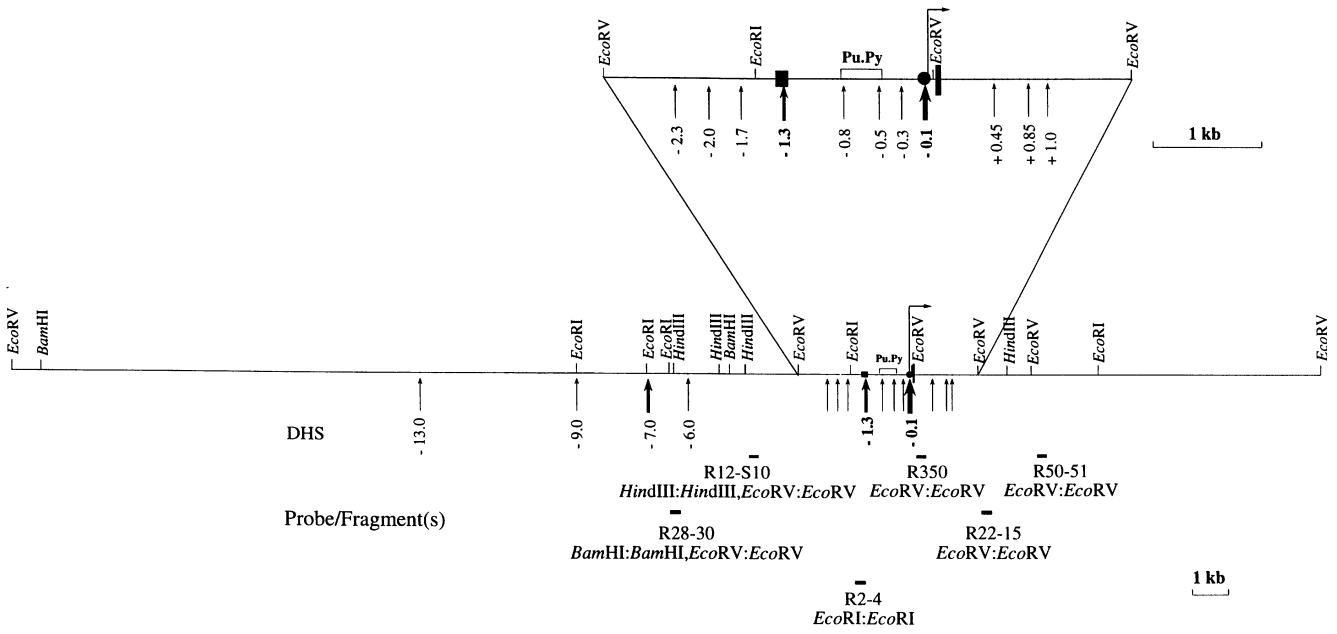

Figure 1 Restriction map of the studied 36 kb rat *CFTR* promoter region and localization of the DHSs

The primary map is drawn to scale and the region containing the 11 DHSs clustered within a 3.5 kb region encompassing the first *CFTR* exon is shown as a magnified segment above the primary map. The DHSs are indicated by a thin or thick arrows according to the strength of each site. The upper arrow indicates the transcription start; the vertical bar represents the first exon. The positions of the DHSs are indicated in kilobases from the ATG initiation codon. Positions of the probes are indicated in the bottom part of the figure together with the fragments tested with each probe. The purine·pyrimidine stretch is shown as Pu·Py. The symbols ■ and ● indicate the TRE and CRE sites within DHS −1.3 and the DHS −0.1 respectively.

GGTCTGAGTAAAAGGGACGAGC; GS2, CCCGAGTGG-GTGGGGGGAATTGGAAGCAAATGACATCACCTCAG; GS2M, CCCGAGTGGGTGGGGGGAATTGGAAGCAAA catCAcCACCTCAG; CREcf, TAGCAAATGACATCACCT-CAGG; CAATicf, TGGGAATTGGAAG. The underlined sites correspond to phylogenetic footprints and/or consensus sites for transcription factors; lower-case letters indicate mutated bases (see below).

S1 hypersensitivity assay

Either the P1–P2 or the PL construct plasmid DNA (5 μg) [13] was precipitated with ethanol. The P1–P2 construct plasmid encompasses mouse sequence from nt −1122 to +15 relative to the initiation codon, a region that includes the rodent-specific Pu·Py stretch, whereas PL is the control plasmid with no *CFTR* sequences [13]. The pellet was equilibrated overnight at 4 °C in 22 μl of 40 mM sodium acetate (pH 4.6)/50 mM NaCl/1 mM $ZnCl_2$. After preincubation for 10 min at 37 °C, 0.5–2 units of S1 nuclease (Promega, Madison, WI, U.S.A.) were added and the incubation was prolonged for 1 h. DNA was purified with a microcolumn (Promega) and recovered in 50 μl of sterile distilled water; 16 μl was then treated with *Nco*I, a restriction enzyme that cleaves uniquely in the vector and not the insert.

Cell culture, plasmid constructs and plasmid-mediated DNA transfer

Caco-2 human colon adenocarcinoma cells and NIH-3T3 murine fibroblast cells were maintained as described [19]. The RV-2.0 construct consists of the mouse minimal promoter sequence (nt −219 to +15 according to the ATG initiation start) linked upstream of a promoterless chloramphenicol acetyltransferase (*CAT*) gene [13]. Plasmid PL served as a background control [13]. Plasmid pBLCAT2, which contains the Herpes simplex virus thymidine kinase gene (*tk*) promoter fused to the *CAT* gene [31] was used to subclone oligonucleotides. The oligonucleotides were synthesized with a *Hin*dIII and a *Bam*HI site at each end to allow their cloning at the *Hin*dIII and *Bam*HI sites of the pBLCAT2 polylinker. pBCAAT, pBCRE and pBCAAT-CRE plasmids contain respectively the inverted CAAT box, or the CRE site, or the two elements together. pBTRE and pBTREM plasmids contain respectively the wild-type and the mutated TRE sequences corresponding to the GS9 and GS9/1+2M probes. In all cases the insert sequences were verified by DNA sequencing. Plasmid DNA was isolated and purified by the Qiagen plasmid purification procedure (Qiagen, Chatsworth, CA, U.S.A.).

Transfections were performed with the lipofectamine protocol (Gibco/BRL, Gaithersburg, MD, U.S.A.). For the experiments testing the whole *CFTR* promoter, which is known to drive a weak expression, 7 μg of a *CAT* construct and 0.7 μg of pCH110, a control plasmid containing a β-galactosidase cDNA driven by the SV40 promoter enhancer, were co-transfected into Caco-2 cells grown to 80% confluence on a 60 mm plate. For the experiments with the *tk* promoter, 2.5 μg of a CAT construct and 0.35 μg of pCH110 were transfected as above but on 35 mm plates. Transfections were done in duplicate and each set of transfections was done at least three times. Cells were harvested 28 h after transfection. cAMP and diacylglycerol transduction pathways were stimulated by adding forskolin or PMA to the medium 20 h after the transfection, at final concentrations of 10 μM and 100 nM respectively. Cells were harvested 8–10 h after the stimulation. Protein extracts, CAT and β-galactosidase assays were performed as described [13]. CAT activities were then normalized to that of the β-galactosidase to account for

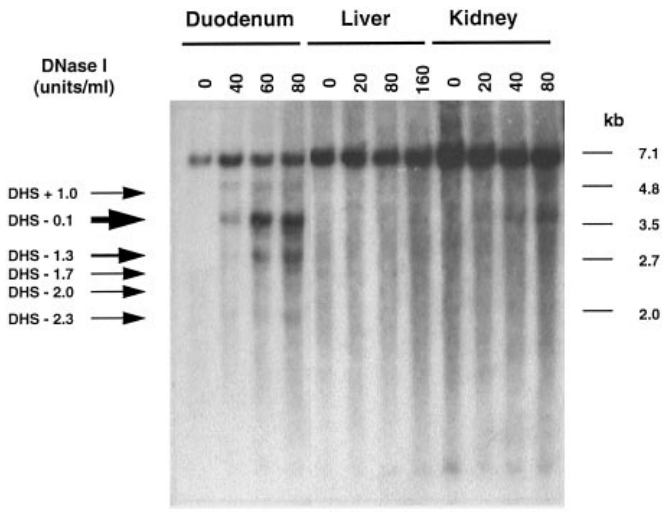

Figure 2 DHSs in the *CFTR* promoter region

An example of a Southern blot corresponding to a *Hin*dIII digest hybridized with the R12-S10 probe is shown. The amount of DNase I is indicated at the top of the blot and ranges from 0 (absence of DNase I) to 160 units/ml. Numbers at the right indicate the migration of DNA size standards in kb. The arrows indicate the DHS fragments. DHSs −0.8, −0.5 and +0.45 are not indicated. The 0.8% agarose gel used did not allow us to resolve DHS fragments −0.3 and +0.85, which co-migrated with DHS fragments −0.1 and +1.0 respectively. A better resolution of these fragments was obtained by using *Eco*RV and *Eco*RI digest hybridized with R350 and R2-4 probes respectively (Table 1).

transfection efficiencies. For the experiments with PMA stimulation, normalization was made on the amount of proteins, as pCH110 is sensitive to PMA.

RESULTS

Fifteen DHSs are identified in *CFTR*-expressing tissues and eleven of them map within a 3.5 kb region encompassing the first *CFTR* exon

To identify DHSs, nuclei from rat tissues (duodenum mucosa, lung, liver and kidney) were incubated at 0 °C with increasing amounts of DNase I (0–320 units/ml). The DNA was then isolated and resolved by Southern blot analysis with the restriction endonucleases *Bam*HI, *Eco*RI, *Eco*RV and *Hin*dIII with the probes R28-30, R12-S10, R2-4, R350, R22-15 and R50-51 (Figure 1). In rodents, *CFTR* is expressed at high level in the duodenum mucosa, moderate level in the lung and kidney, and low level in the liver [5,32]. Fifteen DHSs have been identified within the 36 kb studied region; eleven of them are clustered within a 3.5 kb region encompassing exon 1 of *CFTR* (Figures 1 and 2, and Table 1). One major site (DHS-0.1) is detected in duodenum, lung and kidney but not in liver nuclei and maps to the minimal promoter. All the remaining DHSs are detected only in the duodenum nuclei. Two are of strong intensity and map at 7.0 and 1.3 kb upstream of the ATG (DHSs −7.0 and −1.3) (Figure 2). The last 12 DHSs are relatively weak. Three map in the first intron at positions 0.45 (DHS+0.45), 0.85 (DHS+0.85) and 1 kb (DHS+1.0), whereas the others are localized at positions −0.3 (DHS −0.3), −0.5 (DHS −0.5), −0.8 (DHS −0.8), −1.7 (DHS −1.7), −2.0 (DHS −2.0), −2.3 (DHS −2.3), −6.0 (DHS −6.0), −9.0 (DHS −9.0) and −13.0 kb (DHS −13.0) relative to ATG. DHSs −0.5 and −0.8 map in the Pu·Py stretch that is specific for the rodents [13] (Figure 1 and Table 1). It should be noted that the previously described human DHS at position 181+10 kb [18] has no counterpart within the rat first intron.

Table 1 List of the identified DHS fragments with the probes used for their recognition and the tissues in which they were detected

The DHS fragments are delineated by a restriction enzyme site and a DHS. They are designated by the name of the enzyme and the name of the DHS (see Figure 1) separated by a colon, the first and second items representing the 5′ and 3′ ends of the DHS fragment respectively. The presence of the DHS fragments is indicated by a + sign; its absence is indicated by a zero. The number of + signs corresponds to the intensity of the DHS fragment ranging from major (+++) through strong (++) to weak (+). Abbreviation: n.d., not determined.

DHS fragments	Probes	Duodenum	Liver	Kidney	Lung
*Eco*RV: +1.0	R350	+	0	0	n.d.
*Eco*RV: +0.85	R350	+	0	0	n.d.
*Eco*RV: +0.45	R350	+	0	0	n.d.
*Eco*RI: +0.85, +1.0*	R2-4	+	0	0	0
*Eco*RI: +0.45	R2-4	+	0	0	0
*Eco*RI: −0.1	R2-4	+++	0	+	+
*Eco*RI: −0.3	R2-4	+	0	0	0
*Eco*RI: −0.5	R2-4	+	0	0	0
*Hin*dIII: −2.3	R12-S10	+	0	0	0
*Hin*dIII: −2.0	R12-S10	+	0	0	0
*Hin*dIII: −1.7	R12-S10	+	0	0	0
*Hin*dIII: −1.3	R12-S10	++	0	0	0
*Hin*dIII: −0.8	R12-S10	+	0	0	0
*Hin*dIII: −0.5	R12-S10	+	0	0	0
*Hin*dIII: −0.3, −0.1*	R12-S10	+++	0	+	+
*Hin*dIII: +0.45	R12-S10	+	0	0	0
*Hin*dIII: +0.85, +1.0*	R12-S10	+	0	0	0
−13.0: *Eco*RV	R12-S10	+	0	0	n.d.
−9.0: *Eco*RV	R12-S10	+	0	0	n.d.
−7.0: *Eco*RV	R12-S10	++	0	0	n.d.
−6.0: *Eco*RV	R12-S10	+	0	0	n.d.
−13.0: *Bam*HI	R28-30	+	n.d.	0	0
−9.0: *Bam*HI	R28-30	+	n.d.	0	0
−7.0: *Bam*HI	R28-30	++	n.d.	0	0
−13.0: *Eco*RV	R28-30	+	0	0	n.d.
−9.0: *Eco*RV	R28-30	+	0	0	n.d.
−7.0: *Eco*RV	R28-30	++	0	0	n.d.

* The 0.8% agarose gel did not allow us to resolve these fragments.

Comparison from eight mammalian species sequences reveals eleven phylogenetic footprints within the same 3.5 kb region

To define regions conserved throughout evolution and thus potentially involved in basic mechanisms of *CFTR* regulation, the search for phylogenetic footprints was performed in eight mammalian species representing four orders (Primates, Artiodactylia, Lagomorpha and Rodentia) within the 3.5 kb region where the DHSs were clustered. Among the Primates, four families were represented: the single species of Hominidae (man), the Hylobatidae (gibbon), the Cercopithecoidae (cynomolgus) and the Cebidae (squirrel monkey); one or two members of the other orders were studied: Artiodactyla (cow), Lagomorpha (rabbit) and Rodentia (mouse and rat). Sequence comparisons from 2.5 kb upstream of the ATG initiation codon to 1.3 kb downstream for the first exon in these eight mammalian species allows the identification of 11 phylogenetic footprints (PFs), i.e. regions of at least six contiguous base pairs conserved in all of the species studied (Figure 3). Two of them (PF +1 and +2) are located at +1 kb in the first intron and are spaced by 21 bp. Five of them (PF −1 to −5) encompass a 116 bp region 95 bp upstream of the ATG. This region has been defined as the minimal promoter in transfection assays with reporter genes,

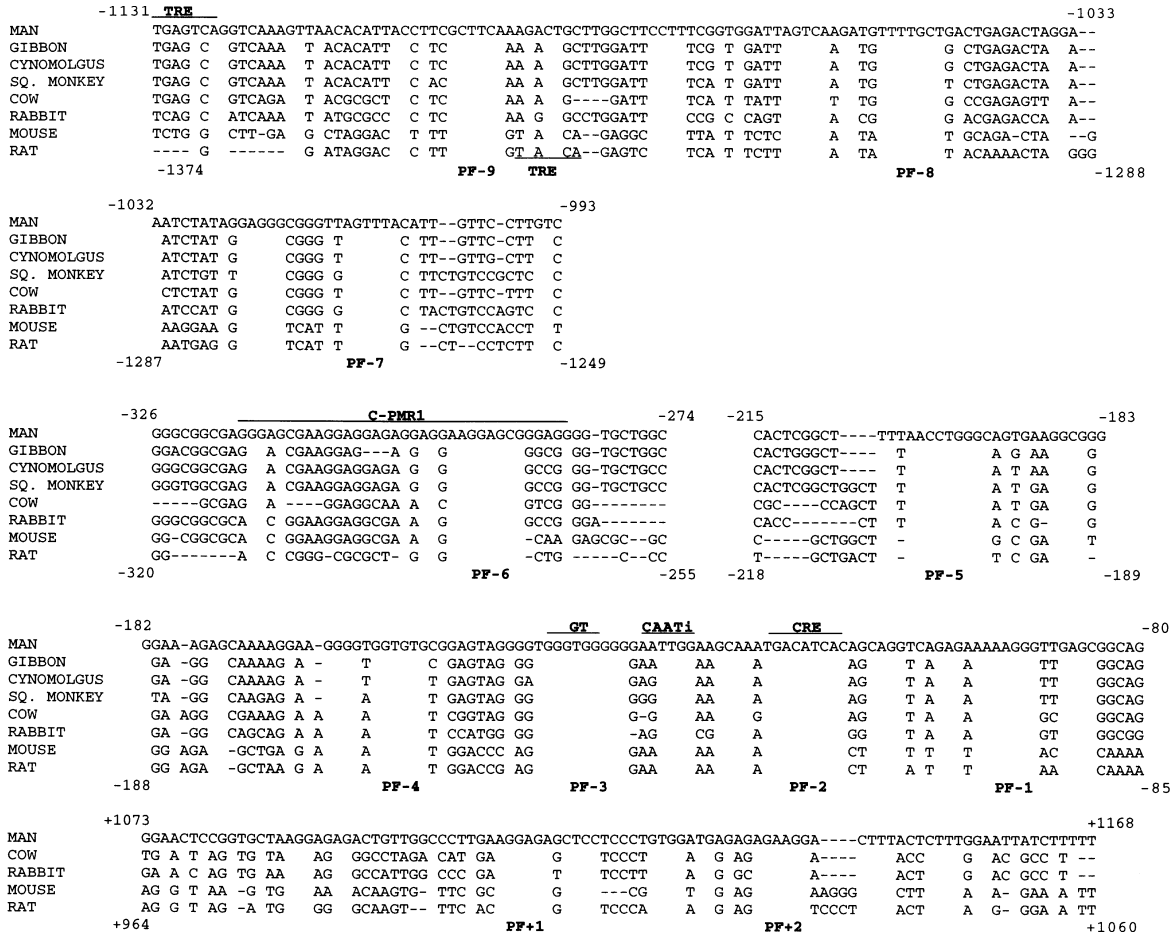

Figure 3 Sequence alignments of the *CFTR* promoter region of a number of primate species representing four families (Hominidae, man; Hylobatidae, gibbon; Cercopithecoidae, cynomolgus; Cebidae, squirrel monkey), cow, rabbit, mouse and rat, showing phylogenetic footprints

Areas of 100% conservation produce a blank space or phylogenetic footprint (PF). Deletions are indicated by a dash. Consensus or known binding sites are noted above or under the alignment. The numberings of the human and rat sequences are indicated according to the ATG initiation start codon. Primate sequences are highly conserved, whereas the most important divergence is observed between rodent and non-rodent sequences. PF +1 and PF +2 were determined without non-human primate sequences as they were available only for the first 0.5 kb within the first intron.

in both human and mouse [12–15]. One phylogenetic footprint (PF −6) is located at position −278. Lastly, three phylogenetic footprints (PF −7 to −9) are localized over a 100 bp region 1.3 kb upstream of the ATG (Figure 3). The number of phylogenetic footprints identified in the region studied is drastically lower than in promoters of the γ and ϵ globin genes [22]. This is due mainly to the divergence of sequences in the mammalian species that we have observed between rodents and non-rodents [21]. Thus it can be postulated that the remaining homologies are under a strong selective pressure and therefore have a critical role in the basic mechanisms of transcriptional *CFTR* regulation that account for the common features observed between human and rodents. In contrast, *cis* regulatory elements found only in one group (i.e. human or rodent) could be involved in species-specific patterns of *CFTR* expression.

DHSs and phylogenetic footprints map conjointly within two main regions

Comparison of the two sets of results (DHSs and phylogenetic footprints) shows that two DHSs of strong intensity found in duodenum, kidney and lung, or in duodenum only, fall within regions that are conserved between species, i.e. within the minimal promoter (PF −1 to −5 and DHS −0.1) and 1.3 kb upstream of the ATG (PF −7 to −9 and DHS −1.3). These two regions were further analysed *in vitro* for protein-binding sites by footprint protection and/or gel-shift assays with probes derived from rodent sequences incubated with rat tissue protein extracts. The choice of probes for the competition assays in the gel shift experiments was dictated by the results of the search for transcription factor recognition sequences.

PF +1 and PF +2 map around DHS +1.0 within the first intron. PF −6 maps roughly at DHS −0.3 and could correspond to the DHSs localized at −200 bp in the human cell lines T84 [15] and HT29 [16]. PF −6 falls within a region that displays Pu·Py strand asymmetry and mirror symmetry [Pu·Py mirror repeat element (PMR)] and has been called C-PMR1 [33]. It has been reported that a 27 kDa nuclear protein binds to the purine-rich strand in this region [33]. PF −6 is localized in the downstream portion of C-PMR1 and the mirror symmetry is mostly respected in all the species studied, except for the mouse sequence. However, a purine stretch of length greater than the 12

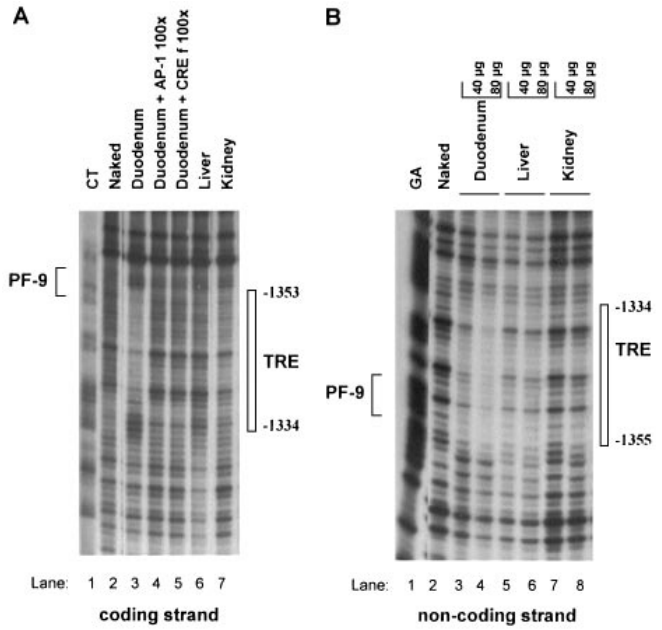

Figure 4 DNase I footprint *in vitro* of the *CFTR* promoter region in which DHS −1.3 and PF −7 to −9 map conjointly

For the coding strand (**A**), 40 μg of nuclear extracts were used; for the non-coding strand (**B**), 40 and 80 μg of nuclear extracts were tested, as indicated at the top. Lanes 2 in (**A**) and (**B**) correspond to the ^{32}P-labelled DNA fragment subjected to DNase I digestion without nuclear proteins. Protected areas are shown by boxes and indicated as numerical nucleotide position relative to the ATG. They were mapped by comparison with the Maxam and Gilbert sequence reactions performed on the identical ^{32}P-labelled DNA fragment [lanes 1 in (**A**) and (**B**)]. The protected region is seen only with duodenum extracts [lane 3 in (**A**) and lanes 3 and 4 in (**B**)] and encompasses a consensus TRE and PF −9. In (**A**) the competition experiments with added unlabelled probes (AP-1, lane 4, and CREf, lane 5) at a 100-fold molar excess before incubating the nuclear extracts and the labelled probe with the DNase I showed a disappearance of the protection.

nucleotides that seems necessary for protein binding [34] is observed in the mouse sequence (Figure 3).

Seven DHSs have no counterpart in phylogenetic footprints. Interestingly, all of them are of weak intensity and are present only in duodenum nuclei. For two of them (DHS −0.5 and −0.8) it can be argued that these sites represent rodent-specific *cis* elements, as they have been found in rat within a region that is present only in rodents, i.e. a 300 bp Pu·Py stretch. This Pu·Py stretch, which has been described previously as a *cis*-negative element [13] was further tested for its ability to confer S1 hypersensitivity. The five remaining DHSs (three within the 5′ flanking region, i.e. DHSs −1.7, −2.0 and −2.3, and two within the first intron, i.e. DHSs +0.4 and 0.85), which contain no phylogenetic footprints, map roughly to DHSs found at positions −1.6 [16] and −3.0 kb [16] in the human HT29 cell line and to the DHSs found within the first intron in T84 cell line [15]. These DHSs could represent species-specific sequences involved in *CFTR* regulation.

In the duodenum, a TRE at position −1346 bp binds a complex including Fra-2, Jun D and a protein immunologically related to Jun/CREB

Within the region where DHS −1.3 and PF −7 to −9 map conjointly, footprinting *in vitro* with a probe encompassing nt −1389 to −1227 identifies a main protected region of 20 bp (−1353 to −1334) on both strands (Figure 4). This protected region is centred on a consensus TRE (TGACTCA) and also encompasses PF −9 (GCTTCA) (Figure 3). This TRE is present only in rodents but another TRE (TG/CAGTCA) is present 31 bp upstream in the non-rodent species (Figure 3). Interestingly, this protected region is detected only with duodenal nuclear extracts (Figures 4A, lane 3, and 4B, lanes 3 and 4), but not with liver or kidney extracts whatever the amount of protein used (40 or 80 μg) (Figures 4A, lanes 6 and 7, and 4B, lanes 5–8).

Gel-shift experiments were done to characterize these protein–DNA interactions further with a double-stranded oligonucleotide GS9 covering nt −1355 to −1333. TRE are usually bound by AP-1 factors [23]. A major shifted complex was observed with the GS9 probe and duodenal protein extracts (Figure 5A, lane 2). Competition experiments showed that this complex is displaced by a 100-fold molar excess of unlabelled homologous GS9 probe, AP-1 probe (Promega), and to a smaller extent, consensus CRE probe from the fibronectin gene (CREf) [35] (Figure 5A, lanes 3, 7 and 8). To discriminate more precisely between the participation of the TRE and PF −9 sites in protein binding, gel-shift experiments were performed with oligonucleotides mutated at the PF −9 site (GS9/1M), the TRE site (GS9/2M) or both sites (GS9/1+2M). An identical shifted complex was observed with the GS9/1M probe, but no complexes were observed with the GS9/2M or the GS9/1+2M probes (results not shown). Furthermore the complex obtained with the GS9 probe is displaced totally by a 100-fold molar excess of cold GS9/1M, but not of GS9/2M nor GS9/1+2M probes (Figure 5A, lanes 4–6). Lastly, competition experiments were performed *in vitro* in a footprinting assay on the coding strand, with duodenum extracts. A 100-fold molar excess of cold AP-1 or consensus CREf probes abolished the 20 bp footprint (Figure 4A, lanes 4 and 5). Thus these results indicate that the TRE is the site involved in protein binding. Because TRE binding AP-1 transcription factors are known to be composed of members of the Fos and Jun protein families, we next performed experiments to determine the components involved in protein binding to the TRE site. Supershift experiments were performed with a rabbit polyclonal anti-c-Fos (K-25) antibody that cross-reacts with Fos B, Fra-1 and Fra-2 (Santa Cruz Biotechnology, Santa Cruz, CA, U.S.A.) and a rabbit polyclonal anti-(c-Jun/AP-1) (D) antibody that cross-reacts with Jun B and Jun D (Santa Cruz Biotechnology). The complex disappeared and supershifted upwards with the anti-c-Fos (K-25) antibody and disappeared with the anti-(c-Jun/AP-1) (D) antibody (Figure 5B, lanes 3 and 8). A partial and uniform disappearance of the complex was observed when using an anti-CREB antibody raised against a synthetic peptide encompassing amino acid residues 121–159 (CREB no. 244) (Marc Montminy, Salk Institute, La Jolla, CA, U.S.A.) (Figure 5B, lane 13). We then tried to determine which members of the known Fos (c-Fos, Fos B, Fra-1 and Fra-2) and Jun (c-Jun, Jun B and Jun D) proteins were present in the complex by using specific antibodies [36,37]. Within the Fos family, only the anti-(Fra-2)-specific antibody markedly affected the binding complex (Figure 5B, lanes 4–7). Within the Jun family, a supershift and a decrease in the upper part of the shifted complex are observed with the anti-(Jun D)-specific antibody and with the mixture of the three anti-Jun-specific antibodies (Figure 5B, lanes 9–12). No modification of the retarded complex was observed with a rabbit preimmune serum used as a control for non-specific effects (results not shown). To unravel the discrepancy between the results obtained with the anti-(c-Jun/AP-1) (D) and the anti-(Jun D)-specific antibodies, supershift experiments were performed with the GS9 probe and protein extracts from the *CFTR* non-expressing cells NIH-3T3, which are known to express Jun D at a high level [38]. The retarded complex obtained with the NIH-3T3 cell extracts

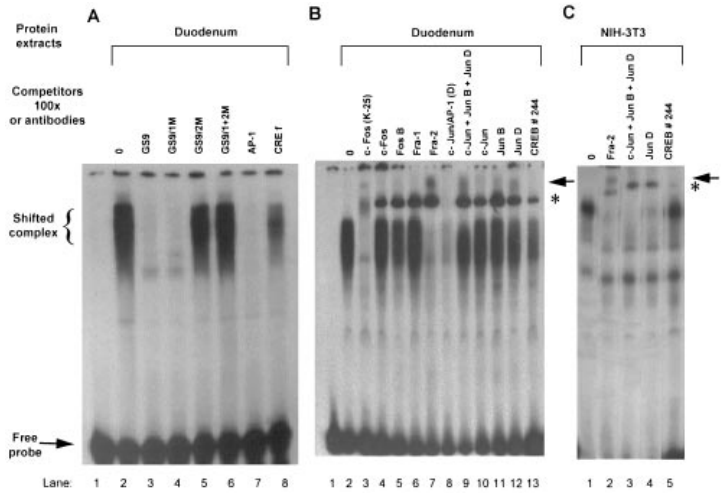

Figure 5 Gel-shift experiments with probes encompassing the TRE site

Protein extracts were from duodenum in (**A**) and (**B**) and from NIH-3T3 in (**C**). The labelled probe was GS9 and the GS9/1M, GS9/2M and GS9/1 + 2M probes were used for competition. The GS9 probe covers nt −1355 to −1333 and encompasses the TRE consensus site and PF −9. GS9/1M, GS9/2M and GS9/1 + 2M correspond to probes in which the PF −9, TRE and both sites were mutated respectively. The complete sequences of the probes are described in the Materials and methods section. The first lanes in (**A**) and (**B**) correspond to the probe alone without protein extract. Lanes 2 in (**A**) and (**B**) and 1 in (**C**) correspond to the retarded complex observed with duodenum or NIH-3T3 nuclear protein extracts respectively. The retarded complex obtained with the NIH-3T3 protein extracts corresponds to the upper part of the duodenum protein extract complex. (**A**) The 100-fold excess unlabelled probes used as competitor are indicated at the top of the autoradiogram. Lanes 3–7, duodenum nuclear extract binding was competed away with 100-fold unlabelled GS9, GS9/1M and AP-1 probes but not with GS9/2M and GS9/1 + 2M probes; 100-fold unlabelled CREf probe, which encompasses the fibronectin CRE consensus site, moderately displaced the retarded complex. (**B**, **C**) Supershift experiments with duodenum and NIH-3T3 cell extracts respectively. Antibodies are indicated at the top of the autoradiogram. Specific antibodies were preincubated for 1 h at 4 °C before performing the electrophoretic mobility-shift assay. (**B**) Lanes 3–7, antibodies against the Fos family members. Only anti-(c-Fos) (K25) and anti-(Fra-2) antibodies markedly affected the binding complex. Lanes 8–12, antibodies against the Jun family members. The binding complex disappeared with the anti-(c-Jun/AP-1) (D) antibody, whereas a supershift and a decrease in the upper part of the shifted complex were observed with the anti-(Jun D)-specific antibody and with the mixture of the three anti-Jun-specific antibodies. Lane 13, CREB no. 244 antibody yielded a partial and uniform disappearance of the binding complex. (**C**) Lanes 2–5, anti-(Fra-2) and anti-(Jun D) antibodies markedly affected the complex, while CREB no. 244 had no effect. Samples were analysed on a 4% (w/v) polyacrylamide gel. The star corresponds to the immunoprecipitation of the probe; the arrow corresponds to a supershifted complex.

corresponds to the upper part of the complex obtained with the duodenum extracts (Figure 5C, lane 1). This complex was completely supershifted by the anti-(Fra-2) antibody. It disappeared with the anti-(Jun D) and anti-(Jun D/c-Jun/Jun D) antibodies but was unaffected by the CREB no. 244 antibody (Figure 5C, lanes 2–5) or by a rabbit preimmune serum (results not shown). These results are consistent with the expression of Jun and Fos protein in NIH-3T3 fibroblasts [36]. They demonstrate the efficiency of the anti-Jun D antibody and the participation of Jun D in the binding complex. Taken together, these findings strongly suggest that the duodenum retarded complex is composed of Fra-2, Jun D and of a Jun/CREB immunologically related protein. In contrast, the complex observed in the *CFTR* non-expressing cell line NIH-3T3 is composed of only Fra-2 and Jun D.

The TRE site acts as a weak transactivator stimulated by PMA in Caco-2 cells

The functional relevance of the TRE was assessed in the human intestinal Caco-2 cell line by using reporter gene experiments. These cells express *CFTR* at high level [16] and, in supershift experiments with the GS9 probe, Caco-2 nuclear extracts produce a pattern roughly identical with that obtained with duodenum extracts (results not shown). Linked to the *tk* promoter, the TRE site has a weak transactivator activity and is moderately stimulated by PMA. Both effects were abolished when the TRE site was mutated (Figure 6).

ATF-1 binds to a CRE-like site at position −124 to −116 bp within the minimal promoter

Footprinting of the minimal promoter region *in vitro* where DHSs −0.1 and PF −1 to −5 conjointly map was performed with a probe encompassing nt −50 bp to −218 bp. On both strands it identified a protected region of 19 bp (−129 to −111 to bp) (Figure 7). This protection was present with all tested tissue proteins (duodenum, lung, liver and kidney). The protected region is centred on a CRE motif (TGACGTCA) degenerated on one nucleotide (the underlined A) TGAC<u>A</u>TCA (CREcf) and corresponds to PF-2 (Figure 3). No other protection was observed when the amount of protein was increased to 80 μg (results not shown). To further characterize this CRE-like motif and to explore other potential protein–DNA interactions within this region, gel shift analysis were performed with double-stranded oligonucleotides covering nt −154 to −109 (GS2) and nt −129 to −90 (GS1). GS1 encompasses the CREcf site and sequences downstream, but not the inverted CAAT (CAATi) element, which has been involved in the cAMP-mediated transcriptional regulation of *CFTR* [39]. GS2 encompasses CREcf and CAATi, and sequences upstream of both elements. With both GS2 and GS1 probes, an identical major gel-shifted band was observed with all protein extracts, whatever the level of *CFTR* expression (Figure 8, lanes 2, 8 and 24). Competition assays with both probes showed that this shifted band is displaced by 10–100-fold molar excess of unlabelled GS1 or GS2 probes, of the CREcf probe that contains the CRE-like site only, and of the consensus CREf probe [35] (Figure 8, lanes 3, 4 and 9–14). The shifted band

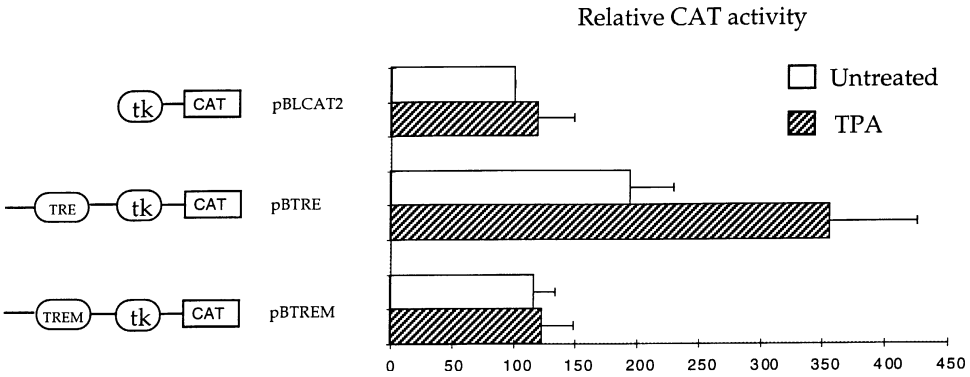

Figure 6 Functional analysis of the TRE site in Caco-2 cells

Relative promoter activity in the Caco-2 cell line by using constructs encompassing the TRE site linked to the *tk* promoter fused to the *CAT* gene (see the Materials and methods section). The results are expressed as promoter activity relative to the activity of the pBLCAT2 construct. pBTRE and pBTREM constructs contained respectively the wild-type and the mutated TRE sequences corresponding to the GS9 and GS9/1 + 2M probes. Cells were stimulated by PMA at a final concentration of 100 nM for 8–10 h. Each value represents the mean ± S.E.M. for at least three independent transfection experiments. Levels of CAT expression were normalized to the amount of proteins (see the Materials and methods section).

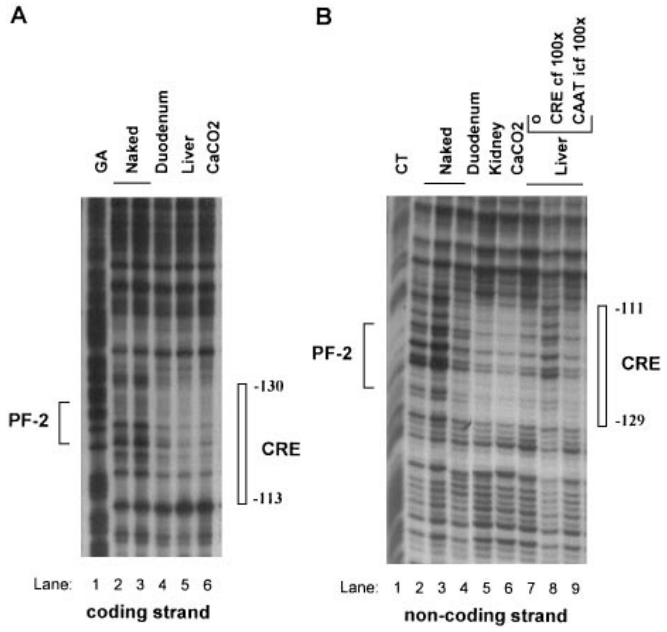

Figure 7 DNase I footprint *in vitro* of the *CFTR* minimal promoter located 250 bp upstream of the ATG

(**A**) Coding strand; (**B**) non-coding strand. Lanes 2 and 3 in (**A**) and (**B**) correspond to the ^{32}P-labelled DNA fragment subjected to DNase I digestion without nuclear extracts. Protected areas are shown by boxes and mapped by comparison with the Maxam and Gilbert sequence reactions [lanes 1 in (**A**) and (**B**)] as in Figure 4. The protected region is centred on a CRE degenerated on one base and PF-2. In (**B**), competition experiments performed by adding unlabelled probes (CREcf and CAATicf) in 100-fold molar excess before incubating with DNase I (lanes 8 and 9) showed a disappearance of the protection for the CREcf probe only. CaCO2 = CaCO-2 cells.

was not displaced by cold GS1M or GS2M probes in which the CREcf sequence was changed for CATCACCA, nor by cold CAATicf or CAATith probes, which contain the inverted CAAT boxes of *CFTR* or of the human tryptophan hydroxylase gene promoter [40] respectively (Figure 8, lanes 15–20, and results not shown). Furthermore the shifted band was not displaced by unlabelled AP-1 or Sp1 probes (results not shown). No shifted band was observed when using rat duodenum nuclear extracts and the GS1M and GS2M probes (Figure 8, lanes 6 and 23). Competition experiments in the footprint assay confirmed the gel-shift data, as a 100-fold molar excess of unlabelled CREcf oligonucleotide abolished the protection, but the CAATicf oligonucleotide did not (Figure 7B, lanes 8 and 9). It can be concluded from these experiments that the CREcf site is the binding site for protein(s) and that neither the CAATi box nor the GT box (GGGTGG) corresponding to PF −3 (Figure 3) is involved in the retarded complex. To identify the protein(s) involved in the shifted band, supershift experiments with antibodies targeted against the CREB/ATF proteins [CREB no. 244 as well as CREB-1 (24H4B), ATF-1 (C41-5.1), ATF-2 (C19) (Santa Cruz Biotechnology)], and to the related bZIP family proteins C/EBP [C/EBPα, C/EBPβ, C/EBPδ and CRP-1 (Santa Cruz Biotechnology)] and AP-1 [c-Fos (K-25) and c-Jun/AP-1 (D)] were performed. The anti-(ATF-1) antibody was the only one to alter DNA–protein interaction, as it partly abolished the shifted band (Figure 8, lanes 24–26, and results not shown). Gel-shift analysis with the GS2 probe and 0.5 μg of purified CREMτ (Paolo Sassone-Corsi, INSERM U184, Illkirch, France) gave a protein–DNA complex of higher molecular mass than the duodenum protein–DNA complex (Figure 8, lane 21). Thus it can be concluded from the above data that ATF-1 binds to the CREcf motif.

The *CFTR* CRE-like motif drives a high basal transcriptional activity but does not promote cAMP transcriptional stimulation of a heterologous promoter in Caco-2 cells

To assess the functional relevance of the CREcf sequences, Caco-2 cell lines were used in reporter gene experiments. The endogenous *CFTR* gene was moderately stimulated by cAMP in Caco-2 cells (results not shown) and extracts from these cells gave a protection identical with that obtained with rat tissue extracts (Figures 7A and 7B, lane 6). A 2–3-fold increase in the basal expression of the reporter gene was observed 8–10 h after forskolin stimulation of Caco-2 cells transfected with a 250 bp mouse minimal *CFTR* promoter linked to the *CAT* gene (RV 2.0) [13]. No CAT activity was detected with the control plasmid PL (Figure 9A). It has been reported that the CAATicf element directs the cAMP-mediated activation of a heterologous pro-

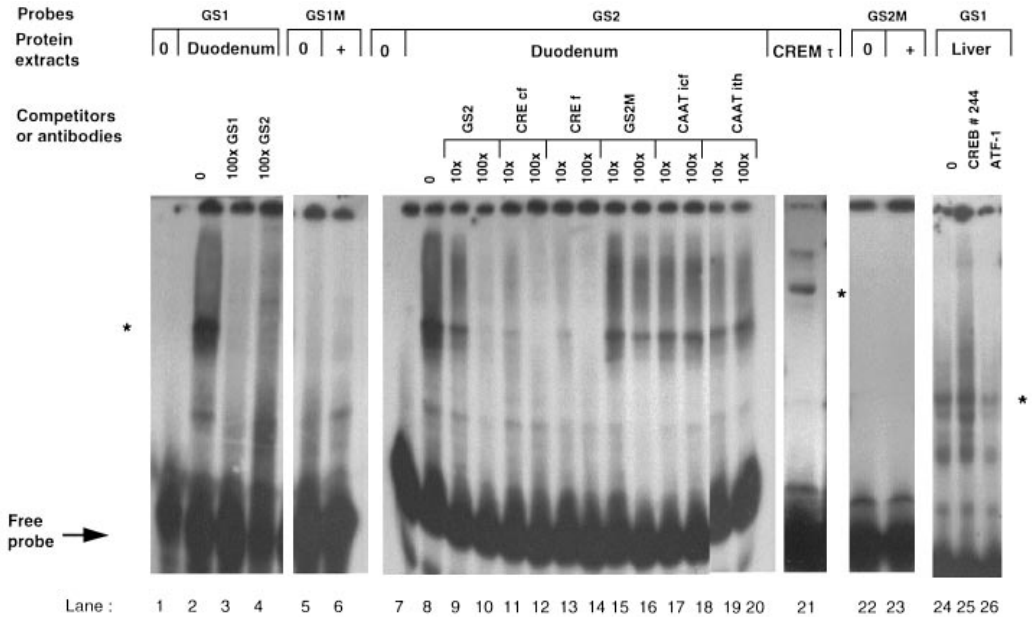

Figure 8 Gel-shift experiments with probes encompassing the CREcf

The labelled probes (GS1, GS1M, GS2 and GS2M) are indicated at the top and correspond to sequences surrounding the CREcf (see squences in the Materials and methods section). In GS1M and GS2M probes, the CREcf has been mutated. Protein extracts are from duodenum (lanes 2–4, 6, 8–20 and 23) or liver (lanes 24–26) except for the GS2 probe, where purified CREMτ was used in one case (lane 21). There were no protein extracts in lanes 1, 5, 7 and 22. The 10-fold or 100-fold excess unlabelled probes used as competitors are indicated at the top of the autoradiography. In lanes 3, 4 and 9–20, duodenum nuclear extract complex was competed away with unlabelled homologous CREcf and CREf probes but not with GS2M, CAATicf or CAATith probes. No retarded complex was observed with duodenum extracts and GS1M (lane 6) or GS2M (lane 23) probes. The retarded complex observed with CREMτ (lane 21) was of higher molecular mass than the duodenum protein–DNA complex (lane 8). Stars indicate the shifted bands. For supershift experiments (lanes 24–26), the nature of the antibodies is indicated at the top of the autoradiogram. Samples were loaded on a 4% (w/v) polyacrylamide gel.

moter in human PANC1 and HeLaS3 cells which do not express *CFTR* [39]. We therefore tested the ability of the CAAT site alone, of the CREcf site alone, and of the two sites together to direct cAMP-mediated activation of the *tk* promoter in Caco-2 cells. Forskolin treatment had no significant effect on the expression of constructs containing CAATicf, CREcf and CAATicf + CREcf sequences. However, an increase in the basal level of expression, as compared with the pBLCAT2 plasmid alone, was seen with the pBCRE and pBCAAT-CRE plasmids (3–6-fold) but not with the pBCAAT plasmid (Figure 9B).

Thus the CREcf is involved in basal transcriptional activity but cAMP regulation of the mouse minimal promoter (RV2.0 construct) involves some other elements than the CREcf and CAAT sites.

The rodent-specific Pu·Py stretch confers S1 nuclease hypersensitivity under conditions of acidic pH and supercoiling, indicating a non-B DNA conformation

The ability of the Pu·Py stretch to induce conformational changes in DNA was tested by its ability to confer S1 nuclease hypersensitivity on superhelical plasmids [41]. The assay consisted of determining whether discrete S1 cleavage sites occur in the Pu·Py stretch of the P1–P2 plasmid construct. Both the control (PL) and test (P1–P2) plasmids were first subjected to increasing levels of S1 digestion (0, 0.5, 1 and 2 units of enzyme) and then to *Nco*I cleavage. Because a single *Nco*I site is present in both plasmids, discrete bands are expected only if S1 nuclease has cleaved at specific sites, otherwise the plasmid would simply linearize. *Nco*I cleavage alone of the control and P1–P2 plasmids generated in both cases a single cleavage product of 4.5 and 5.6 kb respectively, consistent with a linearized plasmid (Figure 10A). Prior digestion with increasing concentrations of S1 resulted in discrete bands of 4.4, 4.3, 1.3 and 1.2 kb, but for the P1–P2 plasmid only. These fragments represent *Nco*I–S1 cleavage products, indicating the presence of two S1 sites, 100 bp apart from each other, within the Pu·Py stretch (Figure 10B).

DISCUSSION

Difficulties in identifying the regulatory elements involved in the transcriptional regulation of *CFTR* might have been due partly to the use of approaches that (1) do not analyse the gene in its chromatin environment or (2) use immortalized cell lines. To overcome these difficulties we have used a combined approach based on the search for DHSs in rat tissues and on evolutionary clues, i.e. phylogenetic footprinting. In this report, we present evidence for the implication of multiple regulatory elements that act in the chromatin structure *in vivo*, some of them being conserved throughout evolution, the others being species-specific. These findings support the notion that some basic mechanisms of *CFTR* regulation, such as the involvement of a CRE-like element in the basal level of transcription, are conserved between species, while species-specific mechanisms also exist.

Among the conserved elements, a CRE-like element (CREcf) and a TRE were identified. Classically, these elements are the ultimate targets for transcriptional control through the cAMP and diacylglycerol pathways, respectively [23–25]. Both pathways have been reported to modulate *CFTR* expression at the transcriptional level in human cell cultures [8–10,42,43]. Various stimuli that lead to a sustained (at least 8 h) elevation of intracellular cAMP levels elicit a specific increase in *CFTR*

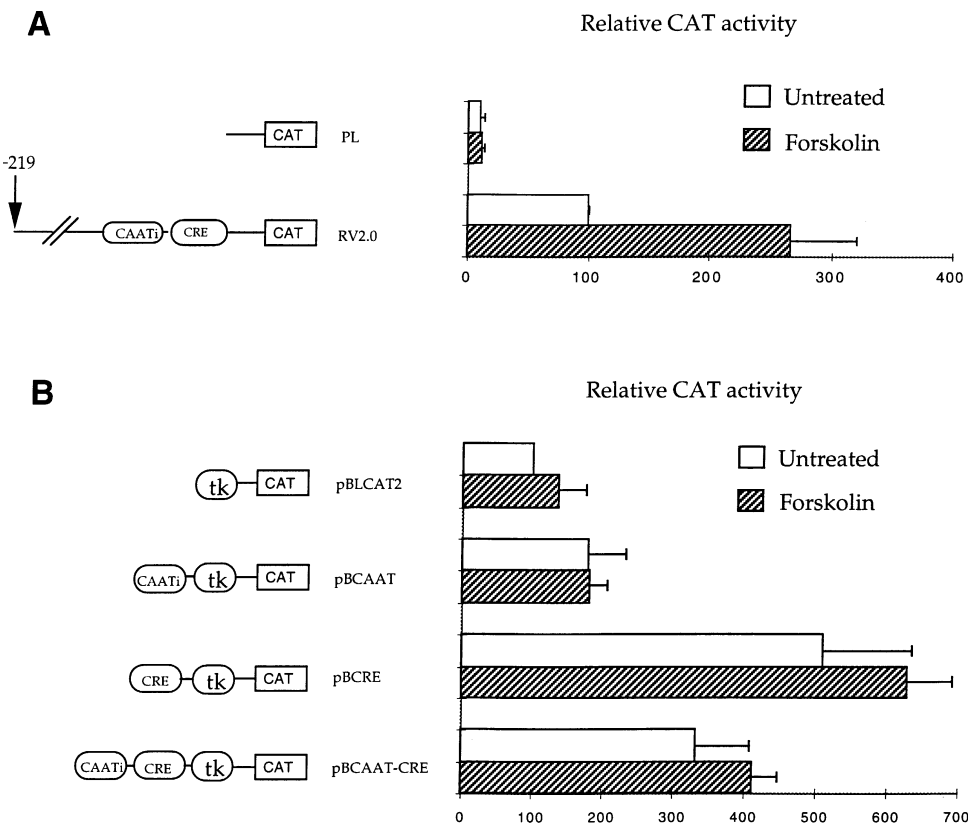

Figure 9 Functional analysis of the CRE site in Caco-2 cells

Relative promoter activity in Caco-2 cells is shown. (**A**) The RV2.0 construct consists of the mouse minimal promoter sequence (nt −219 to +15 from the ATG initiation start) fused to the *CAT* gene. Data are expressed as promoter activity relative to the activity of the RV2.0 construct. (**B**) Constructs correspond to the rat *CFTR* CAATi, CRE, and CAATi and CRE elements linked to the *tk* promoter fused to the *CAT* gene (see the Materials and methods section). In (**A**) and (**B**), cells were stimulated by forskolin at a final concentration of 10 μM for 8–10 h. Each value shown represents the mean ± S.E.M. for at least three independent transfection experiments. Levels of CAT expression were normalized for transfection efficiency by comparison with an internal control plasmid harbouring the β-galactosidase gene (pCH110) as described in the Materials and methods section.

expression in various cell lines including HT-29 and T84 human colonic carcinoma cells [8,42] and CaLu-3 and HTE-1 human lung epithelial cells [43]. The level of inducibility is dependent on the cell type but in all cases remains modest. Several groups have shown that treatment of T84 and HT29 cells with PMA results in a down-regulation of CFTR mRNA [9,10] with a decrease in the transcription rate of *CFTR* [10].

The nuclear effector implicated in the cAMP-dependent activation of *CFTR*, as well as the location and nature of its binding site(s), remain controversial. McDonald et al. [42] showed that the CREcf is implicated in protein kinase A (PKA)-mediated responsiveness of the promoter, PKA being the protein kinase involved in the cAMP pathway. Further studies have demonstrated that CREcf binds CREB and ATF-1 in T84 and CaLu-3 cells [43]. Pittman et al. [39] reported that in human PANC1 cell lines that physiologically do not express *CFTR*, an imperfect and inverted CAAT (CAATi) element located 7 bp upstream of the CREcf (Figure 3) was essential: (1) for the basal expression of an exogeneous *CFTR* promoter linked to a reporter gene, and (2) for the cAMP-mediated transcriptional regulation of a heterologous promoter. They proposed that C/EBP, ATF1 and CREB1 are part of a nuclear protein complex bound to this CAATi sequence [39]. The present results on DNA–protein interaction, with protein extracts from rat tissues that are more physiologically significant than nuclear extract from cell lines, provide no evidence for such a possibility. Indeed, our footprinting experiments *in vitro* show clearly a ubiquitous 20 bp protection that corresponds to PF −2 and is centred on the CREcf. Furthermore we find that the CAATi box is not conserved between species (Figure 3). Our gel-shift experiments are in agreement with the footprinting experiments *in vitro* as they show that the binding site for protein(s) is the CREcf and that, under our conditions, neither the GT box Sp1 binding site [44] nor the CAATi box [39,40], which have been previously implicated in the cAMP transduction pathway, is involved in protein binding. The fast-moving retarded band in the gel-shift experiments as compared with the band obtained with CREMτ and the modest activator effect obtained in the functional assays are in agreement with the supershift experiments that identify ATF-1 as the binding protein [24]. Our experiments in Caco-2 cells with a construct with the mouse minimal promoter sequence extend the evidence of *CFTR* inducibility by cAMP, thereby showing that this mechanism is conserved between human and rodents. However, outside their context, the CREcf and the CAATi box linked to the *tk* promoter fail to direct cAMP-mediated transcription either alone or together, even though similar elements have been shown to co-operate in the human fibronectin gene promoter [45]. Thus it seems that multiple additional unidentified elements located in the 250 bp region upstream of the ATG are critical in directing cAMP-mediated transcription. This might explain why the delay for transcriptional induction by cAMP is long (6–10 h) and why the level of induction is weak (one-half to

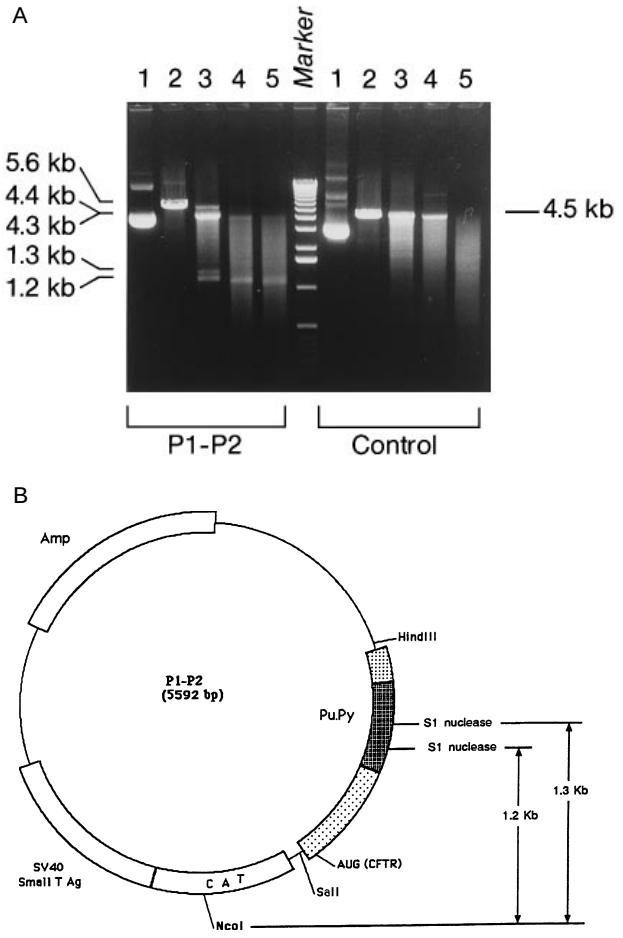

Figure 10 S1 nuclease hypersensitivity assay of the rodent-specific Pu·Py stretch

(**A**) Electrophoretic fractionation of *Nco*I and S1 nuclease cleavage products on a 1% (w/v) agarose gel. The P1–P2 and control (PL) plasmids are shown as superhelical plasmids (lanes 1), linearized with *Nco*I and no S1 nuclease (lanes 2), and with *Nco*I and increasing concentrations of S1 nuclease (lanes 3, 0.5 unit; lanes 4, 1 unit; lanes 5, 2 units). The marker lane is a 1 kb ladder. Complete cleavage with S1 resulted in a 1.2 kb fragment and its associated 4.4 kb fragment. The incomplete cleavage shown by the 1.3 and 4.3 kb fragments demonstrates an additional S1 site. (**B**) Map of the P1–P2 plasmid. The unique *Nco*I site is shown in bold. The two S1 nuclease sites map at 1.2 and 1.3 kb from *Nco*I within the Pu·Py stretch.

one-third), both facts suggesting the contribution of multiple partners as recently suggested [43]. The present work together with the results of McDonald et al. [42] show clearly that the CREcf is involved in the basal expression of *CFTR*. It has to be noted that it is the first time that a CRE site has been reported in a TATA-less promoter, thus raising the possibility that it could act as a positioning element. Interestingly, elements with the same sequence as CREcf have been previously reported in c-*Jun* [46] and the tissue-type plasminogen activator [47] gene promoters. In these two systems, the CRE-like site binds CREB, ATF-2 or c-Jun depending on the cell type [48,49] and seems to be involved in basal expression but not in the classical cAMP-mediated induction of transcription [48,49]. The containing CREcf region is highly sensitive to DNase I in a tissue-specific manner that correlates well with the expression of *CFTR* (duodenum, + + + ; lung, + ; kidney, + ; liver, 0). This contrasts with the ubiquitous character of the protection observed in the footprinting assay *in vitro* and of the retarded complexes in gel-shift experiments. Thus tissue-specific features concerning the CREcf do exist that are observed only within the chromatin structure. A plausible hypothesis is that in *CFTR*-expressing cells, long-range protein–protein interactions promote DNA looping and co-operation between several *cis*-acting elements scattered over at least 14 kb of the gene (Figure 1).

The TRE site that we describe here binds a complex composed of Fra-2, Jun D and a protein immunologically related to Jun/CREB. We propose that this complex has a role in the duodenum-specific expression of *CFTR*. It has been demonstrated in other systems that specific heterodimers of the Fos and Jun family members might have a role in tissue-specific expression [50,51]. Furthermore it has been shown that an additional level of sophistication can be reached by cross-family dimerization of Fos/Jun and ATF/CREB family members and altered binding specificities [52]. Our results indicate that most of the duodenum-specific retarded complex is composed of Fra-2 and a protein immunologically related to Jun/CREB. Fra-2 is expressed ubiquitously but Jun D is highly expressed in the intestine and not in the liver [38,53]. The protein immunologically related to Jun/CREB involved in the duodenum-specific complex should heterodimerize with Fra-2, should be expressed in the duodenum, and the heterodimers it forms with Fra-2 should bind a TRE site with high affinity. Within the CREB/ATF family, the protein most closely related to the Jun family is ATF-2 [24]; Fos/ATF-2 heterodimers have been described [54]. However, we were unable to supershift the duodenum retarded complex with an ATF-2 specific antibody (results not shown). The interaction with the *CFTR* TRE site that we observed merits further investigation to determine the main protein that heterodimerizes with Fra-2. In our transfection experiments in Caco-2 cells, we obtained a weak stimulation of transcription by TPA. This indicates that the decrease in *CFTR* expression that was observed in the *CFTR*-expressing cell-lines HT-29 and T84 [9,10] must involve other unknown PMA-sensitive elements within the endogenous gene.

Among the DHSs that we mapped within the non-conserved elements, two fall in the 300 bp Pu·Py stretch specific of rodents localized 500–550 bp upstream from the ATG. This element has been described to act as a negative element of basal transcription in transient transfection assays [13]. The S1 nuclease assay in the present study shows clearly that this Pu·Py stretch exhibits hypersensitivity to single-strand-specific nucleases at low pH and in supercoiling conditions, indicating a non-B DNA conformation [41]. Furthermore DHS-0.3 maps roughly within a phylogenetic footprint (PF −6) that corresponds, in the human *CFTR* promoter, to a region of non-random Pu·Py strand asymmetry (C-PMR1). C-PMR1 exhibits non-B DNA conformation and a 27 kDa nuclear protein binds to the purine-rich strand of this element [33,34]. The identification of three DHSs, which reflect conditions *in vivo*, specific for *CFTR*-expressing tissue, within non-random purine/pyrimidine strand asymmetry (Pu·Py stretch and C-PMR1) reinforces the biological significance of these elements in the regulation of transcription. Conditions required to generate a non-B DNA structure such as low pH are not expected to pertain *in vivo*. However, accumulating evidence suggests that triplex structures can form at neutral pH and chromatin structure as well as other local protein–DNA interactions result in a high degree of supercoiling in intact cells [55].

We thank Paolo Sassone-Corsi for his invaluable help in the realization of this work; Bernard Grandchamp, Raphaël Scharfmann, Pierre Lehn and Ann Harris for fruitful discussions and encouragements; Bernard Kaltenboeck for computer analysis of the

nucleotide sequences; and Mireille Lambert, Violaine Pinta, Abdelali Tazi and Noufissa Oudrhiri for their help within the binding assays *in vitro*. This work was supported in part by the Association Française de Lutte contre la Mucoviscidose (AFLM).

REFERENCES

1 Welsh, M. J., Tsui, L. C., Boat, T. F. and Beaudet, A. L. (1995) in The Metabolic and Molecular Bases of Inherited Diseases (Scriver, C. R., Beaudet, A. L., Sly, W. S. and Valle, D., eds.), pp. 3799–3876, McGraw-Hill, New York
2 Trezise, A. E. O., Chambers, J. A., Wardle, C. J., Gould, S. and Harris, A. (1993) Hum. Mol. Genet. **2**, 213–218
3 Tizzano, E. F., Chitayat, D. and Buchwald, M. (1993) Hum. Mol. Genet. **3**, 221–224
4 Riordan, J. R., Rommens, J. M., Kerem, B. S., Alon, N., Rozmahel, R., Grzelczak, Z., Zielenski, J., Lok, S., Plavsic, N., Chou, J. L. et al. (1989) Science **245**, 1066–1073
5 Trezise, A. E. O. and Buchwald, M. (1991) Nature (London) **353**, 434–437
6 Engelhardt, J. F., Yankaskas, J. R., Ernst, S. A., Yang, Y., Marino, C. R., Boucher, R. C., Cohn, J. A. and Wilson, J. M. (1992) Nat. Genet. **2**, 240–247
7 Trezise, A. E. O., Linder, C. C., Grieger, D., Thompson, E. W., Meunier, H., Griswold, M. D. and Buchwald, M. (1993) Nat. Genet. **3**, 157–164
8 Breuer, W., Kartner, N., Riordan, J. R. and Cabantchik, Z. I. (1992) J. Biol. Chem. **267**, 10465–10469
9 Trapnell, B. C., Zeitlin, P. L., Chut, C. S., Yoshimura, K., Nakamura, H., Guggino, W. B., Bargon, J., Banks, T. C., Dalemans, W., Pavirani, A. et al. (1991) J. Biol. Chem. **266**, 10319–10323
10 Bargon, J., Trapnell, B. C., Yoshimura, K., Dalemans, W., Pavirani, A., Lecocq, J. P. and Crystal, R. G. (1992) J. Biol. Chem. **267**, 16056–16060
11 Bargon, J., Trapnell, B. C., Chu, C. S., Rosenthal, E. R., Yoshimura, K., Guggino, W. B., Dalemans, W., Pavirani, A., Lecocq, J. P. and Crystal, R. G. (1992) Mol. Cell. Biol. **12**, 1872–1878
12 Yoshimura, K., Nakamura, H., Trapnell, B. C., Dalemans, W., Pavirani, A., Lecocq, J. P. and Crystal, R. C. (1991) J. Biol. Chem. **266**, 9140–9144
13 Denamur, E. and Chehab, F. F. (1994) Hum. Mol. Genet. **3**, 1089–1094
14 Chou, J. L., Rozmahel, R. and Tsui, L. C. (1991) J. Biol. Chem. **266**, 24471–24476
15 Koh, J., Sferra, T. J. and Collins, F. S. (1993) J. Biol. Chem. **268**, 15912–15921
16 Yoshimura, K., Nakamura, H., Trapnell, B. C., Chu, C. S., Dalemans, W., Pavirani, A., Lecoq, J. P. and Crystal, R. G. (1991) Nucleic Acids Res. **19**, 5417–5423
17 Smith, A. N., Wardle, C. J. C. and Harris, A. (1995) Biochem. Biophys. Res. Commun. **211**, 274–281
18 Smith, A. N., Barth, M. L., McDowell, T. L., Moulin, D. S., Nuthall, H. N., Hollingsworth, M. A. and Harris, A. (1996) J. Biol. Chem. **271**, 9947–9954
19 Denamur, E. and Chehab, F. F. (1995) DNA Cell Biol. **14**, 811–815
20 Gross, W. T. (1988) Annu. Rev. Biochem. **57**, 159–197
21 Vuillaumier, S., Kaltenbock, B., Lecointre, G., Lehn, P. and Denamur, E. (1997) Mol. Biol. Evol. **14**, 372–380
22 Gumucio, D. L., Heilstedt-Williamson, H., Gray, T. A., Tarlé, S. A., Shelton, D. A., Tagle, D. A., Slightom, J. L., Goodman, M. and Collins, F. S. (1992) Mol. Cell. Biol. **12**, 4919–4929
23 Vogt, P. K. and Bos, T. J. (1989) Trends Biochem. Sci. **14**, 172–175
24 Sassone-Corsi, P. (1995) Annu. Rev. Cell. Dev. Biol. **11**, 355–377
25 Vallejo, M. (1994) J. Neuroendocrinol. **6**, 587–596
26 Wu, C. (1980) Nature (London) **286**, 854–859
27 Perret, C., L'Horset, F. and Thomasset, M. (1991) Gene **108**, 227–235
28 Higgins, D. G. and Sharp, P. M. (1988) Gene **73**, 237–244
29 Dessen, P., Fondrat, C., Valencien, C. and Mugnier, C. (1990) Comput. Appl. Biosci. **6**, 355–356
30 Lambert, M., Colnot, S., Suh, E. R., L'Horset, F., Blin, C., Calliot, M. E., Raymondjean, M., Thomasset, M., Traber, P. G. and Perret, C. (1996) Eur. J. Biochem. **236**, 778–788
31 Sassone-Corsi, P., Lamph, W. W., Kamp, M. and Verma, I. M. (1988) Cell **54**, 553–560
32 Snouwaert, J. N., Brigman, K. K., Latour, A. M., Malouf, N. N., Boucher, R. C., Smithies, O. and Koller, B. H. (1992) Science **257**, 1083–1088
33 Hollingswoth, M. A., Closken, C., Harris, A., McDonald, C., Paahwa, G. S. and Maher, III, L. J. (1994) Nucleic Acids Res. **7**, 1138–1146
34 McDonald, C. D., Hollingsowth, M. A. and Maher, III, L. J. (1994) Gene **150**, 267–274
35 Maekawa, T., Sakura, H., Kanei-Ishii, C., Sudo, T., Yoshimura, T., Fujisawa, J., Yoshida, M. and Ishii, S. (1989) EMBO J. **8**, 2023–2028
36 Lallemand, D., Spyrou, S., Yaniv, M. and Pfarr, C. M. (1997) Oncogene **14**, 819–830
37 Pfarr, C. M., Mechta, F., Spyrou, G., Lallemand, D., Carillo, S. and Yaniv, M. (1994) Cell **76**, 747–760
38 Ryder, K., Lanahan, A., Perez-Albuerne, E. and Nathans, D. (1989) Proc. Natl. Acad. Sci. U.S.A. **86**, 1500–1503
39 Pittman, N., Shue, G., LeLeiko, N. S. and Walsh, M. J. (1995) J. Biol. Chem. **270**, 28848–28857
40 Boularand, S., Darmon, M. C., Ravassard, P. and Mallet, J. (1995) J. Biol. Chem. **270**, 3757–3764
41 Hanvey, J. C., Klysik, J. and Wells, R. D. (1988) J. Biol. Chem. **263**, 7386–7396
42 McDonald, R. A., Matthews, R. P., Idzerda, R. L. and McKnight, G. S. (1995) Proc. Natl. Acad. Sci. U.S.A. **92**, 7560–7564
43 Matthews, R. P. and McKnight, G. S. (1996) J. Biol. Chem. **271**, 31869–31877
44 Begeot, M., Shetty, U., Kilgore, M., Waterman, M. and Simpson, E. (1993) J. Biol. Chem. **268**, 17317–17325
45 Muro, A. F., Bernath, V. A. and Kornblihtt, A. R. (1992) J. Biol. Chem. **267**, 12767–12774
46 Herr, I., Van Dam, H. and Angel, P. (1994) Carcinogenesis **15**, 1105–1113
47 Medcalf, R. L., Rüegg, V. and Schleuning, W. D. (1990) J. Biol. Chem. **265**, 14618–14626
48 Lamph, W. W., Dwarki, V. J., Ofir, R., Montminy, M. and Verma, I. M. (1990) Proc. Natl. Acad. Sci. U.S.A. **87**, 4320–4324
49 Costa, M. and Medcalf, R. L. (1996) Eur. J. Biochem. **237**, 532–538
50 Boise, L. H., Petryniak, B., Mao, X., June, C. H., Wang, C. Y., Lindsten, T., Bravo, R., Kovary, K., Leiden, J. M. and Thompson, V. (1993) Mol. Cell. Biol. **13**, 1911–1919
51 Mukai, F., Mitani, F., Shimada, H. and Ishimura, Y. (1995) Mol. Cell. Biol. **15**, 6003–6012
52 Hai, T. and Curran, T. (1991) Proc. Natl. Acad. Sci. U.S.A. **88**, 3720–3724
53 Hirai, S. I., Ryseck, R. P., Mechta, F., Bravo, R. and Yaniv, M. (1989) EMBO J. **8**, 1433–1439
54 Hoeffler, J. P., Lustbader, J. W. and Chen, C. Y. (1991) Mol. Endocrinol. **5**, 256–266
55 Xu, G. and Goodridge, A. G. (1996) J. Biol. Chem. **271**, 16008–16019

Received 9 January 1997/10 June 1997; accepted 7 July 1997