Analysis of DNase-I-hypersensitive sites at the 3' end of the cystic fibrosis transmembrane conductance regulator gene (*CFTR*)

Hugh N. NUTHALL*, Danielle S. MOULIN*, Clare HUXLEY† and Ann HARRIS*¹

*Paediatric Molecular Genetics, Institute of Molecular Medicine, Oxford University, John Radcliffe Hospital, Oxford OX3 9DS, U.K., and

†Section of Molecular Genetics, Division of Biomedical Sciences, Imperial College School of Medicine, London SW7 2AZ, U.K.

The cystic fibrosis transmembrane conductance regulator gene (*CFTR*) exhibits a complex pattern of expression that shows temporal and spatial regulation, although the control mechanisms are not fully known. We have mapped DNase-I-hypersensitive sites (DHSs) flanking the *CFTR* gene with the aim of identifying potential regulatory elements. We previously characterized DHSs at -79.5 and -20.9 kb with respect to the *CFTR* translational start site and a regulatory element in the first intron of the gene at $185+10$ kb. We have now mapped five DHSs lying 3' to the *CFTR* gene at $4574+5.4$, $+6.8$, $+7.0$, $+7.4$ and $+15.6$ kb that show some degree of tissue specificity. The

DHSs are seen in chromatin extracted from human primary epithelial cells and cell lines; the presence of the $+15.6$ kb site is tissue-specific in transgenic mice carrying a human *CFTR* yeast artificial chromosome. Further analysis of the $4574+15.6$ kb DHS implicates the involvement of CCAAT-enhancer-binding protein (C/EBP), cAMP-response-element-binding protein (CREB)/activating transcription factor (ATF) and activator protein 1 (AP-1) family transcription factors at this regulatory element.

Key words: DHS, regulation of expression.

INTRODUCTION

The cystic fibrosis transmembrane conductance regulator gene (*CFTR*) [1] is expressed at significant levels in specialized epithelial cells within the pancreas, intestine, bile ducts, urogenital tract, sweat gland and certain regions of the respiratory system [2–5]. The *CFTR* gene shows temporal regulation during human foetal development, although adult patterns of expression are established in many epithelial cells as early as the 12th week of gestation. However, in the lung epithelium, expression levels are much higher *in utero* and a substantial down-regulation of *CFTR* expression occurs after birth [6–8]. Analysis of rodent tissues by *in situ* hybridization have shown gradients of *CFTR* expression in the intestine along both the crypt–villus and proximal–distal axes [9]. *CFTR* expression is also hormonally regulated in the rodent uterus and shows stage-specific expression during spermatogenesis [10]. Hence the *CFTR* gene exhibits a complex, tightly regulated pattern of expression, yet relatively little is known of the complex mechanisms underlying this.

Key tissue-specific control elements are often located in the basal promoter of a gene. Analysis of both the human and mouse *CFTR* gene promoter regions produced inconsistent results, though the results established that the *CFTR* promoter is inefficient in driving reporter genes in assay systems *in vitro* [11–14]; 5' deletion analyses revealed that sequences between -357 and -33 bp upstream of the translation start site contained minimal promoter elements capable of driving basal reporter gene transcription in all studies. Chou et al. [12] reported the presence of a negative element between positions -477 and -409 bp that repressed reporter gene transcription, although this element was not identified in other studies.

Several elements that might be involved in the regulation of transcription are located immediately 5' to the *CFTR* gene.

These include several potential binding sites for the ubiquitous transcription factor Sp1 and other putative binding sites for activator protein 1 (AP-1). There is evidence for the regulation of *CFTR* gene expression *in vitro* by phorbol esters, a group of compounds that are known to induce AP-1 activity [15]. There is a variant cAMP-response element (CRE) TGACaTCA (cfCRE) within the putative basal promoter, at -19 bp upstream; this and a Y-box (inverted CCAAT) element, at -131 bp upstream of the translation start site have both been shown to be functional [16,17]. Importantly, the DNA sequence between -131 and -112 , containing both the Y-box and variant cfCRE elements, lies within a 34 bp sequence that is conserved in mouse, rat and human *CFTR* promoters [18]. However, none of these promoter elements can account for the tissue-specific pattern of *CFTR* expression.

Analysis of chromatin 5' to the *CFTR* gene, between -3000 bp relative to the transcription start site and $+100$ bp into intron 1, revealed the presence of several DNase-I-hypersensitive sites (DHSs) that showed some correlation with *CFTR* expression in cell lines [11,13]. These sites are often associated with regulatory elements. Further analyses of chromatin structure in both long-term cell lines and primary genital duct epithelial cells identified two DHSs at -79.5 kb and -20.5 kb 5' to the *CFTR* translation start codon that did not show a strong correlation with *CFTR* gene expression [19]. A DHS was also found within intron 1 of the *CFTR* gene that correlates well with the levels of expression of the *CFTR* gene in primary cells and long-term cell lines with known levels of *CFTR* expression [20].

Regulatory elements have been characterized downstream of many human genes, so we extended our search for tissue-specific DHSs to DNA lying 3' to the *CFTR* gene. We identified a cluster of four DHSs at $4574+5.4$ – 7.4 kb and a single DHS at $4574+15.6$ kb that were seen only in epithelial cells. The ap-

Abbreviations used: AP-1, activator protein 1; ATF, activating transcription factor; CF, cystic fibrosis; *CFTR* gene, cystic fibrosis transmembrane conductance regulator gene; C/EBP, CCAAT-enhancer-binding protein; CRE, cAMP-response element; CREB, cAMP-response-element-binding protein; CREM, cAMP-responsive element modulator; DHS, DNase-I-hypersensitive site; EMSA, electrophoretic mobility-shift assay; ER, oestrogen receptor; PR, protected region; YAC, yeast artificial chromosome.

¹ To whom correspondence should be addressed (e-mail aharris@molbiol.ox.ac.uk).

pearance of these DHSs did not correlate fully with *CFTR* expression, although they might contain elements that contribute to the tissue specificity of expression.

We evaluated the 4574 + 15.6 kb DHS *in vivo* in transgenic mice. A yeast artificial chromosome (YAC) containing the intact human *CFTR* gene (37AB12) [21] was previously introduced into transgenic mice [22]. The transgenic mice were bred with the 'Cambridge' cystic fibrosis (CF)-null mice to produce mice that only expressed the human *CFTR* transgene [22]. These mice had a normal phenotype, in contrast with the CF-null mice, suggesting that the 37AB12 YAC contained regulatory elements to direct *CFTR* expression in many of the usual sites. Expression of the human transgene in these mice showed cell-type-specific expression, equivalent to that of the endogenous mouse gene, in the intestine and salivary glands. However, it was not observed in some tissues in which human *CFTR* expression would have been expected. In sites in which there was a difference in *CFTR* expression patterns between mouse and humans, the transgene expression followed the mouse pattern. The YAC contains 310 kb of genomic DNA, including 200 kb spanning the coding region of the human *CFTR* gene and 55 kb of upstream flanking sequence [23], hence it is likely to contain at least 50 kb of sequence 3' to the *CFTR* gene. We evaluated the presence of the 4574 + 15.6 kb DHS in chromatin extracted from the YAC transgenic mouse tissues. Appearance of the 4574 + 15.6 kb DHS correlated with *CFTR* expression in several tissues from these transgenic mice. These results suggest that the element found at this site is important in regulating *CFTR* gene expression *in vivo*.

The 4574 + 15.6 kb DHS was analysed further and specific DNA-protein complexes were identified within this region by DNase I footprinting and electrophoretic mobility-shift assays (EMSAs). EMSA and supershift assays identified CCAAT-enhancer-binding protein (C/EBP), cAMP-response-element-binding protein (CREB) and activating transcription factor 1 (ATF1) as transcription factors potentially involved at this DHS.

EXPERIMENTAL

Cell culture

The following cell lines were used: Caco2 [24], HT29 [25], Capan1 [24], Calu3 [26], Panc1 [27] and MCF7 [28] were cultured in DMEM; the lymphoblastoid cell line 37566 was cultured in RPMI 1640; primary foetal vas deferens and epididymis cells [29]

and the simian virus 40 Ori-transformed vas deferens (RVP) and epididymis (REP) cell lines [30] were cultured in CMRL1066. Adult lung was normal tissue including bronchioles, from a site adjacent to a lung carcinoma. Foetal lung was from a 12.5-week foetus.

Reverse-transcriptase-mediated PCR (RT-PCR)

All cell lines studied were tested for *CFTR* mRNA expression by RT-PCR at the time of isolation of nuclei for chromatin analysis. RNA was isolated [31] and RT-PCR was performed as described previously [19].

Generation of probes

Cosmid clones A0726, C0155 and N1536, containing the 3' region of the *CFTR* gene, were identified from a chromosome 7 Lawrist 4 cosmid library [32] (Figure 1) by using the H-probe from the 3' untranslated region of *CFTR* [33]. Relevant cosmid clones were obtained from the Reference Library Database (Max Planck Institute for Molecular Genetics, Berlin, Germany). Restriction fragments of cosmid C0155 were sub-cloned into pBluescript(SK) (InVitrogen) and the inserts from the sub-clones were gel-purified, gene-cleaned and used as probes. The positions of the EH0.9 (0.9 kb *EcoRI/HindIII*), XE1.4 (1.4 kb *XbaI/EcoRI*), EN1.5 (1.5 kb *EcoRI/NheI*), ES1.2 (1.2 kb *EcoRI/StyI*), EB1.6 (1.6 kb *EcoRI/BamHI*) and BE1.3 (1.3 kb *EcoRI/BamHI*) subfragments of cosmid C0155 and the 0.9 kb H-probe are shown in Figure 1. Phage clones TE22 and TE27 [1] were kindly donated by Johanna Rommens.

Amplification and cloning of fragments spanning the 4574 + 15.6 kb DHS

Three fragments were generated spanning the 4574 + 15.6 kb DHS (see Figure 5). 16PQ is a 331 bp fragment that was amplified with primers HUG16P and HUG16Q; 16RS is a 311 bp fragment that was amplified with primers HUG16R and HUG16S, and 16TU is a 333 bp fragment that was amplified with primers HUG16T and HUG16U (Table 1). PCR primers were designed to have either *BamHI* (forward primers) or *XhoI* (reverse primers) restriction sites at their 5' ends for strand-specific labelling. The PCR products were cloned into pCR2.1 (InVitrogen) and sequenced to exclude PCR errors.

Figure 1 Genomic map of the 3' end of the *CFTR* gene

Lines above the gene show the inserts of cosmid clones A0726, C0155 and N1536 and the bacteriophage clones TE22 and TE27 [1]. The positions of the DHSs are indicated by vertical arrows. Below the gene, the map of the region analysed for the presence of DHSs is shown in more detail. The positions of restriction sites for *BamHI*, *EcoRI*, *HindIII*, *XbaI* and *ClaI* are indicated. The scale is in kilobases downstream of the last base of the *CFTR* coding region in exon 24. The positions of DNA probes used for restriction mapping and in the analysis of DHSs are shown.

Table 1 PCR primers and conditions for amplification of fragments spanning the 4574 + 15.6 kb DHS from genomic DNA

The bold nucleotides represent additional sequences at the 5' end of the primer that contained restriction enzyme sites for *Xho*I (P, R and T) or *Bam*HI (Q, S and U).

Primer pair 5' and 3'	Primer sequence (5' → 3')	Annealing temperature (°C)	Extension time (min)	Product size (bp)
HUG16P	CCCTCGAG GAGTAGCTGGTATTACAGGTGTAC	64	1	331
HUG16Q	CGGGATCC TGAGGATTGTTTATGATTGGATC			
HUG16R	CCCTCGAG AACCTCTTTATGGAATCTCC	58	1	311
HUG16S	CGGGATCC AGAACCTAGACAGACTCCTG			
HUG16T	CCCTCGAG TTTGGTTGTTACTTATCTTGAC	58	1	333
HUG16U	CGGGATCC CGTAGCTTCACAGCAAGATG			

Assay of DNase I-hypersensitivity

Chromatin from a panel of cell lines was probed for DNase-I-hypersensitive regions by standard methods [34]. Nuclei were treated in parallel aliquots by digestion with DNase I [0, 15, 30, 60 and 120 units of DNase I (FPLCpure; Pharmacia; 1 unit is equal to approx. 0.3 Kunitz units; 1 unit of DNase I will degrade 1 µg of pBR322 DNA in 10 min at 37 °C in 20 µl of assay buffer) for a constant duration]. DNase-I-treated DNA samples were then cleaved with relevant restriction enzymes, resolved on 0.8 % agarose gels and blotted to Hybond N membranes. Probing of membranes was performed by standard methods.

DNase I footprinting *in vitro*

*Xho*I/*Eco*RI-restricted or *Bam*HI/*Eco*RI-restricted DNA fragments were labelled at one end by fill-in reactions with Klenow DNA polymerase and [α -³²P]dCTP. Nuclear extracts were prepared from cell lines and DNase I footprinting was performed as described previously [20,35].

EMSA

EMSA analysis was performed as described previously [20]. DNA fragments 163 and 164, corresponding to protected region (PR)3 and PR4 respectively at DHS 4574 + 15.6 kb (see Figure 7A) were labelled by fill-in reactions with Klenow DNA polymerase and [α -³²P]dCTP. The consensus C/EBP β oligonucleotide was generated from two oligonucleotides, 5'-AGCTTTAAGA TTGTGCAATGTGACG-3' and 3'-AATCTAACACGTTA CACTGCTCGA-5' (where the C/EBP β consensus is underlined), and labelled as above. Consensus CREB1 (sc-2504) and AP-1 (sc-2501) oligonucleotides were purchased from Santa Cruz (Santa Cruz, CA, U.S.A.) and end-labelled with T4 polynucleotide kinase and [γ -³²P]dATP. Supershift assays were performed by standard methods. Antibodies specific for C/EBP (sc-746), CREB1 (sc-271), ATF1 (sc-243), CREB1/cAMP-responsive element modulator (CREM)/ATF (sc-270), Jun family (sc-44) and Fos family (sc-253) were purchased from Santa Cruz.

Analysis of 4574 + 15.6 kb DHS in transgenic mice

Details of the generation of the mice carrying the 37AB12 YAC and crosses with the 'Cambridge' null mice were as described previously [22]. The T30 mice used in this study were homozygous for the YAC transgene and for the CF knockout. Mouse tissues were collected immediately after death and processed directly for chromatin extraction as described previously [23]. For most tissues, chromatin was extracted from the whole organ. For the

small intestine the samples were enriched in epithelial cells by washing the contents from the intestine, opening it by cutting along its length, and scraping the surface with a microscope slide.

RESULTS

Expression of *CFTR* mRNA

Expression of *CFTR* was measured by RT-PCR in all the cell cultures used in this study at the time that nuclei were isolated for hypersensitivity assays [19]. Although this assay does not provide an accurate quantitative measure of levels of *CFTR* expression, in the presence of an internal control for a housekeeping gene such as glucocerebrosidase it provides a useful gross measure of *CFTR* mRNA. RT-PCRs for each cell type are included in Table 2.

DHSs

To evaluate the efficacy of DNase I treatment, every batch of DNase-I-treated nuclei was assayed with the RA2.2 probe, which detects a constitutive DHS in the α -globin gene cluster in all cell types [36].

A 40 kb length of genomic DNA, extending from the *Bam*HI site at the 3' end of exon 22 (55205 in HSAC000061) to the *Bam*HI site approx. 38 kb downstream of the end of exon 24 (13936 in HSAC0004240), was screened for the presence of DHSs by using cosmid restriction fragments as probes (see Figure 1). This search revealed five regions of chromatin that were hypersensitive to treatment with DNase I.

Cluster of DHSs 5.4–7.4 kb downstream of exon 24 of the *CFTR* gene

DNase-I-treated chromatin from Caco2 cells was digested with *Eco*RI and probed with DNA fragments EH0.9 and XE1.4 (see Figure 1). Figure 2(A) shows a map of the region and the locations of probes and the DHS. In Figure 2(B), the 0.9 kb *Eco*RI/*Hind*III restriction fragment (EH0.9) hybridized to an *Eco*RI restriction fragment of approx. 9.4 kb. Subfragments of approx. 5.8, 5.4 and 3.8 kb can be seen in lanes 2–5 with increasing digestion with DNase I. In Figure 2(C), the 1.4 kb *Xba*I/*Eco*RI restriction fragment (XE1.4) hybridized to the same *Eco*RI restriction fragment of approx. 9.4 kb. Subfragments of approx. 5.6 kb (DHS A), 4 kb (DHS C) and 3.6 kb (DHS D) can be seen in lanes 2–5 with increasing digestion with DNase I.

These results show DHSs downstream of the *CFTR* gene. DNase I-treated nuclei from a range of cell lines were probed with XE1.4 to evaluate whether these DHSs were tissue-specific. The tissue origin of each cell line is shown in Table 2. In Figures

Table 2 Evaluation of the 4574 + 5.4–7.4 kb DHS cluster and the 4574 + 15.6 kb DHS

The levels of *CFTR* expression are gross estimates from RT–PCR.

Cell line	Cell type/origin	Level of <i>CFTR</i> expression (+/–)	Presence of DHS				
			+ 5.4 kb	+ 6.8 kb	+ 7.0 kb	+ 7.4 kb	+ 15.6 kb
Caco2	Colon carcinoma	+++	+	–	+	+	++
HT29	Colon carcinoma	+++	+	–	+	–	++
Capan1	Pancreatic adenocarcinoma	++	+	–	+	–	+
Calu3	Lung adenocarcinoma	+++	–	–	+	–	–
MCF7	Mammary carcinoma	–	+	–	+	–	+
Panc1	Pancreatic adenocarcinoma	–	+	–	+	–	++
37566	Lymphoblastoid	–	–	–	–	–	–
Primary vas deferens	Vas deferens	++	–	+	+	–	++
RVP	Transformed vas	+ / –	–	–	+	–	+
Primary epididymis	Epididymis	++	–	+	+	–	++
REP	Transformed epididymis	+ / –	–	–	+	–	+
	Foetal lung tissue sample	*	+	–	+	–	–
	Adult lung tissue sample	*	+	+	+	+	++

* *CFTR* mRNA expression was not analysed in these samples.

2(D)–2(F) the XE1.4 probe hybridized to an *EcoRI* fragment of 9.4 kb in chromatin from the HT29 (Figure 2D) Capan1 (Figure 2E) and MCF7 (Figure 2F) cell lines. Subfragments of 5.6 and 4.0 kb can be seen in lanes 2–5 (with increasing digestion with DNase I) in all cell lines. In Figures 2(G)–2(I), a subfragment of 4.0 kb can be seen in lanes 2–5 (with increasing digestion with DNase I) in the primary epididymal cells (Figure 2G) and REP (transformed epididymal) (Figure 2H) and Calu3 (Figure 2I) cell lines. Another subfragment of 4.2 kb (DHS B) can be seen in lanes 2–5 in the primary epididymal cells (Figure 2G). Identical subfragments were detected in primary vas deferens and primary epididymal cells (results not shown). In addition, the transformed vas deferens cell line RVP and the REP line showed identical profiles (results not shown). Chromatin from both adult and foetal lung tissue also showed the 4.0 kb subfragment; other subfragments of 5.6, 4.2 and 3.6 kb were evident in chromatin from adult lung (results not shown). No subfragments are apparent in the lymphoblastoid 37566 cell line (Figure 2J) despite the samples revealing a DHS with the RA2.2 probe in the α -globin locus (results not shown).

The 4.0 kb sub-band was present in all epithelial cell lines analysed, irrespective of their status with respect to *CFTR* expression. The 3.6 kb sub-band is clearly present in the colon carcinoma cell line Caco2 and adult lung tissue. The 5.6 kb sub-band is present in Caco2, HT29, Capan1, Panc1 (results not shown) and MCF7 cell lines and in chromatin from foetal and adult lung. A sub-band of 4.2 kb is apparent only in primary cultures of foetal male genital duct epithelial cells and chromatin from adult lung. The results for the different cell lines and tissue samples tested are summarized in Table 2.

As shown in Figure 2(A), the two sets of subfragments position the DHSs at the following locations downstream of the 5' *EcoRI* site: DHS A, 3.8 kb; DHS B, 5.2 kb; DHS C, 5.4 kb; DHS D, 5.8 kb. The DHSs are referred to in kilobases from the last base of the coding sequence of the *CFTR* gene, base number 4574, consistent with CF Genetic Analysis Consortium nomenclature (<http://www.genet.sickkids.on.ca>): hence the 4574 + 5.4 kb DHS (HSAC000061, bases 62735–63135), 4574 + 6.8 kb DHS (HSAC000061, bases 64235–64435), 4574 + 7.0 kb DHS (HSAC000061, bases 64435–64635) and 4574 + 7.4 DHS (HSAC000061, bases 64735–65135) (Figure 1).

DHS 15.6 kb downstream of exon 24 of the *CFTR* gene

To search for the presence of DHSs further downstream of the 4574 + 7.4 kb DHS, DNase-I-treated nuclei from Caco2 cells were digested with the restriction enzyme *XbaI* and probed with XE1.4, which hybridized to a restriction fragment of approx. 9.4 kb. The detection of subfragments of approx. 6.5 and 5.6–6.0 kb with increasing amounts of DNase I (results not shown) indicated the presence of a DHS located approx. 15.6 kb downstream of exon 24 of the *CFTR* gene. The position of the site was confirmed by using the 1.5 kb *EcoRI/NheI* fragment (EN1.5) to probe *EcoRI* digests of DNase I-treated chromatin.

The restriction map of the 4574 + 15.6 kb region is shown in Figure 3(A). In Figure 3(B), EN1.5 is shown to hybridize to an *EcoRI* restriction fragment of approx. 6 kb in Caco2 chromatin. A diffuse subfragment of approx. 4.2–4.6 kb can be seen in lanes 2–5 (with increasing amounts of DNase I). DNase-I-treated nuclei from a number of cell lines and primary genital duct cells were digested with *EcoRI* and probed with EN1.5 or the 1.2 kb *EcoRI/StyI* fragment (ES1.2; see Figure 1) to assess whether the DHSs displayed any tissue-specific properties. In Figure 3, EN1.5 hybridized to a 6 kb *EcoRI* fragment in all cell lines and to a subfragment of 4.2–4.6 kb in lanes 2–5 (with increasing amounts of DNase I) in the HT29 (Figure 3C) and Panc1 (results not shown) cell lines, in primary epididymal cells (Figure 3F) and at a lower level in Capan1 (Figure 3D), MCF7 (Figure 3E) and the REP cell line (Figure 3G). Identical subfragments were detected in primary vas deferens and primary epididymal cells; in addition, RVP and the REP line had identical profiles (results not shown). No subfragments were apparent in the lymphoblastoid 37566 cell line (Figure 3H) or Calu3 (Figure 3I, probed with ES1.2) despite the samples' revealing a DHS with the RA2.2 probe in the α -globin locus (results not shown). The 4.2–4.6 kb subfragment was seen in chromatin prepared from adult lung tissue but not from foetal lung tissue (results not shown). The results for the cell lines tested are summarized in Table 2. To summarize, the 4574 + 15.6 kb DHS was seen in epithelial cell lines that express *CFTR* (Caco2, HT29 and Capan1) and also in lines that do not (MCF7 and Panc1), but not in the lymphoblastoid cell line 37566 or the *CFTR*-expressing Calu3 airway epithelial line. The intensity of the DHS was correlated with *CFTR* expression levels

Figure 2 Detection of DHSs at 4574 + 5.4–7.5 kb

(A) Map of region. Arrows above the line denote the DHS positions. Restriction sites shown above the line are: E, *EcoRI*; H, *HindIII*; X, *XbaI*. Below the line are shown the probes (grey rectangles) and the subfragments produced by DHSs. (B–J) Southern blots of DNase-I-digested chromatin cleaved with *EcoRI* and hybridized with DNA probes. DNA size markers were a 1 kb ladder (Gibco-BRL). In each panel, lane 1 shows DNA prepared from nuclei that had not been treated with DNase I, and lanes 2–5 show DNA prepared from nuclei treated with increasing amounts of DNase I: lane 2, 15 units; lane 3, 30 units; lane 4, 60 units; lane 5, 120 units. (B) Caco2 chromatin probed with the EH0.9 probe. EH0.9 hybridized to an *EcoRI* restriction fragment of approx. 9.4 kb and subfragments of approx. 5.8, 5.4 and 3.8 kb can be seen in lanes 2–5. (C) Caco2 chromatin probed with XE1.4. An *EcoRI* restriction fragment of approx. 9.4 kb and subfragments of approx. 5.6, 4.0 and 3.6 kb can be seen in lanes 2–5. (D–F) Chromatin from HT29 cells (D), Capan1 cells (E) and MCF7 cells (F) probed with XE1.4. An *EcoRI* restriction fragment of approx. 9.4 kb and subfragments of approx. 5.6 and 4.0 kb can be seen in lanes 2–5 in these cell lines. (G–J) Chromatin from primary epididymal cells (G), REP cells (H), Calu 3 cells (I) and 37566 cells (J). In each case XE1.4 hybridized to an *EcoRI* restriction fragment of approx. 9.4 kb. A subfragment of approx. 4.0 kb can be seen in lanes 2–5 in the primary epididymal, REP and Calu 3 cells. A subfragment of 4.2 kb is also present in lanes 2–5 of the primary epididymal cells. No subfragments were seen with the 37566 cell line.

Figure 3 Detection of a DHS at 4574 + 15.6 kb

(A) Map of region. Above the line the arrow and bracket denote the DHS position. Restriction sites shown above the line are: E, *EcoRI*; B, *BamHI*; X, *XbaI*. Below the line are shown the probes (grey rectangles) and subfragments produced by DHS. (B–J) Southern blots of DNase-I-digested chromatin cleaved with *EcoRI* and hybridized with DNA probes. DNA size markers were a 1 kb ladder (Gibco-BRL). In each panel, lane 1 shows DNA prepared from nuclei that had not been treated with DNase I, and lanes 2–5 show DNA prepared from nuclei treated with increasing amounts of DNase I: lane 2, 15 units; lane 3, 30 units; lane 4, 60 units; lane 5, 120 units. (B) Caco2 chromatin probed with EN1.5, which hybridized to an *EcoRI* restriction fragment of approx. 6 kb and diffuse subfragments of approx. 4.2–4.6 kb in lanes 2–5, indicating the presence of a DHS at 4574 + 15.6 kb. (C–H) Chromatin from HT29 (C), Capan1 (D), MCF7 (E), primary epididymis (F), REP (G) and 37566 (H) probed with EN1.5. An *EcoRI* restriction fragment of approx. 6 kb and subfragments of approx. 4.2–4.6 kb can be seen in lanes 2–5 in chromatin from all cells except 37566, although the relative intensity of the subfragment varies. (I) Calu3 chromatin probed with ES1.2, which hybridized to the same 6 kb *EcoRI* restriction fragment; however, no subfragments are seen.

Figure 4 Detection of a DHS at 4574 + 15.6 kb in transgenic mice

(A) Map of the region. (B–E) Southern blots of DNase-I-digested chromatin from transgenic mouse tissue samples digested with *Eco*RI and hybridized with probe ES1.2. DNA size markers were a 1 kb ladder (Gibco-BRL). For each panel, lane 1 shows DNA prepared from nuclei that had not been treated with DNase I, and lanes 2–5 show DNA prepared from nuclei treated with increasing amounts of DNase I: lane 2, 15 units; lane 3, 30 units; lane 4, 60 units; lane 5, 120 units. The probe ES1.2 hybridized to an *Eco*RI restriction fragment of approx. 6.0 kb in all panels and to a subfragment of approx. 4.4 kb in lung and pancreas only.

in primary and transformed genital duct epithelial cells and colon carcinoma cell lines.

With the use of the XE1.4 fragment to probe *Xba*I-digested Caco2 chromatin samples, a subfragment of approx. 5.6–6.0 kb and a second subfragment of approx. 6.5 kb were seen (results not shown). With the use of the EN1.5 or derivative ES1.2 fragments to probe *Eco*RI-digested samples, a diffuse subfragment was seen between 4.2 and 4.6 kb. A distance 5.6–6.0 kb downstream of the *Xba*I site corresponded to 4.2–4.6 kb downstream of the *Eco*RI site. Hence the DHS spanned a region that corresponded to bases 72840–73240 of HSAC000061. The 3' end of exon 24 of the *CFTR* gene corresponded to base 57460, so the DHS lies 15.4–15.8 kb downstream of the *CFTR*. The DHS is called the 4574 + 15.6 kb DHS, where 4574 is the last base in the *CFTR* coding sequence.

Figure 5 Map and sequence of DHS 4574 + 15.6 kb region

(A) The locations of the DHSs downstream of the *CFTR* gene are shown by arrows; the scale is in kb from the 3' end of exon 24. The region spanning the DHS 4574 + 15.6 kb, which was cloned as three 300 bp fragments, is also shown. (B) Sequence of 800 bp spanning the DHS, indicating the primers used for cloning overlapping fragments PQ, RS and TU. Forward primers were P, R and T and reverse primers were Q, S and U.

With the use of the 1.6 kb *Eco*RI/*Bam*HI (EB1.6) and 1.3 kb *Bam*HI/*Eco*RI (BE1.3) probes (see Figure 1) to search for DHSs in the 20 kb downstream of the 4574 + 15.6 kb DHS in Caco2 cells (38 kb 3' to exon 24), no additional sites were found (results not shown).

Analysis *in vivo* in transgenic mice

Figure 4 shows the evaluation of the 4574 + 15.6 kb DHS in chromatin from transgenic mice tissues. The ES1.2 probe detected an *Eco*RI fragment of approx. 6 kb; a subfragment of approx. 4.4 kb was seen with increasing digestion with DNase I (lanes 2–5) in transgenic mouse lung and pancreas (Figures 4B and 4C) and small intestine (results not shown) but not in transgenic mouse kidney (Figure 4D), liver (Figure 4E) or spleen (results not shown). Results presented elsewhere ([22], and D. S. Moulin, A. Manson, C. Huxley, H. N. Nuthall, D. J. Smith and A. Harris, unpublished work) have shown, by mRNA *in situ* hybridization and RT-PCR, that the YAC transgene is expressed in lung, pancreas and small intestine but not in liver or kidney.

DNase I footprint analysis of the 4574 + 15.6 kb DHS *in vitro*

Because the 4574 + 15.6 kb DHS showed some tissue specificity, genomic DNA spanning the DHS was analysed for DNA–protein interactions. Primers for the amplification of genomic DNA spanning the 4574 + 15.6 kb DHS were generated from the *CFTR* genomic sequence (HSAC000061) (Figure 5). Three overlapping fragments of approx. 300 bp covering 800 bp in total were generated by PCR amplification of genomic DNA. The three

Figure 6 DNase I footprinting *in vitro* of the PQ and RS fragments from DHS 4574 + 15.6 kb

(A) Sequence of the PQ-RS fragments showing PRs. (B, C) DNase I footprinting reactions. PRs are represented by solid bars to the right of each panel; triangles indicate hypersensitive sites. Areas of decreased digestion are indicated with dotted lines. (B) PQ (negative strand). Lane 1, AG ladder; lanes 2 and 5, no nuclear extracts; lanes 3 and 4, 20 and 40 μ g of nuclear extract from Caco2 cells respectively. (C) RS (negative strand). Lane 1, AG ladder; lanes 2 and 5, no nuclear extracts; lanes 3 and 4, 20 and 40 μ g of nuclear extract from Caco2 cells respectively.

fragments were called HUG16PQ, HUG16RS and HUG16TU, and are referred to here as PQ, RS and TU.

DNase I footprinting was performed on each labelled fragment with nuclear extracts from the colon carcinoma cell line, Caco2. Multiple PRs were identified throughout the DNA sequence spanned by the fragments PQ and RS, as shown in Figure 6. Some were visible as strongly protected regions, bound by hypersensitive sites (PR3, PR5, PR7, PR10, PR11, PR12, PR13, PR14 and 15; Figures 6B and 6C). Other regions showed protection, with a hypersensitive site at only one end (PR2, PR6, PR7, PR8 and PR9; Figures 6B and 6C). Several regions displayed an extended region of protection between hypersensitive sites (PR4, PR6 and PR16) that might have resulted from the binding of more than one factor, because the footprints contained poorly protected regions and some contained sites that seemed more sensitive to digestion with DNase I (PR4).

DNase I footprinting reactions *in vitro* were performed with nuclear extracts from several cell lines to determine whether any of the PQ and RS footprints were specific to certain cell types. Identical patterns of protection were seen with nuclear extracts from the *CFTR*-expressing colon carcinoma cell lines, Caco2 and HT29, and the lung carcinoma cell line, Calu3, even though Calu3 does not show the DHS. The lymphoblastoid cell line 37566, which does not express *CFTR*, did not produce any regions of protection (results not shown).

Experiments with the TU fragment, labelled on each strand and incubated with nuclear extracts from Caco2 cells, did not show any regions of protection or bases that seemed hypersensitive to digestion with DNase I.

EMSA analysis

Analysis of the DNA sequence for potential binding motifs of transcription factors (TRANSFAC) in the footprinted regions PQ and RS revealed motifs of potential importance in PR3 (C/EBP), PR4 [CRE, AP-1 and oestrogen receptor (ER)], PR9 (COUP and HNF4), PR10 (HNF3), PR14 (C/EBP) and PR16 (AP-1). The cluster of transcription-factor-binding motifs in PR3 and PR4, including potential binding sites for C/EBP, CRE and AP-1, warranted further evaluation. Figure 7(A) shows the sequence of the footprints PR3 and PR4 in the PQ DNase I footprinting reaction. Probes 163, which includes PR3, and 164, which includes PR4, are also shown, as are consensus sites for C/EBP, ER, CRE and AP-1. Figure 7(B) shows an EMSA reaction with 163 as a probe and nuclear extracts from Caco2 in which two complexes were generated (indicated by arrows at the left of Figure 7B). The slower-migrating complex was competed for efficiently by the C/EBP β and 163 oligonucleotides but not by an irrelevant oligonucleotide (results not shown). The faster-migrating complex was competed for by the 163 oligonucleotide but not the C/EBP β oligonucleotide, which stabilized the complex. The contribution of C/EBP β to the 163 complexes was confirmed in supershift reactions with an antibody specific for C/EBP. Figure 7(B) shows supershifting by the C/EBP antibody of the complexes formed between a labelled C/EBP β consensus oligonucleotide and Caco2 nuclear extracts. The complexes formed between the 163 probe and Caco2 nuclear extracts show an equivalent supershift with the C/EBP antibody. Both supershifts are indicated by arrows at the right of Figure 7(B).

1634 PR3 PR4
 GACAGGGTTTCGCCATGTTGGCCAGGCTGGTCTTGAAGTCCTGACCTCATGATCCAC
 CTGTCCCAAAGCGGTACAACCGGTCCGACCAGAACTTCAGGACTGGAGTACTAGGTG

 163 GACAGGGTTTCGCCATGTTGGCCAGGC
 CTGTCCCAAAGCGGTACAACCGGTCCG

cAAAGCGGT (C/EBP)
 (3'-KAANGNNAT-5' consensus)

 164 TGGTCTTGAAGTCCTGACCTCATGATCC
 ACCAGAACTTCAGGACTGGAGTACTAGG

(CRE) TGACcTCA
 (consensus 5'-TGACGTCA-3')

(AP-1 family) TGACcTCA
 (consensus 5'-TGAC TCA-3')

(ER) tGaagTCCTGACC
 (consensus 5'-GGTCANNNTGACC-3')

A.

Figure 7 EMSAs of the PR3 and PR4 regions of the PQ fragment

(A) Sequence of the PR3 and PR4 footprints and the oligonucleotides 163 and 164 used in EMSA. C/EBP, ER, CRE and AP-1 consensus sequences within 163 and 164 are shown. The nomenclature corresponds to IUBMB codes. (B) EMSA with 163 as probe and Caco2 nuclear extracts, showing specific DNA-protein complexes (indicated by arrows at the left), one of which is removed by competition with the 163 oligonucleotide and a C/EBP consensus oligonucleotide and the other by 163 only. Supershift reactions are shown with the C/EBP consensus oligonucleotide and 163 as probes, nuclear extracts from Caco2 cells and an antibody specific for C/EBP. An antibody-induced supershift of the DNA-protein complexes is seen for both probes (indicated by arrows at the right). (C, D) EMSA with 164 as probe and Caco2 nuclear extracts. (C) Specific DNA-protein complexes (indicated by arrows at the left) are partly removed by competition with 164. Competition with CREB and AP-1 consensus oligonucleotides removes the two slow-migrating complexes. (D) Supershift reactions with antibodies specific for CREB and ATF and antibodies that recognize CREB1, CREM and ATF-1, the Jun family and the Fos family showed some inhibition of the DNA-protein complexes between 164 and Caco2 nuclear extracts but did not abolish them completely.

Figure 7(C) shows EMSAs with 164 as a probe and nuclear extracts from Caco2 cells. Three specific complexes are formed (indicated by arrows at the left of Figure 7C) that are partially competed for by the 164 oligonucleotide but not by an irrelevant 33 bp oligonucleotide (results not shown). Competition with the 164 oligonucleotide in other experiments abolished these complexes completely (results not shown). Competition with consensus oligonucleotides for CREB and AP-1 abolished two of the complexes and stabilized the complex with the greatest mobility. Figure 7(D) shows supershift assays with 164 as a probe and antibodies against CREB and ATF-1 specifically, which do not cause a supershift but decrease the abundance of the complexes. These antibodies did not produce marked supershifts when used with consensus oligonucleotides for their respective recognition motifs and purified protein; the complexes were merely inhibited

(results not shown). A supershift reaction with an antibody that is reactive with ATF-1 p35, CREB-1 p43 and CREM-1 (results not shown) produced a similar profile to that seen with the ATF-1-specific antibody. Antibodies reactive with the Jun and Fos families of transcription factors also showed an inhibition of the complexes formed between Caco2 nuclear extracts and the 164 probe rather than a clear supershift.

DISCUSSION

Expression of the *CFTR* gene is regulated by complex mechanisms, which are poorly understood. It seemed probable that the regulatory elements that we had identified previously upstream of the *CFTR* gene ([19,23] and H. N. Nuthall, G. Vassaux, C. Huxley and A. Harris, unpublished work)

and in the first intron ([20]; T. L. McDowell, D. J. Smith, H. N. Nuthall, D. S. Moulin and A. Harris, unpublished work) could not completely account for the regulation of *CFTR* expression. We therefore extended our search for tissue-specific regulatory elements to the sequences downstream of the *CFTR* coding region. Approx. 40 kb of downstream sequence was screened and two clusters of DHSs were identified, one lying 5.4–7.4 kb downstream of the last base of the *CFTR* open reading frame (4574) and one sited 15.6 kb 3' to the last base, termed the 4574+15.6 kb DHS. The cluster of DHSs located between 5.4 and 7.4 kb downstream of the transcript contains four distinct DHSs that are termed the 4574+5.4 kb, 4574+6.8 kb, 4574+7 kb and 4574+7.4 kb DHSs. These sites do not correspond to the duodenum-specific DHSs recently described at the 3' untranslated region of the rat *CFTR* gene [37].

The DHSs at 4574+5.4–7.4 kb, which were initially identified in chromatin from the *CFTR*-expressing Caco2 colon carcinoma cell line, were subsequently evaluated in a number of cell types (Table 2). The 4574+5.4 kb and 4574+7 kb DHSs are seen in many epithelial cells, irrespective of their *CFTR* expression status, suggesting that the elements at these sites might enable *CFTR* expression but not cause it. Of most interest in terms of potential tissue-specific control elements was the 4574+6.8 kb DHS that was detected only in primary genital duct epithelial cells and adult lung. The DHSs in the 4574+5.4–7.4 kb cluster might have a number of roles in the regulation of *CFTR* expression. There is currently no information on the transcriptional termination of the *CFTR* mRNA and it would not be surprising if this group of DHSs were reflecting transcriptional termination sites.

The 4574+15.6 kb DHS was most prominent in chromatin from the Caco2 and HT29 colon carcinoma cell lines and primary genital duct epithelial cells, all of which express *CFTR* endogenously at high levels, and in adult lung. It was also weakly evident in the Capan1 pancreatic adenocarcinoma cell line and transformed genital duct epithelial cells, which express low levels of endogenous *CFTR*. However, the appearance of the 4574+15.6 kb DHS does not show a complete correlation with *CFTR* expression (Table 2).

Evidence for the importance of the 4574+15.6 kb DHS *in vivo* was provided by evaluation of this site in tissues from transgenic mice carrying the human *CFTR* gene on a YAC [22]. These mice have previously been shown to express the human *CFTR* gene in many of the tissues in which the rodent *CFTR* gene is normally expressed [22,23]. The transgene was expressed in small intestine, lung and pancreas but not in kidney, liver or spleen. The 4574+15.6 kb DHS was seen in chromatin extracted from transgenic mouse lung, pancreas and small intestine but was not detected in kidney, liver or spleen (results not shown). The tissue-specific appearance of the DHS therefore reflects the expression pattern of the human *CFTR* gene in these transgenic mice and provides support for an important role for the 4574+15.6 kb DHS in *CFTR* expression *in vivo*. Although the regulation of expression of rodent and human *CFTR* shows substantial divergence [14,18], it seems that the mouse expresses the necessary transcription factors to interact with the regulatory element located within the 4574+15.6 kb DHS.

The question arises of whether the 4574+15.6 kb DHS contains a regulatory element that functions in controlling *CFTR* expression or whether it is associated with a locus lying downstream. Although enhancers might function at distances of greater than 50 kb from a gene promoter, analysis of the chromosome 7 gene map (<http://www.genet.sickkids.on.ca>), suggests that there are no genes within 2 Mb of the 3' end of the *CFTR* gene. It is therefore likely that the 4574+15.6 kb DHS

functions in association with the *CFTR* locus. The availability of the 310 kb human *CFTR* YAC that is capable of driving wild-type expression levels of *CFTR* in human cells [38] and in transgenic mice [22,23] provides a useful tool for investigating this question. This YAC extends approx. 30 kb 3' to the *CFTR* gene, so YAC fragmentation experiments will enable further evaluation of the 4574+15.6 kb DHS *in vivo*.

Evidence for a role *in vivo* for the 4574+15.6 kb DHS justified further analysis *in vitro* of the regulatory element located at this site. The 800 bp flanking the site were evaluated by DNase I footprinting, which revealed a number of PRs. A search for transcription-factor-binding motifs within the PRs revealed sites in PR3 and PR4 that were of particular interest with respect to *CFTR* transcription. These included consensus sites in PR3 for C/EBP, and in PR4 for CRE, AP-1 and ER. *CFTR* is known to show hormonal regulation of expression during the oestrous cycle of rodents [10]. However, the ER site in 164 was a single half-site, with no sites in the adjacent DNA; it was therefore thought unlikely to be important and was not investigated further.

EMSA and gel supershift assays were done with oligonucleotides 163, which spanned PR3, and 164, which spanned PR4, and relevant transcription factor antibodies. Several complexes were generated between 163 and proteins in nuclear extracts from Caco2 cells. Competition with the C/EBP consensus oligonucleotide and supershift experiments with the 163 and C/EBP consensus oligonucleotides and an antibody specific for C/EBP showed that C/EBP was one of the proteins involved in the complex binding to 163. The C/EBPs are a family of transcription factors that have a key role in cellular differentiation. C/EBPs function in a variety of tissues including lung and intestine, sites in which mechanisms of spatial and temporal regulation of *CFTR* expression are likely to be functional. A potential role for C/EBP transcription factors in the expression of the *CFTR* gene would therefore not be unexpected.

The CRE and AP-1 consensus sites in oligonucleotide 164 were evaluated by EMSA with specific competitor oligonucleotides; both the CREB and AP-1 consensus oligonucleotides competed efficiently for two of the three sets of complexes. The CREB/ATF family of transcription factors, which function through binding to the CRE (TGACc/gTCA), can bind in homodimeric and heterodimeric forms. The heterodimers can be generated with other ATF/CREB proteins or with members of the AP-1 transcription factor family. AP-1 is formed by Jun homodimers or Fos/Jun heterodimers that recognize the sequence TGA(C/G)TGA. Supershift experiments with specific antibodies against CREB1, ATF1 and the Jun and Fos families did not cause clear supershifts but did decrease the abundance of the complexes. These results are likely to reflect the interaction of CREB1, ATF1, Jun or Fos family members with the consensus site in 164.

CFTR is a cAMP-activated Cl⁻ ion channel that is also regulated by cAMP at the transcriptional level. A variant CRE element in the basal promoter is conserved across species [18] and is required for *CFTR* promoter activity [16,17]. The detection of additional CRE motifs in *CFTR* regulatory elements would be consistent with cAMP's having an important role in *CFTR* transcription. In addition, phorbol esters, which induce AP-1 activity, are known to affect *CFTR* transcription [15] through undefined regulatory elements.

In conclusion, it is probable that the complete elucidation of the tissue-specific and temporal regulation of *CFTR* expression will be a complex task and that multiple interacting elements are involved. We have identified a cluster of potential regulatory elements between 4574+5.4 and 7.4 kb and an additional one at

4574+15.6 kb 3' of the *CFTR* gene. We previously identified potential control elements in the 5' flanking region of the *CFTR* gene, at -79.5 and -20.9 kb upstream of the ATG, and a regulatory element within the first intron at 185+10 kb. The distance between the -20.9 kb site and the DHS at 4574+15.6 kb is approx. 223 kb. To evaluate the integrated function of these widely spaced elements it will be necessary to analyse them *in vivo*, once they have been integrated into chromatin.

We thank Tarra McDowell and Irina Udalova for help, and Johanna Rommens for the TE22 and TE27 phage. This work was supported by the Cystic Fibrosis Trust, U.K., the Wellcome Trust and the Association Française de Lutte contre la Mucoviscidose. H.N.N. was the recipient of a Cystic Fibrosis Trust studentship and D.S.M. of a Wellcome Trust prize studentship.

REFERENCES

- Rommens, J. M., Iannuzzi, M. C., Kerem, B., Drumm, M. L., Melmer, G., Dean, M., Rozmahel, R., Cole, J. L., Kennedy, D., Hidaka, N. et al. (1989) *Science* **245**, 1059–1065
- Crawford, I., Maloney, P. C., Zeitlin, P. L., Guggino, W. B., Hyde, S. C., Turley, H., Gatter, K. C., Harris, A. and Higgins, C. F. (1991) *Proc. Natl. Acad. Sci. U.S.A.* **88**, 9262–9266
- Denning, G. M., Ostedgaard, L. S., Cheng, S. H., Smith, A. E. and Welsh, M. J. (1992) *J. Clin. Invest.* **89**, 339–349
- Kartner, N., Augustinas, O., Jensen, T. J., Naismith, A. L. and Riordan, J. R. (1992) *Nat. Genet.* **1**, 321–327
- Engelhardt, J. F., Zepeda, M., Cohn, J. A., Yankaskas, J. R. and Wilson, J. M. (1994) *J. Clin. Invest.* **93**, 737–749
- Harris, A., Chalkley, G., Goodman, S. and Coleman, L. (1991) *Development* **113**, 305–310
- Foulkes, A. G. and Harris, A. (1993) *Pancreas* **8**, 3–6
- Treize, A. E., Chambers, J. A., Wardle, C. J., Gould, S. and Harris, A. (1993) *Hum. Mol. Genet.* **2**, 213–218
- Treize, A. E. and Buchwald, M. (1991) *Nature (London)* **353**, 434–437
- Treize, A. E., Linder, C. C., Grieger, D., Thompson, E. W., Meunier, H., Griswold, M. D. and Buchwald, M. (1993) *Nat. Genet.* **3**, 157–164
- Yoshimura, K., Nakamura, H., Trapnell, B. C., Dalemans, W., Pavirani, A., Lecocq, J. P. and Crystal, R. G. (1991) *J. Biol. Chem.* **266**, 9140–9144
- Chou, J. L., Rozmahel, R. and Tsui, L. C. (1991) *J. Biol. Chem.* **266**, 24471–24476
- Koh, J., Sterra, T. J. and Collins, F. S. (1993) *J. Biol. Chem.* **268**, 15912–15921
- Denamur, E. and Chehab, F. F. (1994) *Hum. Mol. Genet.* **3**, 1089–1094
- Trapnell, B. C., Zeitlin, P. L., Chu, 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
- Matthews, R. P. and McKnight, G. S. (1996) *J. Biol. Chem.* **271**, 31869–31877
- Pittman, N., Shue, G., LeLeiko, N. S. and Walsh, M. J. (1995) *J. Biol. Chem.* **270**, 28848–28857
- Vuillaumier, S., Dixmeras, I., Messai, H., Lapoumeroulie, C., Lallemand, D., Gekas, J., Chehab, F. F., Perret, C., Elion, J. and Denamur, E. (1997) *Biochem. J.* **327**, 651–662
- Smith, A. N., Wardle, C. J. and Harris, A. (1995) *Biochem. Biophys. Res. Commun.* **211**, 274–281
- 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
- Anand, R., Ogilvie, D. J., Butler, R., Riley, J. H., Finniear, R. S., Powell, S. J., Smith, J. C. and Markham, A. F. (1991) *Genomics* **9**, 124–130
- Manson, A. L., Treize, A. E., MacVinish, L. J., Kasschau, K. D., Birchall, N., Episkopou, V., Vassaux, G., Evans, M. J., Colledge, W. H., Cuthbert, A. W. and Huxley, C. (1997) *EMBO J.* **16**, 4238–4249
- Moulin, D. S., Nuthall, H., Smith, D. J., Huxley, C. and Harris, A. (1999) *Mol. Med.* **5**, 211–223
- Fogh, J., Wright, W. C. and Loveless, J. D. (1977) *J. Natl. Cancer. Inst.* **58**, 209–214
- Huet, C., Sahuquillo Merino, C., Coudrier, E. and Louvard, D. (1987) *J. Cell. Biol.* **105**, 345–357
- Fogh, J. and Trempe, G. (1975) in *New Human Tumour Cell Lines. Human Tumour Cell Lines In Vivo* (Fogh, J., ed.), pp. 115–159, Plenum Press, New York
- Lieber, M., Mazzetta, J., Nelson Rees, W., Kaplan, M. and Todaro, G. (1975) *Int. J. Cancer* **15**, 741–747
- Soule, H. D., Vazquez, J., Long, A., Albert, S. and Brennan, M. (1973) *J. Natl. Cancer Inst.* **51**, 1409–1416
- Harris, A. and Coleman, L. (1989) *J. Cell. Sci.* **92**, 687–690
- Coleman, L. and Harris, A. (1991) *J. Cell. Sci.* **98**, 85–89
- Chirgwin, J. M., Przybyla, A. E., Macdonald, R. J. and Rutter, W. J. (1979) *Biochemistry* **18**, 5294–5299
- Nizetic, D., Monard, S., Young, B., Cotter, F., Zehetner, G. and Lehrach, H. (1994) *Mamm. Genome* **5**, 801–802
- Chalkley, G. and Harris, A. (1991) *J. Med. Genet.* **28**, 777–780
- Higgs, D. R., Wood, W. G., Jarman, A. P., Sharpe, J., Lida, J., Pretorius, I.-M. and Ayyub, H. (1990) *Genes Dev.* **4**, 1588–1601
- Hollingsworth, M. A., Closken, C., Harris, A., McDonald, C. D., Pahwa, G. S. and Maher, III, L. J. (1994) *Nucleic Acids Res.* **22**, 1138–1146
- Vyas, P., Vickers, M. A., Simmons, D. L., Ayyub, H., Craddock, C. F. and Higgs, D. R. (1992) *Cell* **69**, 781–793
- Dixmeras, I., Lapoumeroulie, C., Pascual-Le Tallec, L., Bens, M., Elion, J., Vanderwalle, A. and Denamur, E. (1998) *Biochem. Biophys. Res. Commun.* **250**, 328–334
- Vassaux, G., Manson, A. L. and Huxley, C. (1997) *Gene Ther.* **4**, 618–623

Received 22 February 1999/30 April 1999; accepted 28 May 1999