Erythroid expression and DNAaseI-hypersensitive sites of the carbonic anhydrase 1 gene

Jane SOWDEN, Mina EDWARDS, Katie MORRISON, Peter H. W. BUTTERWORTH and Yvonne H. EDWARDS

MRC Human Biochemical Genetics Unit, The Galton Laboratory (UCL), Wolfson House, 4 Stephenson Way, London NW1 2HE, U.K.

The carbonic anhydrase 1 gene is expressed in adult human and mouse erythroid cells and colon epithelia from two distinct promoters. We have explored the erythroid promoter for *cis*-acting sequences involved in transcription using DNAaseI as a probe. Two DNAaseI-hypersensitive sites (DHS-1 and DHS-2) have been identified in the distal erythroid promoter in *CA1*-expressing erythroleukaemic cells. These sites are present at low levels in K562 cells, which have a foetal/embryonic phenotype and do not express *CA1*. DHS-1 and DHS-2 are not present in non-erythroid cells, including colon cells, which express *CA1* from the proximal colon promoter. DHS-1 and DHS-2 were also generated in an heterologous *CA1* gene containing 5 kb of erythroid promoter sequence after transfection into erythroid cells, including K562 cells. These transfection studies showed that both this fragment, and an abbreviated 817 bp promoter fragment which contains only DHS-1, were sufficient to confer erythroid-specific expression to a reporter gene. These promoters were active in cell lines expressing *CA1* and in K562 cells. This latter observation implies that a developmental repressor factor is both present in K562 cells and binds to a *cis*-acting sequence that is absent from the sequence 5 kb upstream of the erythroid transcription start site.

INTRODUCTION

Seven genes have been identified which encode carbonic anhydrase (CA) isoenzymes in mammals (Tashian, 1989). The CA isoenzymes all perform vital functions in the cellular diffusion of CO₂, ion transport and pH regulation, but show differences in their sensitivities to CA inhibitors and rates of CO₂ turnover (Khalifah & Silverman, 1991). In recent years significant progress has been made towards understanding the molecular basis for these differences by mutational analyses of recombinant CA protein (Tu *et al.*, 1989; Jewell *et al.*, 1991). The various isoenzymes also exhibit characteristic patterns of cell-type expression, and several studies correlating protein with mRNA levels suggest that these are regulated at the level of gene transcription (Brownson *et al.*, 1988; Brady *et al.*, 1990).

Three of the genes, *CA1*, *CA2* and *CA3*, are clustered in the mammalian genome, near to the centromere, band 3A, in the mouse and within 150 kb of each other on human chromosome 8q22 (Beechey *et al.*, 1990; Lowe *et al.*, 1991). Despite their close proximity and evident structural similarities, these three genes each exhibit a different pattern of expression during differentiation and development. *CA3* is first expressed in the early mesoderm and developing notochord and later in skeletal muscle (Lyons *et al.*, 1991); its proximal promoter region contains binding sites for several muscle-specific transcription factors (Edwards *et al.*, 1992). In contrast, *CA2* is ubiquitously expressed and exhibits a characteristic housekeeping promoter (Shapiro *et al.*, 1987). The third gene, *CA1*, which is the subject of the present paper, is expressed in two major sites: erythroid cells and colon epithelia (Villeva *et al.*, 1985; Lonnerholm, 1984). In red blood cells CA isoenzyme 1 (*CA1*) is the next most abundant protein to haemoglobin (10 µg/mg of Hb).

The *CA1* gene in both human and mouse is transcribed from two promoters with different tissue specificities (Butterworth *et al.*, 1991; Fraser *et al.*, 1989). Structural characterization of the

CA1 gene revealed an unusual 5' architecture (Lowe *et al.*, 1990), with one promoter proximal to the protein coding sequence and an additional distal promoter 36 kb upstream of the protein coding sequence. The distal promoter gives rise to high levels of *CA1* mRNA in reticulocytes and erythroleukemic cells (Brady *et al.*, 1989), whereas the proximal promoter determines expression in colon epithelial cells (Brady *et al.*, 1991). The use of these promoters appears to be exclusive, such that both are never active in the same cell type. We are interested in the mechanisms and factors which specifically activate the *CA1* erythroid promoter and repress the colon promoter in erythroid cells.

Two particular aspects of the regulation of *CA1* in erythroid cells form the focus of these studies. The first is the developmental-stage-specific activation of *CA1* (Brady *et al.*, 1990); the second is the regulation of *CA1* during erythropoiesis (Villeva *et al.*, 1985). We have used several cell lines as model systems to represent these patterns of *CA1* expression. In particular K562 cells have characteristics of an embryonic/foetal erythroid cell and have been used in other studies to model an early developmental stage (Lozzio & Lozzio, 1975; Donovan-Peluso *et al.*, 1987). In addition, murine erythroleukaemic (MEL) cells are useful, since they can be induced with dimethyl sulphoxide (DMSO) to undergo a differentiation process that mimics erythropoiesis (Kabat *et al.*, 1975) and is marked by a suppression of *CA1* transcription (Fraser & Curtis, 1987).

It is well recognized that transcriptionally active promoters and other *cis*-acting sequences acquire an 'open' chromatin conformation that is accessible to *trans*-acting factors. Nuclease-hypersensitive sites are believed to mark such chromatin conformations and have provided valuable indications of *cis*-acting regulatory sequences in numerous systems [for reviews, see Elgin (1981) and Gross & Garrard (1988)]. For example, at least ten DNAase I-hypersensitive sites (DHSs) have been characterized in the chicken lysozyme gene locus and all appear to mark sequences with a range of functions including transcriptional

enhancers and silencers (Fritton *et al.*, 1988). Of particular note are the DHSs of the β -globin locus control region (LCR), which mark out sequences responsible for activating the entire locus in erythroid cells (Grosveld *et al.*, 1987), bind *trans*-acting transcription factors, and confer position-independent high-level gene expression (Ney *et al.*, 1990; Blom van Assendelft *et al.*, 1989).

Using DNAaseI as a probe, we have looked for sequences involved in transcription in and around the *CAI* gene in erythroid and non-erythroid cell lines. These studies have revealed several DHSs, at least two of which are erythroid-specific and can be generated in transfected promoter constructs.

MATERIALS AND METHODS

Cell lines

The following cell lines were used in this study: erythro-leukaemic cell lines K562 (Lozzio & Lozzio, 1975), HEL (Martin & Papayannopoulou, 1982) and MEL (Friend *et al.*, 1971), colon carcinoma LIM1215 (Whitehead *et al.*, 1985), T lymphoblastoid CCRF-CEM (Foley, 1965) and HeLa (Gey *et al.*, 1952). Tissue-culture practice was standard (Freshney, 1983). HeLa, K562, HEL, and CCRF-CEM lines were maintained in Dulbecco modified Eagle medium. MEL C88 (APRT⁻) cells were maintained in α -minimal essential medium supplemented with diaminopurine (50 $\mu\text{g}\cdot\text{ml}^{-1}$). All media were supplemented with 10% (v/v) foetal-calf serum, penicillin (100 i.u.), streptomycin (100 $\mu\text{g}\cdot\text{ml}^{-1}$), amphotericin B (2.5 $\mu\text{g}\cdot\text{ml}^{-1}$) and 2 mM-glutamine.

DNAaseI-hypersensitivity assay

Nuclei were prepared from cultured cells harvested in exponential phase. Typically 7×10^7 cells were washed twice in ice-cold phosphate-buffered saline and twice in ice-cold Buffer I (0.05% (v/v) Nonidet P40/10 mM-NaCl/3 mM-MgCl₂/10 mM-Hepes, pH 7.5). Cellular lysis was confirmed by Trypan Blue staining. The lysate was re-suspended in 10 ml of Buffer II (10 mM-NaCl/3 mM-MgCl₂/10 mM-Hepes, pH 7.5) and nuclei were twice sedimented at 1000 *g* through 10 ml of 30% (w/v) sucrose dissolved in Buffer II at 4 °C for 5 min. DNAaseI-hypersensitivity assays were carried out on isolated nuclei (essentially as described by Enver *et al.*, 1985). Isolated nuclei were re-suspended in ice-cold DNAaseI buffer [100 mM-NaCl/50 mM-Tris (pH 8.0)/3 mM-MgCl₂/0.1 mM-phenylmethanesulphonyl fluoride/1 mM-CaCl₂], and aliquots were digested with a range of DNAaseI concentrations from 0.1 to 10 $\mu\text{g}\cdot\text{ml}^{-1}$. These reaction mixtures were incubated for 30 min at 37 °C before DNAaseI digestion was terminated by the addition of EDTA (to 25 mM). RNAase was added to 20 $\mu\text{g}\cdot\text{ml}^{-1}$ and the mixture was incubated for 30 min at 37 °C; proteinase K and SDS were added to 50 $\mu\text{g}\cdot\text{ml}^{-1}$ and 0.25% (w/v) respectively, with further incubation for 18 h at 60 °C. Samples were extracted with phenol/chloroform and chloroform and subjected to ethanol precipitation. DNAaseI digestion patterns were analysed by Southern-blot analysis. Several fragments were isolated from *CAI* genomic clones (Lowe *et al.*, 1990) for use as hybridization probes: 1.2 kb *HindIII*-*SmaI* fragment, (1.2H-S); 1.4 kb *AvaII* fragment (1.4A); 1.0 kb *XbaI*-*HindIII* fragment (1.0X-H); 2.0 kb *XbaI* fragment (2.0X); 0.8 kb *EcoRI* fragment (0.8E); 700 bp *RsaI*-*RsaI* (0.7R). Fig. 1 shows the relationship of these probes to the genomic map of *CAI* and their positions relative to the erythroid transcription start site.

Plasmids

Two *CAI* promoter fragments, *SauI* to *SspI* (+14 to -817) and *SauI* to *SmaI* (+14 to -5.0 kb), were ligated on to the murine H2K gene and cloned into the G418 selection vector ptkAGPT (Grosveld *et al.*, 1982) to give plasmids *CAI*-*SspI*-H2K and *CAI*-*SmaI*-H2K respectively (see Fig. 5 below). The cloning strategy was to replace the β -globin promoter (-815 to +50) fragment from β -H2K (Antoniou *et al.*, 1988) with *CAI* promoter fragments, using *SauI* and *HindIII* sites flanking the β -globin fragment. In both constructs a *CAI* *SauI* site (at +14) was ligated to a *HindIII* linker in the 5' leader of H2K. A promoterless vector H2K was formed by circularization of the 12 kb linear plasmid tkAGPT (minus the β -globin promoter). *HindIII* digests were used to select *CAI* recombinants and a construct fragment was sequenced to confirm the integrity of the ligation between the *CAI* promoter (at +14) and the leader of the H2K reporter gene (all positions given are relative to the erythroid transcription start site).

Transfections

Plasmid DNA for transfections was purified from CsCl gradients, linearized by digestion with *PvuI*, and ethanol-precipitated. Electroporation parameters were as follows: cultured cells were harvested in exponential phase and re-suspended to a volume of 0.7 ml in 25 mM-Hepes (pH 7.24)/140 mM-NaCl/0.75 mM-Na₂HPO₄ (Smithies *et al.*, 1985) and mixed with *PvuI*-digested plasmid H2K, β -H2K, *CAI*-*SspI*-H2K or *CAI*-*SmaI*-H2K resuspended in 100 μl of the same buffer. Samples were left at room temperature for 10 min, electroporated at 250 V (960 μF) and left at room temperature for a further 5 min before being transferred to culture medium. Geneticin G418 sulphate was added after 48 h. The minimum concentration of Geneticin necessary to kill untransfected cells in 10 days was empirically determined for each cell line. The concentrations used were between 0.2 and 1.0 $\mu\text{g}\cdot\text{ml}^{-1}$, and resistant colonies were usually detected approx. 14 days after transfection.

Conditions were optimized to give at least 50 colonies per transfection. MEL, HEL, HeLa and CCRF-CEM cells ($1.0\text{--}2.0 \times 10^6$) were each transfected with 100 μg of linear plasmid *CAI*-*SspI*-H2K and *CAI*-*SmaI*-H2K DNA respectively and 4×10^6 K562 cells were transfected with 25 μg of the same plasmids. Fewer K562 cells were used per transfection because of their higher transfection efficiency. After selection, transfected populations of MEL cells were divided and half were induced to differentiate by culturing for 3–4 days in the presence of 1.8% (v/v) DMSO. RNA and DNA for analysis was isolated from transfected cell populations using standard methods (Maniatis *et al.*, 1982; Chirgwin *et al.*, 1979).

DNA analysis

Amplification was performed on 1 μg of DNA isolated from transfected cell populations in 100 μl of *Taq* DNA polymerase buffer containing 50 pmol of oligonucleotide primers specific for the amplification of a 226 bp DNA fragment across the *CAI*-H2K construct junction. The sequences of the primers were: CA 5'-CCCACTCTAATCACCACAGG-3' (sense); H2 5'-CAACAGCAGGAGCAGCGTGC-3' (complementary). After 5 min at 95 °C, 2 units of *Taq* DNA polymerase were added, followed by 32 cycles of amplification (denaturation for 15 s at 94 °C, annealing for 30 s at 56 °C and elongation for 30 s at 72 °C). Control amplifications of the same DNA samples were performed using human phosphoglycerate mutase 2 and mouse erythropoietin oligonucleotide primers (Edwards *et al.*, 1989; Abbott, 1992).

To confirm the integrity of the construct after integration into

Fig. 1. DHSs in the *CAI* gene

Open and filled-in boxes depict leader and coding exons respectively (Lowe *et al.*, 1990). The erythroid and colon promoters lie immediately upstream of exons 1a and 1c respectively. DHSs are depicted as vertical arrows and the hybridization probes used in DNAaseI-hypersensitivity assays are indicated by horizontal solid black lines (see the Materials and methods section for a full description of probes). The positions of the probes relative to the erythroid transcription start site are as follows: 1.2H-S at -6.2 kb, 1.4A at -2.4 kb, 1.0X-H at +9 kb, 2.0X at +26.5 kb, 0.8E at +35 kb, 0.7R at +50 kb.

the cellular genome, Southern filters of genomic DNA isolated from transfected cell populations were routinely digested with *EcoRI* and hybridized with ^{32}P -labelled 2.1 kb *EcoRI* fragment containing the *AGPT* (aminoglycosyl 3'-phosphotransferase) gene. Populations of transfected cells with similar average numbers of *CAI*-H2K construct copies were analysed for H2K expression.

RNA analysis

A 150 μg portion of RNA and 3 ng of ^{32}P -labelled oligonucleotide H2 (sp. radioactivity 1×10^8 c.p.m. $\cdot \mu\text{g}^{-1}$) complementary to *CAI*-H2K mRNA (51–71 nt downstream of the erythroid start site in the first exon of H2K) were annealed in 250 mM-KCl/10 mM-Tris/HCl, pH 8.3, by heating to 85 °C for 5 min and 58 °C for 1 h, followed by ethanol precipitation. The pellet was resuspended in 50 mM-Tris/HCl (pH 8.0)/140 mM-KCl/6 mM-MgCl₂/30 mM- β mercaptoethanol/placental RNAase inhibitor (2 units $\cdot \mu\text{l}^{-1}$)/1 mM-(all 4)dNTPs/reverse transcriptase (0.5 units $\cdot \mu\text{l}^{-1}$) and incubated at 50 °C for 1 h. EDTA was then added to 15 mM, followed by RNAase to 0.5 $\mu\text{g} \cdot \text{ml}^{-1}$ and the mixture incubated for 30 min at 37 °C. The purified extension product was analysed on 8% (w/v) polyacrylamide/50% (w/v) urea denaturing gel.

RESULTS

Two cell lines were used as a source of nuclei expressing *CAI*. HEL cells express *CAI* from the distal erythroid-specific promoter (Brady *et al.*, 1989), whereas the colon-carcinoma cell line, LIM1215 (Whitehead *et al.*, 1985), expresses *CAI* from the proximal colon promoter (J. Sowden, unpublished work).

Two erythroid-specific DHSs occur in the 5' flanking region of the *CAI* gene

Southern-blot analysis was used to map the positions of two DHSs in the 5' region of the gene. The probes, 1.4A and 1.2H-S, are derived from a 6.3 kb *HindIII* fragment that includes the *CAI* distal erythroid promoter and immediate upstream sequences (Fig. 1). Both probes detected identical fragments in *HindIII* and in *EcoRI* digestions after DNAaseI treatment. Fig. 2 shows a DNAaseI-hypersensitivity assay performed on nuclei isolated from HEL cells. The parent 6.3 kb *HindIII* fragment is seen in sample 0 without DNAaseI treatment. When nuclei were treated with DNAaseI at increasing concentrations, two novel fragments of 5.8 kb and 4.5 kb appeared, indicating the presence of two DHSs. The 1.4A probe detected a parent *SstI* fragment of 8 kb, which was cleaved by DNAaseI to give fragments of 2.5 kb and 3.8 kb, and an *XbaI* 10.0 kb parent band was cleaved by

DNAaseI to give fragments of 2.8 kb and 4.1 kb. Similar data was obtained using *EcoRI* and other enzymes in the post-DNAaseI digest. The combined restriction mapping data locate two DHSs (DHS-1 and DHS-2) to positions approx. 200 bp and 1.5 kb upstream of the erythroid start site respectively (Figs. 1 and 2).

A comparison of the relative intensities of the DNAaseI fragments suggest that, at higher DNAaseI concentrations, most of the chromosomal DNA was cleaved either at the more 5' site, DHS-2, or at both sites simultaneously, such that the small 2.8 kb *XbaI* fragment or the 2.5 kb *SstI* fragment were the most prominent product. At lower concentrations DNAaseI cleaves at only one or the other of the DHSs or not at all, and both DNAaseI products are clearly detected (Fig. 2).

To establish whether DHS-1 and DHS-2 are tissue-specific phenomena, DNAaseI-hypersensitivity assays were performed in parallel on nuclei isolated from *CAI*-expressing HEL cells, from LIM1215 cells, which express *CAI* from the colon promoter and not the erythroid promoter, and from *CAI*-non-expressing K562 and HeLa cells. A typical experiment is shown in Fig. 3. The appearance of two small DNAaseI products of 8.5 kb and 7.2 kb in DNAaseI/*EcoRI*-digested HEL-cell DNA confirmed the presence of DHS-1 and DHS-2. In contrast, the HeLa and LIM1215

Fig. 2. DHS-1 and DHS-2 in the distal erythroid promoter of *CAI* in HEL chromosomal DNA

DNA isolated from DNAaseI-treated (with concentrations increasing from 0.1 to 6.0 $\mu\text{g} \cdot \text{ml}^{-1}$, indicated by an arrow) HEL-cell nuclei was digested with *SstI*, *XbaI*, *EcoRI* and *HindIII* and hybridized with ^{32}P -labelled 1.4A (see Fig. 1). A restriction map around the erythroid transcriptional start site (+1) is shown. Leader exon 1a is indicated by an open box, DHSs are depicted as vertical arrows, and the positions of sequence motifs fitting the GATA-1 consensus are indicated by —●—. The hybridization probes are indicated below the map by solid black lines.

Fig. 3. Erythroid-specific 5' hypersensitive sites in *CA1* chromosomal DNA

HEL, K562, HeLa and LIM1215 cell nuclei were treated with DNAaseI ($0.1\text{--}6.0\ \mu\text{g}\cdot\text{ml}^{-1}$, indicated by an arrow) and purified DNA was digested with *EcoRI* or *HindIII* and hybridized with 1.4A fragment (Fig. 1). Hybridization with 1.2H-S fragment (Fig. 1) produced the same fragment pattern (results not shown).

samples exhibited only the parental 11.0 kb *EcoRI* fragment at all DNAaseI concentrations until digestion was complete. The analysis of K562 samples was less straightforward. The DNAaseI products were not immediately apparent in K562 DNA. However, prolonged autoradiographic exposure (> 10 days) and higher starting nuclear DNA concentrations revealed a low level of these fragments in digested K562 samples. For example, using $15\ \mu\text{g}$ instead of $10\ \mu\text{g}$ of nuclear DNA per digest, two smaller bands of 5.8 kb and 4.5 kb were detected (Fig. 3), in addition to a parent 6.3 kb *HindIII* fragment after DNAaseI treatment. The DNAaseI products were the same size as those detected in HEL DNA (compare Fig. 2), but always occurred as minority species, even at higher DNAaseI concentrations. These results indicate that a proportion of K562 chromosomal DNA exhibits DHS-1 and DHS-2. This feature was specific to K562, since these sites could not be detected in HeLa or LIM1215 cells at high concentrations of starting nuclear DNA after prolonged autoradiographic exposure.

DHSs downstream of the erythroid start site and in the region of the proximal colon promoter

Restriction-enzyme mapping of DNA from HEL DNAaseI-treated nuclei using 1.0X-H (Fig. 1) as the probe indicated that three additional hypersensitive sites, DHS+1, DHS+2 and DHS+3, are present 7 kb, 7.8 kb and 12.8 kb respectively downstream of the leader exon 1a, within the large 5' intron of *CA1* (Fig. 1). DNAaseI cleavage of the parent 3.5 kb *HindIII* fragment produced two fragments of 2.0 kb and 2.7 kb, whereas cleavage of the overlapping parent 4.5 kb *XbaI* fragment generates a single band of 4.0 kb (Fig. 4). In addition a weak DHS, DHS+4, was detected 3 kb upstream of exon 1b using the 2.0X probe (Fig. 1). These four DHSs were not present in HeLa or K562 cells. Further, DHSs were not found in erythroid cells, either in the coding regions or in the 3' flanking region of *CA1* using probes which spanned these regions (see the Materials and methods section and Fig. 1).

In addition to DHSs specific to erythroid cells, a single site (DHS+5) specific to *CA1*-expressing colon cells (LIM1215) was identified. DHS+5 lies approx. 500 bp upstream of the proximal colon promoter and was detected in the colon-carcinoma cell line LIM1215 using the 0.8E probe (see Figs. 1 and 4). Parent 11 kb

Fig. 4. DHSs downstream of the erythroid start site and in the region of the colon promoter

DNA isolated from DNAaseI-treated HEL-cell nuclei was digested with *XbaI* and *HindIII* and hybridized with the 1.0X-H probe (left-hand side). DNA isolated from DNAaseI-treated HEL and LIM1215 nuclei was digested with *HindIII* and hybridized with the 0.8E probe (right-hand side). A restriction map of *XbaI* (X) and *HindIII* (H) sites is shown. Leader exons 1a and 1c are indicated by open boxes. The erythroid and colon promoters lie immediately upstream of exons 1a and 1c respectively. DHSs are depicted as vertical arrows, and the hybridization probes are indicated below the map by solid black lines.

HindIII and 7 kb *XbaI* fragments were cleaved by DNAaseI to generate an 8 kb fragment (Fig. 4) and a 3 kb fragment respectively. These products were not detected in HEL DNAaseI-treated DNA.

These data from DNAaseI-hypersensitivity assays, when taken together, demonstrate that the *CA1* gene adopts distinctive and alternative chromatin conformations in cells which express *CA1* specifically from either the distal erythroid promoter or the more proximal colon promoter.

DHSs in a transfected *CA1*-H2K gene construct

A series of transfection experiments were carried out to establish the functional significance of DHS-1 and DHS-2 *in vivo*. Two constructs were made in which the *CA1* erythroid promoter was attached to the H2K reporter gene. One, *CA1*-*SmaI*-H2K, included 5 kb of 5' flanking sequence (+14 bp to -5.0 kb) which encompassed four binding sites for the erythroid-specific transcription factor GATA-1 (Brady *et al.*, 1989; Tsai *et al.*, 1989) and both DHS-1 and DHS-2. The other construct, *CA1*-*SspI*-H2K included only 815 bp of 5' flanking sequence (+14 bp to -817 bp) encompassing the same four GATA-1 binding sites but only DHS-1; DHS-2 was not included. The *CA1*-H2K plasmids were stably transfected into five different cell lines, MEL, HEL, K562, HeLa and CCRF-CEM. Transfection was confirmed by PCR amplification of a 226 bp fragment specific for the *CA1*-H2K construct using DNA isolated from pools of Geneticin-resistant cells (Fig. 5).

DNAaseI-hypersensitivity assays were performed on cell populations stably transfected with plasmid *CA1*-*SmaI*-H2K. Post DNAaseI digests were *EcoRI*/*BamHI* in order to distinguish the construct from the endogenous *CA1* DNA. A typical assay using the 1.4A probe is shown in Fig. 6. Two large fragments, which correspond to the endogenous human *CA1* gene (a 9.0 kb *BamHI*-*EcoRI* fragment) and to the heterologous *CA1*-*SmaI*-H2K (a 7.2 kb *EcoRI*-*BamHI* fragment), are seen in DNA from transfected K562 and HeLa without DNAaseI treatment. In MEL cells the mouse *CA1* gene is not detected by the human

Fig. 5. PCR amplification of DNA isolated from untransfected MEL, HEL, K562, HeLa cells and from cell populations including CCRF-CEM, transfected with plasmids *CAI-SspI-H2K* (Ss) and *CAI-SmaI-H2K* (Sm) (the latter are indicated as Sm/MEL, Ss/MEL etc. above each lane, as are the identities of the PCR products)

Amplification was carried out using specific oligonucleotide primers to detect *CAI-H2K* (Ch) which amplify a 226 bp fragment across the *HindIII* site in both constructs, and primers to detect endogenous human phosphoglycerate mutase 2 (Pg) and mouse erythropoietin (Ep) as controls. The minor band of about 300 bp seen in the MEL, HEL and K562 untransfected Ch tracks is a non-specific PCR product. The structure of the *CAI-H2K* constructs are indicated diagrammatically. Abbreviations: Amp, β -lactamase gene of pBR322; tkAGPT, aminoglycosyl 3'-phosphotransferase gene under the control of the herpes-simplex-virus thymidine kinase gene promoter.

Fig. 6. HeLa, K562 and MEL cells were stably transfected with the plasmid *CAI-SmaI-H2K*

Nuclei were isolated from populations of stable transfectants and treated with DNAaseI ($0.1\text{--}3.0\text{ }\mu\text{g}\cdot\text{ml}^{-1}$, increasing concentrations of DNAaseI as indicated by an arrow), and purified DNA was double-digested with *EcoRI* and *BamHI* and hybridized with ^{32}P -labelled 1.4A (see Fig. 1). C1 and C2 are untransfected samples of K562 and MEL DNA respectively. Fragments attributed to cleavage at DHS-1 and DHS-2 are indicated by small arrows ($\leftarrow$).

probe, so that only the transfected construct is seen in untreated cells (Fig. 6).

Two erythroid-specific and *CAI-H2K*-specific DNAaseI products, 4.8 kb and 3.5 kb, were detected in chromosomal DNA from transfected MEL cells and HEL cells (HEL-cell results not shown). These bands were not seen in HeLa-cell DNA and are distinguished in size from the bands which would be expected from DNAaseI digestion of the endogenous gene. Restriction mapping indicated that the positions of these two DHS, in the *CAI* promoter of the *CAI-H2K* gene are the same as DHS-1 and DHS-2 in the endogenous gene in HEL chromosomal DNA. Thus DHS-1 and DHS-2 are generated when cloned linearized *CAI* DNA is integrated into heterologous erythroid cells.

A striking feature of these studies is that, whereas DHS-1 and DHS-2 are only weakly evident in the endogenous unexpressed *CAI* gene of K562 cells, transfection of K562 cells with *CAI-SmaI-H2K* resulted in the generation of these DHS to the same extent as in equivalent transfections of *CAI*-expressing cells (for example, compare with MEL, Fig. 3).

In transfected samples of K562 and HeLa, but not MEL cells,

two additional fragments (of 5.0 kb and 6.8 kb) were detected. The intensities of these fragments do not relate to DNAaseI concentrations, since they are found both in the control (no DNAaseI) and DNAaseI-treated samples; nor do the fragments appear to derive from the endogenous *CAI* gene, since they were not detected after DNAaseI treatment of untransfected cells. Thus they appear to be consequent upon the introduction of the *CAI-H2K* chimaeric gene into the human cells and perhaps endogenous nuclease activity. The nature of these fragments still requires investigation.

***CAI* 5' flanking sequences confer erythroid-specific expression on a reporter gene**

In order to demonstrate whether the generation of DHS-1 and DHS-2 in transfected *CAI* constructs was associated with transcriptional activity, RNA was isolated from pools of transfectant cells and analysed for the presence of *CAI-H2K* transcripts. An oligonucleotide, H2 (see the Materials and methods section), within the first exon of *CAI-H2K* was used as a primer in RNA extension analyses.

These experiments showed that both *CAI-H2K* constructs were transcriptionally active in an erythroid-specific manner. Deletion of DHS-2 and 4.2 kb of upstream sequence (*CAI-SspI-H2K*) did not alter the level of *CAI-H2K* expression. Of special note was the finding that *CAI-H2K* mRNA was detected in transfected K562 cells which do not normally express *CAI*. All of these features are demonstrated in Fig. 7. A 71 nt band corresponding to *CAI-H2K* mRNA was detected in MEL, K562 and HEL cells transfected with both *CAI-SspI-H2K* and *CAI-SmaI-H2K*, but not in transfected HeLa or CCRF-CEM cells (HEL-cell results not shown). A 56 nt product in MEL cells corresponds to mRNA transcribed from the endogenous murine H2K gene and provides an internal control; otherwise no extension products were primed from untransfected control RNA samples (C in Fig. 7). Comparisons of the levels of the 71 nt transcript relative to each other (and to the control H2K signal in MEL cells) suggest that similar levels of RNA are transcribed from both constructs. A promoterless H2K plasmid (–P in Fig. 7) was also transfected, but did not transcribe H2K.

Fig. 7. Primer extension analyses of *CAI*-H2K erythroid transcripts in RNA isolated from K562, MEL, HeLa and CCRF-CEM populations transfected with *CAI*-*Sspl*-H2K (Ss) and *CAI*-*Smal*-H2K (Sm) and promoterless construct H2K (-P)

C is control RNA from untransfected cells. MEL cells are shown separately transfected with β -H2K (β), and both before (-) and after (+) induction with DMSO. The 85nt, 71nt and 56nt products are derived from β -H2K, *CAI*-H2K and H2K mRNA respectively, as predicted from sequence data and as described in the text. Below is a Northern blot of the same MEL transfectant RNA (20 μ g), before and after DMSO treatment, hybridized with 32 P-labelled mouse *CAI* cDNA (Fraser *et al.*, 1989).

CAI promoter activity during DMSO induction of MEL cells

Stable populations of MEL cells transfected with both *CAI* promoter constructs were induced to undergo erythroid differentiation using DMSO (Kabat *et al.*, 1975). The efficiency of DMSO induction was confirmed by the down-regulation of the endogenous mouse *CAI* mRNA and activation of the β -globin promoter in a parallel transfection of the β -H2K plasmid (Antoniou *et al.*, 1988) (Fig. 7). Primer extension analyses (Fig. 7) showed that the efficiency of the *CAI* promoter in both the *CAI*-H2K constructs was unaffected by DMSO, as judged by the intensity of the 71nt *CAI*-H2K transcript, which was present at similar levels before and after treatment (the 56nt band corresponding to the endogenous H2K transcript acts as an internal control for mRNA levels, since H2K transcription does not alter with MEL-cell differentiation; Wright *et al.*, 1984).

These results suggest that the mechanism which down-regulates the endogenous mouse *CAI* mRNA during DMSO-induced erythroid differentiation (Fraser & Curtis, 1987) is not effective on the *CAI*-H2K promoter, and implies that these *CAI*-H2K constructs do not include necessary *cis*-acting sequences.

DISCUSSION

The presence of two promoters in the *CAI* gene, each possessing a distinctive set of regulatory elements, is conserved between mouse and humans (Fraser *et al.*, 1989; Brady *et al.*, 1991). The colon promoter is poorly characterized, and in general very little is known about transcription factors regulating the expression of intestinal genes. A sequence element important for the tissue-specific expression of the intestinal fatty-acid-binding

protein has been identified and is found within the promoters of other intestinal genes (Sweetser *et al.*, 1987), but does not occur within the colon promoter of *CAI*. Our characterization of the colon promoter is preliminary; previous sequence analyses located a TATA box 28 bp upstream of the transcription-initiation site (Brady *et al.*, 1991), and we now describe a region 500 bp upstream of the first coding exon which is marked by a colon-specific DHS and is potentially important in transcription.

The other promoter is erythroid-specific and has been characterized in more detail; it is known to contain sequence elements which bind the erythroid transcription factor GATA-1 (Brady *et al.*, 1989; Tsai *et al.*, 1989; Zon *et al.*, 1990). In the present paper we have focused on the association between the transcriptional activity of this erythroid promoter and its local chromatin conformation. Two DHSs, DHS-1 and DHS-2, identified upstream of the erythroid start site in chromosomal DNA, were shown to be both erythroid-specific and independent of transcriptional activity. These sites were present in *CAI*-expressing erythroid cells and at low frequency in *CAI*-non expressing K562 cells. K562 cells are considered to have a foetal/embryonic erythroid phenotype with differentiation blocked at an early stage of erythropoiesis; indeed Enver *et al.* (1988) suggest that K562 cells are uncommitted haematopoietic stem cells. The relatively infrequent occurrence of DHS-1 and DHS-2 in K562 endogenous *CAI* may indicate that K562 cells are a heterogeneous population in which few cells have generated erythroid-specific DHS in their chromosomal DNA; perhaps these latter cells represent those further committed to the erythroid lineage. The notion that the cell line is heterogeneous is supported by an earlier observation that sequences around the *CAI* erythroid promoter in DNA from K562 cells always exhibit partial methylation (Lowe *et al.*, 1990). We propose that the appearance of DHS-1 and DHS-2 in certain K562 cells is an early marker for commitment to the erythroid lineage and precedes the expression of *CAI*.

In addition to the characterization of DHS in endogenous *CAI* chromosomal DNA we have shown that DHS-1 and DHS-2 are specifically generated in transcriptionally active heterologous *CAI* genes transfected into erythroid cells. There are reports of a few other cases where DHSs have been generated in an appropriate tissue-specific manner, for example after the integration of the elastase I (Hammer *et al.*, 1987) or β -globin (Radice & Constantini, 1986) transgenes into the mouse germ line. Currently the mechanism whereby DHSs are generated is not well understood. Significant progress has been made towards understanding this process in inducible systems; for example the generation of a DHS in the mouse mammary-tumour-virus promoter is associated with binding of hormone-receptor complex and nuclear factor-1 at the hormone-responsive element and this may be concomitant with nucleosome displacement (Cordingley *et al.*, 1987; Pina *et al.*, 1990). In the case of *CAI*, the GATA-1 transcription factor may play a role in the generation of erythroid-specific DHS, involving nucleosome displacement and increasing accessibility to other *trans*-acting factors. GATA-1 is thought to play a major role in the commitment of cells to the erythropoietic pathway (Pevny *et al.*, 1991) and occurs in all erythroid cells exhibiting DHS-1 and DHS-2 in both endogenous and in transfected cloned *CAI* genes. Furthermore, both DHS-1 and DHS-2 are located in the immediate region of GATA-1 motifs that fit the consensus A/TGATA/TA at -218, -190 and -1502 relative to the transcription start site, as indicated in Fig. 2.

The *CAI*-H2K promoter assays reported here showed that a minimal *CAI* promoter fragment (-815 bp to +14 bp) lacking DHS-2 contains sufficient sequence information to confer erythroid-specific expression to an H2K reporter gene. This

fragment contains four GATA-1 binding sites in addition to binding sites for other ubiquitously expressed transcription factors (Brady *et al.*, 1989). In HeLa cells, transcriptional activity or generation of DHSs is not observed, presumably because these cells lack essential erythroid-specific factors. The addition of DHS-2 to the minimal promoter did not significantly alter the *CAI*-H2K mRNA levels in erythroid cells. These experiments have not excluded the existence of enhancer elements downstream of the start site which may interact with DHS-2 sequences, and indeed our studies show that *CAI* is a complex gene with a further four erythroid DHSs and a second (colon) promoter (Brady *et al.*, 1991) downstream of the erythroid promoter.

The observation that both *CAI* promoter constructs were transcriptionally active when transfected into K562 cells where the endogenous *CAI* promoter is repressed was unexpected. This finding suggests that a *cis*-acting sequence responsible for *CAI* repression must be absent from sequence 5 kb upstream of the erythroid transcription start site and implies that a developmental repressor factor is present in K562 cells. This interpretation is compatible with the data of Brady *et al.* (1990), who found that the human *CAI* gene in K562 cells was *trans*-activated when these cells were fused with adult phenotype MEL cells, if the presence of MEL *trans*-acting factor(s), which modify the activity of the K562 repressor protein, is invoked.

We thank Mike Antoniou for the β -H2K plasmid. This work was supported by the Medical Research Council and the Wellcome Trust.

REFERENCES

- Abbott C. (1992) *Mamm. Genome* **2**, 106–109
- Antoniou, M., deBoer, E., Habets, G. & Grosveld, F. (1988) *EMBO J.* **7**, 377–384
- Beechey, C., Tweedie, S., Spurr, N., Ball, S., Peters, J. & Edwards, Y. (1990) *Genomics* **6**, 692–696
- Blom van Assendelft, G., Hanscombe, O., Grosveld, F. & Greaves, D. R. (1989) *Cell* **56**, 969–977
- Brady, H. J., Sowden, J. C., Edwards, M., Lowe, N. & Butterworth, P. H. W. (1989) *FEBS Lett.* **257**, 451–456
- Brady, H. J., Edwards, M., Linch, D. C., Knott, L., Barlow, J. H. & Butterworth, P. H. W. (1990) *Br. J. Haematol.* **76**, 135–142
- Brady, H. J., Lowe, N., Sowden, J. C., Edwards, M. & Butterworth, P. H. W. (1991) *Biochem. J.* **277**, 903–905
- Brownson, C., Isenberg, H., Brown, W., Salmons, S. & Edwards, Y. (1988) *Muscle Nerve* **11**, 1183–1189
- Butterworth, P., Barlow, J., Brady, H., Edwards, M., Lowe, N. & Sowden, J. (1991) in *The Carbonic Anhydrases, Cellular Physiology and Molecular Genetics* (Dodgson, S., Tashian, R., Gros, G. & Carter, N., eds.), pp. 197–207, Plenum Press, New York
- Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. (1979) *Biochemistry* **18**, 5294–5299
- Cordingley, M., Riegel, A. & Hager, G. (1987) *Cell* **48**, 261–270
- Donovan-Peluso, M., Acuto, S., Swanson, M., Dobkin, C. & Bank, A. (1987) *J. Biol. Chem.* **262**, 17051–17057
- Edwards, Y. H., Sakoda, S., Schon, E. & Povey, S. (1989) *Genomics* **5**, 948–951
- Edwards, Y. H., Tweedie, S., Lowe, N. & Lyons, G. (1992) in *The Molecular Biology of Muscle* (El-Haj, A., ed.), Society for Experimental Biology, in the press
- Elgin, S. C. R. (1981) *Cell* **27**, 413–415
- Enver, T., Brewer, A. C. & Patient, R. K. (1985) *Nature (London)* **318**, 680–683
- Enver, T., Zhang, J.-W., Anagnou, N. P., Stamatoyannopoulos, G. & Papayannopoulou, T. (1988) *Mol. Cell Biol.* **8**, 4917–4926
- Foley, G. (1965) *Cancer Res.* **18**, 522–529
- Fraser, P. J. & Curtis, P. (1987) *Genes Dev.* **1**, 855–861
- Fraser, P., Cummings, P. & Curtis, P. (1989) *Mol. Cell Biol.* **9**, 3308–3313
- Freshney, R. I. (1983) *Culture of Animal Cells – A Manual of Basic Techniques*, Alan R. Liss, New York
- Friend, C., Scher, W., Holland, J. G. & Sato, T. (1971) *Proc. Natl. Acad. Sci. U.S.A.* **68**, 378–382
- Fritton, H. P., Jantzen, K., Igo-Kemenes, T., Nowock, J., Strech-Jurk, U., Theisen, M. & Sippel, A. E. (1988) in *Architecture of Eukaryotic Genes* (Kahl, G., ed.), pp. 333–353, VCH, Weinheim
- Gey, G. O., Coffman, W. D. & Kubicek, M. T. (1952) *Cancer Res.* **12**, 264–265
- Gross, D. S. & Garrard, W. T. (1988) *Annu. Rev. Biochem.* **57**, 159–197
- Grosveld, F., Lund, T., Murray, E., Mellor, A., Dahl, H. & Flavell, R. (1982) *Nucleic Acids Res.* **10**, 6715–6732
- Grosveld, F., Blom van Assendelft, G., Greaves, D. R. & Kollias, G. (1987) *Cell* **51**, 975–985
- Hammer, R., Swift, G., Ornitz, D., Quaife, C. & Palmiter, R. (1987) *Mol. Cell Biol.* **7**, 2956–2967
- Jewell, D. A., Tu, C. K., Paranawithana, S. R., Tanhauser, S. M., LoGrasso, P. V., Laipis, P. J. & Silverman, D. N. (1991) *Biochemistry* **30**, 1484–1490
- Kabat, D., Sherton, C., Evans, L., Bigley, R. & Koler, R. (1975) *Cell* **5**, 331–338
- Khalifah, R. G. & Silverman, D. (1991) *Carbonic Anhydrase Kinetics and Molecular Function in the Carbonic Anhydrases; Cellular Physiology and Molecular Genetics* (Dodgson, S., Tashian, R., Gros, G. & Carter, N., eds.), pp. 49–70, Plenum Press, New York
- Lonnerholm, G. (1984) *Ann. N.Y. Acad. Sci.* **429**, 369–381
- Lowe, N., Brady, H. J., Barlow, J. H., Sowden, J. C., Edwards, M. & Butterworth, P. H. W. (1990) *Gene* **93**, 277–283
- Lowe, N., Edwards, Y. H., Edwards, M. & Butterworth, P. H. W. (1991) *Genomics* **10**, 882–888
- Lozzio, C. B. & Lozzio, B. B. (1975) *Blood* **45**, 321–334
- Lyons, G. E., Buckingham, M., Tweedie, S. & Edwards, Y. H. (1991) *Development* **111**, 233–244
- Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
- Martin, P. & Papayannopoulou, T. (1982) *Science* **216**, 1233–1235
- Ney, P. A., Sorrentino, B. P., McDonagh, K. T. & Nienhuis, A. W. (1990) *Genes Dev.* **4**, 993–1006
- Pevny, L., Simon, M. C., Robertson, E., Klein, W., Tsai, S.-F., D'Agati, V., Orkin, S. H. & Contantini, F. (1991) *Nature (London)* **349**, 257–260
- Pina, B., Bruggemeier, U. & Beato, M. (1990) *Cell* **60**, 719–731
- Radice, G. & Constantini, F. (1986) *Nucleic Acids Res.* **14**, 9765–9780
- Shapiro, L. H., Venta, P. J. & Tashian, R. E. (1987) *Mol. Cell Biol.* **7**, 4589–4593
- Smithies, O., Gregg, R. G., Boggs, S., Koralewski, M., Kucherlapati, R. (1985) *Nature* **317**, 230–234
- Sweetser, D. A., Birkenmeier, E. H., Klisak, I. J., Zollman, S., Sparkes, R. S., Mohandes, T., Lusi, A. J. & Gordon, J. (1987) *J. Biol. Chem.* **262**, 2458–2467
- Tashian, R. (1989) *BioEssays* **10**, 186–192
- Tsai, S.-F., Martin, D. I. K., Zon, I. Z., D'Andrea, A. D., Wong, G. G. & Orkin, S. H. (1989) *Nature* **339**, 446–451
- Tu, C. K., Silverman, D. N., Forsman, C., Jonsson, B. H. & Linskog, S. (1989) *Biochemistry* **28**, 7913–7918
- Villeval, J. L., Testa, U., Vinci, G., Tonthat, H., Bettaieb, A., Titeux, M., Cramer, P., Edelman, L., Rochant, H., Breton-Gorius, J. & Vainchenker, W. (1985) *Blood* **66**, 1162–1170
- Whitehead, R., Marcrae, F. A., St. John, D. J. & Ma, J. (1985) *J. Natl. Cancer Inst.* **74**, 759–765
- Wright, S., Rosenthal, A., Flavell, R. & Grosveld, F. (1984) *Cell* **38**, 265–273
- Zon, L. I., Tsai, S.-F., Burgess, S., Matsudaira, P., Bruns, G. A. P. & Orkin, S. H. (1990) *Proc. Natl. Acad. Sci.* **87**, 668–672