Nucleic Acids Research, 1993, Vol. 21, No. 10 2465–2471

Cell type-specific interactions of transcription factors with a housekeeping promoter *in vivo*

Genevieve Stapleton, Maria Patrizia Somma[1] and Patrizia Lavia[1]*
AFRC Centre for Genome Research, University of Edinburgh, King's Buildings, West Mains Road, Edinburgh EH9 3JQ, UK and [1]Centro di Genetica Evoluzionistica del CNR, c/o Dipartimento di Genetica e Biologia Molecolare, Università 'La Sapienza', Rome 00185, Italy

Received December 24, 1992; Revised and Accepted April 16, 1993

ABSTRACT

Mammalian housekeeping promoters represent a class of regulatory elements different from those of tissues-specific genes, lacking a TATA box and associated with CG-rich DNA. We have compared the organization of the housekeeping *Htf9* promoter in different cell types by genomic footprinting. The sites of *in vivo* occupancy clearly reflected local combinations of tissue-specific and ubiqitous binding factors. The flexibility of the *Htf9* promoter in acting as the target of cell-specific combinations of factors may ensure ubiquitous expression of the *Htf9*-associated genes.

INTRODUCTION

Over 10.000 genes in higher eukaryotes are thought to encode proteins with 'housekeeping' functions, which are required for survival, growth and duplication of all cells. Such genes are active in all developmental stages and tissues, despite the differences in the transcriptional apparatus of different cell types. Mammalian housekeeping promoters lie within 1–2 kb CG-rich DNA stretches (1–2), exceeding by far the average promoter size. Functional studies indicate that they represent a class of regulatory elements distinct from those of tissue-specific genes. An obvious difference is the absence of the TATA box, which results in heterogeneous transcription initiation (1) often on both DNA strands (3–5). On the other hand, elements that are potential targets of factors with CG-rich recognition sequences (such as Sp1, AP2, MLTF or E2F) occur frequently.

The mouse *Htf9* locus contains a typical housekeeping promoter (3) shared by two genes, *Htf9a* and *Htf9c*, that are transcribed from complementary DNA strands in opposite directions. Both genes are evolutionary conserved, and their expression in all tissues and cell lines suggests that they both encode proteins with basic functions (6). The *Htf9a* gene is the mouse homolog of a novel yeast gene (*SFO1*) recently identified in a search for mutations suppressing the mating deficiency of *fus1* mutants, which are defective in cell fusion (J.Trueheart and J.Thorner, manuscript in preparation). The product of the *Htf9c* gene remains unidentified.

In a previous study the *Htf9* promoter was characterized by combining protein-binding and deletion mapping assays (7, 8). The results indicated two novel features of this promoter. Firstly, we identified multiple factor-binding sites resulting in a complex architecture, yet only a subset was required for expression of a reporter gene in both orientations (7): thus the *Htf9* promoter elements were redundant. This characterization is consistent with results obtained from deletion analysis of unrelated housekeeping promoters (5, 9–12), whose transcriptional activity is also confined to short DNA fragments. Secondly, alternative promoter elements were required for activity in different cell types (8). We have hypothesized that different elements are trans-activated in a cell-specific manner by factors varying from type to type: thus, the apparent redundancy of binding sites might in fact provide the structural basis for ubiquitous expression.

To assess this hypothesis we have examined the organization of the native*Htf9* promoter in different cells by genomic footprinting. The results presented here show that the interactions between factors and single *Htf9* elements in vivo differ in cell lines of different origin and with different levels of specialization. This gives rise to different combinations, involving both ubiquitous and cell-specific factors, which are assembled in a cell type-specific manner.

MATERIALS AND METHODS

Cell lines

The following cell lines were used: mouse C3H/10 T 1/2 (ATCC CCL 226) and NIH/3T3 fibroblasts (ATCC CRL 1658), rat C6 glioma (ATCC CCL 107), mouse S-20Y neuroblastoma (13), rat H4-II-E-C3 hepatoma (ATCC CRL 1600) and human HepG2 hepatocyte carcinoma cells (ATCC HB 8065). Cells were grown in D-MEM medium supplemented with L-glutamine and 10% fetal calf serum under 5% CO_2. Growth conditions for the embryonic stem cell line CP1 and inhibition of differentiation with differentiation inhibitory activity (DIA/LIF) were described in (14).

Gel-shift assays

Protein extracts were prepared following published methods (15, 16). Gel shift assays were carried out as previously described (7, 8) using routinely 1.10^4 cpm of end-labeled probe and 1 μg of double-stranded poly (dI.dC) as a non-specific competitor.

* To whom correspondence should be addressed

In vivo footprinting

The protocol for genomic footprinting using ligation-mediated PCR (LM-PCR) (17) was kindly provided by G. Pfeiffer and generally followed. Cell monolayers were exposed to dimethyl sulfate (DMS) concentrations ranging from 0.05% to 0.5%. In most experiments the effective concentrations fell between 0.1 and 0.3%. Following exposure to DMS cells were collected and DNA extracted by standard methods. After piperidine cleavage the DNA was processed for LM-PCR. To circumvent compressions or stops in regions of high guanine content, Sequenase was replaced with Taq polymerase in the extension step and the temperature was raised to 74°C. PCR-amplified products were run on 60 cm gradient gels and electroblotted onto GeneScreen (Dupont) membranes (18). DNA was UV cross-linked under a 30 Watt lamp from a 80 cm-distance. Single-stranded probes were prepared by primer extension using 4 μg of pH9.2 plasmid containing *Htf9* annealed to 0.5 pmol of the appropriate primer, in the presence of 100 μM dNTPs, 50 μCi of α-^{32}P-dCTP and Sequenase. Filters were hybridized and washed as described in (18). In experiments with whole tissues and DNAse I as the cleaving agent nuclei were isolated through polyamine-free sucrose gradients (19). DNase I digestions and terminal-transferase reactions before LM-PCR were carried out as described in (20).

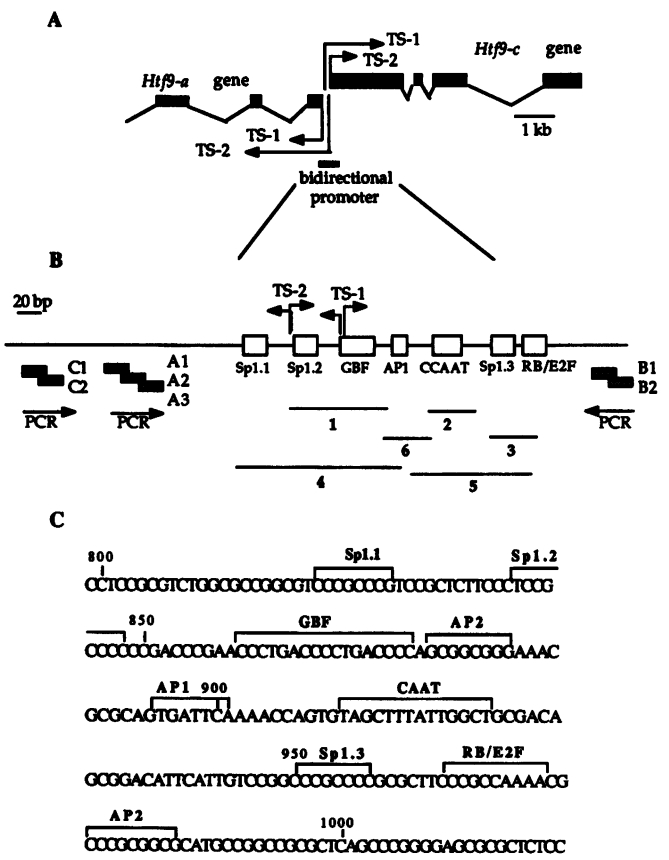

Figure 1. A. Map of the *Htf9* locus, showing the arrangement of the first exons of the *Htf9a* (remaining exons are not located) and *Htf9c* genes. Major transcriptional start sites (TS) of both genes are arrowed. **B.** The bidirectional promoter region in detail: open boxes represent the protein-binding sites identified in vitro (7). The position of the PCR primers (shaded boxes) is shown. Lines represent restriction fragments used in gel shift assays. **C.** Sequence of the *Htf9* promoter. Potential protein-binding sites are indicated.

Oligonucleotides

Two combinations (labeled A and C) of nested primers were copied from the upper strand and one set (B) from the lower strand of *Htf9* (accession number X05830). HPLC-purified primers (Oswel Service, Department of Chemistry, University of Edinburgh) had the following sequences: A1 (5'-ACGGTCT-CGACTAGCTCA-3'), A2 (5'-TAGCTCAGGCGTCGC-TGGC-3'), A3 (5'-TGGCTCGGCGGCCTCTGG-3'), C1 (5'-CCTACCCCACCCGGACCT-3') and C2 (5'-GGACC-TCGCTGTACCTT-3'), corresponding to positions 734–751, 745–763, 760–777, 680–697 and 692–708; B1 (5'-TGCGCGCGTGGGTCGCCG- 3') and B2 (5'-TCGCCGGG-AGAGCGCGCT- 3') correspond to positions 1039–1022 and 1027–1010. The universal linker was composed of a partially complementary 25-mer and 11-mer of respective sequence 5'-GCGGTGACCCGGGAGATCTGAATTC-3' and 5'-CTA-GACTTAAG-3'. HPLC-purified oligonucleotides for in vitro assays (Genenco Service, Dipartimento di Genetica e Biologia Molecolare, Universitê di Roma) were as follows: 5'-CGTCA-CAAAGAGGCGGGGCTATGCGCATAG-3' and its complement represent a high-affinity Sp1-binding site; 5'-AGC-TTGAACCCTGACCCCTGACCCCAGCA-3' and its complement represent the *Htf9* GBF-binding site; 5'-CGCTGT-CGGAGCCAATAAAGCTACAC-3' and its complement represent the *Htf9* CCAAT site. Double-stranded oligonucleotides carrying the c/EBP (5'-GATCCAGGAATTACGAAATGG-AGGAG-3') and the E2F- (5'-TAGTTTTCGCGCTTAAA-TTTGA-3') binding sites were kindly given by A. Vitelli (IRBM Istituto di Ricerche in Biologia Molecolare, Pomezia, Italy) and A. Felsani (CNR, Istituto di Tecnologie Biomediche, Rome, Italy) respectively. AP1- (5'-TTCCGGCTGACTCATC-AAGCG-3') and AP2- (5'-GATCGAACTGACCGCCCGC-GGCCCGT-3') binding oligonucleotides were from Promega.

RESULTS

The mouse *Htf9* locus contains two divergent genes that are both ubiquitously transcribed from a shared bidirectional promoter (3). In the promoter region (Fig. 1) several factor-binding sites were identified in vitro (7). Transient expression assays showed that the CCAAT box was a major promoter element in hepatoma cells while being dispensable in fibroblasts; conversely, a major contribution of the Sp1.2 site to the overall promoter activity was observed in fibroblasts (8). To establish wheter these different requirements reflected cell-type specific patterns of trans-activation, we compared the in vivo interactions of nuclear factors with the *Htf9* promoter in different cell lines. To this aim cells were subjected to in vivo footprinting aided by ligation-mediated PCR. Cell lines with different biological characteristics and transcriptional abilities were chosen, though the process of in vitro adaptation may have somewhat quenched the original differences between cell types. The following lines were examined: NIH/3T3 and C3H/10T 1/2 cells, independently derived from mouse embryonic fibroblasts, with the 10T 1/2 cells retaining a broader pluripotency; rat liver hepatoma H4-II-E-C3 cells, expressing a large number of hepatic functions, such as albumin, tyrosine aminotransferase, transferrin, prothrombin; rat C6 cells, derived from a glial tumor and expressing the S-100

protein, typycal of vertebrate neural tissues; mouse S-20Y cells, a cholinergic neuroblastoma line retaining the choline acetyltransferase enzymatic activity (13); mouse CP1 embryonic stem cells, whose differentiation was prevented by differentiation inhibitory activity DIA/LIF (14). DMS titration experiments were initially carried out to establish the effective concentrations in each cell line and for particular sites. Thereafter three concentrations were routinely used in each experiment and a trend analysis enabled us to appreciate the protection of individual sites.

Protected *Htf9* elements in H4-II-E-C3 hepatoma cells

The *in vivo* pattern of the *Htf9* promoter in hepatoma nuclei is shown in Fig. 2. A protection was observed at position 859–875, overlapping one major transcriptional origin of both divergent genes. The protected element consisted of a direct repeat (GGG-ACTGGGGACTGGGG) recognized by a novel factor called G-binding factor, GBF (8; G. Di Matteo and P. L., unpublished). In vivo the element was protected on both strands and the binding was accompanied by the appearance of an additional band (position 856) at the 5′ edge of the protection (Fig. 2A).

Another footprint was evident over the sequence GAGCC-AATAAAGCTA (position 911–925), harboring a CCAAT box (Fig. 2A and 2B). The protection was flanked by hypersensitive bands on either side (positions 940 and 914 respectively, Fig. 2B). Fig. 2C shows the CCAAT site in different cell lines: a distinct footprint was seen in H4 cells only. Because H4 cells are derived from rat hepatoma, rat DNA was sequenced in vivo to assess any divergence which may have occurred between mouse and rat: in the rat sequence one A to G transition (position 926) was found 5′ to, but not affecting, the CCAAT site. Rat C6 cells, derived from a glial tumor, were also analysed as a species-specific control: no CCAAT protection was observed (Fig. 2C). On the other hand, a footprint of similar extension to that seen in H4 cells was mapped in mouse liver nuclei using DNase I (not shown): thus the CCAAT protection is indeed restricted to cells of hepatic origin regardless of the species.

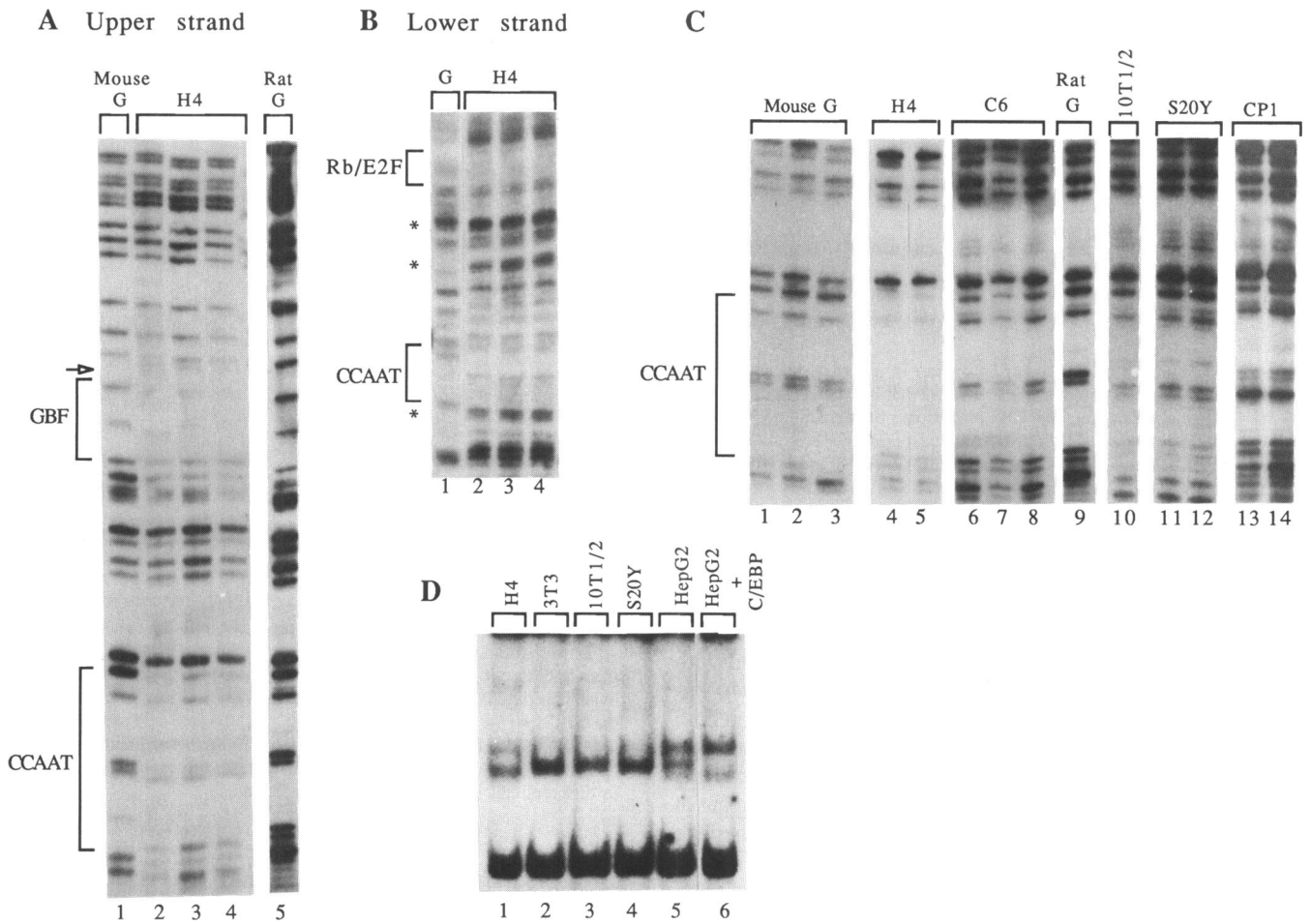

Figure 2. A. Genomic footprints on the *Htf9* promoter in H4 cells. 1. Mouse control DNA; 2–4: H4 cells exposed to 0.1, 0.2 and 0.3% DMS; 5: rat control DNA. Footprints at the GBF and CCAAT sites are bracketed. The arrowhead marks an extraband 5′ to the GBF protection. **B.** Genomic footprint of CCAAT box on the lower strand. 1: mouse control DNA. 2–4: H4 cells treated as for panel A. Asterisks mark DMS-hypersensitive sites. **C.** *In vivo* footprinting of the CCAAT box (upper strand) in all cell types. 1–3: mouse control DNA; 4–5: rat H4 hepatoma cells; 6–8: rat glioma C6 cells; 9: rat control DNA; 10: mouse 10 T1/2 fibroblasts; 11–12: neuroblastoma S-20Y cells; 13–14: mouse CP1 cells. DMS concentrations ranged between 0.1 and 0.3%. **D.** Gel shift assay of *Htf9*- probe 2 carrying the CCAAT box (see Fig. 1). Extract types (4 μg) are indicated above each lane. In lane 6 the reaction was preincubated with a 50-fold excess of c/EBP-binding oligonucleotide prior to addition of the probe.

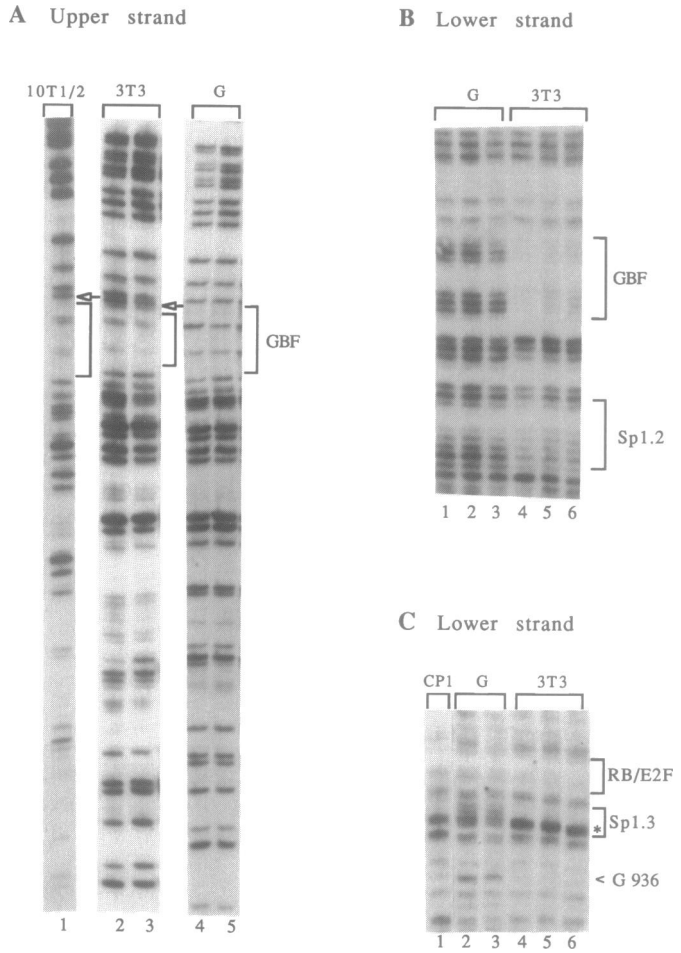

Figure 3. *In vivo* footprints on the *Htf9* promoter in fibroblast lines. **A.** Upper strand. 1: 10T 1/2 cells exposed to 0.3% DMS. 2–3: 3T3 cells exposed to. 0.1 and 0.2% DMS. 4–5: G ladder. The GBF protection is bracketed; the 5′ extraband is arrowed. **B.** Sites Sp1.2 and GBF (lower strand). DMS concentrations ranged between 0.1 and 0.3%. **C.** Sites Sp1.3 and E2F (lower strand). 1. CP1 cells (0.3% DMS); 2–3. G ladder control; 4–6: 3T3 cells treated as for panel B. Asterisks mark hyperreactive bands. Protein-binding sites are indicated. No binding activity was identified corresponding to the G396 protection in 3T3 cells (see text).

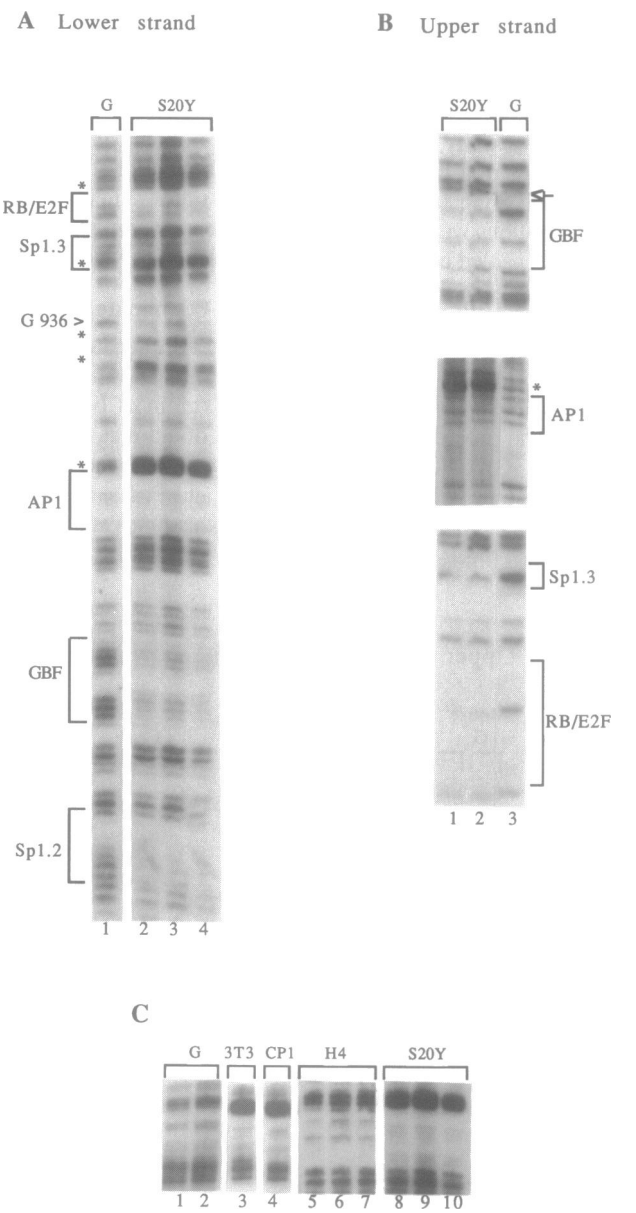

Figure 4. *In vivo* footprints on the *Htf9* promoter in S-20Y neuroblastoma cells. **A.** Lower strand. Lanes 1–2: G ladders from naked DNA controls 3–5: DNA from S-20Y cells exposed to 0.1, 0.2 and 0.3% DMS. Protections are bracketed, hypersensitive bands are marked by an asterisk. **B.** Upper strand. Footprints of the GBF-, AP1-, Sp1.3 and E2F-binding sites. 1–2: S-20Y cells were exposed to 0.1 and 0.3% DMS. 3: G control ladder. The arrowhead marks the extraband 5′ of the GBF protection; asterisks mark hypersensitive bands. **C.** The AP1 site in different cell lines. Cell types are indicated.

The potential binding ability of the CCAAT site was assessed in gel-shift assays. A synthetic 26-mer copied from the protected CCAAT window gave an identical shift upon incubation with all extracts (not shown); however, when we assayed a 40 bp-fragment (probe 2 in Fig. 1) carrying flanking sequences on both sides of the CCAAT box, both ubiquitous and hepatocyte-enriched protein complexes were detected (Fig. 2D): thus the *Htf9* sequences surrounding the core CCAAT box participate in stabilizing the binding of a hepatic factor. Several hepatic CCAAT-binding factors are known (21, 22). The largest family of liver-enriched factors includes the c/EBP protein group (22), whose binding specificity includes CCAAT boxes, the enhancer core motif and c-AMP response elements. We have established that the Htf9-CCAAT factor does not belong to the c/EBP family by site-specific competition assays (Fig. 2D), but have not pursued its identity any further.

Finally, a distinct footprint was seen at positions 964–974 (Fig. 2B) protecting the sequence TTTGGCGG, which matches a high-affinity site for the $p105^{RB}$/E2F complex (23). The protection was also visible on the complementary strand (not shown).

Protected *Htf9* elements in 3T3 and 10 T 1/2 fibroblasts

In both 3T3 and 10T 1/2 fibroblast lines protection of the GBF-binding site was visible on both strands (Fig. 3A and 3B), as in hepatoma cells. Three potential Sp1 sites occur in the *Htf9* promoter (see map in Fig. 1). No interaction was seen with site Sp1.1, whose sequence matches a reported medium-affinity binding site (24). On the other hand, sites Sp1.2 (CTCCGCCCCC, 840–849) and Sp1.3 (CCCGCCCC, 950–957) both match high-affinity sites for Sp1 (24). In vitro binding of fibroblast extracts to both Htf9 sites was very efficient and was competitively inhibited by an excess of Sp1-binding

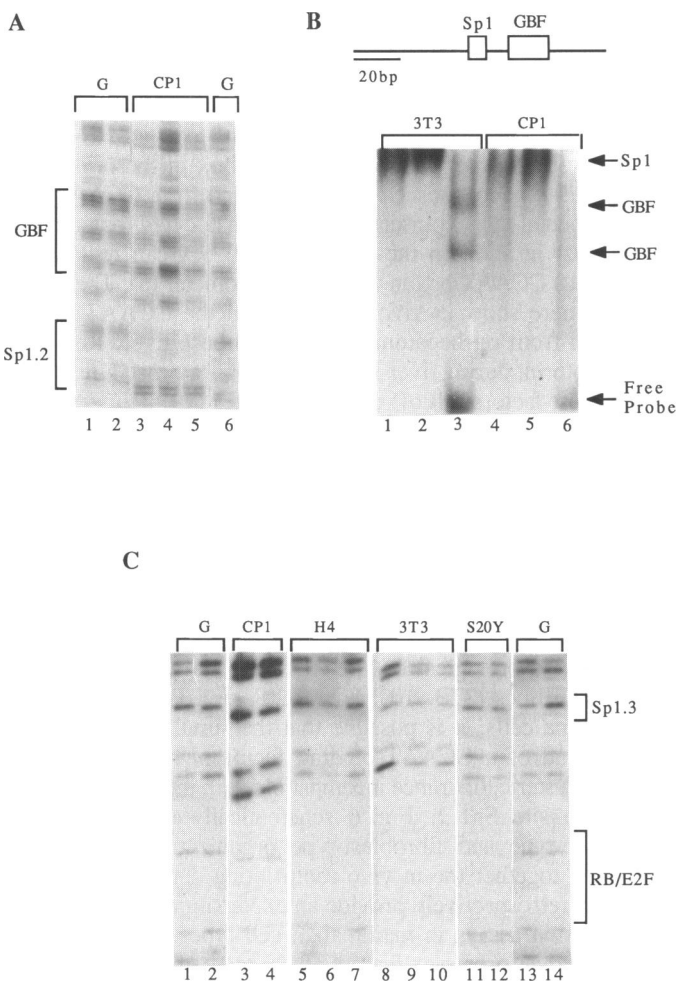

Figure 5. *In vivo* footprints in CP1 embryonic stem cells. **A.** Sp1.2 and GBF sites (lower strand). 1, 2 and 6: G ladder; 3–5: ES cells exposed to 0.1, 0.2 and 0.3% DMS: site Sp1.2 is footprinted while site GBF is not. **B.** Gel shift assay of probe 1, carrying both Sp1.2 and GBF sites. Lanes 1 and 2: 3 μg and 5 μg of 3T3 extract; lane 3: 5 μg of 3T3 extract preincubated with a 50-fold excess of Sp1-binding oligonucleotide: two complexes unaffected by the Sp1 competitor represent different froms of GBF complexes (G.Di Matteo, unpublished); Lanes 4–6: same reactions as in 1–3 using ES cell extracts: no complex is formed after preincubation with the Sp1 oligonucleotide. Migration of the free probe is indicated. **C.** The E2F site in different cell lines (upper strand). Cell types are indicated above each lane. DMS concentrations were between 0.1 and 0.5% DMS. Lanes 1–2 and 13–14: control G ladders. All lines but CP1 show a distinct E2F footprint (bracketed).

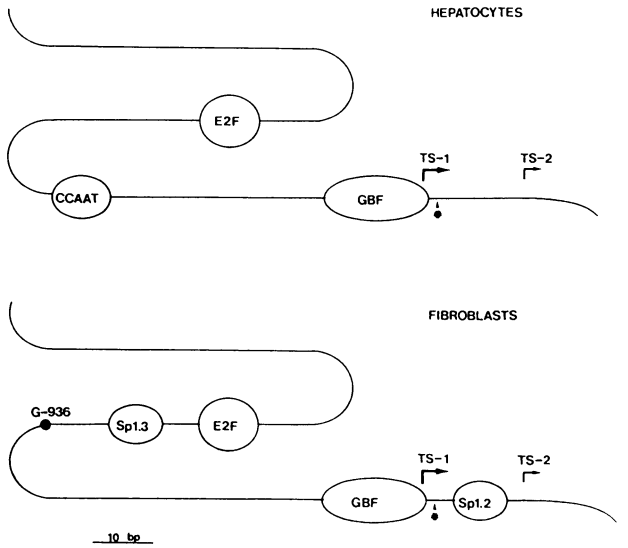

Figure 6. Summary of the *in vivo* protections detected at the *Htf9* promoter in hepatoma cells and in fibroblasts: the orientation of *Htf9-a* transcription is shown. TS-1 and TS-2 indicate the preferred sites of transcription initiation (see ref.3). The asterisk shows the hypersensitive band flanking the GBF protection.

oligonucleotide. In vivo protection of both sites Sp1.2 (Fig. 3B) and Sp1.3 (Fig. 3C) was evident on the G-rich lower strand.

The region surrounding site Sp1.3 showed an altered organization compared to the control ladder on the lower strand. Three footprints were individually distinguishable (Fig.3C). Inspection of the DNA sequence (Fig. 1C) revealed close recognition sites for known factors: site Sp1.3 (950–957) was preceded by the E2F-binding sequence TTTGGCGG (964–974) on the lower strand; the latter overlapped with a potential AP2-binding site (976–985) on the complementary strand. Site-specific competition assays were carried out to disentangle the binding events: binding to Htf9-probe 3 (see Fig.1B) was competitively inhibited by an excess of both Sp1- and E2F-binding oligonucleotides, while remaining unaffected by an AP2-binding competitor (not shown). Thus, both E2F and Sp1 factors are involved in binding to the region and respectively generate the footprints at sites 964–974 (E2F) and 950–957 (Sp1), flanked by hypersensitive bands at positions 946–947. No further binding activity was identified *in vitro*, though the G-residue at position 936 was clearly quenched *in vivo* (Fig. 3C). On the upper strand only the E2F footprint was apparent (see below and Fig. 5C).

Protected *Htf9* elements in S-20Y neuroblastoma cells

In S-20Y neuroblastoma nuclei the GBF-binding site was fully protected on both strands and was bordered by the characteristic 5′ band seen in all other cell lines (Fig. 4A-B). Adjacent to that site an element of similar sequence (TGATTC, position 895–900) to the recognition site for AP1 (TGAGTC) occurs, which was found to be a genuine AP1-binding site by gel-shift assays using a specific competitor (not shown). The AP1 site was protected and flanked by hypersensitive bands in S-20Y cells (Fig. 4A–B). No protection was seen in other examined cell lines (Fig. 4C).

The target sites for Sp1 and E2F showed a similar in vivo occupancy to that observed in fibroblast cells: both Sp1 sites were protected on the lower strand (Fig. 4A); the E2F site was protected on both strands (Fig. 4A–4B); adjacent footprints at sites E2F, Sp1.3 and G 936 on the lower strand generated an in vivo ladder similar to that seen in fibroblasts (compare Fig. 3C and 4A).

Protected *Htf9* elements in CP1 embryonic stem cells

In all somatic cell lines examined thus far the GBF site had been found to be similarly protected in vivo. An exception is represented by the CP1 embryonic stem cell line, in which no GBF footprint was detected (Fig. 5A); the adjacent Sp1.2 site was protected by an independent binding event on the lower strand. In vitro assays confirmed that embryonic cell extracts do not in fact express GBF (Fig. 5B). The pattern of the Sp1.3/E2F region in undifferentiated CP1 cells also differed from that seen

in other cell types: site Sp1.3 was protected on the the lower strand as in fibroblasts and neuroblastoma cells (Fig. 4C); in contrast the E2F site was not protected: therfore the lower strand footprint was of limited extension, and no protection at all was seen on the complementary strand. Fig. 5C shows that the lack of E2F protection is characteristic of ES cells.

DISCUSSION

During the process of differentiation and development the pool of transcription factors is diversified in different cell types. In addition the availability of the so-called general factors also fluctuates, as single promoters have different requirements for the factors involved in basal complexes (25). Housekeeping promoters are ubiquitously active despite these variations. *In vitro* evidence for cell-specific interactions was reported for the promoters of the *Htf9* (8), aldolase (26), B-polymerase (27) and Na,K-ATPase (28) housekeeping genes. Based on the in vitro results, one can predict that the trans-activated elements in housekeeping promoters vary from cell to cell: ubiquitous expression will result from cell-specific patterns of transcriptional activation.

Genomic footprinting provides the most reliable picture of the interactions occurring within the cell (rev. 29) and has proved extremely powerful in pinpointing the onset of activation in inducible (30–32) and developmentally regulated genes (18, 33–35). The technique has also depicted the alternative organization of the X-linked PGK promoter in its active and inactive state (20, 36). In these studies the occupancy of critical cis-active sequences appeared to be an all-or-nothing event related to the acquisition of an open chromatin conformation which triggered transcription. We have employed genomic footprinting to study the native organization of the *Htf9* promoter. We have regarded as informative the *Htf9* sites showing unambiguous protections in at least one cell type—though these may not represent the only functionally relevant elements—and have compared their status in cells expressing different specialized functions. An important point emerging from this study is that *in vivo* protections of the *Htf9* promoter reflected cell type-specific combinations of factors. In contrast, in vitro studies using isolated promoter elements often reflected all potential binding events.

Htf9, like most housekeeping promoters, lacks a TATA box; TATA-less promoters are thought to assemble transcriptional complexes via interactions differing from those occurring at TATA-containing promoters (37) and involving distinct initiators which are only beginning to be identified (38, 39). In this study one major divergent origin of transcription, coinciding with the GBF-binding element, was found to be fully protected on both strands in all differentiated cell lines. The GBF element is included in the shortest DNA fragment required for transcription in both orientations (7). Protection of the GBF site was accompanied by the appearance of a flanking band, which might either indicate a displacement of the G residue 856 at the 5′ border of the protected site, or cleavage of an exposed adenine at position 857, which would also be methylatable by DMS. Both possibilities suggest an altered chromatin structure around the transcription start site. It is noteworthy that the preferrred site of *Ht9-a* RNA initiation is located exactly at 857 (3). It is possible that GBF is a novel member of the protein group involved in transcription initiation in the absence of a TATA box—this question is currently being addressed in our laboratory. Embryonic stem cells, which are the only type lacking GBF, are known to express specific variants of certain 'somatic' factors: the elements required for promoter activity often differ in embryonic and differentiated cells. It will be interesting to ask whether the absence of GBF alters transcription initiation of the *Htf9-* transcripts in ES cells.

The *Htf9* promoter carries a CCAAT box in the orientation of the *Htf9-a* gene transcripition. Deletion of the CCAAT box is detrimental for transcription in hepatocytes but not in fibroblasts; in addition the Htf9-CCAAT box can replace the endogenous CCAAT box in the albumin promoter (8). The results reported here show in vivo occupancy of the CCAAT box in the nuclei from rat hepatoma and from mouse liver. Retention of the footprint during liver nuclei isolation—a procedure during which most factors fall off their target sites (19)—suggests that the interaction is remarkably stable.

On the other hand, a promoter fragment containing only the GBF and Sp1.2 sites was previously found to be sufficient for high expression in fibroblasts. In the present study site Sp1.2 appeared to be effectively protected on the lower strand. The presence of one single guanine in the site made the protection difficult to assess on the upper strand, thus we could not establish whether the asymmetrical footprint at the Sp1.2 site reflects a technical or a biological feature. Site Sp1.2 was footprinted in several cell lines (3T3, 10T 1/2, S-20Y and ES cells) but not in hepatoma cells: it is possible that the simultaneous binding of GBF and of a CCAAT-binding factor in hepatoma nuclei generates a steric hindrance incompatible with the further binding of Sp1 to site Sp1.2. Fig. 6 schematically summarizes the hepatocyte-type and fibroblast-type organization of the *Htf9* promoter: together the in vivo footprinting data are consistent with, and retrospectively provide an explanation for, the results of expression assays in mammalian cell types.

A few novel promoter features have also emerged from this study. A cell-type restricted interaction was seen at the AP1 site in S-20Y cells though no functional analysis was carried out in this cell line. The footprint did not require TPA-induction as reported for the in vivo binding of AP1 to the *c-fos* promoter (40). The footprint might be due to a neuronal AP1 subtype indistinguishable from the AP1 factor in gel-shift assays, or to the interaction of neuronal factor(s) with AP1 (or the AP1-related protein) which might stabilize the binding to the *Htf9* promoter.

The distal promoter region showed an interesting organization, which has not yet been functionally dissected but may have a significant role in control of *Htf9-a* transcription. A complex pattern was seen in neuroblastoma and fibroblast cells. Alternating hypersensitive and protected elements indicated that more than one protein were involved in the footprint. Sp1 was one of the factors involved and protected the Sp1.3 box on the lower strand in fibroblasts, neuroblasts and embryonic stem cells—the site could not be unambiguously resolved in hepatoma nuclei because of its compression between the E2F and CCAAT protections. The element TTTTGGCGGG was also protected in fibroblasts and neuroblastoma cells. The protected site represents a perfect match to sites selected in vitro by complexes containing the $p105^{RB}$(retinoblastoma) protein (23)—to which the E2F factor is also thought to associate—and is identical to the E2F site in the *c-myb* promoter (rev. 41). The involvement of E2F is compatible with the lack of protection in undifferentiated embryonic stem cells, as published evidence reported very low levels of the embryonal E2F-like factor DRTF1 in teratocarcinoma cells prior to retinoic acid-induced differentiation (42–43). To our knowledge this is the first demonstration of

the in vivo occupancy of an E2F-binding site in proliferating cells. In the light of our observations that *Htf9-a* is expressed in cycling cells in which the $p105^{RB}$ protein was sequestrated by the E1A oncogene, while being repressed in cells transfected with E1A mutants failing to interact with $p105^{RB}$ (A.Bressan, M.Caruso, A.Felsani and P.L., unpublished), occupancy of the $p105^{RB}$/E2F target is very intriguing.

In fibroblast and neuroblastoma nuclei the promoter ladder appeared to be altered beyond site Sp1.3 on the lower strand. No corresponding alteration was seen on the upper strand, which parallels the strand-specificity of the Sp1.3 footprint. The altered organization was apparent in cell types in which both Sp1.3 and E2F-binding sites were occupied, and in our opinion may reflect a local distortion resulting from the simultaneous binding of E2F and Sp1 to adjacent sites, rather than the binding of a novel protein to an independent sequence. Sp1.3 and E2F are separated by the sequence 5'-GGAAGCGCGG-3', which is part of the recently described SCE element separating the CG-box 1 from the E2F site in the *dhfr* promoter (44). The SCE is not a protein-binding site, yet exerts a repressive function and is thought to generate a distortion of the DNA structure characteristically framed by hypersensitive sites: this would prevent the 'cross-talk' between the Sp1 and E2F proteins when *dhfr* expression is not required. It is tempting to suggest that the altered structure in the distal region of *Htf9* also indicates an element of structural distortion which could serve a similar function and inhibit *Htf9-a* expression in G0 cells.

Our current understanding of transcriptional regulation is based on combinatorial models. The*Htf9* promoter shows an intrinsic flexibility to act as the target of different combinations of factors, which may enable it to direct transcription in cells equipped with different transcriptional machineries. The cooperation between cell-specific and ubiquitous factors in activating different sets of regulatory elements may represent a generalised mechanism maintaining the ubiquitous expression of housekeeping functions during differentiation.

ACKNOWLEDGEMENTS

The *in vivo* footprinting experiments were carried out at the AFRC Centre for Genome Research (Edinburgh, UK) in the laboratory of R.Lathe. We are grateful to G.Pfeiffer and H.Saluz for protocols and help in setting up this work. We also thank A.Smith for ES cells, A.Vitelli and A.Felsani for specific oligonucleotides, A.Bressan and G.Di Matteo for many contributions to the work, and J.Trueheart and J.Thorner for communicating results before publication. R.Perry, H.Saluz and R.Lathe are gratefully acknowledged for their critical comments on this manuscript. During part of this work P.L. was on leave of absence from Centro di Genetica Evoluzionistica del CNR (Rome, Italy) and was partly supported by a Senior fellowship in the framework of the BRIDGE program of the European Communities. This work was supported by Progetto Finalizzato Ingegneria Genetica, Fondazione Istituto Pasteur-Cenci Bolognetti and Programma Bilaterale 104039/04/9210648 Consiglio Nazionale delle Ricerche (Italy), and by the AFRC (UK).

REFERENCES

1. Dynan, W. S. (1986) *Trends Genet.* **6**, 196–198.
2. Bird, A.P. (1987) *Trends Genet.* **3**, 342–347.
3. Lavia, P., Macleod, D. and Bird, A.P. (1987) *EMBO J.* **6**, 2773–2779.
4. Linton, J.P., Yen, J., Selby, E., Chen, Z., Chinsky, J.M., Lin, K., Kellems, R. E. and Crouse, G. F. (1989) *Mol. Cell. Biol.* **9**, 3058–3072.
5. Lennard, A. and Fried, M. (1991) *Mol. Cell. Biol.* **11**, 1281–1294
6. Bressan, A., Somma, M. P., Lewis, J., Santolamazza, C., Copeland, N., Gilbert, D., Jenkins, N. and Lavia, P. (1991) Gene **103**, 201–209.
7. Somma, M. P., Pisano, C. and Lavia, P. (1991) *Nucleic Acids Res.* **19**, 2817–2824.
8. Somma, M. P., Gambino, I. and Lavia, P. (1991) *Nucleic Acids Res.* **19**, 4451–4458.
9. Melton, D., Mc Ewan, C., Mc Kie, A. and Reid, A. (1986) *Cell* **44**, 314–328.
10. Osborne, T., Gil, G.,. Brown, M. S, Kowal, R. and Goldstein, J. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 3614–3618.
11. Atchinson, M.L., Meyuhas, O. and Perry, R.P. (1989) *Mol. Cell. Biol.* **9**, 2067–2074.
12. Smith, M. L., Mitchell, P. J. and Crouse, G. F. (1990) *Mol. Cell. Biol.* **10**, 6003–6012.
13. Kirshner, J. A., Markelonis, G. J., and Max, S. R. (1986) *J. Neurochem.* **46**, 817–821.
14. Smith, A. G. (1991) *J.Tiss. Cul. Meth.***13**, 89–94.
15. Dignam, J.D., Lebovitz, R.M. and Roeder, R. (1983) *Nucleic Acids Res.* **11**, 1475–1489.
16. Andrew, N. and Faller, D. V. (1991) *Nucleic Acids Res.* **19**, 2499.
17. Pfeiffer, G., Steigenwald, S., Mueller, P. R., Wold, B. and Riggs, A. D. (1989) *Science* **246**, 810–813.
18. Saluz, H. and Jost, J. P. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 2602–2606.
19. Mirkovitch, J. and Darnel, J. E. (1991) *Genes Dev.* **5**, 83–93.
20. Pfeiffer, G. and Riggs, A. D. (1991) *Genes Dev.* **5**, 1102–1113.
21. Raymondjean M., Cereghini S. and Yaniv M. 1988, Proc. Natl. Acad. Sci. USA **85**: 757–761.
22. Faisst S. and Meyer S., 1992. Nucl. Acids Res. **20**: 3–26).
23. Chittenden, T., Livingston, D. and Kaemlin, W.G. (1991) *Cell* **65**, 1073–1082.
24. Kadonaga, T., Jones K. A. and Tjian R. (1986) *Trends Biochem.* **11**, 20–23.
25. Parvin, J. D., Timmers, H. T. and Sharp, P. A. (1991) *Cell* **68**, 1135–1144.
26. Concordet, J.P., Maire, P., Kahn, A. and Daegelen, D. (1991) *Nucleic Acids Res.* **19**, 4173–4180.
27. Englander, E. W. and Wilson, S. H. (1990) *Nucleic Acids Res.* **18**, 919–928.
28. Suzuki-Yagawa, Y., Kawakami K. and Nagano K. (1992) *Mol. Cell. Biol.* **12**, 4046–4055.
29. Saluz, H, Wiebauer K. and Wallace, A. (1991) *Trends Genet.* **7**, 207–211.
30. Mueller, P. R., Salser, S. J. and Wold, B. (1988) *Genes Dev.* **2**, 412–427.
31. Saluz, H., Feavers, I. M., Jiricny, J. and Jost, J. P. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 6697–6700.
32. Abravaya, K., Phillips, B. and Morimoto, R. I. (1991) *Mol. Cell. Biol.* **11**, 586–592.
33. Becker, P.B., Ruppert, S. and. Schutz, G. (1987) *Cell* **51**, 435–443.
34. Mueller, P. R. and Wold, B. (1989) *Science* **246**, 780–786.
35. Rigaud, G., Roux, J., Pictet, R. and Grange, T. (1991) *Cell* **67**, 977–986.
36. Pfeiffer, G., Tanguay, R.L., Steigerwald, S. and Riggs, A.D. (1990) *Genes Dev.*4, 1277–1287.
37. Pugh, B.F. and Tjian, R. (1991) *Genes Dev.* **5**, 1935–1945.
38. Smale, T. S. and Baltimore, D. (1989) *Cell* **57**, 103–113.
39. Means, A. L. and Farnham, P. J. (1990) *Mol. Cell. Biol.* **10**, 653–661.
40. Dey, A., Nebert, D. and Ozato K. (1991) *DNA Cell Biol.* **10**, 537–544.44.
41. Nevins, J.R. (1992) *Science* **25**, 424–429.
42. La Thangue, N. B., Thimmappaya, B. and Rigby P. W. J. (1990) *Nucleic Acids Res.* **18**, 2929–2938.
43. Hamel, P. A., Gill, R. M., Phillips, R. A. and Gallie, B. L. (1992) *Mol. Cell. Biol.* **12**, 3431–3438.
44. Pierce, A.J., Jambou, R.C., Jensen D.E. and Azizkhan, J.C. (1992) *Nucleic Acids Res.* **20**, 6583–6587.