Nucleic Acids Research, 1996, Vol. 24, No. 13 **2597–2605**

Comparative and functional analysis of the AP2 promoter indicates that conserved octamer and initiator elements are critical for activity

Peter C. Creaser, David A. D'Argenio and Trevor Williams*

Department of Biology, Yale University, PO Box 208103, New Haven, CT 06520-8103, USA

Received February 5, 1996; Revised and Accepted May 17, 1996

GenBank accession nos X95234–X95236 (incl.)

ABSTRACT

AP-2 is a developmentally-regulated transcription factor expressed in ectodermal cell lineages. The AP-2 protein is essential for neural tube, craniofacial and body wall morphogenesis and has been implicated in oncogenesis. Here we report the isolation of the AP-2 promoter from human, mouse and chicken. The initiation sites for the human gene have been mapped in a variety of cell lines, including several derived from breast tumours. Initiation occurs just upstream of an IR3-like repetitive element, present in the human and mouse genes, but absent in chicken. The *cis*-acting elements responsible for promoter activity in human HeLa cells have been mapped both *in vivo* and *in vitro*. The proximal promoter contains binding sites for transcription factors AP-2, NF-1 and octamer proteins, but lacks a TATA box motif. Functional analysis demonstrates that the octamer binding site is the critical component of basal promoter activity. In addition, the promoter relies on an initiator element for efficient start site utilization. There is an excellent correlation between the requirement for the initiator and octamer elements in transcription assays and the conservation of these *cis*-acting sequences between chicken, mouse and human.

INTRODUCTION

The transcription factor AP-2 is a critical regulatory molecule required for vertebrate development. The AP-2 protein binds as a homodimer to the consensus recognition sequence GCCNNNGGC that is an important *cis*-regulatory element for a variety of cellular genes and viral genomes, including keratin, proenkephalin and MMTV (1–3). In addition, the pattern of AP-2 expression is regulated both temporally and spatially during mouse embryogenesis beginning around day E8.5 (4). The majority of AP-2 RNA and protein are found in ectodermal cell lineages, including the neurepithelium and neural crest (4–6). The tissue-specific distribution of AP-2 correlates with the differential expression of the gene in various cell lines. For example, AP-2 mRNA and protein are present in epithelial-derived HeLa cells, but absent from HepG2 hepatoma cells (7). Furthermore, AP-2 expression is induced in teratocarcinoma cell lines undergoing differentiation in response to the morphogen retinoic acid (6–9). The importance of this transcription factor for normal mammalian embryogenesis was recently revealed by generating mice containing a homozygous disruption of the AP-2 gene. These AP-2-null mice die at birth of severe developmental abnormalities including exencephaly, craniofacial defects and a failure of ventral body wall closure (10) .

The AP-2 protein is also associated with cellular transformation. In PA1 human teratocarcinoma cells, the addition of an AP-2 expression vector leads to anchorage independent growth (11). Furthermore, the AP-2 protein is highly expressed in a particular class of human breast cancer cell lines. Approximately 30% of human breast cancers contain high levels of c-erbB2, a tyrosine kinase receptor that causes mammary carcinoma in animal model systems (12–14). The region of the c-erbB2 promoter responsible for increased expression in human breast cancer cell lines has been mapped (15) and identified as an AP-2 responsive element (16). Moreover, there is a correlation between the presence of AP-2 and elevated levels of c-erbB2 protein in human breast cancer cell lines, strongly suggesting that AP-2 may be directly responsible for c-erbB2 overexpression (16). Therefore, the mechanism of regulation of the transcription factor AP-2 may provide an important key to understanding the aetiology of mammary carcinoma. In this study we have begun to address this question by characterizing the AP-2 promoter from three vertebrate species and identifying the *cis*-regulatory elements critical for its basal level of expression.

MATERIALS AND METHODS

Isolation and sequencing of genomic clones

An *Eco*RI–*Nco*I fragment of the human AP-2 cDNA spanning from nucleotides (nt) 1–1289 (7) was used to probe a human male circulating lymphocyte genomic cosmid library and a mouse C57 Black/6 female liver genomic cosmid library (Stratagene). Restriction fragments surrounding the 5′-end of the AP-2 cDNA were subcloned from positive cosmids and sequenced on both strands. The chicken promoter sequence was obtained by PCR from genomic DNA (gift of H. Belting) using primers based on mouse/human homology. Flanking sequence was then obtained using inverse PCR as follows. Genomic DNA was digested with the restriction enzymes *Nla*III, *Taq*I, or *Pst*I plus *Nsi*I, followed by religation at low DNA concentration. PCR was then performed

* To whom correspondence should be addressed

using the primers CK6 (GGATC GGACC CTCTC CCGCC GACCC) and CK7 (GCTTT ACCCG CAGCC GGAGC GCCTC ATTAG C). The sequence was obtained directly by cycle sequencing (Perkin Elmer) and was identical for all three enzyme digests.

Plasmid constructs

The plasmids designated +275 were made by fusing the *Sau*3AI site at +275 of the human genomic sequence, located 8 nt upstream of the translational initiator, to the *Bgl*II site of pBLCAT3Δ (8,17). The plasmids designated +37 were generated by fusing an *Ecl*136II site at +37 of the human genomic sequences to the T4 DNA polymerase repaired *Xho*I site in pBLCAT3Δ. The extent of upstream sequences is as follows: –500 is an *Xho*I partial digest and –190 is an *Xho*I digest into the *Sal*I site of pBLCAT3Δ; –151 is an *Xba*I digest into the *Xba*I site of pBLCAT3Δ; –122 and –99 are *Stu*I and *Hae*III digests into the T4 DNA polymerase repaired *Sal*I site of pBLCAT3Δ; –82, –49 and –10 were made using PCR and introduce a *Sal*I site that was ligated into the *Sal*I site of pBLCAT3Δ. The relevant sequences at the junction sites are; gtc gac TAC TGG CGA (–82), gtc gac tTA ATG AGG (–49); gtc gac GGA AAA GTT (–10). The Δoct, Δctf and LS templates were made in the –151 backbone using PCR or direct insertion of double-stranded oligonucleotides. All constructs were confirmed by sequencing.

In vitro transcription reactions

HeLa nuclear extract was prepared according to Dignam (18) as modified by Wildeman (19). Each *in vitro* transcription reaction (20 µl) contained 4 µl nuclear extract in 10 mM HEPES–KOH (pH 7.9), 0.5 mM DTT, 0.1 mM EDTA, 1 mM $MgCl_2$, 10 mM KCl, 10% glycerol, 4 mM spermidine, 0.5 mM each rNTP. Where appropriate, each reaction contained 200 ng template DNA linearized with *Nco*I. α-Amanitin (5 µg/ml) was included where indicated. Reactions proceeded for 60 min at 30°C and were then analyzed by S1 nuclease mapping or primer extension. All transcription reactions were performed independently at least twice. Primer extension analysis was conducted using a kinase labelled 24 nt CAT primer 5′- GCCAT TGGGA TATAT CAACG GTGG-3′ which was annealed in water to the transcription products at 37°C for 10–20 min. Primer extension reactions proceeded for 60 min at 37°C in 1× buffer (Gibco BRL), 10 mM DTT, 0.5 mM each dNTP, 20 µg/ml actinomycin D, 0.5 U/µl RNasin (Promega) and 2.4 U/µl superscript II reverse transcriptase (Gibco BRL).

S1 nuclease mapping

Total RNA was prepared by the guanidine isothiocyanate procedure followed by centrifugation through a caesium step gradient (20). RNA from *in vitro* transcription reactions was isolated as above. S1 probe used for mapping either endogenous transcripts, or RNA generated by *in vitro* transcription of templates extending to +275, was derived from the –49/+275 plasmid. This plasmid was kinase labelled at an *Ngo*MI site at +195, followed by an *Sph*I restriction digest, which cuts in the pBLCAT3 polylinker sequences upstream of the AP-2 insert. The S1 probe used for mapping transcripts derived by *in vitro* transcription of templates extending to +37 was generated from the –190/+37 plasmid. This plasmid was kinase labelled at a *Pvu*II site in the CAT gene, followed by a second digest with *Sph*I as above. In both instances, the strands were then separated on a standard sequencing gel. Hybridization between probe and RNA was performed at 42°C for at least 8 h in 50% formamide, 40 mM Pipes pH 6.4, 1 mM EDTA and 0.4 M NaCl before S1 nuclease digestion at 37°C for 30 min in the absence of carrier DNA (20).

DNase I footprinting

The plasmid used for footprinting was the human genomic clone from the *Xho*I site at –190 to a *Bam*HI site in the transcribed region of the gene, cloned into pBluescript II (Stratagene). This plasmid was kinase labelled at the *Acc*65I site in the vector polylinker, then restricted with *Ecl*136II (+37) and the footprinting probe subsequently purified by PAGE. DNAse I footprinting with HeLa nuclear extract and purified HeLa AP-2 or CTF protein was as described previously (7).

EMSA

HeLa nuclear extracts were made according to the procedure of Hurst (21). Assays were performed as described previously (3). Where appropriate, the reactions were incubated with either the competitor oligonucleotides for 10 min, or with 1 µl of an anti-oct1 or anti-oct2 antiserum (Santa Cruz Biotechnology) for 60 min on ice, before addition of the kinase labelled AP octamer probe.

Transfections

Co-transfections were performed by the calcium phosphate method as previously described (3) using 18 µg of AP-2 /+37 CAT3Δ reporter construct and 12 µg pUC118 DNA as carrier. Transfections were performed at least twice in duplicate, and the extracts normalized for protein concentration prior to the CAT assay.

RESULTS

Identification of the AP-2 promoter

The human AP-2 coding region was used to isolate four independent human genomic cosmid clones, and the sequence encompassing the 5′ boundary of the AP-2 cDNAs was then determined (shown diagrammatically in Fig. 1; see also Fig. 10). This analysis revealed the presence of an ~140 nt copy of a CT-rich repeat sequence in the 5′ non-coding region of the AP-2 transcript. This repeat, based on the triplet TCC, is a member of the IR3-like repetitive element family that was initially described in the genome of EBV, but has since also been identified in a number of cellular genes (22). It might be expected that this repetitive element could complicate the analysis of the AP-2 promoter. First, repetitive elements often harbour internal promoter elements and so it could function as an integral part of the AP-2 upstream regulatory region. Secondly, the presence of an IR3-like repeat would make it more difficult to analyze transcription of the gene by techniques such as RNase protection because of the presence of other cellular transcripts containing related sequences. Therefore, to determine which region corresponded to the AP-2 promoter, S1 nuclease mapping was utilized to locate the 5′-end of the AP-2 mRNA from human HeLa cells. In contrast to a uniformly labelled probe, typical of an RNase protection, any of the S1 probe annealing to related repeats in other transcripts will have the label present in the unique sequence removed by nuclease digestion. Thus, the spurious IR3 hybrids will be eliminated from the analysis and only genuine AP-2 transcripts will be observed.

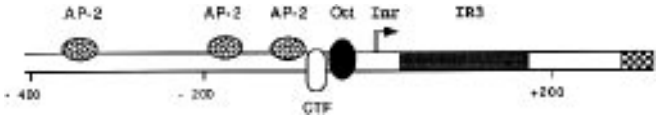

Figure 1. Diagrammatic representation of the human AP-2 promoter sequence. Nucleotide numbers are given based on the transcription initiation site at +1 (arrow). The IR3-like repetitive sequence (shaded), the position of the coding sequences (chequered box), and transcription factor binding sites (ovals), are shown.

The probe utilized for S1 nuclease mapping was end-labelled in unique sequence downstream of the IR3-like repeat, and extends ~250 nt upstream. Several nucleotides derived from linker sequence were included at the 3′-end of the probe so that any transcription initiating further upstream would generate a 244 nt band that could be distinguished from the 261 nt undigested probe (Fig. 2). The S1 mapping data obtained using RNA from HeLa cells, from which the AP-2 protein was originally isolated, show a series of protected fragments of 185–195 nt in length (Fig. 2, lane 1). In contrast, these transcripts are not observed with RNA derived from HepG2 cells (Fig. 2, lane 2), which lack AP-2 mRNA and protein (7). These data suggest that the transcription start sites for AP-2 map to 30–40 nt upstream of the IR3-like repeat.

The data from S1 mapping are consistent with transcriptional initiation, but could also represent a splice acceptor site. Therefore, to determine directly if the AP-2 upstream region possessed promoter activity, the ability of this region to direct accurate transcriptional initiation was tested in an *in vitro* transcription system derived from HeLa cell nuclei. A plasmid containing the AP-2 upstream region from –151/+275 was added to an *in vitro* transcription system. The resulting RNA transcripts were then subjected to S1 analysis using the same probe employed for mapping endogenous HeLa transcripts. The data show that the 5′-ends of the RNA produced in tissue culture cells, or by *in vitro* transcription, map to the same major start sites (Fig. 2; compare lanes 1 and 4). Therefore, this region of AP-2 can act as the promoter and has the ability to direct the specific pattern of initiation seen with the endogenous gene. Further mapping of the start site by S1 mapping, RNAse protection and primer extension analyses position the 5′-end of the AP-2 mRNA near the C designated +1 (see below and Fig. 10). Taken together, these results indicate that we have mapped the major AP-2 start sites and identified a fragment of the AP-2 gene that functions *in vitro* as the promoter.

We next determined the activity of the endogenous AP-2 promoter in a variety of other human cell types. In particular, several breast cancer cell lines were analyzed because of the correlation between the presence of AP-2 protein and elevated levels of c-erbB2 transcripts in these cells (16). In all cell lines where AP-2 RNA was detected, it initiated at the same sites as those used in HeLa cells, and no transcripts originated from within the repetitive element (Fig. 3).

Characterization of AP-2 promoter activity in cell-free extracts

The functional organization of the AP-2 promoter was further analyzed by using a series of 5′ deletion mutants as templates for *in vitro* transcription. Figure 4 lane 2 shows the basal amount of endogenous AP-2 RNA present in the HeLa nuclear extract used

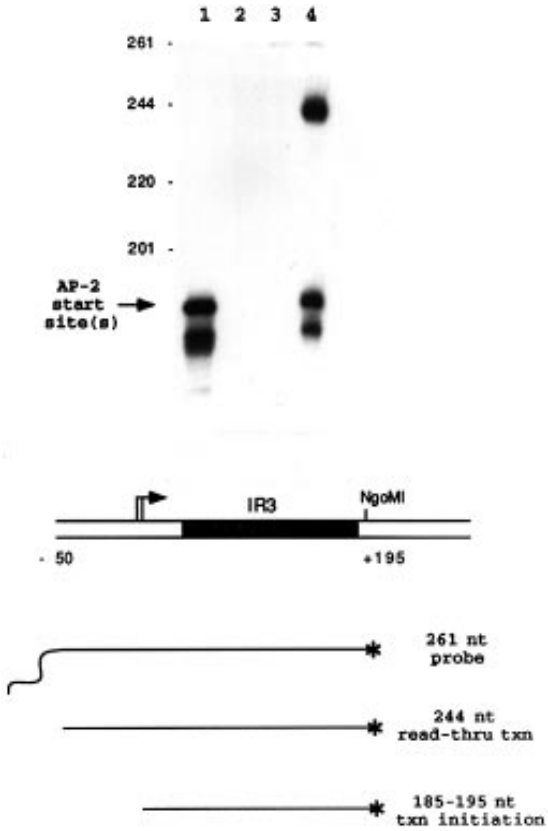

Figure 2. Characterization of the AP-2 transcription initiation sites. (**Top**) S1 nuclease mapping of HeLa RNA (lane 1), HepG2 RNA (lane 2), tRNA (lane 3) and RNA derived from the AP-2 promoter by *in vitro* transcription (lane 4). Size markers are indicated at the left. (**Bottom**) Diagrammatic representation of the AP-2 promoter, the S1 probe and the observed products. The 244 nt signal (lane 4) represents RNA polymerase transcripts initiating upstream of the AP-2 promoter in vector sequences.

for the transcription reactions in the absence of any added template DNA. Inclusion of the longest template tested, –190/+275, caused a marked increase in the 185–195 nt products derived from the AP-2 promoter (Fig. 4; compare lanes 2 and 8). These AP-2 transcripts are sensitive to the low concentrations of α-amanitin that specifically inhibit RNA polymerase II (compare lanes 1 and 8). In contrast, the 244 nt products initiating upstream of the AP-2 sequences are not eliminated by this concentration of α-amanitin, indicating that they are products of other RNA polymerases. Successive deletion of upstream sequences between –190 and –82 had no effect on the amount of *in vitro* transcription observed (Fig. 4, lanes 4–8). However, deletion to –49 caused a severe reduction of transcription (Fig. 4, lane 3). These experiments map the major determinants for *in vitro* transcription of the AP-2 promoter downstream of –82.

Primer extension analysis was utilized to analyze the expression of templates containing deletions nearer to the start site. To increase the resolution of start site mapping a new set of promoter templates was derived in which the AP-2 sequences upstream of nt +37, which lies immediately prior to the IR3-like repeat, were fused to the CAT gene. *In vitro* transcription products derived from these new AP-2 templates were analyzed by primer extension analysis using an oligonucleotide within the CAT gene (Fig. 5). The primer extension analysis indicated that the majority

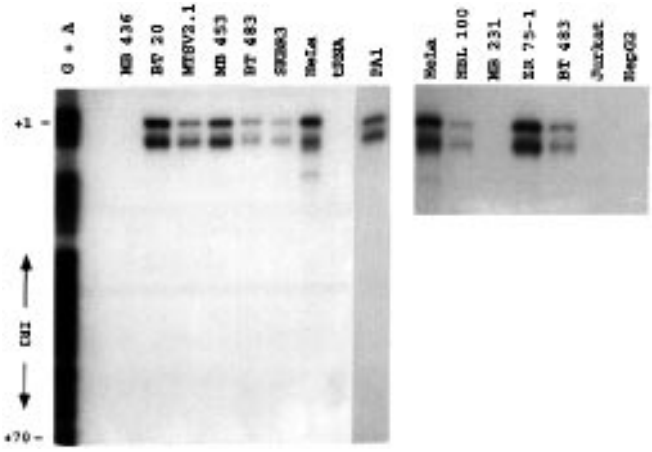

Figure 3. S1 mapping of the AP-2 promoter in human cell lines. MDA MB 436, HBL100, MDA MB 231, BT20, MTSV2.1 are breast cancer cell lines that have low levels of c-erbB2 expression. MDA MB 453, BT483, SKBR3, ZR 75-1 are breast cancer cell lines that are high expressors of c-erbB2. PA1 and Jurkat are teratocarcinoma and T cell lines, respectively. A G+A sequencing ladder of the probe and the position of the IR3-like repeat are indicated. The presence of intact RNA was confirmed in all samples using an S1 probe specific for the transcription factor Sp1 (data not shown).

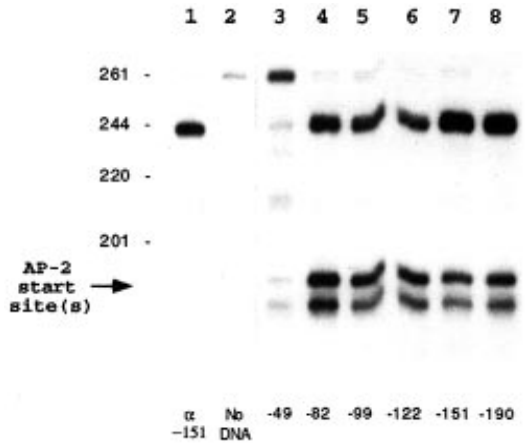

Figure 4. Analysis of the AP-2 promoter. 5′ deletion mutants were tested for promoter activity using a coupled *in vitro* transcription: S1 mapping protocol. Promoter templates used were; –151/+275 with α-amanitin (lane 1), no template (lane 2), –49/+275 (lane 3), –82/+275 (lane 4), –99/+275 (lane 5), –122/+275 (lane 6), –151/+275 (lane 7) and –190/+275 (lane 8). The S1 probe is derived from the –49 construct and therefore generates a 261 nt signal from upstream transcripts when used for the –49 template (lane 3), in contrast to the 244 nt product observed with other templates.

of transcripts mapped to an initiation site in the TC-rich region around +1, in agreement with the S1 nuclease mapping data. The absence of the other slightly shorter products, observed only in the S1 nuclease analysis, suggests that these latter molecules were produced by enzyme nibbling of the AT-rich region just downstream of the start site. Also in agreement with the S1 mapping data, a significant drop in the level of product was observed when the sequences between –82 and –49 were deleted (Fig. 5, lanes 3 and 4). Finally, the more extensive removal of the sequences upstream of –10 completely eliminates specific transcription (Fig. 5, lane 2).

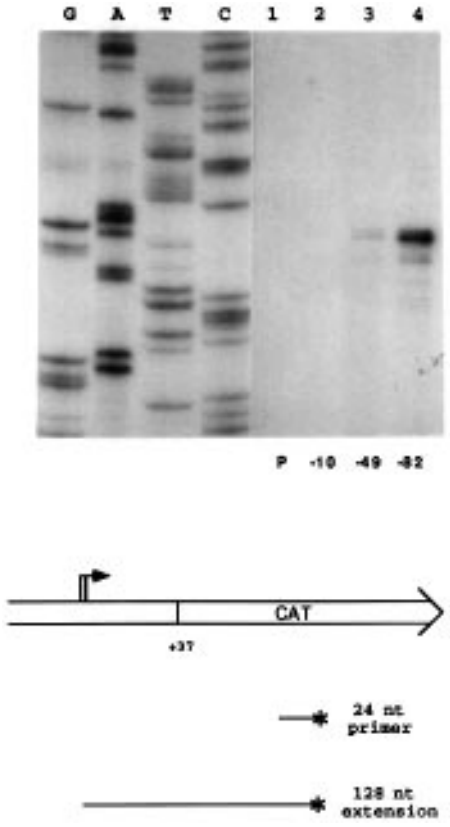

Figure 5. Analysis of the AP-2 promoter by coupled *in vitro* transcription/primer extension assay. Results obtained with pBLCAT3Δ (lane 1), –10/+37 (lane 2), –49/+37 (lane 3) and –82/+37 (lane 4) are shown next to a sequencing ladder of the AP-2 promoter/CAT fusion template generated using the same primer. The template, primer and products are illustrated schematically beneath.

Identification of proteins interacting with the AP-2 promoter

DNase I footprinting assays were performed using the nuclear extract employed in the *in vitro* transcription studies to identify proteins that interacted with the basal promoter region (Fig. 6). The most noticeable footprint occurred between nucleotides –39 and –56, which maps to the critical region of the promoter identified in the transcription assays (Fig. 6, left). This region conforms strongly to the octamer consensus sequence, differing by only 1 nt from the canonical binding site (23).

The potential octamer binding site at –49 was further characterized by gel mobility shift experiments using an oligonucleotide corresponding to this sequence. Incubation of this probe with HeLa nuclear extract produced one retarded band that was supershifted with an anti-oct1 antiserum, but not an antiserum against oct2 (Fig. 7, lanes 2, 3 and 4 respectively). Moreover, the appearance of the retarded band was specifically inhibited by consensus octamer competitor DNAs, but not by mutant octamer sequences or unrelated competitor DNAs (Fig. 7, lanes 5–15). These data indicate that in HeLa extract, oct1 binds strongly to the AP-2 promoter.

No other strong binding activity was observed using crude nuclear extract. However, DNase I footprinting analysis of the proximal promoter sequences using purified transcription factors

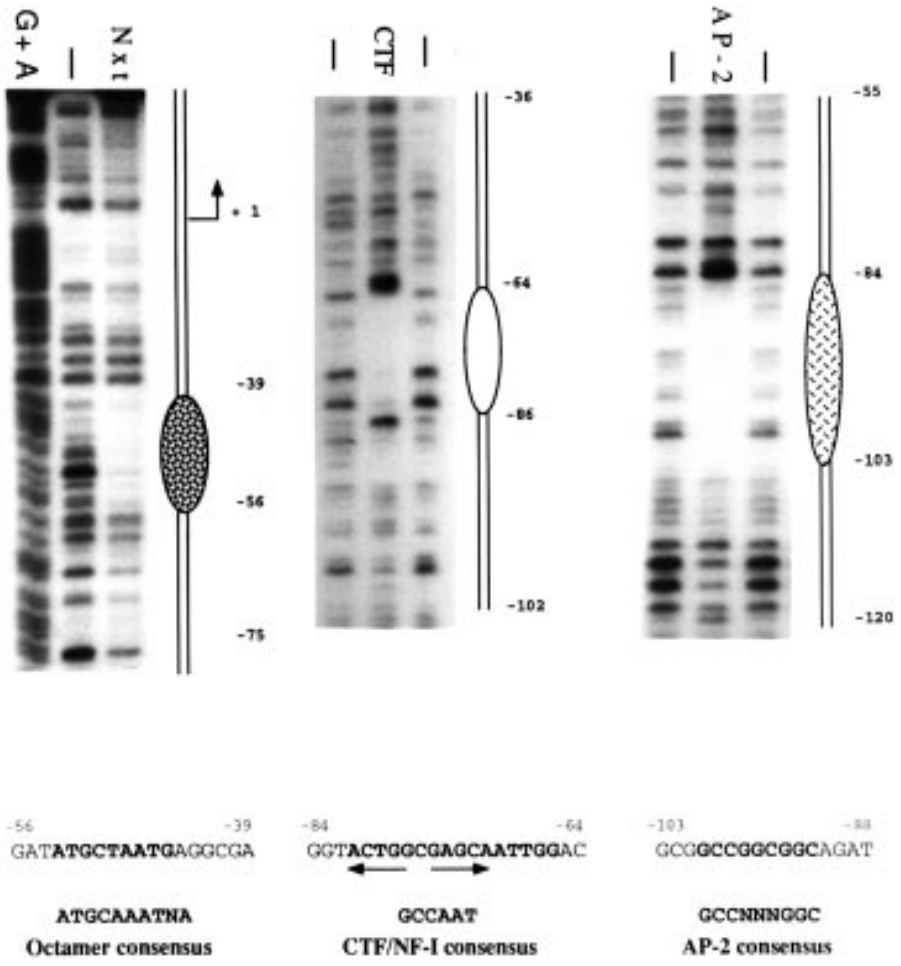

Figure 6. DNase I footprinting analysis of the AP-2 promoter. (**Left**) DNase I footprint of the AP-2 promoter using crude HeLa cell nuclear extract (Nxt) compared with a no protein control (–) and a G+A sequencing ladder. The sequence of the protected region and the octamer consensus are shown beneath. Data obtained with purified CTF/NF-1 (**middle**) and AP-2 (**right**) are also shown.

revealed a binding site for CTF/NF-1 (30) between nucleotides –64 and –86 (Fig. 6). In addition, three binding sites for AP-2 itself were identified. The most distal of these sites, at –336 (GCCCCAGGC), was documented in a previous study of the AP-2 gene (24). Our analysis also located two additional AP-2 sites, more proximal to the start site, at –165 (GTTCGCGGC) and –95 (GCCGGCGGC) (Fig. 6, right and data not shown). However, deletion of these three AP-2 sites did not appear to be critical for *in vitro* transcription of the AP-2 promoter (Fig. 4; compare lanes 4 through 6).

The activity of the AP-2 promoter relies on octamer and initiator elements

Next, a series of specific point mutations was constructed that disrupted discrete *cis*-acting elements in the context of the intact promoter. In light of the DNA binding data, the marked loss of activity seen *in vitro* with the –49 template (Figs 4 and 5) might result from mutation of either the CTF/NF-1 site or the octamer motif. In the –49 template, the CTF/NF-1 site is completely removed and the octamer sequence GATATGCTAATGA is replaced with GTCGACTTAATGA by the juxtaposition of upstream vector sequences. Therefore, to determine the importance of the octamer and CTF/NF-1 sites in the normal context of surrounding AP-2 sequences, more specific substitution mutations of these sites were engineered into the –151 template. These constructs were then used for *in vitro* transcription assays and the products analyzed by S1 nuclease mapping. As shown in Figure 8, the alteration of the CTF/NF-1 site showed no decrease in the level of transcription. In contrast, the disruption of the octamer site at –49 severely diminished transcription of the normally active –151 promoter template *in vitro* (Fig. 8, Δoct), demonstrating that the oct sequence is a critical component of the AP-2 promoter.

Because there is no apparent TATA-box motif located upstream of the start site it was possible that the AP-2 promoter relied on an initiator element for efficient transcription. Therefore, linker scanning analysis was used to examine the requirement for sequences around the start site. Substitution mutations were engineered into the –151/+37 promoter background. These templates were then used for *in vitro* transcription and the products analyzed by primer extension (Fig. 8). Alteration of sequences either between –9/–5 or downstream of +15 produced only a slight affect on initiation (data not shown). In contrast, alteration of the sequences between +3 and +8 resulted in a greatly reduced level of initiation (Fig. 8; LS2). An almost complete loss of promoter activity is observed with the more extensive substitution of sequences between –2 and +8 (LS4). These data

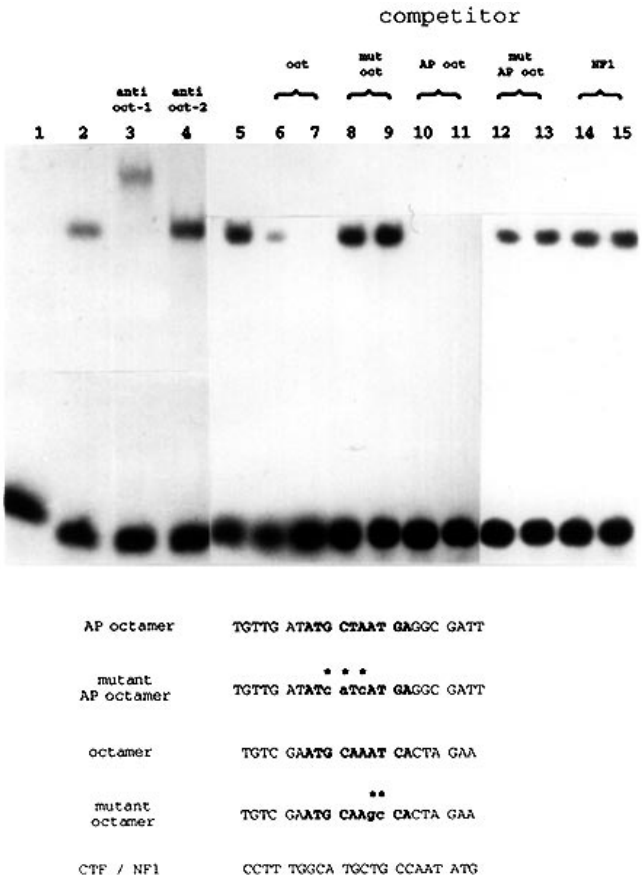

Figure 7. Identification of the octamer site in the AP-2 promoter by EMSA. The sequence of the labelled probe (AP octamer) and various competitor oligonucleotides are given, showing octamer homology (bold type) and substitutions (lower case and asterisk). The probe alone is shown in lane 1. All other lanes contain HeLa nuclear extract either alone (lanes 2 and 5), or with; anti-oct1 antiserum (lane 3), anti-oct2 antiserum (lane 4), 1 and 10 ng octamer competitor (lanes 6 and 7), 1 and 10 ng mutant octamer (lanes 8 and 9), 1 and 10 ng AP octamer (lanes 10 and 11), 1 and 10 ng mutant AP octamer (lanes 12 and 13) and 1 and 10 ng CTF/NF1 (lanes 14 and 15).

demonstrate that the sequences around the initiation site are a critical component of the AP-2 promoter.

In vivo activity of the AP-2 promoter

To confirm and extend the results obtained with *in vitro* transcription, the AP-2 promoter constructs were transfected into HeLa cells. Varying lengths of the AP-2 promoter extending to +37 were fused to the CAT gene in the pBLCAT3Δ background. Figure 9 shows that when the AP-2 promoter spanning from –151/+37 is inserted into this vector there is an ~5-fold increase in the level of CAT activity relative to pBLCAT3Δ alone. Analysis of the reporter gene transcripts from the transfected cells indicated that they utilized the correct initiation sites (data not shown). The further addition of upstream sequences to the –151 template, up to –5 kb, did not cause any significant increase in the level of CAT expression. Similarly, deletion of sequences to –82 does not affect relative expression levels (data not shown). Therefore, the AP-2 sites are not required for basal activity, in agreement with the data obtained in cell free extracts. In contrast, mutation of the octamer sequence abolished expression of the CAT gene (Fig. 9). Therefore, the upstream sequences that appear to be critical for *in vitro* transcription of the AP-2 gene also appear to be the most important in the HeLa co-transfection assay.

Comparison of AP-2 promoter sequences between species

We compared the sequence of the AP-2 promoter from different vertebrate species to determine which *cis*-acting elements were conserved. The human AP-2 coding region was used to isolate a mouse genomic cosmid clone. Subsequently, sequences conserved between these two species were used to design primers to isolate the chicken genomic sequence. A comparison between the human, mouse and chicken data is shown in Figure 10.

Both the human and mouse 5′ untranslated regions contain an IR3-like repeat, but this element is absent from the chicken. Upstream of this repetitive element there is a high degree of sequence identity between all three species, especially surrounding the initiation site and octamer element. Just upstream from the CTF/NF-1 site the homology between human and mouse is still very high; the proximal AP-2 site is identical in human and mouse, but is less conserved in chicken. Further upstream from this site, the degree of sequence identity between mouse and human is greatly reduced (data not shown). These comparative data indicate that the *cis*-acting DNA sequences critical for transcription are highly conserved between these three species. Therefore, this comparison strongly supports the role of the octamer and initiator elements as fundamental control sequences for the AP-2 promoter.

DISCUSSION

In this report we have isolated and compared the AP-2 promoter from human, mouse and chicken species. In addition, the critical elements for AP-2 expression in HeLa cells have been identified and characterized in cell free extracts or in the context of intact cells. There is a high degree of conservation between the promoter sequences, especially in the vicinity of the start site. Both human and mouse promoters contain a CT-rich IR3-like repetitive element (6,22,24) between the transcriptional start site and the translational initiation site. The conservation of the large IR3-like element between these two species suggests that it was inserted into the transcribed region of AP-2 before the divergence of these organisms and raises the possibility that it may perform some function for this transcription unit. However, the IR3-like repeat is absent from the chicken AP-2 gene. Furthermore, this repeat is not necessary for AP-2 promoter function as it can be deleted from the human gene without affecting initiation *in vitro* and we find that no transcripts originate from within this element either *in vivo* or *in vitro*. Alternatively, this repetitive element, which will be present in the 5′ untranslated region of the AP-2 RNA, could be important for regulating message stability or translational efficiency.

Upstream of the IR3-like repeat the promoters display a high degree of sequence identity, especially around the start site and octamer motif. None of the promoters contain a recognisable TATA box sequence upstream of the initiation site. Furthermore, AP-2 does not resemble TATA-less GC-rich promoters that use a number of dispersed start sites (25). Instead, AP-2 is similar to promoters that do not contain an obvious TATA-box or Sp1 binding sites, yet initiate at only a few discrete sites (25–30). These promoters, typified by the terminal deoxynucleoside transferase gene, contain a *cis*-acting initiator element (Inr)

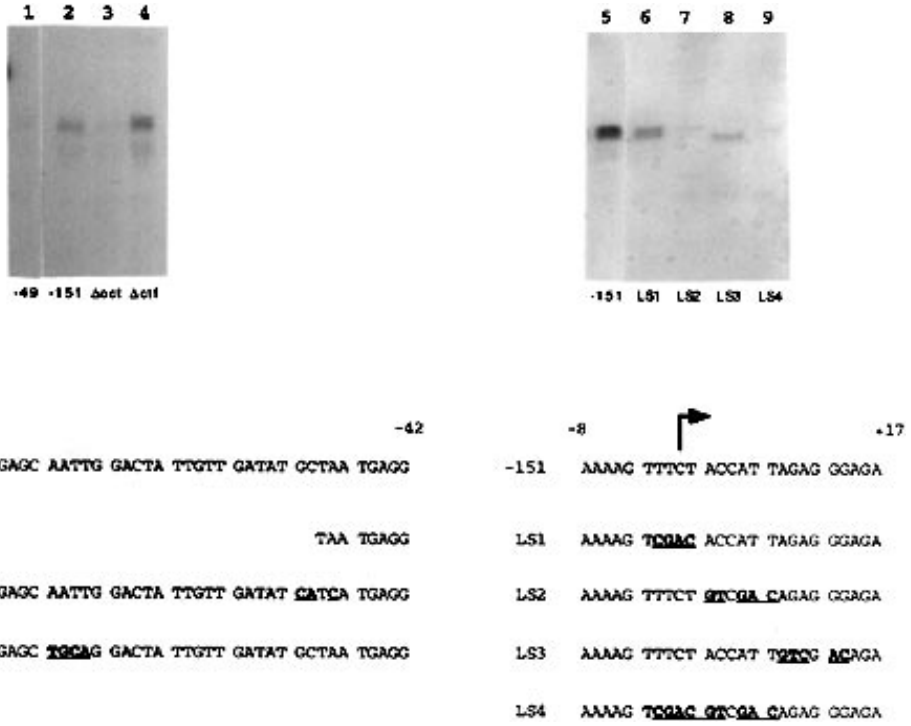

Figure 8. Transcriptional activity of linker scanner mutant templates in nuclear extracts. (**Left**) Results from a coupled *in vitro* transcription: S1 nuclease assay using mutant AP-2 promoter templates targeting the octamer and CTF/NF-1 binding sites. Below, the AP-2 promoter sequence is given in comparison with the mutants (substitutions shown by underlined bold lettering). (**Right**) Similar analysis, using primer extension assays, of mutations around the start site.

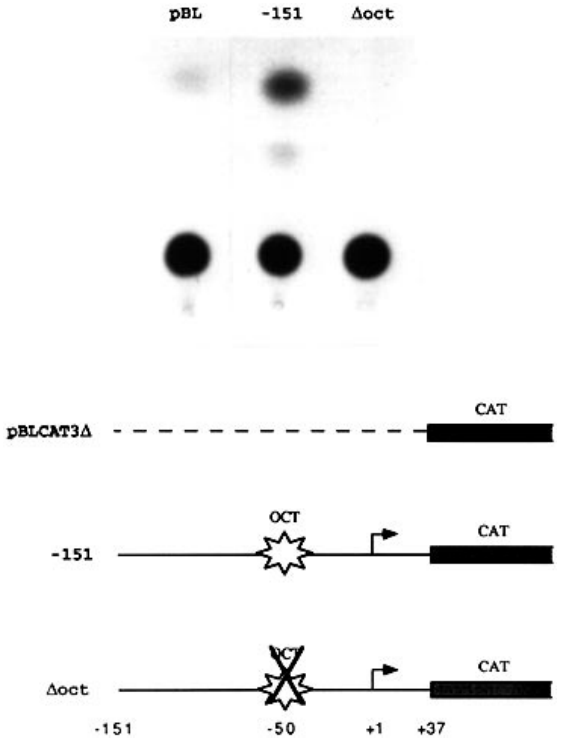

Figure 9. *In vivo* analysis of the AP-2 promoter. (**Below**) Schematic representation of the templates. In the pBLCAT 3Δ template, upstream vector sequences (dashed line) are adjacent to the CAT gene, instead of AP-2 promoter sequences (solid line). (**Top**) representative CAT assay results. The Δoct mutation is identical to that shown in Figure 8.

centered around the start site which is important for both the accuracy and level of transcription initiation (25,30). Similarly, mutation of sequences encompassing the AP-2 start site leads to a loss of activity, indicating that this sequence is an integral component of the promoter and not simply a passive recipient of upstream information. Indeed, the region between +1 and +10, CTACCATTAG, has limited homology to the consensus initiator sequence PyPyPyPyCANTPyPy which acts as a binding site for the TFIID complex (25,31–33). The AP-2 start site differs from the classical Inr element in that AP-2 transcription does not begin at the C of the CANT consensus but at a C residue 4 nt upstream. However, the potential Inr in the AP-2 promoter is consistent with previous experiments in which initiator elements can direct polymerase to commence transcription at adjacent nucleotides (28,34). In the future, a detailed analysis of this region of the AP-2 promoter in a heterologous context will help to define its potential function as an initiator element.

The AP-2 initiator region alone is not sufficient to direct significant transcription in isolation. The most critical upstream element for basal promoter activity is the octamer sequence at –49 (23). Deletion or mutation of this *cis*-acting element profoundly reduces promoter function in a cellular or cell-free milieu. Supershift experiments show that this site is exclusively bound by the ubiquitous transcription factor oct-1 in HeLa cell extracts. In contrast, in PA1 and NT2 teratocarcinoma cell extracts, several additional complexes are observed that are not supershifted by an oct-1 antisera (T. Williams, unpublished observations). This raises the possibility that part of the tissue-specificity of AP-2 transcription in development is accomplished by the variety of octamer binding proteins that have different cell-type distributions (35). The octamer site present in the immunoglobulin enhancer

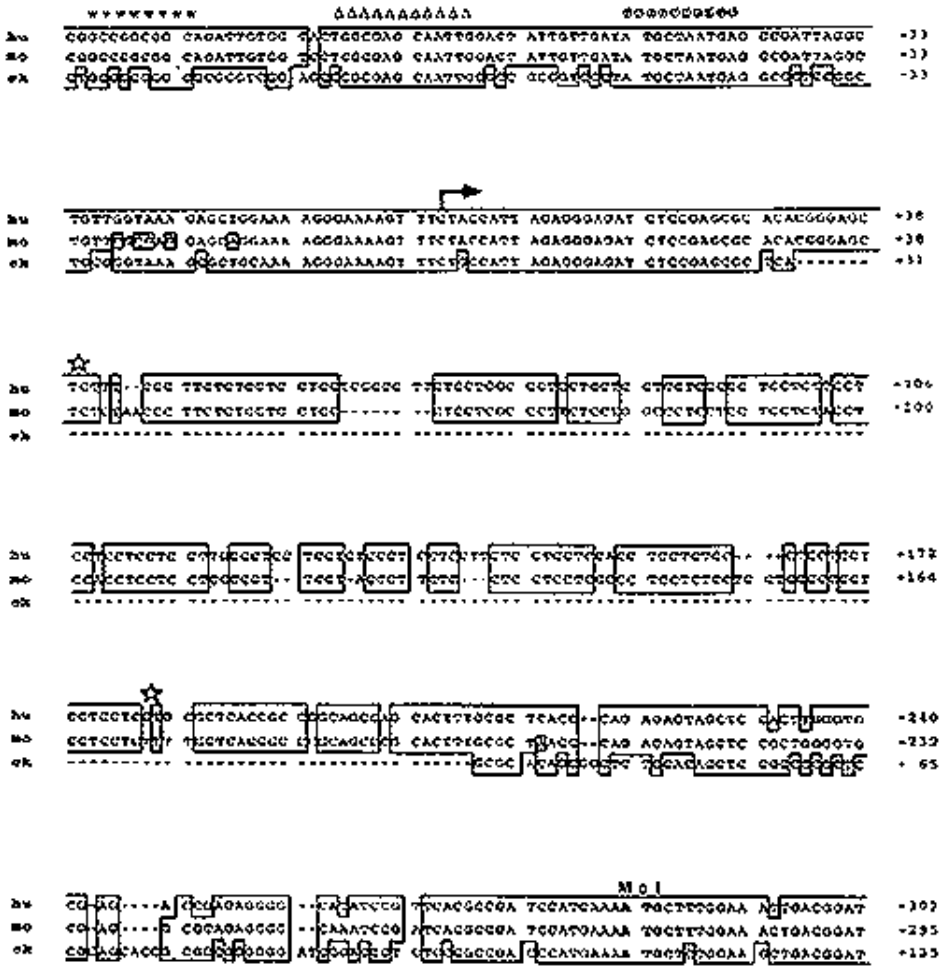

Figure 10. Comparison between human, mouse and chicken AP-2 promoter sequences. Identical nucleotides are boxed. Nucleotide numbers are based on the transcription initiation site at +1 (arrow). The translation initiation codon (Met), octamer (øøøø), CTF/NF-1 (ΔΔΔΔ), and AP-2 (****) sites are indicated. The IR3-like repeat boundaries are marked by open stars.

that imparts cell-type specific expression by interaction with the oct-2 protein present in B cells would be the classic precedent for such a model (23).

The AP-2 promoter also contains several binding sites for the transcription factor AP-2 itself, only the most distal of which was identified in a previous study of the genomic structure of AP-2 (24). The current analysis of deletion mutants shows that these sites are not critical for basal level expression, either in nuclear extracts or when transfected into tissue culture cells. However, exogenous AP-2 supplied by transfection can activate reporter constructs containing the two distal AP-2 binding sites (24) or the AP-2 promoter itself (T. Williams, unpublished). An AP-2 promoter template with the proximal binding site is activated 4-fold greater than one without, and the presence of all three AP-2 sites results in a further slight increase in AP-2 responsiveness (T. Williams, unpublished). Therefore, the presence of these binding sites suggests that the AP-2 gene could be autoregulatory. One implication of these findings is that, if inappropriately activated, AP-2 may function to keep itself switched on. Alternatively, agents that inhibit the AP-2 protein may be able to reduce the expression of the AP-2 gene. The AP-2 sites present in the promoter might also act as a mechanism that enables the AP-2 gene to be regulated by other AP-2 related proteins (36).

The activity of the endogenous AP-2 promoter in various cell lines has also been addressed. In cell lines expressing AP-2, the RNA originates at the same initiation sites used in HeLa cells. This contrasts with the previous mapping of AP-2 start sites by RNase protection and RACE PCR, where the major start sites were postulated to lie within the IR3-like element (24). However, the S1 mapping of the start sites in the present study provides much greater resolution than this earlier work. Furthermore, we supply direct evidence that the region encompassing the initiation sites we have characterized functions as a promoter element, both in cell free extracts and in tissue culture cells.

Because of a potential link between AP-2 and breast carcinoma, promoter activity was examined in several breast cancer cell lines (16). These studies indicated that AP-2 transcripts are present in all the breast cancer cell lines which have elevated levels of c-erbB2 (MDA MB 453, BT 483, SKBR3 and ZR75-1). Of the five breast cancer lines that are not c-erbB2 overexpressors, three also contain AP-2 transcripts, while two others do not, suggesting that the presence of AP-2 mRNA is not sufficient to cause

c-erbB2 overexpression. Promoter activity in the T cell line Jurkat was also examined because of the potential regulation of HTLV-I by AP-2 (37,38). No AP-2 signal derived from the endogenous promoter was detected. Conflicting reports exist concerning the presence of AP-2 protein in various T cell lines (37,38) and further analysis of T cells will be necessary to resolve these discrepancies and determine if AP-2 is involved in HTLV-I expression.

In the present study we have not attempted to address the sequences responsible for the retinoic acid-inducibility or tissue-specific activity of the AP-2 promoter. However, HepG2 cells, which do not express endogenous AP-2, can utilize the AP-2 promoter when supplied by transient co-transfection (data not shown). Examination of the endogenous AP-2 promoter in several cell lines indicates that it is hypomethylated in cell lines that utilize the promoter, such as HeLa and ZR75-1 cells. In contrast, cell lines that do not use the endogenous promoter, such as HepG2 and Jurkat, are hypermethylated as determined by restriction enzyme digestion with a panel of ^{me}CpG sensitive enzymes (J. Bernstein, unpublished observations). These data suggest that an inactive chromosomal context for the AP-2 promoter may provide one component of its tissue-specific activity. Alternatively, the AP-2 gene could contain a distal enhancer element that functions in concert with these conserved promoter elements to achieve tissue-specific expression. Therefore it may be necessary to utilize transgenic animal experiments to identify the *cis*-acting elements responsible for tissue-specific expression of AP-2. The isolation of the AP-2 promoter and the characterization of the *cis*-acting elements necessary for its basal expression provides an excellent basis for this future analysis.

ACKNOWLEDGEMENTS

We are grateful to Zhiling Jiang for the synthesis of oligonucleotides and Helen Hurst for RNA samples. We thank Rick Austin, Mark Biggin, Adrian Hayday, Helen Hurst, Bill Segraves, Timothy Nottoli and Jian Zhang for critical reading of the manuscript and encouragement throughout. We are also indebted to Steve Smale for helpful discussion. This work was funded in part by grant GM 46770 from the National Institutes of Health. T.W. is a Pew Scholar in the Biomedical Sciences. The accession numbers for the chicken, human and mouse promoter sequences are X95234, X95235 and X95236, respectively.

REFERENCES

1 Mitchell,P.J., Wang,C. and Tjian,R. (1987) *Cell*, **50**, 847–861.
2 Imagawa,M., Chiu,R. and Karin,M. (1987) *Cell*, **51**, 251–260.
3 Williams,T. and Tjian,R. (1991) *Genes Dev.*, **5**, 670–682.
4 Mitchell,P.J., Timmons,P.M., Hébert,J.M., Rigby,P.W.J. and Tjian,R. (1991) *Genes Dev.*, **5**, 105–119.
5 Byrne,C., Tainsky,M. and Fuchs,E. (1994) *Development*, **120**, 2369–2383.
6 Meier,P., Koedood,M., Philipp,J., Fontana,A. and Mitchell,P.J. (1995) *Dev. Biol.*, **169**, 1–14.
7 Williams,T., Admon,A., Lüscher,B. and Tjian,R. (1988) *Genes Dev.*, **2**, 1557–1569.
8 Lüscher,B., Mitchell,P.J., Williams,T. and Tjian,R. (1989) *Genes Dev.* **3**, 1507–1517.
9 Philipp,J., Mitchell,P.J., Malipiero,U. and Fontana,A. (1994) *Dev. Biol.*, **165**, 602–614.
10 Zhang,J., Hagopian-Donaldson,S., Serbedzija,G., Elsemore,J., Plehn-Dujowich,D., McMahon,A.P., Flavell,R.A. and Williams,T. *Nature*, **381**, 238–241; Schorle, H., Meier, P., Buchert, M., Jaenisch, R. and Mithcell, P.J. (1996) *Nature*, **381**, 235–238..
11 Kannan,P., Buettner,R., Chiao,P.J., Yim,S.O., Sarkiss,M. and Tainsky,M.A. (1994) *Genes Dev.*, **8**, 1258–1269.
12 Muller,W.J., Sinn,E., Pattengale,P.K., Wallace,R. and Leder,P. (1988) *Cell*, **54**, 105–115.
13 Lemoine,N.R. and Wright,N.A. (1993) In *The Molecular Pathology of Cancer.* Cancer Surveys, Vol. **16**, CSHL Press.
14 Guy,C.T., Webster,M.A., Schaller,M., Parsons ,T.J., Cardiff,R.D. and Muller,W.J. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 10578–10582.
15 Hollywood,D.P. and Hurst,H.C. (1993) *EMBO J.*, **12**, 2369–2375.
16 Bosher,J.M., Williams,T. and Hurst,H.C. (1995) *Proc. Natl. Acad. Sci. USA*, **92**, 744–747.
17 Luckow,B. and Schütz,G. (1987) *Nucleic Acids Res.*, **15**, 5490.
18 Dignam,J.D., Lebovitz,R.M. and Roeder,R.G. (1983) *Nucleic Acids Res.*, **11**, 1475–1489.
19 Wildeman,A.G., Sassone-Corsi,P., Grundström,T., Zenke,M. and Chambon,P. (1984) *EMBO J.*, **3**, 3129–3133.
20 Ausubel,F.M., Brent,R., Kingston,R.E., Moore,D.D., Seidman,J.G., Smith,J.A. and Struhl,K. (1989) *Current Protocols in Molecular Biology.* Greene Publishing Associates and Wiley-Interscience.
21 Hurst,H.C. (1991) In Collins,M. (eds) *Methods in Molecular Biology,* Humana Presss, NY, pp. 141–151.
22 Heller,M., Flemington,E., Kieff,E. and Deininger,P. (1985) *Mol. Cell. Biol.*, **5**, 457–465.
23 Herr,W. (1992) Oct-1 and Oct-2, In McKnight,S.L. and Yamamoto,K.R. (eds) *Transcriptional Regulation*, **2**. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 1103–1135.
24 Bauer,R., Imhof,A., Pscherer,A., Kopp,H., Moser,M., Seegers,S., Kerscher,M., Tainsky,M.A., Hofstaedter,F. and Buettner,R. (1994) *Nucleic Acids Res.*, **22**, 1413–1420.
25 Smale,S.T. (1994) In Conaway,R.C. and Conaway,J.W. (eds) *Transcription, Mechanisms and Regulation.* Raven Press Ltd., New York, pp. 63–81.
26 Chen,H., Vinnakota,R. and Flint,S.J. (1994) *Mol. Cell. Biol.*, **14**, 676–685.
27 Murphy,J.T., Skuzeski,J.T., Lund,E., Steinberg,T.H., Burgess,R.R. and Dahlberg,J.E. (1987) *J. Biol. Chem.*, **262**, 1795–1803.
28 Pal,S.K., Zinkel,S.S., Kiessling,A.A. and Cooper,G.M. (1991) *Mol. Cell. Biol.*, **11**, 5190–5196.
29 Propst,F., Rosenberg,M.P., Iyer,A., Kaul,K., and Vande Woude,G.F. (1987) *Mol. Cell. Biol.*, **7**, 1629–1637.
30 Smale,S.T. and Baltimore,D. (1989) *Cell*, **57**, 103–113.
31 Kaufmann,J. and Smale,S.T. (1994) *Genes Dev.*, **8**, 821–829.
32 Purnell,B.A., Emanuel,P.A. and Gilmour,D.S. (1994) *Genes Dev.*, **8**, 830–842.
33 Verrijzer,C.P., Yokomori,K., Chen,J.-L. and Tjian,R. (1994) *Science*, **264**, 933–941.
34 Zenzie-Gregory,B., O'Shea-Greenfield,A. and Smale,S.T. (1992) *J. Biol. Chem.*, **267**, 2823–2830.
35 Schöler,H.R. (1991) *Trends Genet.*, **7**, 323–329.
36 Moser,M., Imhof,A., Pscherer,A., Bauer,R., Amselgruber,W., Sinowatz,F., Hofstädter,F., Schüle,R. and Buettner,R. (1995) *Development*, **121**, 2779–2788; Chizaud, C., Oulad-Abdelghani, M., Bouillet, P., Décimo, D., Chambon, P., and Dollé, P. (1996) *Mech. Dev.*, **54**, 83–94.
37 Nyborg,J.K. and Dynan,W.S. (1990) *J. Biol. Chem.* **265**, 8230–8236.
38 Muchardt,C., Seeler,J.-S., Nirula,A., Gong,S. and Gaynor,R. (1992) *EMBO J.* **11**, 2573–2581.