Nucleic Acids Research, 1994, Vol. 22, No. 8 1413–1420

# The genomic structure of the human AP-2 transcription factor

Reinhard Bauer, Axel Imhof, Armin Pscherer, Heidrun Kopp, Markus Moser, Silvia Seegers, Monika Kerscher, Michael A.Tainsky[1], Ferdinand Hofstaedter and Reinhard Buettner*
Department of Pathology, University of Regensburg Medical School, D-93042 Regensburg, Germany and [1]The University of Texas M.D.Anderson Cancer Center, 1515 Holcombe Boulevard, Houston, TX 77030, USA

Received January 13, 1994; Revised and Accepted March 15, 1994 EMBL accession number X77343

## ABSTRACT

**The transcription factor AP-2 is encoded by a gene located on chromosome 6 near the HLA locus. Here we describe the genomic organization of the AP-2 gene including an initial characterization of the promoter. We have mapped two mRNA initiation sites, the entire exon – intron structure and located two polyadenylation sites. The mature AP-2 mRNA is spliced from 7 exons distributed over a region of 18 kb genomic DNA. A recently cloned inhibitory AP-2 protein is generated by alternative usage of a C-terminal exon. The proline-rich transactivation motif is encoded by a single exon within the N-terminal region in contrast to the complex DNA binding and dimerization motif which involves amino acid residues located on four different exons. The sites of mRNA initiation are located 220 and 271 bases upstream from the ATG translation start site. Although the promoter contains no canonical sequence motifs for basal transcription factors, such as TATA-, CCAAT- or SP-1 boxes, it mediates cell-type-specific expression of a CAT reporter gene in PA-1 human teratocarcinoma cells and is inactive in murine F9 teratocarcinoma cells. We demonstrate that the promoter of the AP-2 gene is subject to positive autoregulation by its own gene product. A consensus AP-2 binding site is located at position – 622 with respect to the ATG. This site binds specifically to bacterially expressed AP-2 as well as to multiple proteins, including AP-2, present in PA-1 and HeLa cell nuclear extracts. A partial AP-2 promoter fragment including the AP-2 consensus binding site is approximately 5-fold transactivated by cotransfection of an AP-2 expression plasmid.**

## INTRODUCTION

The transcription factor AP-2 was first isolated from HeLa cells by affinity chromatography using specific binding sites within the SV40 and the human metallothionein IIa promoter (1). Screening of a HeLa cDNA library with oligonucleotide probes predicted from partial peptide sequences led to the isolation of the AP-2 cDNA (2). Subsequently the gene was mapped to a single site on chromosome 6 near the HLA locus (3). A number of functional sites with a consensus palindromic core sequence, 5'-GCCNNNGGC-3'(4), that differ considerably in their affinity to purified AP-2 protein (5) have been identified in enhancer regions of viral and cellular genes. Molecular fusion of AP-2 and Gal-4 protein domains revealed an N-terminal proline and glutamine-rich transactivation motif (6). The DNA binding domain is located within the C-terminal half of the AP-2 protein and is composed of a hypothetical helix-span-helix motif which is both necessary and sufficient for homodimer formation (4).

AP-2 activity is subject to regulation through a number of different signal transduction pathways. Phorbol esters and signals that enhance cyclic AMP levels induce AP-2 activity independent of protein synthesis, whereas retinoic acid treatment of teratocarcinoma cell lines results in a transient induction of AP-2 mRNA levels on a transcriptional level (7,8). We have recently shown that an alternative AP-2 protein, AP-2B, that differs in its C-terminus from the originally cloned protein, AP-2A, is a negative regulator of transcriptional activation (8). Therefore, the endogenous AP-2 activity within the cell nucleus results from a complex integration of different signal transduction pathways that target transcriptional, posttranscriptional and posttranslational AP-2 regulation.

AP-2 is a key transcriptional regulator involved in mediating programmed gene expression both during embryonic morphogenesis and adult cell differentiation. A restricted spatial and temporal expression pattern during murine embryogenesis has been determined by in situ hybridization. In particular, regulated AP-2 expression was observed in neural crest-derived cell lineages and in facial as well as limb bud mesenchyme at times when they are known to be developmentally retinoid sensitive (9). Furthermore, analyses of AP-2 expression in embryonic and adult skin have implicated a causal role for AP-2 in regulation of keratin gene expression during skin differentiation (10) and the keratin transcription factor KTF-1 has been shown to be similar or identical to AP-2 (11). A number of genes expressed at high levels in epidermal tissues, including keratins and E-cadherin (12), contain cis-regulatory AP-2 sites within their

*To whom correspondence should be addressed

promoters. More recently we have identified high levels of AP-2 expression in several prostate cell lines (13).

Both mRNAs AP-2A and AP-2B are encoded by a single gene located on chromosome 6p22.3-pter. Here we report the molecular cloning of the AP-2 gene and describe its organization including the complete exon−intron structure. Since exons provide frequently functional modules for complex polypeptides, we analyse relationships between the exon structure and hypothetical protein domains determined previously by Williams and Tjian (4,6). Further, we have cloned and sequenced 2 kb of the AP-2 promoter. This sequence confers expression to a CAT reporter gene in the human teratocarcinoma cell line PA-1 and is inactive upon transfection into the AP-2 negative murine teratocarcinoma cell line F9. Analysis of the promoter will lead to a molecular characterization of transcriptional control of the AP-2 gene and unravel mechanisms of programmed gene expression during morphogenesis and cell differentiation. As a first step we have studied a consensus AP-2 binding site within the AP-2 promoter. Our data provide evidence that the AP-2 gene is subject to autoregulation by its own gene product.

## MATERIALS AND METHODS

### Isolation and orientation of recombinant genomic AP-2 clones

As described previously (8), we have isolated four independent overlapping recombinant phages spanning the entire AP-2 locus from a human placenta genomic DNA library in the phage lambda FixII (Stratagene, Heidelberg, Germany) using an N-terminal cDNA fragment as a probe. The complete inserts were purified from these phages as NotI fragments, subjected to incomplete digestion with the restriction enzyme XbaI, then separated by agarose gel electrophoresis and transferred onto Nylon membranes. Since the polylinkers located at the left and right side of the genomic DNA insert contain T3 and T7 primer sequences, respectively, we hybridized the Southern blots with radiolabeled T3 and T7 primers. Starting from the smallest hybridizing fragment we were able to measure the distances between the XbaI restriction site and thereby the relative location of the XbaI fragments within the genomic DNA fragments.

### DNA sequence analysis

DNA sequencing was performed in part by dideoxy chain termination according to the standard Sanger protocol (14) and in part by cycle sequencing (15) using an automatic DNA sequencer (Applied Biosystems, Boulder, CO, USA). Double stranded plasmid DNA isolated by alkaline lysis of *E.coli* minicultures (16) served as template and 18mer oligonucleotides complementary to insert DNA were used as primers along with T3 and T7 primers. Computer analysis of nucleotide sequences was performed using the University of Wisconsin Genetics Computer Group software (17).

### Isolation of mRNA and RNase protection analysis

Total cellular RNA was isolated from PA-1 6928 cells by guanidium thiocyanate lysis and centrifugation through cesium chloride (18). 1 mg of total RNA was selected twice by oligo-d(T) chromatography using commercially available spin columns (Pharmacia, Freiburg, Germany) resulting in 20 $\mu$g of poly(A)$^+$-RNA.

RNAase protection assays were performed as described in detail (16). Briefly, from a genomic DNA template spanning the XbaI site at base −435 and the ApaI site at base +396 with respect to the ATG translation start codon (fig. 2b) single stranded sense and antisense RNA probes with a specific activity of $1-2\times10^9$ cpm/$\mu$g were transcribed using T3 and T7 RNA polymerases. 1 $\mu$g poly(A)$^+$-RNA together with $5\times10^5$ cpm probe was redissolved in 30 $\mu$l of 40 mM PIPES (pH 6.4), 1mM EDTA (pH 8.0), 0.4 M NaCl and 80 % formamide for 10 minutes at 85 °C. After hybridization overnight at 43 °C the mixture was digested in 300 $\mu$l of 2 $\mu$g/ml RNAase T1 and 40 $\mu$g/ml RNAase A for 60 minutes at 30 °C. Finally, protein was removed from the mix by a short treatment with proteinase K and two phenol/chloroform extractions and the nucleic acids were separated on a 5 % polyacrylamide/urea gel next to a sequencing reaction.

### RACE-PCR

The AP-2 mRNA 5'-end was directly cloned using a commercially available kit for rapid amplification of cDNA ends, RACE-PCR (Clontech, Palo Alto, CA, USA). The procedure has been described previously (19). Briefly, the oligonucleotide 5' TCG CAG TCC TCG TAC TTG ATA TTA TC 3' (primer 1) matching bases +44 to +19 with respect to the ATG served as a primer to synthesize first strand cDNA from 1 $\mu$g of poly(A)$^+$-RNA prepared from PA-1 clone 6928 cells. The quality of cDNA synthesis was verified using a nested primer pair matching residues −6 to −32 (5' TGG ATC GGC GTG AAC GGA TAT GCC CCT 3'; primer 2) and −71 to −54 (5' CCC AGA GAG TAG CTC CAC 3'; primer 3) with respect to the ATG for PCR which resulted in a 65 bp amplification product. A second control PCR using primer 1 and primer 3 resulted in the expected 115 bp product. After alkaline hydrolysis of the RNA and purification of the cDNA a 3'-modified anchor primer (5' P-CAC GAA TTC ACT ATC GAT TCT GGA ACC TTC AGA GG-$NH_3$ 3') was ligated to the cDNA by T4 RNA ligase. Finally, the resulting cDNA products were amplified by PCR using an unmodified anchor sequence and the nested primer 2. The products were analysed on an 1.8 % agarose gel and subcloned into the sequencing plasmid pCR-Script (Stratagene).

### Plasmids, transfections and CAT assays

A promoter fragment spanning nucleotides −6 to −2013 with respect to the ATG translation start codon was amplified by PCR and subcloned into the SalI/BamHI sites of the plasmid pBLCAT3 (20). A second 0.7 kb genomic fragment spanning the XbaI sites at positions −1145 and −435 was ligated into the XbaI site of the plasmid pBLCAT2 which contains the CAT gene under the control of a minimal TK promoter (20). 5 $\mu$g of CAT plasmids were transfected into PA-1 clone 6928 and clone 9117 cells as well as into F9 cells by calcium phosphate precipitation (21). For coexpression of AP-2A protein the full length AP-2A cDNA was ligated into the EcoRI site of the plasmid pCMX-pL1 which was derived from the vector pCMX by modifying the multiple cloning site (22). 2 $\mu$g of the plasmid pCMX-$AP_2A$ were cotransfected with the pBLCAT2 construct. The CMV promoter was chosen to mediate AP-2 expression because it is active both in PA-1 and F9 cells and its activity is independent from endogenous AP-2 levels. The transfection procedure, tissue culture conditions and characteristics of the PA-1 cell lines clone 6928 and clone 9117 have been described in detail previously (23,24). Equal amounts of protein were used for CAT assays (25) and the transfection efficiency was controlled by cotransfection of an LTR-lacZ plasmid. To perform the assays in a comparable range of enzyme activity the amount of protein

lysate used for the $SV_2CAT$ and the RSV-CAT assays was reduced to one third.

### Gel mobility-shift assays

A genomic DNA fragment spanning nucleotides −668 to −562 including the AP-2 consensus site at position −621 (fig. 3) was amplified by PCR, subcloned into the vector pCR-Script (Stratagene) and excised for gel shifts. $2\times10^4$ cpm of phosphate labeled DNA was used per shift assay. Competition experiments were performed with 50-fold excess of unlabeled homologous DNA fragment or a synthetic consensus AP-2 site from the human metallothionein IIA promoter (5′ AGG AACT GAC CGC CCG CGG CCC GTG TGC AGA G 3′; ref.5). Gel shift conditions and protocols for purification of nuclear extracts and bacterially expressed AP-2 protein have been described (8).

### DNAase I protection analysis

The genomic DNA fragment from −668 to −562 that was used for gel mobility-shift assays was excised from the plasmid pCR-Script using the SacII and EcoRI restriction sites and labeled at the EcoRI site by a Klenow fill-in reaction with $^{32}$P-dATP and cold dTTP. Finally, the probe was recovered from a 5 % polyacrylamide gel. $1.5\times10^5$ cpm were subjected to DNAase I digestion following precisely a standard protocol (26). Protection was performed with 1 and 5 μg purified AP-2 protein, respectively.

## RESULTS

Four overlapping genomic clones spanning 35 kb of the AP-2 gene locus were isolated and the relative positions of XbaI restriction sites were determined (see Materials and Methods). From these clones a continuous stretch of 20154 bases genomic DNA covering the entire coding region including approximately 2 kb of the promoter was sequenced on both strands and submitted to the EMBL data library (accession no. X77343). A graph of the genomic locus indicating the relative positions of all exons is shown in figure 1. The coding nucleotides determined from the genomic DNA match perfectly with the cDNA sequences that have been obtained from two independent sources previously (HeLa cells and PA-1 cells; ref. 2,8).

The mature AP-2A mRNA is spliced from 7 exons and the AP-2B mRNA from 5 exons. The sequences of the exon−intron boundaries are shown in table 1. Examination of the sequence of the fifth exon and intron revealed that the alternative AP-2B mRNA is generated by a failure to splice after the fifth exon. Thereby, a continuous transcript is processed that encodes for an alternative C-terminal open reading frame. We have shown previously that the protein AP-2B by itself is incapable of binding DNA but inhibits transcriptional activation by the protein AP-2A. The fifth splice site that is ignored when AP-2B mRNA molecules are processed represents an intact splice donor site matching to the consensus sequence as well as other splice sites within the AP-2 gene. Therefore, the alternative splicing of AP-2B mRNA is an active process and not due to ambiguity of a splice site.

Next, we performed two sets of experiments to determine the initiation site of AP-2 mRNA. An XbaI/ApaI genomic DNA fragment spanning nucleic acids residues −435 to +396 was subcloned into the plasmid pBluescript and served as a template to transcribe $^{32}$P-UTP-labelled RNA probes for RNAase protection assays. T7 and T3 RNA polymerases were used to transcribe antisense and sense probes, respectively. As shown in figure 2a an RNA fragment of approximately 285 bases and a second very weak band, approximately 325 bases in size, were protected specifically when the T7 antisense probe was used. The T3 sense RNA transcript did not protect any RNA fragment.

Since we obtained two protected fragments we needed to investigate whether the 5′ untranslated mRNA was spliced from two different exons. Therefore, RACE-PCR amplifications of the 5′ end of the AP-2 mRNA end were performed (see Materials and Methods) resulting in a reproducible pattern of 2 different bands between 240 and 400 bp detected on ethidium bromide stained agarose gels (fig.2c). When the amplification products were transfered to Southern blots and probed with the 0,8 kb XbaI/ApaI genomic fragment the 240 bp fragment yielded the strongest hybridization signal and an additional slightly larger,

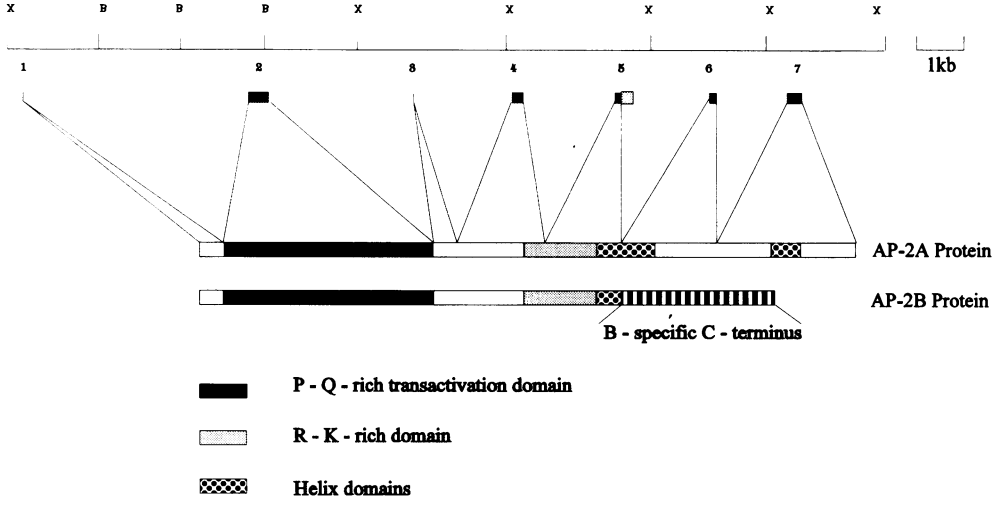


**Figure 1.** Exon−intron structure of the AP-2 gene and the relative position of all exons encoding the AP-2A and AP-2B proteins. The alternative C-terminal exon 5 of AP-2B is indicated by a shaded box. The solid bar above the AP-2 exons represents a continuous stretch of 20154 bases that has been sequenced entirely and submitted to the EMBL data library (X77343).

**Table 1.** Sequences around the exon−intron boundaries of AP-2

| Exon number | Exon size (bp) | Sequence at exon - intron junction | |
|---|---|---|---|
| | | 5'splice donor | 3'splice acceptor |
| 1 | 45* | GAC TGC GAG gta agc | gtc cct ctc ttc tag GAC CGT CAC |
| 2 | 435 | GAG GTC CCG gta aga | atc ctt ttg ttg cag CAT GTA GAA |
| 3 | 52 | AAG AAA G gta agc | cct ctg tct ccg cag GC CCC GTG |
| 4 | 232 | CTC CGG AG tga ggc | ata ttt gtg aaa cag G GCG AAG |
| 5 | 119 | GTA GAG G gta agc | gtc ccc tct tgc cag GA GAA GCT |
| 6 | 142 | GCT ACA AA gta agg | acc ctt gtc ctg cag A CAG ATA |
| 7 | 285* | CAC AGA AAG tga ggc Stop | |

Exon sequences are printed in capital letters and intron sequences in lower case letters. Nucleic acids matching the consensus splice junction are underlined. The complete coding sequences of AP-2A and AP-2B cDNAs have been published in reference 8. The sizes of the first and the last exon, marked by an asterisk, refer to the coding residues only.

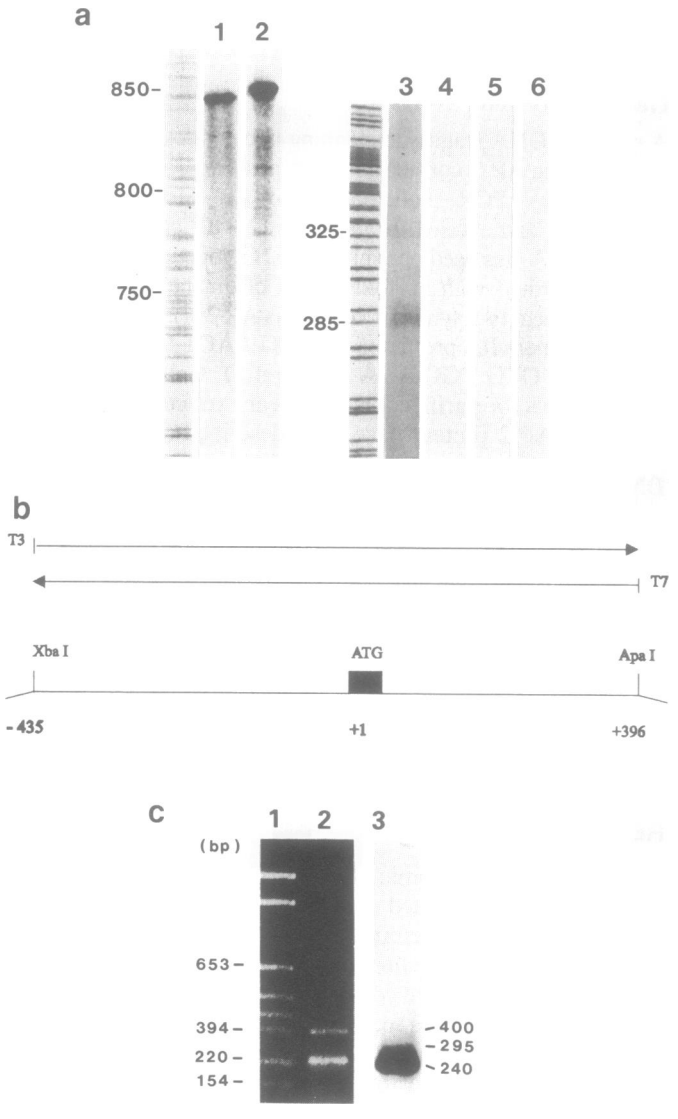


**Figure 2.** Determination of the mRNA initiation sites of the AP-2 gene. (**a**) RNAase protection assay. Antisense (T7 RNA polymerase) and sense (T3 RNA polymerase) riboprobes were hybridized to 6928 cell RNA. A 285 nucleotide fragment and a second very weak 325 nucleotide fragment were protected specifically from RNAase digestion in the T7 riboprobe. Lane 1: free probe T7; lane 2: free probe T3; lane 3: RNAase digested probe T7; lane 4: RNAase digested probe T3; lane 5: RNAase digested T7 probe hybridized to tRNA; lane 6: RNAase protected T3 probe hybridized to tRNA. (**b**) Location of the genomic fragment that served as a template to transcribe antisene (T7) and sense (T3) riboprobes for RNAase protection experiments. (**c**) Direct cloning of the N-terminal AP-2 cDNA end by RACE-PCR. First strand cDNA was synthesized from 6928 poly(A)$^+$-RNA using an antisense primer in the first coding exon (see Materials and Methods). After removal of the mRNA an anchor primer was ligated to the 3′ end of the cDNA and the resulting DNA fragments were amplified by PCR. The amplification products were visualized on a 1.8 % agarose gel (lane 2), transfered to a Southern blot and hybridized to the XbaI/ApaI genomic DNA (lane 3).

weak signal was observed. This larger signal represented a DNA fragment of very low abundancy that was not detected easily by ethidium bromide staining. In contrast the 400 bp fragment detected by ethidium bromide staining did not hybridize with the XbaI/ApaI probe. One tenth of the RACE-PCR mix was ligated into the plasmid pCR-Script by a shotgun cloning procedure resulting in a total of 9 colonies, 5 of which hybridized specifically with the 0.8 kb XbaI/ApaI genomic probe. We then analyzed all 9 clones by sequencing entirely their respective inserts. One of the inserts started at nucleic acid position −271 upstream from the AP-2 ATG translation start codon and 4 others at residue −220 (fig.3). As expected all of these five RACE-PCR clones ended with primer 2 at residue −6 (see Materials and Methods). The remaining 4 clones that did not hybridize to the 5′ genomic AP-2 fragment contained the 400 bp fragment that was amplified by the RACE-PCR (fig.2c). This fragment had no sequence similarity to the 5′ genomic AP-2 sequence except from the primer and therefore must result from nonspecific priming. The cDNA fragments that were amplified by RACE-PCR and the RNAase protected RNA fragments are in good agreement with respect to their size and relative abundance. Furthermore, all RACE-PCR clones contained a continuous stretch of 5′ genomic sequence that was identical to the sequence obtained from the genomic clones. Thus, we conclude that there are two AP-2 mRNA initiation sites at −271 and −220 and that the 5′ untranslated mRNA together with the first 45 coding nucleic acid residues are located on a single exon. Figure 3 displays the 5′ genomic sequence and the location of the two mRNA start sites.

Inspection of the genomic DNA sequence 3′ adjacent to the TGA stop codon of the AP-2A reading frame revealed five putative polyadenylation sites (AATAAA) located 416, 934, 1339, 1601 and 1616 bases downstream from the TGA, respectively. Only a single polyadenylation signal was found 1477 bases downstream from the TAA stop codon of the reading frame AP-2B. To determine which sites are used we have sequenced entirely the 3′ termini of eight different AP-2 cDNA clones that were obtained by screening an oligo d(T)-primed cDNA library from PA-1 clone 6928 cells. As reported previously (8), one encodes for the protein AP-2B and 7 others for the protein AP-2A. 2 out of 7 AP-2A cDNAs originated upstream from the first polyadenylation signal 416 bases 3′ from the stop codon and the 5 others upstream from the second polyadenylation signal 934 bases 3′ from the stop codon (data not shown). The AP-2B cDNA clone originated upstream from the polyadenylation site 1477 bases 3′ of the respective stop codon. We did not find any indication for an intervening intron located within the 3′ noncoding sequence. Therefore, we conclude that the two main

```
-650 AGGGTAAGGG AGAAGTGCTA GAAGCTGGGC CCCAGGCGTG GCGCTTGGAC CTTGAGCAGT
-590 GCCGGGCTTT GCAGAAGGCA TCTCTGGAGA AAAAAGGAAG CGGAGAAAGC AAGGAAGGGG
-530 AATGTGGCGG AATTGGGGTA AGAGAGGGGG TGTTGGGGCA GGGCTCGCGG CGGAGCTCGA
-470 GGAAGGTTTT ATCAGTTCCG CAGGCTGGAG TGGCTCTAGA AGCAAGTTCG CGGCGTAGAG
-410 AGGCCTTTAT ACCTGTGTCG GAGCGGCCGG CGCAGATGTG GTACTTGGCG AGCAATTGGA
-350 CTATTGTTGA TATGCTAATG AGGCGATTAG GCTGTTGGTA AAGAGCTGGA AAAGGGAAAA
-290 GTTTCTACCA TTAGAGGGAG ATCTCCGAGC GCACACGGGA GCTCTTTCCC TTCTCTCCTC
-230 CTCCTCGCCC TTCTCCTCGC CCTCCTCCTC CTCCTCGCCC TCCTCTTCCT CCTCCTCCTC
-170 CTTGCCCTCC TCCTCTCCCT CCTCCTTCTC CTCCTCCACC TCCTCTCCCT CCTCCTCCTC
-110 CTCCTCCTGC GCTCACCGCC GGAGCCAGCA CTTTGGCTAC CCAGAGAGTA GCTCCACCGT
 -50 TAGAGGTGCG AGACCGAGAG GGGCATATCC GTTCACGCCG ATCCATGAAA ATG
                                  initiation sites of transcription
```

**Figure 3.** Sequence of the AP-2 promoter from nucleic acid residue −650 to the ATG translation start codon at residue +1. Two initiation sites of mRNA transcription at positions −271 and −220 are marked by arrows. Consensus binding sites for AP-2 at −622, GCF at −384 and a TC-rich sequence from −240 to −100 are underlined. The complete 5′ flanking sequence starting at residue −2013 is part of the sequence submitted to the EMBL data library (X77343).

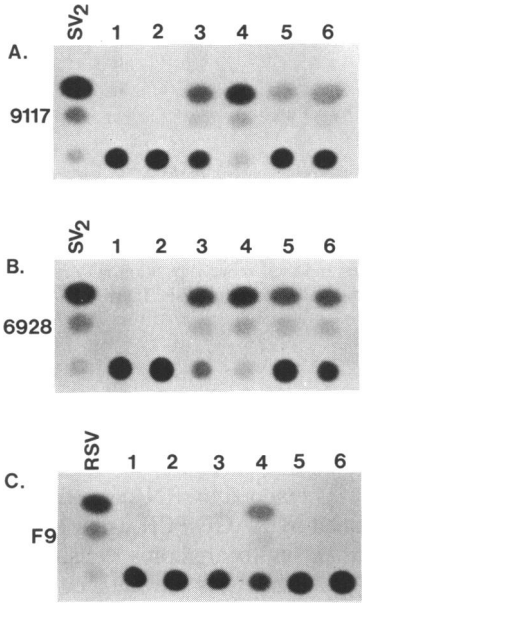


**Figure 4.** Analysis of the AP-2 promoter by transient transfections of CAT reporter plasmids. (**A**) Transfections into PA-1 clone 9117 cells, (**B**) transfections into PA-1 clone 6928 cells and (**C**) transfections into F9 cells. 5 μg of the following plasmids were transfected: (1) $pBLCAT_2$; (2) $pBLCAT_3$; (3) AP2−0.7kb; (4) AP-2−0.7kb cotransfected with $pCMX\text{-}AP_2A$; (5) AP2−2.0kb; (6) AP2−2.0kb followed by treatment with 10 μM retinoic acid for 36 hours after transfection. For a positive control 5 μg of the plasmid $pSV_2CAT$ ($SV_2$) were transfected into 9117 and 6928 cells and 5 μg RSV-CAT (RSV) into F9 cells.

AP-2A transcripts that are observed on Northern blots both in murine and human tissues (27) originate from usage of two different polyadenylation sites located 416 and 934 bases downstream from the stop codon, whereas the AP-2B mRNA uses a single polyadenylation signal located 1477 bases downstream from the AP-2B protein stop codon.

To study developmental and tissue specific cis-regulatory factors controlling AP-2 mRNA expression we have started to analyse the AP-2 promoter. A 2 kb 5′ genomic fragment spanning nucleic acid residues −6 to −2013 with respect to the ATG start codon was subcloned into the plasmid pBLCAT3 (20) and transiently transfected into two subclones that were derived from the human teratocarcinoma cell line PA-1, clones 6928 and 9117 (28). We have shown recently that ras-transformed PA-1 cells, including 6928 cells that are stably transfected with a single copy mutated N-ras oncogene, express approximately sixfold increased levels of AP-2 mRNA. In contrast, 9117 cells that do not harbour any ras gene mutations express low levels of AP-2 mRNA that can be increased transiently by treatment with retinoic acid (8, Kannan *et al.* unpublished data). Further, we transfected the AP-2 promoter-CAT construct into the AP-2 negative murine teratocarcinoma cell line F9. The results of the CAT assays show that the 2 kb genomic AP-2 fragment confers cell type-specific gene expression in PA-1 teratocarcinoma cells but is inactive in F9 cells (fig. 4, lane 5). In addition the AP-2 promoter fragment mediated approximately 5-fold increased CAT activity in 6928 cells as compared to clone 9117. However, treatment of 9117 and 6928 cells with retinoic acid for 36 hours after transfection did not alter significantly the activity of the CAT reporter (fig. 4, lane 6).

To identify potential cis-regulatory elements the 2 kb promoter fragment was analysed for homologies to known consensus binding sites of transcription factors. We did not find any known sequence motifs for basal transcription factors, such as TATA-, CCAAT- or SP-1 boxes, within reasonable distance from the mRNA start sites. Initiation occurs just upstream from a long TC-rich region between residues −240 and −100 that contains 28 times the sequence TCC and in addition 7 times the sequence TCCC or GCCC (figure 3). Further, we found consensus binding sites for the transcription factors GCF starting at position −384 (CCGGCGC) and AP-2 starting at position −622 (GCCNNNGGC).

The AP-2 site was analysed in more detail. Therefore, we cloned an XbaI−XbaI fragment from nucleic acid residue −1145 to −435 into the plasmid pBLCAT2 and transfected it into PA-1 cells. We observed moderate CAT expression in 9117 cells but 4-fold higher levels in 6928 cells (fig.4, lane 3). Cotransfection of an AP-2 expression plasmid enhanced the activity of the AP-2 promoter fragment in 9117 cells approximately 5-fold whereas a moderate but reproducible approximately twofold stimulation was observed in 6928 cells (fig.4, lane 4). Because the plasmid pBLCAT2 provides a minimal TK promoter it allows to analyse enhancer elements within gene promoters independent of their own transcription initiation site. Therefore, we transfected the pBLCAT2 construct containing the partial AP-2 promoter fragment including the AP-2 site into the AP-2 negative teratocarcinoma cell line F9. Results shown in figure 4 lanes 3 and 4 demonstrate that the pBLCAT2 construct alone is almost silent in F9 cells but is approximately 10-fold transactivated by coexpression of AP-2A protein.

To demonstrate specific protein−DNA interactions directly at the AP-2 site we performed gel mobility-shift assays using a radiolabeled genomic DNA fragment from nucleotides −668 to −562 including the AP-2 site. The results of the gel shift experiments (fig.5) clearly revealed that both bacterially expressed AP-2 and proteins present in PA-1 cell nuclear extracts bind to the AP-2 site. Using purified recombinant protein we observed four shifted bands all of which were competed by an excess of unlabeled binding site. The band shift pattern differed when nuclear extracts from PA-1 or HeLa cells were used. Therefore,

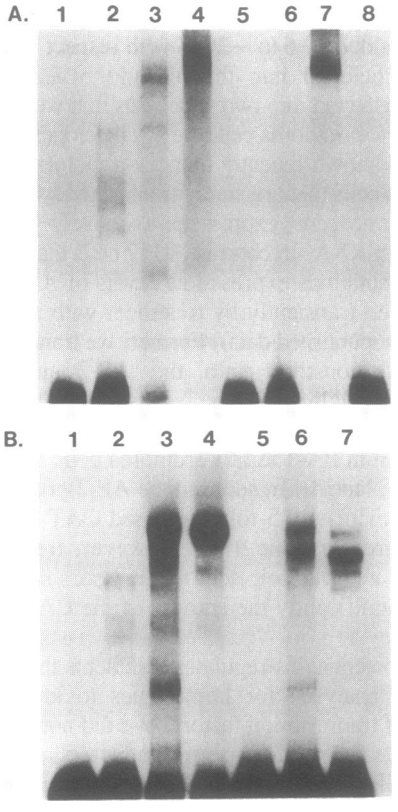


**Figure 5.** Analysis of specific protein DNA interactions in the AP-2 promoter by gel mobility shift assays. (**A**) Gel shifts using the labeled AP-2 site from the AP-2 promoter (see Materials and Methods) and a 50-fold excess of unlabeled homologous binding site for competition. The following proteins were added: (1) 10 μg BSA; (2) 1 ng recombinant AP-2A purified from *E.coli*; (3) 2 μg 9117 nuclear extract; (4) 5 μg 9117 nuclear extract; (5) 1 ng recombinant AP-2A competed with unlabeled binding site; (6) 5 μg 9117 nuclear extract competed with unlabeled binding; site (7) 5 μg of HeLa cell nuclear extract; (8) 5 μg HeLa nuclear extract competed with unlabeled binding site. (**B**) Gel shifts using the labeled AP-2 site from the AP-2 promoter and for competition 50-fold excess of an unlabeled AP-2 binding site from the huMTIIa promoter. The following proteins were added: (1) 10 μg BSA; (2) 1 ng recombinant AP-2A; (3) 10 μg 9117 cell nuclear extract; (4) 10 μg HeLa cell nuclear extract; (5) 1 ng recombinant AP-2A protein competed with unlabeled huMTIIa site; (6) 10 μg of 9117 cell nuclear extract competed with unlabeled huMTIIa site; (7) 10 μg of HeLa cell nuclear extract competed with unlabeled huMTIIa site.

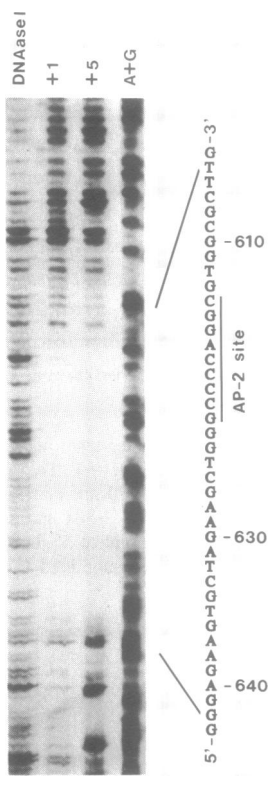


**Figure 6.** DNAaseI footprint analysis of the AP-2 consensus binding site at position −622 in the AP-2 promoter. Lanes from left to right show unprotected DNAaseI digest and reactions supplemented with 1 and 5 μg AP-2A protein. The A+G sequence reaction products were used as a molecular weight marker. The protected nucleic acid residues and their relative position with respect to the ATG protein start codon are indicated at the right (see figure 3).

the specificity of binding was further analysed by competition with 50-fold excess of a synthetic AP-2 site from the human metallothionein promoter (5). This site competed entirely all bands shifted by the recombinant AP-2 protein and reduced significantly the shift activity observed with PA-1 and HeLa cell nuclear extracts. However, a substantial part of the shift activity was not competed even by a high excess of unlabeled huMTIIA site.

To verify further a direct protein-DNA interaction at the DNA fragment from −668 to −562 we performed DNAaseI protection analysis using purified recombinant AP-2A protein. We found a specific footprint covering nucleotides −643 to −604 located around the AP-2 consensus site at position −622 (fig. 6).

## DISCUSSION

Here we report on the molecular cloning of the genomic AP-2 locus, describe its organization and give an initial characterization of cis-regulatory elements within the 5′ genomic region. The exon structure of the AP-2 gene was determined by sequencing continuously all 7 genomic regions complementary to the AP-2A and AP-2B cDNAs including the six intervening introns. The 5′ mRNA end was mapped by RNAase protection experiments and directly cloned by RACE-PCR using poly(A)$^+$-RNA from PA-1 6928 human teratocarcinoma cells. These experiments revealed that AP-2A mRNAs are initiated from two sites 220 and 271 bases upstream from the translation start codon. The downstream start site at -220 appears to be used preferentially in teratocarcinoma cells.

We have analysed previously the pattern of AP-2 mRNA expression in human and murine embryonic cells by Northern blot hybridization (8, 27). In both species we detected an identical pattern of at least two approximately 2.0 and 2.8 kb AP-2A mRNAs and one approximately 3.2 kb AP-2B mRNA. Assuming polyadenylation tails between 100 to 400 bases these sizes match well with our results obtained from genomic clones. The two AP-2A mRNAs arise from usage of two different polyadenylation signals at a constant ratio of 1:3 and encode for identical open reading frames. It has been noticed by Williams *et al.* that Northern signals appear slightly dishomogeneous (2). Due to the two mRNA initiation sites the Northern bands represent mRNAs that differ in size by approximately 50 bases.

Exons are frequently functional modules in protein architecture that may be used in different proteins and sometimes in different functional modes such as the zinc-finger-motif. A comparison

of the exon structure with data available from a series of deletion mutants and Gal4 fusion proteins (4,6) reveals a very complex protein architecture. Transactivation is mediated by a proline- and glutamine-rich region located on a single exon between amino acids 16 and 150. We have analysed whether this protein motif is present in other proteins. A comparison to the EMBL data library revealed matches to the transcription activation domains of CTF/NF1 (29) and OTF-2 (30) which have been discussed in detail previously (4). However, these matches are very low in score and mainly due to the frequent proline and glutamine residues. Thus, we are unable to decide whether the proline-glutamine-rich motif has evolved by divergent evolution of a common ancestor or by convergent evolution of different peptides. No apparent function can be allocated to the small first and third exons. Since a slight increase in transcription activation was observed after omission of the first 51 amino acids (4) it is possible that the first exon is involved in negative regulation of transactivation.

It has been shown that functional DNA binding of AP-2 requires dimerization and is dependent on amino acids located on exons 4 to 7 between residues 204 and 409 (4,6). Williams and Tjian suggested that DNA binding involves a positively charged protein region between amino acids 227 and 278 which are located on two different exons, i.e. 4 and 5. Omission of bases in exon 5, 6 and 7 also eliminates the ability of the AP-2A protein to dimerize. Thus, two different protein functions, dimerization and DNA binding, are mediated by an overlapping set of exons. Possibly, efficient DNA binding and dimerization occur cooperatively and have therefore evolved from the same functional protein modules.

Our analysis of the genomic region located 5′ upstream from the AP-2 gene from nucleic acid −6 to −2013 with respect to the ATG translation start codon revealed that this sequence contains at least part of the cis-regulatory elements required for cell type-specific AP-2 gene expression. Promoter activity was found in the human teratocarcinoma cell line PA-1 but no activity was detected in murine F9 teratocarcinoma cells which do not express significant levels of AP-2. Interestingly, we have observed recently that PA-1 cells transformed by mutated ras genes express increased levels of AP-2 mRNAs in comparison to other PA-1 cells (Kannan *et al.* unpublished data). We now demonstrate that the 2 kb AP-2 promoter fragment is approximately four times as active in clone 6928 cells as in clone 9117 cells. 6928 cells were derived from wild-type PA-1 cells by stable transfection of a single copy mutated N-ras oncogene (31). This raises the interesting possibility that ras-controlled signal transduction pathways regulate a cis-acting element located on the 2 kb AP-2 promoter fragment.

Lüscher *et al.* have analysed the induction of AP-2 in N-Tera 2 human teratocarcinoma cells by retinoic acid and shown by nuclear run-on assays that it occurs on a transcriptional level (7). We have observed a very similar pattern of induction in PA-1 clone 9117 cells (8). However, treatment with retinoic acid does not alter significantly the activity of the 2 kb CAT reporter construct and therefore other cis-regulatory elements including the retinoic acid response element must be located either further upstream or possibly downstream from the first exon.

A computer analysis of the AP-2 promoter revealed only few putative binding sites for transcription factors including an AP-2 site at nucleic acid residue −622 (2,4,6), a GCF binding site at −384 (32, 33) and multiple TCC elements between residues −240 and −100 (34). The transcription factor GCF has been implicated in regulation of the epidermal growth factor receptor promoter (32) and is widely expressed in many human tissues and cell lines. Since it has been shown that the expression of GCF is rapidly induced by phorbol esters (33) it is possible that the AP-2 promoter is subject to regulation by phorbol esters in tissues that express GCF.

Among the putative transcription factor binding sites within the AP-2 promoter we have analysed the AP-2 consensus site at residue −622 in more detail. Our transfection data provide evidence that the protein AP-2A activates its own promoter. A similar positive autoregulatory loop was detected in the c-jun promoter and has been implicated in prolonging the effects of transient gene induction by stimulatory signal transduction pathways (35). Binding of AP-2A protein to the site at −622 was verified using both gel mobility-shift assays and DNAaseI protection analyses. The results demonstrate that purified AP-2A protein binds specifically to the site and is competed by the homologous AP-2 site as well as by a second well studied AP-2 site from the human metallothionein IIa promoter. Using purified AP-2A protein we observed a footprint covering nucleotides −643 to −604 located asymmetrically around the AP-2 binding site. The geometry of the protection pattern suggests that the homodimeric AP-2 protein complex does not touch evenly both half sides of the core binding element.

Interestingly, the gel shift pattern differed when PA-1 clone 6928 nuclear extracts rather than purified AP-2A protein were used. While competition with the homologous AP-2 site eliminated all gel shifting activity we observed only a partial reduction when an excess of unlabeled huMTIIa site was used. These data indicate that multiple transcription factors interact at or closely nearby the AP-2 autoregulatory element.

## ACKNOWLEDGEMENTS

We are indebted to Dr Schüle (Freiburg) for helpful discussions, reading critically the manuscript and providing the plasmid pCMX-PL1. This work was supported in part by the DFG and the Wilhelm Sander-Stiftung. A.I. is supported by a predoctoral fellowship from the Graduiertenkolleg Therapieforschung Onkologie.

## REFERENCES

1. Mitchell, P.J., Wang, C., and Tjian, R. (1987) *Cell* 50, 847–861.
2. Williams, T., Admon, A., Lüscher, B., and Tjian, R. (1988) *Genes Dev.* 2, 1557–1569.
3. Gaynor, R.B., Muchardt, C., Xia, Y.R., Kliask, I., Mohandas, T., Sparkes, R.S., and Lusis, A.J. (1991) *Genomics* 10, 1100–1102.
4. Williams, T., and Tjian, R. (1991). *Genes Dev.* 5, 670–682.
5. Imagawa, M., Chiu, R., and Karin, M. (1987) *Cell* 51, 251–260.
6. Williams, T., and Tjian, R. (1991) *Science* 251, 1067–1071.
7. Lüscher, B., Mitchell, P.J., Williams, T., and Tjian, R. (1989) *Genes Dev.* 3, 1507–1517.
8. Buettner, R., Kannan, P., Imhof, A., Bauer, R., Yim, S.O., Glockshuber, R., VanDyke, M.W., and Tainsky, M.A. (1993) *Mol. Cell. Biol.* 13, 4174–4185.
9. Mitchell, P.J., Timmons, P.M., Hebert, J.M., Rigby, P.W., and Tjian, R. (1991) *Genes Dev.* 5, 105–119.
10. Leask, A., Byrne, C., and Fuchs, E. (1991) *Proc. Natl. Acad. Sci. USA* 88, 7948–7952.
11. Snape, A.M., Winning, R.S., and Sargent, T.D. (1991) *Development* 113, 283–293.
12. Behrens, J., Löwrick, O., Klein-Hitpass, L., and Birchmeier, W. (1991) *Proc. Natl. Acad. Sci. USA* 88, 11495–11499.
13. Kopp, H., Masters, J.R.W., Wieland, W., Hofstaedter, F., and Buettner, R. (1993) *Verh. Dt. Ges. Pathol.* 77, 129–132.

14. Sanger, F., Nicklen, S., and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA* 74, 5463–5467.
15. Protocol provided with the DyeDeoxy Terminator Cacle Sequencing Kit. Applied Biosystems, Colorado, USA. Part Number 401388. Also in: Halloran, N., Du, Z., Wilson, R.K. (1992) *In* Griffin, H.G., and Griffin, A.M. (Eds.) Methods in Molecular Biology, Vol.10. The Humana Press, Clifton, NJ.
16. Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) *Molecular Cloning: A Laborartory Manual* (2nd edition). Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
17. Devereux, J., Haeberli, P., and Smithies, O. (1984) *Nucl. Acids Res.* 12, 387–395.
18. Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J., Rutter, W.J. (1979) *Biochemistry* 18, 5294–5299.
19. Edwards, J.B.D.M., Delort, J., and Mallet, J. (1991) *Nucl. Acids Res.* 19, 5227–5232.
20. Luckow, B., and Schütz, G. (1987) *Nucl.Acids Res.* 15, 5490.
21. Graham, F.L., and van der Eb, A.J. (1973) *Virology* 52, 456–467.
22. Umesono, K., Murakami, K.K., Thompson, C.C., and Evans, R.M. (1991) *Cell* 65, 1255–1266.
23. Buettner, R., Yim, S.O., Hong, Y.S., Boncinelli, E., and Tainsky, M.A. (1991) *Mol. Cell. Biol.* 11, 3573–3583.
24. Tainsky, M.A., Yim, S.O., Krizman, D.B., Kannan, P.J., Mukhopadyay, T., and Buettner, R. (1991) *Oncogene* 6, 1575–1582.
25. Gorman, C.M., Moffat, L.F., and Howard, B.H. (1982) *Mol. Cell. Biol.* 2, 1044–1053.
26. Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., and Struhl, K. (eds.) Current Protocols in Molecular Biology. Vol.2, pp 12.4.1–12.4.6, John Wiley and Sons, New York, 1990.
27. Moser, M., Pscherer, A., Bauer, R., Imhof, A., Seegers, S., Kerscher, M., and Buettner, R. (1993) *Nucl. Acids Res.* 21, 4844.
28. Tainsky, M.A., Krizman, D.B., Chiao, P.J., Yim, S.O., and Giovanella, B.C. (1988) *Anticancer Res.* 8, 899–914.
29. Mermod, N., O'Neill, E.A., Kelly, T.J., and Tjian, R. (1989) *Cell* 58, 741–753.
30. Gerster, T., Balmaceda, C.G., and Roeder, R. (1990) *EMBO J.* 9, 1635–1643.
31. Tainsky, M.A., Cooper, C.S., Giovanella, B.C., and Vande Woude, G.F. (1984) *Science* 225, 643–645.
32. Kageyama, R., and Pastan, I. (1989) *Cell* 59, 815–825.
33. Johnson, A.C., Kageyama, R., Popescu, N.C., and Pastan, I. (1992) *J. Biol. Chem.* 267, 1689–1694.
34. Lafyatis, R., Denhez, F., Williams, T., Sporn, M., and Roberts, A. (1991) *Nucl. Acids. Res.* 19, 6419–6425.
35. Angel, P., Hattori, K., Smeal, T., and Karin, M. (1988) *Cell* 55, 875–885.