

Transcriptional Analysis of Minute Virus of Mice P₄ Promoter Mutants

JEONG K. AHN,¹ BRIAN J. GAVIN,^{2†} GYANENDRA KUMAR,^{2‡} AND DAVID C. WARD^{1,2*}

Departments of Human Genetics^{2} and Molecular Biophysics and Biochemistry,¹
Yale University School of Medicine, 333 Cedar Street, New Haven, Connecticut 06510*

Received 21 April 1989/Accepted 10 August 1989

A series of 5' deletion, internal deletion, and linker-scanning mutants of the minute virus of mice P₄ promoter were constructed and analyzed for transcriptional activity in nuclear extracts of mouse A92L fibroblasts. A GC box and a TATA box essential for in vitro transcription from the P₄ promoter were localized between nucleotides 150 and 180 (−55 to −25 relative to the primary RNA start site). Although this region also exhibited homologies to other transcriptional control elements, the simian virus 40 enhancer, and the adenovirus E1A enhancer, only the GC box and TATA box appear functional. These two motifs also play an essential role in vivo, although additional upstream sequences (between −139 and −55) are required for optimal transcription. DNase I footprinting, competitive gel retardation assays, and UV-photocrosslinking were used to identify Sp1-like proteins of 95 and 120 kilodaltons in A92L extracts that interact with the GC box of the minute virus of mice P₄ promoter.

Minute virus of mice (MVM) is a DNA virus of the family *Parvoviridae*. The viral genome is a linear single-stranded molecule, approximately 5 kilobases (kb) long (9), containing terminal palindromic sequences which are essential for viral replication. MVM, because of its limited genetic capacity, is highly parasitic of host cell functions for all aspects of its life cycle. Unlike other small DNA viruses such as simian virus 40 (SV40) or polyomavirus, MVM is unable to force the host cell to enter the S phase of the cell cycle. However, since MVM requires a cellular function expressed transiently during the S phase for the early stages of viral replication (42), productive infection occurs only in rapidly dividing cell populations. In addition, productive infection by MVM, and parvoviruses in general, appears to be highly dependent on the differentiated state of the host cell (43), particularly in vivo (39).

The cell specificity demonstrated by MVM in vivo has been reproduced in tissue culture by using the prototype strain of MVM, MVM(p), and a spontaneously arising immunosuppressive variant, MVM(i) (12). In culture, MVM(i) productively (lytically) infects a mouse T-cell lymphoma line, S-49, whereas MVM(p) productively infects mouse fibroblast cell lines, such as A92L (A9 cells). At the genomic sequence level these viruses are 96% homologous (1, 34), yet neither virus strain can productively infect the other's host cell (44). It has been suggested that the block in a restrictive infection [i.e., MVM(p) into S49 cells] is mediated at the level of transcriptional initiation (E. M. Gardiner, B. A. Spalholz, D. C. Ward, and P. Tattersall, manuscript in preparation). This block appears to be due to the lack of a positively acting cell-specific protein (44) which, in a productive infection, may interact with some component of the viral capsid to allow initiation of viral transcription (14). For this reason, we have initiated a detailed analysis of the factors which control MVM transcription.

The transcriptional organization of MVM is relatively simple (Fig. 1). The virus contains two overlapping transcription units with RNA start sites at map units 4 and 38 (30) and a single polyadenylation site at ca. map unit 95 (7). The promoters for these transcription units, termed P₄ and P₃₈, produce the three major viral mRNA classes R1, R2, and R3. The R1 (4.8-kb) and R2 (3.3-kb) transcripts are both initiated at map unit 4 and code for the two viral nonstructural proteins, NS-1 and NS-2, which are essential for viral replication (9) and appear to transactivate the P₃₈ promoter (10, 32, 33). The R3 mRNAs (3.0 kb) are initiated at map unit 38 and produce the viral coat proteins VP-1 and VP-2 via differential splicing (28). A third coat protein, VP-3, is produced by specific proteolytic cleavage of VP-2 (9). Clemens and Pintel (8) have shown that the production of viral mRNAs is organized into a classic early-late strategy, with R1 and R2 being produced before R3 in a productive infection.

To define the functional elements responsible for the control of transcription from the P₄ promoter of MVM(p), we constructed a series of 5' deletion, internal deletion, and linker-scanning mutants. Here we report the construction of the mutant set, the initial studies on their transcriptional activity in A9 cells in vivo and in vitro, and the identification of proteins that interact with high affinity with a functional GC box motif in the P₄ promoter.

MATERIALS AND METHODS

Construction of mutants. (i) pMpC mutants. pMM984 is the infectious MVM(p) clone which contains the entire MVM genome (27). pMpC was constructed by subcloning the MVM(p) EcoRI C fragment (nucleotides 1 to 1084) from pMM984 as a BamHI-EcoRI fragment into pUC18 between the BamHI and EcoRI sites. For mutant constructions, restriction sites in the virus are indicated by the nucleotide number and those in the plasmid polylinker are indicated by PL. For pMpCΔ60 construction, pMpC was digested with PstI (PL/nucleotide [nt] 415) and the ca. 450-base-pair (bp) fragment containing the P₄ promoter was isolated. This fragment was digested with Fnu4HI (nt 54), and the resulting fragments were blunt ended by treatment with Klenow

* Corresponding author.

† Present address: Department of Cell and Developmental Biology, Roche Institute of Molecular Biology, Nutley, NJ 07110.

‡ Present address: Department of Molecular Biology and Genetics, Wayne State University School of Medicine, Detroit, MI 48201.

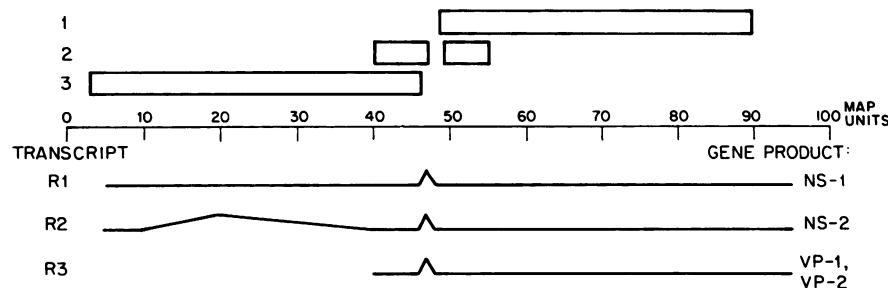


FIG. 1. Transcription map of MVM. The open boxes above the genomic map indicate the large open reading frames found within the virion DNA. The transcripts generated from the virus are indicated below the map. R1 and R2 are initiated at nt 204 and 205 from the P₄ promoter. R3 is initiated at ca. nt 2005 from the P₃₈ promoter. The primary translation product(s) of each mRNA is indicated to the right.

polymerase. This mixture was next digested with *Nco*I (nt 260) and the ca. 200-bp promoter containing fragment was isolated. For pMpCΔ136 and pMpCΔ170 construction, pMpC was digested with *Nco*I (nt 260) and *Hinc*II (PL) and the resulting fragments were gel purified. The ca. 270-bp fragment containing the P₄ promoter and 55 bp of coding sequence was subjected to partial digestion with *Dra*I, and the partial digestion products were gel purified. Deleted promoter fragments were reconstructed into the 3'-flanking viral sequence by ligating the blunt (either *Dra*I or Klenow-treated *Fnu*4HI-*Nco*I) fragments into the large fragment isolated from the pMpC *Hinc*II-*Nco*I double digest. Correct constructions were identified by both restriction mapping and dideoxy sequencing. This analysis demonstrated that the two *Dra*I mutants were as expected. The *Fnu*4HI construction appeared to have lost 6 bp during construction; thus, the deletion endpoint was nt 60 rather than nt 54.

(ii) pMB mutants. For the wild-type P₄ construct pMB415, a *Bam*HI-*Pst*I fragment (nt 1 to 415) containing the P₄ promoter was subcloned from pMM984 into the pBluescript⁺ (Stratagene Inc.) vector.

For 5' deletion mutants, 4 pmol of pMB415 was linearized by complete digestion with *Sac*I and *Bam*HI. Unidirectional deletions were generated from the *Bam*HI site by digestion with 120 U of exonuclease III, which is specific for 5' overhanging ends. Samples were taken every 30 s from 0 to 10 min, and reactions were stopped by the addition of an equal volume of 40 mM EDTA. After phenol-chloroform extraction, each sample was ethanol precipitated and suspended in 1× mung bean nuclease buffer (30 mM sodium acetate [pH 5.0], 50 mM NaCl, 1 mM ZnCl₂, 5% glycerol). After 0.25 U of mung bean nuclease had been added, the reactions were incubated at 30°C for 30 min to make the deletion endpoint blunt and then stopped by the addition of an equal volume of 40 mM EDTA. The reaction products were analyzed by agarose gel electrophoresis, and plasmid DNAs containing the desired deletions were excised from the gel and isolated by electroelution. To ensure that the deletions were blunt ended, we treated the isolated bands with Klenow polymerase for 30 min at 37°C. *Bgl*II linkers (10 bp; 5'-GAAGATCTTC-3') were ligated to the deleted plasmids, and excess *Bgl*II linkers were removed by digestion with *Bgl*II followed by agarose gel purification. The 5' deletion mutants were then circularized by intramolecular self ligation. This results in the insertion of a *Bgl*II linker at the 5' deletion endpoint. Since the *Sac*I site was untouched by exonuclease III, the sequences upstream of the deleted end remained the same for all pMB 5' deletion mutants.

For 3' deletion mutants, pMB415 was linearized by complete digestion with *Apa*I and *Hind*III. Deletion mutants

were then constructed by using the exonuclease III and mung bean nuclease method described above. Another set of 3' deletion mutants were also constructed by using BAL 31 nuclease to resect *Nco*I (viral nt 260)-digested pMB415. After the BAL 31 reaction, MVM sequences 3' to the *Nco*I site were removed from the deleted plasmids by *Hind*III digestion. Following agarose gel purification, the deleted plasmids were blunt ended by treatment with Klenow polymerase. *Bgl*II linkers were ligated to the plasmids, and excess linkers were removed by *Bgl*II digestion. 3'-Deleted linear plasmids were then purified by agarose gel electrophoresis and self-ligated. The deletion endpoints of all 5' and 3' deletion mutants were determined by dideoxy DNA sequence analysis of T3 and T7 RNA transcripts as recommended by the manufacturer (Stratagene).

(iii) Linker-scanning and internal deletion mutants. Linker-scanning mutants were constructed by combining pairs of 5' and 3' deletion mutants whose endpoints were separated by 10 bp, the same size as a *Bgl*II linker. 5' deletion mutants were digested with *Bgl*II and *Kpn*I, and the small *Bgl*II-*Kpn*I fragments were subcloned into the large fragment of the appropriate *Bgl*II-*Kpn*I-digested 3' deletion mutant. Internal deletion mutants were constructed by the same method, with appropriate combinations of 5' and 3' deletion mutants.

(iv) P₄β mutants. The parent plasmid for these constructs, pCH110-x, was derived from the eucaryotic β-galactosidase expression vector pCH110 (Pharmacia, Inc.) which contains the lacZ gene driven by the SV40 early promoter. pCH110-x was constructed by partially digesting pCH110 with *Pvu*II and isolating the linear molecules. An 8-bp *Xba*I linker was ligated to the blunt ends, excess linkers were removed by digestion with *Xba*I, and the molecules were recircularized by self-ligation. The resulting plasmids were then screened for the insertion of an *Xba*I restriction site upstream of the SV40 early promoter. This allowed easy removal of the SV40 promoter by a *Hind*III-*Xba*I double digest. The vector for subcloning the P₄ promoter mutants was prepared by digesting pCH110-x with *Hind*III, filling in the overhanging ends with Klenow polymerase, digesting with *Xba*I, and gel purifying the large fragment containing the β-galactosidase expression cassette but lacking the promoter.

P₄ promoter fragments were prepared as follows. Wild type pMB415 was digested with *Taq*I, and a 656-bp fragment containing the intact P₄ promoter was gel purified. This fragment was blunt ended with Klenow polymerase and digested with *Xba*I, and the 250-bp promoter fragment was gel purified. For preparation of pMpCΔ mutants, pMpC deletion mutants were digested with *Neol*-*Hind*III and the promoter-containing fragments were purified. These fragments were then blunt ended as described above and ligated

to 10-bp *Xba*I linkers. The mixture was digested with *Taq*I, treated with Klenow polymerase, and digested with *Xba*I to remove excess linkers, and the *Xba*I-blunt (*Taq*I) promoter-containing fragments were gel purified. These fragments were then ligated into the pCH110 vector prepared as described above.

Preparation of nuclear extracts. Nuclear extracts were prepared from mouse A9 fibroblasts and human HeLa cells as previously described (18). A9 cells were cultured in suspension by using Joklik modified minimal essential medium with 5% heat-inactivated fetal calf serum (FCS). HeLa cells were suspension cultured in Autopow-modified minimal essential medium with 5% FCS. For each nuclear extract, the protein concentration was determined by the modified Bradford method (4); the concentrations generally ranged from 8 to 10 mg/ml for HeLa extracts and 5 to 7 mg/ml for A9 extracts.

Each extract was initially characterized for transcriptional activity by using an in vitro runoff transcription assay. Runoff transcription was carried out in a final volume of 20 μ l containing 12 mM N-2-hydroxyethylpiperazine-N'-2-ethanesulfonic acid (HEPES; pH 7.9), 12% glycerol, 60 mM KCl, 120 μ M EDTA, 300 μ M dithiothreitol, 300 μ M phenylmethylsulfonyl fluoride, 200 μ M ATP, 200 μ M CTP, 200 μ M UTP, 10 μ M GTP, 5 μ Ci of [α -P³²]GTP, 5 mM creatine phosphate, and nuclear extract. Each extract was optimized for three variable parameters: MgCl₂ concentration, DNA template concentration, and nuclear extract concentration. These generally ranged from 6 to 8 mM MgCl₂, 25 to 50 μ g of template per ml, and 50 to 60% extract for HeLa cells and 10 to 12 mM MgCl₂, 15 to 25 μ g of template per ml, and 30 to 50% extract for A9 cells. DNA templates were prepared by digesting plasmids with *Eco*RI (pMpC mutants) or *Pst*I (pMB mutants). The concentration of DNA templates was determined by using a Hoefer fluorometer. After incubation at 30°C for 60 min, the reactions were stopped by adding 5 μ l of stop buffer (25 mM EDTA, 2.5% sodium dodecyl sulfate [SDS], 2.5 μ g of carrier tRNA per ml). After the addition of 2 μ l of proteinase K (10 mg/ml), the reaction mixture was incubated at 65°C for 10 min. The labeled RNA was then ethanol precipitated and suspended in 10 μ l of TE8 buffer (10 mM Tris hydrochloride [pH 8.0], 1 mM EDTA). After the addition of 10 μ l of urea-dye mix (0.1% bromophenol blue plus 0.1% xylene cyanol in 10 M urea), the samples were loaded on a 5% polyacrylamide gel containing 8 M urea. In general, A9 extracts showed ca. 10% of the transcriptional activity of HeLa extracts.

Primer extension analysis. Unlabeled RNA was transcribed in vitro by using either linearized or supercoiled templates as indicated. Transcription conditions were identical to those used for runoff transcription, except that the nucleoside triphosphates were at 600 μ M and contained no labeled nucleotides. Ethanol-precipitated RNA was suspended in 20 μ l of DNase mixture, which contained 40 mM Tris hydrochloride (pH 7.5), 10 mM NaCl, 6 mM MgCl₂, 20 U of RNase inhibitor, and 1 U of RQ-1 DNase (Promega Biotech) per μ g of template DNA. After incubation at 37°C for 30 min, the DNase mixture was brought to 100 μ l with double-distilled H₂O, extracted with equal volumes of phenol and chloroform/isoamylalcohol (24:1), and ethanol precipitated. RNA pellets were lyophilized and suspended in 25 μ l of hybridization buffer containing 300 mM NaCl, 10 mM Tris hydrochloride (pH 7.4), 1 mM EDTA, and 2 ng of an end-labeled oligonucleotide complementary to the transcribed strand. The hybridization mixture was incubated at 65°C for 30 min and 42°C for 45 min and precipitated by

addition of 2.5 vol of ethanol. After ethanol precipitation, primer-hybridized RNA was suspended in 15 μ l of reverse transcription buffer, which consisted of 50 mM Tris hydrochloride (pH 8.3), 150 mM KCl, 10 mM MgCl₂, 20 mM β -mercaptoethanol, 20 μ g of bovine serum albumin per ml, 10 mM dithiothreitol, 660 μ M deoxynucleoside triphosphates, and 12 U of avian myeloblastosis virus reverse transcriptase. Reverse transcription was carried out at 42°C for 30 min, and the cDNA was ethanol precipitated. The samples were suspended in 6 μ l of formamide dye and loaded onto an 8% polyacrylamide sequencing gel containing 8 M urea. For sequencing markers, pMBA142 was transcribed with T7 RNA polymerase. In vitro-generated RNA was then hybridized to the same end-labeled oligonucleotide as used for primer extension analysis and sequenced as previously described.

Transient-transfection assays. All transfection experiments were carried out in triplicate. Calcium phosphate DNA precipitates (27) were made in batch and divided among three plates. A9 fibroblasts (10⁶ cells per 60-mm dish) were first plated in Dulbecco modified Eagle medium (DME)-5% heat-inactivated FCS (DME-FCS) and incubated overnight. The following morning, cells were fed with 5 ml of DME-10% heat-inactivated FCS-1× nonessential amino acids (DME-FCS-acids), and 4 h later 0.5 ml of precipitate containing 15 μ g of reporter construct and 5 μ g of pSV2CAT DNA was added to the media. Control transfections contained 20 μ g of pBR322 or calf thymus DNA. After an additional 4 h, the medium was removed, and the cells were shocked with 1 ml of 20% glycerol in DME-FCS for 1 min at room temperature, washed twice with 5 ml of DME-FCS, and fed with 5 ml of DME-FCS-acids containing 5 mM sodium butyrate. At 16 h after addition of medium plus butyrate, the cells were fed with fresh DME-FCS-acids. At 48 h after addition of the precipitates, the medium was removed, and the cells were washed with phosphate-buffered saline and scraped off the plates in 0.5 ml of buffer containing 40 mM Tris hydrochloride (pH 7.4), 150 mM NaCl, and 1 mM EDTA. The cells were then pelleted and suspended in 50 μ l of 250 mM Tris hydrochloride (pH 7.8). Extracts were prepared by three freeze-thaw cycles followed by clarification by centrifuging for 10 min at 4°C.

β -Galactosidase assays were carried out in a total volume of 100 μ l essentially as described previously (38). Reactions containing 10 μ l of extract were incubated at 37°C until all experimental reactions were within the linear range and then stopped by the addition of 300 μ l of 100 mM lactose. Chloramphenicol acetyltransferase (CAT) assays were carried out by standard methods (16). Following thin-layer chromatography, CAT activity was quantitated by excision and scintillation counting of acetylated and nonacetylated material. Each extract was assayed twice for both β -galactosidase and CAT activity. Each β -galactosidase value was normalized against each CAT value, and then the resulting four ratios were averaged for each extract. The β -galactosidase/CAT ratios were then expressed as percentages relative to the wild-type activity.

DNase I footprinting analysis. (i) **MVM templates.** Both wild-type and linker-scanning mutants of the P₄ promoter were used. To make the noncoding strand probe, each construct was digested with *Bam*HI, treated with alkaline phosphatase, and labeled with [³²P]PO₄ by using T4 polynucleotide kinase and [γ -³²P]ATP. After digestion with *Nco*I, the labeled *Bam*HI-*Nco*I fragment (260 bp; nucleotides 1 to 260) was isolated from a 1% agarose gel. For the coding-strand probe, each construct was digested with *Pst*I, treated

with alkaline phosphatase, and labeled with [γ -³²P]ATP. After digestion with *Bam*HI, the labeled 415-bp *Pst*I-*Bam*HI fragment was gel purified.

A preincubation step was carried out under essentially the same conditions as those for transcription reactions, with the exception of nucleotides and the following modifications: 30% nuclear extract was used with 10 mM MgCl₂ and 1 μ g of poly(dI-dC). After preincubation at 0°C for 30 min, 1 μ l of labeled probe (5 fmol) was added to 19 μ l of preincubation mixture and incubated at 0°C for an additional 30 min. After the addition of 20 μ l of dilution solution (5 mM CaCl₂, 10 mM MgCl₂), 1 μ l of freshly diluted DNase I (100 μ g/ml; Worthington Diagnostics) was added, and the mixture was then incubated at 0°C for 1 min. The DNase I reaction was stopped by adding 100 μ l of stop buffer (500 μ g of tRNA per ml, 20 mM EDTA, 20 mM Tris hydrochloride [pH 7.5], 0.6% SDS, 20 μ g of proteinase K), and the mixture was then incubated at 65°C for 10 min. Following phenol-chloroform extraction, the DNA was ethanol precipitated and suspended in 10 μ l of formamide dye. Each sample was loaded on a 6% polyacrylamide wedge gel containing 8 M urea. A+G and G Maxam-Gilbert sequencing reactions were performed on the pMB415 coding and noncoding probes as sequence markers (26).

(ii) **SV40 templates.** The plasmid RI-Bam-e-4, which contains the SV40 early region, was digested with *Eco*RI, and the 5' ends were dephosphorylated with calf intestine alkaline phosphatase and labeled with [³²P]PO₄ by using T4 polynucleotide kinase and [γ -³²P]ATP. Following digestion with *Hind*III, the uniquely end-labeled probe containing the origin/promoter region (418 bp) was isolated by nondenaturing polyacrylamide gel electrophoresis. DNase I footprinting analysis was carried out under the same salt and buffer conditions as described above. In a 20- μ l reaction volume, 15 fmol of probe (*Hind*III-*Eco*RI), 3 μ g of sonicated calf thymus DNA, and the indicated amounts of competitor DNA (cold duplex 1) were incubated with 20 μ g of HeLa nuclear extract at 30°C for 30 min. Subsequent processing and analysis of the reaction mixture were as outlined above.

Gel retardation assays. (i) **Preparation of probe.** Two complementary oligonucleotides corresponding to nt 137 to 170 and 173 to 140 (5' → 3') of the MVM(p) P₄ promoter were synthesized and hybridized. The resulting duplex was end labeled with T4 polynucleotide kinase and [γ -³²P]ATP (Amersham). This was followed by treatment with Klenow polymerase in the presence of cold dATP and TTP to fill in the overhanging ends.

(ii) **Preparation of competitors.** The P₄ competitors were prepared by digestion of pMB415 with *Bam*HI and *Pst*I. The 415-bp fragment containing the intact P₄ promoter and 210 nt of 3'-flanking sequence (MVM nt 1 to 415) was gel purified and digested with *Nco*I to yield two fragments of 260 nt (1 to 260) and 155 nt (260 to 415). These fragments were isolated as above and used as the P₄ homologous and P₄ heterologous competitors (see Fig. 5). The SV40 competitor was prepared by digestion of the plasmid RI-Bam-e-4, which contains the SV40 early region from nt 346 to 1/5243 to 2533 as an *Eco*RI/*Bam*HI insert in pBR322, with *Nco*I to release the SV40 origin region. A 296-bp fragment containing the 21-bp and 72-bp repeats but lacking the early promoter TATA box (see Fig. 5) was isolated as above. To separate the SV40 transcriptional control elements, we digested pCH110-x with *Xba*I and *Hind*III to release the SV40 early promoter as a 347-bp fragment. This fragment was isolated as above and further digested with *Sfa*NI, which cuts within the 72-bp repeat (at nt 115 and 182). The three fragments generated,

which contain either a 72-bp repeat or the three 21-bp repeats (see Fig. 5), were then gel purified. The heterologous adenovirus major late promoter fragment was isolated from plasmid p Δ , which contains the intact major late promoter, by digestion with *Hind*III and *Bam*HI, and then the 457-bp promoter-containing fragment was isolated. This fragment contains 260-bp 5' and 197-bp 3' of the major late RNA start site (see Fig. 5).

For competition studies with internal deletion and linker-scanning mutants of the P₄ promoter, competitors were used as supercoiled plasmids. As a wild-type control, the plasmid pMB415 was used. The concentrations of all competitors were determined by using a Hoefer minifluorimeter as specified by the manufacturer.

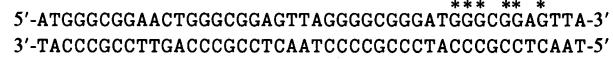
(iii) **Reaction conditions.** All assays were run essentially as described by Fried and Crothers (13) with the following modifications. A final reaction volume of 20 μ l containing 4 fmol of labeled probe, 4 μ g of protein, and 1 μ g of poly(dI-dC). Reaction conditions were identical to those described above for transcription, except that nucleotides and creatine phosphate were omitted and the MgCl₂ concentration was reduced to 6 mM. Competitor DNAs were added prior to initiation of the reaction, and the amount of poly(dI-dC) was adjusted to account for the weight of the added competitor. Reactions were initiated by the addition of protein and incubated for 30 min at 30°C. Following incubation, 15 μ l of each reaction was loaded onto a 4% polyacrylamide gel (30:1, 0.25× TBE [13]) which had been prerun for 2 h and run at 10 V/cm. Following electrophoresis, gels were dried under vacuum and autoradiographed.

Preparation of duplex 1 and duplex 2. Complementary oligonucleotides corresponding to nt 51 to 83 and 94 to 52 (5' → 3') of the SV40 genome and nt 137 to 158 and 173 to 140 (5' → 3') of the MVM genome were synthesized and hybridized together. For use as UV photocrosslinking probes, the resulting duplexes were treated with Klenow polymerase in the presence of dATP, dCTP, TTP, and [α -³²P]dGTP. This results in labeling of a single GC box motif.

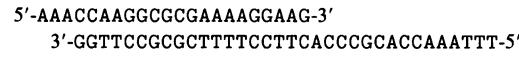
For the SV40 44-mer (duplex 1),



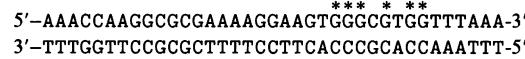
↓ Klenow, dATP, dCTP, TTP, [α -³²P]dGTP



For the MVM 37-mer (duplex 2),



↓ Klenow, dATP, dCTP, TTP, [α -³²P]dGTP



For use as competitor molecules, the above duplexes were prepared as outlined, with the exception of the Klenow polymerase fill-in, in which cold dGTP was used. In addition, a 42-bp synthetic duplex encompassing the adenovirus type 2 upstream stimulatory factor (USF) binding site (37) from nt -90 to -48 relative to the RNA start site was constructed as a heterologous competitor.

UV cross-linking of internally ³²P-labeled oligoduplex-protein complex. In a 20- μ l reaction volume, 40 fmol of internally labeled oligoduplex was incubated with 10 μ g of

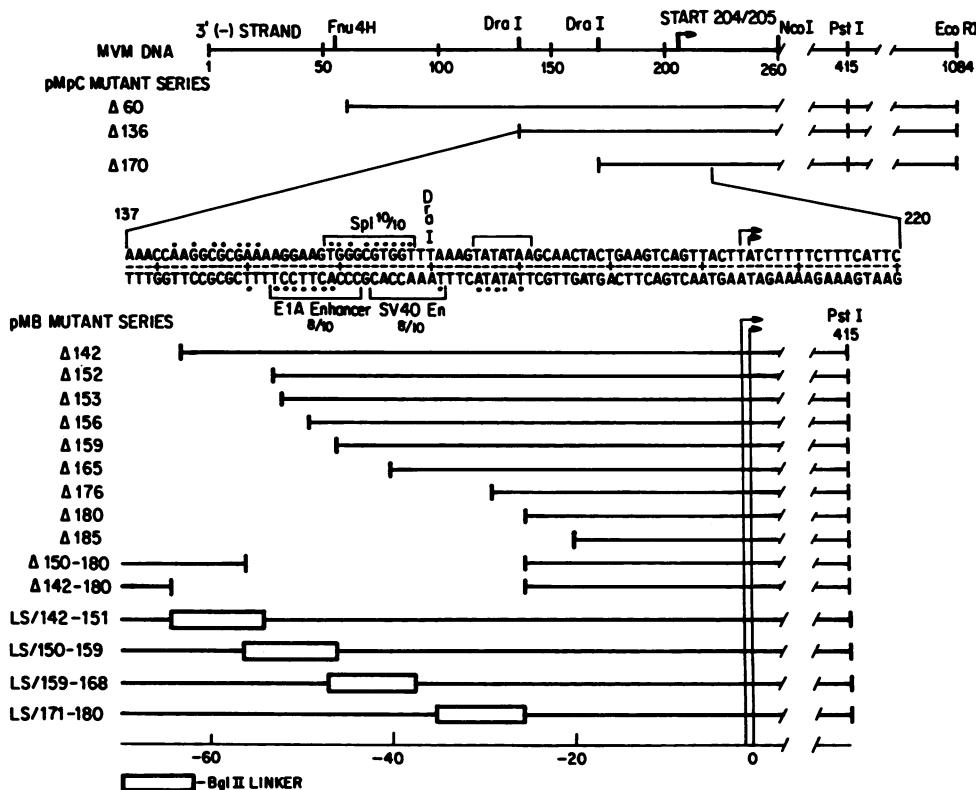


FIG. 2. Schematic representation of P₄ promoter mutants. For the pMpC mutant series, the line labeled MVM DNA is representative of the insert in the wild-type clone pMpC. The location of the Fnu4HI site is nt 54, and the two DraI sites are at nt 136 and 170. In the 5' deletion mutants, all sequences downstream of the deletion endpoint are identical to the wild type. The boxes on the sequence denote homologies to known transcriptional control elements, and the arrows indicate the major transcriptional start sites. For the pMB mutant series, the wild-type construct contains viral sequences from nt 1 to 415 (BamHI-PstI). The deletion endpoints of the pMB mutant series are aligned with the sequence. Each 5' deletion mutant has a 10-bp BglII linker attached at the deletion endpoint. Internal deletions also contain a BglII linker in the deleted region. The nucleotides which are altered in each linker-scanning mutant are indicated by dots on the P₄ sequence.

protein (nuclear extract or Sp1-enriched fractions) and 0.5 µg of poly(dI-dC) under conditions identical to those used in the gel retardation assays. The reaction mixture was incubated at 30°C for 30 min and then irradiated with a hand-held UV lamp from 2 in. (ca. 5 cm) above the mixture (ca. 4.5 J/m² per s) for an additional 30 min at 4°C. The samples were treated with 2 µg of DNase I for 1 min, 10 µl of SDS loading dyes (300 mM dithiothreitol, 6% SDS, 30% glycerol, 0.25% bromophenol blue) was added, and the reactions were electrophoresed without heating on a SDS-10% polyacrylamide gel. Following electrophoresis, the gel was dried under vacuum, and the prestained protein marker bands were marked with radioactive ink and autoradiographed. The cross-linking procedure was optimized by cross-linking of duplex 1 with the Sp1-enriched fraction of HeLa extract. UV irradiation for 40 min and addition of 200 mM KCl to the reaction mixture resulted in the maximum formation of the 95-kilodalton (kDa) photoadduct. However, to maintain consistency with the binding of footprinting assays, cross-linking was routinely carried out in 60 mM KCl.

The 1.5-kb SV40 DNA fragment used as a competitor in cross-linking experiments was prepared by digesting pRI-Bam-e-4 with Pvul and BstXI and subjecting it to gel purification.

Fractionation of nuclear extract. HeLa nuclear extracts were fractionated essentially as described by Dynan and Tjian (11) to yield Sp1-enriched and Sp1-devoid fractions. A 2.5-ml portion of nuclear extract containing 15 mg of protein

was loaded onto a 2-ml heparin-Sepharose column which had been equilibrated in buffer D (20 mM HEPES [pH 7.9], 20% glycerol, 100 mM KCl, 0.2 mM EDTA, 0.5 mM phenylmethylsulfonyl fluoride, 0.5 mM dithiothreitol). The column was washed with 4 ml of buffer D, 4 ml of buffer D containing 0.2 M KCl, and 4 ml of buffer D containing 0.4 M KCl. The 0.4 M step fraction was then dialyzed against two changes, 50 volumes each, of buffer D and loaded onto a 2-ml DEAE-Sepharose column equilibrated with buffer D. The DEAE column was then washed with an additional 2 ml of buffer D, and the total flowthrough fraction (ca. 6 ml) was loaded onto a 2-ml P-11 phosphocellulose column which had been equilibrated in buffer D. The flowthrough fraction was recycled over the column twice, and then the column was washed with 4 ml of buffer D. Proteins were eluted by sequential steps of 4 ml buffer D-0.35 M KCl and 4 ml of buffer D-0.6 M KCl to yield Sp1-enriched and Sp1-devoid fractions, respectively.

RESULTS

Analysis of the pMpC mutant series. To elucidate the cis-acting transcriptional control elements of the MVM P₄ promoter, 5' deletion, internal deletion, and linker-scanning mutants were constructed as described in Materials and Methods. The first set of mutants to be analyzed were the pMpC mutant series (Fig. 2). The parent plasmid, pMpC, contained the entire P₄ promoter and ca. 880 bp of 3'-

flanking sequence of MVM(p) (nt 1 to 1084). 5' deletions were made by using convenient restriction sites upstream of the transcription initiation site, previously reported to be at nt 205 ± 5 (2). To accurately assess the functionality of these mutants in the normal milieu within which the P₄ promoter functions, i.e., murine fibroblasts, we prepared nuclear extracts from mouse A9 fibroblasts and used them to transcribe the promoter constructs. In vitro-generated transcripts were then analyzed by using a primer extension protocol. This approach allowed mapping of the P₄ RNA start sites to the nucleotide and detection of any changes in the site of initiation which might be caused by the deletions.

The RNA start sites occurred at nt 204 and 205 (Fig. 3A), which is consistent with those previously predicted from *in vivo* (2, 14) and *in vitro* (30) data. The deletion of sequences upstream of these start sites resulted in sequential ca. 25% decreases in transcriptional activity but had no apparent effect on the site of initiation (Fig. 3A). This suggested the presence of multiple *cis*-active control elements which contribute to the level of P₄ activity. Analysis of the activity of the mutants in nuclear extracts of human HeLa cells gave very similar results. However, sequences upstream of nt 136 (-69) appear to be less important in these extracts, since their deletion reduced transcription to ca. 77% of wild-type levels (data not shown), as opposed to a twofold reduction in the murine extracts.

Analysis of *in vivo* reporter constructs. To determine whether the A9 nuclear extracts faithfully reproduced *in vivo* events, we constructed reporter plasmids by fusing the pMpC 5' deletion mutants to a eucaryotic β-galactosidase expression cassette as described in Materials and Methods. To eliminate the possibility of generating potentially inactivate or toxic fusion proteins from the mutant plasmids, the 3' fusion site of the MVM inserts was upstream of any viral ATG codons. Therefore, all translation initiation should occur at the correct site within the β-galactosidase cassette.

The P_{4β} mutants were first assayed for transcriptional activity *in vitro* in A9 extract. These mutants also initiated transcription at nucleotides 204 and 205, as observed with the pMpC constructs (data not shown) and exhibited the same relative transcriptional efficiency, despite a difference of ca. 850 bp of 3'-flanking sequence. The activity of the P_{4β} mutants was next assayed *in vivo* by using a transient-transfection assay. To control for variabilities in transfection frequency, experimental constructs were cotransfected with the reporter plasmid pSV2CAT as an internal control. Extracts of transfected cells were assayed for both β-galactosidase and CAT activity, and β-galactosidase activity was then normalized to CAT. Data from a representative experiment are shown in Fig. 4, paired with the appropriate *in vitro* data. Comparison of the *in vivo* and *in vitro* data shows that although these mutants exhibit the same general pattern of transcriptional expression, the magnitudes of some of the changes are strikingly different. Although deletion of the sequences from -205 to -145 appears to have a minimal effect in both situations, deletion from -145 to -69 results in a threefold reduction of transcription *in vivo* but only a ca. 25% decrease *in vitro*. Thus, deletion of sequences upstream of -69 reduces transcriptional activity to ca. 20% of wild-type activity *in vivo*. Furthermore, deletion of the 34-bp region from -68 to -35, which reduces transcription ca. 5-fold *in vitro*, results in a ca. 15-fold drop in activity relative to the activity of the -69 mutant and a ca. 80-fold reduction relative to wild-type activity *in vivo*. Although sequences upstream of -69 clearly play a role in modulating the transcriptional activity of the P₄ promoter, sequences be-

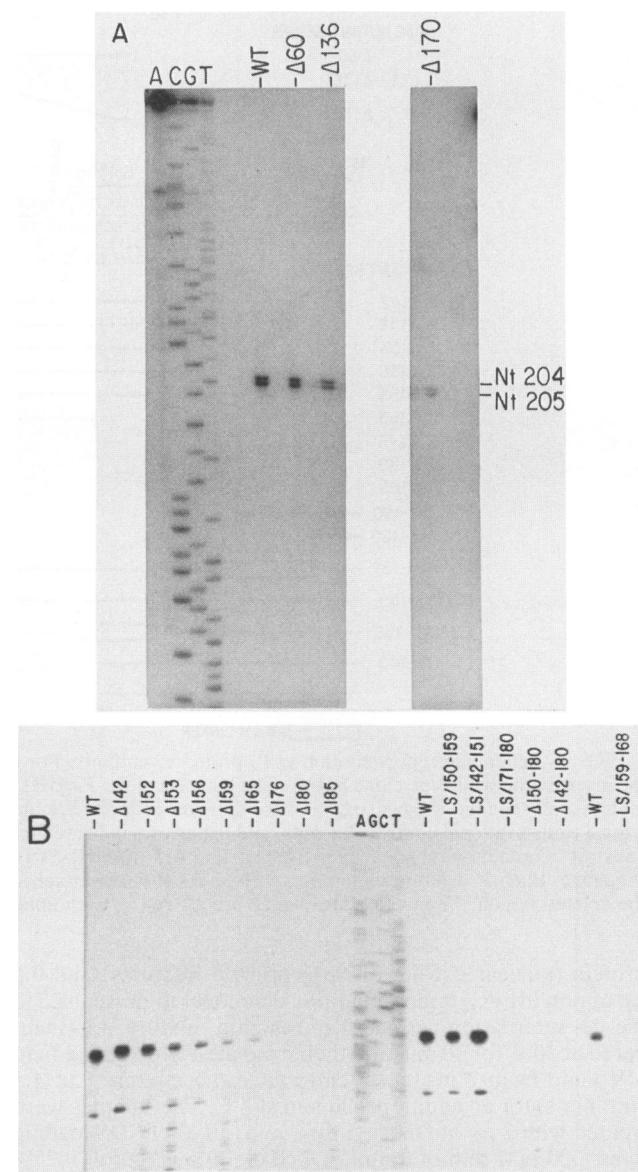


FIG. 3. Transcription of P₄ promoter mutants in A9 cell nuclear extract. (A) Transcription of pMpC mutants. Linearized pMpC mutant templates were transcribed, and the *in vitro*-generated transcripts were analyzed by using a primer extension protocol as described in Materials and Methods. The primer used for both the sequencing ladder and primer extension corresponded to nt 285 to 259 (5' to 3') of the MVM(p) sequence. The major RNA start sites are indicated. Triplicate reactions were quantitated by excising the specifically initiated bands and analyzing the incorporated radioactivity in a liquid scintillation counter. The activities of the samples were expressed as percentages of the average wild-type value. The average of the triplicate reactions are as follows: pMpC, 100%; pMpCΔ60, 73%; pMpCΔ136, 48%; pMpCΔ170, 24%. (B) Transcription of pMB, internal deletion, and linker-scanning mutant series. Transcripts generated from linearized DNAs were analyzed by using a primer extension protocol; the primer used was as described above. Reactions containing mutant LS/159-168 were from a separate experiment; thus, comparison with the other mutants must be made relative to the appropriate wild-type control.

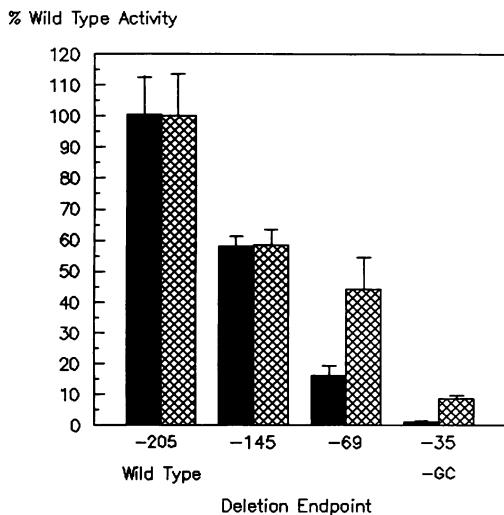


FIG. 4. Comparison of the transcriptional activity of the P₄ β mutants in vivo and in vitro. The results presented are the average of triplicate reactions (in vitro) or triplicate transfections (in vivo) expressed as percentages of wild-type (P₄ β) activity. Analysis of in vivo activity is described in Materials and Methods, and in vitro reactions were quantitated as described in the legend to Fig. 3. The actual values of each point are as follows. In vivo (■): wild type, 100%; -145 (Δ 60), 58%; -69 (Δ 136), 16%; -35 (Δ 170), 1%. In vitro (▨): wild type, 100%; -145, 58%; -69, 44%; -35, 8%.

tween -69 and -35 appear to be essential for basic promoter function both in vitro and in vivo (see below). An examination of the sequences within the 34-bp region from nt 136 to 170 (-69 to -35) revealed homologies to a number of known *cis*-acting transcriptional regulatory elements (Fig. 2). These elements were the adenovirus E1A enhancer motif (nt 153 to 162 [17]), the SV40 enhancer core (nt 164 to 170 [45]), and the GC box binding site for the HeLa cell transcription factor Sp1 (nt 159 to 168 [5]). However, of these homologies, only the GC box showed a perfect match to the published consensus sequence. We therefore chose to concentrate our initial molecular dissection of the P₄ promoter on this small but potentially complex control region.

Analysis of the pMB mutant series. To further analyze the putative *cis*-acting control elements within the region from -68 to -35, we constructed the pMB mutant series (Fig. 2). The parent plasmid for these mutants, pMB415, contained a 415-bp fragment of MVM(p) (nt 1 to 415) that had the intact P₄ promoter and 210 bp of 3'-flanking sequence. Primer extension analysis of reactions containing the wild-type construct, pMB415, showed a pattern of initiation identical to that seen for pMpC and P₄ β (compare Fig. 3A and B). Deletion up to the E1A enhancer motif of -53 (pMB Δ 152) retains the majority of wild-type transcriptional activity, whereas deletion to -46 (pMB Δ 159) results in a significant drop in activity to ca. 11% of wild-type activity. This deletion (pMB Δ 159) removes the majority of the E1A homology and the first nucleotide of the GC box. When most of the GC box sequences were deleted (to -40; pMB Δ 165), the transcriptional activity was further reduced to 7.3% of wild-type activity. It is worth noting that pMB Δ 165 showed less transcriptional activity than pMpC Δ 170, in which an additional 5 nt were deleted. This is most probably due to the different sequences which are brought proximal to the deletion endpoint in the two mutants. Finally, another mutant,

pMB Δ 176 (deletion to -29), which deletes the entire GC box and 2 nt of the TATA box, showed no detectable activity.

This analysis suggested, but did not prove, that the homologies to the GC box and the TATA box were essential elements of the P₄ promoter. It should be considered, however, that there are certain limitations when interpreting data from 5' deletion mutants, as demonstrated by the difference in activity of pMpC Δ 170 and pMB Δ 165. Sequential deletion of sequences upstream of an RNA start site allows general mapping of functional domains but makes it difficult to assign roles for specific motifs. Therefore, a series of linker-scanning and internal deletion mutants were constructed to separately address the roles of the putative functional elements identified in the P₄ promoter.

Analysis of internal deletion and linker-scanning mutants. Internal deletion mutants were constructed by joining appropriate 5' and 3' deletion mutants in the pMB series through their common Bg/II linker (Fig. 2). pMB Δ 142-180 (deletion from -63 to -25) and pMB Δ 150-180 (-55 to -25) were both totally inactive, indicating that the sequences from -55 to -25 play an essential role in P₄ transcription. This region contained the core sequences of the GC box, TATA box, E1A enhancer, and SV40 enhancer motifs. However, deletions of this type affect the spacing of the sequences adjacent to the deletion endpoints, which may affect the normal interaction of any functional elements lying upstream or downstream of the deletion site. Therefore, a more detailed analysis of these sequences was obtained by testing a number of linker-scanning mutants with mutations located from -63 to -25 (nt 142 to 180) (Fig. 3B). In linker-scanning mutants, the change of spacing and alignment of neighboring sequences are minimized by substituting 10 bp of a Bg/II linker (5'-GAAGATCTTC-3') for 10 bp of MVM sequence. LS/142-151, which substituted sequences upstream of the clustered transcription motifs, showed almost the same transcriptional activity as that of the wild type (pMB415). When most of the E1A enhancer motif and the first nucleotide of GC box sequences were substituted (LS/150-159), the transcriptional activity still remained at 84% of wild-type activity in A9 extract. However, in LS/159-168, in which 9 of 10 nt in the GC box were changed, transcriptional activity was reduced to 25% of wild-type activity. Surprisingly, LS/171-180, which altered most of the TATA box sequences, completely eliminated transcriptional activity. Thus, in A9 nuclear extracts, the functional elements within this 34-bp region of the P₄ promoter appear to be the TATA and GC motifs.

Gel retardation analysis. To examine the DNA-protein interactions within this region of the P₄ promoter, we performed gel retardation analyses (13) as described in Materials and Methods. The probe used as a 37-bp double-stranded oligonucleotide, corresponding to nt 137 to 173 of the MVM(p) genome (Fig. 5), which contains the E1A enhancer, the SV40 enhancer, and the GC box homologies. Only one of these elements, the GC box, has been shown to interact with a specific *trans*-acting protein (5).

Incubation of the P₄ probe in mouse A9 cell extracts generally resulted in the formation of multiple DNA-protein complexes, a major band with lower mobility and a number of minor faster-migrating species (Fig. 6). To determine whether these complexes were specific, we performed competition analyses with the various competitor fragments described in Fig. 5. Competition with a 10-fold molar excess of the intact P₄ promoter reduced complex formation by ca. 90%; with a 40-fold molar excess the complex was almost undetectable (Fig. 6, lanes 6 to 8). In contrast, competition

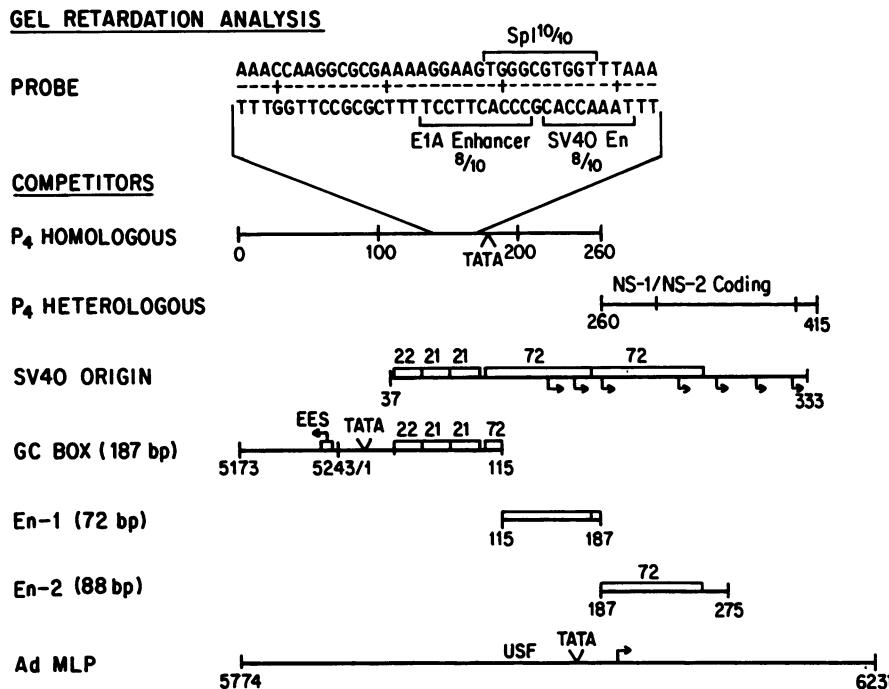


FIG. 5. Gel retardation probe and competitor DNA fragments. Preparation of the probe and competitor DNA fragments is described in Materials and Methods. Abbreviations: EES, SV40 early early start site; Ad MLP, adenovirus major late promoter.

with an equivalent amount of a MVM heterologous sequence had no apparent effect on binding (Fig. 6, lanes 10 to 12). Similar results were observed when a completely heterologous competitor, the adenovirus major late promoter (Fig. 6, AdML, lanes 2 to 4), was used. These results indicated that the P₄ probe was binding a factor(s) which interacted specifically with this 37-bp region in the intact promoter.

To determine whether the GC box was mediating binding,

we tested a fragment containing the prototypical GC box, the SV40 early promoter, for its ability to compete for P₄-binding activity. The SV40 early promoter did compete (Fig. 6, lanes 14 to 16), but, surprisingly, not as well as the homologous competitor, although the SV40 promoter fragment has six Sp1-binding sites (15). The SV40 competitor fragment also contained the two 72-bp enhancer repeats. Therefore, it was possible that the factor(s) was binding to the SV40 enhancer homology in the P₄ probe and that the observed competition was due to the 72-bp repeats. DNA fragments containing only a 72-bp repeat or the three 21-bp repeats (GC boxes) were isolated and used as competitors in gel retardation assays; only the 21-bp repeats compete with the P₄ probe for binding (Fig. 7). Previous studies had indicated that MVM(p) DNA fragments were unable to function as enhancers in vivo (P. Tattersall and E. M. Gardiner, personal communication). Therefore, we believed that the homology to the E1A enhancer core found within our probe was coincidental. Although our results with in vitro transcription (Fig. 4) were somewhat inconclusive on this point, a plasmid containing the E1A promoter-enhancer region (nt 1 to 452; L.-S. Chang, personal communication) showed no significant competition for the murine P₄ binding activity by gel retardation analysis (data not shown). These observations further supported the contention that the GC box was mediating factor binding at the P₄ promoter.

An identical series of binding competition experiments were carried out by using nuclear extracts of human HeLa cells (Fig. 8). These experiments demonstrate that binding in the HeLa cell extracts results in the formation of a single major complex similar in mobility to the major complex formed in A9 extract. Competition analyses showed that as with the A9 DNA-protein complexes, only the P₄ promoter and SV40 early promoter fragments which contain GC boxes compete. This correlates with our transcriptional results,

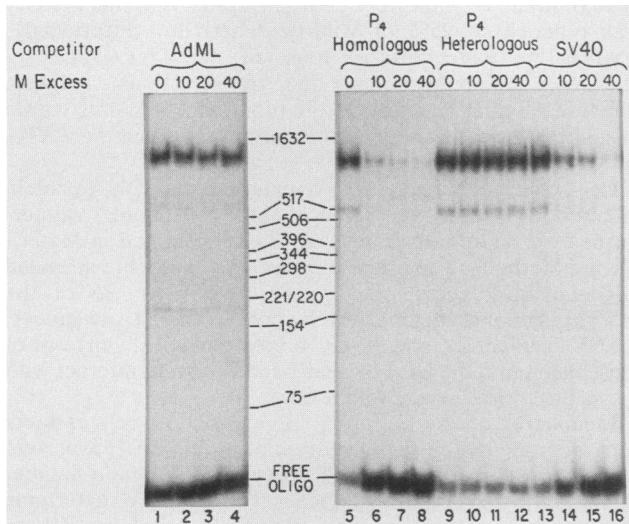


FIG. 6. Competition binding analysis. Competition binding analyses were carried out by using A9 nuclear extract as described in Materials and Methods. The probe and competitor DNA fragments are depicted in Fig. 5. The molar excess of competitor used in each reaction is indicated at the top of each lane. Markers are an end-labeled *Hinf*I digest of pBR322.

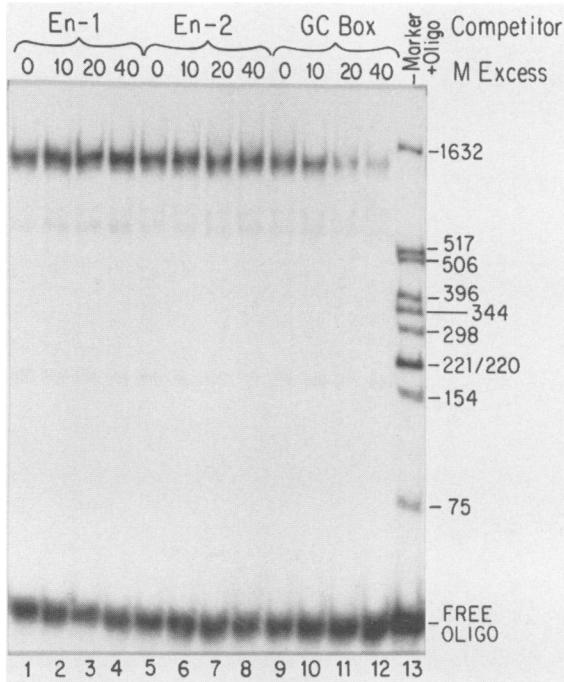


FIG. 7. The GC box sequence mediates binding to the P₄ probe. Competition binding analyses were carried out as described in Materials and Methods by using A9 nuclear extracts. The probe and competitor DNA fragments are depicted in Fig. 5. Markers are an end-labeled *Hinf*I digest of pBR322.

which showed that the P₄ promoter functioned identically in both extracts.

Mutation of the P₄ GC box affects binding. To examine the effect of mutations in and around the P₄ GC box on binding, we used the linker-scanning and internal deletion mutants (Fig. 9) as competitors in binding assays. Substitution of a 10-bp linker which eliminates most of the E1A enhancer homology and the first T of the Sp1 site (LS/150–159) significantly reduces competition relative to the wild-type competitor (Fig. 9; compare lanes 2 to 5 with lanes 6 to 8). Movement of this linker 10 bp upstream (LS/142–151) results in an increase in binding efficiency of the competitor, but competition is still somewhat reduced relative to that of the wild type (lanes 9 to 11). If the linker is inserted on the other side of the GC box, eliminating the P₄ TATA box (LS/171–180), the efficiency of competition is increased relative to LS/142–151, but the competition is still not as efficient as that of the wild-type competitor (lanes 12 to 14). Finally, the internal deletions, which remove most of the P₄ probe homology, drastically reduce competition (lanes 15 to 20).

If binding of the P₄ probe is mediated solely by the GC box, we would expect that a promoter carrying a linker substitution of this motif would be completely noncompetitive. At high concentrations, this mutant, LS/159–168, still weakly competes for the P₄ binding activity (Fig. 9, lanes 25 to 27). This is consistent with the *in vitro* transcription assays. We believe that this residual binding activity is due to a "pseudo" GC box located from -63 to -54 (nt 142 to 151). This interpretation is supported by the competitive activity of other linker scanning and internal deletion mutants. For example, Δ150–180, which eliminates all of the readily apparent transcriptional control elements, is still weakly competitive, similar to the results obtained with

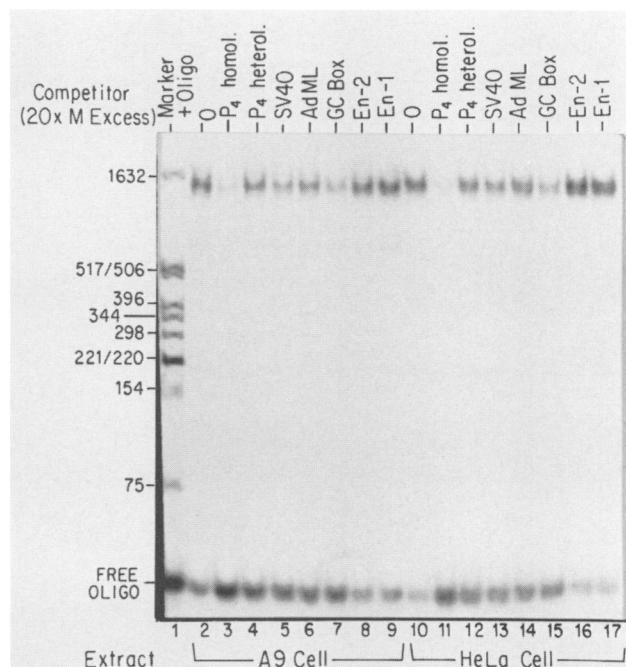


FIG. 8. Direct comparison of DNA-protein complexes formed in A9 and HeLa nuclear extracts. Competition binding analyses were carried out as described in Materials and Methods. Probe and competitor DNAs are depicted in Fig. 5. Nuclear extract used in each set of reactions is indicated. Markers are an end-labeled *Hinf*I digest of pBR322. Abbreviation: Ad ML, adenovirus major late promoter; En, enhancer.

LS/159–168. However, Δ142–180, which extends the 5' boundary of the deletion to include the pseudo box, further reduces competition. In addition, a mutant which specifically eliminates the motif, LS/142–151, also reduced binding to the competitor relative to wild-type levels.

DNase I footprint analysis. To directly identify the sequences within this region which were mediating protein binding, we carried out footprinting analyses on the wild type and linker-scanning mutants. When the wild-type promoter was used, both HeLa and A9 extracts completely protected the GC box but not the TATA box. The footprint patterns were similar but not identical (Fig. 10A and B), suggesting that a murine Sp1-like molecule(s) binds to the MVM GC box in A9 extract in a manner similar to Sp1 binding in HeLa extract. In mutant LS/159–168, which substitutes the GC box, there was no detectable footprint (Fig. 10C). However, substitution of the *Bgl*II linker did change the nuclease digestion pattern of this region, as would be expected. Thus, protein-DNA interactions in this region of the promoter appear to be mediated by the GC box sequence.

Mutant LS/171–181, which altered the TATA box sequences, had a different digestion pattern around the TATA box region owing to the linker substitution. When nuclear extract was added, a new hypersensitive site became apparent and there was only slight protection of the GC box. This suggested that the altered TATA sequences might influence protein-DNA interaction in the GC box region.

A number of common nuclease-hypersensitive sites were also observed (Fig. 10B and C). These sites mapped to non-base-pairing regions of the palindromic sequences which form the 3' hairpin structure that functions as a

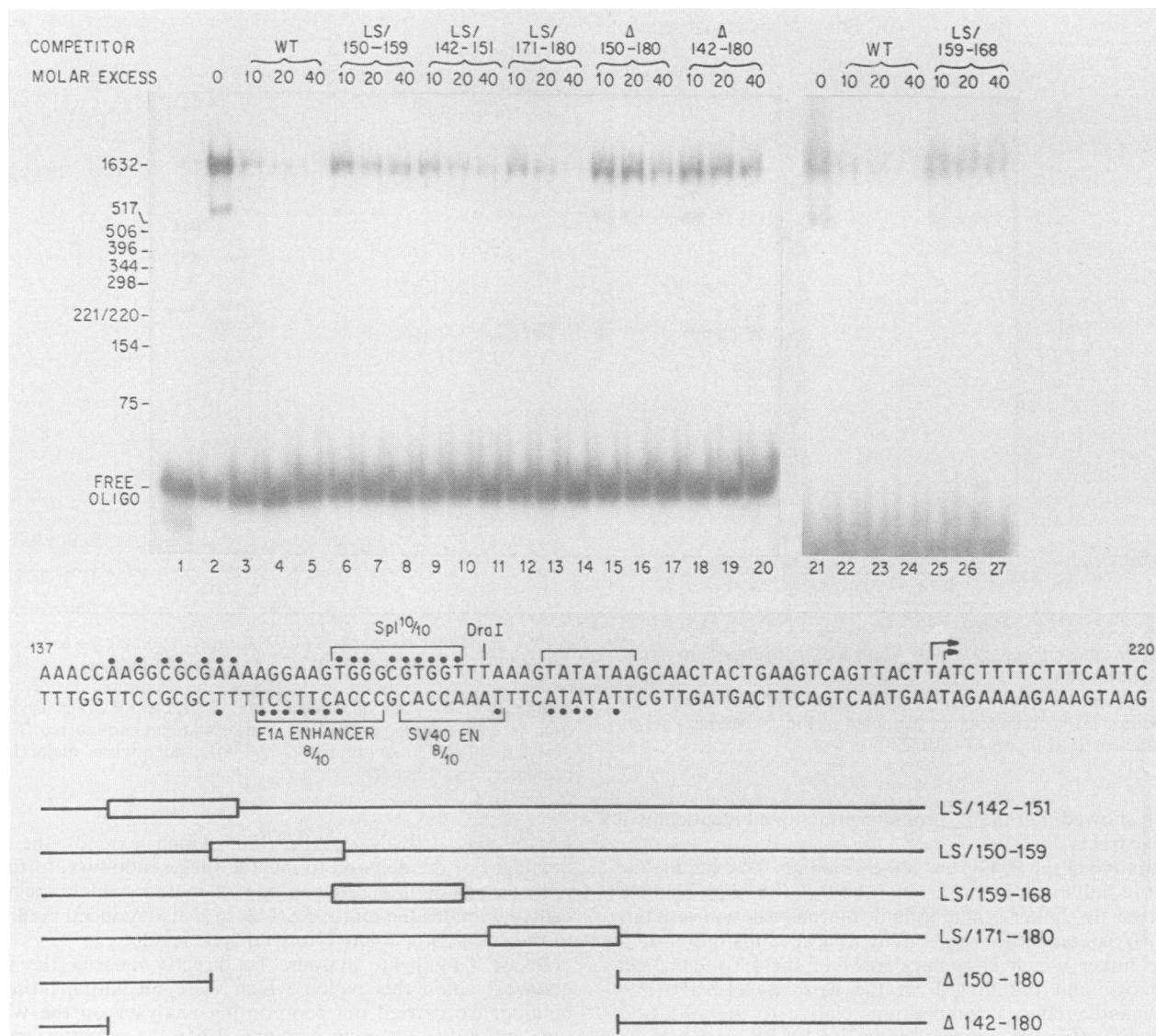


FIG. 9. Mutation of the P₄ GC box affects protein binding. The top panel shows competition binding analyses with mutant P₄ promoter containing plasmids as competitors. The 37-bp probe (nt 137 to 170) is depicted in Fig. 5, and reaction conditions are described in Materials and Methods. The lower panel depicts the sequence of the P₄ promoter surrounding the RNA start sites and the structures of mutant promoters used as competitors. Boxes on the sequence indicate homologies to known *cis*-acting control elements; the arrows indicate the RNA start sites. All competitors contain viral sequences from nt 1 to 415, except as indicated. The boxes in the linker-scanning mutants represent a 10-bp *Bgl*II linker which has been substituted for the viral sequence. Nucleotide changes generated by this substitution are represented by dots on the DNA sequence.

replication origin (34). This suggests that specific sequences within the palindrome are recognized by factors present in A9 nuclear extracts and that these may reflect interactions which are involved in viral DNA replication.

Identification of P₄-binding protein(s) by UV-photocross-linking. To identify the sequence-specific proteins binding to the MVM P₄ GC box, we used UV-photocrosslinking experiments. It has been previously shown that UV irradiation can induce stable photoadducts between the amino group of a protein and the base of a nucleic acid (35). If this photoadduct is formed with a ³²P-labeled oligoduplex, analysis of the reaction products by SDS-polyacrylamide gel electrophoresis followed by autoradiography should identify the protein(s) in a cell extract which is cross-linked to the DNA. Furthermore, if only one protein-binding motif in the oligo is

internally radiolabeled and the photoadduct is treated with DNase I following irradiation, the only proteins identified should be those which interact with the radiolabeled motif. To test the validity of these premises, we first applied this method to the identification of the well-characterized HeLa cell transcription factor, Sp1.

A 42-bp oligonucleotide duplex representing two copies of the SV40 21-bp repeat was synthesized as described in Materials and Methods. To confirm that this oligomer, designated duplex 1, was capable of interacting with Sp1, we tested it as a competitor for the binding of Sp1 to the SV40 21-bp repeats. Figure 11A (lanes 1 to 4) shows DNase I footprinting of the SV40 early promoter in the presence of 0-, 5-, 10-, and 20-fold molar excess, respectively, of duplex 1. As can be seen by comparison with lane 5, which shows the

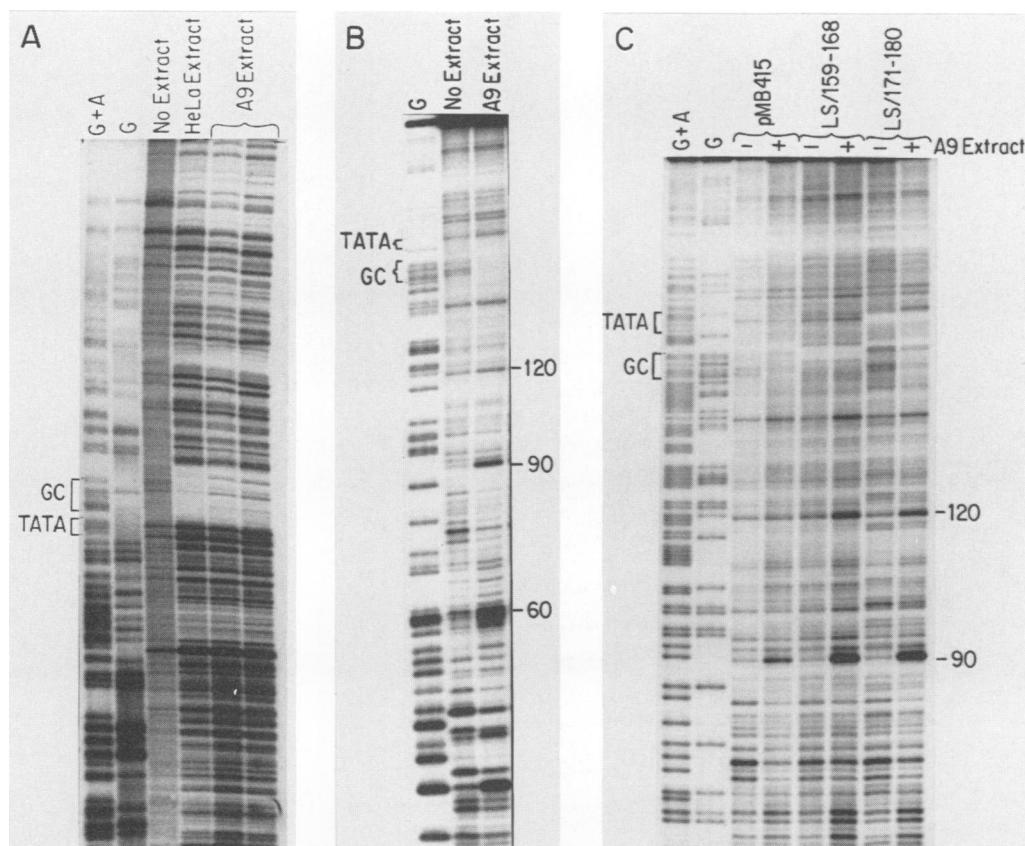


FIG. 10. DNase I footprint analysis of wild-type and mutant P₄ promoters. Preparation of probe and reaction conditions are described in Materials and Methods. (A and B) Footprinting of the wild-type P₄ promoter labeled on the coding and noncoding strands, respectively. (C) Footprinting of the noncoding strand of the wild-type promoter and two linker-scanning mutants. LS/159-168 disrupts the GC box, and LS/171-180 disrupts the TATA box. The relative positions of the GC and TATA motifs are indicated. Nuclease-hypersensitive sites which appear on the noncoding strand are indicated in panels B and C.

pattern of DNase I cleavage in the absence of any cell extract, increasing molar excess of duplex 1 reduce the footprinting of the 21-bp repeat region. These competition experiments, as well as gel retardation analyses (data not shown), clearly indicate that duplex 1 does interact with Sp1 in a sequence-specific manner and could be used as an appropriate DNA segment for the photoidentification of Sp1 or related DNA-binding proteins.

Duplex 1 was internally labeled such that only one GC box was radioactive (see Materials and Methods). UV cross-linking of this duplex in HeLa extract should result in the formation of a radioactive photoadduct with only one molecule of Sp1. Photocrosslinking of duplex 1 in crude HeLa cell extract revealed a number of cross-linked bands, including a photoadduct of ca. 95 kDa (Fig. 11B, lane 1), which is similar to the molecular mass of Sp1 as reported by Briggs et al. (5). Competition experiments (data not shown) suggested that the only specific interaction was the ca. 95-kDa cross-linked product. To confirm that this photoadduct resulted from cross-linking of duplex 1 with Sp1, we fractionated HeLa nuclear extract by the method of Dynan and Tjian (11) without S300 chromatography to yield Sp1-enriched and Sp1-devoid fractions. Upon photocrosslinking of these fractions with duplex 1, only the Sp1-enriched fraction resulted in the formation of a 95-kDa photoadduct (Fig. 11B, lane 3). Furthermore, competition analysis (data not shown) revealed that cross-linking of the 95-kDa protein is specifically

reduced with increasing amounts of a 1.5-kb DNA fragment containing the SV40 early promoter.

To identify the proteins which interact with the P₄ GC box, we constructed an internally labeled oligonucleotide duplex, duplex 2, corresponding to nt 137 to 173 (-68 to -32) of MVM(p) (see Materials and Methods). UV photocrosslinking of duplex 2 in HeLa cell extract resulted in the formation of a 95-kDa photoadduct similar to that seen with duplex 1 (compare Fig. 11B and 12A). The formation of this complex was not affected by the presence of a 42-bp oligo duplex which contains the binding site for the HeLa cell transcription factor USF (37) (Fig. 12A, lane 3). However, it was inhibited by the presence of excess homologous duplex (Fig. 12A, lane 2) or excess duplex 1, which carries in its framework four Sp1-binding sites (Fig. 12A, lane 4). Therefore, the ca. 95-kDa complex most probably results from the cross-linking of duplex 2 and Sp1. It is interesting that although duplex 1 has more Sp1-binding sites than duplex 2, the latter always acted as a better competitor, similar to the results obtained with gel retardation.

Duplex 2 was then UV cross-linked with nuclear extract from mouse A9 cells. In contrast to HeLa extracts, this resulted in the formation of several photoadducts (Fig. 12B). Cross-linking experiments in the presence of various cold competitor duplexes revealed that only the ca. 95- and ca. 120-kDa photoadducts were reduced by using the homologous competitor. The 42-bp duplex representing the SV40

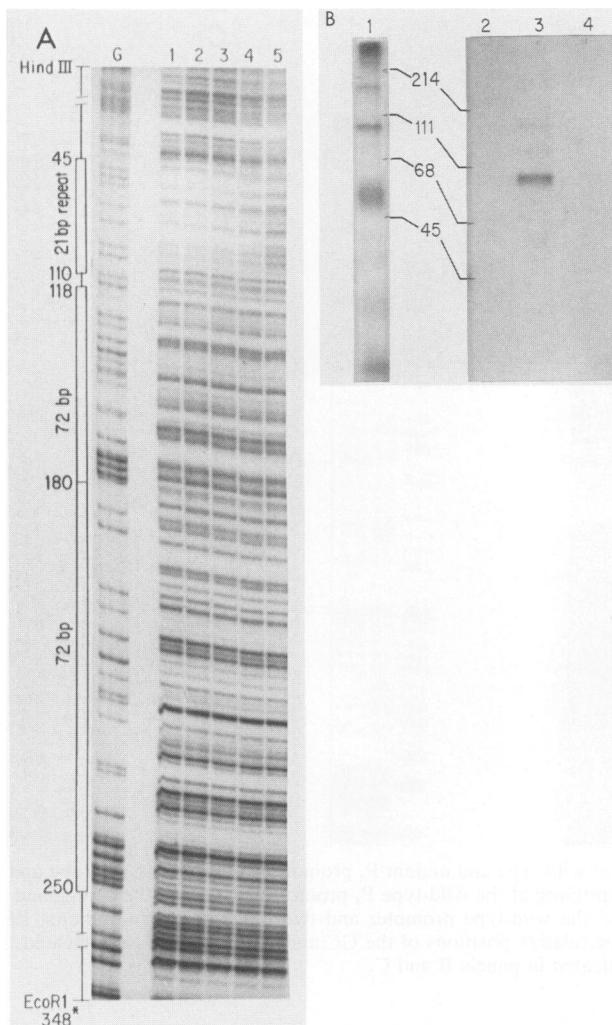


FIG. 11. Binding and UV cross-linking of duplex 1 to Sp1 from HeLa cells. (A) DNase I footprinting of the coding strand of the SV40 origin region in the presence of an increasing molar excess of cold duplex 1. Preparation of probe and reaction conditions are described in Materials and Methods. Lanes: G, Maxam-Gilbert sequencing G reaction; 1, no cold duplex; 2, 5× molar excess; 3, 10× molar excess; 4, 20× molar excess; 5, no protein. (B) UV photocrosslinking of labeled duplex 1 with crude and chromatographically fractionated HeLa nuclear extract. Lanes: 1, crude nuclear extract; 2, phosphocellulose flowthrough fraction; 3, phosphocellulose fraction eluted with 0.35 M KCl; 4, phosphocellulose fraction eluted with 0.6 M KCl. For details of cross-linking procedures and chromatography, see Materials and Methods.

21-bp repeats did inhibit cross-linking (lane 4), but not as well as the homologous duplex (lane 2), even though it had four GC boxes. Thus, in murine cells the P₄ GC box binds both a 95-kDa protein, similar in size to the human Sp1, and a 120-kDa protein, which may represent a modified form of the 95-kDa protein (see Discussion) or a novel GC box-binding protein.

DISCUSSION

By analyzing the transcriptional activities of mutants of the MVM P₄ promoter, we have established that a 30-bp region from -55 to -25 (nt 150 to 180) relative to the major transcription start site contains DNA elements essential for promoter activity in vitro (Fig. 3A and 4). The functional sequence motifs within this region appear to be a TATA box and a GC box. These elements also play essential roles in vivo (Fig. 3B); however, optimal in vivo transcription requires additional sequences upstream of these motifs. Iden-

tification and characterization of the upstream transcriptional control elements are presently under way.

The GC box sequence in the MVM P₄ promoter is a 10-of-10 match (5'-TGGGCGTGTT-3') to the consensus-binding site for the human transcription factor Sp1 (5). According to the reported relative Sp1-binding affinity for various GC box consensus sequences, the P₄ GC box should be a weak binding site for Sp1 (24). Gel retardation (Fig. 6 to 9) and DNase I footprinting (Fig. 10) analyses demonstrated that a cellular *trans*-acting factor(s) interacts with this motif in both human and murine nuclear extracts. Interestingly, this interaction is of very high affinity, being stable in the presence of 10 µg (10⁵-fold mass excess) of nonspecific DNA poly(dI-dC) (data not shown). Furthermore, competition analyses (Fig. 6 to 8) have demonstrated that DNA fragments containing the P₄ GC box compete more efficiently for the protein which binds this motif than the SV40 21-bp repeats which contain six GC boxes. A 10-fold molar excess

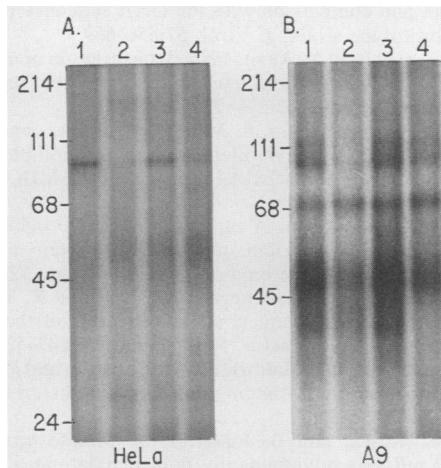


FIG. 12. Photocrosslinking of duplex 2 in HeLa (A) and A9 (B) nuclear extracts. Lanes: 1, no competitor; 2, 40× molar excess of cold probe duplex; 3, 40× molar excess of a synthetic adenovirus major late promoter USF binding site; 4, 40× molar excess of duplex 1.

of a wild-type P₄ promoter fragment reduced binding by ca. 90%, whereas a similar molar excess of a SV40 21-bp repeat fragment resulted in only ca. 50% reduction of the bound complex. This difference in apparent affinity becomes quite significant when one considers that five of the six SV40 GC boxes can be occupied at once. Therefore, the molar excess of SV40 competitor in these reactions in terms of GC box equivalents is 50-, 100-, and 200-fold. Thus, it appears that the SV40 fragment exhibits at least a 10-fold-lower affinity for the P₄ binding factor(s) in both extracts (compare Fig. 2, lanes 6 to 8 with lanes 14 to 16 and Fig. 4, lanes 3 and 5 with 11 and 13). Taken together with our transcriptional data, this suggests that a murine Sp1-like molecule(s) activates P₄ transcription by binding to the GC box with extremely high affinity.

Compared with other Sp1-responsive promoters, P₄ was less dramatically affected by mutation of its GC box (24). LS/159–168, which completely changed the GC box sequence (5'-TGGCGTGTT-3') by replacing it with a *Bgl*II linker (5'-GAAGATCTTC-3'), reduced P₄ in vitro transcription to ca. 25% of wild-type levels in A9 nuclear extract. As expected, DNase I footprint analysis of this mutant showed no protection of the GC region in either extract. Interestingly, there is another sequence just upstream of the apparently functional GC box which also showed homology to this motif. We initially did not believe that this sequence, 5'-AAGCGCGAA-3' (-63 to -54), was functional, because of its extreme divergence from the published consensus sequence 5'-^{GG}_{TA}GGCG^{GGGC}_{TAAT}-3'. In addition, mutant LS/142–151, which eliminates this sequence, showed about the same transcriptional activity as the wild-type did. DNase I footprinting also showed no protection of this region in either extract, suggesting that this motif plays a minimal role, if any, in vitro. However, gel retardation competition analyses (Fig. 9) suggest a weak interaction in this region with the same protein(s) which binds to the P₄ GC box. Therefore, we believe that in the absence of the high-affinity binding site, the pseudo GC box may be functional, thus explaining the 40% transcriptional activity of LS/159–168. An alternative but less likely explanation for these observations is outlined below.

Overlapping the pseudo GC box is a binding site for the cellular transcription factor E2F, 5'-TTTCGCGC-3' (-53 to -60). E2F has been shown to be induced by the adenovirus E1A protein (47) and is involved in the control of transcription of a number of viral early protein genes. However, in cells which do not express E1A, E2F is present at almost undetectable levels. In addition, mutant LS/142–151, which substitutes this region, shows essentially wild-type transcriptional activity. Therefore, we do not believe that this motif is functional. Experiments are currently under way to resolve this question.

In the in vitro system, the TATA sequence appears to be the essential transcriptional control element for the P₄ promoter. LS/171–180, which destroys the TATA homology, was completely devoid of transcriptional activity. Despite its critical role in transcriptional activation of P₄, DNase I footprinting showed no protection of this region when our crude nuclear extracts were used. This is not surprising in light of a recent report (29) on the purification of the human TATA box-binding protein TFIID. These investigators have demonstrated that the relatively weak binding affinity of TFIID makes it impossible to see a footprint unless partially purified fractions are used.

The proximity of the GC box and TATA box and the requirement of both motifs for transcriptional activity suggests that the GC box-binding protein(s) might interact with the TATA box-binding protein(s) to stimulate activity of the P₄ promoter. Indeed, mutant LS/171–180, which destroys the TATA homology, also affects binding to the GC box (Fig. 9), consistent with reports (31, 41) that protein-protein interactions are critical for efficient transcriptional control. Furthermore, transcription of the SV40 early start sites requires stereospecific alignment of the GC and TATA boxes (41), suggesting a need for interaction between the GC box (Sp1)- and TATA box (TFIID)-binding proteins. Using our mutant panel, we can now begin to address this question in P₄ by altering the spatial orientation of the two binding sites with respect to each other.

We did not expect the observed difference in binding affinity between the SV40 promoter and the P₄ promoter for a number of reasons. Prior to this report, the SV40 21-bp repeats had been defined as the prototypical high-affinity Sp1-binding sequence. Comparison of a large number of viral and cellular GC boxes had demonstrated a number of high-affinity sequences which are not contained in the SV40 21-bp repeats. However, none had shown greater affinity for Sp1 than the high-affinity SV40 GC boxes. There are a number of possibilities that could account for the difference in binding affinity seen for the two sequences. First, it is possible that we have identified an extremely high-affinity Sp1-binding sequence. We believe that this possibility is somewhat unlikely, because of all the GC boxes examined for binding affinity, the closest to the MVM sequence, 5'-TGGCGGGGT-3' from herpes simplex virus thymidine kinase (24), shows only medium affinity for Sp1. In addition, the P₄ GC box contains a G-to-T transversion at position 7 of the 10-bp consensus sequence. This position has been shown to be protected from dimethyl sulfate methylation by Sp1 and is therefore assumed to be involved in sequence recognition by the protein. However, an Sp1 site found in the human immunodeficiency virus long terminal repeat (21) contains this mutation and a G-to-A transition at position 2 but still binds Sp1 with high affinity. In addition, the identical GC box sequence has been shown to be essential for control of the adenovirus EIIA-late promoter and is protected from

DNase I digestion by a partially purified preparation of HeLa Sp1 (3).

A second possible explanation is that the P₄ GC box interacts with a modified form of Sp1 which is different from that used by SV40. It is interesting that both the P₄ and SV40 probes cross-link a 95-kDa peptide when incubated with HeLa extracts (Fig. 11 and 12). This correlates with the reported molecular mass of one of the affinity-purified Sp1 polypeptides (5). However, the same investigators (20) reported that high-resolution analysis of their affinity-purified material revealed microheterogeneity of both the 95- and 105-kDa species; they attribute this to posttranslational modification. Tjian and colleagues have demonstrated that Sp1 is modified by O-linked glycosylation (20), and that analysis of Sp1 cDNAs has revealed possible sites for phosphorylation and N-linked glycosylation (22). The O-linked sugars do not appear to affect DNA binding in any way, but do play a role in regulating the ability of Sp1 to activate transcription. It is possible that other modifications could be used to regulate the affinity of Sp1 for different sequences. For example, modification of binding activity by phosphorylation has been reported for both polymerase III (19) and polymerase II (40, 46) transcription factors.

A third possibility is that the protein(s) bound by the P₄ GC box is not the prototypical Sp1. It has been reported that low-stringency Southern blots of fragments of the Sp1 cDNA reveal multiple cross-hybridizing bands, suggesting the existence of a family of related genes (23). This is not surprising, given the ubiquitous nature of the GC box as a transcriptional control element. One could imagine that this element plays a very basic role in polymerase II transcriptional control and that only a single *trans*-acting factor would therefore be necessary for function. However, a more dynamic role for this motif would require not only the diversity generated by modification of a single binding protein but also interaction with multiple, mutually exclusive factors which could mediate different effects. The interaction of multiple regulatory proteins with closely related sequences is not without precedent. For example, a *cis*-active regulatory sequence, also thought to play a relatively basic role in transcriptional control, the CCAAT box, has now been shown to interact with at least four different *trans*-acting factors (6, 25, 36). In the light of these observations, it is intriguing that the P₄ probe cross-linked a novel 120-kDa peptide in mouse cell extracts. It is possible that this is homologous to the 105-kDa Sp1 peptide seen in HeLa cell extracts and that, together with the common 95-kDa peptide, these are the mouse equivalent of human Sp1. However, given the preferential affinity for the P₄ GC box, we believe that it is just as likely that these peptides are closely related but not identical to Sp1. Before it can be found whether these proteins are the mouse equivalent of human Sp1, they must be purified; such studies are currently under way.

ACKNOWLEDGMENTS

Plasmids R1-Bam-e-4 and pΔΔ were gifts from the laboratories of S. Weissman and J. Steitz, respectively. We thank Bill Morgan and Zach Pitluk for sharing their ideas and technical expertise.

This work was supported by Public Health Service grants R01-AI19973 (to D.C.W.) and R29-GM38228 (to G.K.) from the National Institutes of Health. B.J.G. was supported by National Research Service Award 5 T32 GM07499.

LITERATURE CITED

1. Astell, C. R., E. M. Gardiner, and P. Tattersall. 1986. DNA sequence of the lymphotropic variant of minute virus of mice,

MVM(i), and comparison with the DNA sequence of the fibrotropic prototype strain. *J. Virol.* **57**:656–669.

2. Ben-Asher, E., and Y. Aloni. 1984. Transcription of minute virus of mice, an autonomous parvovirus, may be regulated by attenuation. *J. Virol.* **52**:266–276.
3. Bhat, G., L. SivaRaman, S. Murthy, P. Domer, and B. Thimmappa. 1987. In vivo identification of multiple promoter domains of adenovirus EIIA-late promoter. *EMBO J.* **6**:2045–2052.
4. Bradford, M. M. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**:248–254.
5. Briggs, M. R., J. T. Kadonaga, S. P. Bell, and R. Tjian. 1986. Purification and biochemical characterization of the promoter-specific transcription factor, Sp1. *Science* **234**:47–52.
6. Chodosh, L. A., A. S. Baldwin, R. C. Carthew, and P. A. Sharp. 1988. Human CCAAT-binding proteins have heterologous subunits. *Cell* **53**:11–24.
7. Clemens, K. E., and D. Pintel. 1987. Minute virus of mice (MVM) mRNAs predominantly polyadenylate at a single site. *Virology* **160**:511–514.
8. Clemens, K. E., and D. J. Pintel. 1988. The two transcription units of the autonomous parvovirus minute virus of mice are transcribed in a temporal order. *J. Virol.* **62**:1448–1451.
9. Cotmore, S. F., and P. Tattersall. 1987. The autonomously replicating parvoviruses of vertebrates. *Adv. Virus Res.* **33**:91–173.
10. Doerig, C., B. Hirt, P. Beard, and J.-P. Antonietti. 1988. Minute virus of mice non-structural protein NS-1 is necessary and sufficient for trans-activation of the viral P₃₉ promoter. *J. Gen. Virol.* **69**:2563–2573.
11. Dynan, W. S., and R. Tjian. 1983. Isolation of transcription factors that discriminate between different promoters recognized by RNA polymerase II. *Cell* **32**:669–680.
12. Engers, H. D., J. A. Louis, R. H. Zuber, and B. Hirt. 1981. Inhibition of T-cell mediated functions by MVM(i), a parvovirus closely related to minute virus of mouse. *J. Immunol.* **127**:2280–2285.
13. Fried, M. G., and D. M. Crothers. 1981. Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis. *Nucleic Acids Res.* **9**:6505–6526.
14. Gardiner, E. M., and P. Tattersall. 1988. Evidence that developmentally regulated control of gene expression by a parvoviral allotropic determinant is particle mediated. *J. Virol.* **62**:1713–1722.
15. Gidoni, D., J. T. Kadonaga, H. Barrera-Saldana, K. Takahashi, P. Chambon, and R. Tjian. 1985. Bidirectional SV40 transcription mediated by tandem Sp1 binding interactions. *Science* **230**:511–517.
16. Gorman, C. M., L. F. Moffat, and B. H. Howard. 1982. Recombinant genomes which express chloramphenicol acetyltransferase in mammalian cells. *Mol. Cell. Biol.* **2**:1044–1051.
17. Hearing, P., and T. Shenk. 1983. The adenovirus type 5 E1A transcriptional control region contains a duplicated enhancer element. *Cell* **33**:695–703.
18. Heintz, N., and R. G. Roeder. 1984. Transcription of human histone genes in extracts from synchronized HeLa cells. *Proc. Natl. Acad. Sci. USA* **81**:2713–2717.
19. Hoeffer, W. K., R. Kovelman, and R. G. Roeder. 1988. Activation of transcription factor IIIC by the adenovirus E1A protein. *Cell* **53**:907–920.
20. Jackson, S. P., and R. Tjian. 1988. O-glycosylation of eukaryotic transcription factors: implications for mechanisms of transcriptional regulation. *Cell* **55**:125–133.
21. Jones, K. A., J. T. Kadonaga, P. A. Luciw, and R. Tjian. 1986. Activation of the AIDS retrovirus promoter by the cellular transcription factor, Sp1. *Science* **232**:755–759.
22. Kadonaga, J. T., K. C. Carner, F. R. Masiarz, and R. Tjian. 1987. Isolation of cDNA encoding transcription factor Sp1 and functional analysis of the DNA binding domain. *Cell* **51**:1079–1090.
23. Kadonaga, J. T., A. J. Courey, J. Ladika, and R. Tjian. 1988. Distinct regions of Sp1 modulate DNA binding and transcrip-

- tional activation. *Science* **242**:1566–1570.
24. Kadonaga, J. T., K. A. Jones, and R. Tjian. 1986. Promoter-specific activation of RNA polymerase II transcription by Sp1. *Trends Biochem. Sci.* **11**:20–23.
 25. Landschulz, W. H., P. E. Johnson, E. Y. Adashi, B. J. Graves, and S. L. McKnight. 1988. Isolation of a recombinant copy of the gene encoding C/EBP. *Genes Dev.* **2**:786–800.
 26. Maxam, A. M., and W. Gilbert. 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods Enzymol.* **65**:499–560.
 27. Merchlinsky, M. J., P. J. Tattersall, J. J. Leary, S. F. Cotmore, E. M. Gardiner, and D. C. Ward. 1983. Construction of an infectious molecular clone of the autonomous parvovirus minute virus of mice. *J. Virol.* **47**:227–232.
 28. Morgan, W. R., and D. C. Ward. 1986. Three splicing patterns are used to excise the small intron common to all minute virus of mice RNAs. *J. Virol.* **60**:1170–1174.
 29. Nakajima, N., M. Horikoshi, and R. G. Roeder. 1988. Factors involved in specific transcription by mammalian RNA polymerase II: purification, genetic specificity and TATA box-promoter interactions of TFIID. *Mol. Cell. Biol.* **8**:4028–4040.
 30. Pintel, D., D. Dadachanji, C. R. Astell, and D. C. Ward. 1983. The genome of minute virus of mice, an autonomous parvovirus, encodes two overlapping transcription units. *Nucleic Acids Res.* **11**:1019–1038.
 31. Ptashne, M. 1988. How eukaryotic transcription activators work. *Nature (London)* **335**:683–689.
 32. Rhode, S. L., III. 1985. *trans*-Activation of parvovirus P38 promoter by the 76K noncapsid protein. *J. Virol.* **55**:886–889.
 33. Rhode, S. L., III, and S. M. Richard. 1987. Characterization of the *trans*-activation-responsive element of the parvovirus H-1 P38 promoter. *J. Virol.* **61**:2807–2815.
 34. Sahli, R., G. K. McMaster, and B. Hirt. 1985. DNA sequence between two tissue-specific variants of the autonomous parvovirus, minute virus of mice. *Nucleic Acids Res.* **13**:3617–3633.
 35. Saito, I., and T. Matuura. 1985. Chemical aspects of UV induced crosslinking of proteins to nucleic acids: photoreactions with lysine and tryptophan. *Chem. Res.* **18**:134–141.
 36. Santoro, C., N. Mermod, P. C. Andrews, and R. Tjian. 1988. A family of human CCAAT-box-binding proteins active in transcription and DNA replication: cloning and expression of multiple cDNAs. *Nature (London)* **334**:218–224.
 37. Sawadogo, M., and R. G. Roeder. 1985. Interaction of a gene-specific transcription factor with the adenovirus major late promoter upstream of the TATA box region. *Cell* **43**:165–175.
 38. Settleman, J., and D. DiMaio. 1988. Efficient transactivation and morphologic transformation by bovine papilloma genes expressed from bovine papillomavirus/simian virus 40 recombinant virus. *Proc. Natl. Acad. Sci. USA* **85**:9007–9011.
 39. Siegl, G. 1984. Biology and pathogenicity of autonomous parvoviruses, p. 297–362. *In* K. I. Berns (ed.), *The parvoviruses*. Plenum Publishing Corp., New York.
 40. Sorger, P. K., M. J. Lewis, and H. R. B. Pelham. 1987. Heat shock factor is regulated differently in yeast and HeLa cells. *Nature (London)* **329**:81–84.
 41. Takahashi, K., M. Vigneron, H. Matthes, A. Wildeman, M. Zenke, and P. Chambon. 1986. Requirement of stereospecific alignments for initiation from the simian virus 40 early promoter. *Nature (London)* **319**:121–126.
 42. Tattersall, P. 1972. Replication of the parvovirus MVM. I. Dependence of virus multiplication and plaque formation on cell growth. *J. Virol.* **10**:586–590.
 43. Tattersall, P. 1978. Susceptibility to minute virus of mice as a function of host cell differentiation, p. 131–149. *In* D. C. Ward and P. Tattersall (ed.), *Replication of mammalian parvoviruses*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
 44. Tattersall, P., and J. Bratton. 1983. Reciprocal productive and restrictive virus-cell interactions of immunosuppressive and prototype strains of minute virus of mice. *J. Virol.* **46**:944–955.
 45. Weiher, H., M. Konig, and P. Gruss. 1983. Multiple point mutations affecting the simian virus 40 enhancer. *Science* **219**: 626–631.
 46. Yamamoto, K. K., G. A. Gonzalez, W. H. Biggs III, and M. R. Montminy. 1988. Phosphorylation-induced binding and transcriptional efficacy of nuclear factor CREB. *Nature (London)* **334**:494–498.
 47. Yee, A. S., R. Reichel, I. Kovacs, and J. R. Nevins. 1987. Promoter interaction of the E1A-inducible factor E2F and its potential role in the formation of a multi-component complex. *EMBO J.* **6**:2061–2068.