JOURNAL OF VIROLOGY, Dec. 1991, p. 6661–6670
0022-538X/91/126661-10$02.00/0

Vol. 65, No. 12

Unusual Sp1-GC Box Interaction in a Parvovirus Promoter

ZACHARY W. PITLUK[1] AND DAVID C. WARD[1,2]*

Departments of Molecular Biophysics and Biochemistry[1] and Genetics,[2] Yale University School of Medicine, 333 Cedar Street, New Haven, Connecticut 06510-8005

Received 24 May 1991/Accepted 30 August 1991

The P_4 promoter of the parvovirus minute virus of mice contains a single degenerate GC box sequence which binds the transcription factor Sp1 with high affinity. The two protomers of murine Sp1 were affinity purified, and their interactions with the P_4 promoter were examined. Several unusual features were observed. Methylation interference experiments demonstrated that Sp1 makes contacts with both DNA strands, including the central guanine as well as an adenine residue on the cytidine-rich strand of the GC box. UV photocross-linking revealed that the 95- and the 105-kDa protomers cross-link exclusively to opposite strands of the GC box. These results suggest that the phosphorylation of the 95-kDa Sp1 protomer results in a change in the way Sp1 is positioned on the P_4 GC box and identifies a high-affinity GC box motif.

The transcription factor Sp1 regulates the basal level of expression of many cellular and viral genes (22). Currently, it is thought that either the 95- or the 105-kDa protomer of Sp1 binds to the 10-nucleotide DNA sequence element, the GC box, by specifically contacting guanine residues within the element. The apparent molecular weight difference between the two protomers is due to the phosphorylation state (15). A GC box consensus sequence has been derived and divided into high- and low-affinity sequences on the basis of competition studies for Sp1 binding (20). Both protomers show equal and high affinity for the GC box sequences in the 21-bp repeats of simian virus 40 (SV40) despite the difference in phosphorylation state between the two protomers (15). The Sp1-DNA interaction usually involves the contacting of guanine residues only on the guanine-rich strand. One exception to this general rule is the human immunodeficiency virus long terminal repeat promoter's GC boxes I and III in which guanine residues on the cytidine-rich strand of the GC box were shown by methylation protection experiments to be contacted by Sp1 (18).

The P_4 promoter of the parvovirus minute virus of mice [MVM(p)] is also activated by Sp1 in an apparently unusual manner. In vitro, the P_4 promoter is a strong promoter, yet it was shown by deletion mutagenesis to be primarily activated by a single nonconsensus GC box and a consensus TATA element (1). Moreover, in crude nuclear extracts, the MVM P_4 GC box was also found to be as strong a competitor for Sp1 binding as the high-affinity GC boxes in the 21-bp repeats of SV40. The consensus sequence which ranks the strength of GC boxes failed to predict the strength of the P_4 GC box. To define the Sp1-P_4 GC box interaction further, we have purified murine Sp1 (mSp1) from A92L nuclear extracts and analyzed the molecular basis of the high-affinity interaction between mSp1 and the MVM(p) P_4 GC box. Our results indicate that Sp1 binds to the P_4 promoter in an unusual manner. These observations led us to reanalyze the composition of GC box sequences found in promoters to elucidate conserved sequence motifs in the GC box element which might influence Sp1 binding affinity.

* Corresponding author.

MATERIALS AND METHODS

Plasmids and oligonucleotides. The plasmid pMB415 was made by ligating the *Bam*HI-*Pst*I P_4 promoter-containing fragment from MVM(p) into the Bluescript KS(+) vector. The plasmid pP4− was constructed by ligating the *Bam*HI-*Pst*I P_4 promoter-containing fragment into the BS KS (−) vector. The SV40 early promoter was derived from the plasmid pSV2XCAT which has an *Xba*I linker placed in the *Acc*I site in pSV2CAT.

Chromatography resins. WGA sepharose was prepared by coupling wheat germ agglutinin (WGA) (Sigma) to commercially available CNBr-activated Sepharose CL-4B (PL Biochemicals) according to the manufacturer's instructions. The final amount of WGA coupled was 5 mg/ml of resin. Oligonucleotide Sepharose was prepared by coupling the oligonucleotide duplex BG 1/2 (containing the sequence from nucleotide [nt] 137 to nt 173 with three non-base-paired adenines at its left end [Fig. 1A]) to CNBr-activated Sepharose CL-4B according to the protocol of Arndt-Jovin et al. (3). The final concentration of oligonucleotide coupled was 100 μg/ml of resin.

Cell culture and nuclear extracts. Murine A92L fibroblasts were grown in spinner flasks in Jokliks Dulbecco modified Eagle medium (GIBCO-BRL) supplemented with 5% fetal calf serum. The cells were harvested at a density of 5×10^5, and nuclear extracts were prepared according to the method of Dignam (1) except for the following modifications. All buffers contained the following protease inhibitors: phenylmethylsulfonyl fluoride, metabisulfite, and benzamidine at 0.5 mM; EDTA and EGTA [ethylene glycol-bis(β-aminoethyl ether)-*N*,*N*,*N*′,*N*′-tetraacetic acid] at 0.1 mM; and aprotinin, leupeptin, and pepstatin at 50 μg/ml]. The cells were lysed in hypotonic buffer (20 mM HEPES [*N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid; pH 7.9], 10 mM KCl, 1.5 mM $MgCl_2$, 20% glycerol) that also contained 1% (vol/vol) Nonidet P-40 (NP-40; Calbiochem) and 2% (wt/vol) polyvinyl alcohol (PVA) (Fluka) to stabilize the nuclei. The nuclei were pelleted at 10,000 × *g* for 15 min and gently resuspended in high-salt buffer (0.6 M KCl, 20 mM HEPES, 1.5 mM $MgCl_2$, 20% glycerol, 0.1% NP-40, 10% [wt/vol] sucrose [Schwartz-Mann]). The nuclei were extracted on a nutator for 1 h at 4°C. The nuclear debris was removed by centrifuging at 80,000 × *g* in a SW-50 rotor for 30 min. The supernatant was dialyzed against 100 volumes of buffer D (20

mM HEPES, 0.1 M KCl, 20% glycerol. 0.1% NP-40) which was changed three times, once every 40 min; the total dialysis time was 2 h.

Purification of mSp1. The freshly prepared nuclear extract was applied to a 4-ml WGA Sepharose column at 15 ml/h. The flowthrough fraction was aliquoted into Eppendorf tubes and frozen in liquid nitrogen. The column was washed with 16 ml of buffer D. The bound proteins were eluted by adding 4 ml of buffer D containing 0.3 M *N*-acetyl glucoseamine (Sigma), mixing the resin, and allowing the resin to stand for 10 min prior to elution. The column was washed with 4 ml of buffer D-GlcNAc solution. Then, $ZnCl_2$ was added to the eluted proteins to a final concentration of 10 μM, and salmon sperm DNA was added to a final concentration of 0.4 μg/ml of eluate. No magnesium was added to any of the buffers used in the oligonucleotide sepharose chromatography. The protein solution was placed on a nutator for 10 min. After nutation, the protein was applied to a 1-ml column of BG1/2 oligonucleotide Sepharose at a flow rate of 15 ml/h. The eluate from this step was recycled over the column twice. The column was washed with 4 volumes of buffer D. Proteins were eluted by adding 1.2 ml of buffer D containing 0.6 M KCl, mixing the resin with a pipette tip, waiting 10 min, and eluting the released proteins. The remainder of the proteins were collected by washing the column with 1 ml of 0.6 M KCl–buffer D. The eluted proteins were diluted to a KCl concentration of 0.1 M, $ZnCl_2$ was added to 10 μM, and salmon sperm DNA was added to 5 ng/ml. The protein solution was reapplied to a fresh BG1/2 sepharose column, and the above steps were repeated. The final eluted protein was aliquoted into 50- and 100-μl aliquots and rapidly frozen in liquid nitrogen. The protein composition of various stages of the purification was analyzed by sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis (PAGE) (26), and the gels were stained by the method of Ansorge (2).

DNase I footprinting. The MVM P_4 promoter footprinting template is a *Bam*HI-*Nco*I fragment. It is prepared by restriction digestion of pMB415 (1) with either *Nco*I (C rich) or *Bam*HI (G rich), followed by dephosphorylation with calf intestinal phosphatase. After three phenol-chloroform-isoamyl alcohol (PCIA) extractions and two cycles of ethanol precipitation, the concentration of the template DNA was determined by UV spectrophotometry. A typical footprinting probe labelling reaction involved 2 pmol of template which was kinased by using T4 polynucleotide kinase and [γ-^{32}P]ATP (NEN; >6,000 Ci/mmol) followed by a 3-h restriction digest with either *Bam*HI or *Nco*I. An aliquot of the labelled DNA was removed, and its specific activity was determined by trichloroacetic acid precipitation. The labelled DNA was purified by PCIA, passed through a Biogel P-6 spin column, and resolved on a native Tris-borate-EDTA–5% polyacrylamide gel. The 260-bp promoter fragment was excised from the gel and eluted by diffusion overnight with a buffer containing 0.5 M NH_4 acetate, Tris-HCl [pH 7.5], 0.1 mM EDTA, and 0.1% SDS. The eluted DNA was purified by PCIA extraction and ethanol precipitation and was passed through a Biogel P-6 spin column. The concentration of the labelled, purified probe was determined by measuring the radioactivity by scintillation spectrometry of an aliquot and dividing by the specific activity.

For the footprint binding reactions, the probe was mixed with 4 μg of crude nuclear extract or 1 μl of WGA or DNA Sepharose fractions. Each reaction contained 60 mM KCl, 6 mM $MgCl_2$, 12 mM HEPES [pH 7.9], 12% glycerol, and 2% PVA, and 1 μg, 100 ng, or 10 ng of $(dI\text{-}dC)_n$ was added to the reaction for the respective protein sources. The mixture was incubated for 15 min at 30°C. The rest of the protocol has been described (9). The footprinting reactions were resolved on a 6% polyacrylamide–urea (19:1) sequencing gel. Sequencing ladders were prepared by previously described protocols (21).

The footprinting competition assays were conducted by mixing the competitor with the end-labelled probe before adding the affinity-purified Sp1. The remainder of the protocol was the same as the footprinting reaction outlined above. The competitors which were used are as follows. The P_4 competitor was an oligonucleotide with the P_4 promoter sequences from nucleotide 137 to 173. The SV40 early promoter 21-bp repeat competitor fragment is 187 bp long and contains the SV40 sequences between nt 5173 and 115, the *Sfa*NI-*Hin*dIII fragment. The SV40 early promoter enhancer competitor fragment is 72 bp long and contains the SV40 sequences between nt 115 and 187, the *Sfa*NI-*Xba*I fragment. The adenovirus major late competitor was an oligonucleotide containing the sequences from nt −90 to −48 relative to the initiation of transcription.

The SV40 fragments were derived by digesting the plasmid pSV2XCAT (containing an 8-bp *Xba*I linker inserted into the SV40 early promoter *Acc*I site of pSV2CAT) (1) with either *Xba*I or *Hin*dIII and then labelled by kinasing. To produce the uniquely end-labelled footprinting probe, the *Xba*I-labelled plasmid was digested with *Bgl*II and the *Hin*dIII end-labelled plasmid was digested with *Pvu*II. The probes were purified and quantitated as outlined above. The SV40 templates were DNase I footprinted following the same protocol as for the MVM P_4 promoter templates.

MPE footprinting. The P_4 promoter fragments were footprinted with methidium propyl-EDTA (MPE) (gift from P. Dervan) by the method outlined by Briggs et al. (6).

Methylation interference. The templates used in DNase I footprinting reaction were also used in the methylation interference reaction. The DNA was methylated according to the method of Hendrickson and Schlief (14). The conditions for the mobility shift were the same as in the DNase I footprinting reaction discussed above. The bound and free bands were localized by autoradiography of the wet gels, excised, and eluted by electroelution overnight into TEA electrophoresis buffer (26) containing 0.01% SDS. The eluted samples were PCIA extracted twice and precipitated by using ethanol with 10 μg of yeast rRNA as carrier. The samples were cleaved with piperidine, lyophilized, resuspended in formamide dyes, and run on a 6% (19:1) sequencing gel.

Mobility shift. The templates and DNA binding conditions used for mobility shift reactions were the same as those of the DNase I footprinting experiments (see above). Any competitor DNAs in a particular reaction were added with the labelled template at the beginning of the binding reaction. After the incubation at 30°C, the DNA-protein complexes were resolved on a vertical agarose gel and electrophoresed at 80 V at room temperature. The buffer in the gel was 20 mM TEA electrophoresis buffer with 0.05% NP-40. For the P_4 promoter, the agarose (Bio-Rad) concentration used was 2% and the gel was run for 3 h. A similar gel was used for the SV40 early promoter fragments except that the agarose concentration was 1.5%. The gels were run in 20 mM TEA electrophoresis buffer. The gel was cast between frosted glass plates with 1-mm vinyl spacers.

The competitors used in the mobility shift assay were as follows. The P_4 promoter is the 415-bp *Bam*HI-*Pst*I frag-

ment. The adenovirus major late promoter fragment is a 457-bp fragment spanning from nt 5774 to 6231 in the adenovirus 5 genome. The SV40 early promoter 21-bp repeat competitor fragment (*Sfa*NI-*Hin*dIII) is 187 bp long and contains the SV40 sequences between nt 5173 and 115.

In vitro transcription and primer extension. The standard transcription reaction (20 μl) was described previously (1). The primer extension conditions were described by Jones et al. (19). The transcription reconstitution reactions were as described by Jackson and Tjian (17). The primer that was used was complementary to nt 259 to 285 in the MVM(p) genome. Particular amounts of materials are defined in the figure legends.

UV photocross-linking. Single-stranded M13 DNA was isolated by superinfecting JM101 cells which harbored either pMB415 or pP4− with the helper phage R408 according to the manufacturer's instructions (Stratagene). Either the T3 or the T7 promoter primer was annealed to the purified single-stranded DNA and the primers were extended by using the Kleenow fragment of DNA Pol1 with 5-Br-dUTP, [α-^{32}P]dGTP, dCTP, and dATP as substrates. The *Bam*HI-*Nco*I fragment was isolated by restriction digestion and purification on a native 5% polyacrylamide gel. The fragment was used in a mobility shift assay as described above. The remainder of the protocol has been described by Gallinari et al. (10). The samples were run on a SDS–6% polyacrylamide (30:1) gel. Dimethylpimelimidate (DMP) for cross-linking was prepared by the method of Cover (8). Samples were treated with DMP by adding 1 μl of neutralized 20 mM DMP to the binding reactions (20 μl) and allowing the mixture to react for 10 min at 30°C. These samples were processed in the same manner as the non-DMP-treated samples.

RESULTS

Purification of mSp1. The two mSp1 protomers were purified from mouse A92L fibroblasts to near homogeneity by using a procedure based on the method outlined for the human Sp1 (hSp1) proteins (17). One significant change was that the DNA affinity resin used an oligonucleotide containing sequences from the P_4 promoter (Fig. 1A, BG1/2) coupled as a single duplex oligomer to Sepharose CL-4B beads. The purification was initially hampered by the persistence of a potent protease activity which contaminated even the final highly enriched affinity fraction. Thus, high levels of protease inhibitors were required throughout the purification scheme (see Materials and Methods) to ensure the integrity of the proteins. Moreover, the crude extract apparently has high levels of glycosidic enzymes, since prolonged dialysis or a freeze-thaw step resulted in significant amounts of Sp1 flowing through a WGA-Sepharose affinity column. Consequently, the freshly prepared nuclear extracts were immediately fractionated, and the entire purification was completed in 1 day. The purified protein appears as a doublet on a silver-stained SDS-polyacrylamide gel (Fig. 1B, lane A2). The purification was followed by DNase I footprinting assays (Fig. 1C). Because of the inclusion of high levels of the peptide protease inhibitor leupeptin, it was difficult to calculate overall yields of mSp1 by spectrophotometric dye binding assays, yet, as judged by silver staining, the yield from 3×10^9 cells ranged from 10 to 20 μg.

Further characterization of the purified protein demonstrated that it was biologically active and biochemically similar to hSp1. Purified mSp1 was capable of enhancing transcription from the P_4 promoter in a depleted crude nuclear extract (Fig. 2, lane 2 versus lanes 5, 7, and 9). Purified mSp1 cross-reacted strongly with a polyclonal anti-Sp1 antiserum raised against HeLa cell Sp1 (data not shown). By using the galactosyl transferase assay (16), purified mSp1 was strongly labelled with [^{3}H]galactose, as were some of the proteolytic degradation products (data not shown). As a further check on the activity of mSp1, we demonstrated that mSp1 forms footprints on the late strand of the GC boxes in the SV40 early promoter (Fig. 3).

The MVM P_4 GC box is a high-affinity Sp1 binding site. In a mobility shift competition assay, the P_4 GC box was originally shown to compete strongly for Sp1 binding by using either crude human (HeLa cell) or crude murine (A92L cell) nuclear extracts (1). Since the P_4 promoter contains homologies to three overlapping transcription factor binding site (Fig. 1A), it was possible that the affinity of Sp1 for the P_4 promoter was due to an interaction of Sp1 with another transcription factor(s). Therefore, purified mSp1 was used in a mobility shift competition assay to compare the competitive strength of the GC boxes in the SV40 21-bp repeats to the MVM P_4 GC box using either an end-labelled P_4 promoter fragment (Fig. 4A) or an SV40 early promoter fragment as the reporter (Fig. 4B). With either probe, it is clear that the P_4 promoter is as avid a competitor as the SV40 21-bp repeats, especially when one considers that there are six GC boxes, including two of high affinity in the SV40 promoter fragment used as a competitor (20). The avidity of mSp1 for the P_4 GC box was also tested in a competition assay using DNase I footprinting (Fig. 4C), in which the competitors were oligonucleotides. Again, both the SV40 and MVM promoters exhibit high-affinity binding. The results of these experiments support the earlier observation (1) that Sp1 binds tightly to the MVM P_4 promoter.

The mobility shift assay also provided some insight into the interaction of Sp1 with the SV40 21-bp repeats. In Fig. 4B, the P_4 promoter fragment was labelled, incubated with Sp1, and the P_4-Sp1 complex was run as a molecular weight marker alongside the labelled SV40 early promoter fragment. There were two discrete complexes formed on the SV40 fragment. The lower band (B1) has the same mobility as the band formed by the P_4 promoter fragment which corresponds to one Sp1 protomer binding to the SV40 21-bp repeats. The upper band (B2) corresponds to two molecules of Sp1 binding. The results of methylation interference studies on the shifted bands lend support to there being several different partially filled GC boxes in each band, because no specific pattern of interference was seen in either band (24). Apparently, either one of the two high-affinity GC boxes in the 21-bp repeats can be bound by Sp1 in the lower complex.

Sp1 contacts both strands of the P_4 GC box. Methylation interference experiments were conducted to define the guanine and adenine residues in the P_4 GC box which were contacted by Sp1. The results of these experiments are shown in Fig. 5. For stable binding to the P_4 GC box, mSp1 contacts seven guanine residues on the G-rich plus strand, including a guanine which is outside of the normal decamer consensus sequence limit. The large methylation interference footprint is consistent with the results of MPE-Fe(II) footprinting in which a 17-bp footprint extending beyond the decamer limits of the GC box is seen (Fig. 6). Interestingly, methylation interference experiments conducted on the C-rich minus strand revealed that the central guanine and a peripheral adenine are contacted by mSp1 (Fig. 5B). We obtained similar results when purified hSp1 was used instead of mSp1 in the methylation interference assays (data not shown). The large number of residues that Sp1 contacts in

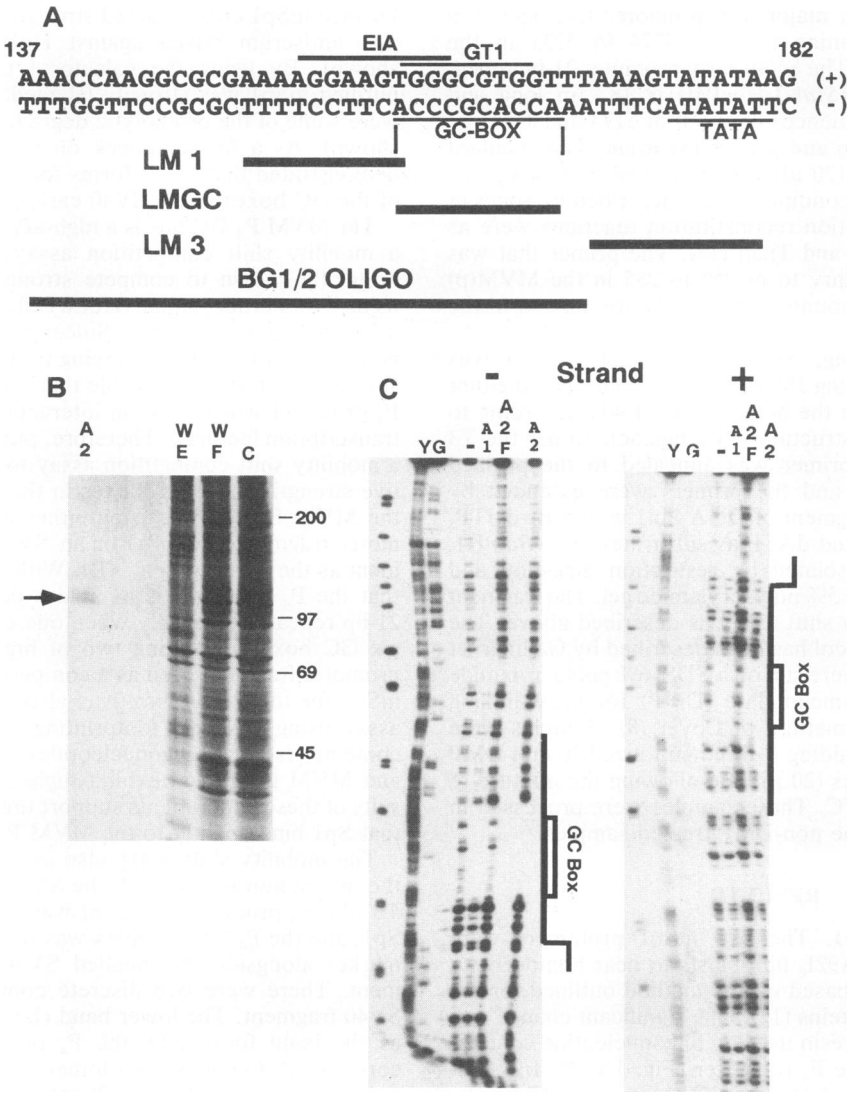

FIG. 1. MVM P_4 promoter sequence and purification of mSP1. (A) A partial sequence of the P_4 promoter of MVM(p); transcriptional initiation occurs at nt 205. The plus (+) and minus (−) strands are indicated. LM-1 (nt 150 to 159), LM-GC (nt 159 to 168), LM-3 (nt 171 to 180) are *Bgl*II linker insertion mutants (1); promoter sequences which are replaced by the linker are denoted by the dark line. BG1/2 is an oligonucleotide which contains sequences (nt 137 to 173) from the P_4 promoter (underlined) used to purify mSp1. The E1A sequence is an 8 of 10 nucleotide match to the E1A enhancer element (13). The GT1 sequence is a seven of eight nucleotide match to the GT1 element in the SV40 early promoter (28). The GC box is a ten of ten match to the GC box consensus sequence (20). (B) Aliquots of murine fibroblast A92L nuclear extract (C), WGA flowthrough fraction (WF), WGA GlcNAc eluate (WE), and a TCA precipitate of 50 μl of a 0.6 M KCl fraction from the second DNA Sepharose column (A2) were electrophoresed on a SDS–6 to 12% polyacrylamide gradient gel and silver stained. The affinity fraction (lane A2) contains a prominent doublet of bands at 95 and 105 kDa (arrow) and two fainter bands at ~67 and 45 kDa which cross-react with a polyclonal anti-hSp1 antibody and therefore appear to be proteolytic degradation products of full-length Sp1. (C) Autoradiogram of DNase I footprinting assay of affinity fractions. Lane Y is a C+T sequencing reaction. Lane G is a G sequencing reaction. The top strand (+) and the bottom strand (−) of the P_4 promoter are indicated. The lanes are marked as follows: −, no protein; A1, 1 μl of 0.6 M eluate from DNA Sepharose column 1; A2F, 1 μl of flowthrough fraction from DNA Sepharose column 2; A2, 1 μl of 0.6 M KCl fraction from DNA Sepharose column 2. The position of the GC box is indicated by the open rectangle.

the P_4 GC box led us to inquire about the number of Sp1 protomers which were binding to this GC box; usually, one protomer of Sp1 contacts only 3 or 4 guanine residues on the G-rich strand of a GC box.

Sp1 protomers assymmetrically cross-link to the P_4 GC box. We used the mobility shifted P_4-mSp1 complex in Fig. 4 as the starting point for the UV photocross-linking of mSp1 to the P_4 promoter fragment. This complex had already been demonstrated (Fig. 4 and 5) to be the result of a specific P_4-mSp1 interaction. The UV photocross-linking scheme that we used combines two methods which were recently reported (10, 12). This method, summarized in Fig. 7A, involves generating promoter fragments which are substituted with Br-dUTP and [α-^{32}P]dGTP on one strand by primer extension of a primer annealed to a single-stranded M13 template. The promoter containing fragment is isolated by restriction digestion, purified, and used in a mobility shift assay, and the resultant DNA-protein complexes are photo-

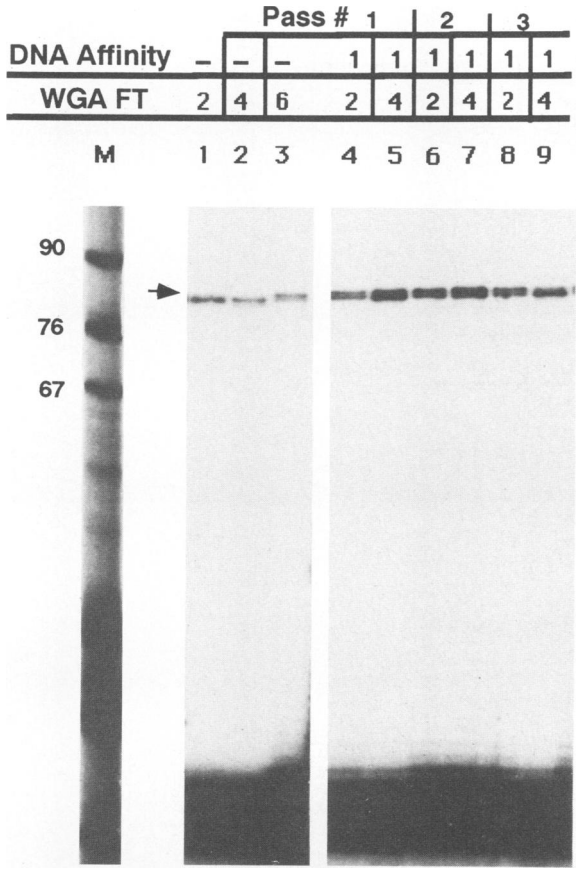

FIG. 2. Primer extension analysis of enhanced P_4 promoter transcription. Transcription reconstitution from the P_4 promoter. The amounts of WGA flowthrough (WGA FT) fraction (an Sp1 depleted fraction) and the affinity-purified Sp1 A2 fraction (DNA Affinity) used in the reactions are indicated above each lane. Lanes 1 to 3, WGA flowthrough fraction alone (2, 4, and 6 μl, respectively); Lanes 4, 6, and 8, 2 μl of WGA FT fraction added; Lanes 5, 7, and 9, 4 μl of WGA FT fraction. Lanes 4 and 5 were supplemented with affinity-purified mSp1 (1 μl of 0.6 M eluate from DNA Sepharose column 1). Lanes 6 and 7 were supplemented with mSp1 (1 μl of 0.6 M KCl fraction from DNA Sepharose column 2). Lanes 8 and 9 were supplemented with 1 μl of 0.6 M KCl fraction from a third DNA Sepharose column. Compare lane 2 (depleted) to lanes 5, 7, and 9 (supplemented). The primer extension product of a correctly initiated P_4 transcript is 82 nucleotides long. Lane M is radioactively end-labelled pBR322 *Msp*I fragments.

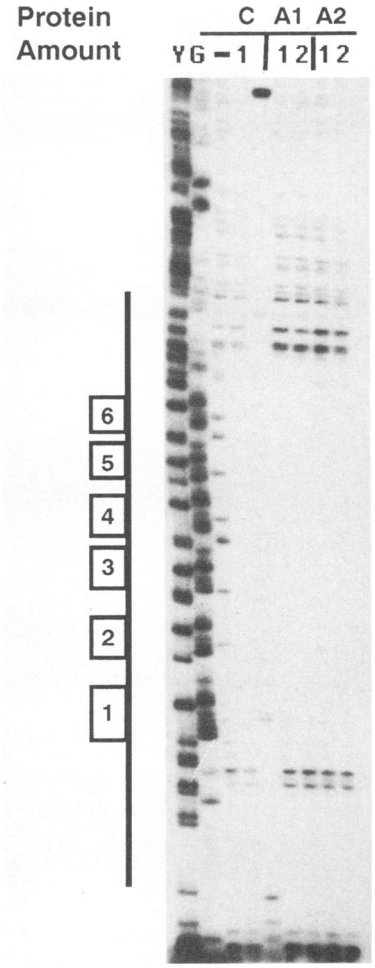

FIG. 3. DNase I footprinting assay of mSp1 bound to the late strand of the SV40 early promoter. Four femtomoles of the G-rich strand of the SV40 early promoter was end labelled and incubated with no protein (–), 1 μl of crude A92L nuclear extract (lane C1), 1 and 2 μl of WGA Sepharose GlcNAc eluate (lanes A1-1 and A1-2, respectively), and 1 and 2 μl of 0.6 M KCl fraction from DNA Sepharose column 2 (lanes A2-1 and A2-2, respectively). Lanes Y and G are the C+T and G Maxam and Gilbert sequencing reactions, respectively.

cross-linked in the gel. Advantages of this method include the following: (i) the template is randomly labelled and there are no assumptions as to which guanine residues will be contacted by a DNA binding protein; (ii) because M13 clones containing either the plus or minus strands can be isolated for selective strand labelling, it allows the examination of possible contacts on both strands; and (iii) the specificity of the method originates from the isolation of a particular mobility-shifted band for cross-linking analysis. Since the P_4 promoter forms only one band in the mobility shift assay when mixed with purified Sp1 and this band was shown by methylation interference to be a specific interaction with the GC box, we were confident that the interactions that were being detected in the cross-linking assay were specifically those of Sp1 binding to the GC box of the P_4 promoter.

Figure 7B shows the results of cross-linking mSp1 to the P_4 GC box. When the G-rich plus strand was substituted with Br-dUTP and [α-^{32}P]dGTP, only the 95-kDa protomer of mSp1 cross-linked to the P_4 GC box; the 105-kDa protomer was not detected. In contrast, when the C-rich minus strand was labelled, the 105-kDa protomer cross-linked to the GC box. We concluded from these results that the two Sp1 protomers are contacting opposite strands of the P_4 GC box. In Fig. 7B, the lanes marked DMP + show the results of an attempt to chemically cross-link the 95- and 105-kDa protomers together by using DMP (1.1-nm linker) and then mobility shifting and UV photocross-linking to the P_4 promoter in the gel matrix. We had hoped to see the appearance of a new band of approximately 200 kDa, which would indicate that the two promoters of Sp1 had been cross-linked together. However, we were unable to detect any dimeric species of Sp1 which cross-linked to the labelled DNA.

Both the photocross-linking and the methylation interference assays were performed on the same mobility-shifted complex. Thus, the results of the two types of experiments

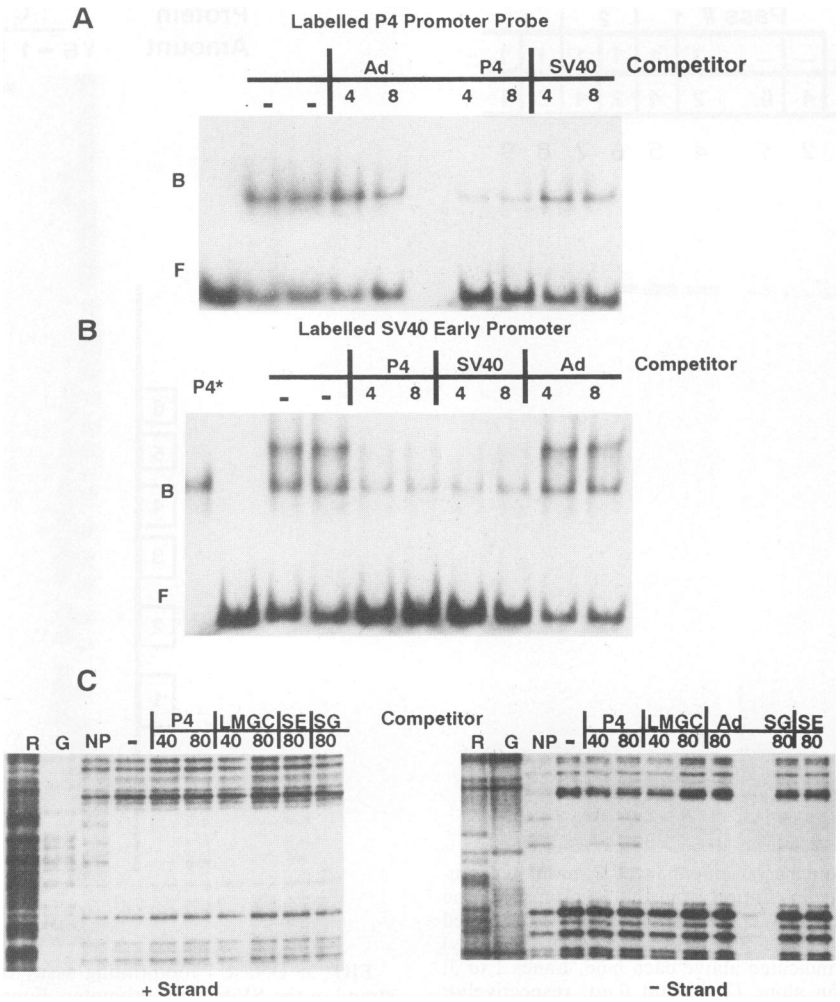

FIG. 4. Mobility shift and DNase I footprinting competition assays. (A) Mobility shift of 4 fmol of end-labelled P_4 promoter fragment (*Bam*HI-*Nco*I, 260 bp) bound by 1 μl of affinity-purified (passage two) mSp1. The lanes contain no competitor (−), cold adenovirus major late promoter fragment (Ad), cold P_4 promoter fragment (P4), or cold SV40 early promoter fragment (SV40). The competitors were present at either 4 or 8 fmol. Competitors are described in Materials and Methods. (B) Mobility shift of 4 fmol of end-labelled SV40 early promoter fragment (*Hin*dIII-*Pvu*II, 342 bp) bound by affinity-purified (passage two) mSp1. The competitors were the same as described above. The lane labelled P4* contains 4 fmol of end-labelled P_4 promoter fragment bound by 1 μl of affinity-purified (passage two) mSp1. (C) DNase I footprinting competition of affinity-purified (passage two) mSp1 binding to the P_4 promoter (4 fmol of *Bam*HI-*Nco*I fragment). The plus (+) and minus (−) strands are indicated. Lanes R and G, Maxam and Gilbert A+G and G sequencing reactions, respectively. Lanes: NP, no protein; −, no competitor; P4, cold P4 promoter fragment at 40 and 80 fmol; LMGC, cold GC box linker scanning mutation promoter fragment at 40 and 80 fmol; Ad, cold adenovirus major late promoter at 80 fmol; SG, SV40 21-bp repeats at 80 fmol; SE, SV40 enhancer at 80 fmol. Competitors are described in Materials and Methods.

taken together imply that either both promoters of Sp1 are binding to the P_4 GC box individually and that their conformations differ such that they are cross-linked to opposite strands or that both promoters of Sp1 are binding simultaneously to the P_4 GC box and are unreactive to cross-linking by DMP.

GC box consensus sequence. The current Sp1 consensus sequence predicted that the P_4 GC box would be a low-affinity binding site, yet the experimental evidence showed that it was a strong recognition site for Sp1. In an effort to resolve the paradox of these findings, we decided to analyze the distribution of sequences found in GC boxes. Figure 8 shows an analysis of the distribution of nucleotides in 64 promoter-associated GC boxes which have been reported in the literature. In this analysis we did not rank the sequences in order of their Sp1 affinity; such a method had failed to predict the strength of the P_4 GC box. The distribution of nucleotides in the GC boxes surveyed produces essentially the same consensus sequence as the one previously published (20) despite there being 34 different GC box sequences. The guanines at positions 3, 4, and 6 are absolutely conserved in all of GC boxes surveyed. Guanine residues 3, 4, and 6 have been shown to be important residues by methylation interference and protection for Sp1 binding (5, 11, 18, 20).

In examining the sequences collected in Fig. 8, we noted that there are nine GC boxes with a thymine at position 7. Interestingly, the thymine at position 7 correlates with Sp1 making contacts on both strands of the GC box element in all three of the GC boxes which have been characterized.

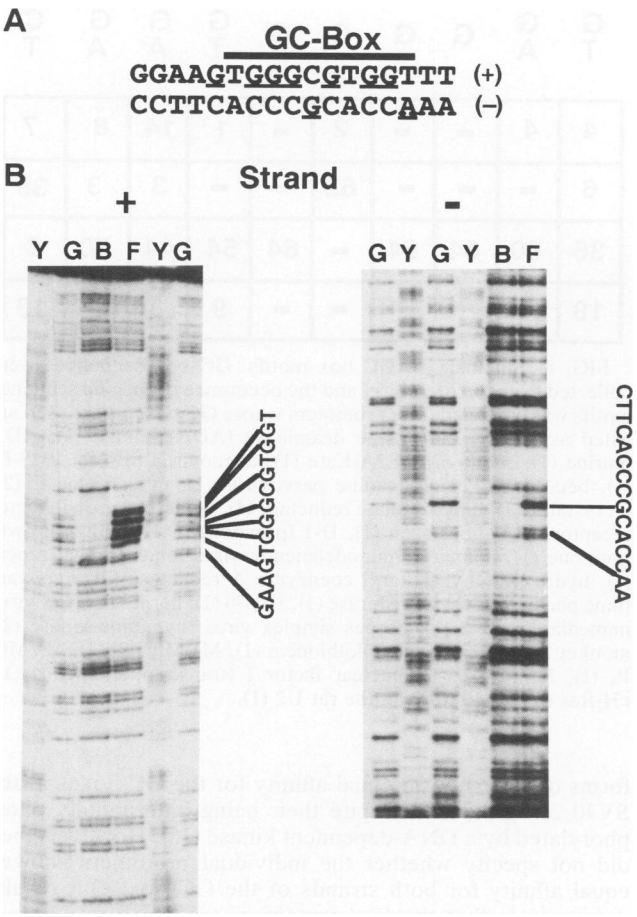

FIG. 5. Methylation interference. (A) Summary of residues (underlined) which interfere with mSp1 binding to the P_4 GC box when methylated. The conditions used in the mobility shift assay are described in Materials and Methods. (B) Methylation interference results. The P_4 promoter G-rich (+) and C-rich (−) strands are indicated. Lane Y, C+T sequencing ladder; lane G, G sequencing ladder; lane B, bound fraction; lane F, free fraction. The positions of the guanine residues which are contacted is indicated on the right side of each panel of lanes.

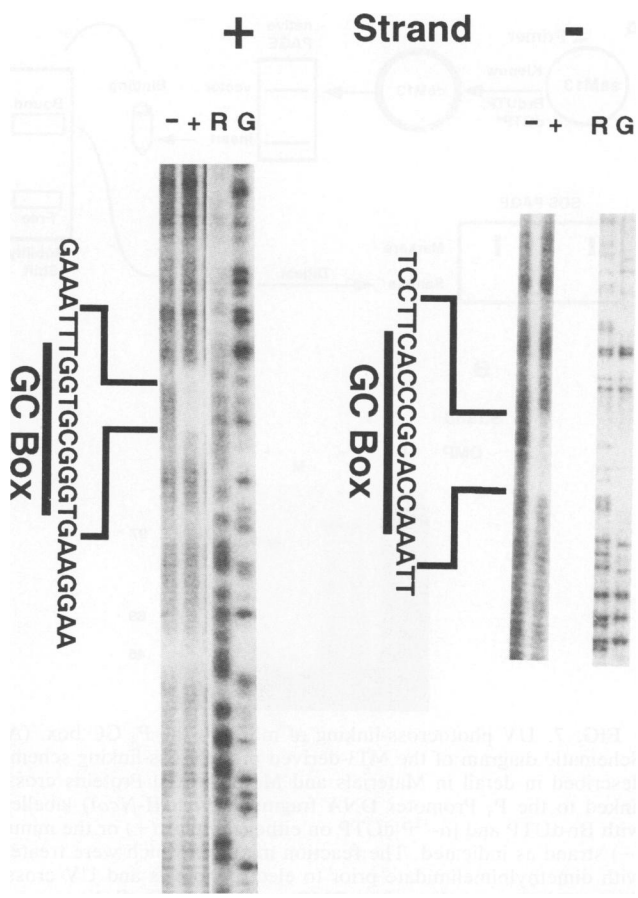

FIG. 6. MPE footprinting of P_4 promoter. The P_4 promoter G-rich (+) strand and C-rich (−) strand are indicated. Lane −, no protein; lane +, 2 μl of 0.6 M KCl fraction from DNA Sepharose column 2; lane R, A+G sequencing ladder; lane G, G sequencing ladder. The reaction conditions are described in Materials and Methods.

Moreover, both the P_4 and human immunodeficiency virus GC box 3 are high affinity (18). We propose that the thymine residue at position 7 of the GC box is indicative of a subclass of high-affinity GC boxes.

DISCUSSION

P_4 promoter is GC box activated. The main conclusion of the P_4 promoter mutagenesis studies was that the P_4 promoter is a GC box-driven promoter (1). The purification of mSp1 as the dominant protein binding to the P_4 promoter and the demonstration that affinity-purified mSp1 is able to enhance transcription from the P_4 promoter in a depleted nuclear extract substantiate this conclusion. The low level of transcription that persists from the P_4 promoter may be due to close (20 nucleotides) positioning of the P_4 TATA box to the RNA initiation site. This basal activity is consistent with the results of chloramphenicol acetyltransferase assays with a deletion mutant of the P_4 promoter which has all of the sequences upstream of the TATA box deleted yet still retains low-level activity (1). Further proof of this hypothesis is found in the results of recent viral reconstruction experiments from the Pintel laboratory (23). The *Bgl*II linker scanning mutations in the P_4 promoter (Fig. 1A) were reconstructed into a full-length, infectious MVM(p) clone, and after transfection, the transfected cell supernatant was assayed for the production of viable virus in a plaque assay. The linker mutants replacing the GC box (LM-GC) and the TATA box (LM-3) were lethal; no detectable virus was recovered. A linker mutant in the sequence immediately upstream of the GC box, LM-1, reduced the yield of recoverable virus 25-fold. Since the linker mutations reside in viral sequences which appear to exclusively regulate early gene transcription and are not likely to be involved in packaging and virus assembly, the results demonstrate that in vivo the P_4 promoter requires an intact GC box and TATA box for activity. Furthermore, the finding that sequences adjacent to the GC box decamer are required for in vivo activity is supported by the results of the methylation interference experiments (Fig. 5). The experiments revealed that one of the guanine residues replaced by the linker is contacted by Sp1. Indeed, repeated attempts to map the potential contact residues within the LM-GC and LM-1 mutants by methylation interference and UV photocross-linking experiments using purified mSp1 were unsuccessful (24), indicating that

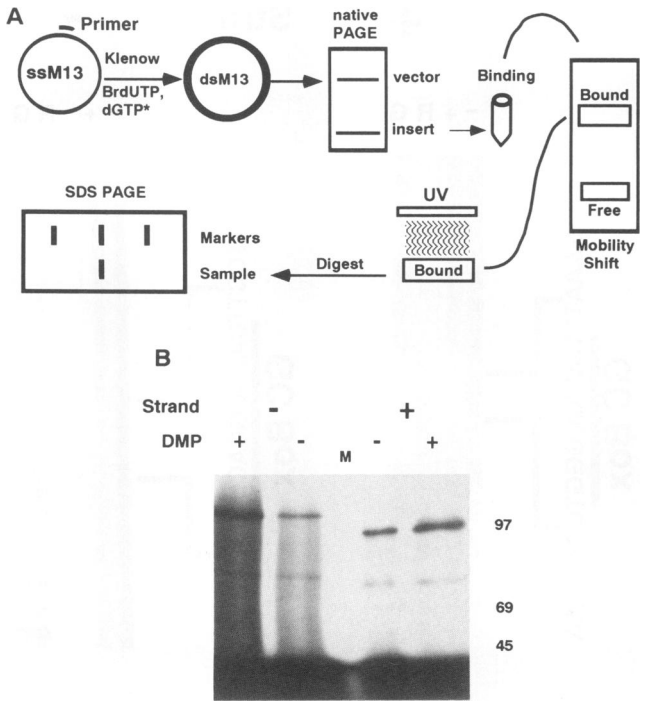

FIG. 7. UV photocross-linking of mSp1 to the P_4 GC box. (A) Schematic diagram of the M13-derived probe cross-linking scheme described in detail in Materials and Methods. (B) Proteins cross-linked to the P_4 Promoter DNA fragment (*Bam*HI-*Nco*I) labelled with Br-dUTP and [α-^{32}P]dGTP on either the plus (+) or the minus (−) strand as indicated. The reaction mixtures which were treated with dimethylpimelimidate prior to electrophoresis and UV cross-linking(DMP) are indicated as DMP +. The 105-kDa Sp1 protomer specifically binds the minus strand while the 95-kDa protomer binds the plus strand.

G T	G A	G	G	C	G	G T	G A	G A	C T
4	4	-	-	2	-	1	14	8	7
6	-	-	-	62	-	-	3	3	36
36	60	64	64	-	64	54	44	50	8
18	-	-	-	-	-	9	3	3	13

FIG. 8. Summary of GC box motifs. GC box sequences were collected from the literature, and the occurence of specific sequence motifs was compiled. The promoters whose GC boxes were used are listed as follows: adenosine deaminase (ADA) human (3), ADA murine (1), adenovirus E2A-Late (1), adenovirus pIX (1), B19-P6 (1), beta-DNA Pol (3), canine parvovirus (1), cytochrome C (2), δ-crystallin (1), dihydrofolate reductase (4), epidermal growth factor receptor (1), γ-fibrinogen (1), H-1 (parvovirus) (1), human growth hormone (1), human immunodeficiency virus long terminal repeat (3), hydroxymethyl glutaryl coenzyme A reductase (2), hypoxanthine phosphoribosyltransferase (1), Hsp70 (1), herpes simplex virus immediate-early 3 (5), herpes simplex virus thymidine kinase (2), monkey promoter (3), metallothionein (1), MVM(p) P_4 (1), MVM(i) P_4 (1), MVM P_{38} (1), nuclear factor I (transcription factor) (1), cH-Ras (6), SV40 (6), and the rat U2 (1).

both linker insertion mutations disrupt contacts necessary for mSp1 binding.

Different classes of GC boxes. At this time, there is sufficient evidence to define two classes of high-affinity GC boxes. One class is certainly the class with a guanine at position 7, where one protomer of Sp1 binds to each GC box element (7). This class of GC boxes is recognized with equal affinity by both protomers of Sp1, and the proteins contact only the G-rich strand. We suggest that the P_4 GC box is representative of another class of high-affinity GC box with a thymine at position 7, where Sp1 makes contacts with guanines on both strands of the GC box element. All of the members of this class that have been characterized are contacted by Sp1 on both strands. With respect to this point, the recent work of Thiesen and Bach is illuminating (27). They used a nitrocellulose filter binding assay with purified Sp1 to select oligonucleotides with high-affinity GC boxes from a pool of oligonucleotides which had random sequences. Their analysis offers a neutral assessment of Sp1 DNA sequence affinity. In accordance with our proposal, one of the two high-affinity oligonucleotides that was selected, S3, contains a thymine at position 7 (AGGGCGT ATA).

Sp1 switches or heterodimers? An unusual feature of Sp1's binding to the P_4 GC box is the asymmetric cross-linking of the two Sp1 protomers to the positive and negative strands. Recently, Jackson et al. reported that the 95- and 105-kDa forms of Sp1 showed equal affinity for the GC boxes in the SV40 21-bp repeats despite their being differentially phosphorylated by a DNA-dependent kinase (15). However, they did not specify whether the individual protomers showed equal affinity for both strands of the GC box. Our results demonstrate that the two protomers cross-link to different sequences in the P_4 GC box. Since the cross-linked samples were exhaustively digested with 100 U of DNase I (35 μU are used in a DNase I footprinting reaction) and micrococcal nuclease prior to SDS-PAGE, we feel that this would remove any residual non-cross-linked DNA which could artifactually increase the apparent molecular weight of the cross-linked species. Further proof that the two bands are representative of distinct protein species is found in the protease sensitivity of the two protomers. Initial attempts to cross-link proteins to the P_4 promoter revealed that the 95-kDa protomer cross-linked to the G-rich strand when it was radioactively labelled. With the label on the opposite strand, a faint band at 105 kDa was visible, as were several discrete lower-molecular-weight bands corresponding to proteolytic degradation products of the full-length Sp1 protomers; similar degradation products have been reported by others (6). When the concentration of protease inhibitors was raised to 50 μg/ml, the 105-kDa band increased in intensity, while the lower-molecular-weight bands decreased in intensity. In contrast, the intensity of the 95-kDa band remained unchanged. Collectively, the results demonstrate that the Sp1 95-kDa protomer makes contacts which allow it to be cross-linked to the G-rich strand of the GC box and reciprocally, the 105-kDa protomer makes different contacts which allow it to be cross-linked to the C-rich strand. Furthermore, since the M13 cross-linking experiments were conducted with affinity-purified protein under conditions which do not favor phosphorylation of the 95-kDa protomer, it is unlikely that a 95-kDa protomer contacting the C-rich strand was preferentially phosphorylated and converted into the 105-kDa protomer during the incubation reaction.

We have considered two different models of Sp1 binding to the P_4 GC box to explain the footprinting and cross-linking data. In the first model, Sp1 binds to the P_4 promoter as a monomer contacting both strands of the P_4 GC box as suggested by the methylation interference assay. In this context, the mobility shift patterns in Fig. 4 show that the lowest bound state of either the 21-bp repeats or the P_4 promoter contains one Sp1 protomer. The differences in the cross-linking susceptibility of the two protomers suggests that the conversion of the 95-kDa protomer into the 105-kDa protomer changes the proximity of reactive residues on Sp1 relative to the Br-dUTP residues in or flanking the P_4 GC box. Whether the phosphorylation of Sp1 changes the cross-linking susceptibility by electrostatically repelling a reactive intermediate or causes a conformational switch in Sp1 is unclear. However, using a mobility shift assay, Jackson et al. report that both isoforms of Sp1 bind the SV40 high-affinity GC box with equal avidity (15).

An alternate model is that two Sp1 protomers are simultaneously bound to the P_4 GC box as a dimer. Previously, Courey et al. have postulated that Sp1 could not form dimers when both Sp1 protomers were bound to adjacent GC boxes (7). However, they did detect the presence of a subpopulation of Sp1 which can dimerize in solution. The possibility exists that by using the P_4 GC box sequences in the DNA affinity column, we have enriched a subpopulation of Sp1 which can form dimers. Still, there are no data to suggest that the two protomers of Sp1 have different affinities for sequences in the GC box, and this is an inherent weakness in the dimer model. Because we were using a mixture of both protomers in our assays, we are unable to directly test whether both isoforms of Sp1 are required to produce the observed methylation interference pattern or even the mobility shift pattern.

The first model is the simplest explanation of our data. Our inability to cross-link together the two promoters of Sp1 in the gel-shifted band argues against the simultaneous presence of both protomers on the same P_4 GC box. Additionally, the absence of a protein-bound band with a faster mobility than either the P_4 GC box or the SV40 21-bp repeats in the mobility shift assay indicates that either Sp1 binds as a dimer to all GC boxes or that it binds as a single protomer. Moreover, the dimer model is predicated on the untested assumption that a single protomer of Sp1 can't contact both strands of the P_4 GC box. Different cross-linking approaches will need to be tried in order to absolutely prove or disprove the presence of Sp1 dimers on the GC box. Additionally, the differences in the mobility of the cross-linking adducts may be simply due to the cross-linking of a larger piece of DNA to Sp1 when it cross-links to the C-rich strand as opposed to the G-rich strand. If this were true, it does not explain the susceptibility of the C-rich cross-linking adduct to proteases in the original binding reaction.

Thus, these results highlight an unusual manner in which Sp1 can bind to a GC box and raise the possibility that different conformations of Sp1 protomers are involved in a general mechanism of regulating transcription from GC boxes. Characterization of the way in which phosphorylation influences Sp1's interaction with the rest of the transcription apparatus awaits the purification of the adaptor molecules which link Sp1 and transcription factor IID (25).

ACKNOWLEDGMENTS

We thank Peter Dervan for the gift of MPE. We thank S. Jackson and R. Tjian for the generous gift of anti-hSp1. We thank D. Pintel for communicating results prior to publication. We thank Don Crothers and Joan Steitz for their critical reading of the manuscript and valuable comments.

This research was funded by NIH grant AI19973 (D.C.W.).

REFERENCES

1. **Ahn, J. K., B. J. Gavin, G. K. Kumar, and D. C. Ward.** 1989. Transcriptional analysis of minute virus of mice P_4 promoter mutants. J. Virol. **63**:5425–5432.
2. **Ansorge, W.** 1985. Fast and sensitive detection of protein and DNA bands by treatment with potassium permanganate. J. Biochem. Biophys. Meth. **11**:13–20.
3. **Arndt-Jovin, D. J., T. M. Jovin, W. Bahr, A.-M. Frischauf, and M. Marquardt.** 1975. Covalent attachment of DNA to agarose. Eur. J. Biochem. **54**:411–418.
4. **Baeuerle, P. A., and D. Baltimore.** 1988. IκB: a specific inhibitor of the NF-κB transcription factor. Science **242**:540–546.
5. **Blundell, M. C., and C. R. Astell.** 1989. A GC-box motif upstream of the B19 parvovirus unique promoter is important for in vitro transcription. J. Virol. **63**:4814–4823.
6. **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.
7. **Courey, A. J., D. A. Holtzman, S. P. Jackson, and R. Tjian.** 1989. Synergistic activation by the glutamine-rich domains of human transcription factor Sp1. Cell **59**:827–836.
8. **Cover, J. A., J. M. Lambert, C. M. Norman, and R. R. Traut.** 1981. Identification of proteins at the subunit interface of the *Escherichia coli* ribosome by cross-linking with dimethyl 3,3′-dithiobis(proprionimidate). Biochemistry **20**:2843–2852.
9. **Dynan, W. S., and R. Tjian.** 1983. The promoter-specific transcription factor Sp1 binds to upstream sequences in the SV40 early promoter. Cell **35**:79–87.
10. **Gallinari, P., F. La Bella, and N. Heintz.** 1989. Characterization and purification of H1TF2, a novel CCAAT-binding protein that interacts with a histone H1 subtype-specific consensus element. Mol. Cell. Biol. **9**:1566–1575.
11. **Gidoni, D., W. S. Dynan, and R. Tjian.** 1984. Multiple specific contacts between a mammalian transcription factor and its cognate promoters. Nature (London) **312**:409–413.
12. **Gilmour, D. S., T. J. Dietz, and S. C. R. Elgin.** 1990. UV cross-linking identifies four polypeptides that require the TATA box to bind to the *Drosophila hsp70* promoter. Mol. Cell. Biol. **10**:4233–4238.
13. **Hearing, P., and T. Shenk.** 1983. The adenovirus type 5 E1A transcriptional control region contains a duplicated enhancer element. Cell **33**:695–701.
14. **Hendrickson, W., and R. Schlief.** 1985. A dimer of AraC contacts three adjacent major groove regions of the *ara1* DNA site. Proc. Natl. Acad. Sci. USA **82**:3129–3133.
15. **Jackson, S. P., J. J. MacDonald, S. Lees-Miller, and R. Tjian.** 1990. GC box binding induces phosphorylation of Sp1 by a DNA-dependent protein kinase. Cell **63**:155–165.
16. **Jackson, S. P., and R. T. Tjian.** 1988. *O*-Glycosylation of eukaryotic transcription factors: implications for mechanisms of transcriptional regulation. Cell **55**:125–133.
17. **Jackson, S. P., and R. Tjian.** 1989. Purification and analysis of RNA polymerase II transcription factors by using wheat germ agglutinin affinity chromatography. Proc. Natl. Acad. Sci. USA **86**:1781–1785.
18. **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.
19. **Jones, K. A., K. R. Yamamoto, and R. Tjian.** 1985. Two distinct transcription factors bind to the HSV thymidine kinase promoter *in vitro*. Cell **42**:559–572.
20. **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.
21. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
22. **Mitchell, P. J., and R. Tjian.** 1989. Transcriptional regulation in

mammalian cells by sequence specific DNA binding proteins. Science **245**:371–378.
23. **Pintel, D. (University of Missouri, Columbia).** 1991. Personal communication.
24. **Pitluk, Z. W., and D. C. Ward.** Unpublished observations.
25. **Pugh, F. B., and R. Tjian.** 1990. Mechanism of transcriptional activation by Sp1: evidence for coactivators. Cell **61**:1187–1197.
26. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
27. **Thiesen, H.-J., and C. Bach.** 1990. Target detection assay (TDA): a versatile procedure to determine DNA binding sites as demonstrated on SP1 protein. NAR **18**:3203–3209.
28. **Xiao, J.-H., I. Davidson, M. Macchi, R. Rosales, M. Vigneron, A. Staub, and P. Chambon.** 1987. In vitro binding of several cell-specific and ubiquitous nuclear proteins to the GT-1 motif of the SV40 enhancer. Genes Dev. **1**:794–807.