41. P. M. Whitington *et al.*, *Science* **215**, 973 (1982); J. R. Sanes, J. G. Hildebrand, D. J. Prescott, *Dev. Biol.* **52**, 121 (1976).
42. S. E. Blackshaw, J. G. Nicholls, I. Parnas, *J. Physiol. (London)* **326**, 261 (1982); R. K. Murphey and C. A. Lemere, *Science* **224**, 1352 (1984).
43. C. M. Bate, *Nature (London)* **260**, 54 (1976); D. Bentley and H. Keshishian, *Science* **218**, 1082 (1982); C. S. Goodman, J. A. Raper, R. K. Ho, S. Chang, *Symp. Soc. Dev. Biol.* **40**, 275 (1982); R. Levinthal, E. Macagno, C. Levinthal, *Cold Spring Harbor Symp. Quant. Biol.* **40**, 321 (1975).
44. M. J. Bastiani and C. S. Goodman, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 1849 (1984).
45. C. S. Goodman, M. O'Shea, R. McCaman, N. C. Spitzer, *Science* **204**, 1219 (1979); C. M. Bate, *Nature (London)* **260**, 54 (1976); J. A. Raper, M. J. Bastiani, C. S. Goodman, *J. Neurosci.* **3**, 20 (1983); J. E. Sulston and H. R. Horvitz, *Dev. Biol.* **56**, 160 (1977).
46. I. A. Meinertzhagen and A. Frölich, *Trends Neurosci.* **6**, 223 (1983); P. J. Stephens and C. K. Govind, *Brain Res.* **212**, 476 (1981).
47. R. K. Murphey and S. G. Matsumoto, *Science* **191**, 564 (1976); J. Palka, *Trends Neurosci.* **7**, 455 (1984).
48. This article is based on five workshops organized by P.R. at the Neurosciences Institute of the Neurosciences Research Program in New York during 1983–84. Each session was moderated by one of the four authors, who are listed alphabetically. The following scientists participated in the workshops: F. Bonhoeffer, D. Bray, M. Constantine-Paton, G. Edelman, E. Frank, J. Freeman, E. Gall, C. Goodman, Z. Hall, W. Harris, M. Hollyday, R. Lund, E. Macagno, R. Murphey, J. Nicholls, J. Palka, E. Rubel, G. Stent, M. Stryker, D. Trisler, H. Van der Loos, and D. Willshaw. The workshops were supported by the Neurosciences Research Foundation.

RESEARCH ARTICLE

Bidirectional SV40 Transcription Mediated by Tandem Sp1 Binding Interactions

David Gidoni, James T. Kadonaga, Hugo Barrera-Saldaña
Keikichi Takahashi, Pierre Chambon, Robert Tjian

Abstract. *The 21–base pair repeat elements of the SV40 promoter contain six tandem copies of the GGGCGG hexanucleotide (GC-box), each of which can bind, with varying affinity, to the cellular transcription factor, Sp1. In vitro SV40 early RNA synthesis is mediated by interaction of Sp1 with GC-boxes I, II, and III, whereas transcription in the late direction is mediated by binding to GC-boxes III, V, and VI.*

The pattern of gene expression in mammalian cells requires thousands of genes to be turned on and off in a temporally and spatially regulated manner. The critical conditions suitable for regulating the expression of a gene product often occur at the level of transcription. To understand the mechanisms of transcriptional regulation in animal cells, we have used DNA tumor viruses such as SV40 because they provide a relatively simple and valuable model for studying transcriptional specificity. Important *cis*-regulatory elements of the SV40 early promoter have been mapped, and reconstituted in vitro transcription reactions have allowed us to identify and isolate specific cellular factors that recognize and bind to the viral promoter. We now report our analysis of the interaction of a sequence-specific DNA binding protein that activates bidirectional transcriptional initiation from the SV40 promoter region.

The early genes of simian virus 40 (SV40) are expressed shortly after infection, whereas the late genes are maximally activated only after the onset of viral DNA replication and repression of viral early transcription by T antigen (*1, 2*). Analysis of viral promoter mutants both in vivo and in vitro have established that a region of approximately 300 base pairs (bp) adjacent to the origin of DNA replication contains multiple *cis*-regulatory elements responsible for directing transcription of both early and late viral messenger RNA (mRNA) synthesis. Mutational analyses of the viral transcriptional control sequences have revealed that the major early promoter consists of three 21-bp repeated elements preceded by a stretch of AT-rich sequences, and early transcription has been shown to initiate predominantly from distinct sites located 20 to 30 nucleotides downstream from the AT-rich region (*3–11*). In addition, enhancer elements that stimulate SV40 early transcription in vivo are located within the 72-bp repeated sequences, which lie 110 to 250 bp upstream from the early transcription start sites (*4, 7, 12–15*). Late viral transcription appears to be under the direction of multiple regulatory elements and exhibits a heterogeneous population of start sites scattered throughout the control region with a major initiation site at nucleotide 325 and several minor ones located at various positions (*16*). The 21-bp repeats that constitute a major promoter element for early transcription also appear to be a component of the late promoter (*7, 9, 11, 17–19*). In particular, a minor late transcript initiating at nucleotide 170 is strongly dependent on the 21-bp repeated sequences in vitro (*9, 11, 19*). Transcriptional analysis of various plasmid templates containing the 21-bp repeats in an inverted orientation relative to the AT-rich TATA homology confirm the observation that this promoter sequence can potentiate transcription in a bidirectional manner (*10, 11, 20, 21*).

To understand the relation of these various *cis*-acting regulatory sequences to the cellular transcription machinery that must recognize and interact with them, we previously identified the protein factors responsible for activating SV40 RNA synthesis in a cell-free transcription system (*22, 23*). Fractionation of crude HeLa cell extracts resulted in the identification of a transcription factor, Sp1, that binds specifically to a hexanucleotide sequence, GGGCGG (GC-box), that is tandemly repeated six times in the 21-bp repeats of SV40 (*23, 24*). Recently, Sp1 has been shown to activate transcription and bind to the GC-box sequences present in several other viral and cellular promoters, in-

D. Gidoni, J. T. Kadonaga, and R. Tjian are in the Department of Biochemistry, University of California, Berkeley 94720. H. Barrera-Saldaña, K. Takahashi, and P. Chambon are at LGME/CNRS et U.184/INSERM, Institut de Chimie Biologique, Faculte de Medicine, Strasbourg 67085 Cedex, France. The current address of D. Gidoni is Advanced Genetics Sciences, Inc., 6701 San Pablo Avenue, Oakland, California 94608. The current address of H. Barrera-Saldaña is Departamento Bioquimica, Facultad Medicina, V.A.N.L., Monterrey, N.L., Mexico. The current address of K. Takahashi is Medical Institute for Bioregulation, Kyushu University, Maedashi 3-1-1, Higashi-ku, Fukuoka 812, Japan.

cluding herpesvirus IE-3, IE-4/5, and TK promoters, the AIDS virus long terminal repeat (LTR), two monkey genomic promoters, the mouse dihydrofolate reductase promoter, and human metallothionein promoters (*24–28*). Footprint analysis (protection of DNA from degradation by deoxyribonuclease) of the binding of Sp1 to SV40 DNA indicates that an extended region of approximately 70 bp is protected from deoxyribonuclease I (DNase I) digestion (*23, 24*), and methylation protection experiments showed that Sp1 makes close contacts with several clusters of guanine residues on one strand of the SV40 GC-box elements (*24*). Deletion of these GC-box sequences abolishes the ability of the promoter to respond to Sp1 (*23*). Because the 21-bp repeats exhibit the interesting property of mediating transcription in both the early and late directions, we wanted to further understand how the interaction of the transcription factor Sp1 might lead to bidirectional transcriptional activation.

Accordingly, we performed DNA binding experiments and reconstituted in vitro transcription assays with Sp1 and mutant templates bearing clustered base substitutions in the SV40 promoter to address the following questions concerning the interaction of Sp1 with its recognition sequences. (i) Which specific GC-box elements within the 21-bp repeats are responsible for directing early transcription and which are responsible for late direction transcription? (ii) Does the extended Sp1 binding region revealed by DNase footprinting and dimethyl sulfate methylation protection experiments represent a single binding event or multiple binding events? (iii) If there are tandem but distinct binding sites within the Sp1 footprint region, can we correlate any specific binding sites with the differential activation of SV40 early and late transcription?

GC-box requirements for SV40 early transcription. To identify the Sp1 binding sequences that direct early transcription, we analyzed a series of mutant promoters in a reconstituted in vitro reaction. These altered DNA templates, designated the pSVGC series, were constructed by site-directed mutagenesis of the SV40 transcriptional control region as described in the legend to Fig. 1. Each template contains GGGCGG to GGAAAG base substitutions within either single or multiple GC-box elements (Fig. 1). The in vitro transcription reactions contained active fractions of partially purified HeLa RNA polymerase II and the general transcription fraction, Sp2, along with variable amounts of the Sp1 fraction (*22*). The RNA products synthesized from wild-type and mutant templates were analyzed and quantitated by primer extension with the use of a 5′ end-labeled synthetic oligonucleotide fragment complementary to the SV40 early transcripts (Fig. 1).

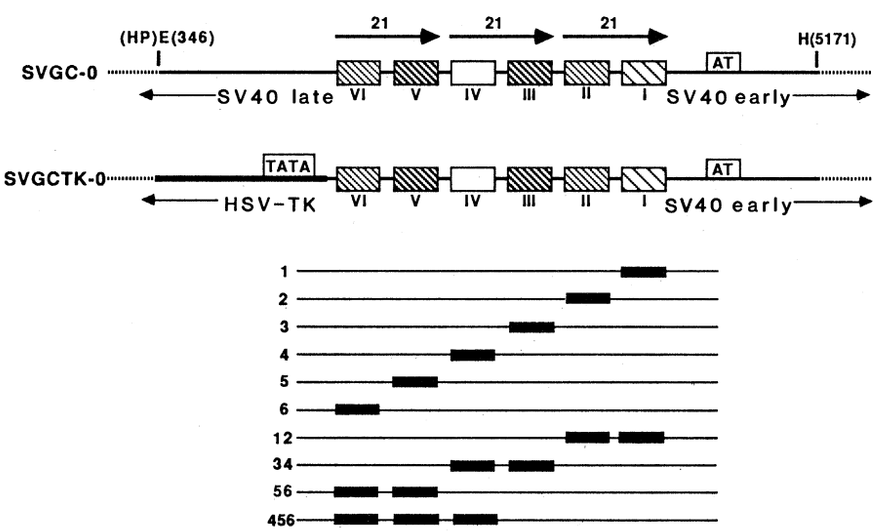


Fig. 1. Schematic diagrams of SV40 templates containing mutations within single or multiple GC-box elements. The site-directed mutations have been constructed by Barrera-Saldaña *et al.* (*37*). The pSVGC-0 wild-type template, which is derived from pSEG0 (*37*), contains the SV40 control elements that reside within the Hpa II–Hind III viral DNA fragment [corresponding to SV40 nucleotides 346 and 5171 (*1*)] cloned into the Eco RI–Hind III polylinker sites of the pUC12 plasmid (*38*). Restriction endonucleases Hpa II, Hind III, and Eco RI are designated as HP, H, and E, respectively [corresponding SV40 nucleotide numbers are according to (*1*)]. The pSVGCTK-0 wild-type template is a promoter fusion construct in which the herpes simplex virus (HSV) thymidine kinase (TK) gene and TATA-box (to nucleotide −32 upstream from the RNA initiation site) was fused in the late orientation, six nucleotides downstream from the SV40 21-bp repeats. The dashed lines represent the pUC12 plasmid, and the thick line designates HSV-TK sequences. The six copies of the sequence GGGCGG (GC-boxes) in the SV40 21-bp repeat region are designated by rectangles. The increasing darkness of the hatched rectangles represents increasing affinity of the GC-boxes for Sp1. GC-box IV is designated by an open rectangle because Sp1 bound at GC-box V appears to prevent efficient binding of the factor to GC-box IV. The solid rectangles indicate GGGCGG to GGAAAG mutations within the GC-boxes. Each mutant template is identified by its respective number corresponding to the mutated GC-box.

As shown previously, transcription of early RNA from the wild-type SV40 template is strongly dependent on the presence of the promoter-specific factor, Sp1 (Fig. 2). This Sp1-dependent transcription is severely affected by mutations in GC-boxes I, II, and III, each of which reduces transcription to approximately 10 to 20 percent of the wild-type activity. In contrast, mutations in GC-boxes IV, V, or VI have very little, if any, detrimental effect upon in vitro transcription from the early promoter (Fig. 2A, lanes a to n, and Table 1). The greatest decrease in early transcription is observed with a double mutant that alters GC-boxes I and II, whereas lesions in boxes V and VI have little effect on transcription, and changes in GC-boxes III and IV produce an intermediate phenotype (Fig. 2A, right panel, lanes o to t, and Table 1). These results suggest that GC-boxes I, II, and III, which are located proximal to the start sites of SV40 early transcription, contain the *cis*-acting elements that mediate early viral transcription.

GC-box requirements for SV40 late direction transcription. To identify the sequence elements within the 21-bp repeats that mediate transcription in the opposite direction of the early RNA, that is, the late direction, we used a set of templates (designated as the pSVGCTK series) that have the 21-bp repeats fused to the TATA-box and coding sequences of a heterologous gene, the thymidine kinase (TK) gene of herpes simplex virus (Fig. 1). Transcription from these hybrid templates is under the control of the SV40 21-bp repeats, while the TK TATA-box element directs the RNA initiation site. The complication of detecting transcripts with multiple initiation sites, as it is normally observed with RNA synthesis in the SV40 late direction, is thus avoided. The primary late direction transcript detected by this assay represents a minor species that originates from nucleotide 170 rather than the major late transcript at 325. As a consequence, our experiments have been designed to measure transcription in the late direction and will not specifically

address many of the unresolved issues concerning late viral transcriptional regulation such as the early-to-late switch.

Sp1 activation of transcription in the late direction was assayed by primer extension with a 5′ end-labeled synthetic oligonucleotide complementary to sequences within the coding region of the TK gene (Fig. 1). Transcription of the hybrid templates was highly dependent on the presence of Sp1 (Fig. 2B, a to j), whereas a control template that contains the TK-TATA box and downstream gene sequences but lacks the Sp1 GC-box binding region does not direct any detectable RNA synthesis. Mutations in GC-box elements III, V, and VI cause the greatest decrease in activity, whereas lesions in GC-boxes I, II, and IV have very little effect on transcription (Fig. 2B and Table 1). In addition to the multiple GC-box mutations shown in Fig. 2B, we have also found a similar pattern of mutant phenotypes with single GC-box mutations, both with SV40-TK fusions as well as bonafide SV40 templates directing late transcription and initiating from nucleotide 170, which were found to be dependent on Sp1 activation in vitro. Moreover, we designed a set of in vitro reactions in which bidirectional transcription is mediated by the 21-bp repeats and transcription from the early and late directions was assayed simultaneously with distinct primers (Fig. 2C). The results confirm that GC-boxes I and II are most important for activating early transcription, whereas GC-boxes V and VI play a major role in late direction transcript, and GC-box III appears to have a secondary effect on transcription in both directions. These in vitro transcription results suggest that individual GC-box sequences within the 21-bp control elements have a differential effect on the bidirectional transcription of SV40 early and late direction RNA.

Table 1. Effects of site-directed mutations within the Sp1 binding sequences on bidirectional transcription. The levels of Sp1-dependent transcription from the promoter mutant series (described in Fig. 1 and indicated by their respective numbers) are expressed as the percentage of their corresponding wild-type (pSVGC-0 and pSVGCTK-0) SV40 promoter transcription levels.

Template	Early mRNA (%)	Late mRNA (%)
0*	100	100
1	21	95
2	24	73
3	11	25
4	80	90
5	83	63
6	75	16
12	5	51
34	20	26
56	46	8
456	30	2

*Wild type.

Binding of Sp1 to the 21-bp repeats. The involvement of specific GC-box elements in activating bidirectional SV40 transcription prompted us to examine more directly the interaction of Sp1 with the individual GC-box sequences. We have therefore carried out deoxyribonuclease (DNase) I footprinting experiments with each of the mutant templates to see if the Sp1 binding region could be subdivided into distinct binding sites that correlated with the differential activation of early and late transcription. With relatively high concentrations of Sp1, the wild-type template exhibited a distinct region protected from DNase digestion (footprint) that extends for 70 bp and covers the entire length of the three tandem 21-bp repeats (Fig. 3, panel 0). This footprint pattern is similar to that which we previously observed (*23, 24*), except that the higher concentrations of Sp1 in these experiments gave a slightly larger footprint because GC-boxes I and II were protected more completely. Mutations in GC-boxes I, II, or V prevent binding of Sp1 to the altered GC-box sequences (Fig. 3, panels 1, 2, and 5). Base substitutions in GC-box III eliminate protection of both GC-boxes III and IV (Fig. 3, panel 3). In contrast, alterations in GC-boxes IV and VI appear to have very little effect on the binding of Sp1, and the footprint pattern is essentially indistinguishable from that of wild-type DNA (Fig. 3, panels 4 and 6). As expected, the double sets of mutations in GC-boxes I and II, III and IV, and V and VI each eliminate the binding of Sp1 to

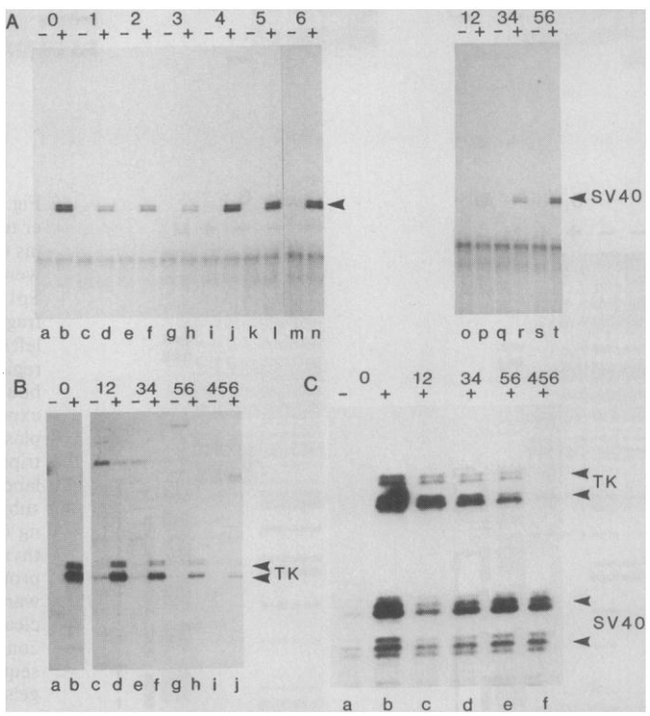


Fig. 2. Analysis of Sp1 directed SV40 transcription in the early and late directions. (A) Mutational analysis of transcription in the early direction. The pSVGC-0 (wild type) template and its related GC-box mutant series (described in Fig. 1 and indicated above each lane in the autoradiogram) were subjected to reconstituted in vitro transcription assays followed by primer extension analysis as described below. To detect SV40 early RNA, a specific complementary oligonucleotide was used (see Fig. 1). The primer extension bands indicating specific SV40 early RNA start sites are designated by arrowheads. (B) Mutational analysis of transcription in late direction. The pSVGCTK promoter-fusion templates (described in Fig. 1 and identified above each lane in the autoradiogram) were subjected to in vitro transcription experiments as described below. To detect late directed RNA synthesis, the transcripts were hybridized to a HSV-TK RNA specific primer (see Fig. 1) and extended as described below. The primer extension bands indicating specific TK RNA (transcribed in the late direction of the SV40 promoter) are marked by arrowheads. (C) Simultaneous mutational analysis of transcription in both early and late directions. The pSVGCTK mutant promoter series were subjected to in vitro transcription reconstitution assays. Transcripts were concomitantly annealed with both TK and SV40 early-specific primers, extended and visualized as described below. The primer extension products indicating accurate TK and SV40 transcripts are designated by arrowheads. The Sp1 was partially purified by fractionation of a human (HeLa) whole cell extract (*22*). Each transcription reaction contained 150 ng of template DNA, endogenous RNA polymerase II (25 μg of total protein) and the general transcription factor Sp2 [2.4 μg of total protein (*22, 23*)]. The promoter specific factor Sp1 was either omitted (−) or added (+) in 5 ng of protein (10 μl) per reaction. The transcription reaction was allowed to proceed for 40 minutes at 30°C and a single-stranded, end-labeled (either SV40 or TK) specific primer (Fig. 1) was hybridized at 53° or 60°C, respectively, for 1 hour and extended with AMV reverse transcriptase for 45 minutes at 37°C. The final products were subsequently analyzed on denaturing 8 percent polyacrylamide gels and visualized by autoradiography.

the altered sequences without affecting the interaction at adjacent binding regions (Fig. 4, panels 12, 34, and 56). The triple mutant with base substitutions in GC-boxes IV, V, and VI gives a footprint pattern that is identical to the double mutant in GC-boxes V and VI, which is in accord with the previous observation that a lesion in GC-box IV has no phenotype. These findings suggest that the Sp1 footprint region is comprised of at least three binding sites with different affinities for the factor and that the interaction of Sp1 with these sites is somewhat complex.

Although the DNase footprint studies provided strong evidence for multiple independent Sp1 domains within the SV40 21-bp repeats, we were not able to discern either the internal structure of the binding or the details of the sequence requirements for the Sp1 interaction at these sites. We have therefore carried out dimethyl sulfate (DMS) methylation protection studies with various GC-box mutant DNA's to investigate, at higher resolution, the interaction of Sp1 with its recognition sequences. Protection of specific guanine residues from methylation at the N7 position by the binding and close proximity of Sp1 was determined as described (*24*). In accord with our previous observations (*24*), wild-type DNA exhibits strong protection of GC-boxes III and V, intermediate protection of GC-boxes II and VI, weak protection of GC-box I, and very weak, if any, protection of GC-box IV (Fig. 5, panel 0). Mutations in GC-boxes I, II, III, IV, or VI prevent binding of Sp1 to the altered GC-box sequences (Fig. 5, panels 1 to 4 and 6). Interestingly, mutations in GC-box V prevent binding of Sp1 to altered sequences but increase protection of GC-box IV (Fig. 5, panel 5). Also, under the conditions used in these studies, there is no evidence for cooperative binding of Sp1 to adjacent GC-box sequences. These methylation protection experiments confirm that there are multiple independent binding sites for Sp1 within the SV40 early promoter and suggest that each of the six GC-boxes is capable of binding Sp1 and

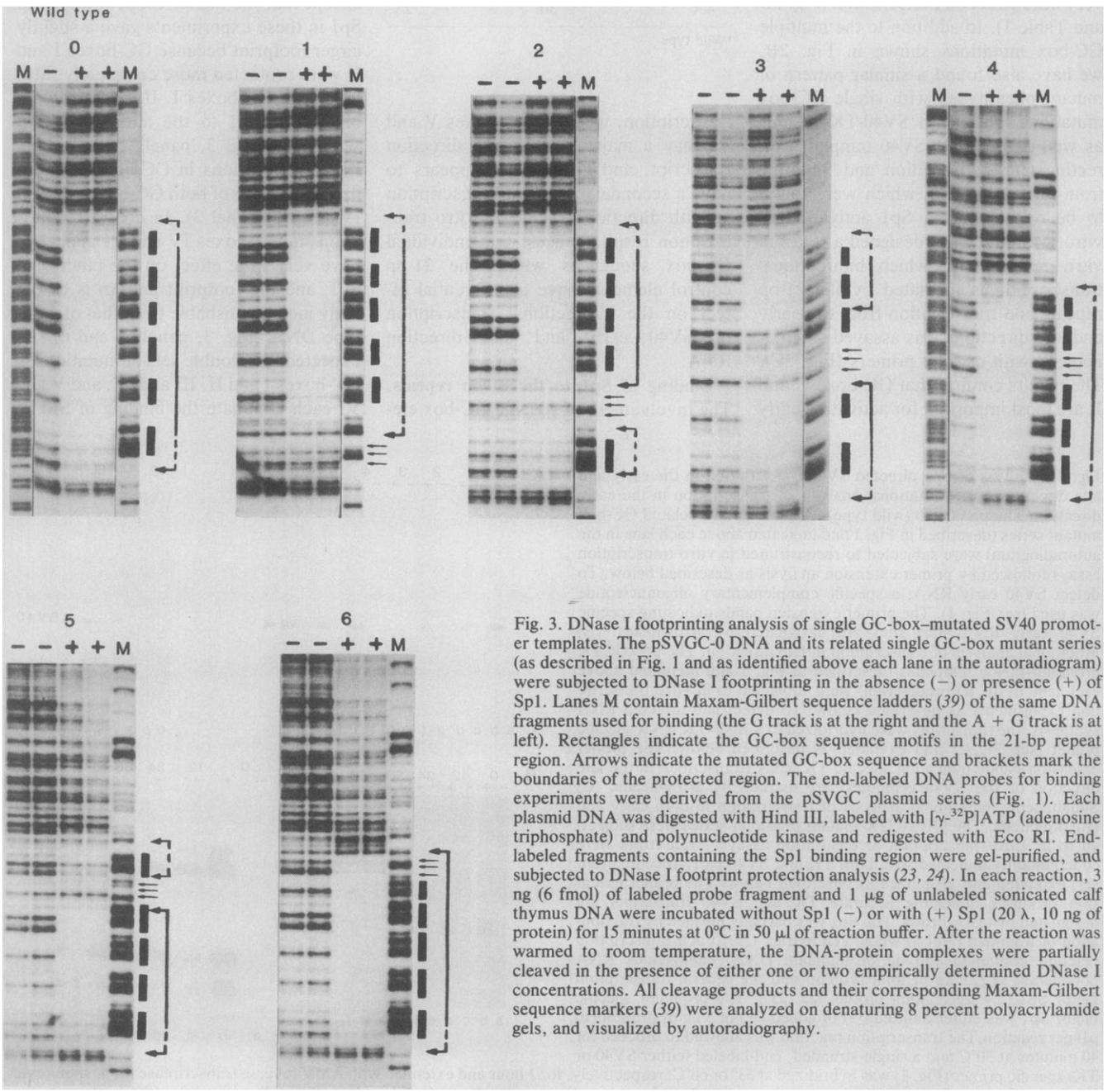


Fig. 3. DNase I footprinting analysis of single GC-box–mutated SV40 promoter templates. The pSVGC-0 DNA and its related single GC-box mutant series (as described in Fig. 1 and as identified above each lane in the autoradiogram) were subjected to DNase I footprinting in the absence (−) or presence (+) of Sp1. Lanes M contain Maxam-Gilbert sequence ladders (*39*) of the same DNA fragments used for binding (the G track is at the right and the A + G track is at left). Rectangles indicate the GC-box sequence motifs in the 21-bp repeat region. Arrows indicate the mutated GC-box sequence and brackets mark the boundaries of the protected region. The end-labeled DNA probes for binding experiments were derived from the pSVGC plasmid series (Fig. 1). Each plasmid DNA was digested with Hind III, labeled with [γ-^{32}P]ATP (adenosine triphosphate) and polynucleotide kinase and redigested with Eco RI. End-labeled fragments containing the Sp1 binding region were gel-purified, and subjected to DNase I footprint protection analysis (*23*, *24*). In each reaction, 3 ng (6 fmol) of labeled probe fragment and 1 μg of unlabeled sonicated calf thymus DNA were incubated without Sp1 (−) or with (+) Sp1 (20 λ, 10 ng of protein) for 15 minutes at 0°C in 50 μl of reaction buffer. After the reaction was warmed to room temperature, the DNA-protein complexes were partially cleaved in the presence of either one or two empirically determined DNase I concentrations. All cleavage products and their corresponding Maxam-Gilbert sequence markers (*39*) were analyzed on denaturing 8 percent polyacrylamide gels, and visualized by autoradiography.

that the affinity of the factor for the different GC-boxes is variable. An interesting complication is that Sp1 bound to GC-box V appears to prevent efficient binding of the factor to adjacent GC-box IV.

Both DNase footprinting and DMS methylation protection experiments suggest that each of the six tandem GC-boxes in the SV40 21-bp repeats is an Sp1 binding site. On the other hand, studies on the herpes simplex virus (HSV) IE-3 promoter have shown that Sp1 can bind to isolated, single GC-box sequences and activate transcription (*26*). If one protomer of Sp1 binds to one GC-box sequence, then it is expected that the Sp1 footprint of an isolated 21-bp repeat sequence, which is a double GC-box binding site, will be significantly larger than a footprint of a single GC-box site. This possibility was tested as follows. An Sp1 binding site that contains one 21-bp repeat (which has GC-boxes III and IV) was chemically synthesized, cloned into a plasmid vector, and then used as a probe for DNase footprinting with Sp1. The footprint observed with this isolated double GC-box binding site was roughly 33 to 35 bp (Fig. 4, panel 2GC), which is almost twice as large as the 18- to 20-bp footprints that have been characterized with the five HSV IE-3 Sp1 binding sites that contain a single GC-box (*26*). Thus, it seems likely that a protomer of Sp1 interacts with one GC-box.

In summary, DNA binding studies of Sp1 on the SV40 21-bp repeats suggest that the 70-bp footprint region consists of six closely spaced binding sites, GC-boxes I-VI, each of which can interact individually with a protomer of Sp1. Efficient binding to GC-box IV appears to occur only when binding to GC-box V is prevented by mutation of the GGGCGG hexanucleotide. As a consequence, for wild-type DNA, there can probably be, at most, five Sp1 protomers bound to the six GC-boxes.

Bidirectional activation of transcription by Sp1. Activation of transcription from the SV40 early and late promoters involves a complex set of specific protein-DNA interactions. A number of studies have identified the *cis*-acting elements that mediate SV40 promoter recognition, and recent biochemical fractionation of transcription extracts has allowed the isolation of a cellular factor, Sp1, that selectively activates SV40 transcription through specific binding to the viral promoter (*22, 23*). The initial DNase footprinting experiments revealed that an approximately 70-bp region containing the three 21-bp repeat elements and six GC-box sequences is specifically recognized and bound by Sp1 (*23, 24*). In addition, DMS methylation protection studies showed that within the Sp1 binding region, there are multiple clusters of guanine residues in the major groove of DNA that are in close contact with Sp1 (*24*). We now report that the Sp1 binding region in SV40 is comprised of six independent binding sites, each of which contains a single GC-box element that is capable of binding a protomer of Sp1. Also, a maximum of five Sp1 protomers appears to be able to bind simultaneously to the six GC-boxes, and the affinities for the different sites vary considerably. For example, GC-boxes III and V have the strongest affinity for Sp1, GC-boxes II and VI are intermediate, GC-box I is weak, and GC-box IV is essentially not accessible to the factor when GC-box V is occupied (Fig. 6). Transcription of mutant templates bearing clustered base substitutions in the GC-box elements indicates that interaction of Sp1 with three of the binding sites, GC-boxes I, II, and III, is largely responsible for mediating early transcription whereas binding to

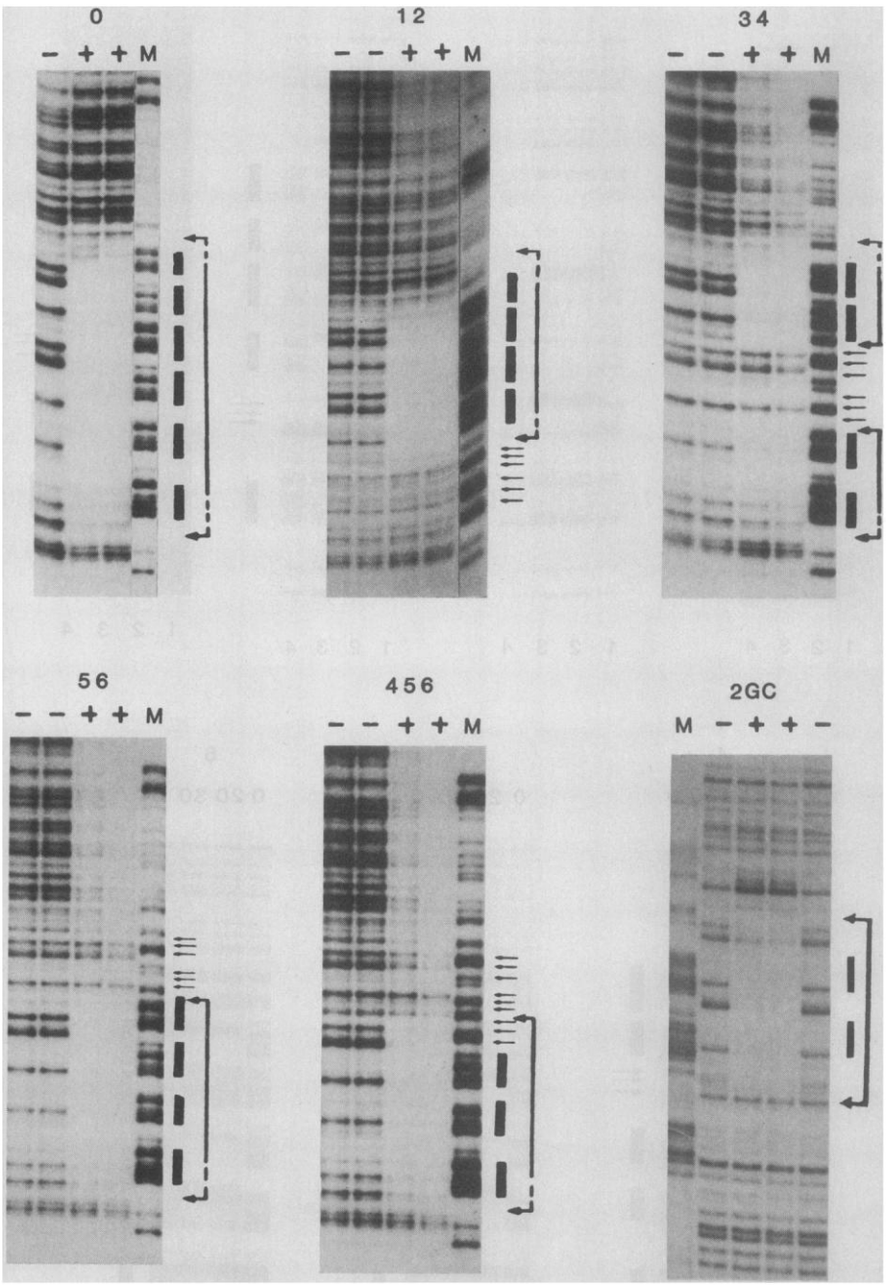
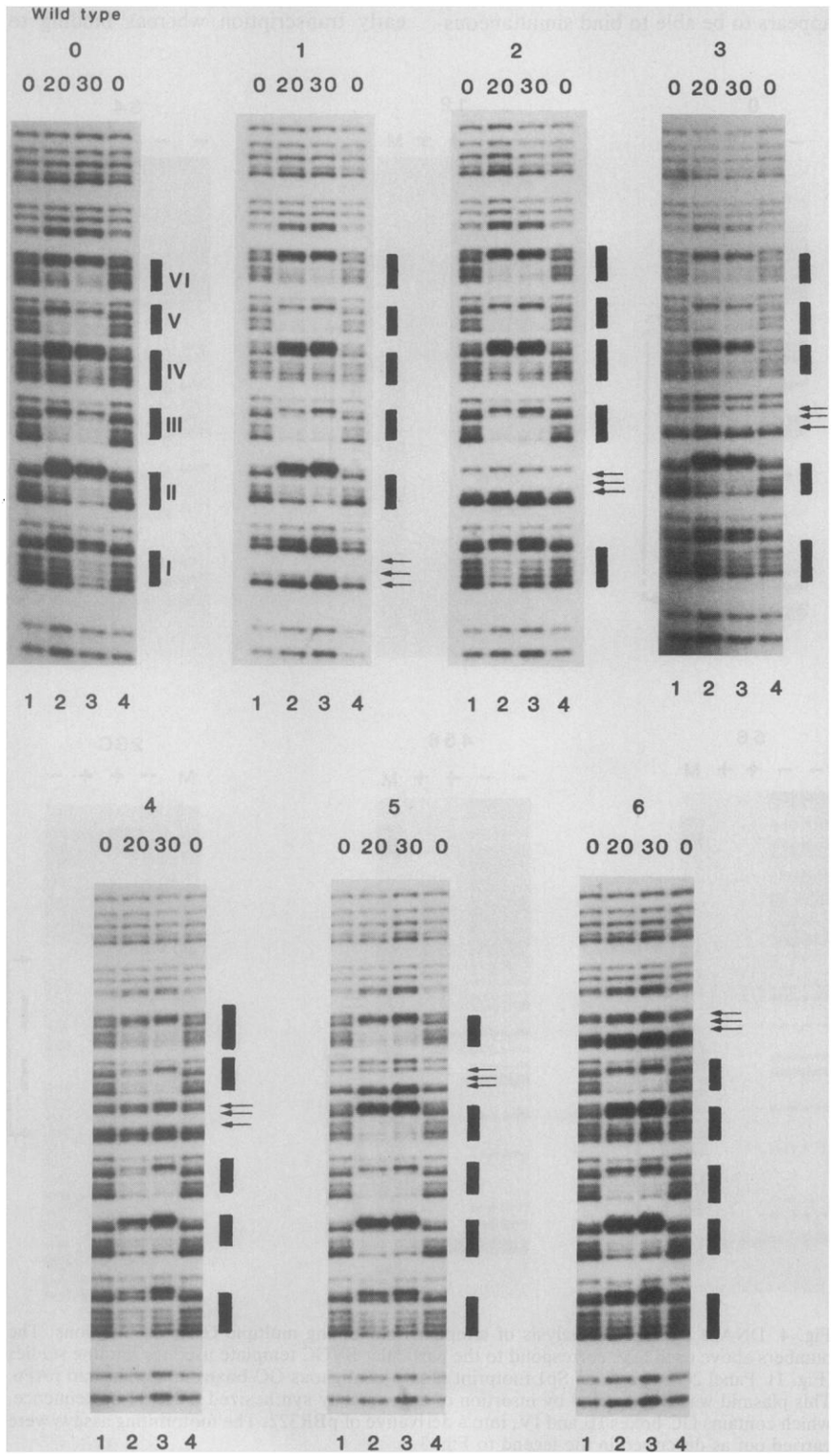


Fig. 4. DNAse I footprint analysis of templates containing multiple GC-box mutations. The numbers above each lane correspond to the particular SVGC template used for binding studies (Fig. 1). Panel 2GC shows an Sp1 footprint of two contiguous GC-boxes in the plasmid pAP6. This plasmid was constructed by insertion of a chemically synthesized SV40 21-bp sequence, which contains GC-boxes III and IV, into a derivative of pBR322. The footprinting assays were carried out as described in the legend to Fig. 3.

GC-boxes III, V, and VI predominantly affects late direction transcription. Thus, as shown in Fig. 6, SV40 bidirectional transcription in vitro appears to be differentially affected by the interaction of a positive transcription factor at five separate binding sites.

A comparison of the Sp1 footprint on single GC-box binding sites (18 to 20 bp) with the footprints on double GC-box sites (33 to 35 bp) suggests that a protomer of Sp1 can interact with a single GC-box. These experiments were carried out with both partially purified Sp1 as well as nearly homogeneous Sp1 [estimated 95 percent purity (*29*)]. Also, the DNase footprint pattern of the wild-type SV40 21-bp repeats is identical when using either the partially purified or nearly homogeneous Sp1 (*29*) and strongly confirms that Sp1 is responsible for the DNA binding properties described here.



The idea that a single GC-box can constitute an Sp1 binding site has been independently confirmed by footprint analysis and transcription studies of several other viral and cellular promoters, including the herpes simplex virus immediate-early genes IE-3 and IE-4/5 (*26*) and a delayed-early herpes gene, thymidine kinase (*27*); the mouse dihydrofolate reductase gene (*28*); human metallothionein genes (*30*); and the major LTR promoter of the AIDS virus (*31*). In all of these promoters, upstream elements contain multiple Sp1 binding sites, and in several cases a region containing only one GC-box element is sufficient to specify a strong binding site for Sp1. A comparison of the available Sp1 binding sites that have been studied thus far also suggests that the GC-box element, which has been defined as the hexanucleotide, GGGCGG, represents only a portion of a longer decanucleotide consensus sequence, $^{G}_{T}$GGGCGG$^{GGC}_{AAT}$ (*32*).

The interaction of Sp1 with GC-box IV is somewhat complicated. First, DNase footprint studies showed that mutation of GC-box III eliminates Sp1 binding to both GC-boxes III and IV (see Fig. 3, panel 3). Second, in DMS methylation protection experiments, efficient binding to GC-box IV was observed only when binding to GC-box V was prevented by mutation (compare panel 0 with panel 5 in Fig. 5). The simplest interpretation of these results is that Sp1 bound to GC-box V (a high affinity site) prevents efficient binding of the factor to GC-box IV (a very low affinity site) due to steric constraints. In support of this hypothesis, the Sp1 recognition sequences in GC-boxes IV and V, as determined from the decanucleotide consensus sequence described above, overlap by one nucleotide, whereas the recognition sequences that contain the other GC-boxes in the

Fig. 5. Dimethyl sulfate methylation protection of SV40 promoter templates containing single GC-box mutations. The numbers above each lane correspond to the particular SVGC templates used for binding (Fig. 1). Each reaction contained either 0, 20 μl, or 30 μl of Sp1 fraction as indicated. Rectangles, the GGGCGG sequence motifs in the SV40 21-bp repeat region. Arrows, GC-boxes converted to GGAAAG by site-directed mutagenesis (Fig. 1). In all methylation protection experiments, end-labeled DNA probe and unlabeled carrier DNA were incubated with Sp1 as described in the legend to Fig. 3. Partial cleavage of N-7 guanine residues was carried out by incubation of the protein-DNA complexes with 1 to 4 μl of dimethyl sulfate for 1 minute at 0°C (*24*). The reactions were terminated according to the Maxam-Gilbert sequencing procedure (*39*), and all samples were analyzed on a denaturing 8 percent polyacrylamide sequencing gel.

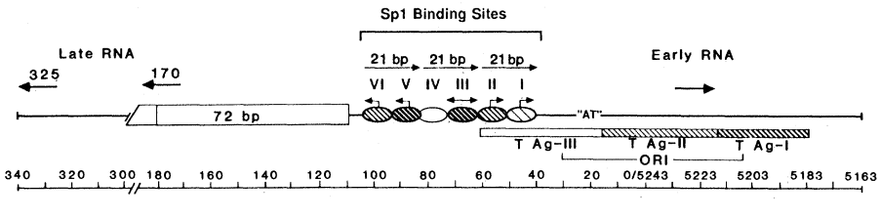


Fig. 6. Binding of Sp1 to the SV40 21-bp repeats. The 21-bp repeats and flanking SV40 sequences are shown and numbered according to (*1*). The six GC-boxes in the 21-bp repeats are designated I to VI, and the direction of transcription that is affected by each Sp1 binding site is indicated by the arrows that are immediately above the GC-boxes. Sp1 protomers are shown as hatched ovals, and the increasing darkness of the lines represents increasing affinity of Sp1 for the GC-boxes. An open oval is at GC-box IV because Sp1 bound to GC-box V appears to prevent efficient binding of the factor to GC-box IV. The binding sites for large T antigen (T Ag), the origin of replication (ORI), and a portion of the 72-bp repeated elements are also shown.

21-bp repeats do not overlap. Also, the in vitro transcription data on the wild-type and mutant GC-box IV templates (see Table 1) can be rationalized within the framework of this model. First, mutation of GC-box IV does not significantly reduce the level of transcription from either the early or late directions, which is consistent with the negligible binding of Sp1 to GC-box IV. Second, transcription in the late direction from the triple GC-box IV-V-VI mutant is much lower than transcription from the double GC-box V-VI mutant. Although this latter result appears to be inconsistent with the previous finding, it is, in fact, exactly what is expected if efficient binding of Sp1 to GC-box IV occurs only when GC-box V is mutated because the factor will bind to the GC-box V-VI mutant but not to the GC-box IV-V-VI mutant. Third, the surprisingly slight reduction in the level of late direction transcription caused by mutations in GC-box V (Table 1) could be the consequence of increased binding of Sp1 to GC-box IV due to the absence of the factor at GC-box V. Also, these data suggest that binding of Sp1 to GC-box IV can have an effect on transcription for the mutant template, even though it is probably not required in the wild-type situation.

The discovery of six closely spaced but apparently independent Sp1 binding sites provides information that influences our view of how the bidirectionality of these promoter elements may actually function. For example, it seems evident that the binding site proximal to the start site of transcription has the greatest influence in activating RNA synthesis for a given direction of transcription. Thus, the functional bidirectionality of the 21-bp repeats could actually be due to multiple directional elements. Alternatively, each unit Sp1 binding sequence could have the ability to activate transcription in an orientation-independent manner with the direction of RNA synthesis determined by the positioning of the Sp1 binding site relative to other promoter elements, such as a TATA-box or other sequences that are required for initiation of transcription. Although GC-boxes I and VI are weak Sp1 binding sites, they are important for specifying transcription in the early and late directions of SV40, respectively. A similar arrangement of a weak Sp1 binding site proximal to the initiation site for transcription has been observed in the HSV-TK (thymidine kinase) promoter, in which the first distal element contains a GC-box sequence that binds Sp1 very weakly (*27*) but has a large effect on transcription in vivo and in vitro (*27, 33, 34*). The significance of positioning a weak factor binding site proximal to the initiation site is not clear, but a reasonable rationale could be to allow better control in modulating transcriptional efficiency.

Both DNase footprinting and DMS methylation studies indicate that the six GC-boxes in the SV40 promoter have different affinities for Sp1. Interestingly, GC-boxes I and II, which are weak Sp1 binding sites, overlap the SV40 T antigen binding site (*2*). It is conceivable that T antigen and Sp1 compete for the same sequences and that autoregulation of SV40 early transcription could, in part, be mediated by the binding of T antigen to site III, which could prevent Sp1 from binding to GC-boxes I and II. Our previous experiments indicated, however, that autoregulation by T antigen is largely dependent on the interaction of the repressor with T antigen binding sites I and II (*3, 35, 36*), which do not overlap with the binding sites for Sp1, and that efficient repression of early transcription occurs at concentrations of T antigen significantly below those required for effective binding to T antigen site III. We do not understand all of the protein-protein and protein-DNA interactions that occur in the SV40 promoter region, and it will be particularly interesting to study the possible relation between Sp1 and other proteins and transcription factors that may be involved in the mechanism of switching from early to late transcription during lytic SV40 infection.

References and Notes

1. J. Tooze, *Molecular Biology of Tumor Viruses*, (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., ed. 2, 1981).
2. R. Tjian, *Cell* **26**, 1 (1981).
3. R. M. Myers, D. C. Rio, A. K. Robbins, R. Tjian, *ibid.* **25**, 373 (1981).
4. C. Benoist and P. Chambon, *Nature (London)* **290**, 304 (1981).
5. D. J. Mathis and P. Chambon, *ibid.*, p. 310.
6. P.K. Ghosh, P. Lebowitz, R. J. Frisque, Y. Gluzman, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 100 (1981).
7. M. Fromm and P. Berg, *J. Mol. Appl. Genet.* **111**, 457 (1982).
8. P. Lebowitz and P. K. Ghosh, *J. Virol.* **41**, 449 (1982).
9. U. Hansen and P. A. Sharp, *EMBO J.* **2**, 2293 (1983).
10. D. Baty, H. Barrera-Saldaña, R. D. Everett, M. Vigneron, P. Chambon, *Nucleic Acids Res.* **12**, 915 (1984).
11. M. Vigneron, H. Barrera-Saldaña, D. Baty, R. D. Everett, P. Chambon, *EMBO J.* **3**, 2373 (1984).
12. J. Banerji, S. Rusconi, W. Schaffner, *Cell* **27**, 299 (1981).
13. P. Gruss, R. Dhar, G. Khoury, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 943 (1981).
14. P. Moreau *et al.*, *Nucleic Acids Res.* **9**, 6047 (1981).
15. M. Fromm and P. Berg, *Mol. Cell. Biol.* **3**, 991 (1983).
16. P. K. Ghosh, V. B. Reddy, J. Swinscoe, P. Lebowitz, S. M. Weissman, *J. Mol. Biol.* **126**, 813 (1978).
17. S. W. Hartzell, B. J. Byrne, K. N. Subramanian, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 23 (1984).
18. J. Brady, M. Radonovich, M. Thoren, G. Das, N. P. Salzman, *Mol. Cell. Biol.* **4**, 133 (1984).
19. D. C. Rio and R. Tjian, *J. Mol. Appl. Genet.* **2**, 423 (1984).
20. R. D. Everett, D. Baty, P. Chambon, *Nucleic Acids Res.* **11**, 2447 (1983).
21. D. Gidoni and R. Tjian, unpublished results.
22. W. S. Dynan and R. Tjian, *Cell* **32**, 669 (1983).
23. ———, *ibid.* **35**, 79 (1983).
24. D. Gidoni, W. S. Dynan, R. Tjian, *Nature (London)* **312**, 409 (1984).
25. W. S. Dynan, J. Saffer, W. S. Lee, R. Tjian, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 4915 (1985).
26. K. A. Jones and R. Tjian, *Nature (London)* **317**, 179 (1985).
27. K. A. Jones, K. R. Yamamoto, R. Tjian, *Cell* **42**, 559 (1985).
28. K. A. Jones, J. Kadonaga, P. Luciw, R. Tjian, unpublished data; W. Dynan, S. Sazer, R. Tjian, R. Schimke, unpublished data; W. Lee, M. Karin, W. Dynan, R. Tjian, unpublished data.
29. J. T. Kadonaga and R. Tjian, unpublished data.
30. W. Lee, M. Karin, W. Dynan, R. Tjian, unpublished data.
31. K. Jones, J. Kadonaga, P. Luciw, R. Tjian, unpublished data.
32. J. T. Kadonaga, K. A. Jones, R. Tjian, *TIBS*, in press.
33. S. L. McKnight and R. Kingsbury, *Science* **217**, 316 (1982).
34. ———, A. Spense, M. Smith, *Cell* **37**, 253 (1984).
35. D. Rio, A. Robbins, R. Myers, R. Tjian, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 5706 (1980).
36. D. Rio and R. Tjian, *Cell* **32**, 1227 (1983).
37. H. Baldera-Saldaña *et al.*, in preparation.
38. J. Vieira and J. Messing, *Gene* **19**, 259 (1982).
39. A. M. Maxam and W. Gilbert, *Methods Enzymol.* **65**, 499 (1980).

40. We thank Kathy Jones for helpful comments on the manuscript, Bill Dynan for assistance with the initial stages of this work, and Karen Ronan for typing the manuscript. This work was funded by grants from the National Cancer Institute to R.T. D.G. was supported by an EMBO fellowship, and J.T.K. is a fellow of the Miller Institute for Basic Research in Science.

8 August 1985; accepted 10 October 1985