JOURNAL OF VIROLOGY, Feb. 1991, p. 565–574
0022-538X/91/020565-10$02.00/0

Vol. 65, No. 2

Herpes Simplex Virus Transactivator ICP4 Operationally Substitutes for the Cellular Transcription Factor Sp1 for Efficient Expression of the Viral Thymidine Kinase Gene

ANTHONY N. IMBALZANO,[1,2] DONALD M. COEN,[3] AND NEAL A. DELUCA[1,2]*

Dana-Farber Cancer Institute[1] and Departments of Biological Chemistry and Molecular Pharmacology[3] and Microbiology and Molecular Genetics,[2] Harvard Medical School, Boston, Massachusetts 02115

Received 24 August 1990/Accepted 22 October 1990

The herpes simplex virus type 1 (HSV-1) ICP4 protein is a transcriptional activator of many eucaryotic RNA polymerase II promoters. The HSV-1 thymidine kinase gene (*tk*) promoter is induced by ICP4 and contains binding sites for the cellular transcription factors TFIID, Sp1, and CCAAT-binding proteins, each of which affects expression of the *tk* gene. In this study, the effects of mutations in these sites on the transcription of *tk* in the presence and absence of ICP4 were determined during viral infection. Only the TATA box was necessary for efficient expression in the presence of ICP4; however, ICP4 apparently can still induce *tk* transcription even when the TATA box is disrupted. Alteration of the Sp1 sites had a minor effect on ICP4-induced expression in comparison to a large effect in the absence of ICP4, indicating that ICP4 can operationally substitute for the function of the transcription factor Sp1. In addition, *tk* was still expressed with the kinetics of an early gene in the absence of binding sites for Sp1 and CCAAT-binding proteins.

Transcription of eucaryotic genes by RNA polymerase II is dependent upon a defined set of general transcription factors (10, 14, 34, 53, 54) and is often modified by additional DNA binding proteins that interact with specific recognition elements near the mRNA start site and cause an increase or decrease in transcription rate (27, 40). The regulation of polymerase II transcription in eucaryotic viruses by *trans*-acting proteins specified by the viral genome has served as a valuable paradigm for understanding eucaryotic gene transcription. The activation of transcription by viral *trans*-acting proteins presumably could involve direct binding to the promoter or the modification of the quantity (51) or net activity (33, 42) of preexisting cellular transcription factors.

Herpes simplex virus (HSV) encodes a 170-kDa phosphoprotein referred to as ICP4, which is critically required for activating transcription from most HSV genes (11, 49, 61). Extensive mutational analysis has identified multiple regions of the ICP4 protein that are associated with its *trans*-acting functions (8, 9, 45, 46, 56). One of these regions confers site-specific DNA-binding properties to the protein. ICP4 was first shown to interact with the sequence ATCGTCNNNNYCGRC, where N is any nucleotide, Y is a pyrimidine and R is a purine, (13, 41) and later shown to have the ability to interact specifically and directly with many different DNA sequences found in or near HSV gene promoters (25, 39). Most mutant ICP4 proteins that fail to bind to DNA in vitro are impaired for transactivation (9, 46, 56). The thymidine kinase gene (*tk*) of HSV type 1 (HSV-1) is readily transactivated by ICP4 in transient transfection assays (5, 7, 15, 43) and in the background of viral infection (29, 31, 48). Several functional and biochemical studies reduce the probability that DNA binding in or near the promoter contributes to ICP4 transactivation of the thymidine kinase gene. There are no relatively strong binding sites having the sequence ATCGTCNNNNYCGRC in the vicinity of the *tk* promoter (13), although weaker binding sites that flank the promoter have been reported (25). In addition, several studies were unable to detect induction-specific sequences in or near the promoter (2, 4, 12), and it is possible to isolate mutant ICP4 molecules that are unable to interact with ICP4 binding sites flanking the viral thymidine kinase promoter without affecting induction (25).

Possible interactions between ICP4 and specific cellular proteins that interact with the thymidine kinase promoter have not been identified. Initial evaluation of the *tk* promoter in heterologous, uninduced systems led to the identification of four *cis*-acting sequence elements (37) that later were shown to be recognition sites for cellular transcription factors (28). These include a TATA box, a CCAAT box, and two Sp1 sites. Subsequent analyses demonstrated that these same elements promote *tk* expression during viral infection (2, 4, 12). However, the specific contribution of the TATA box, CCAAT box, and Sp1 sites in ICP4-mediated induction was not assessed.

The requirement for the recognition sequences for these cellular transcription factors in the *tk* promoter in ICP4-mediated *trans* induction during HSV infection was determined in the present study. This was accomplished by inserting *tk* promoter mutations that alter these sites and that have previously been shown to affect *tk* transcription into their homologous positions in the genome of an ICP4-deficient virus. Expression from the altered *tk* promoter was measured and compared with that from the wild-type promoter following infection of Vero cells to observe transcription in the absence of ICP4 and following infection of E5 cells, which provide ICP4 in *trans*, to observe ICP4-induced transcription.

MATERIALS AND METHODS

Virus and cells. Maintenance of Vero and E5 cell lines has been described previously (5, 11). The wild-type strain of HSV-1 and the parent for all mutant viruses is KOS. Strain n12 contains a nonsense mutation in both copies of the ICP4 coding sequence at amino acid 251 (9) and was propagated

* Corresponding author.

on E5 cells as described previously (6, 8). E5 cells are a Vero cell line that express complementing levels of ICP4 during HSV infection (8). n12 viruses containing linker-scanning (LS) mutations in the *tk* promoter are described below.

Plasmids. The following plasmids were used to incorporate *tk* promoter mutations into the n12 background: pLS/ts−29/−18, pLS/ts−95/−85, pLS/ts+5/+15, and pLS/ts−111/−101//−56/−46 (2, 4), pd/ts−111−18, which was generated by ligation of the 4.4-kb *Bam*HI fragment of pLS/ts−111/−101//−56/−46 with the 3.0-kb *Bam*HI fragment of pLS/ts−29/−18, and pd/ts−111−46, which was generated by ligating the 4.4- and 3.0-kb *Bam*HI fragments of pLS/ts−111/−101//−56/−46.

Generation and screening of recombinant viruses. Recombinant viruses were generated by marker transfer (17, 44) and selection for the temperature-sensitive TK phenotype in a manner similar to that described by Coen et al. (4), except that infectious viral DNA was derived from the mutant n12 and E5 cells were used. Plaques resistant to acyclovir at 39°C were screened by Southern blot (58) to confirm the presence of the *tk* promoter mutations. Positive isolates were plaque-purified three times without selection, retested by Southern blot, and used to generate a viral stock.

Isolation of infected-cell RNA. Approximately 5×10^6 confluent Vero or E5 cells were infected at a multiplicity of infection of 10 PFU/cell and were harvested at 6 h postinfection when phosphonoacetic acid (PAA), a DNA synthesis inhibitor, was used or at 4 h postinfection, a time before the onset of viral DNA replication (data not shown). The measured induction ratios were the same when DNA synthesis was inhibited with PAA or when RNA was harvested before the onset of viral DNA synthesis (data not shown). The monolayers were rinsed on ice with cold TBS (140 mM NaCl, 5 mM KCl, 25 mM Tricine [pH 7.4], 0.7 mM $CaCl_2$, 0.5 mM $MgCl_2$), solubilized, and extracted with 3 ml of 4.0 M guanidine isothiocyanate solution (52) per sample (one or two 100-mm petri dishes). The total RNA contained in the lysate was purified by centrifugation through a 2-ml cushion of 5.7 M cesium chloride in a Beckman SW 50.1 rotor for 18 h at 32,000 rpm at 20°C. RNA pellets were resuspended in 0.3 M sodium acetate, ethanol precipitated, resuspended in 0.3 M sodium acetate, and ethanol precipitated again. Following a wash in 70% ethanol, the pellet was resuspended in diethylpyrocarbonate-treated water, and the A_{260} was measured to determine the concentration.

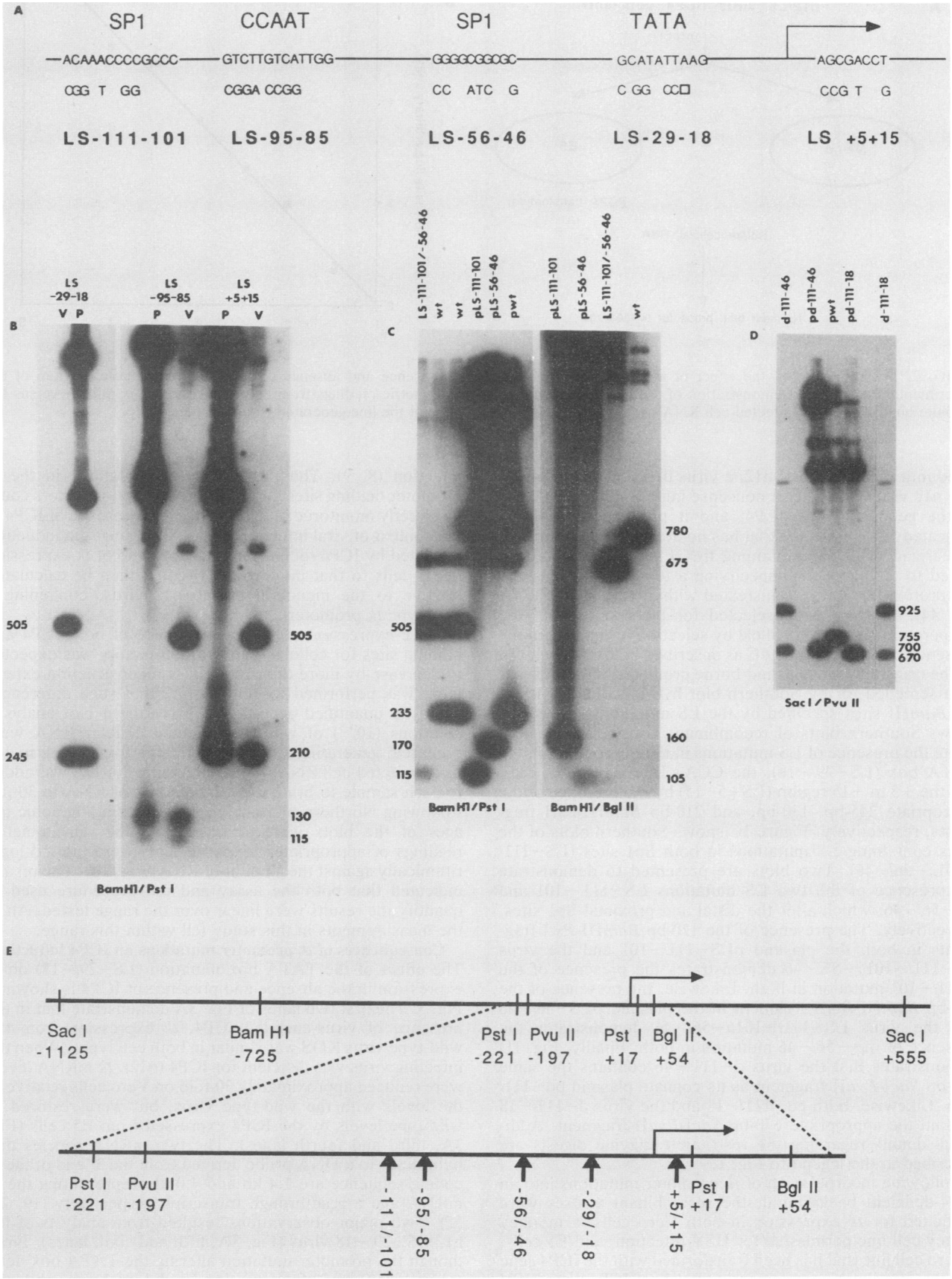
Northern (RNA) blots. Unless otherwise indicated, 20 or 25 μg of total infected-cell RNA was separated by electrophoresis on 1.3% formaldehyde–agarose gels as described before (16, 52) with constant buffer recirculation. Samples were blotted overnight onto nitrocellulose filters in 10× SSC (1× SSC is 0.15 M NaCl, 0.15 M sodium citrate) (52), washed in 4× SSC, and baked under vacuum for 2 h at 80°C. The baked filters were soaked in hybridization solution (50% formamide, 5× SSC, 20 mM Tris hydrochloride [pH 7.5], 1× Denhardt's solution [52], 25 μg of salmon testes DNA per ml, 100 μg of tRNA per ml, and 10% dextran sulfate) at 42°C for 6 h and then hybridized overnight at 42°C in hybridization solution with a ^{32}P-labeled, gel-purified *Sac*I-*Sma*I fragment from the coding region of the *tk* gene. This fragment spans +555 to +1217 relative to the start site of *tk* transcription and was chosen because it does not contain sequences from the overlapping UL24 gene or adjacent gH gene. Filters were washed twice in 2× SSC–0.1% sodium dodecyl sulfate (SDS) for 20 min at room temperature, followed by two 30-min washes in 0.1× SSC–0.1% SDS at 68°C and two 30-min washes in 0.1× SSC–0.1% SDS at 84°C. Dried filters were exposed to preflashed XAR-5 film. Densitometric measurements of appropriately exposed autoradiograms were made with a Hoeffer GS300 densitometer, and the data were processed with compatible Hoeffer software for the Macintosh II computer. All blots were stripped by boiling in distilled H_2O for 15 to 20 min and reprobed for UL42, another early mRNA, as a control. A ^{32}P-labeled, gel-purified *Pst*I fragment internal to the UL42 coding sequence was used as a probe. All probes used in this study were labeled with ^{32}P by nick translation with *Escherichia coli* polymerase I (32).

Primer extension. From 25 to 50 μg of RNA was hybridized with a 25-bp oligonucleotide encoding +51 to +76 of the *tk* noncoding strand. The extension reaction was performed essentially as described by McKnight et al. (36), and extension products were separated by electrophoresis on a 10% denaturing polyacrylamide gel. A 25-bp oligonucleotide encoding +72 to +97 of the noncoding strand of the ICP8 gene was included in some of the reactions as an internal control.

RESULTS

Strategy to determine the effects of *tk* promoter mutations on ICP4-mediated induction. To determine the effect of the *cis*-acting sites and the cellular proteins that bind to them on ICP4-mediated induction, LS mutations (2, 4, 37) and deletion mutations that alter these sites (Fig. 1A) were incorpo-

FIG. 1. Incorporation of *tk* promoter mutations into n12, an ICP4-deficient virus. Shown are Southern blots probed with a plasmid containing wild-type *tk* promoter sequences from −500 to +54. (A) Schematic of the *tk* promoter, indicating the Sp1 sites, CCAAT box, and TATA box as well as the sequence of the LS mutations of McKnight and Kingsbury (37) that alter each of these sites or a sequence downstream of the mRNA start site that previously was shown to affect transcription (4). An open square indicates that the nucleotide was deleted. (B) *Pst*I-*Bam*HI digest of n12 viruses (lanes V) containing LS mutations in the TATA box, the CCAAT box, and between +5 and +15. The 245-, 210-, and 130-bp bands indicated on the figure correspond to fragments from the *Pst*I site at −221 to the *Bam*HI site marking the LS mutation. The 115-bp band present in the LS−95/−85 lanes represents the fragment from the *Bam*HI site at the mutation to the *Pst*I site at +17. Corresponding fragments from the LS−29/−18 and LS+5/+15 digests were too small to transfer efficiently. The 505-bp bands present in the viral DNA lanes correspond to DNA fragments from the *Pst*I site to a *Bam*HI site at −725. Since the plasmid DNAs (lanes P) do not contain sequences beyond −500, these bands are not present. (C) Genotype of n12 LS−111/−101//−56/−46. The left panel is a *Pst*I-*Bam*HI digest marking the presence of the LS−111/−101 mutation, which alters the distal Sp1 site. The presence of a *Bam*HI site at −106 (LS−111/−101) cleaves the 235-bp *Pst*I fragment of the wild-type (wt) promoter (pwt) into two bands of approximately equal size (115 bp). The right panel shows a *Bgl*II-*Bam*HI digest in which the 105-bp band confirms the presence of the LS−56/−46 mutation, which alters the proximal Sp1 site. (D) Two deletion mutants, d−111/−46, which lacks both Sp1 sites and the CCAAT box, and d−111/−18, which lacks the TATA box as well, were created. Shown here are *Pvu*II-*Sac*I digests, which generate bands that span the deletions (−200 to +555). The 925-bp bands present in the viral DNA lanes correspond to fragments from the *Pvu*II site at −200 to a *Sac*I site at −1125. (E) Relevant restriction sites corresponding to the enzymes used in the above digests relative to the *tk* promoter. The vertical arrows show the positions of the *Bam*HI sites specified by the indicated LS mutations.


A
SP1
CCAAT
SP1
TATA
ACAAACCCCGCCC
GTCTTGTCATTGG
GGGCGGCGC
GCATATTAAG
AGCGACCT
CGG T GG
CGGA CCGG
CC ATC G
C GG CC□
CCG T G
LS-111-101
LS-95-85
LS-56-46
LS-29-18
LS +5+15
B
LS -29-18 V P
LS -95-85 P V
LS +5+15 P V
505
245
505
210
130
115
BamH1/Pst I
C
LS-111-101/-56-46
wt
wt
pLS-111-101
pLS-56-46
pwt
pLS-111-101
pLS-56-46
LS-111-101/-56-46
wt
505
235
170
115
780
675
160
105
BamH1/Pst I
BamH1/Bgl II
D
d-111-46
pd-111-46
pwt
pd-111-18
d-111-18
925
755
700
670
Sac I/Pvu II
E
Sac I -1125
Bam H1 -725
Pst I -221
Pvu II -197
Pst I +17
Bgl II +54
Sac I +555
Pst I -221
Pvu II -197
-111/-101
-95/-85
-56/-46
-29/-18
+5/+15
Pst I +17
Bgl II +54

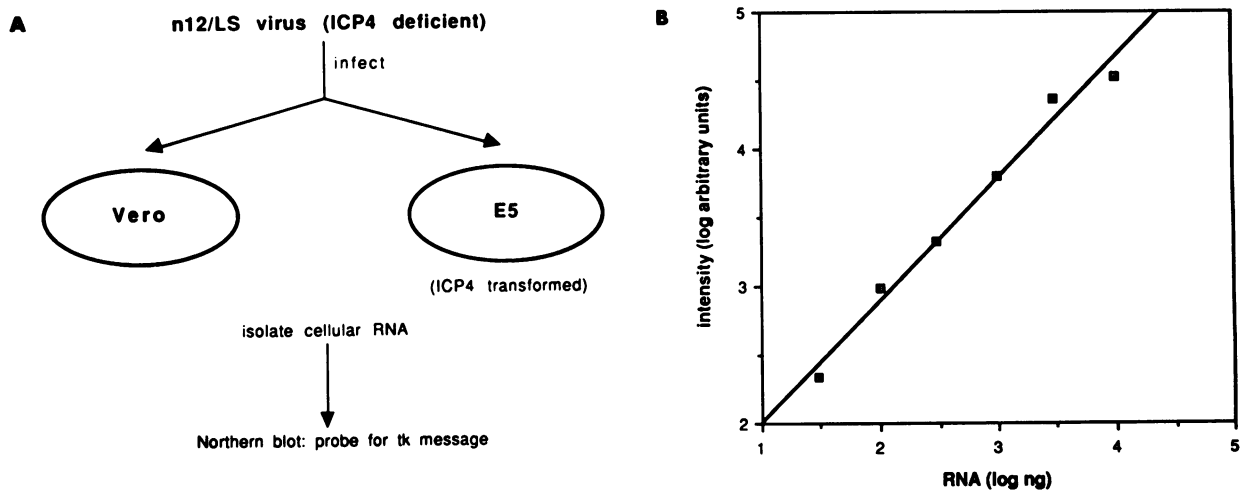


FIG. 2. Determination of the effect of *tk* promoter mutations in the presence and absence of ICP4. (A) Schematic diagram of the experimental design. (B) Demonstration of the linear relationship of densitometric readings from autoradiograms of *tk* mRNA versus the absolute quantity of total infected-cell RNA used for the analysis. The slope of the line generated is approximately 1.0.

rated into the *tk* locus of n12, a virus deficient for ICP4 (9). The n12 virus possesses a nonsense mutation in both copies of the gene encoding ICP4, and it produces a severely truncated ICP4 molecule that has no detectable function (9). Linearized plasmids containing the *tk* promoter mutations linked to a *tk* sequence specifying a temperature-sensitive TK protein (4) were cotransfected with infectious n12 DNA (17, 44). Progeny were selected for incorporation of the temperature-sensitive *tk* allele by selection for growth in the presence of acyclovir at 39°C as described by Coen et al. (4).

The transfer of the plasmid-borne promoter mutations was then screened for by Southern blot hybridization to detect the *Bam*HI sites specified by the LS mutations. Figure 1B shows Southern blots of recombinant viruses that demonstrate the presence of LS mutations in the *tk* promoter in the TATA box (LS−29/−18), the CCAAT box (LS−95/−85), and the +5 to +15 region (LS+5+15) by the presence of the appropriate 245-bp, 130-bp, and 210-bp *Bam*HI-*Pst*I fragments, respectively. Figure 1C shows Southern blots of the virus containing LS mutations in both Sp1 sites (LS−111/−101//−56/−46). Two blots are presented to demonstrate the presence of the two LS mutations LS−111/−101 and LS−56/−46, which alter the distal and proximal Sp1 sites, respectively. The presence of the 120-bp *Bam*HI-*Pst*I fragments in both the plasmid pLS−111/−101 and the virus LS−111/−101//−56/−46 demonstrates the presence of the −111/−101 mutation in both. Likewise, the presence of the 105-bp *Bam*HI-*Bgl*II fragment in the plasmid pLS−56/−46 and the virus LS−111/−101//−56/−46 demonstrates the presence of the −56/−46 mutation in both. Finally, Fig. 1D demonstrates that the virus d−111/−46 contains the same 700-bp *Sac*I-*Pvu*II fragment as its cognate plasmid pd−111/−46. Likewise, both pd−111/−18 and the virus d−111/−18 contain the appropriate 670-bp *Sac*I-*Pvu*II fragment. Additional details regarding the restriction enzyme digests are described in the legend to Fig. 1.

Following incorporation of *tk* promoter mutations into an ICP4-deficient background, the recombinant viruses were evaluated for *tk* expression in both Vero cells, a monkey kidney cell line permissive for HSV infection, and E5 cells, a Vero cell line that has been transformed with the ICP4 gene and that produces complementing levels of ICP4 upon HSV infection (8, 9). Thus, the effect of mutations in the *tk* promoter binding sites for cellular transcription factors could be directly monitored in the absence and presence of ICP4 in the context of viral infection (Fig. 2A). Changes in induction mediated by ICP4 (defined here as the ratio of *tk* expression in E5 cells to that in Vero cells) could then be calculated relative to the induction of the n12 virus, containing a wild-type *tk* promoter.

Since expression of *tk* in the absence of both ICP4 and binding sites for cellular transcription factors was expected to decrease by more than 100-fold, a reconstruction experiment was performed to demonstrate that such differences could be quantified accurately by Northern blot analysis. Dilutions ($10^{0.5}$) of wild-type-infected cellular RNA were prepared, generating samples containing from 30 μg to 100 ng of infected-cell RNA. Mock-infected-cell RNA was added to each sample to bring the total amount of RNA to 30 μg. Following Northern blot analysis, the autoradiographic images of the blots were scanned, and the densitometric readings of appropriately exposed films were plotted logarithmically against the amount of RNA (Fig. 2B). The results indicated that both the assay and the procedure used to quantify the results were linear over the range tested. All of the measurements in this study fell within this range.

Consequences of *tk* promoter mutations on ICP4 induction. The effect of the TATA box mutation (LS−29/−18) on *tk* expression in the absence and presence of ICP4 is shown in Fig. 3. The first two lanes of Fig. 3A demonstrate that in the presence of virus-encoded ICP4, *tk* expression from the wild-type virus KOS was similar in both cell types. When the infecting virus was deficient for ICP4 (n12), *tk* mRNA levels were reduced approximately 30-fold on Vero cells relative to the levels with the wild-type virus, but were restored to wild-type levels by the ICP4 expressed from E5 cells (Fig. 3A, third and fourth lanes). The two mRNA species that hybridized to a DNA probe derived from the 3′ end of the *tk* coding sequence are 1.4 kb and 4.0 kb, representing the *tk* mRNA and a readthrough transcript, respectively (19, 20, 55). Two major observations resulted from analysis of the n12 LS−29/−18 virus (Fig. 3A, fifth and sixth lanes). First, though the promoter mutation altering the TATA box deleteriously affected *tk* expression, it did so by roughly the

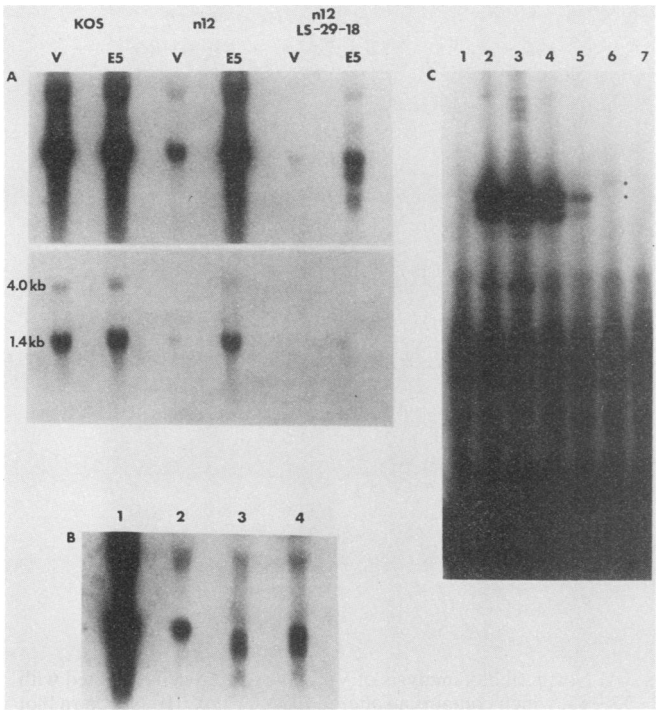


FIG. 3. Effect of the TATA box on ICP4-induced *tk* expression. (A) Northern blot analysis was performed on 25 μg of RNA from Vero (V) or E5 cells infected with n12 LS−29/−18. The top panel is a longer exposure of the bottom panel. (B) Northern blot analysis of *tk* mRNA levels from 25 μg of RNA from n12-infected E5 cells (lane 1), n12-infected Vero cells (lane 2), n12 d−111/−18-infected E5 cells (lane 3), or n12 LS−29/−18-infected E5 cells (lane 4). (C) Primer extension analysis of n12 LS−29/−18. RNA samples include mock-infected E5 cells (lane 1), KOS-infected E5 (lane 2) and Vero (lane 3) cells, n12-infected E5 (lane 4) and Vero (lane 5) cells, and n12 LS−29/−18-infected E5 (lane 6) and Vero (lane 7) cells. RNA samples were generated following infection in the presence of PAA (300 μg/ml).

same extent in both cell types relative to n12. Thus, the mutation in the TATA box apparently does not greatly alter the *trans*-inducing ability of ICP4. Second, this mutation generated two additional *tk* mRNA species that measured 1.2 kb and 1.0 to 1.1 kb in length. Gross over-exposure of the filter in Fig. 3A showed that these species were present in the n12 LS−29/−18-infected Vero cell lane as well (data not shown), indicating that the occurrence of these messages was a function of the LS mutation and did not depend on the presence of ICP4.

Since the origin of these novel mRNAs was unclear, a virus, n12 d−111/−18, that lacked all four of the *cis*-acting sites was generated (see Materials and Methods and Fig. 1D), and accumulation of *tk* transcripts was examined under induced conditions (Fig. 3B). The expression of the full-length 1.4-kb message decreased, while the expression of the novel mRNA species was unaffected (compare lanes 3 and 4). Expression of the novel mRNAs, then, unlike the full-length transcript, is independent of the upstream sequence elements which mediate binding of Sp1 and CCAAT binding protein and may arise from utilization of TATA-like sequences downstream from the proper mRNA initiation site that are not ordinarily recognized in the presence of the wild-type TATA box. To ascertain whether any of the apparent full-length transcripts observed in n12 LS−29/−18-infected cells were initiated in the vicinity of the proper mRNA start site, primer extension analysis was performed (Fig. 3C). The extension products of 73 and 76 bp are consistent with previous observations for the wild-type *tk* transcript (36, 37). In the n12 LS−29/−18 background, extension products indicative of slightly different initiation sites were observed with RNA from E5 cells only (lane 6). Since the primer extension assay was not as sensitive as the Northern blots, this analysis was not effective for the detection of *tk* transcription in the absence of both ICP4 and the TATA box.

The contribution of the CCAAT box and the Sp1 sites to the level of ICP4-induced expression in shown in Fig. 4A. The alteration in the CCAAT box (LS−95/−85) gave rise to an approximately three- to fourfold decrease in *tk* expression in both cell types, and therefore its effect on ICP4 induction was minimal. Similar conclusions were reached for the effect of an LS mutation spanning +5 to +15 (data not shown and Fig. 5), a region shown previously to affect *tk* transcription levels in infected cells (4). In contrast, the level of *tk* expression was not affected equally in the presence and absence of ICP4 when both Sp1 sites were mutated. In the presence of ICP4, *tk* mRNA expressed from n12 LS−111/−101//−56/−46 was reduced only 2- to 3-fold, while in the absence of ICP4 the same mutation resulted in an approximately 20-fold relative reduction (Fig. 4A and 5). The level seen in the presence of ICP4 is consistent with previous measurements of *tk* expression from double Sp1 LS mutants in the background of wild-type virus (2). The levels of *tk* mRNA in n12 LS−111/−101//−56/−46-infected E5 and Vero cells resulted in an induction ratio approximately 10-fold greater than that observed for the wild-type promoter (Fig. 5B). Therefore, the requirement for the Sp1 sites, and probably Sp1, is much greater in the absence of ICP4 than in its presence. The levels of absolute expression and induction by ICP4 are summarized in Fig. 5.

The contribution of all of the upstream elements of the *tk* promoter (two Sp1 sites and CCAAT box) to the level of ICP4-induced expression was examined further by using the virus n12 d−111/−46 (see Materials and Methods and Fig. 1D). In this virus, both of the Sp1 sites and the CCAAT sequence are deleted from the *tk* promoter, leaving only the TATA box and untranslated leader. *tk* expression from this virus in the presence of ICP4 was similar to that from the virus n12 LS−111/−101//−56/−46 (Fig. 4B). Therefore, in the absence of binding sites for Sp1 and CCAAT binding proteins, the TATA box by itself appears to be sufficient for ICP4-induced expression to levels approaching 30% of that of the wild-type promoter.

Kinetics of *tk* expression in the absence of upstream elements. The expression of *tk* from mutant templates altered in the sites for upstream factors in the absence of viral DNA synthesis indicates that the mutant promoters retained the early kinetics characteristic of the wild-type promoter. To confirm this observation, the kinetics of expression of *tk* from the wild type, LS−111/−101//−56/−46, and d−111/−46 were determined by Northern blot and primer extension in the presence and absence of PAA. Assuming that promoter mutations do not differentially affect mRNA stability, comparisons of the amount of *tk* mRNA present in infected cells at different times postinfection reflect the relative activities of the promoters as a function of time. Expression levels of another early gene, that for ICP8, were monitored simultaneously as a control in the primer extension analysis (Fig. 6 and 7). In both the presence and absence of viral

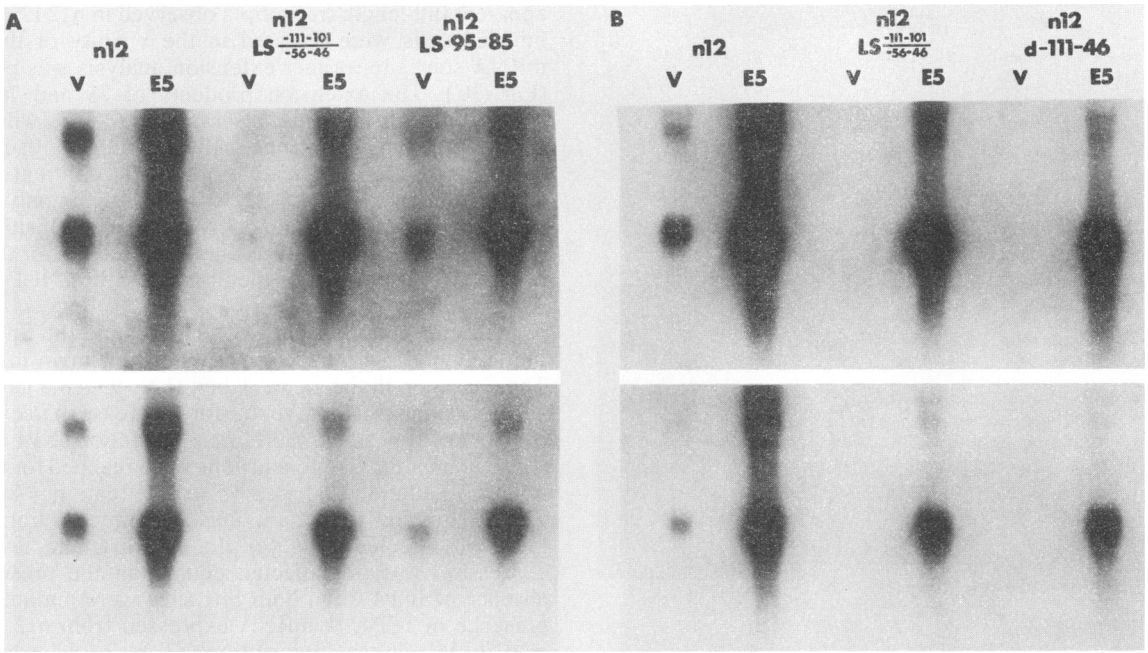


FIG. 4. ICP4 induction in the absence of Sp1 sites and the CCAAT box. (A) Northern blot analysis of Vero (V) and E5 cells infected with n12 LS−111/−101//−56/−46, which contains altered Sp1 sites, and n12 LS−95/−85, which contains an altered CCAAT box. (B) Northern blot analysis of n12 d−111/−46-infected Vero and E5 cells. The top panels are longer exposures of the bottom panels. RNA was extracted from infected cells at 4 h postinfection.

DNA synthesis, when both Sp1 sites were altered or when all three upstream binding sites were deleted, *tk* mRNA was detected at levels ranging from 15 to 30% of that in the wild type. The observations that all the mutants tested expressed *tk* in the absence of DNA synthesis and that the accumulation of *tk* mRNA declined demonstrate that all the promoters retain the properties of an early gene. These data indicate that the binding sites for the upstream factors Sp1 and

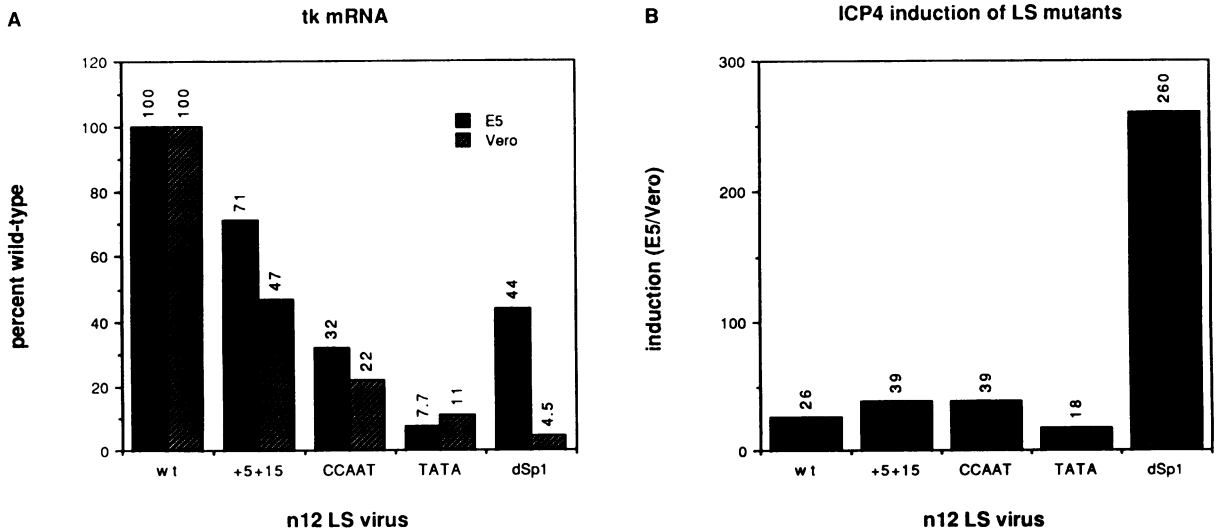


FIG. 5. Levels of expression from *tk* promoter mutants in the presence and absence of ICP4. (A) Relative expression of *tk* from n12 and n12 viruses containing LS mutations in the *tk* promoter in E5 (solid bars) and Vero (hatched bars) cells. Expression of the n12 virus in each cell type was normalized to 100%. To obtain levels of *tk* expression in Vero cells relative to expression in E5 cells, divide the value shown above the Vero bars by 26, the value for ICP4 induction of the n12 virus with a wild-type *tk* promoter. Data represent the average from two experiments. (B) Resulting induction ratios for each virus. Induction is defined as the ratio of the *tk* mRNA level in E5 cells to that in Vero cells. Samples indicated are n12 (wt), n12 LS+5/+15 (+5+15), n12 LS−95/−85 (CCAAT), n12 LS−29/−18 (TATA), and n12 LS−111/−101//−56/−46 (dSp1). The filters used to obtain the above values were stripped by boiling in distilled water for 15 min and reprobed with a ^{32}P-labeled *Pst*I fragment internal to the coding region of the UL42 gene (also an early gene) as a control to normalize for small variations in gel loading and effective multiplicity of infection (data not shown). The levels of UL42 RNA never varied by more than 20%.

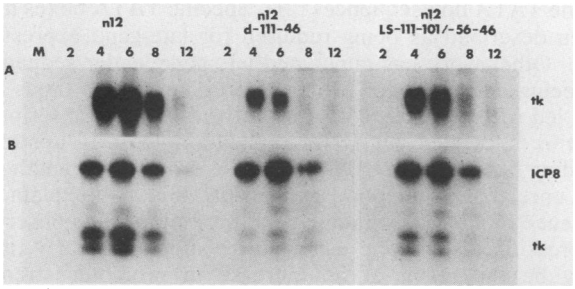


FIG. 6. Time course of *tk* expression in E5 cells infected with n12 or n12 viruses lacking the Sp1 sites (LS−111/−101//−56/−46) or both the Sp1 sites and the CCAAT box (d−111/−46). The numbers across the top of the figure indicate the time postinfection (hours) that the RNA was extracted. (A) Expression of *tk* as monitored by Northern blot. (B) Expression of *tk* as monitored by primer extension analysis. mRNA levels of ICP8, another early gene, were measured as an internal standard.

CCAAT binding proteins are not necessary for expression of *tk* as an early gene.

DISCUSSION

The requirement for binding sites for cellular transcription factors and possibly for the involvement of the proteins that interact with them in the ICP4-mediated induction of the *tk* promoter was evaluated in the context of viral infection. These studies indicate that induction of *tk* expression by ICP4 is independent of the individual *cis* sites that interact with the factors TFIID, Sp1, and CCAAT binding protein. In addition, the requirement for the Sp1 sites was greatly reduced by the presence of ICP4, indicating that ICP4 can operationally substitute for Sp1. Furthermore, the sites for Sp1 and the CCAAT binding proteins were not required for the induction of *tk* as an early gene of HSV.

Effect of upstream DNA-binding activating proteins on ICP4 induction. Given the readily inducible nature of n12 LS−111/−101//−56/−46 and n12 d−111/−46, it appears that Sp1 and CCAAT binding protein are not required for induction of *tk* by ICP4, as was inferred previously (2, 4).

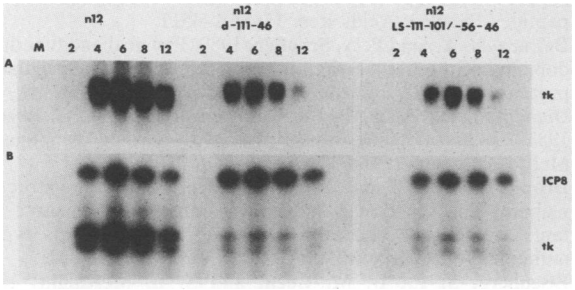


FIG. 7. Time course of *tk* expression in E5 cells infected with n12 or n12 viruses lacking the Sp1 sites (LS−111/−101//−56/−46) or both the Sp1 sites and the CCAAT box (d−111/−46) in the absence of DNA synthesis. Monolayers of E5 cells were infected with the indicated virus in the presence of PAA (400 μg/ml). It was previously determined that PAA at 300 μg/ml was sufficient to completely inhibit viral DNA synthesis. The numbers across the top of the figure indicate the time postinfection (hours) that the RNA was extracted. (A) Expression of *tk* as monitored by Northern blot. (B) Expression of *tk* as monitored by primer extension analysis. ICP8 mRNA levels were monitored as an internal standard.

However, the relative contributions of the Sp1 sites and the CCAAT box differ depending upon the presence or absence of ICP4. This is readily seen in the data presented in Fig. 4A, which shows that similar levels of *tk* mRNA accumulated in the presence of ICP4 with either n12 LS−111/−101//−56/−46 (double Sp1 mutant) or n12 LS−95/−85 (CCAAT mutant); however, accumulation in the absence of ICP4 was greatly reduced with n12 LS−111/−101//−56/−46 compared with n12 LS−95/−85. Since the mutation in the CCAAT box had the same effect on expression in both the presence and absence of ICP4, it can be inferred that the CCAAT box is probably not involved in the mechanism of ICP4 induction. With a similar method of analysis, the same conclusions were reached for the contribution of Sp1 in the mechanism of E1A activation of the adenovirus E1B promoter. In that study, mutation of the E1B Sp1 site resulted in similar decreases in both induced and uninduced expression (62).

In contrast, it appears that ICP4 can operationally substitute for Sp1, in that absolute *tk* expression in the absence of Sp1 sites is only reduced by a factor of two to three in the presence of ICP4, compared to more than 20 in the absence of ICP4 (Fig. 5). This could occur if the mechanisms of ICP4 and Sp1 function involve interactions with common proteins. It has been shown recently that Sp1 functions via interaction with an as yet unidentified protein that also contacts TFIID (47, 50). Such proteins have been termed coactivator or adaptor proteins (47, 50). A possible interpretation of the above data is that ICP4 may interact with the Sp1 adaptor or TFIID. The larger induction ratio observed in the absence of Sp1 sites indicates that ICP4 may be more proficient for transactivation of the *tk* promoter in the absence of Sp1 bound at its recognition site. ICP4 and Sp1 (or its adaptor) need not interact with the same surface of a common protein, as shown by the observation that their action is not totally exclusive, indicating that both may be capable of functioning simultaneously on a single promoter. It is also possible to entertain a more general model in which the interactions between ICP4 and the transcriptional machinery bypass or reduce the requirement for upstream factors.

An alternative interpretation of the observed results is that Sp1 may be directly modified in the presence of ICP4 so that it stimulates transcription of *tk* in the apparent absence of DNA binding to the promoter. In addition, the above results indicate that the Sp1 sites are not required for induction by ICP4 but do not eliminate the possibility that ICP4 can act with or through Sp1, provided that ICP4 can interact with or modify Sp1. This may account for the ability of ICP4 to induce HSV early genes, which frequently contain Sp1 binding sites, to a greater extent than it induces HSV late genes, which lack such sites.

Requirement for the TATA box in ICP4-mediated induction. Alteration of the TATA box in the *tk* promoter appears to generally lower expression efficiency without affecting ICP4 inducibility. This result differs from that obtained by similar genetic studies with adenovirus, which examined the effect of a TATA mutation on E1A activation of the E1B promoter. Those studies showed that E1B transcription was reduced by disruption of the TATA box but that the remaining transcription was not inducible by E1A (62). While this result may indicate that ICP4 can function independently of TFIID, it remains possible that the residual *tk* transcription observed in n12 LS−29/−18 is promoted by TFIID despite the absence of its wild-type recognition site. Recent work showing that the transcription of TATA-less promoters still depends on TFIID supports this idea (57). In addition, the

fact that ICP4 can operationally substitute for Sp1 suggests that TFIID may be required for ICP4-mediated induction. This would also be consistent with previous biochemical studies of the pseudorabies virus IE protein (1), which has extensive amino acid sequence similarity to the ICP4 protein of HSV (3). These studies indicated that pseudorabies virus IE functions by stabilizing TFIID-promoter interactions (1).

The observations that the upstream regions of the *tk* promoter are dispensable for transactivation suggest that ICP4 functions through interactions with general transcription factors or RNA polymerase II itself. Since ICP4 can operationally substitute for Sp1, one of the general transcription factors that ICP4 functions through may be TFIID. Given that ICP4 is a large, elongated protein having a native molecular mass of 350 kDa (38), it may interact with several factors. Biochemical reconstruction experiments will help clarify these possibilities.

Requirements for *tk* expression as an early gene. Properties of HSV early genes include a requirement for immediate-early protein synthesis and the ability to be expressed in the absence of viral DNA replication (23, 24). The existence of upstream binding sites for cellular transcription factors (Sp1 and CCAAT) in early-gene promoters (21, 59) and the apparent lack of such sites in functional HSV late promoters (21, 22, 26) have suggested a causal relationship between these sites (and presumably the factors that bind to them) and early-gene expression. Further supporting this theory is the observation that a late gene can be expressed in the absence of viral DNA synthesis upon the addition of the upstream region from an early promoter (21, 35).

The results of the present study indicate that in the absence of the upstream Sp1 and CCAAT binding sites, the *tk* promoter not only remained responsive to ICP4 induction but directed a significant level of transcription in the absence of DNA synthesis. Thus, neither induction by ICP4 nor expression as an early gene depends upon the presence of upstream binding sites for cellular transcription factors such as Sp1 or the CCAAT binding protein. Perhaps these factors act to maximize expression of early genes to levels required for viral growth. Prior work has reported that a *tk* promoter that lacks these sites is not inducible by ICP4 and does not promote detectable levels of expression (21). In these experiments, a glycoprotein C (gC) gene was constructed that contained the region between −37 and +52 from the *tk* promoter in place of −144 to +124 of the gC promoter, and its expression was measured by Northern blot at 7 h postinfection. A possible explanation for this discrepancy is shown in Fig. 6A of the present study, which shows that the detection of *tk* RNA synthesized from d−111/−46 by Northern blot is dependent on the time that the RNA was extracted; at 6 h postinfection, significant levels of *tk* RNA were detected, while at 8 h the signal was not detectable. Another possibility is that the context in which a promoter element exists within the viral genome may affect its activity. In contrast to the previous studies that examined gC-*tk* chimeras at the gC locus in the viral genome, the present study examined mutations of the *tk* gene in the context of its natural position in the viral genome.

The data presented herein demonstrate that the *tk* TATA box and surrounding sequences are sufficient for expression of an early gene. Given that a virus containing a deletion from −12 to +189, which removes the cap site and the entire leader sequence of the *tk* promoter, still produces TK (18), it appears that the *tk* TATA box alone may constitute an early-gene promoter. The differences in times of expression of early and late genes may therefore result from differences in the TATA box sequences (21). Specific TATA boxes have been described as being required for late-gene expression (21). Other work has implicated late-gene leader sequences as being necessary for proper expression at late times and has led to the proposal of a negative block to transcription that is alleviated by DNA replication (35). If upstream binding factors serve only to increase promoter strength and are unrelated to temporal expression, this may explain the absence of such sites in late-gene promoters. The observed expression of chimeric early-late promoters at early times may be the result of an increase in promoter strength sufficient to overcome the putative block to late-gene transcription.

ACKNOWLEDGMENTS

We thank Colton Smith and Alyssa Shepard for valuable discussions and for reviewing the manuscript.

This work was supported by Public Health Service grants AI26126 to D. Coen and AI27431 to N. DeLuca.

REFERENCES

1. **Abmayer, S. M., J. L. Workman, and R. G. Roeder.** 1988. The pseudorabies immediate early protein stimulates in vitro transcription by facilitating TFIID:promoter interactions. Genes Dev. **2**:542–553.
2. **Boni, J., and D. M. Coen.** 1989. Examination of the roles of transcription factor Sp1-binding sites and an octamer motif in *trans*-induction of the herpes simplex virus thymidine kinase gene. J. Virol. **63**:4088–4092.
3. **Cheung, A. K.** 1989. DNA nucleotide sequence analysis of the immediate early gene of pseudorabies virus. Nucleic Acids Res. **17**:4637–4646.
4. **Coen, D. M., S. P. Weinheimer, and S. L. McKnight.** 1986. A genetic approach to promoter recognition during trans induction of viral gene expression. Science **234**:53–59.
5. **DeLuca, N. A., M. A. Courtney, and P. A. Schaffer.** 1984. Temperature-sensitive mutants of herpes simplex virus type 1 ICP4 permissive for early gene expression. J. Virol. **52**:767–776.
6. **DeLuca, N. A., A. McCarthy, and P. A. Schaffer.** 1985. Isolation and characterization of deletion mutants of herpes simplex virus type 1 in the gene encoding immediate-early regulatory protein ICP4. J. Virol. **56**:558–570.
7. **DeLuca, N. A., and P. A. Schaffer.** 1985. Activation of immediate-early, early, and late promoters by temperature-sensitive and wild-type forms of herpes simplex virus type 1 protein ICP4. Mol. Cell. Biol. **5**:1997–2008.
8. **DeLuca, N. A., and P. A. Schaffer.** 1987. Activities of herpes simplex virus type 1 (HSV-1) ICP4 genes specifying nonsense peptides. Nucleic Acids Res. **15**:4491–4511.
9. **DeLuca, N. A., and P. A. Schaffer.** 1988. Physical and functional domains of the herpes simplex virus transcriptional regulatory protein ICP4. J. Virol. **62**:732–743.
10. **Dignam, J. D., P. L. Martin, B. S. Shastry, and R. G. Roeder.** 1983. Eukaryotic gene transcription with purified components. Methods Enzymol. **104**:582–598.
11. **Dixon, R. A. F., and P. A. Schaffer.** 1980. Fine-structure mapping and functional analysis of temperature-sensitive mutants in the gene encoding the herpes simplex virus type 1 immediate-early protein Vmw175. J. Virol. **36**:189–203.
12. **Eisenberg, S. P., D. M. Coen, and S. L. McKnight.** 1985. Promoter domains required for expression of plasmid-borne copies of the herpes simplex virus thymidine kinase gene in virus-infected mouse fibroblasts and microinjected frog oocytes. Mol. Cell. Biol. **5**:1940–1947.
13. **Faber, S. W., and K. W. Wilcox.** 1986. Association of the herpes simplex virus regulatory protein ICP4 with specific nucleotide sequences. Nucleic Acids Res. **14**:6067–6083.
14. **Flores, O., E. Maldonado, Z. Burton, J. Greenblatt, and D. Reinberg.** 1988. Factors involved in specific transcription by mammalian RNA polymerase II: RNA polymerase II-associating protein 30 is an essential component of transcription factor

IIF. J. Biol. Chem. **263**:10812–10816.
15. **Gelman, I. H., and S. Silverstein.** 1985. Identification of immediate-early genes that trans-activate the virus thymidine kinase gene. Proc. Natl. Acad. Sci. USA **82**:5265–5269.
16. **Goldberg, D. A.** 1980. Isolation and partial characterization of the *Drosophila* alcohol dehydrogenase gene. Proc. Natl. Acad. Sci. USA **77**:5794–5798.
17. **Graham, F. L., and A. J. Van der Eb.** 1973. A new technique for the assay of infectivity of human adenovirus 5 DNA. Virology **52**:456–457.
18. **Halpern, M. E., and J. R. Smiley.** 1984. Effects of deletions on expression of the herpes simplex virus thymidine kinase gene from the intact viral genome: the amino terminus of the enzyme is dispensable for catalytic activity. J. Virol. **50**:733–738.
19. **Harris-Hamilton, E., and S. L. Bachenheimer.** 1985. Accumulation of herpes simplex virus type 1 RNAs of different kinetic classes in the cytoplasm of infected cells. J. Virol. **53**:144–151.
20. **Holland, L. E., R. M. Sandri-Goldin, A. L. Goldin, J. C. Glorioso, and M. Levine.** 1984. Transcriptional and genetic analysis of the herpes simplex virus type 1 genome: coordinates 0.29 to 0.45. J. Virol. **49**:947–959.
21. **Homa, F. L., J. C. Glorioso, and M. Levine.** 1988. A specific 15-bp TATA box promoter element is required for expression of a herpes simplex virus type 1 late gene. Genes Dev. **2**:40–53.
22. **Homa, F. L., T. M. Otal, J. C. Glorioso, and M. Levine.** 1986. Transcriptional control signals of a herpes simplex virus type 1 late (γ_2) gene lie within bases −34 to +124 relative to the 5′ terminus of the mRNA. Mol. Cell. Biol. **6**:3652–3666.
23. **Honess, R. W., and B. Roizman.** 1974. Regulation of herpes virus macromolecular synthesis. I. Cascade regulation of three groups of viral proteins. J. Virol. **14**:8–19.
24. **Honess, R. W., and B. Roizman.** 1975. Regulation of herpes virus macromolecular synthesis: sequential transition of polypeptide synthesis requires functional viral polypeptides. Proc. Natl. Acad. Sci. USA **72**:1276–1280.
25. **Imbalzano, A. N., A. A. Shepard, and N. A. DeLuca.** 1990. Functional relevance of specific interactions between herpes simplex virus type 1 ICP4 and sequences from the promoter-regulatory domain of the viral thymidine kinase gene. J. Virol. **64**:2620–2631.
26. **Johnson, P. A., and R. D. Everett.** 1986. The control of herpes simplex virus type 1 late gene transcription: a TATA box/cap site region is sufficient for fully regulated activity. Nucleic Acids Res. **14**:8247.
27. **Johnson, P. F., and S. L. McKnight.** 1989. Eukaryotic transcriptional regulatory proteins. Annu. Rev. Biochem. **58**:799–839.
28. **Jones, K. A., K. R. Yamomoto, and R. Tjian.** 1985. Two distinct transcription factors bind to the HSV thymidine kinase promoter in vitro. Cell **42**:559–572.
29. **Kit, S., D. R. Dubbs, and P. A. Schaffer.** 1978. Thymidine kinase activity of biochemically transformed mouse cells after superinfection by thymidine kinase negative, temperature sensitive herpes simplex virus mutants. Virology. **85**:456–463.
30. **Kristie, T. M., and B. Roizman.** 1986. α4, the major regulatory protein of herpes simplex virus type 1, is stably and specifically associated with promoter-regulatory domains of α genes and of selected other viral genes. Proc. Natl. Acad. Sci. USA **83**:3218–3222.
31. **Leung, W.-C.** 1978. Evidence for a herpes simplex virus-specific factor controlling the transcription of deoxypyrimidine kinase. J. Virol. **27**:269–274.
32. **Maniatis, T., A. Jeffrey, and D. G. Kleid.** 1975. Nucleotide sequence of the rightward operator of phage λ. Proc. Natl. Acad. Sci. USA **72**:1184–1188.
33. **Marton, M. J., S. B. Baim, D. A. Ornelles, and T. Shenk.** 1990. The adenovirus E4 17-kilodalton protein complexes with the cellular transcription factor E2F, altering its DNA-binding properties and stimulating E1A-independent accumulation of E2 mRNA. J. Virol. **64**:2345–2359.
34. **Matsui, T., J. Segall, P. A. Weil, and R. G. Roeder.** 1980. Multiple factors required for accurate initiation of transcription by RNA polymerase II. J. Biol. Chem. **255**:11992–11996.
35. **Mavromara-Nazos, P., and B. Roizman.** 1989. Delineation of regulatory domains of early (β) and late (γ_2) genes by construction of chimeric genes expressed in herpes simplex virus 1 genomes. J. Virol. **86**:4071–4075.
36. **McKnight, S. L., E. R. Gavis, R. Kingsbury, and R. Axel.** 1981. Analysis of transcriptional regulatory signals of the HSV thymidine kinase gene: identification of an upstream control region. Cell **25**:385–398.
37. **McKnight, S. L., and R. Kingsbury.** 1982. Transcriptional control signals of a eukaryotic protein-coding gene. Science **217**:316–324.
38. **Metzler, D. W., and K. W. Wilcox.** 1985. Isolation of herpes simplex virus regulatory protein ICP4 as a homodimeric complex. J. Virol. **55**:329–337.
39. **Michael, N., D. Spector, P. Mavromara-Nazos, T. M. Kristie, and B. Roizman.** 1988. The DNA-binding properties of the major regulatory protein α4 of herpes simplex virus. Science **239**: 1531–1534.
40. **Mitchell, P. J., and R. Tjian.** 1989. Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. Science **245**:371–378.
41. **Muller, M. T.** 1987. Binding of the herpes simplex virus type 1 gene product ICP4 to its own transcription start site. J. Virol. **61**:858–865.
42. **Muller, U., M. P. Roberts, D. A. Engel, W. Doerfler, and T. Shenk.** 1990. Induction of transcription factor AP-1 by adenovirus E1A protein and cAMP. Genes Dev. **3**:1991–2002.
43. **O'Hare, P., and G. S. Hayward.** 1985. Evidence for a direct role for both the 175,000- and 110,000-molecular-weight immediate-early proteins of herpes simplex virus in the *trans*-activation of delayed-early promoters. J. Virol. **53**:751–760.
44. **Parris, D. S., R. A. F. Dixon, and P. A. Schaffer.** 1980. Physical mapping of herpes simplex virus type 1 *ts* mutants by marker rescue: correlation of the physical and genetic maps. Virology **100**:275–287.
45. **Paterson, T., and R. D. Everett.** 1988. Mutational dissection of the HSV-1 immediate early protein Vmw175 involved in transcriptional transactivation and repression. Virology **166**:186–196.
46. **Paterson, T., and R. D. Everett.** 1988. The regions of the herpes simplex virus type 1 immediate-early protein Vmw175 required for site-specific DNA binding closely correspond to those involved in transcriptional regulation. Nucleic Acids Res. **16**: 11005–11025.
47. **Peterson, M. G., N. Tanese, B. F. Pugh, and R. Tjian.** 1990. Functional domains and upstream activator properties of cloned human TATA binding protein. Science **248**:1625–1630.
48. **Post, L. E., S. Mackem, and B. Roizman.** 1981. Regulation of α genes of herpes simplex virus: expression of chimeric genes produced by fusion of thymidine kinase with α gene promoters. Cell **24**:555–565.
49. **Preston, C. M.** 1979. Abnormal properties of an immediate-early polypeptide in cells infected with herpes simplex virus type 1 mutant *ts*K. J. Virol. **32**:357–369.
50. **Pugh, B. F., and R. Tjian.** 1990. Mechanism of transcriptional activation by Sp1: evidence for coactivators. Cell **61**:1187–1197.
51. **Saffer, J. D., S. P. Jackson, and S. J. Thurston.** 1990. SV40 stimulates expression of the trans-acting factor Sp1 at the mRNA level. Genes Dev. **4**:659–666.
52. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
53. **Samuels, M., A. Fire, and P. A. Sharp.** 1982. Separation and characterization of factors mediating accurate transcription by RNA polymerase II. J. Biol. Chem. **257**:14419–14427.
54. **Sawadogo, M., and R. G. Roeder.** 1985. Factors involved in specific transcription by human RNA polymerase II: analysis by a rapid and quantitative in vitro assay. Proc. Natl. Acad. Sci. USA **82**:4394–4398.
55. **Sharp, J. A., M. J. Wagner, and W. C. Summers.** 1983. Transcription of herpes simplex virus genes in vivo: overlap of a late-gene promoter with the 3′ end of the early thymidine kinase gene. J. Virol. **45**:10–17.
56. **Shepard, A. A., A. N. Imbalzano, and N. A. DeLuca.** 1989.

Separation of primary structural components conferring autoregulation, transactivation, and DNA-binding properties to the herpes simplex virus transcriptional regulatory protein ICP4. J. Virol. **63**:3714–3728.

57. **Smale, S. T., M. C. Schmidt, A. J. Berk, and D. Baltimore.** 1990. Transcriptional activation by Sp1 as directed through TATA or initiator: specific requirement for mammalian transcription factor IID. Proc. Natl. Acad. Sci. USA **87**:4509–4513.
58. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–517.
59. **Su, L., and D. M. Knipe.** 1987. Mapping of the transcriptional initiation site of the herpes simplex virus type 1 ICP8 gene in infected and transfected cells. J. Virol. **61**:615–620.
60. **Tedder, D. G., R. D. Everett, K. W. Wilcox, P. Beard, and L. I. Pizer.** 1989. ICP4 binding sites in the promoter and coding regions of the herpes simplex virus gD gene contribute to activation of in vitro transcription by ICP4. J. Virol. **63**:2510–2520.
61. **Watson, R., and J. B. Clements.** 1980. A herpes simplex virus type 1 function continuously required for early and late virus RNA synthesis. Nature (London) **285**:329–330.
62. **Wu, L., D. S. E. Rosser, M. C. Schmidt, and A. Berk.** 1987. A TATA box implicated in E1A transcriptional activation of a simple adenovirus 2 promoter. Nature (London) **326**:512–515.