JOURNAL OF VIROLOGY, Nov. 1995, p. 7291–7294 Vol. 69, No. 11
0022-538X/95/$04.00+0

Initiator Elements and Regulated Expression of the Herpes Simplex Virus Thymidine Kinase Gene

W. JAMES COOK,[1] SEAN M. LIN,[1] NEAL A. DELUCA,[2] AND DONALD M. COEN[1]*

Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115,[1] and Department of Molecular Genetics and Biochemistry, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania 15261[2]

Received 15 June 1995/Accepted 28 July 1995

Regulated expression of thymidine kinase mRNA in herpes simplex virus mutants harboring thymidine kinase promoters that lacked functional TATA boxes was largely unaffected by additional sequence alterations around the transcriptional start site. A strong initiator element increased the regulated expression of a TATA-containing promoter by 50% but did not affect that of the TATA-less promoter. Thus, initiator elements exert only small effects in this promoter context.

Expression of certain eukaryotic TATA-less promoters depends on specific sequences known as initiator (inr) elements that surround the mRNA start site (27). Activities attributed to inr elements include the ability to independently direct RNA polymerase II transcription from a specific initiation site and the ability to mediate the activity of at least some upstream activators in the absence of a TATA box (27). In addition, inr elements can cooperate with a TATA box to enhance transcription (23, 27). Sequences encompassing the start site, including some that show sequence similarity to well-characterized inr elements, are important for the expression of certain herpes simplex virus (HSV) genes (14, 15, 17, 21, 26, 29). Because these HSV genes belong to the late kinetic classes, one interesting property that appears to separate late HSV genes from those expressed early is the presence or absence, respectively, of inr-like elements (13–15, 17, 21, 26, 29).

The HSV thymidine kinase gene (*tk*) is expressed as an early gene (1, 25, 32). Like that of most HSV genes, expression of *tk* requires the viral regulatory protein ICP4 (8, 10, 24). One mechanism by which ICP4 induces *tk* involves the *tk* TATA box. Results from multiple studies support a model in which at least part of the ICP4 induction of *tk* involves ICP4 enhancement of TFIID binding to the *tk* TATA box (3, 19, 28). However, when the *tk* TATA box was disrupted with the LS-29/-18 mutation (Fig. 1), which abolishes the binding of the TATA-box binding protein (3, 19), ICP4 still induced *tk* (18). We hypothesized (3) that the mechanism by which the TATA-less LS-29/-18 promoter is induced by ICP4 may involve DNA sequences located near the *tk* transcriptional start site. This was based on results showing that ICP4 induced another TATA-less HSV promoter through an inr element (13). Moreover, *tk* bp +5 to +15 (Fig. 1) were shown to be important for *tk* expression in the presence of ICP4 (2). While mutation of the +5-to-+15 element in the context of the intact *tk* TATA box did not eliminate ICP4 induction (18), it remained possible that this region near the *tk* start site is critical for the induction of the TATA-less LS-29/-18 promoter.

Construction and analysis of double mutations. To test whether we could affect the induction of the LS-29/-18 promoter by mutating sequences that span the *tk* start site, we constructed a series of recombinant viruses that were ICP4 deficient and that also contained the LS-29/-18 TATA box disruption mutation and one of several linker-scanning disruptions of the *tk* sequence from -7 to +36 (2, 22) (Fig. 1). The plasmids pLS/ts-29/-18//-7/+3, pLS/ts-29/-18//+5/+15, and pLS/ts-29/-18//+16/+36+10 were constructed by transferring the 154-bp *Mlu*I-*Mlu*I fragments consisting of *tk* bp -15 to +139 from plasmids pLS/ts-7/+3, pLS/ts+5/+15, and pLS/ts+16/+36+10 into the analogous position in pLS/ts-29/-18 (2). DNA from these three plasmids was cotransfected with infectious HSV DNA from the ICP4-deficient mutant *n*12 (9) into E5 cells (a Vero cell line that expresses complementing levels of ICP4 upon HSV infection [7, 8]) as described previously (18). Selection for temperature-sensitive acyclovir resistance and identification of selected plaques were done as described previously (2). Southern analysis and direct PCR sequencing (11) were performed to verify the presence of only the desired mutations (4). The *n*12 *tk*(Ts) virus (3), which contains the wild-type *tk* promoter, is designated here as the wild type. The *n*12 LS-29/-18 virus has been described previously (18). The sequences of the *tk* core promoters analyzed in this report are shown in Fig. 1.

Levels of *tk* expression and ICP4 induction of *tk* promoters containing the LS mutations were determined by infecting Vero and E5 cells with the recombinant *n*12 viruses and measuring *tk* mRNA levels by Northern (RNA) analysis as described previously (3). In agreement with previous results (18), the LS-29/-18 promoter expressed levels of *tk* mRNA in both the absence and presence of ICP4 that were approximately 10% of that expressed by the wild type, and therefore ICP4 induction of the LS-29/-18 promoter was only slightly lower than that of the wild type (14-fold and 18-fold, respectively) (Fig. 2A and Table 1). In viruses containing the LS-29/-18//-7/+3, LS-29/-18//+5/+15, and LS-29/-18//+16/+36+10 promoters, *tk* mRNA levels in both the absence and presence of ICP4 were similar to that of the LS-29/-18 virus (Fig. 2B and Table 1). The resulting levels of ICP4 induction measured with LS-29/-18//-7/+3 (14-fold), LS-29/-18//+5/+15 (13-fold), and LS-29/-18//+16/+36+10 (12-fold) were indistinguishable from that of LS-29/-18 (14-fold). These results indicate that *tk* sequences spanning the mRNA start site do not function as inr elements and that ICP4 induction of the TATA-less LS-29/-18 promoter is unaffected by disruption of these start site sequences.

Effects of a strong inr element on *tk* expression. To determine if replacing the *tk* mRNA start site in the LS-29/-18

* Corresponding author. Phone: (617) 432-1691. Fax: (617) 432-3833. Electronic mail address: dcoen@warren.med.harvard.edu.

```
                      -31                            +1                         +27
wild-type tk          TCGCATATTAAGGTGACGCGTGTGGCCTCGAACACCGAGCGACCCTGCAGCGACCCGC
LS-29/-18             --C-GG--CC*-----------------------------------------------
LS-29/-18//-7/+3      --C-GG--CC*-------------C-GGATCCGG------------------------
LS-29/-18//+5/+15     --C-GG--CC*------------------------CC-GAT--GG-------------
                                                                           (+36)
LS-29/-18//+16/+36+10--C-GG--CC*-------------------------------------C-GAT--G--

                                                     +1
wild-type tk          TCGCATATTAAGGTGACGCGTGTGGCCTCGAACACCGAGCGACCCTGCAGCGACCCGC
TdT Inr               -----------------------------TC-TT-T-GAGA-----------------
LS-29/-18//TdT Inr    --C-GG--CC*------------------TC-TT-T-GAGA-----------------
consensus Inr                                      YYANTCY
                                                       A
```

FIG. 1. The *tk* promoter mutations used in this study. DNA sequence changes in the LS-29/-18 promoter and the promoters containing the double LS mutations are shown relative to the wild-type *tk* promoter sequence (top group), with dashes representing identical residues. An asterisk denotes a deleted base pair. The +16/+36+10 mutation replaces the *tk* sequence between bp +16 and +36 with a 10-bp *Bam*HI linker (22), and therefore the last dash shown for the LS-29/-18//+16/+36+10 promoter represents *tk* bp +36. Sequence changes in the TdT Inr promoter and the LS-29/-18//TdT Inr promoter are shown relative to the wild-type promoter sequence (bottom group). The consensus inr (20) is shown for purposes of comparison (Y = pyrimidine and N = any base).

promoter with a functional inr element would affect *tk* expression, we inserted the inr of the well-characterized terminal deoxynucleotidyl transferase promoter (TdT Inr) (27) into the *tk* promoter. *tk* bp -2 to +10 were replaced with TdT Inr sequences (Fig. 1) by mutually primed PCR mutagenesis (3, 6). Subcloning of the final PCR fragment was done as described previously (3) to produce p*tk*/tsTdT Inr. The 154-bp *Mlu*I-*Mlu*I

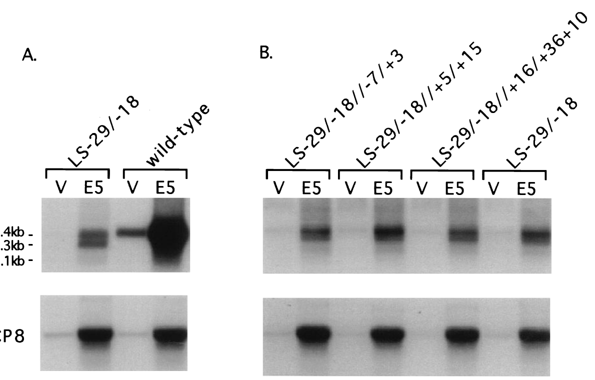


FIG. 2. Disruption of the *tk* mRNA start site does not affect *tk* expression or ICP4 induction of the LS-29/-18 promoter. Northern blot analysis was conducted with 5 μg of total RNA which was harvested at 5 h following the infection of Vero (V) and E5 cells with 10 PFU of the indicated *n*12 viruses per cell. Northern blots were first probed with a ^{32}P-labeled, 0.66-kb *Sac*I-*Sma*I fragment that spanned *tk* bp +555 to +1217, and then the blots were stripped and reprobed for *ICP8* message with a 1.4-kb *Pst*I fragment that spanned the *ICP8* coding sequences (12). The LS-29/-18 mutation was previously shown to only slightly alter the start site of the major 1.4-kb *tk* transcript, with one of two major start sites being the same as that used by the wild type (18). However, viruses containing LS-29/-18 generated additional *tk* mRNA species measuring 1.3 and 1.1 kb that were not originated from the normal *tk* mRNA start site (18). The smaller *tk* transcripts were also visible in LS-29/-18-infected Vero cells upon longer exposures of the autoradiographs (4, 18). Only the 1.4-kb band was quantified in the wild-type virus, while the 1.4- and 1.3-kb doublet bands were quantified together in viruses containing the LS-29/-18 mutation. (A) *tk* mRNA accumulation in the virus containing the LS-29/-18 promoter was compared with that in the virus containing the wild-type *tk* promoter. (B) *tk* mRNA accumulation in the viruses containing the double LS mutations was compared with that in the LS-29/-18 virus. While the *tk* mRNA signals in Vero cells appear faint in this figure, these signals were easily detected upon longer exposures of the autoradiographs and were readily quantified with a PhosphorImager. Results are tabulated in Table 1. This and subsequent figures were generated with the programs Adobe Photoshop and Canvas on a Macintosh computer following the scanning of the autoradiographs with a Microtek scanner.

TABLE 1. Relative *tk* expression and ICP4 induction

*n*12 *tk*(Ts) virus	Relative *tk* mRNA expression in[a]:		Induction[b]
	Vero (without ICP4)	E5 (with ICP4)	
LS-29/-18	1.0	14 ± 1.1	14
LS-29/-18//-7/+3	0.85 ± 0.37	12 ± 2.7	14
LS-29/-18//+5/+15	1.2 ± 0.25	15 ± 2.2	13
LS-29/-18//+16/+36+10	1.1 ± 0.29	13 ± 3.4	12
LS-29/-18//TdT Inr	1.2 ± 0.17	15 ± 0.75	13
Wild type[c]	1.0	18 ± 1.6	18
TdT Inr	1.1 ± 0.17	27 ± 2.7	25

[a] *tk* mRNA expression of viruses relative to the value measured for the LS-29/-18 virus (top group) or the wild type (bottom group) following infection of Vero cells (without ICP4). Values ± standard deviations represent the averages of three to six measurements. All values were normalized to *ICP8* mRNA values as described in the legend to Fig. 2. The values were obtained by quantifying relevant bands on Northern blots with a PhosphorImager (Molecular Dynamics).

[b] Ratio of normalized *tk* mRNA expression in E5 cells (with ICP4) to that in Vero cells (without ICP4).

[c] These values are taken from reference 3.

fragment containing the TdT Inr sequences was transferred from p*tk*/tsTdT Inr to the analogous position in pLS/ts-29/-18 to produce pLS/ts-29/-18//TdT Inr. Recombinant *n*12 viruses containing the *tk*/tsTdT Inr and LS-29/-18//TdT Inr alleles were generated as described above.

Using Northern analysis, we measured *tk* mRNA levels and ICP4 induction in the *tk*/tsTdT Inr virus relative to those of the wild-type *tk* virus and those of the LS-29/-18//TdT Inr virus relative to those of LS-29/-18. In the *tk*/tsTdT Inr virus, the TdT Inr increased *tk* expression reproducibly in the presence of ICP4 (by approximately 50%) but not in the absence of ICP4 (Fig. 3A and Table 1). The resulting ICP4 induction level measured with this allele was 1.4-fold higher than that of the wild type. This suggests that in the presence of ICP4, the TdT Inr cooperates with the functional *tk* TATA box to enhance levels of *tk* expression. In contrast, incorporation of the TdT Inr into the TATA-less LS-29/-18 promoter had no significant effect on *tk* expression in either the absence or presence of ICP4 (Fig. 3B and Table 1). Thus, the resulting ICP4 induction exhibited by the LS-29/-18//TdT Inr promoter was not in-

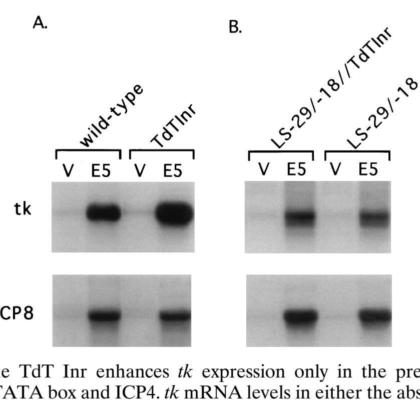
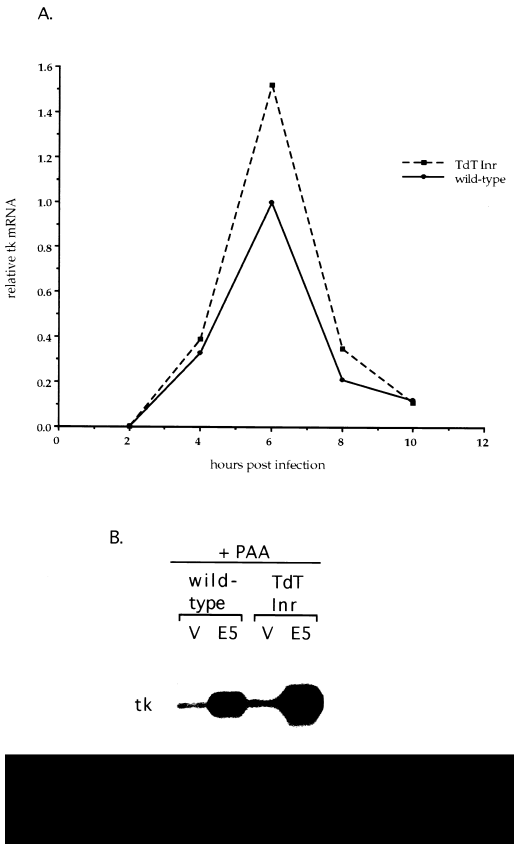


FIG. 3. The TdT Inr enhances *tk* expression only in the presence of the functional *tk* TATA box and ICP4. *tk* mRNA levels in either the absence of ICP4 (Vero [V] cells) or its presence (E5 cells) were measured by Northern analysis as described in the legend to Fig. 2. (A) *tk* mRNA expression in the *tk*/tsTdT Inr virus was compared with that in the virus containing the wild-type promoter. (B) *tk* mRNA expression in the virus containing the LS-29/-18//TdT Inr promoter was compared with that in the LS-29/-18 virus. Results are tabulated in Table 1.



creased relative to that of LS-29/-18 (13-fold and 14-fold, respectively). These findings indicate that the TdT Inr enhances *tk* expression only in the presence of a functional *tk* TATA box and ICP4.

The TdT Inr does not affect the kinetics of *tk* expression. Because inr-like elements are important for the transcription of certain HSV genes that are expressed late in infection (13–15, 17, 21, 26, 29), we investigated whether the insertion of the TdT Inr into *tk* alters the early kinetic profile of *tk* expression so that it resembles late gene kinetics. This was accomplished by comparing the time course of *tk* expression in the *tk*/tsTdT Inr virus with that of the wild type. Total RNA was harvested at time points ranging from 2 to 10 h following the infection of E5 cells and was subjected to Northern analysis. The time course of *tk* mRNA accumulation in the *tk*/tsTdT Inr virus was similar to that of the wild type (Fig. 4A). The rapid decrease in *tk* mRNA accumulation at late times in infection (after 6 h) was observed with both the *tk*/tsTdT Inr and wild-type viruses, which is consistent with the kinetic profile of an early gene.

Transcription of HSV late genes is dependent on viral DNA replication. Therefore, drugs that inhibit DNA replication eliminate or severely reduce expression of late genes (16, 25, 30). To determine if inhibition of viral DNA replication affects expression of the TdT Inr allele, we infected Vero and E5 cells with the *tk*/tsTdT Inr and wild-type viruses in the presence of 0.4 mg of phosphonoacetic acid (PAA) per ml, which inhibited viral DNA replication and late gene expression (5). Relative to that of the wild type, expression of the TdT Inr promoter in the presence of PAA was similar to that measured in the absence of the drug (1.1-fold higher in Vero cells and 1.5-fold higher in E5 cells in both the absence of PAA [Fig. 3A and Table 1] and its presence [Fig. 4B]). The resulting level of ICP4 induction of the TdT Inr promoter in the presence of the drug was therefore similar to that measured in the absence of the drug (1.4-fold higher than that of the wild type in both cases). These results indicate that PAA has no effect on the levels of *tk* expression or on ICP4 induction exhibited by the *tk*/tsTdT Inr allele and thus confirm that the replacement of *tk* bp -2 to +10 with the TdT Inr has no effect on the kinetics of *tk* expression.

The *tk* promoter lacks a functional inr element. The results outlined above suggest that the *tk* promoter lacks a functional inr element. Furthermore, our results suggest that a strong inr is unable to function in the context of the TATA-less *tk* LS-29/-18 promoter. It should be noted that under experimental conditions different from ours, the TdT Inr functioned regardless of the sequences present at the TATA box position (31). Because our assay was conducted with HSV-infected cells, it is possible that viral factors acting at the *tk* start site may cause the TdT Inr to function differently than in an in vitro assay or transfected cells.

Despite the presence of strong inr-like elements in many HSV late genes, we have shown that the presence of the strong TdT Inr did not alter the early kinetics of *tk*, at least in the context of *tk* upstream regulatory sites. Therefore, while the presence of inr-like elements in late genes offers an attractive explanation for aspects of late kinetics, a strong inr by itself is insufficient for imparting late kinetics on the early *tk* gene. Previous studies showed that *tk* was still expressed as an early gene when all distal promoter elements in *tk* were deleted (18) and when the *tk* TATA box was replaced with the TATA box of the late HSV gC gene (19). Furthermore, disruption of the *tk* TATA box with the LS-29/-18 mutation did not alter the early kinetics of *tk* (5). Taken together, these results suggest that the kinetics of *tk* expression are not independently controlled by either the distal *tk* elements or the core promoter elements. One explanation for the lack of *tk* expression late in infection is that *tk* lacks *cis* elements that are present in late genes and that permit transcription to continue at late times. Alternatively, it remains possible that sequences located downstream of *tk* bp +36 control the turnoff of *tk* expression at late times. We are testing this latter possibility.

Mechanism of ICP4 induction. The mechanism by which the TATA-less LS-29/-18 promoter is induced by ICP4 does not involve *tk* sequences that span the mRNA start site (Fig. 2 and Table 1). This indicates that ICP4 retains the ability to induce *tk* even when the two elements that constitute a prototype core promoter, the TATA box and the transcriptional start site, are mutated. Additionally, because ICP4 induction of *tk* does not require distal cellular activator sites (18), induction of *tk* does not appear to depend solely on either specific core promoter sequences or on distal sites. ICP4 activation of *tk* may involve multiple and redundant mechanisms, some of which are independent of *tk* promoter elements. For example, ICP4 may interact with transcription factors that do not make sequence-specific contacts with DNA. These targets might include such factors as TFIIB, TFIIA, and pol II that do not recognize DNA or perhaps even TFIID, which is required for the transcription of TATA-less promoters (33).

We thank David Knipe for the plasmid containing *ICP8* and Anthony Imbalzano and Louane Hann for helpful discussions.