JOURNAL OF VIROLOGY, Dec. 1982, p. 939–949
0022-538X/82/120939-11$02.00/0

Vol. 44, No. 3

Structural Features of the Herpes Simplex Virus α Gene 4, 0, and 27 Promoter-Regulatory Sequences Which Confer α Regulation on Chimeric Thymidine Kinase Genes

SUSAN MACKEM AND BERNARD ROIZMAN

Kovler Viral Oncology Laboratories, The University of Chicago, Chicago, Illinois 60637

Received 2 July 1982/Accepted 12 August 1982

Previous studies have shown that herpes simplex virus genes form three groups, α, β, and γ, whose expression is coordinately regulated and sequentially ordered in a cascade fashion. Chimeric genes constructed by fusion of the coding and 5′ nontranslated leader sequences of the thymidine kinase (TK) gene to the sequences upstream from the site of initiation of transcription of α genes 4 and 27 are regulated as α genes and are induced in cells converted to TK^+ phenotype by infection with TK^- virus. In α gene 4 (S. Mackem and B. Roizman, Proc. Natl. Acad. Sci. U.S.A. 79:4917–4921, 1982), both the promoter and the regulatory region are separable and movable. The promoter permits expression but not induction when fused to TK in the noncoding leader region of the gene. The regulator, when fused to the promoter of an expressible but noninducible portion of the natural β TK, renders the gene inducible as an α gene; it consists of multiple regulatory units acting cumulatively. In this paper, we report on the precise site of initiation of transcription of α gene 0 within the inverted *b* sequences of the L component of viral DNA. We also report the following. (i) The chimeric gene consisting of the coding and 5′ nontranslated leader regions of the TK gene fused to portions of the domain of α gene 0 extending largely upstream from the site of initiation of transcription of α gene 0 was regulated in the same fashion as the α 4- and α 27-TK chimeras. The regulatory region in the α gene 0 is largely upstream from nucleotide −140. (ii) The promoter-regulatory regions of α genes 0, 4, and 27 share TATA sequences, A+T-rich (consensus) sequences occurring in regulating regions of α genes 0 and 4 in more than one copy, and multiple G+C-rich inverted repeats. The relation of these sequences to the function of the promoter-regulatory regions of the α genes is discussed.

The herpes simplex virus 1 (HSV-1) genome consists of at least three groups of genes, α, β, and γ, whose expression is coordinately regulated and sequentially ordered in a cascade fashion (19). To date, five α genes, 0, 4, 22, 27, and 47, have been identified and mapped (1, 20, 24, 33, 38, 45, 46). They are the first set of genes expressed during productive infection (19, 20, 25) and are transcribed by the host RNA polymerase II (7). This study, and earlier reports from this laboratory, centered on the selective recognition and regulation of α gene expression in infected cells. To facilitate the analyses of α gene expression, we constructed a series of chimeric genes produced by fusion of coding and selected noncoding sequences of the TK gene to sequences derived from the domains of the α genes. In the first report (36), we determined that a chimeric gene consisting of the domain of the TK coding and the 5′ noncoding leader sequences up to nucleotide +50, when fused to the sequences contained between the nucleotides +33 to −4,500 of the α gene 4 in the proper transcriptional orientation, behaved like an α gene when recombined into the HSV genome. Specifically, the α-TK gene was transcribed in cells treated with an inhibitor of protein synthesis before and during infection. In cells converted to TK^+ phenotype, the chimeric TK gene was induced by infection with TK^- virus. However, unlike that of the natural, β TK gene, the induction of the α chimeric TK gene by infection with TK^- virus occurred in the absence of protein synthesis and did not require the presence of a functional product of the α gene 4. The sequence, defined as the promoter-regulatory region of α gene 4, which imparted to the TK gene the regulatory properties of the α genes, was located between the nucleotide +33 to −4,500 of the α gene 4.

In subsequent studies (26, 27), we demonstrated the following. (i) Chimeric TK genes contain-

ing promoter-regulatory regions of α gene 27 were regulated like the α 4-TK gene chimeras. In this instance, the promoter-regulatory region was contained between the nucleotides +55 and −270 of the α gene 27. (ii) The promoter-regulatory region of the α gene 4 contains physically separable and movable promoter and regulatory sequences. Thus, chimeric genes containing the α gene 4 sequences extending from +33 to −110, when fused to the domain of the TK gene extending 5′ to nucleotide +50, permitted the expression of the chimeric TK; however, the gene could not be induced. The α gene 4 sequence extending upstream from position −110 conferred α-specific induction when fused to an expressible but noninducible β TK gene extending 5′ to nucleotide −80. This α regulatory region (upstream of −110) required the presence of either an α or a β promoter for its function. (iii) Since the amount of induction observed with α TK chimeras was proportional to the length of α sequences extending upstream from −110, it was concluded that the α 4 regulator region contains multiple regulatory units, which are capable of acting cumulatively.

In this paper, we show that the chimeric genes consisting of appropriate domains of the TK gene fused to the promoter-regulatory region of the α gene 0 cannot be differentiated from the α gene 4- and 27-TK chimeras with respect to the requirements for induction in converted TK^+ cells. We also report on the nucleotide sequences of the promoter-regulatory regions of the α genes 0, 4, and 27. Although the promoter regions show no significant common nucleotide sequences other than the TATA sequence, a consensus sequence and GC-rich inverted repeated sequences are present in all regulatory regions.

MATERIALS AND METHODS

Viruses and cells. The isolation and properties of HSV-1(F), the procedures for the maintenance and growth of cells, and the propagation of virus have been previously described (12, 40). All DNA clones were derived from HSV-1(F), and RNA for all experiments was isolated from Vero cells infected with this strain. B2006 is an HSV-1 TK^- virus (10) obtained from S. Kit. The construction and properties of *ts*502Δ305, temperature sensitive in α gene 4 and containing a 700-base pair (bp) deletion in the thymidine kinase gene, have been previously described (36). Ltk^- cells (21) obtained from S. Kit were used for transformation experiments with chimeric plasmids.

Cloning of DNA. The methods employed for the cloning and purification of cloned DNAs were as previously described (35), except that for ligation of DNA restriction fragments with unmatched protruding terminal nucleotides, the DNAs were first treated with T4 polymerase to generate blunt ends (43).

Isolation and analysis of RNA. The procedures for the isolation and purification of cytoplasmic RNA from Vero cells infected with HSV-1(F) in the presence of cycloheximide were as previously described (24). For 5′ end analysis, 5 μg of polyadenylated [poly(A^+)] RNA from HSV-1-infected cells was hybridized to about 0.003 pmol of specific 5′ end-labeled, strand-separated HSV DNA fragment and then was digested with either S1 nuclease (S1) or exonuclease VII (exo VII), all according to conditions as previously given (26).

Conversion of Ltk^- cells to TK^+ phenotype and assay for TK activity. Ltk^- cells were converted to TK^+ phenotype with purified plasmid DNAs by the calcium-phosphate precipitation procedure (16). TK^+ cells were selected and maintained in HAT medium containing 10^{-5} M hypoxanthine, 1.6×10^{-5} M thymidine, and 4.4×10^{-7} M methotrexate. Procedures for infection of converted Ltk^+ cells and for extraction and assay of thymidine kinase activity accumulating in these cells by using [^{3}H]thymidine as substrate were as previously described (26, 36).

Labeling, strand separation, and sequence analysis of DNA. The procedures for 5′ end labeling cloned DNA fragments, for purification of uniquely labeled fragments by secondary enzyme digestion or strand separation, for base-specific chemical sequencing reactions, and for polyacrylamide sequencing gels were all performed according to the protocol of Maxam and Gilbert (29). The standard reactions used were as follows: G, dimethyl sulfate followed by piperidine treatment; A+G, pyridine formate followed by piperidine treatment; C+T, hydrazine without salt followed by piperidine treatment; C, hydrazine with high salt followed by piperidine treatment; A>C, heating with alkali at 90°C followed by piperidine treatment. All sequences were verified by multiple determinations of the sequence of both strands of the DNA.

RESULTS

Site of initiation of transcription of α gene 0. In previous studies (24), the site of initiation of transcription of α gene 0 was approximately mapped to within a 900-bp *Rsa*I-*Sac*I fragment extending from 600 to 1,500 bp to the right of the *Bam*HI B-SP junction in the orientation shown in Fig. 1. To locate the α gene 0 transcription initiation site more precisely, the 900-bp *Rsa*I-*Sac*I fragment (cloned in pRB336) was end labeled with [γ-^{32}P]ATP and polynucleotide kinase. The purified single strand labeled specifically at the *Rsa*I site was hybridized to poly(A^+)-selected RNA extracted from infected cells, and the hybrids were digested with either S1 or exo VII. The DNA products protected from digestion by S1 and by exo VII comigrated at a position 174 nucleotides from the *Rsa*I site (Fig. 2). This would place the α gene 0 transcription initiation site about 774 bp to the right of the *Bam*HI B-SP junction, in agreement with previous results.

Construction and regulation of α gene O-TK chimeric genes in TK^+ converted cells. To determine whether the α gene 0 contains the same type of regulatory signals as do α genes 4 and 27,

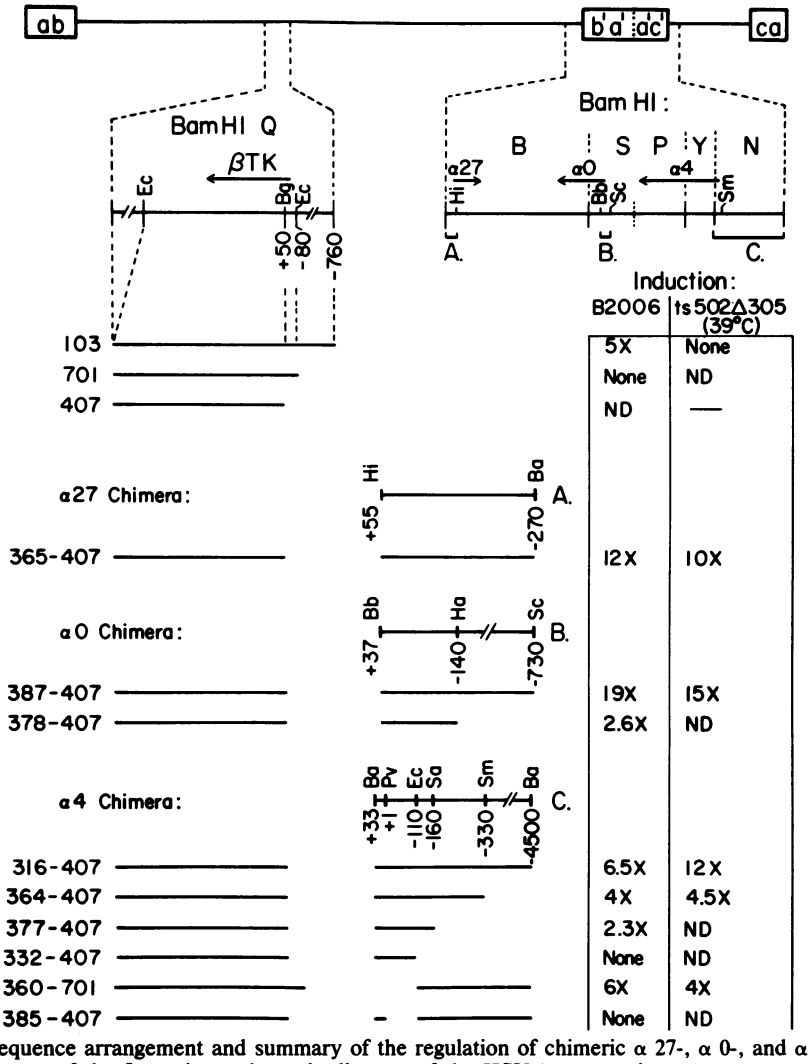

FIG. 1. Sequence arrangement and summary of the regulation of chimeric α 27-, α 0-, and α 4-TK genes. Shown at the top of the figure is a schematic diagram of the HSV-1 genome in prototype arrangement (39). Below, the *Bam*HI Q, B, SP, Y, and N fragments are shown in expanded form, with the domains and orientations of the TK gene and of α genes 27, 0, and 4 indicated (1, 5, 23, 24, 45). Pertinent restriction endonuclease cleavage sites shown are abbreviated as follows: Ba = *Bam*HI; Bb = *Bbv*I; Bg = *Bgl*II, Ec = *Eco*RI; Ha = *Hae*II; Hi = *Hin*fI; Pv = *Pvu*I; Sc = *Sac*I; Sm = *Sma*I. The domains and orientations of the components of the chimeric TK fragments are as shown. The chimeras terminated to the left at either the *Eco*RI site or the *Bam*HI site of the *Bam*HI Q fragment, both of which are located considerably downstream from the poly(A) addition site of the TK gene (30, 42). The inset box summarizes the induction of these genes by B2006, and by *ts*502Δ305 at 39°C, as the maximum-fold increase in TK activity above the baseline level (ND, not done). The sequence arrangements and regulation of natural TK gene fragments shown are summarized from reference 27. Fragment 103 contains an intact, inducible β TK gene; 701 contains an expressible but noninducible β TK extending to −80; and 407 contains a TK fragment extending to +50 (*Bgl*II site) in the transcribed noncoding leader of the gene. A, B, and C show expanded maps for α gene 27, 0, and 4 sequences employed in chimeric constructions. A. Sequence arrangement and regulation of α 27-TK chimera, summarized from reference 26. Note that the expanded α 27 promoter-regulatory region (A.) is shown in the opposite orientation relative to the *Bam*HI maps at the top of the figure. B. Sequence arrangement and regulation of α 0-TK chimeras (from data in Fig. 3). Sequence 387-407 (cloned as pRB387) was constructed by insertion of the *Bbv*I-*Sac*I fragment shown from *Bam*HI SP (pRB336) in the *Bgl*II site of pRB407 (containing the 407 TK fragment shown). The construction necessitated that pRB387 contain, in addition, about 100 bp adjacent to the *Bam*HI SP-Y junction. Not shown is 388-407 (cloned as pRB388), which is identical to pRB387, but the insert is in the opposite orientation. Sequence 378-407 (cloned as pRB378) was constructed by insertion of the *Bbv*I-*Hae*II fragment shown (from pRB336) in the *Bgl*II site of pRB407. C. Sequence arrangement and regulation of α 4-TK chimeras, summarized from reference 27. As illustrated by cells converted to TK^+ phenotype with fragment 360-701, the α regulator sequence can confer α-type regulation when fused to a TK gene containing β promoter sequences but lacking a functional β regulator. However, when the α promoter region was deleted, as in 385-407, there was no regulation, indicating that the α regulator sequences upstream of −110 require a promoter region to function as a regulator.

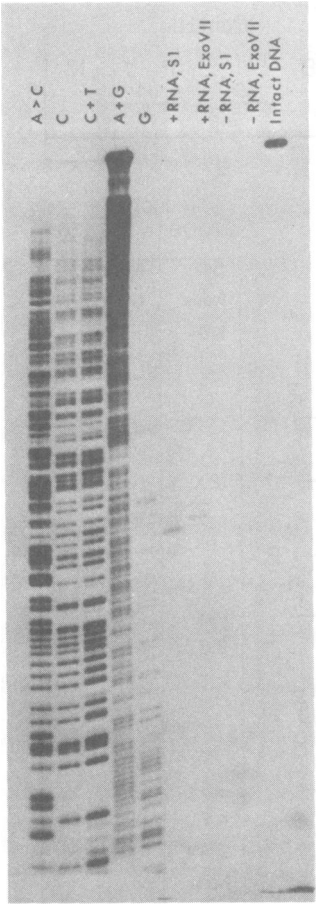

FIG. 2. Autoradiogram of the products of S1 and exo VII digestion of α RNA–HSV-1 DNA hybrids. Poly(A$^+$)-selected RNA from infected cells was hybridized to the appropriate strand of 5′-^{32}P end-labeled *Rsa*I-*Sac*I DNA fragment derived from *Bam*HI SP (located from 600 to 1,500 bp from the left terminus of *Bam*HI SP). The α 0 mRNA-DNA hybrids, and the control of DNA annealed by itself, were digested with either S1 or exo VII. The products were subjected to electrophoresis on an 8% polyacrylamide–8.3 M urea gel. The products of base-specific chemical sequencing reactions of the same DNA that was used for hybridizations were also run on the gel to serve as size markers (lanes 1 to 5), along with the intact, untreated DNA that was used. Lane 1, A>C reaction; lane 2, C reaction; lane 3, C+T reaction; lane 4, A+G reaction; lane 5, G reaction ; lane 6, S1 digest of RNA-DNA hybrids; lane 7, exo VII digest of RNA-DNA hybrids; lane 8, S1 digest of DNA hybridized in the absence of RNA; lane 9, exo VII digest of DNA hybridized in the absence of RNA; lane 10, intact, untreated DNA.

fragments containing sequences located upstream from the α 0 transcription initiation site and including part of the noncoding leader of the α 0 gene were fused to the *Bgl*II site of a TK gene fragment (no. 407, Fig. 1) which contained all of the TK coding sequences and a part of the noncoding leader of the TK gene (beginning at position +50 from the transcription initiation site) (30, 42). The resulting chimeric genes were used to convert Ltk$^-$ cells to the TK$^+$ phenotype. The TK$^+$ cell lines were infected with B2006 (a TK$^-$ virus) to determine whether the resident chimeric TK gene responded to viral regulatory signals, and with *ts*502Δ305 at 39°C to differentiate α from β induction. Cells containing the wild-type β TK gene (no. 103, Fig. 1) served as a control for differentiation between α- and β-regulated TK.

To assess whether sequences containing putative α gene 0 promoter-regulatory regions could confer α-type induction on TK expression, a *Bbv*I-*Sac*I fragment extending from +37 to −730 bp with respect to the α 0 transcription initiation site was fused, in both orientations, to the 5′ terminus of the 407 TK fragment. The chimera in which the fragments were fused in the correct transcriptional orientation was designated 387–407 (Fig. 1). When cells were converted to TK$^+$ phenotype with 387-407, the resident TK was induced about 19-fold above control levels after infection of the cells with B2006 virus (Fig. 3A). Furthermore, infection of the converted cells with *ts*502Δ305 at 39°C (Fig. 3B) demonstrated that the 387-407 chimera was regulated as an α gene. In contrast, when the α gene 0 promoter-regulatory region was fused to the 5′ end of fragment 407 in the inappropriate orientation (388-407), although the chimera converted cells to TK$^+$ phenotype, the resident TK was totally unresponsive to viral regulatory signals after infection of the cells with B2006 virus (Fig. 3A). As would be expected based on previous studies (26), conversion of the TK 407 fragment to a regulated α gene required that the α gene 0 promoter-noncoding leader be fused in the normal transcriptional orientation.

Previous studies of the regulation of α 4-TK chimeras have shown that sequences conferring induction on expression of the chimeras are located considerably upstream (>110 bp) from the α 4 transcription initiation site (27). In an attempt to determine whether this is also the case for the α gene 0 regulatory region, a chimeric gene containing α 0 sequences from +37 to −140 (no. 378, Fig. 1) was fused to the 5′ end of the TK 407 fragment in the proper transcriptional orientation. The 378-407 chimera resident in converted cells was inducible at a low level (2.6-fold over controls) after infection with B2006 (Fig. 3A). Thus, removal of sequences upstream of position −140 from the α 0 transcription initiation site resulted in about a sevenfold reduction in the regulatory responsiveness of the chimeric gene. The induction of 378-407 at a low level may be a reflection of the presence of

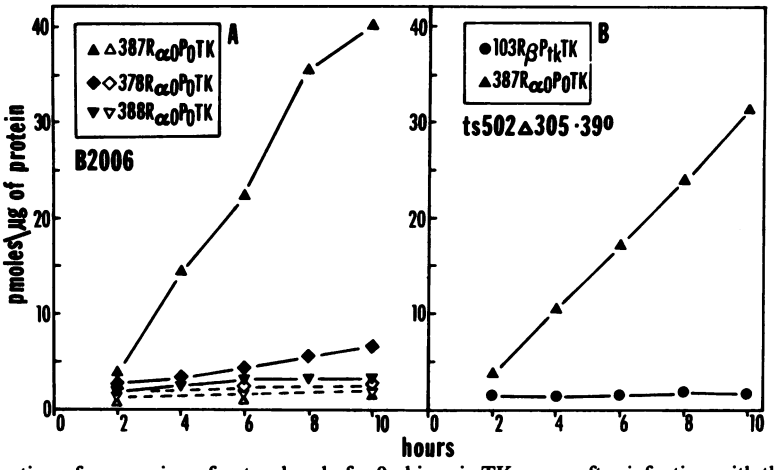

FIG. 3. Induction of expression of natural and of α 0-chimeric TK genes after infection with the TK⁻ viruses HSV-1 (C101) B2006 (A) or *ts*502Δ305 (B). Cells converted to TK⁺ phenotype with the plasmid DNAs indicated in the boxes in each panel were superinfected with 5 PFU of B2006 or of *ts*502Δ305 virus per cell. Extracts from cells harvested at the times shown postinfection were assayed for TK activity by conversion of [^{3}H]thymidine into thymidylate. For infections with *ts*502Δ305, the cells were exposed to virus for 1 h at 10°C and transferred to 39°C (zero-time). The level of TK activity present in cells at different times postinfection is expressed as the number of picomoles of thymidine phosphorylated per microgram of protein in 30 min at 37°C. Closed symbols designate time points for infected cell lines, and open symbols designate time points for mock-infected cell lines. The cell lines (inset boxes) were designated according to the numerical designations of the natural or chimeric TK fragment shown in Fig. 1, followed by a descriptive designation of the components of the TK gene: $R_{\alpha 0}$ = α0 regulatory region; $P_{\alpha 0}$ = α0 promoter region; R_β = β regulatory region; and P_{tk} = β TK promoter region. The induction of TK activity by infection with B2006 indicates that the resident TK gene responds to viral regulatory signals; induction by *ts*502Δ305 at 39°C indicates that the TK gene is regulated as an α gene.

multiple regulatory units capable of acting cumulatively, as appears to be the case for α gene 4 regulation (27).

Nucleotide sequence of the promoter-regulatory regions of α genes 0 and 27. To identify structural characteristics common to the α promoter and regulatory regions which might have a role in α gene regulation, the nucleotide sequence from +176 to −323 in the domain of α gene 0 and from +251 to −271 in the domain of α gene 27 were determined according to the strategy shown in Fig. 4B and 5B. To illustrate the presence and size of multiple inverted repeats, the nucleotide sequences of the nontemplate strands are shown in Fig. 4 and 5 as though they are able to form cruciform structures. Also shown, for purposes of comparison (Fig. 6), is the sequence (+183 to −332) of a more extended portion of the domain of α gene 4 than that published previously (27). These sequences may cover only a portion of the region responsible for α regulation, inasmuch as addition of α gene 4 sequences beyond −330 (chimera 316-407 in Fig. 1) further increased the level of induction.

The sequences of each of the three α genes contain a homolog of the previously described TATA sequence (6, 15) beginning at about 25 to 26 bp upstream from the α transcription initiation site (Fig. 4 to 6). Each of the three heptanucleotide homologs differs from the consensus sequence in the presence of G residues. Only α gene 0 contains a sequence resembling the CAAT sequence present at position −70 to −80 of some eucaryotic genes (4, 11).

The α gene 4, 0, and 27 sequences share two distinctive features: directly repeated, homologous, AT-rich sequences present in one to several copies in each gene (compared in Table 1), and multiple G+C-rich inverted repeated sequences illustrated as cruciform structures in Fig. 4 to 6. The AT-rich directly repeated sequences, which are numbered in Table 1 relative to the α transcription initiation sites, contain a central 11-nucleotide-long region that is highly conserved, flanked by regions that are partially conserved. The nucleotides for each sequence are also grouped into stretches showing greater and lesser degrees of conservation. In deriving a consensus sequence, only sequences located in regions of α gene 4, 0, and 27 that have been experimentally shown to contain regulatory function were included (above the dashed line). The table also shows a similar sequence which was found at the 3′ end of the noncoding leader region of α gene 27 (which was not contained in any of the chimeric constructions) and homologous sequences present in the region upstream from the transcription initiation site of α gene 47,

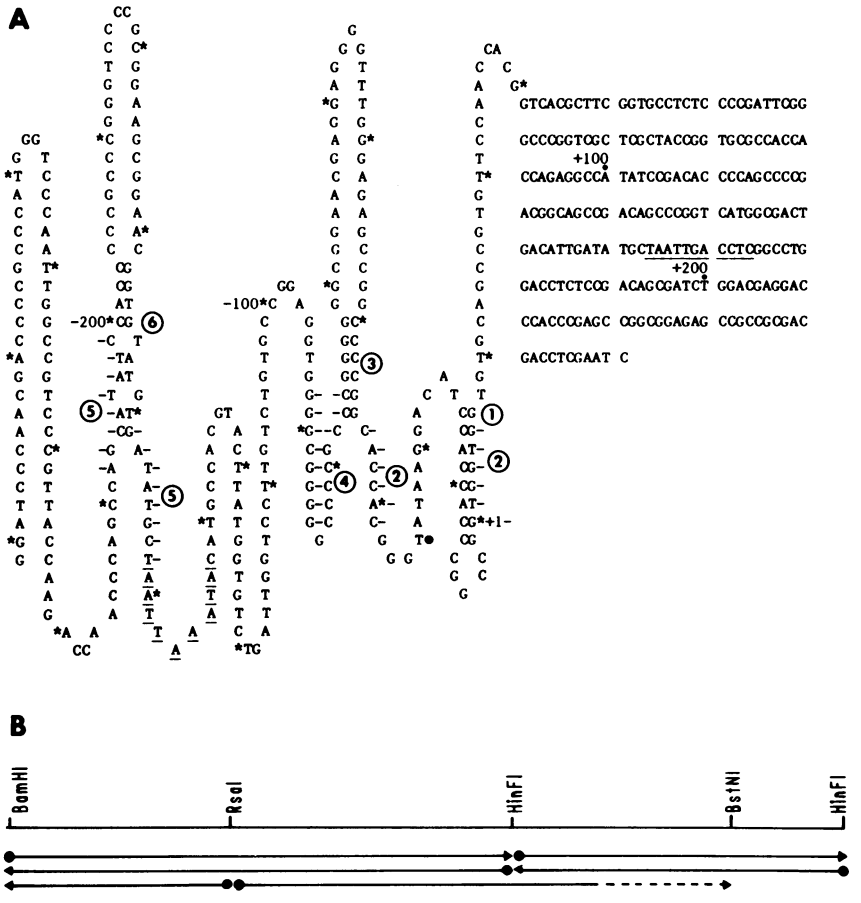

FIG. 4. Nucleotide sequence of the promoter-regulatory region and noncoding leader of α gene 27. (A) The sequence of the nontemplate strand extends from the *Bam*HI site at −271 bp (far left) to the *Hin*fI site at +251 bp (far right). The sequence is numbered in sets of 100, with asterisks or spaces at every 10th nucleotide, relative to the transcription initiation site (previously mapped by S1 analysis [26]), which is shown as position +1. A homolog of the TATA sequence (6, 15) occurs at −25 bp and is denoted at the 5′ residue by a large dot. Major inverted-repeated sequences are shown in cruciform structures designated by numerals, with dashes indicating alternate sequences capable of base pairing. The underscored, homologous, repeated sequences are compared in Table 1 and discussed in the text. (B) The top line is a restriction enzyme map of the sequence shown above. Below the map, the sequencing strategy used is shown schematically. The dots represent labeled termini that were sequenced in the direction of the arrows extending from them.

recently sequenced by other investigators (47) (shown below the dashed line in Table 1). The significance of finding a homologous sequence downstream from the transcription initiation site of α gene 27 is unclear; recent work on the effect of viral enhancers on genes transcribed by RNA polymerase II (2, 8, 34, 48) and studies on transcription by RNA polymerase III (13, 41) suggest that sequences capable of influencing gene expression need not necessarily be located upstream from the transcription initiation sites.

A second, striking feature of the sequences 5′ to the transcription initiation sites of α genes 4, 0, and 27 is the presence of multiple inverted repeats (Fig. 4 to 6). Although these inverted repeats do not share obvious primary sequence homology, they are generally GC rich; they predominate in regions downstream and are interspersed among the directly repeated AT-rich sequences discussed above. The presence of inverted repeated sequences has also been noted in the region upstream from the transcription initiation site of α gene 47 (47).

DISCUSSION

In this report, we show that the regulation of chimeric genes containing the promoter-regulatory region of α gene 0 could not be functionally differentiated from that of α genes 4 and 27. We also report on the nucleotide sequences of pro-

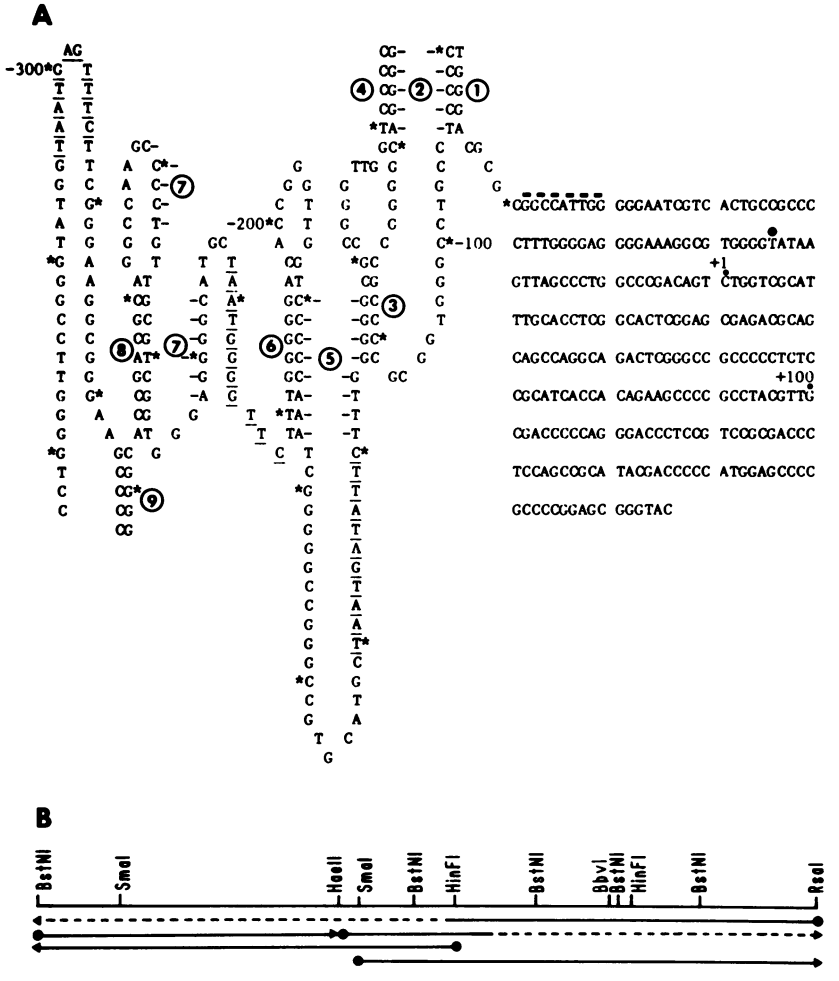

FIG. 5. Nucleotide sequence of the promoter-regulatory region and noncoding leader of α gene 0. (A) The sequence of the nontemplate strand extends from the *Bst*NI site at −323 bp (far left) to the *Rsa*I site at +176 bp (far right). The sequence is numbered in sets of 100, with asterisks or spaces at every 10th nucleotide, relative to the transcription initiation site, which is shown as position +1. Homologs of the TATA sequence (denoted at 5′ residue by a large dot) and the CAAT sequence (denoted by the heavy dashed line) (4, 6, 11, 15) occur at −25 and −79 bp, respectively. Major inverted-repeated sequences are shown in cruciform structures designated by numerals, with dashes indicating alternate sequences capable of base pairing. The underscored, homologous, repeated sequences are compared in Table 1 and discussed in the text. (B) The top line is a restriction enzyme map of the sequence shown above. The map is in the same orientation as the sequence presented above but in opposite orientation relative to the map shown in Fig. 1. Below the map, the sequencing strategy used is shown schematically. The dots represent labeled termini that were sequenced in the direction of the arrows extending from them.

moter and regulatory regions of α genes 0, 4, and 27. At this point, it is both convenient and desirable to summarize the major functional and structural features of the expression of α genes.

Characteristics of the regulation of α genes. Figure 1 summarizes the regulation of chimeric TK genes generated by fusion of different subsets of the α promoter-regulatory regions from α genes 0, 4, and 27 to appropriate portions of the domain of the TK gene. From the results of this and previous studies (26, 27, 36), as summarized in the figure, several conclusions can be drawn regarding the characteristics of α gene regulation.

(i) We are currently aware of at least three structural elements and at least two virus-specific signals involved in the regulation of α genes. The first structural element, which we have designated as the promoter, facilitated the expression of chimeric TK genes but did not impart upon the chimeric gene sensitivity to regulatory signals. In α gene 4, this region

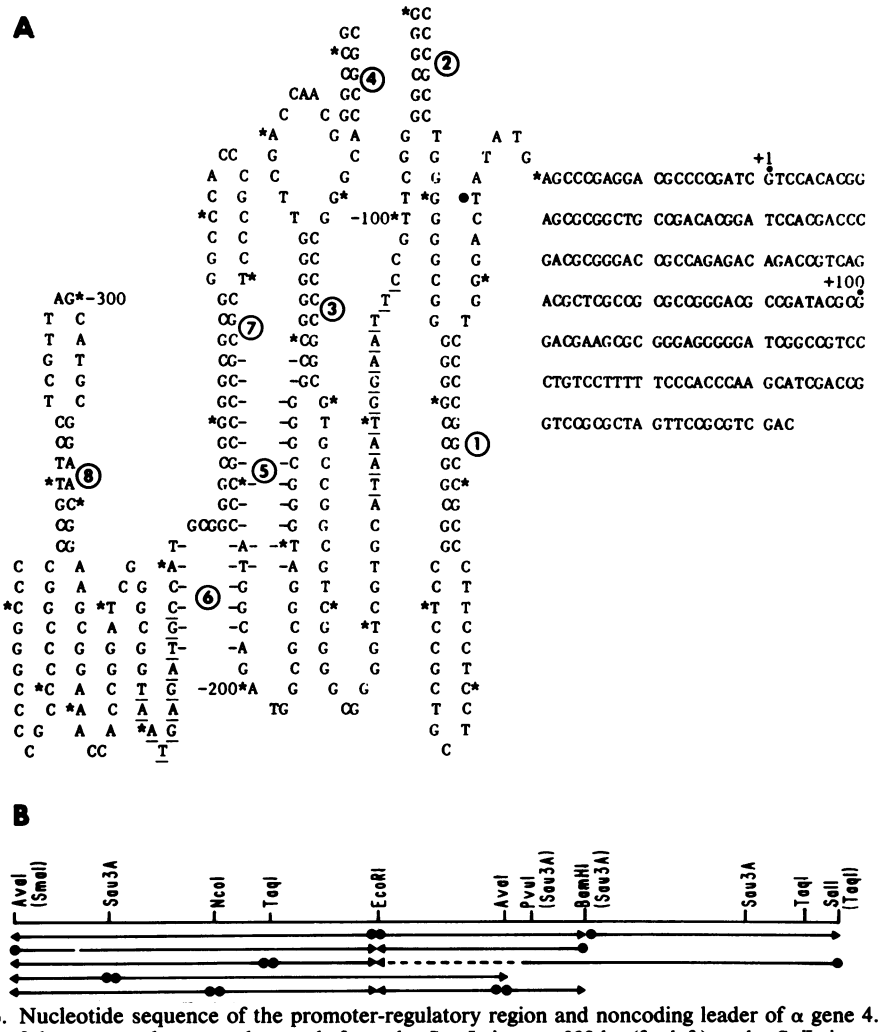

FIG. 6. Nucleotide sequence of the promoter-regulatory region and noncoding leader of α gene 4. (A) The sequence of the nontemplate strand extends from the *SmaI* site at −332 bp (far left) to the *SalI* site at +183 bp (far right). The sequence is numbered in sets of 100, with asterisks or spaces at every 10th nucleotide, relative to the transcription initiation site (previously mapped by S1 analysis [27]), which is shown as position +1. A homolog of the TATA sequence (6, 15) occurs at −26 bp and is denoted at the 5′ residue by a large dot. Major inverted-repeated sequences are shown in cruciform structures designated by numerals, with dashes indicating alternate sequences capable of base pairing. The underscored, homologous, repeated sequences are compared in Table 1 and discussed in the text. (B) The top line is a restriction enzyme map of the sequence shown above. The map is in the same orientation as the sequence presented above but in opposite orientation relative to the map shown in Fig. 1. Below the map, the sequencing strategy used is shown schematically. The dots represent labeled termini that were sequenced in the direction of the arrows extending from them.

extended upstream from the site of initiation of transcription to about −110 bp. In α gene 0, it extended upstream to a position less than −140 bp. In cells converted to TK^+ phenotype, the chimeric TK gene containing only this portion of the α genes did not respond to infection of the cells by higher levels of TK expression. The second element, extending upstream from nucleotide −110 bp in α gene 4 and from at least −140 bp in α gene 0, imparted upon chimeric genes sensitivity to signals introduced into the cells by the virus and may be designated as the induction sequence. In cells converted to TK^+ phenotype, the chimeric genes containing these elements were expressed at a much higher level after infection with TK^- virus. Furthermore, there may be several induction elements responding concurrently and cumulatively to the infection. This conclusion is based on the correlation between the induced enzyme activity and the size

of the DNA fragment extending upstream from −110 bp in the domain of α gene 4 that was fused to the appropriate promoter sequence.

The third element is as yet imprecisely defined. It has been demonstrated that fusion of the α gene 4 promoter-regulatory region to the body of the chick ovalbumin gene converts that gene into an α-regulated gene (37). Current studies by C. Herz and B. Roizman (unpublished data) indicate that in cells converted to TK^+ phenotype by a plasmid containing the natural TK gene and the α-ovalbumin gene, the expression of α-ovalbumin specified by the resident plasmid closely paralleled the expression of α gene 4 by the infecting virus. Thus, several hours after infection of these cells with *ts* mutants in α gene 4, the expression of both α-ovalbumin and α gene 4 product continued virtually undiminished at the nonpermissive temperature but was shut off late in infection at the permissive temperature. In light of the evidence that the shutoff of α gene expression is mediated at the transcriptional level (44), these results suggest that the promoter-regulatory region of α gene 4 contains a structural element recognized by virus-specific signals for shutoff of α gene expression.

(ii) The chimeric α TK genes respond to at least two signals. The first of the two known signals acts early after infection. On the basis of the observation that the induction can take place in the presence of cycloheximide and is independent of α gene 4 expression, it has been suggested that the signal is a gene product brought into the cell by the infecting virus. It should be noted that the only known activities associated with the virions are a protein kinase (22) and a function responsible for an early shutoff in host protein synthesis (14). It remains to be seen whether the induction of α gene expression is caused by one of these or by a different *trans*-acting viral product. The second signal is expressed after the onset of α gene expression and results in the late shutoff of expression of all α genes. It has been suggested on the basis of studies done with a single mutant in α gene 4 that this gene regulates itself (44). However, the possibility that this mutant has additional defects has not been rigorously excluded.

(iii) There is currently little or no information on the interaction of the various signals and structural elements which govern the expression of α genes. Because the amount of the chimeric TK gene product induced by infection was related to the size of the regulatory region of the gene, it could be argued that the signal gene product interacts directly with the regulatory region or with a host product which interacts directly with the regulatory region.

TABLE 1. Directly repeated A+T-rich sequences present in the promoter-regulatory regions of α genes[a]

Consensus:	$^{5'}$G[C]AT GN TAAT GA[G]A TTC [C]TT GN GGG$^{3'}$
α Gene 0:	$^{-310}$GTAT Gg TAAT GAGt TTC TTc Gg GAA^{-286}
	$^{-228}$GCAT Gc TAAT GgGg TTC TTT Gg GGG^{-204}
	$^{-166}$GCAT Gc TAAT GAtA TTC TTT Gg GGG^{-142}
α Gene 4:	$^{-268}$Gggc Gg TAAT GAGA TgC CaT Gc GGG^{-244}
	$^{-119}$cgtg cA TAAT GgAA TTC CgT Tc GGG $^{-95}$
α Gene 27:	$^{-158}$ATAT Gc TAAT tAAA TaC ATg cc AcG^{-134}

	$^{+158}$ATAT Gc TAAT TgAc cTC ggc cT GGA^{+182}
α Gene 47:	$^{-397}$GCAT Gc TAAc GAGg AAC ggg cA GGG^{-375}
	$^{-343}$cggc Gg TAAT GAGA TaC gag cc ccG^{-321}
	$^{-178}$GCcg Gg TAAa AgAA gTg AGA Ac GcG^{-154}

[a] The possible consensus sequence shown at the top of the table was derived by comparison of the six sequences above the horizontal dashed line, for reasons indicated in the text. The central 11-nucleotide region that is most highly conserved is indicated by the bracket above the consensus sequence. Residues shown as large capitals conform to the consensus, whereas those shown as small capitals deviate from the consensus. The homologs shown for α gene 47 are taken from sequence data by Watson and Vande Woude (47).

ulatory regions of α genes 0, 4, and 27. The α genes 0, 4, and 27 contain sequences common to other viral and eucaryotic cell genes as well as shared sequences which have not been identified in other genes.

Among the sequences shared by many viral and eucaryotic genes, the α genes 0, 4, and 27 exhibit a TATA homolog (6, 15) at the predicted distance from the site of initiation of transcription. The TATA homologs found in the α genes differ, however, from the consensus in G residues. Only one of the α genes contains a sequence somewhat homologous to the CAAT consensus sequence (4, 11). Although the TATA sequence appears to specify the precise initiation of transcription (3, 9, 28), the CAAT appears to be dispensable (17), although in a few instances it has been correlated with efficiency of gene expression (31, 32).

Two distinctive features of the regulatory-promoter regions of α genes were not previously noted in viral or eucaryotic genes. These are a consensus sequence (Table 1) and GC-rich inverted repeats. Although the figures showing the sequences were designed to emphasize the number, positioning, and base pair matching of the inverted repeats, it is not at all clear that the secondary structures depicted in the figures actually form. It seems significant that all α genes

contain both the consensus sequence and the inverted repeats and that in α genes 0 and 4 the consensus sequences appear interspersed among the inverted repeats. At least in these genes, the multiplicity of consensus sequences and of inverted repeats is consistent with evidence for multiple induction sequences. However, it remains to be determined which, if any, of these sequences correspond to the functional elements described in the preceding section.

Irrespective of the identity of the α induction sequences, the observation that they are located at a considerable distance from the transcription initiation site and are apparently separate from the α promoter function suggests a similarity to the enhancer sequences which have been described for papova (2, 8) and retroviruses (34) and possibly for the H2A histone gene (18). These sequences also occur as repeats and have been found to stimulate gene expression from considerable distances both upstream and downstream of the gene to which they are linked. Although functionally similar, enhancer sequences in different genes described to date share no obvious sequence homology with each other (48). A unique feature of the α induction sequences is that they can respond in the absence of de novo viral protein synthesis to signals brought into the cells by the infecting virus. In this respect, they differ from the enhancer sequences cited above. Moreover, it remains to be determined whether the viral α induction sequences are capable of functioning from variable distances both upstream and downstream from the gene to which they are linked.

ACKNOWLEDGMENTS

We thank Mayer Stein for invaluable technical assistance.

S.M. is a Public Health Service Medical Scientist Program trainee (PHS-2-T32 GM07281-C06). These studies were supported by Public Health Service grants CA-08494 and CA-19264 from the National Cancer Institute and grant MV-2-P from the American Cancer Society.

LITERATURE CITED

1. **Anderson, K. P., R. H. Costa, L. E. Holland, and E. K. Wagner.** 1980. Characterization of herpes simplex virus type 1 RNA present in the absence of de novo protein synthesis. J. Virol. **34**:9–27.
2. **Banerji, J., S. Rusconi, and W. Schaffner.** 1981. Expression of a β-globin gene is enhanced by remote SV40 DNA sequences. Cell **27**:299–308.
3. **Benoist, C., and P. Chambon.** 1981. In vivo sequence requirements of the SV40 early promoter region. Nature (London) **290**:304–310.
4. **Benoist, C., K. O'Hare, R. Breathnach, and P. Chambon.** 1980. The ovalbumin gene-sequence of putative control regions. Nucleic Acids Res. **8**:127–142.
5. **Colbere-Garapin, F., S. Chousterman, F. Horodniceanu, P. Kourilsky, and A.-C. Garapin.** 1979. Cloning of the active thymidine kinase gene of herpes simplex virus type 1 in Escherichia coli K-12. Proc. Natl. Acad. Sci. U.S.A. **76**:3755–3759.
6. **Corden, J., B. Wasylyk, A. Buchwalder, P. Sassone-Corsi, C. Kedinger, and P. Chambon.** 1980. Promoter sequences of eukaryotic protein-coding genes. Science **209**:1406–1414.
7. **Costanzo, F., G. Campadelli-Fiume, L. Foa-Tomasi, and E. Cassai.** 1977. Evidence that herpes simplex virus DNA is transcribed by cellular RNA polymerase B. J. Virol. **21**:996–1001.
8. **deVilliers, J., and W. Schaffner.** 1981. A small segment of polyoma virus DNA enhances the expression of a cloned β-globin gene over a distance of 1400 base pairs. Nucleic Acids Res. **9**:6251–6264.
9. **Dierks, P., A. van Ooyen, N. Mantei, and C. Weissman.** 1981. DNA sequences preceding the rabbit β-globin gene are required for formation in mouse L cells of β-globin RNA with the correct 5′ terminus. Proc. Natl. Acad. Sci. U.S.A. **78**:1411–1415.
10. **Dubbs, D. R., and S. Kit.** 1964. Mutant strains of herpes simplex deficient in thymidine kinase-inducing activity. Virology **22**:493–502.
11. **Efstratiadis, A., J. W. Posakony, T. Maniatis, R. M. Lawn, C. O'Connell, R. A. Spritz, J. K. DeRiel, B. G. Forget, S. M. Weissman, C. C. Slightom, and N. J. Proudfoot.** 1980. The structure and evolution of the human β-globin gene family. Cell **21**:653–668.
12. **Ejercito, P. M., E. D. Kieff, and B. Roizman.** 1968. Characterization of herpes simplex virus strains differing in their effects on social behavior of infected cells. J. Gen. Virol. **2**:357–364.
13. **Engelke, D. R., S.-Y. Ng, B. S. Shastry, and R. G. Roeder.** 1980. Specific interaction of a purified transcription factor with an internal control region of the 5S RNA genes. Cell **19**:717–728.
14. **Fenwick, M. L., and M. J. Walker.** 1978. Suppression of the synthesis of cellular macromolecules by herpes simplex virus. J. Gen. Virol. **41**:37–51.
15. **Gannon, F., K. O'Hare, F. Perrin, J. P. LePennec, C. Benoist, M. Cochet, R. Breathnach, A. Royal, A. Garapin, B. Cami, and P. Chambon.** 1979. Organization and sequences at the 5′ end of a cloned complete ovalbumin gene. Nature (London) **278**:428–434.
16. **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**:456–467.
17. **Grosschedl, R., and M. L. Birnstiel.** 1980. Identification of regulatory sequences in the prelude sequences of an H2A histone gene by the study of specific deletion mutants in vivo. Proc. Natl. Acad. Sci. U.S.A. **77**:1432–1436.
18. **Grosschedl, R., and M. L. Birnstiel.** 1980. Spacer DNA sequences upstream of the T-A-T-A-A-A-T-A sequences are essential for promotion of H2A histone gene transcription in vivo. Proc. Natl. Acad. Sci. U.S.A. **77**:7102–7106.
19. **Honess, R. W., and B. Roizman.** 1974. Regulation of herpesvirus macromolecular synthesis. I. Cascade regulation of the synthesis of three groups of viral proteins. J. Virol. **14**:8–19.
20. **Jones, P. C., G. S. Hayward, and B. Roizman.** 1977. Anatomy of herpes simplex virus DNA. VII. α RNA is homologous to noncontiguous sites in both the L and S components of viral DNA. J. Virol. **21**:268–276.
21. **Kit, S., D. R. Dubbs, L. J. Piekarski, and T. C. Hsu.** 1963. Deletion of thymidine kinase activity from L cells resistant to bromodeoxyuridine. Exp. Cell Res. **31**:297–312.
22. **Lemaster, S., and B. Roizman.** 1980. Herpes simplex virus phosphoproteins. II. Characterization of the virion protein kinase and of the polypeptides phosphorylated in the virion. J. Virol. **35**:798–811.
23. **Locker, H., and N. Frenkel.** 1979. *Bam*I, *Kpn*I, and *Sal*I restriction enzyme maps of the DNAs of herpes simplex virus strains Justin and F: occurrence of heterogeneities in defined regions of the viral DNA. J. Virol. **32**:429–441.
24. **Mackem, S., and B. Roizman.** 1980. Regulation of herpesvirus macromolecular synthesis: transcription-initiation sites and domains of α genes. Proc. Natl. Acad. Sci. U.S.A. **77**:7122–7126.

25. **Mackem, S., and B. Roizman.** 1981. Regulation of herpesvirus macromolecular synthesis: temporal order of transcription of α genes is not dependent on the stringency of inhibition of protein synthesis. J. Virol. **40**:319–322.
26. **Mackem, S., and B. Roizman.** 1982. Regulation of α genes of herpes simplex virus: the α 27 promoter-thymidine kinase chimera is positively regulated in converted L cells. J. Virol. **43**:1015–1023.
27. **Mackem, S., and B. Roizman.** 1982. Differentiation between α promoter and regulator regions of herpes simplex virus 1: the functional domains and sequence of a movable α regulator. Proc. Natl. Acad. Sci. U.S.A. **79**:4917–4921.
28. **Mathis, D., and P. Chambon.** 1981. The SV40 early region TATA box is required for accurate *in vitro* initiation of transcription. Nature (London) **290**:310–315.
29. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
30. **McKnight, S. L.** 1980. The nucleotide sequence and transcript map of the herpes simplex virus thymidine kinase gene. Nucleic Acids Res. **8**:5949–5964.
31. **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.
32. **Mellon, P., V. Parker, Y. Gluzman, and T. Maniatis.** 1981. Identification of DNA sequences required for transcription of the human α1-globin gene in a new SV40 host-vector system. Cell **27**:279–288.
33. **Morse, L. S., L. Pereira, B. Roizman, and P. A. Schaffer.** 1978. Anatomy of herpes simplex virus (HSV) DNA. X. Mapping of viral genes by analysis of polypeptides and functions specified by HSV-1 × HSV-2 recombinants. J. Virol. **26**:389–410.
34. **Payne, G. S., J. M. Bishop, and H. E. Varmus.** 1982. Multiple arrangements of viral DNA and an activated host oncogene in bursal lymphomas. Nature (London) **295**:209–214.
35. **Post, L. E., A. J. Conley, E. S. Mocarski, and B. Roizman.** 1980. Cloning of reiterated and nonreiterated herpes simplex 1 sequences as BamHI fragments. Proc. Natl. Acad. Sci. U.S.A. **77**:4201–4205.
36. **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.
37. **Post, L. E., B. Norrild, T. Simpson, and B. Roizman.** 1982. Chicken ovalbumin gene fused to a herpes simplex virus α promoter and linked to a thymidine kinase gene is regulated like a viral gene. Mol. Cell. Biol. **2**:233–240.
38. **Preston, V. G., A. J. Davison, H. S. Marsden, M. C. Timbury, J. H. Subak-Sharpe, and N. M. Wilkie.** 1978. Recombinants between herpes simplex virus types 1 and 2: analyses of genome structures and expression of immediate early polypeptides. J. Virol. **28**:499–517.
39. **Roizman, B.** 1979. The organization of the herpes simplex virus genomes. Annu. Rev. Genet. **13**:25–57.
40. **Roizman, B., and P. G. Spear.** 1968. Preparation of herpes simplex virus of high titer. J. Virol. **2**:83–84.
41. **Sakonju, S., D. F. Bogenhagen, and D. D. Brown.** 1980. A control region in the center of the 5S RNA gene directs specific initiation of transcription: 1. The 5′ border of the region. Cell **19**:13–25.
42. **Wagner, M. J., J. A. Sharp, and W. C. Summers.** 1981. Nucleotide sequence of the thymidine kinase gene of herpes simplex virus type 1. Proc. Natl. Acad. Sci. U.S.A. **78**:1441–1445.
43. **Wartell, R. M., and W. S. Reznikoff.** 1980. Cloning of DNA restriction endonuclease fragments with protruding single-stranded ends. Gene (Amst.) **9**:307–319.
44. **Watson, R. J., 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.
45. **Watson, R. J., C. M. Preston, and J. B. Clements.** 1979. Separation and characterization of herpes simplex virus type 1 immediate-early mRNA's. J. Virol. **31**:42–52.
46. **Watson, R. J., M. Sullivan, and G. F. Vande Woude.** 1981. Structures of two spliced herpes simplex virus type 1 immediate-early mRNA's which map at the junctions of the unique and reiterated regions of the virus DNA S component. J. Virol. **37**:431–444.
47. **Watson, R. J., and G. F. Vande Woude.** 1982. DNA sequence of an immediate-early gene (IE mRNA-5) of herpes simplex virus 1. Nucleic Acids Res. **10**:979–991.
48. **Yaniv, M.** 1982. Enhancing elements for activation of eukaryotic promoters. Nature (London) **297**:17–18.