JOURNAL OF VIROLOGY, Feb. 1986, p. 629–637
0022-538X/86/020629-09$02.00/0

Vol. 57, No. 2

The Terminal *a* Sequence of the Herpes Simplex Virus Genome Contains the Promoter of a Gene Located in the Repeat Sequences of the L Component

JOANY CHOU AND BERNARD ROIZMAN*

The Marjorie B. Kovler Viral Oncology Laboratories, The University of Chicago, Chicago, Illinois 60637

Received 27 August 1985/Accepted 25 October 1985

The herpes simplex virus DNA genome consists of two covalently linked components, L and S. The unique sequences of the L component are flanked by 9-kilobase-pair inverted repeat sequences *ab* and *b'a'*, whereas those of the S component are flanked by 6.5-kilobase-pair inverted repeat sequences *c'a'* and *ca*. We report that the 500-base-pair *a* sequence contains the promoter-regulatory domain and the transcription initiation site of a diploid gene, the coding sequences of which are located in the *b* sequences of the inverted repeats of the L component. The chimeric gene constructed by fusion of the *a* sequence to the coding sequences of the thymidine kinase gene and recombined into the viral genome was regulated as a γ_1 gene. The size of the protein predicted from its sequence is 358 amino acids; it was designated as infected cell protein (ICP) 34.5. Thus, the inverted repeats flanking the unique sequences of the L component contain two genes specifying ICP0 and ICP34.5, respectively. Moreover, in addition to the *cis*-acting sites for the inversion of L and S components relative to each other, for cleavage of unit length DNA molecules from head-to-tail concatemers, and for packaging of the DNA into capsids, the *a* sequence also contains the promoter-regulatory domain and transcription initiation sites of a gene.

The herpes simplex virus type 1 (HSV-1) genome, a linear double-stranded DNA of 150 kilobase pairs (kbp; 14) consists of two components, L and S. Each component consists of unique sequences, U_l and U_s, flanked by inverted repeats (29). The inverted repeats of the L component, designated as *ab* and *b'a'*, are each approximately 9 kbp, whereas those of the S component, *c'a'* and *ca*, are each approximately 6.5 kbp (36). An unusual property of the HSV-1 genome is that the L and S components can invert such that DNA extracted from infected cells or virions consists of four equimolar populations differing solely in the orientation of the L and S components relative to each other (10). Earlier studies have shown that the sequence shared by the inverted repeats of the L and S components, designated the *a* sequence (36) and located in the same orientation at the termini of the DNA and in an inverted orientation at the junction between the L and S components, is the *cis*-acting site for the inversion of the L and S components (6, 24–26, 31), cleavage of the DNA from head-to-tail concatemers, and packaging of unit-length DNA (35). We report here that the *a* sequence contains the promoter-regulatory sequence and the transcription initiation sites for a gene located in the *b* sequence of the inverted repeats of the L component. Relevant to the results described in this report are the following.

The *a* sequence of the HSV-1 strain F [HSV-1(F)] consists of several elements. Starting from the inverted repeat *b'* sequence of the L component to the inverted *c'* sequence of the S component (see Fig. 1A), these elements consist of a terminal 20-base-pair (bp) direct repeat (DR) designated DR1b, a 58-bp unique sequence (U_b), a 12-bp DR (DR2) repeated 19 to 22 times, a 37-bp DR (DR4) repeated two to three times, a 65-bp unique sequence (U_c), and the terminal 20-bp DR1c (25, 26). Only a single copy of the *a* sequence is present at the S component terminus of the DNA, but one to more than five directly repeated copies of the *a* sequence may be present at the L component terminus and at the junction between the L and S components (16, 38). Directly repeated copies of the *a* sequence share the intervening DR1 (25, 26).

The HSV-1 genes form three groups, the expression of which is coordinately regulated and sequentially ordered in a cascade fashion (11, 12). The α genes are expressed first, and functional α-gene products are required for the expression of β genes. The products of β genes shut off α-gene expression and turn on the expression of γ genes. The γ proteins, mostly structural components of the virion, turn off the expression of β genes. At least one γ protein designated as an α *trans*-inducing factor and introduced into cells during infection has recently been shown to induce α-gene transcription (1, 5, 15, 17–19, 27a, 28). The γ genes are heterogeneous; the γ_1 gene expression is reduced but not abolished by inhibitors of DNA synthesis, whereas the expression of γ_2 genes is stringently dependent on viral DNA synthesis (7, 11, 30, 37). To facilitate analyses and identification of the regulatory groups to which a specific gene belongs, we introduced the technique of fusing the promoter-regulatory domain of the specific gene to coding sequences of the thymidine kinase (TK) gene, a β gene (28). The chimeric gene can be introduced into the viral genome by recombination through homologous flanking sequences or into the cellular genome by transfection and selection of transformed cells that carry a suitable marker. The regulation of the chimeric TK gene can be studied in cells infected with the recombinant virus or by infection of the cells that carry the chimeric gene with TK^- virus under appropriate conditions.

MATERIALS AND METHODS

Cells and viruses. HSV-1(F) is the prototype HSV-1 strain used in this laboratory (9). Rabbit skin cells, originally obtained from J. McLaren, were used for transfection with viral DNAs. The selection for TK^+ recombinants was done in human $143TK^-$ cells overlaid with HAT medium (8.0 μM

* Corresponding author.

thymidine, 0.1 mM hypoxanthine, 0.44 μM methotrexate). Recombinant virus stocks were prepared, and titers were determined in Vero cells.

Transformation of 143TK$^-$ cells with the pSV2-neo vector and its derivatives was done as described previously (32), and the selection was done in the presence of the antibiotic G418 (400 μg/ml; GIBCO Laboratories, Grand Island, N.Y.).

Enzymes and cloning. Restriction enzymes were obtained from Bethesda Research Laboratories, Inc., Gaithersburg, Md.; Pharmacia Fine Chemicals, Piscataway, N.J.; and New England Biolabs, Inc., Beverly, Mass., and were used according to the instructions of the suppliers. T4 DNA ligase and calf intestinal alkaline phosphatase were from Boehringer Mannheim Biochemicals, Indianapolis, Ind. After ligation, plasmid DNAs were transformed into *Escherichia coli* c600 cells. pSV2-neo vector was obtained from P. Berg, Stanford University. pUC9 and pUC13 vectors and their host JM83 (34) were obtained from J. Messing through Bethesda Research Laboratories. The construction of the plasmids relevant to these studies is described where appropriate.

Preparation of probes and hybridizations. For detection of mRNA the plasmid containing the probe DNA described in Results was nick translated with ^{32}P with the aid of a kit from New England Nuclear Corp. (Boston, Mass.) by the specifications of the manufacturer. The preparation and labeling of the probe for S1 nuclease analyses is also described in Results.

RNA studies. Cytoplasmic RNA extracted from infected Vero cells with 20 PFU of HSV-1(F) per cell as described previously (13) was electrophoretically separated in an agarose gel, transferred to a nitrocellulose sheet, and hybridized to 100,000 to 500,000 cpm of nick-translated DNA probe. Poly(A)$^+$ RNA was purified by oligo(dT) cellulose chromatography by the procedure of Jenkins and Howett (13). For S1 nuclease analyses, 150 μg of cytoplasmic RNA extracted as described above was hybridized with a DNA probe labeled at one end as described below. The DNA-RNA hybrid was subjected to S1 nuclease digestion and electrophoresis along with a sequencing ladder of the DNA fragment used in this hybridization, as described previously (13).

DNA sequencing analysis. The nucleotide sequence of HSV-1(F) gene 34.5 was determined by the method of Maxam and Gilbert (20). ^{32}P was obtained from ICN Pharmaceuticals Inc., Irvine, Calif. T4 polynucleotide kinase was purchased from New England Biolabs. Chemically treated fragments were subjected to electrophoresis on an 8% sequencing gel.

Assay for TK activity. HSV-1(F)Δ305 virus contains a 700-bp deletion in the TK gene (28). The procedures for the assay of TK activity with [^{3}H]thymidine (specific activity, 71.2 Ci/mmol; New England Nuclear) were as described previously (28).

RESULTS

HSV-1 terminal *a* sequence acts as a promoter-regulatory sequence for expression of HSV-1 genes. The structure of the constructs made to test the ability of the *a* sequence to function as a promoter-regulator for expression of the HSV-1 TK gene is shown in Fig. 1. The key construct pRB3119 (6) contained a *Bam*HI fragment that carried an intact *a* sequence flanked by the sequence GGATCCCCGG at the DR1b terminus and the sequence CACAGGTGTAAC CCCGGAATTCCCCGGATCC at the DR1c terminus. The purpose of the *Eco*RI cleavage site outside the *a* sequence but adjacent to DR1c was to permit rapid identification of the relative orientation of the *a* sequence. This construct was used in two series of experiments.

In the first series of experiments, the *Bam*HI fragment from pBR3119 was fused in both orientations to the *Bgl*II-*Bam*HI subfragment of the *Bam*HI Q fragment of HSV-1(F). The *Bgl*II-*Bam*HI subfragment of *Bam*HI-Q contains a portion of the 5′-transcribed noncoding and coding domains of the TK genes, extending downstream from nucleotide +50 at the *Bgl*II site (21, 28). The three constructs shown in Fig. 1 were cloned into the *Bam*HI site of the pSV2-neo plasmid (32). The plasmids carrying the chimeric *a* sequence-TK (*a*-TK) genes were then used to transform LTK$^-$ cells from G418 resistance. The resulting G418-resistant cell lines were then infected with the mutant HSV-1(F)Δ305 containing a 700-bp deletion in the coding sequences of the TK gene (28), and the cell lysates were tested for TK activity. HSV-1(F)Δ305 virus induced the resident *a*-TK in pRB3198 but not the resident TK gene introduced into the LTK$^-$ cells as pRB3197 or pRB3196 (Fig. 2A). As would be expected, the cells transformed by the vector plasmid pSV2-neo lacked TK activity. The induced resident TK gene was fused to the *a* sequence oriented such that the DR1 nearest to the *b* sequence (DR1b) was adjacent to the 5′-transcribed noncoding domain of the TK gene. The noninduced TK constructs either had no *a* sequence inserted (pRB3196) or were fused such that the DR1c was adjacent to the 5′-transcribed noncoding domain of the TK gene.

In the second series of experiments, the *Bam*HI fragment from pRB3119 containing the intact *a* sequence was inserted into the *Bgl*II site of the *Bam*HI Q fragment in an orientation such that DR1b was adjacent to the 5′-noncoding and -coding domains of the TK gene. The resulting construct (pRB3198) was cotransfected with intact HSV-1(F)Δ305 DNA into rabbit skin cells, and TK$^+$ virus progeny was selected and checked for the presence of the intact *a* sequence insert by *Bam*HI digestion. The recombinant selected for these studies (R3383) has been described previously (6). Specifically, a phenotypic property of this recombinant is that the *a* sequence insertion mediates additional inversions within the genome of the recombinant.

Figure 2B and C show the temporal patterns of expression of the TK genes of wild-type [HSV-1(F)] and of recombinant virus R3383, respectively, in 143TK$^-$ cells in the presence and absence of phosphonoacetate (250 μg/ml of medium), which is a potent inhibitor of HSV DNA synthesis. The results indicate that the temporal patterns of expression of the natural βTK gene resident in wild-type virus and the chimeric *a*-TK gene resident in the R3383 recombinant virus are similar, except that the TK activity of R3383 continues to increase as late as 16 h postinfection. However, whereas the induction of the expression of the *a*-TK chimeric gene was slightly reduced in the presence of phosphonoacetate, that of the βTK gene was actually increased in the presence of the drug. The slight decrease in activity in the presence of inhibitors of DNA synthesis is characteristic of and operationally defines the gene as belonging to the γ_1 group. Results of these experiments suggest that the *a* sequence contains a promoter-regulatory sequence and a transcription initiation site located at or near the DR1b terminus and therefore that a gene compatible with γ_1 regulation is located near the *a* sequence in the *b* inverted repeat sequence flanking the unique sequences of the L component.

Mapping of transcripts and demonstration of transcription initiation sites in the *a* sequence by S1 nuclease analysis. Two

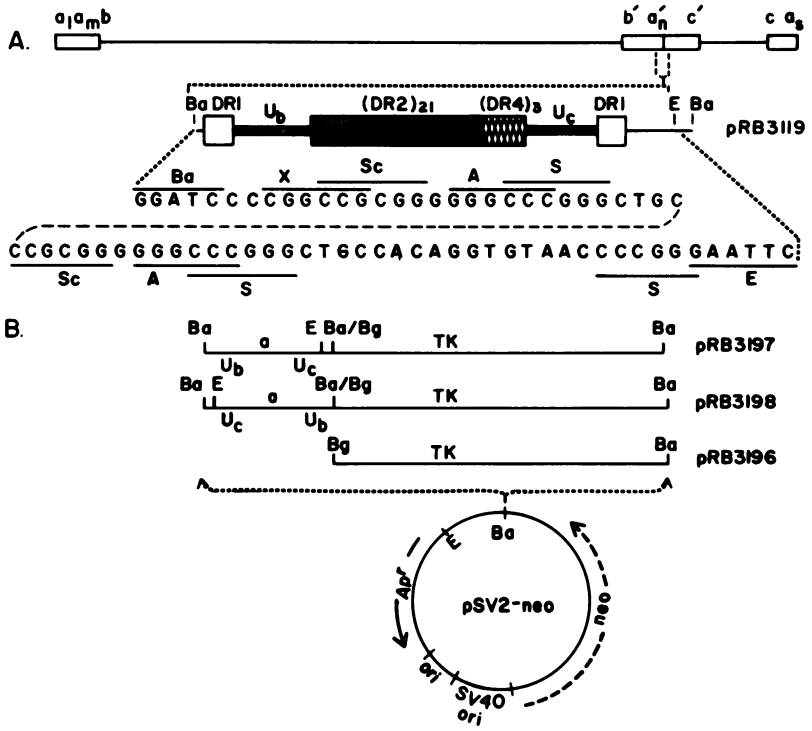

FIG. 1. Sequence arrangements and plasmid constructions. (A) Schematic diagram of the sequence arrangement in the HSV-1 genome and in the *a* sequence cloned as a *Bam*HI fragment. a_l and a_s refer to the terminal *a* sequence of the L and S components, respectively. The boxes on the top line represent the reiterated sequences (*a*, *b*, and *c*) flanking the unique sequences represented by a line. Subscript n represents one or more additional *a* sequences, and subscript m represents zero to greater than five additional *a* sequences. *b* and *c* are reiterated sequences that appear both at the genome termini and the junction region of the L and S components. The second line shows a schematic representation of the sequence arrangement of the *Bam*HI fragment cloned as pRB3119. DR1, U_b, DR2, DR4, and U_c are components of the *a* sequence described in the text. The third line represents the sequence at the termini of the construct shown in line 2; the dashed line represents the remainder of the *a* sequence. Not shown are the nine bases between the *Eco*RI and *Bam*HI sites at the right terminus of the construct. The dashed line represents the *a* sequence. Abbreviations: X, *Xma*III; Sc, *Sac*II; A, *Apa*I; S, *Sma*I; E, *Eco*RI. (B) Construction of recombinant pSV2-neo plasmids containing the *Bam*HI DNA fragment carrying the *a* sequence fused in both orientations to the *Bgl*II-*Bam*HI fragment carrying the 5′-transcribed noncoding and the coding sequences of the TK gene. The orientation of the *a* sequence in pRB3197 and pRB3198 was determined by the *Eco*RI cleavage site marker at the *c* terminus of the *a* sequence. pRB3196 carried an insert that lacked the *a* sequence. Except for the orientation and the absence of the *a* sequence in pRB3196, all inserts into pSV2-neo vector were isogenic. Abbreviations: Bg, *Bgl*II; Ba, *Bam*HI; E, *Eco*RI.

series of experiments were done. The purpose of the first was to determine whether RNA homologous to the reiterated *b* sequences immediately adjacent to the *a* sequence was present in infected cells. Experiments involving hybridization of electrophoretically separated RNA transferred to nitrocellulose sheets with the ^{32}P-labeled *Bst*EII-*Sac*I subfragment of the *Bam*HI S fragment (Fig. 3C) revealed the presence of three poly(A)$^+$ RNAs containing sequences homologous to the labeled DNA probe (Fig. 4). Of the three, two species of 1.3 and 5.2 kilobases (kb) accumulated uniquely in infected cells, whereas the third species of 3.2 kb was present in both infected and uninfected cells (Fig. 4B). It is noteworthy that the intensity of the host RNA species that hybridized with the viral DNA probe was higher in the infected than in the uninfected cells; this may reflect the extent of purification of the poly(A)$^+$ RNA rather than an absolute increase in host mRNA after infection.

The objective of the second series of experiments was to map the transcription-initiation sites of viral RNA homologous to the DNA in this region of the HSV-1 genome. In this experiment, pRB143 plasmid DNA (26) carrying the L component terminal *Bam*HI S fragment was cleaved with *Nco*I and *Bam*HI (Fig. 3C) and 5′ end labeled. The *Nco*I-*Bam*HI terminal fragment was purified and cleaved with *Ava*II. The *Ava*II site is located in the U_c sequence of the *a* sequence, whereas the *Nco*I site is located in the *b*-reiterated sequence 100 bp from the DR1b terminus of the *a* sequence (Fig. 3B). The fragment was denatured and hybridized to total cytoplasmic RNA extracted from Vero cells at 13 h postinfection with HSV-1(F). The DNA protected from digestion with S1 nuclease was subjected to electrophoresis under sequencing conditions, as was a sample of the undigested hybrid and a sequencing ladder of the *Nco*I-*Ava*II fragment used in the protection experiment. Two fragments differing by 14 nucleotides were protected from S1 nuclease digestion (Fig. 5). The results of this experiment indicate that the RNA that accumulated in the cytoplasm of infected cells is homologous to sequences located at the DR1b and U_b sequences of the inverted repeats of the L component. The larger protected DNA fragment (fragment 1) mapped the transcription initiation site in U_b, whereas the smaller fragment (fragment 2) mapped it in DR1b. We noted at this point that sequence analyses failed to show inverted repeats which could account for the presence of two DNA fragments protected from S1 nuclease digestion, as is the case for the transcription initiation site of the α27 gene (18).

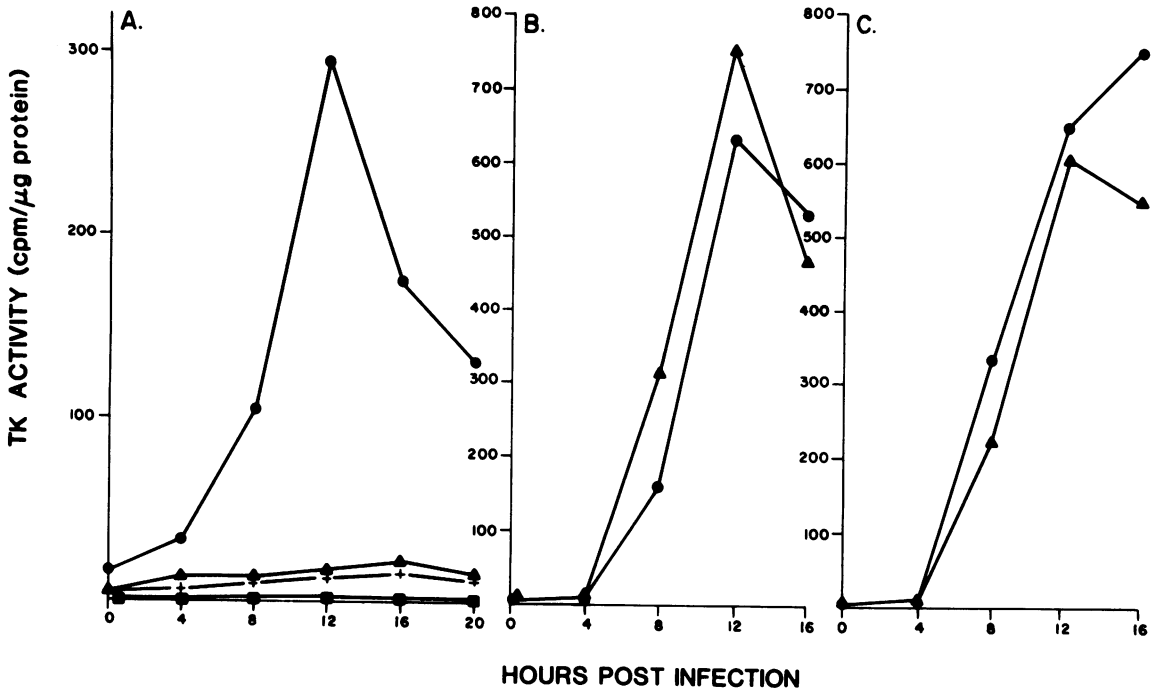

FIG. 2. TK activity per microgram of protein as a function of time for postinfection or mock infection. (A) Enzyme activity in LTK$^-$ cell lines converted to resistance to G418 by transfection with the pSV2-neo plasmid carrying the inserts shown in Fig. 1 and infected with 10 PFU of HSV-1(F)Δ305 per cell. The cell lines were designated according to the plasmid with which they were transfected, i.e., L3197, L3198, L3196, and LSV2-neo (converted to resistance to G418 with the pSV2-neo plasmid vector). Cytoplasmic extracts were assayed for TK activity at 0, 4, 8, 12, 16, and 20 h postinfection. Symbols: ●, L3198 cells; ▲, L3197 cells; +, L3196 cells; ■ LSV2 cells. (B) Enzyme activity in 143TK$^-$ cells infected with 5 PFU of HSV-1(F) per cell in the presence (▲) or absence (●) of phosphonoacetate (250 μg/ml of medium; Abbott Laboratories, North Chicago, Ill.). (C) Enzyme activity in 143TK$^-$ cells infected with 5 PFU of recombinant virus R3383 per cell in the presence (▲) or absence (●) of phosphonoacetate (250 μg/ml of medium). R3383 carries an *a* sequence in the *Bgl*II site of the TK gene in such orientation that U_b of the *a* sequence is proximal to the TK structural gene.

Sequence of the putative gene regulated by the promoter-regulatory domain of the *a* sequence. Earlier, this laboratory reported the nucleotide sequences of the *a* sequence (25) and the transcription initiation site and the promoter-regulatory domain of the α0 gene (19) located in its entirety in the *b*-reiterated sequence flanking the unique sequences of the L component of HSV-1(F) DNA. In this study we sequenced the domain of the *b* sequence located between the *a* sequence and the α0 gene. The sequencing strategy is shown in Fig. 3, and the fragments generated by this scheme were sequenced by the method of Maxam and Gilbert (20). Both strands were sequenced in each direction, and most domains were sequenced more than once. Codon usage analyses were done to assist in ensuring that the open reading frames were not artifacts of faulty sequencing. We sequenced 1,524 bp; to facilitate analyses of the results, Fig. 6 shows the sequence of 1,794 bp, i.e., from the last DR2 in the *a* sequence to the transcription initiation site of the α0 gene. The additional 270 bp are those described by Macken and Roizman (19). The highlights of the sequence (Fig. 6) are as follows.

(i) The nearest sequence which remotely resembles a TATAA box is the sequence TTTAAA located approximately 15 bp upstream from the second putative transcription initiation site. The transcription initiation site of the larger mRNA (fragment 1) maps at the end of the TTTAAA sequence (Fig. 4 and 5).

(ii) The first reading frame contains methionine codons at nucleotides 1215, 1392, and 1419 and a termination codon at nucleotide 1263. The second reading frame contains six methionine codons at nucleotides 191, 209, 421, 1040, 1259, and 1280. The relevant termination codons are at nucleotides 275 and 1223; although several termination codons precede the domain of the α0 gene, there are no methionine codons between the last relevant termination codon at nucleotide 1223 and the domain of the α0 gene. The largest predicted protein encoded between nucleotides 421 and 1223 would contain 267 amino acids. The third reading frame contains an initiation site at nucleotide 145 and termination codons at nucleotides 1219 and 1267, another methionine codon at nucleotide 1327, and a termination codon at nucleotide 1366. The largest protein predicted by this reading frame is a protein of 358 amino acids.

(iii) The significant predictive feature of the sequence is a 9-bp sequence repeated 10 times, starting with nucleotide 664 and terminating at nucleotide 753. This repeat codes for Asp-Pro-Arg in frame 1 (although it is not preceded by a methionine codon), Arg-Pro-Pro in frame 2, and Pro-Ala-Thr in frame 3. A synthetic peptide consisting of the oligomer Ala-Thr(Pro-Ala-Thr)$_9$Pro induced in rabbits an antibody which reacted with an infected cell protein with an apparent molecular weight of 43,500 in denaturing polyacrylamide gels (M. Ackermann, J. Chou, M. Sarmiento, R. Lerner, and B. Roizman, manuscript in preparation). This observation supports the hypothesis that translation begins in frame 3 at the first AUG codon after the transcription initiation sites and terminates at nucleotide 1219.

(iv) The apparent molecular weight of the protein (43,500) that reacted with the antibody to the synthetic peptide is larger than the molecular weight (37,054) of the protein predicted to be encoded by reading frame 3. This is consis-

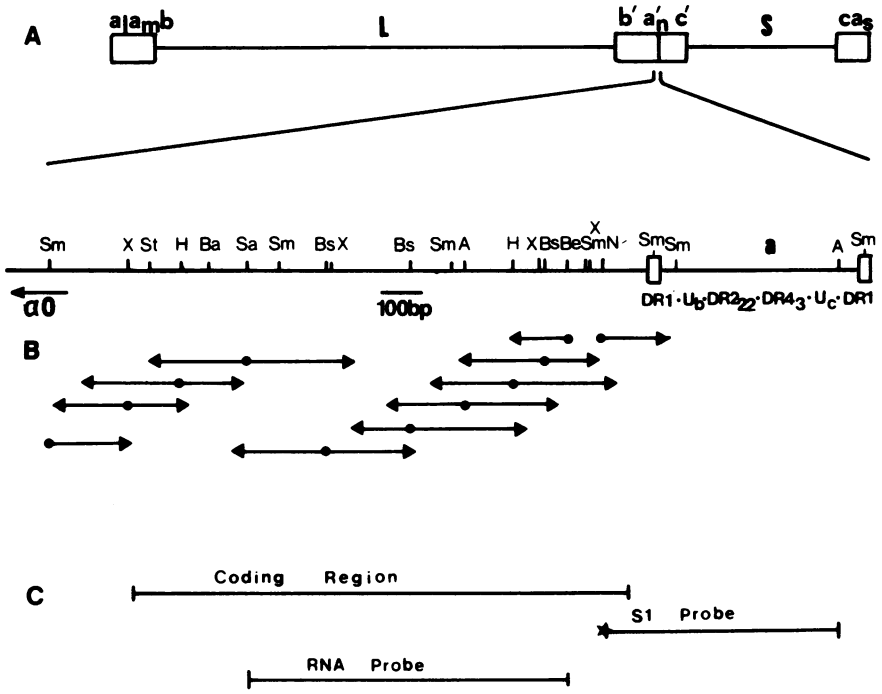

FIG. 3. Sequencing strategy and sequence arrangement of the HSV-1(F) genome domain between the *a* sequence and the site of initiation of the α0 gene transcription. (A) The top line shows the location of the sequence domain in the HSV-1 genome. The second line shows the restriction endonuclease sites and the domains of the sequenced region. Abbreviations: Sm, *Sma*I; *A*, *Ava*II; N, *Nco*I; X, *Xma*III; Be, *Bst*II; Bs, *Bss*HII; H, *Hin*fI; Sa, *Sac*I; Ba, *Bal*I; St, *Stu*I. The *Sma*I site at the far left of the map designates position −270 of the α0 gene domain reported by Macken and Roizman (19). (B) Sequencing strategy across the region shown in line 2 of part A. Dots represent the end-labeled nucleotide location and the start of each sequencing run. Arrows point in the direction of each run. The nucleotide sequence to the right of the *Bst*EII cleavage site was reported by Mocarski and Roizman (25). (C) The coding region of open reading frame 3, the domain of the *Nco*I-*Ava*II fragment probe used in the S1 nuclease analyses, and the *Bst*EII-*Sac*I fragment used to probe electrophoretically separated cytoplasmic RNA transferred to a nitrocellulose sheet.

tent with the observation that the electrophoretic mobility of the HSV-1 proteins that are rich in proline is that of proteins with molecular weights 10 to 25% greater than those predicted from the nucleotide sequence. The size of this protein is consistent with and could be encoded by the 1.3-kb cytoplasmic RNA detected by hybridization (see above).

(v) The special features of the protein predicted by the third reading frame, in addition to the Pro-Ala-Thr repeat, include a run of seven aspartic acids. The amino acid composition of the protein listed in Table 1 indicates a high proline (21%), alanine (17%), and arginine (14%) content and single methionine, isoleucine, tyrosine, and lysine residues.

(vi) The polyadenylation site available for the mRNA encoded by the third reading frame is the sequence ATTAAA which immediately follows the termination codon of the predicted sequence of the protein.

(vii) An interesting feature of the nucleotide sequence shown in Fig. 5 is the presence of numerous direct repeats within the coding sequence of the predicted protein and downstream, between the coding sequence of this protein and the coding sequence of the α0 gene. In addition to the CCCGCGACC sequence (R3) that is repeated 10 times, the coding sequence contain four other repeats. These include a 19-bp sequence (R1) that is repeated two times, an 11-bp sequence (R2) that is repeated two times, an 11-bp sequence (R4) that is repeated two times, and a 12-bp sequence (R5) that is repeated two times. Between the 3′ terminus of the coding sequence of this protein and the transcription initiation site of the α0 gene there are four sequences containing 25 (R6), 17 (R7), 35 (R8), and 41 (R9) bp, each of which is repeated twice. Mackem and Roizman (19) have reported that the promoter-regulatory domain of the α0 gene between nucleotides +1 and −330 contained three homologs of AT-rich sequences that were shown to be necessary for the induction of α genes by the α-*trans*-inducing factor (15) and numerous GC-rich inverted repeat sequences. The nucleotide sequence (Fig. 6) shows the presence of three homologs of the sequence GGTATGGTAAT which contain a portion of the AT-rich sequence. This is followed by nucleotide sequences that represent variants of the AT-rich sequence reported by Macken and Roizman (19) and Kristie and Roizman (15). The homolog GGTATGGTAAT most distal from the transcription initiation site of the α0 gene is at nucleotide −583 and overlaps with the terminus of the predicted coding sequences of the protein.

DISCUSSION

Results of earlier mapping studies have shown that the inverted repeats *ab* and *b'a'* of the L component and *c'a'* and *ca* of the S component contain the α genes 0 and 4, respectively, in their entirety. The discovery of a new diploid gene located in the *b* sequence of the L component was unanticipated. The following points are relevant to the results presented in this study.

The finding that the *a* sequence contains the promoter-regulatory domain and the transcription initiation sites of a gene was unexpected, inasmuch as the *a* sequence has already been shown to contain the *cis*-acting sites for the inversion of the L and S components (6, 24–26, 31) for the

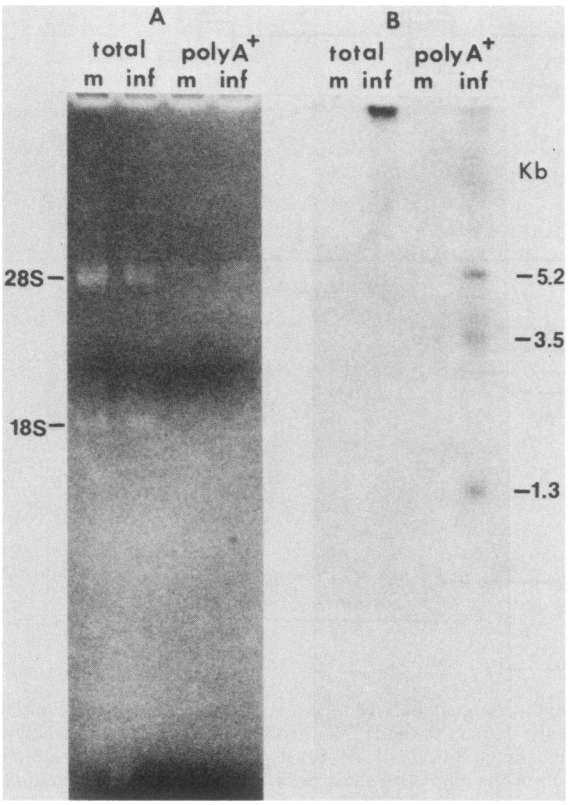

FIG. 4. Images of cytoplasmic RNA from infected (inf) and mock-infected (m) cells separated by electrophoresis. (A) Photographic images of ethidium bromide-stained 1.4% Formalin agarose gel. The lanes representing total RNA each contain 8 μg of total cytoplasmic RNA extracted from Vero cells at 13 h post-mock infection or post-infection with 20 PFU of HSV-1(F) per cell. The prominent RNA species visualized by ethidium bromide staining are 28S, 18S, and 5S. The lanes representing poly(A)$^+$ RNA each contain 8 μg of RNA purified by chromatography on oligo(dT) linked to cellulose from cytoplasmic RNA from mock-infected and infected cells. (B) Autoradiographic image of the RNA in part A after transfer to a nitrocellulose sheet and hybridization with the ^{32}P-labeled *Bst*EII-*Sac*I DNA fragment (Fig. 3C) of pRB143. This fragment contained approximately two-thirds of the total coding sequence of the gene identified in this study. Kb, kilobases.

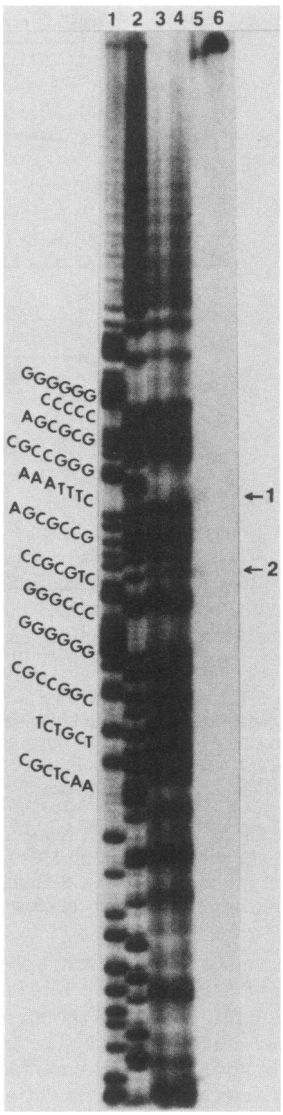

FIG. 5. Autoradiographic image of the ^{32}P-labeled DNA probe fragments protected from S1 nuclease digestion by hybridization with homologous RNA from the cytoplasm of infected cells. Lanes 1 through 4, DNA sequencing ladder of the *Nco*I-*Ava*II DNA probe labeled at the *Nco*I site; lane 1, G; lane 2, A > C; lane 3, C; lane 4, C + T; lane 5, DNA fragments protected from S1 nuclease digestion; lane 6, ^{32}P-labeled DNA-RNA hybrids not digested with S1 nuclease. In these experiments 150 μg of cytoplasmic RNA harvested from HSV-1(F)-infected Vero cells at 13 h postinfection was hybridized to a purified end-labeled *Nco*I-*Ava*II fragment from pRB143 shown in Fig. 3. The RNA-DNA hybrid was digested with S1 nuclease and subjected to electrophoresis. The numbers 1 and 2 on the right of the figure refer to DNA fragments protected from S1 digestion.

cleavage of the unit-length HSV-1 DNA from head-to-tail concatemers and for packaging of the DNA into capsids (35). Preliminary studies suggest that the promoter-regulatory domain of the gene is in the U_b and in the 12-bp repeat DR2 sequences (J. Chou and B. Roizman, manuscript in preparation). The nucleotide sequences of these domains are highly conserved among *a* sequences derived from different strains of HSV-1 (8, 23, 25).

The existence of a bona fide gene driven and regulated by a promoter-regulatory domain in the *a* sequence rests on the finding of (i) orientation-dependent *cis*-acting sites that confer transcription initiation and inducibility to an *a*-TK chimeric gene, (ii) cytoplasmic RNA homologous in sequence to the predicted transcribed domain of the gene, (iii) sequencing data indicating an open reading frame consistent with the size of the protein detected in infected cell lysates, and (iv) the presence in infected cell lysates of proteins of predicted apparent molecular weight reactive with antibody made against the Pro-Ala-Thr repeat unit predicted by the sequence. As noted above, the sequence TTTAAA is the nearest equivalent to a TATAA box and it maps closer to the second putative transcription initiation site than would have been predicted for a bona fide TATAA sequence. The absence of a bona fide punctuation sequence is not unusual per se; it may explain the apparent multiple initiation sites for the RNA. The polyadenylation signal ATTAAA occurs immediately after the termination codon rather than at some distance downstream, but its sequence is the only observed

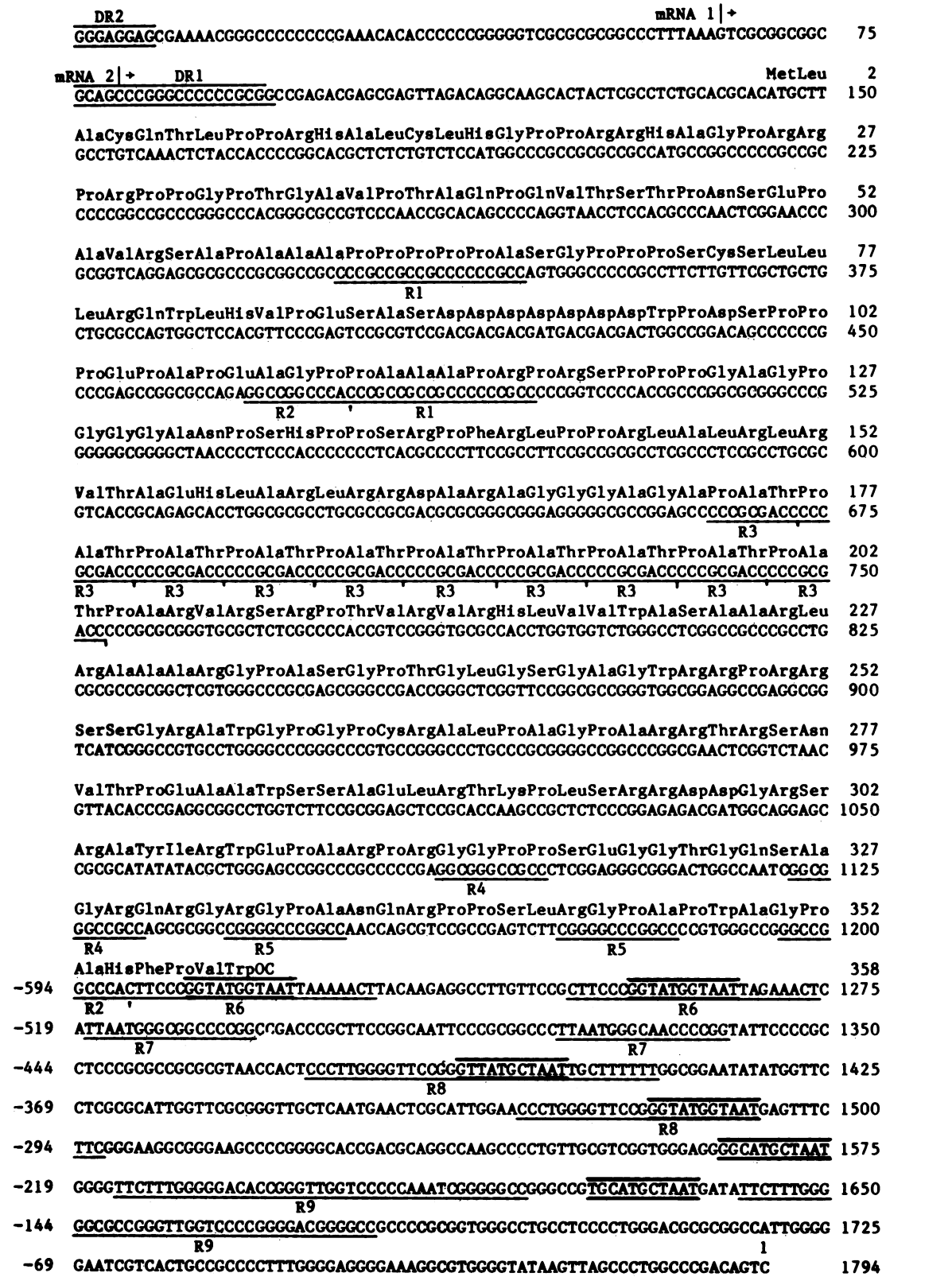

FIG. 6. The nucleotide sequence of the DNA fragment from the far left DR2 shown in Fig. 3 to the transcription initiation site of the α0 gene. The amino acid sequence of open reading frame 3 predicted from the nucleotide sequence is designated by the three-letter code. Reiterated sequences are underlined. The AT-rich homologs (consensus sequence GGTATGCTAAT) in the promoter regulatory domain of the α0 gene are indicated by heavy lines above and below the sequence.

TABLE 1. Amino acid composition of the protein predicted by reading frame 3

Amino acid	No. of residues	% total
Alanine	59	16.5
Asparagine	4	1.1
Cysteine	4	1.1
Glutamine	7	2.0
Histidine	8	2.2
Leucine	21	5.9
Phenylalanine	2	0.6
Serine	26	7.3
Tryptophan	9	2.5
Valine	12	3.3
Arginine	49	13.7
Aspartic acid	11	3.1
Glutamic acid	9	2.5
Glycine	36	10.1
Isoleucine	1	0.3
Lysine	1	0.3
Methionine	1	0.3
Proline	75	20.9
Threonine	22	6.1
Tyrosine	1	0.3

variant of the consensus AATAAA signal noted by Berget (2) in a survey of 61 polyadenylation signals.

The smaller 1.3-kb RNA species detected only in the infected cells and homologous to the domain of the gene demonstrated in this study are of sufficient size to encode the product of the gene predicted in this study. The larger mRNA is of sufficient size to represent a readthrough terminating at the 3′ terminus of the α0 gene. Nothing is presently known of the host RNA-containing sequences homologous to the domains of the gene reported here.

It could be predicted that the function of a gene, the promoter-regulatory domain of which overlaps with sequences carrying *cis*-acting sites of global significance to the genome (e.g., origins of DNA synthesis, *cis* sites for DNA packaging into capsids, site-specific DNA recombination sites), would be to act in conjunction with these sites. This is the case for genes involved in site-specific recombination. Examples of genes mapping next to such sites have been documented in the *H2* locus of *Salmonella*, in the G loop of phage Mu, near the integration site of phage lambda, and at the site of genome rearrangements of the yeast 2μm circle plasmid (3, 4, 22, 27, 33, 39, 40). Nothing is known of the function of this gene, and experiments to determine its role in the reproductive cycle of the virus are in progress. Studies with the antibody to the synthetic peptide indicate that the 43,500-apparent-molecular-weight protein carrying the antigenic determinants contained in the $(Ala\text{-}Thr\text{-}Pro)_{10}$ repeat is soluble, accumulates late in infection predominantly in the cytoplasm of the infected cell, and is hydrophilic in nature, as could be predicted by the Asp and Arg residues. The antibody did not react with proteins extracted from mock-infected or HSV-2-infected cells (Ackermann et al., in preparation). The presence of a single methionine and its comigration in denaturing polyacrylamide gels with one of the polypeptide bands of the infected cell polypeptide (ICP35) family may explain the failure to detected it in earlier studies. We have designated this protein as the ICP34.5. The patterns of regulation of theexpression of this gene and its slight sensitivity to phosphonoacetate suggest that it belongs to the γ_1 regulatory class.

ACKNOWLEDGMENTS

This study was supported by Public Health Service grants CA-08494 and CA-19264 from the National Cancer Institute and grant MV-2T from the American Cancer Society. J.C. is a U.S. Public Health Service predoctoral trainee (GM 07197).

LITERATURE CITED

1. **Batterson, W., and B. Roizman.** 1983. Characterization of the herpes simplex virion-associated factor responsible for the induction of alpha genes. J. Virol. **46**:371–377.
2. **Berget, S. M.** 1984. Are U_4 small nuclear ribonucleoproteins involved in polyadenylation? Nature (London) **309**:179–182.
3. **Broach, J. R., V. R. Guarascio, and M. Jayaram.** 1982. Recombination within the yeast plasmid 2μm circle is site-specific. Cell **29**:227–234.
4. **Bukhari, A., and L. Ambrosio.** 1978. The invertible segment of bacteriophage Mu DNA determines the adsorption properties of Mu particles. Nature (London) **271**:575–577.
5. **Campbell, M. E. M., J. W. Palfreyman, and C. M. Preston.** 1984. Identification of herpes simplex virus DNA sequences which encode a *trans*-acting polypeptide responsible for stimulation of immediate early transcription. J. Mol. Biol. **180**:1–19.
6. **Chou, J., and B. Roizman.** 1985. The isomerization of the herpes simplex virus I genome: identification of the *cis*-acting site and probable recombination site within the domain of the α sequence. Cell **41**:803–811.
7. **Conley, A. J., D. M. Knipe, P. C. Jones, and B. Roizman.** 1981. Molecular genetics of herpes simplex virus. VII. Characterization of a temperature sensitive mutant produced by in vitro mutagenesis and defective in DNA synthesis and accumulation of γ polypeptides. J. Virol. **37**:191–206.
8. **Davison, A. J., and N. M. Wilkie.** 1981. Nucleotide sequences of the joint between the L and S segments of herpes simplex virus type 1 and 2. J. Gen. Virol. **55**:315–331.
9. **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.
10. **Hayward, G. S., R. J. Jacob, S. C. Wadsworth, and B. Roizman.** 1975. Anatomy of herpes simplex virus DNA: evidence for four populations of molecules that differ in the relative orientations of their long and short segments. Proc. Natl. Acad. Sci. USA **72**:4243–4247.
11. **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.
12. **Honess, R. W., and B. Roizman.** 1975. Regulation of herpesvirus macromolecular synthesis: sequential transition of polypeptide synthesis requires functional viral polypeptides. Proc. Natl. Acad. Sci. USA **72**:1276–1280.
13. **Jenkins, F. J., and M. K. Howett.** 1984. Characterization of mRNAs that map in the *Bgl*II N fragment of the herpes simplex virus type 2 genome. J. Virol. **52**:99–107.
14. **Kieff, E. D., S. L. Bachenheimer, and B. Roizman.** 1971. Size, composition, and structure of the DNA of herpes simplex virus subtypes 1 and 2. J. Virol. **8**:125–132.
15. **Kristie, T. M., and B. Roizman.** 1984. Separation of sequences defining basal expression from those conferring α gene recognition within the regulatory domains of herpes simplex virus 1 α genes. Proc. Natl. Acad. Sci. USA **81**:4065–4069.
16. **Locker, H., and N. Frenkel.** 1979. *Bam*HI, *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**:424–441.
17. **Mackem, S., and B. Roizman.** 1982. Differentiation between the promoter and regulator regions of herpes simplex virus 1: the functional domains and sequence of a movable α regulator. Proc. Natl. Acad. Sci. USA **79**:4917–4921.
18. **Mackem, S., and B. Roizman.** 1982. Regulation of α genes of herpes simplex virus: the α27 promoter-thymidine kinase chimera is positively regulated in converted L cells. J. Virol. **43**:1015–1023.

19. **Mackem, S., and B. Roizman.** 1982. Structural features of the herpes simplex virus α gene 4, 0, and 27 promoter-regulatory sequences which confer α regulation of chimeric thymidine kinase genes. J. Virol. **44**:939–949.
20. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
21. **McKnight, S. L.** 1980. The nucleotide sequence and transcript map of the herpes simplex virus thymidine kinase gene. Nucleic Acids Res. **8**:5949–5964.
22. **Mizuuchi, K., R. Weisberg, L. Enquist, M. Mizuuchi, M. Bunaczynska, C. Foeller, P. L. Hsu, W. Ross, and A. Landy.** 1980. Structure and function of the phage lambda att site: size, int-binding sites, and location of crossover point. Cold Spring Harbor Symp. Quant. Biol. **45**:429–437.
23. **Mocarski, E. S., L. P. Deiss, and N. Frenkel.** 1985. The nucleotide sequence and structural features of a novel Us-*a* junction present in a defective herpes simplex virus genome. J. Virol. **55**:140–146.
24. **Mocarski, E. S., L. E. Post, and B. Roizman.** 1980. Molecular engineering of the herpes simplex virus genome: insertion of a second L-S junction into the genome causes additional genome inversions. Cell **22**:243–255.
25. **Mocarski, E. S., and B. Roizman.** 1981. Site specific inversion sequence of herpes simplex virus genome: domain and structural features. Proc. Natl. Acad. Sci. USA **78**:7047–7051.
26. **Mocarski, E. S. and B. Roizman.** 1982. The structure and role of the herpes simplex virus DNA termini in inversion, circularization and generation of virion DNA. Cell **31**:89–97.
27. **Nash, H. A.** 1977. Integration and excision of bacteriophage lambda. Curr. Top. Microbiol. Immunol. **78**:171–200.
27a. **Pellet, P. E., J. L. C. McKnight, F. J. Jenkins, and B. Roizman.** 1985. The nucleotide sequence and predicted amino acid sequence of a protein encoded in a small herpes simplex virus DNA fragment capable of trans-inducing α genes. Proc. Natl. Acad. Sci. USA **82**:5870–5874.
28. **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 the thymidine kinase with α gene promoters. Cell **24**:555–565.
29. **Sheldrick, P., and N. Berthelot.** 1975. Inverted repetitions in the chromosome of herpes simplex virus. Cold Spring Harbor Symp. Quant. Biol. **39**:667–678.
30. **Silver, S., and B. Roizman.** 1985. γ_2-thymidine kinase chimeras are identically transcribed but regulated as γ_2 genes in herpes simplex virus genomes and as β genes in cell genomes. Mol. Cell. Biol. **5**:518–528.
31. **Smiley, J., B. Fong, and W. C. Leung.** 1981. Construction of a double jointed herpes simplex virus DNA molecule: inverted repeats are required for segment inversion and direct repeats promote deletions. Virology **111**:345–362.
32. **Southern, P. J., and P. Berg.** 1982. Transformation of mammalian cells to antibiotic resistance with a bacterial gene under control of the SV40 early region promoter. J. Mol. Appl. Genet. **1**:327–341.
33. **van de Putte, P., S. Cramer, and M. Giphart-Gussler.** 1980. Invertible DNA determines host specificity of bacteriophage Mu. Nature (London) **286**:218–222.
34. **Vieira, J., and J. Messing.** 1982. The pUC plasmid, an M13 mp7-derived system for insertion, mutagenesis and sequencing with synthetic universal primers. Gene **19**:259–268.
35. **Vlazny, D. A., A. Kwong, and N. Frenkel.** 1982. Site specific cleavage packaging of herpes simplex virus DNA and the selective maturation of nucleocapsids containing full length viral DNA. Proc. Natl. Acad. Sci. USA **79**:1423–1427.
36. **Wadsworth, S., R. J. Jacob, and B. Roizman.** 1975. Anatomy of herpes virus DNA. II. Size, composition and arrangement of inverted terminal repetitions. J. Virol. **15**:1487–1497.
37. **Wagner, E.** 1985. Individual HSV transcripts: characterization of specific genes, p. 45–105. *In* B. Roizman (ed.), The herpesviruses, vol. 3. Plenum Publishing Corp., New York.
38. **Wagner, M. M., and W. C. Summer.** 1978. Structure of the joint region and the termini of the DNA of herpes simplex virus type 1. J. Virol. **27**:374–387.
39. **Zieg, J., M. Silverman, H. Hilmen, and M. I. Simon.** 1977. Recombinational switch for gene expression. Science **196**: 170–172.
40. **Zieg, J., and M. I. Simon.** 1980. Analysis of the nucleotide sequence of an invertible controlling element. Proc. Natl. Acad. Sci. USA **77**:4196–4200.