

# Localization and comparative nucleotide sequence analysis of the transforming domain in herpes simplex virus DNA containing repetitive genetic elements

(minimal transforming region/Z-DNA elements/G+C-rich sequences/insertion sequences)

CLINTON JONES, JENNIFER ORTIZ, AND RAXIT J. JARIWALLA\*

Laboratory of Viral Carcinogenesis, Armand Hammer Cancer Research Center, Linus Pauling Institute of Science and Medicine, Palo Alto, CA 94306

Communicated by Linus Pauling, July 14, 1986

**ABSTRACT** The 7.5-kilobase *Bam*HI E fragment (*Bam*HI-E) of herpes simplex virus type 2 (HSV-2) DNA (map position 0.533-0.583) encodes the 144-kDa subunit of ribonucleotide reductase and induces the neoplastic transformation of immortalized cell lines. To define the minimal transforming region of *Bam*HI-E, a series of subclones were constructed that spanned the entire fragment. These subclones were assayed for focus formation in Rat-2 cells. Removal of the promoter region from the viral 144-kDa-protein gene left the transforming activity of DNA clones intact. A 481-bp *Pst* I-*Sal* I subclone of *Bam*HI-E was capable of inducing focus formation and tumorigenic conversion. The nucleotide sequence of this fragment and the colinear nontransforming region of HSV-1 DNA was determined and compared. Striking differences were detected in the structure and organization of repeated sequence elements. Specifically, transforming HSV-2 DNA contains multiple regions of alternating purines and pyrimidines, G+C-rich sequences that are potential binding sites for transcription factor Sp1, and insertion-like sequence elements that are interrupted by base substitutions in nontransforming HSV-1 DNA. These results define a distinct transforming domain in HSV-2 DNA composed of repetitive elements implicated in gene rearrangement and activation.

Three distinct, nonhomologous transforming regions (*mtr*I, *mtr*II, and *mtr*III) have been mapped in the genomes of herpes simplex viruses (HSV-1 and HSV-2) (1-6). The *mtr*III region in HSV-2 DNA is defined by the *Bgl* II C fragment (*Bgl* II-C, 0.42-0.58 map unit), which induces the multistep transformation of nonestablished diploid cells (3, 4). The left-hand 64% subfragment of *Bgl* II-C causes cellular immortalization without tumorigenesis (4), whereas the right-hand 30% (the *Bam*HI E subfragment, *Bam*HI-E) induces tumorigenic transformation of immortalized cell lines (7, 8). The minimal transforming region of the 7.5-kilobase (kb) *Bam*HI-E has not been mapped. *Bam*HI-E contains the entire protein-coding sequence of the 144-kDa subunit of ribonucleotide reductase (9, 10), but it has not been established that the viral protein or a truncated derivative is necessary to induce neoplastic transformation. Cells transformed by *Bam*HI-E DNA occasionally express the 144-kDa polypeptide, but continuous expression of this protein is not required to maintain the transformed phenotype (7, 8).

The present study was designed to (i) localize and sequence the minimal transforming region of HSV-2 *Bam*HI-E and (ii) identify the DNA sequence elements responsible for the transforming potential of *Bam*HI-E. In this paper, we present a detailed molecular and sequence analysis of HSV-2 *Bam*HI-E that led to the identification of the minimal trans-

forming domain. A preliminary report of these studies has been presented.<sup>†</sup>

## MATERIALS AND METHODS

**Transformation Assays.** Growth conditions for Rat-2 cells, DNA transfection, and assays for focus formation, anchorage independence, and tumorigenicity were as described by Jariwalla *et al.* (7).

**Construction of Recombinant Subclones.** For construction of *mtr*III subclones of HSV-2 strain 333, recombinant plasmid pGH17a (7), containing *Bam*HI-E, was digested with *Bam*HI/*Pst* I. Restriction fragments were separated by agarose gel electrophoresis, electrophoretically eluted, purified by two phenol extractions, and precipitated by ethanol. The *Pst* I fragments A and C (see Fig. 1) were ligated into the *Pst* I site of phage M13 mp8 to produce mEA-11 and mEC-23. Fragments B and D were not stable in M13mp8 or -mp9 phage vectors. Thus, B and D were cloned into the *Pst* I site of pBR322, yielding pEB-12 and pED-6, respectively. The terminal *Bam*HI/*Pst* I fragments of *Bam*HI-E, E and F, were cloned into *Bam*HI/*Pst* I-digested M13mp8 and designated mEE-3 and mEF-1. The 821-base-pair (bp) and 481-bp fragments obtained by digestion of *Pst* I-C with *Sal* I were cloned in M13mp8 and designated mCPS-5 and mCPS-15, respectively. The 0.7- and 2.2-kb *Bam*HI fragments of HSV-1 strain F, which are homologous to the *Pst* I-C region of HSV-2, were isolated from a recombinant plasmid containing the *Hind*III-K region (0.52-0.59 map unit) and cloned in M13mp19. The size, identity, and orientation of the subclones were verified by restriction analysis of recombinant plasmid or replicative form (RF) phage DNA and Southern blot (11) hybridization of gel-purified HSV inserts to each clone. All pBR322 clones were propagated in *Escherichia coli* strain HB101, and plasmid DNA was isolated as described (7). Recombinant phage were grown in *E. coli* JM101, and phage DNA was prepared as described (12).

**DNA Sequence Analysis.** Sequences of the HSV-2 and HSV-1 clones were determined by the dideoxy chain-termination method of Sanger *et al.* (13), using single-stranded M13 recombinant phage DNA as templates (12). DNA sequences were assimilated and analyzed with the aid of International Biotechnologies (New Haven, CT) and Inteligenetics (Palo Alto, CA) computer programs.

## RESULTS

**Localization of the Smallest Transforming Region.** As HSV-2 *Bam*HI-E is rather large (7.5 kb), our approach to defining the minimal transforming region consisted of localization of

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: HSV-1 and HSV-2, herpes simplex virus types 1 and 2; bp, base pair(s); kb, kilobase(s).

\*To whom reprint requests should be addressed.

<sup>†</sup>Jones, C., Ortiz, J. & Jariwalla, R. J. (1986) American Society for Virology Annual Meeting, June 23-26, 1986, University of California, Santa Barbara, CA.

the smallest transforming fragment, followed by DNA sequence analysis. The overall strategy was to subclone and analyze restriction fragments from enzyme digests that left the transforming activity of *Bam*HI-E DNA intact. Fig. 1 shows the physical map of relevant restriction enzyme sites determined in our laboratory and the locations of the 5.0-kb transcript and the coding region of the 144-kDa protein determined in other studies (9, 10). Digestion of cloned *Bam*HI-E DNA with restriction endonucleases *Pst* I, *Sal* I, *Sac* I, and *Pvu* II followed by assay of total digests in Rat-2 cells did not result in significant reduction of transforming activity (data not shown). The six individual *Pst* I fragments were subcloned in plasmid pBR322 or M13 phage vectors and assayed for transformation. The results of three independent experiments (Fig. 2) indicated that only the 1.3-kb *Pst* I-C fragment derived from the interior of *Bam*HI-E exhibited significant transforming activity. Further experiments were designed to confirm this result and to localize the transforming activity in a smaller fragment.

The *Pst* I-C region overlaps with *Sal* I-B and *Sal* I-C fragments (Fig. 1) of which only *Sal* I-B manifested transforming activity in two independent trials. Restriction enzyme *Sal* I cuts *Pst* I-C into left-hand 821-bp and right-hand 481-bp subfragments, which were isolated, cloned, and used in transformation assays. Transforming activity was consistently detected in the 481-bp *Sal* I-*Pst* I subclone, at a higher level compared to *Bam*HI-E (Fig. 2). The 821-bp *Pst* I-*Sal* I subclone and vector plasmid or replicative form (RF) phage DNAs did not induce significant focus formation. These results place the minimal transforming region within the *Sal*

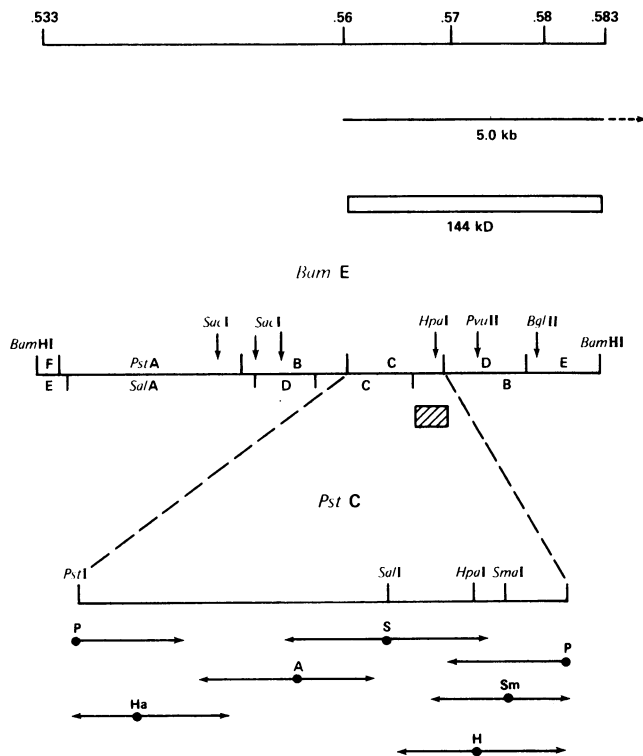


FIG. 1. Structural organization of the HSV-2 *Bam*HI-E and *Pst* I-C regions. From top to bottom are the following: map coordinates along the HSV-2 genome; positions of the mRNA (arrow) and 144-kDa ribonucleotide reductase protein (box); physical map of restriction sites within *Bam*HI-E (*Pst* I fragments A-F and *Sal* I fragments A-E are delimited by vertical lines above and below the map, respectively); hatched box shows the location of the minimal transforming region determined in this study); physical map of restriction sites within *Pst* I-C and sequencing strategy employed in this study (arrows). Abbreviations for restriction sites: P, *Pst* I; Ha, *Hae* III; A, *Alu* I; S, *Sal* I; H, *Hpa* I; Sm, *Sma* I.

I-B/*Pst* I-C overlap, mapping between coordinates 0.567 and 0.570 on the HSV-2 genome.

To determine whether *Pst* I-C and its 481-bp subclone induced full or partial transformation, individual foci were isolated and their phenotypic properties were determined. Cell lines derived from foci induced by either DNA fragment formed large colonies in soft agar and induced progressively growing tumors in immunocompetent syngeneic Fischer rats (Fig. 2). In contrast, lines derived from small, rare foci induced by plasmid or phage vectors did not grow efficiently in soft agar and were nontumorigenic in immunocompetent syngeneic animals.

**Sequence Analysis of the HSV-2 Transforming Region.** To deduce the nucleotide sequence of the *Pst* I-C region (map position 0.562–0.570) of HSV-2 strain 333, we employed the strategy shown in Fig. 1. Fragments spanning the *Sal* I, *Hpa* I, *Sma* I, and *Pst* I restriction sites and *Alu* I and *Hae* III subfragments of *Pst* I-C, cloned in M13 mp8 and mp9, were used. Sequence analysis of *Pst* I-C (Fig. 3) revealed the following features: (i) the complete fragment is 1302 bp long; (ii) a portion of the coding region for the 144-kDa protein (10) is located within *Pst* I-C, but the transcriptional regulatory signals (ATAAAA box and 22-bp A+C-rich sequence) of the 144-kDa early promoter are not contained within this fragment, and the termination codon for the 144-kDa protein and the polyadenylation signal for the 5.0-kb message also lie outside of *Pst* I-C; (iii) there are short open reading frames of up to 232 residues long in the nontransforming 821-bp region and short frames up to 154 residues in the reverse orientation of the 481-bp transforming region. However, no upstream TATA or CAAT boxes are present.

A striking feature of the *Pst* I-C sequence is the presence and clustering of repetitive DNA elements in the minimal transforming region (Fig. 3). Four distinct elements are recognized, as summarized in Table 1. In a 320-bp segment within the transforming *Sal* I-B/*Pst* I-C region, there are five stretches of alternating purine/pyrimidines, three CCGCCC boxes, five purine-rich regions, and three pyrimidine-rich elements. Except for the latter, the remaining elements are present at low abundance and at scattered locations within the 821-bp nontransforming *Pst* I-C/*Sal* I-C subfragment. The occurrence of repetitive elements with the potential to form stem-loop structures in the 481-bp *Sal* I-B/*Pst* I-C region is indicated in Fig. 4. In the five stem-loops, there is only one base mismatch in each stem sequence. Of potential interest is the structure containing a 62-bp loop and a stem flanked by direct repeats consisting of 7 of 10 matching nucleotides.

**Sequence Analysis of the Nontransforming HSV-1 Region.** Two independent groups have mapped the transforming region of HSV-1 DNA between map coordinates 0.30 and 0.45 (1, 2). This region does not exhibit homology to the transforming *mtr*II and *mtr*III regions of HSV-2 DNA. Furthermore, previous attempts by other investigators to detect transforming activity in the 0.50–0.60 region of HSV-1 were negative despite similarity in the genetic organization of the ribonucleotide reductase subunits encoded by the two strains (9, 10, 14). The cloned *Hind*III-K fragment (0.52–0.59 map unit) of HSV-1 did not exhibit significant transforming activity in Rat-2 cells (data not shown). To identify any sequence differences between the two strains that may explain the transforming potential of HSV-2, the HSV-1 *Hind*III-K clone was cleaved with *Bam*HI and a Southern blot was probed with HSV-2 *Pst* I-C DNA. Two hybridizing *Bam*HI fragments, of 0.7 and 2.2 kb, were isolated, cloned in M13mp19, and sequenced. The DNA sequence was aligned and compared with the colinear *Sal* I-B/*Pst* I-C region of HSV-2 DNA (Fig. 3). Within the 500-bp region, there are 71 nucleotide differences. Of potential significance are the

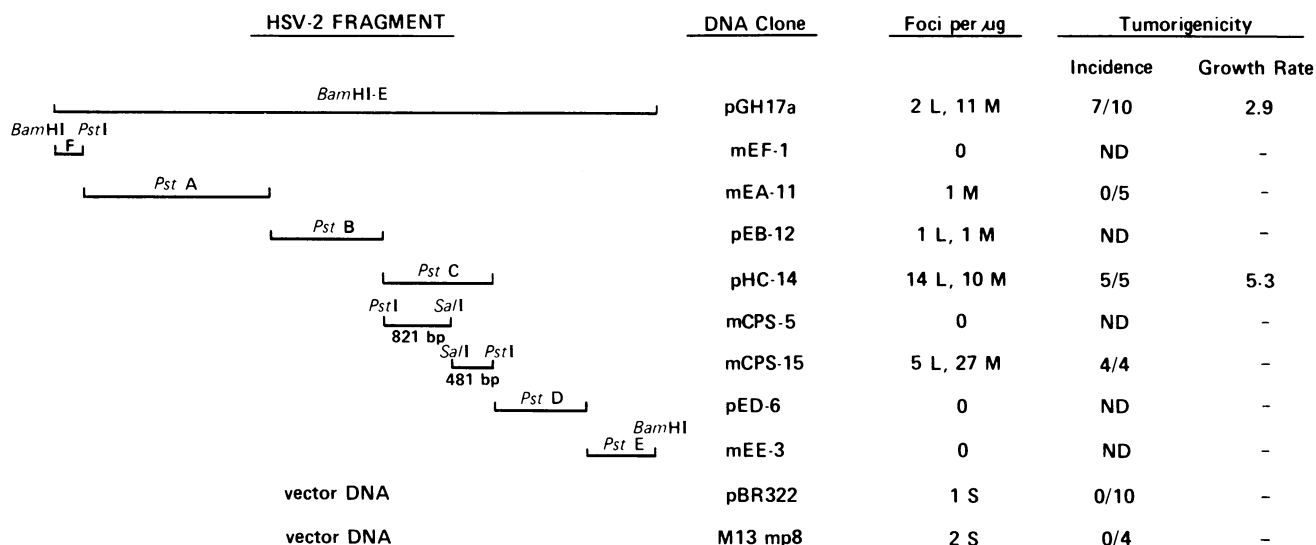


FIG. 2. Transforming potential of HSV-2 *Bam*HI-E subclones. Focus-formation data are the average of three experiments (focus size: L, large; M, medium; S, small). Tumorigenicity of isolated foci was assayed in syngeneic immunocompetent rats, and growth rate is expressed as  $\text{cm}^2/\text{week}$ . ND, analysis not done.

effects of these base changes on the structure and arrangement of repetitive elements (Fig. 3). Unlike transforming HSV-2 DNA, which contains five stretches of alternating purines/pyrimidines and three CCGCC boxes in a 320-bp region, nontransforming HSV-1 contains only one of each element. The influence of HSV-1 nucleotide differences on the formation of stem-loops is shown in Fig. 4. Each stem sequence is interrupted by two or more mismatched bases, and the flanking direct repeats contain multiple nucleotide substitutions.

DISCUSSION

The molecular analysis of the HSV-2 *Bam*HI-E fragment presented in this paper has localized the minimal transforming region to a 481-bp DNA segment. This segment contains repetitive sequences which map within *Sal* I-B/*Pst* I-C (map coordinates 0.567-0.570). The transforming sequences are derived from the coding region of the 144-kDa viral protein that lacks the upstream promoter and 3' transcription termination signals. Thus, it seems unlikely that a truncated viral

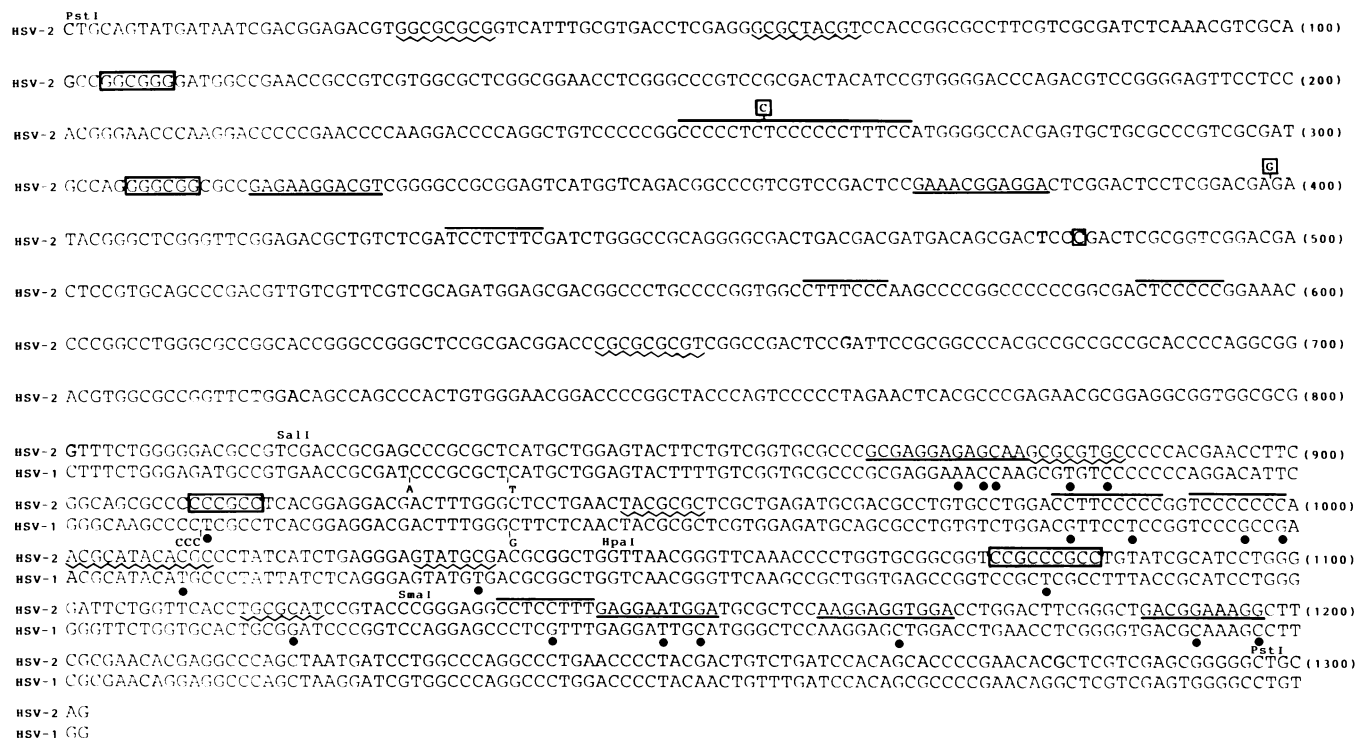


FIG. 3. Nucleotide sequence of the transforming HSV-2 domain and comparison with nontransforming HSV-1 DNA. The HSV-2 sequence presented is of the 1.3-kb *Pst* I-C region, in the prototypic orientation. Minimal transforming region extends from the *Sal* I to the *Pst* I site (nucleotides 822-1302). Single boxed nucleotides in HSV-2 DNA signify differences existing with the Swain and Galloway sequence (10). Four classes of repetitive elements are indicated: pyrimidine-rich (underlined), purine-rich (overlined), alternating purine/pyrimidine (zigzag underlined), and CCGCC sites (boxed). Base substitutions within repeats in HSV-1 DNA are shown by dots; the insertions are shown by vertical lines.

Table 1. Comparison of HSV-2 repeat elements to HSV-1 sequence

Spl binding site		Alternating purine/pyrimidine ( $n \geq 7$ )		Purine-rich ( $\geq 9/10$ )	
Sequence*	Position	Sequence	Position	Sequence†	Position
GGGGCGGGGC T AAT	(Consensus)	GCGCGTGC T C	879-886	GAGGAGAGCAAG A C	868-879
GGGGCGGAGT A	910-919	TACGCGC	946-952	GAGGGAGT C	1023-1030
CGGGCGGACA A	1078-1087	ACGCATACACGC T	1001-1012	GAGGAATGGA T C	1144-1153
CGGGCGGACC A	1082-1073	GTATGCG T	1029-1035	AAGGAGGTGGA C	1162-1172
		TGCGCAT G	1115-1121	GACGGAAAGG C C	1188-1197
				GGAAGGGGGC C A C	981-990
				CAGGGGGGG C C	991-999
				GGAGGAAA C	1136-1143

HSV-1 residues differing from the HSV-2 sequence are shown below the sequence.

\*Residues differing from the consensus sequence are underlined. First two sequences are direct complements, whereas third sequence (residues 1082-1073) is inverted complement of sequences at corresponding positions shown in Fig. 3.

†Last three purine-rich sequences are direct complements of pyrimidine-rich sequences shown in Fig. 3.

protein is synthesized or required to initiate neoplastic transformation. The significance of small open reading frames in the opposite sequence orientation remains to be determined. The repeated sequences in the transforming domain resemble DNA elements known to play a role in genetic recombination and control of gene activity in mammalian cells. The data are consistent with a mechanism of transformation in which repetitive HSV-2 DNA sequences interact with cellular genes, resulting in DNA rearrangements

and gene activation associated with neoplastic transformation.

Nucleotide sequence analysis of the transforming *Pst* I-C fragment subcloned from HSV-2 strain 333 revealed >99% identity to the sequence of the corresponding region from the same HSV-2 strain recently deduced by Swain and Galloway (10). Only three single-nucleotide differences were detected in a sequence of 1302 bases (Fig. 3). By comparison, about 14% base mismatch was detected when HSV-1 DNA was

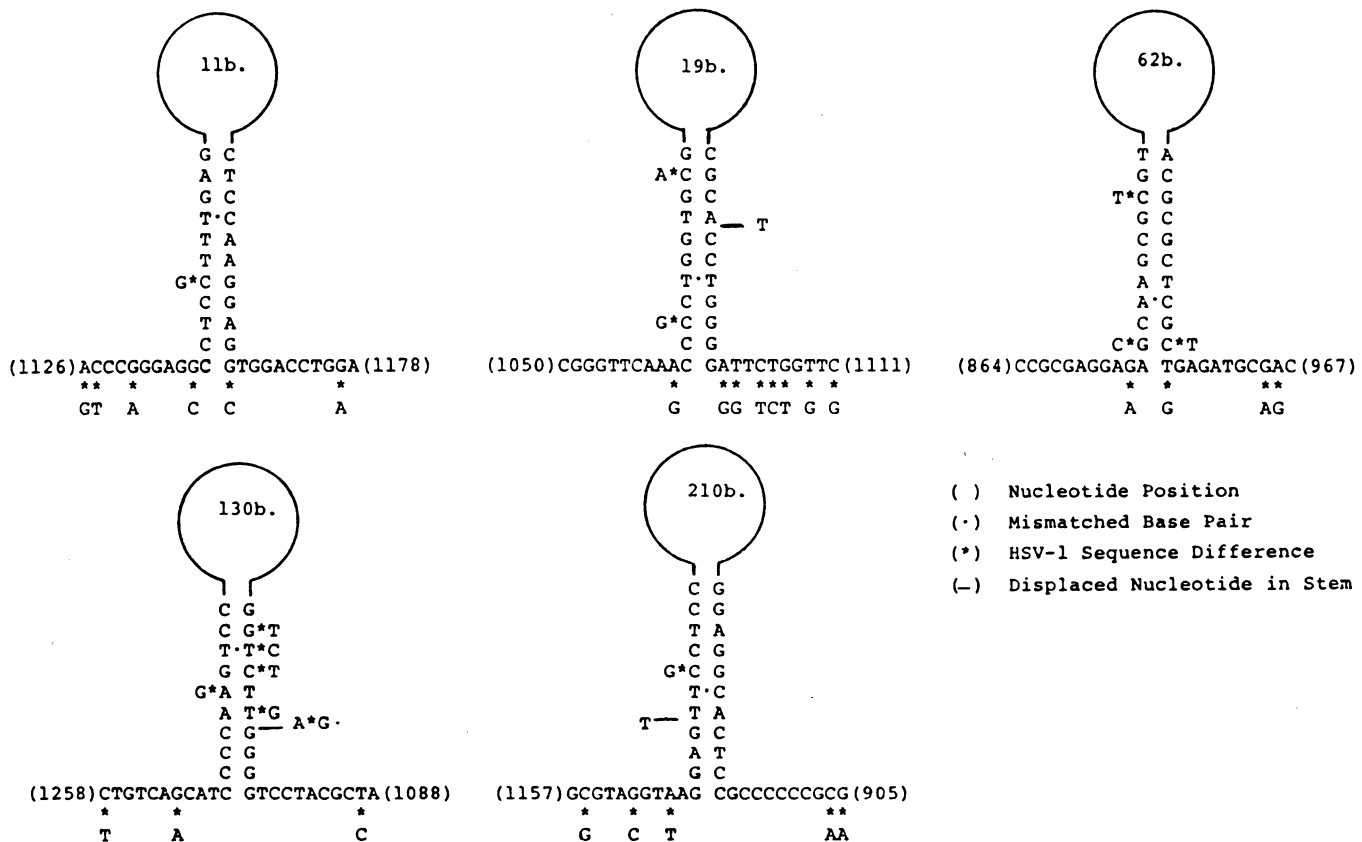


FIG. 4. Potential stem-loop structures in the minimal transforming domain of HSV-2 DNA.

aligned with the sequence of the transforming HSV-2 domain. These base substitutions in HSV-1 DNA disrupt the structure of repeated elements (Table 1) and stem-loops (Fig. 4) seen in HSV-2 DNA.

Is the 144-kDa HSV-2 protein involved in transformation? Huszar and Bacchetti (15) have proposed that the ribonucleotide reductase function associated with this protein may initiate transformation by inducing mutations in cellular genes. The catalytic site of the reductase has been mapped to the carboxyl terminus, which is encoded outside the *Pst* I-C region, indicating that the enzyme activity is not required to induce transformation. Swain and Galloway detected a structurally unique domain of 355 hydrophilic amino acid residues in the amino terminus of the 144-kDa HSV-2 protein that was absent from the ribonucleotide reductases encoded by Epstein-Barr virus, *E. coli*, and mice and poorly conserved in the HSV-1 enzyme (10). Our results indicate that this hydrophilic domain is also not required to induce transformation. Thus, the enzyme *Pst* I cleaves within this domain, resulting in its structural disruption, without reduction of transforming activity. Second, the *Sal* I-C fragment, which encodes an intact hydrophilic domain, did not exhibit significant transforming activity. Third, the minimal 481-bp fragment encodes only 16 of the 355 hydrophilic amino acids.

What elements are responsible for transformation? Our results support a role for noncoding repeated sequence motifs involved in DNA transposition and modulation of cellular gene activity. Thus, the HSV-2 transforming domain contains four candidate elements (Table 1, Fig. 4). Alternating purine/pyrimidine sequences occur in stretches of 7–12 nucleotides and are repeated five times within a 241-bp region in the transforming HSV-2 domain. This element occurs only twice in the 500-bp nontransforming HSV-1 region. The purine/pyrimidine sequences can form Z-DNA, which has been implicated in homologous recombination (16) and destabilization of chromatin structure (17). The cluster of Z-DNA tracts may be a hot spot for recombination and can contribute a transforming function by exerting cis-acting effects on expression of cellular genes located in the vicinity of viral insertion sites. Both up-regulation (18) and shut-off (19) of specific host genes have been detected in cells transformed by HSV-2 DNA. Alternatively, recombination of transforming viral sequences with host DNA, followed by excision, replication, and reinsertion, can also lead to alteration of cellular genes. Data in support of this mechanism are the rescue from HSV-2-transformed cells of extrachromosomal plasmids containing portions of HSV-2 *Bam*HI-E fused to cellular sequences (7). The same transformed cells also exhibit amplification and rearrangements of specific cellular DNA segments (7).

Another element capable of modulating gene expression is the hexanucleotide CCGCCC. This motif is present in three copies within a 175-bp stretch of the HSV-2 domain, occurring once as an inverted sequence. The GGGGCGGAGT element (residues 910–919) matches perfectly the consensus sequence that binds the Sp1 transcription factor (20). The other two G+C-rich decanucleotides in the transforming domain exhibit strong homology to the Sp1 consensus sequence, differing by 2 or 3 bp (Table 1). Oncogenes of the *ras* gene family are regulated by four to eight 10-bp G+C-rich sequences (20). Recently, three copies of the inverted se-

quence CCGCC were detected in the strong enhancer of human cytomegalovirus DNA (21).

Purine- or pyrimidine-rich motifs in the minimal transforming region resemble enhancer core sequences and may play a role in enhancer function. The stem-loop structures (Fig. 4) may contribute a transforming function by promoting DNA insertions and transpositions associated with neoplastic conversion. A likely candidate for these reactions is the structure containing a 61-bp loop and a stem flanked by direct repeats (Fig. 4). The *Bgl* II-N *mtrII* region of HSV-2 DNA also contains a stem-loop structure resembling an insertion element within its 730-bp minimal transforming region (5). It is not known whether the various repeated elements can cooperate with each other and whether they are recognized by one or several cellular factors.

**Note Added in Proof.** The 481-bp transforming HSV-2 fragment hybridizes intensely to dispersed repetitive sequences of the human and monkey genomes. Within the HSV-2 transforming domain there is a region of 25 nucleotides (residues 928–906) that exhibits strong homology to regions in the simian virus 40 (SV40) origin of replication (*ori*) and to the "consensus" *Alu* repeat sequence that falls within SV40 *ori* (22).

We thank Dr. E. Blaisdell for help with computer analysis of DNA sequences, Dr. E. Zuckerkindl for constructive discussions, and S. Schwoebel for processing the manuscript. Part of this research was supported by the Robert Sabin Fund.

1. Reyes, G., LaFemina, R., Hayward, S. & Hayward, G. (1980) *Cold Spring Harbor Symp. Quant. Biol.* **44**, 629–641.
2. Camacho, A. & Spear, P. G. (1978) *Cell* **15**, 993–1002.
3. Jariwalla, R. J., Aurelian, L. & Ts'o, P. O. P. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 2279–2283.
4. Jariwalla, R. J., Aurelian, L. & Ts'o, P. O. P. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 5902–5906.
5. Galloway, D. A., Nelson, J. A. & McDougall, J. K. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 4736–4740.
6. Cameron, I. R., Park, M., Dutia, B. M., Orr, A. & Macnab, J. C. M. (1985) *J. Gen. Virol.* **66**, 517–527.
7. Jariwalla, R. J., Tanczos, B., Jones, C., Ortiz, J. & Salimi-Lopez, S. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 1738–1742.
8. Hayashi, Y., Iwasaka, T., Smith, C. C., Aurelian, L., Lewis, G. K. & Ts'o, P. O. P. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 8493–8497.
9. McLauchlan, J. & Clements, J. B. (1983) *EMBO J.* **2**, 1953–1961.
10. Swain, M. A. & Galloway, D. A. (1986) *J. Virol.* **57**, 802–808.
11. Southern, E. M. (1975) *J. Mol. Biol.* **98**, 503–517.
12. Schrier, P. H. & Cortese, R. (1979) *J. Mol. Biol.* **129**, 169–172.
13. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
14. Anderson, K. P., Frink, R. J., Devi, G. B., Gaylord, B. H., Costa, R. H. & Wagner, E. K. (1981) *J. Virol.* **37**, 1011–1027.
15. Huszar, D. & Bacchetti, S. (1983) *Nature (London)* **302**, 76–79.
16. Kmiec, E. B. & Holloman, W. K. (1986) *Cell* **44**, 545–554.
17. Nickol, J., Behe, M. & Felsenfeld, G. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 1771–1775.
18. Macnab, J. C. M., Orr, A. & La Thangue, N. B. (1985) *EMBO J.* **4**, 3223–3228.
19. Leavitt, J., Gunning, P., Kedes, L. & Jariwalla, R. J. (1985) *Nature (London)* **316**, 840–842.
20. Ishi, S., Kadonaga, J., Tjian, R., Brady, J. N., Merlino, G. T. & Pastan, I. (1986) *Science* **232**, 1410–1413.
21. Boshart, M., Weber, F., Jahn, G., Dorsch-Hasler, K., Fleckenstein, B. & Schaffner, W. (1985) *Cell* **41**, 521–530.
22. Johnson, E. M. & Jelinek, W. R. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 4660–4664.