JOURNAL OF VIROLOGY, May 1987, p. 1559–1570
0022-538X/87/051559-12$02.00/0

Vol. 61, No. 5

Multiple Tandemly Repeated Binding Sites for Cellular Nuclear Factor 1 That Surround the Major Immediate-Early Promoters of Simian and Human Cytomegalovirus

KUAN-TEH JEANG,[1]† DAN R. RAWLINS,[2]‡ PHILIP J. ROSENFELD,[2] JAMES H. SHERO,[1] THOMAS J. KELLY,[2] AND GARY S. HAYWARD[1]*

The Virology Laboratories, Department of Pharmacology and Molecular Sciences[1] and Department of Molecular Biology and Genetics,[2] The Johns Hopkins University School of Medicine, Baltimore, Maryland 21205

Received 20 October 1986/Accepted 6 January 1987

We show that the large DNA genomes of human and simian cytomegaloviruses (HCMV and SCMV, respectively) each contain multiple binding sites for purified cellular nuclear factor 1 (NF1) protein. Examination of the major immediate-early (IE) gene region in the *Hin*dIII H fragment of SCMV(Colburn) by filter binding assays showed that it competed 45-fold better than the single adenovirus type 2 binding site for NF1 protein and that it contained at least two distinct binding loci. Direct DNase I footprinting analyses of the 5′ upstream locus detected at least 20 adjacent NF1-binding sites located between positions −600 and −1300 relative to the IE94 mRNA start site. DNA sequence analysis of the region revealed a conserved consensus NF1 recognition element (T)TGG(C/A)N_5GCCAA embedded within each of 23 highly diverged 30-base-pair tandem repeats, together with a second downstream cluster of five consensus NF1-binding sites between positions +470 and +570 in the large first intron. Two separate NF1-binding loci were also found in the equivalent IE68 gene of HCMV(Towne) DNA, but in this case the DNA sequence and competition filter binding experiments indicated a maximum of only four to five consensus binding sites encompassing the promoter-enhancer region. In transient expression assays, neither the isolated upstream IE94 tandem repeats nor a synthetic single-copy consensus NF1-binding site acted as transcriptional *cis* activators or enhancers when placed adjacent to the simian virus 40 minimal early region promoter. We conclude that the large and complex 5′ upstream promoter-regulatory region for the SCMV IE94 gene comprises two distinct domains. The previously described four sets of 13- to 18-base-pair interspersed repeat elements between −55 and −580 provide most of the high basal transcriptional strength, whereas the arrangement of further upstream tandemly repeated NF1-binding sites may contribute significantly to the expanded biological host range for expression of SCMV IE94 compared with HCMV IE68.

The major immediate-early (IE) genes of both human cytomegalovirus (HCMV) and simian cytomegalovirus (SCMV) differ from all other genes of these viruses by being transcribed abundantly in infected permissive human diploid fibroblast cells even in the presence of cycloheximide to block de novo protein synthesis (18, 36). Furthermore, the IE gene products are the only viral proteins expressed after infection of nonpermissive rodent fibroblasts (16, 21), both promoters give strong basal levels of transcription in in vitro assays (38; K.-T. Jeang, Ph.D. thesis, The Johns Hopkins University, Baltimore, Md., 1984) and the intact phosphorylated IE94 protein product from SCMV(Colburn) is expressed abundantly in stable DNA-transfected mouse Ltk^+ cell lines receiving plasmid DNA containing an isolated 10.8-kilobase (kb) viral DNA fragment (17). These unusual properties, together with our expectations that these IE gene products play a key role in CMV gene regulation, led us to examine the overall structure and DNA sequence of the SCMV IE94 gene together with its associated upstream promoter-regulatory region.

IE transcription in cells infected with the African green monkey-derived SCMV(Colburn) (17; Jeang, Ph.D. thesis) proceeds leftward from map units 0.73 to 0.71 on the 220-kb linear double-stranded DNA genome (20) and gives rise to a major poly(A)$^+$ cytoplasmic mRNA species of 2.3 kb. DNA sequencing and S1 RNA analysis revealed a complex intron-exon structure consisting of three small 5′ exons and one large 3′ exon (Jeang, Ph.D. thesis; K.-T. Jeang, T. Lietman, Y.-N. Chang, and G. S. Hayward, manuscript in preparation). The map location of the HCMV IE68 gene (36) and overall intron-exon structure of the major 1.9-kb IE mRNA species of HCMV (1, 34) are remarkably similar to those of SCMV(Colburn). However, the major SCMV IE94K and HCMV IE68K proteins encoded by these mRNAs differ considerably in apparent molecular weight and possess only limited overall amino acid homology, although both have characteristic large glutamic acid-rich domains in their COOH-terminal region. In contrast to the relatively weak homology in the coding regions, the 5′ nontranslated leader sequences of the two IE mRNAs are very highly conserved. Both 5′ upstream promoter-regulatory regions between positions −55 and −550 consist of complex arrangements of four separate sets of multicopy 13- to 18-base-pair (bp) interspersed repetitive elements whose sequences are highly conserved between the two viruses, although the relative organization of the different repeat units within each region has been altered markedly (38; Jeang, Ph.D. thesis). The 5′ promoter-regulatory region from IE94 drives expression of beta interferon from a hybrid IE94-interferon construct in microinjected *Xenopus* oocytes or of chloramphenicol ace-

* Corresponding author.
† Present address: Laboratory of Molecular Virology, National Cancer Institute, Bethesda, MD 20892.
‡ Present address: Department of Microbiology and Immunology, School of Medicine, Emory University, Atlanta, GA 30322.

tyltransferase (CAT) from hybrid IE94-CAT constructs in DNA-transfected Vero cells to titers and activities that are 2 to 3 orders of magnitude higher than those produced by herpes simplex virus thymidine kinase-interferon or thymidine kinase-CAT hybrid genes (K.-T. Jeang, J. D. Mosca, M. Pizzorno, J. Shero, D. Gay, and G. S. Hayward, submitted for publication). Boshart et al. (5) have described the equivalent complex upstream region from HCMV(AD169) as the strongest enhancer yet encountered in eucaryotic cells when selected in an "enhancer trap" assay as a substitute for the two 72-bp repeats of simian virus 40 (SV40).

The direct requirement for several cellular proteins for initiation of adenovirus DNA synthesis in vitro has been described by several groups (6, 11, 25). One of these proteins, referred to as nuclear factor 1 (or NF1), has been shown to protect a 26-bp conserved binding site adjacent to an essential A+T-rich sequence within the DNA replication origin of many adenoviruses (10, 26, 30). The purified NF1 proteins appear to act in a tightly position-specific and synergistic manner in concert with other cellular proteins and the adenovirus terminal-binding protein. In the present report, we describe the existence of a new set of tandemly repeated elements further upstream from the SCMV IE94 promoter and present evidence for specific binding of the cellular NF1 protein to these repeats. Additional NF1-binding sites surrounding both the IE94 and IE68 promoters are described, and some assessment of the possible contributions of these binding sites to the basal strength of the CMV IE promoters is presented.

MATERIALS AND METHODS

Viral and plasmid DNAs. The SCMV(Colburn) isolate, although reportedly obtained originally from a human patient, is clearly of African green monkey origin by DNA structural and sequence homology studies (12, 15, 20). HCMV(Towne) and SCMV(Colburn) were grown in human diploid fibroblast cells from stocks passaged at a low multiplicity of infection. Cytoplasmic virions were recovered from clarified Dounce homogenates by centrifugation in sucrose density gradients. Purified virion DNA was prepared after DNase treatment of the virions by lysis with 1% sodium dodecyl sulfate–0.1 M EDTA–0.1 M Tris hydrochloride (pH 8.4), phenol extraction, and equilibrium banding in CsCl density gradients followed by ethanol precipitation and extensive dialysis.

The plasmid pTJ148 was described previously by Jeang et al. (17) and contains the 10.8-kb *Hin*dIII H fragment from SCMV(Colburn) including the complete IE94 gene. Two *Bam*HI-*Hin*dIII subclones of pTJ148, which include parts of the first two exons of the coding region and approximately 3,500 bp of 5′ sequences upstream from the transcriptional start site of the IE94 gene, were inserted into pBR322 (pTJ207) or the 1.9-kb poison-minus vector pKP45 (pKCMV), respectively. A number of 5′ and 3′ deletion derivatives of pTJ148 or pTJ207 that were used for mapping and DNA sequence analysis are described in Fig. 4b. The 19.5-kb *Xba*I E fragment of HCMV(Towne) in pMSDT-E (obtained from M. Stinski, University of Iowa) contains the complete IE68 gene with 4.6 kb of upstream 5′-flanking sequences from the viral genome. pSV2-CAT, pSV0-CAT, and pA10-CAT were obtained from G. Khoury (National Cancer Institute) and are described by Laimins et al. (23). Plasmids pTJ278, pTJ279, and pTJ280 are hybrid CAT gene constructs containing 990, 340, and 260 bp, respectively, from immediately 5′ to the IE94 transcriptional start site and including the *Sac*I site at +30 inserted into pCATB′ at the *Bam*HI site (see Fig. 4c). The 3′ deletion constructs pTJ273 and pTJ275 contain upstream sequences from −940 or −216 to −3500 from the IE94 gene linked 5′ to the minimal SV40 early-region promoter at the *Bam*HI site in pA10-CAT. To test for enhancer function we used the pV21-CAT plasmid (constructed by Keith Peden and Jordan Kreidberg, The Johns Hopkins School of Medicine) in which one complete copy of the SV40 72-bp enhancer was removed from pSV2-CAT (in a pKP54 background) by cleavage with *Sph*I followed by rejoining with a *Sac*I linker. A monomer 26-bp synthetic oligonucleotide containing the consensus NF1-binding site was then inserted in either orientation at the *Sac*I linker site in pV21-CAT to give pV21(BS)-CAT and pV21(SB)-CAT. Construction of the plasmids pKB67 and pKB67-88, which contain either 1 or 88 copies, respectively, of the 67 terminal base pairs of the adenovirus type 2 (Ad2) genome inserted into pKP45, was described by Rosenfeld and Kelly (32).

NF1 protein and nitrocellulose filter binding assays. The NF1 protein preparation used for these studies was partially purified up to the single-stranded DNA-cellulose fraction from uninfected HeLa cells as described by Rawlins et al. (30). After cleavage of the plasmid or viral DNAs with the appropriate restriction enzymes, the DNA fragments generated were 3′ labeled at their termini by filling in 5′ overhanging ends with deoxynucleoside triphosphates (one or two of which were ^{32}P labeled). Either *Micrococcus luteus* DNA polymerase or *Escherichia coli* DNA polymerase I (Klenow fragment) was used in the labeling procedure. Protein-DNA binding was done at 4°C for 30 min in a reaction mixture (40 to 50 μl) containing 150 mM NaCl. The reaction mixtures also routinely contained 50 mM HEPES (*N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid) (pH 7.5), 5 mM $MgCl_2$, 1 mM dithiothreitol, 250 μg of bovine serum albumin per ml, 125 μg of tRNA per ml, 10 to 20 fmol of [^{32}P]DNA, and binding protein. The reaction mixture was filtered through nitrocellulose and washed five times with 6 to 8 volumes of wash buffer (150 or 200 mM NaCl, 5 mM $MgCl_2$, 1 mM dithiothreitol, 10% glycerol). DNA bound to the filters was eluted with 25 mM HEPES (pH 7.5)–10 mM NaCl–0.2% sodium dodecyl sulfate. After gel electrophoresis, ethanol fixation, and drying, the DNA was visualized by autoradiography. For quantitation of binding in the competition assay, protein-DNA complexes were determined directly by scintillation counting of the dried nitrocellulose filters after filtration and washing (32).

DNase footprinting, sequencing analysis, and deletions. Protein-mediated protection of regions of the test DNA fragments from cleavage with DNase I was determined as described previously (29, 30). DNA sequencing reactions were done by a modified chemical sequencing procedure (2, 24). For the 5′ deletion series from pTJ148, the plasmid was opened with *Nco*I, digested back with *Bal* 31 nuclease, and rejoined to the *Sal*I site of pBR322 after the addition of *Sal*I linkers. The 3′ deletion series involved *Bal* 31 nuclease action at the *Bam*HI site of pTJ207 followed by rejoining directly after addition of *Bam*HI linkers.

CAT assays. Cell harvesting and CAT assays were performed as described previously (9, 28). Under our routine conditions, using 0.2 μCi of [^{14}C]chloramphenicol, the assay was linear with respect to enzyme concentration until at least 10^5 cpm appeared as the chloramphenicol-3-acetate product and linear with respect to time for at least 60 min. For quantitative estimates of CAT activity, the appropriate

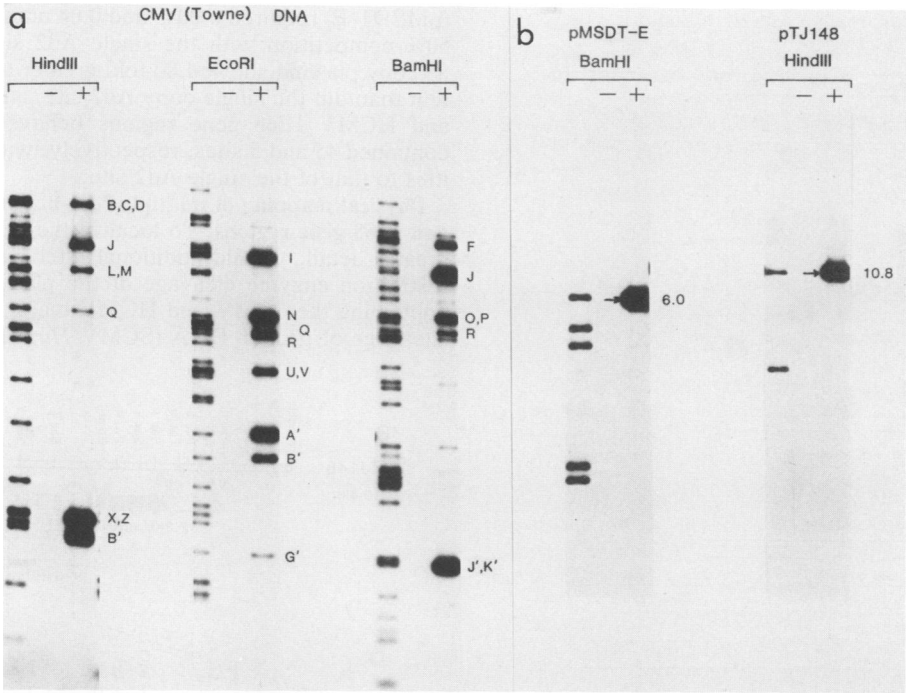


FIG. 1. Binding of HeLa cell NF1 to CMV DNA. Binding reactions containing NF1 and ^{32}P-labeled restriction fragments of cloned or virion DNA were filtered through nitrocellulose. The DNAs retained on the filters were eluted and electrophoresed on a 1.4% agarose gel. In each set lane 1 represents input DNA, lane 2 shows the DNA retained on the filter in the absence of NF1 (−), and lane 3 shows the DNA retained on the filter in the presence of NF1 (+). (a) Genomic HCMV(Towne) DNA cleaved with *Hin*dIII, *Eco*RI, and *Bam*HI. Confirmed or possible positive fragments are identified by the standard letter nomenclature (20, 39). (b) Recombinant cloned plasmid DNAs containing the major IE gene regions of HCMV and SCMV: HCMV(Towne) IE68 represented by pMSDT-E cleaved with *Bam*HI; SCMV(Colburn) IE94 represented by pTJ148 cleaved with *Hin*dIII. Numbers to right of lanes are in kilobases.

sections were cut from the thin-layer chromatography plate, and the amount of ^{14}C radioactivity was determined in a liquid scintillation counter.

RESULTS

Detection of NF1-binding sites in CMV DNA. During a filter binding assay survey of the frequency of occurrence of NF1-binding sites in mammalian DNA viruses, we found that several binding loci with various affinities occurred within the 240-kb genome of HCMV(Towne). One of these loci appeared to lie within the overlapping *Hin*dIII C (21.3 kb), *Bam*HI J (9.9 kb), and *Eco*RI I (10.9 kb) fragments encompassing the major IE gene of that virus (Fig. 1a). Further analysis with the HCMV(Towne) DNA *Xba*I E fragment cloned in a bacterial plasmid (pMSDT-E) confirmed that the IE68 gene region within the 6.0-kb *Bam*HI subfragment of pMSDT-E represented a strong NF1-binding locus (Fig. 1b). A similar positive result was also obtained with the cloned 10.8-kb *Hin*dIII H fragment (pTJ148) which contains the intact major IE94 gene region from SCMV(Colburn). We do not yet know whether the map locations of the other NF1-binding sites in HCMV are associated with features of any particular interest, but the sites within the major IE gene regions were expected to have potential functional significance.

Presence of very high affinity or multicopy binding sites in the SCMV IE region. To examine the relative affinity of the CMV IE region binding sites for the protein and also to confirm that NF1 binding within the CMV IE gene regions represented the same kind of sequence-specific interaction found in the adenovirus binding site, we carried out a quantitative competition filter binding assay (Fig. 2). This type of experiment measures the relative abilities of test DNA samples to compete with single and multicopy forms of the Ad2 consensus NF1 site for binding to a constant amount of the purified NF1 protein (32). The control Ad2 DNA

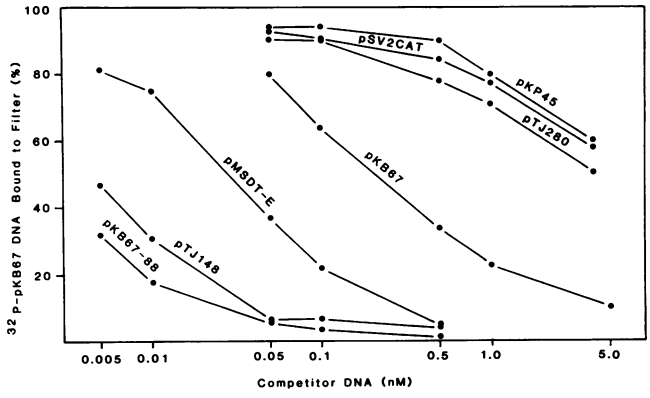


FIG. 2. Competition filter binding assay for comparing relative affinities of different NF1-binding loci. The test plasmid DNA samples used included those containing a single copy (pKB67) and 88 tandem copies (pKB67-88) of the 67-bp Ad2 terminal fragment, the major IE68 gene of HCMV (pMSDT-E), the major IE94 gene of SCMV (pTJ148), and control samples representing vector plasmid DNA (pKP45), an SV40-CAT plasmid construct (pSV2-CAT), and the promoter-regulatory region of IE94 (5′ to −260) fused to CAT DNA (pTJ280). The assay measures the amount of ^{32}P-labeled pKB67 DNA bound in the presence of a constant amount of NF1 protein and increasing amounts of unlabeled competitor DNA.

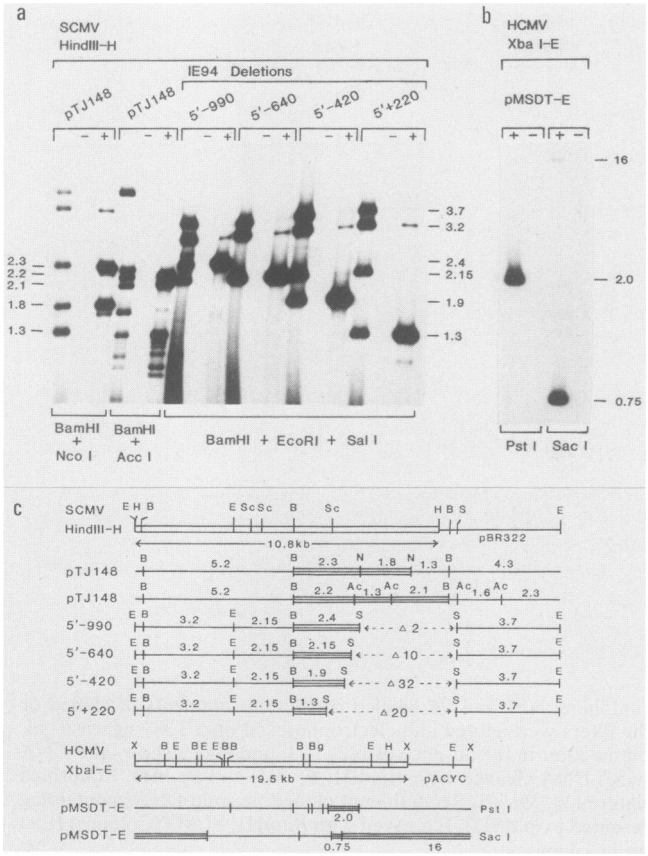


FIG. 3. More detailed mapping by filter binding assays of the location of NF1-binding sites in the IE94 and IE68 regions. (a) SCMV *Hin*dIII-H. Three-lane sets are as described in the legend to Fig. 1. From left to right: pTJ148 cleaved with *Bam*HI plus *Nco*I; pTJ148 cleaved with *Bam*HI plus *Acc*I; IE94 deletion plasmids, IE94 5′ to −990 (pTJ148 Δ2), IE94 5′ to −640 (pTJ148 Δ10), IE94 5′ to −420 (pTJ148 Δ32), and IE94 5′ to −220 (pTJ148 Δ20), all cleaved with a mixture of *Bam*HI, *Eco*RI, and *Sal*I. Several minor NF1-binding subfragments derived from the 1.3-kb species of pTJ148 DNA appear in the *Bam*HI-plus-*Acc*I-digested sample, but their exact origin is unknown. (b) HCMV *Xba*I-E. Two-lane sets (input samples omitted). pMSDT-E, *Pst*I digestion; pMSDT, *Sac*I digestion. (c) Relative map locations of the positive fragments (solid bars) observed in each sample set in panels a and b above. Nucleotide positions given represent the locations of the *Sal*I linkers marking the endpoints of the pTJ148 5′ deletions. Fragment sizes are given in kilobase pairs. Ac, *Acc*I; B, *Bam*HI; E, *Eco*RI; H, *Hin*dIII; N, *Nco*I; Sc, *Sac*I; S, *Sal*I.

plasmids used in these studies were a single copy of the 67-bp terminus of Ad2 DNA (pKB67) and an engineered 88 tandem repeat copy of the 65-bp sequence (pKB67-88). The amount of NF1 protein used in the assay gave 50% filter binding of ^{32}P-labeled pKB67 DNA in the presence of 0.2 nmol of competitor unlabeled pKB67 DNA. Control vector plasmids pKP45 and pSV2-CAT lacking binding sites required at least 7 nmol of DNA to reduce binding to pKB67 to 50%, whereas the 88-copy plasmid required only 0.0025 nmol to reduce binding to 50% (extrapolated from the data shown in Fig. 2). The use of a hybrid IE94-CAT plasmid (pTJ280) containing only 290 bp of the IE94 gene proximal promoter region (positions −260 to +30) gave no significant binding. In comparison, the plasmids containing the complete IE94 gene (pTJ148) and the complete IE68 gene (pMSDT-E) required 0.0045 and 0.04 nmol, respectively, for 50% competition with the single Ad2 site. Therefore, the 88-copy plasmid showed 80-fold-greater affinity for the protein than did the single-copy Ad2 site, and the SCMV IE94 and HCMV IE68 gene regions behaved as though they contained 45 and 5 sites, respectively, with equivalent affinities to that of the single Ad2 site.

Physical mapping of multiple NF1-binding loci within IE94 and IE68 gene regions. To localize the NF1-binding sites in greater detail, we did additional filter binding assays after restriction enzyme cleavage of the plasmid DNA samples containing the SCMV and HCMV major IE genes (Fig. 3). Cleavage of pTJ148 DNA (SCMV *Hin*dIII-H) within the 5′

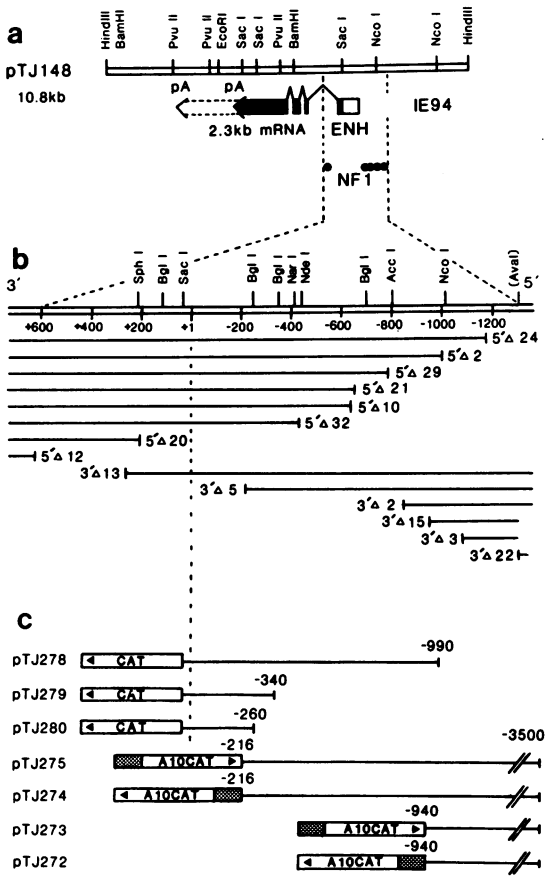


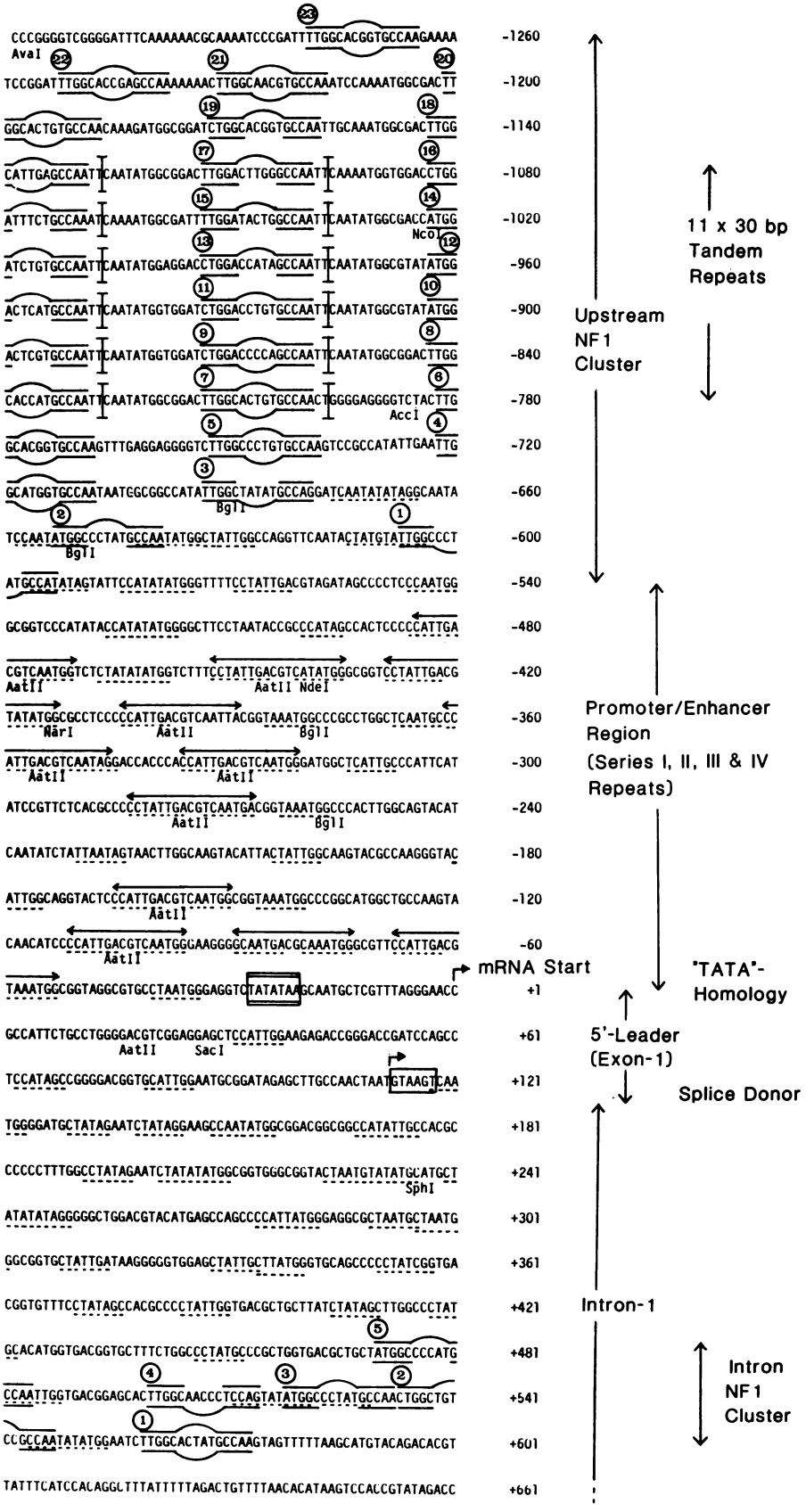
FIG. 4. Physical map of SCMV IE94 promoter-regulatory region. (a) The upper portion of the diagram shows the location and structure of the SCMV(Colburn) IE94 gene and 2.3-kb spliced mRNA (solid bar) relative to the enhancerlike region (open bar), NF1-binding loci (solid circles), and key restriction enzyme cleavage sites within the *Hin*dIII H DNA fragment (pTJ148). (b) An expanded view of the 2,000-bp region containing NF1-binding sites and encompassing the promoter-regulatory region. The boundaries of eight 5′ deletion clones of pTJ148 (*Sal*I linkers) and six 3′ deletion clones of pTJ207 (*Bam*HI linkers) used in the DNA sequencing analysis are shown. Distances are given in base pairs upstream (5′) or downstream (3′) from the major leftward mRNA start site at position +1 (Jeang, Ph.D. thesis; Jeang et al., submitted). (c) Structure of the hybrid CAT constructs containing portions of the IE94 upstream regions (see Fig. 8A). Open bars represent the bacterial CAT gene coding region in the standard cassette from pCATB′ with SV40 early antigen-derived splicing and poly(A) signals (28). Hatched bars denote the minimal SV40 early-region promoter from pA10-CAT including the 21-bp repeats and multiple Sp1-binding sites.

upstream regions of IE94 showed that some binding sites mapped upstream from the *Nco*I site at −1020 and on both sides of the *Acc*I site at −2100, whereas other sites mapped downstream from −1020 and on both sides of the *Acc*I site at −790. Furthermore, the use of 5′ deletion constructs ending with *Sal*I linkers inserted at positions −990, −640, −420, and +220 relative to the mRNA start site in the IE94 gene all gave positive *Sal*I-to-*Bam*HI fragments, revealing an additional locus mapping between position +220 and the *Bam*HI site at +1440. Within the 19.5-kb *Xba*I E fragment of HCMV(Towne) in plasmid pMSDT-E all the binding proved to be localized within a single 2.0-kb *Pst*I fragment mapping between positions +950 and −1000 relative to the IE68 mRNA start site, although this could be subdivided into two regions lying in adjacent *Sac*I fragments of 16 and 0.75 kb mapping to the right and left, respectively, of position −20. Note that the upstream site in IE68 gave a very weak signal relative to the downstream site in this experiment, but we have consistently observed that sites within DNA fragments as large as 15 kb give signals only 1/5th to 1/10th of the intensity as those obtained from the same sites when located within DNA fragments below 1 to 2 kb in size.

Far upstream repeated motifs in IE94 promoter-regulatory region. The relative location and orientation of the IE94 gene coding exons and promoter-regulatory regions within the SCMV(Colburn) *Hin*dIII H fragment are shown in Fig. 4a. The proximal upstream promoter-regulatory region of the primate CMV IE genes consists almost exclusively of four sets of multiple repetitive elements that are highly conserved in sequence but vary considerably in relative location and organization between IE68 and IE94 (5, 38; Jeang, Ph.D. thesis; Jeang et al., submitted). At least 3 and up to 11 copies of each repeat motif occur within a 460-bp region (positions −55 to −510) in HCMV IE68 and within an equivalent 530-bp region in SCMV IE94. Surprisingly, further sequence analysis of the more distal IE94 upstream region revealed additional repetitive structure. A restriction site map of the entire 5′ upstream region and proximal NH_2-terminal intron region of IE94 is presented in Fig. 4b together with the locations of a series of 5′ and 3′ deletion clones prepared from pTJ148. The sequencing analysis was done both by Maxam and Gilbert (24) chemical analysis from the few natural cleavage sites (*Sac*I, *Bgl*I, *Ava*I) or from the linker sites (*Sal*I or *Bam*HI) in the deletion plasmids and by dideoxynucleotide analysis of M13 derivatives of these deletion clones. The deduced nucleotide sequence of the entire 1,980-bp region from −1280 to +600 relative to the IE94 mRNA start site is presented in Fig. 5. The interspersed repeat elements (series I, II, III, and IV) all lie between positions −50 and −580 (Jeang, Ph.D. thesis; Jeang et al., submitted). Between positions −820 and −1150 we recognized a central 11-copy, 30-bp tandemly repeated sequence which is bounded on both sides by additional copies of closely related sequence elements. Furthermore, although the nucleotide sequence of individual repeat units is relatively poorly conserved, a total of 23 copies of a core consensus element with the partially palindromic sequence $TTGGN_6GCCAA$ occurs at approximately 30-bp intervals throughout this region. An overlapping repeated element with the sequence CCAATTCAATATGG is also highly conserved among the 11-copy, 30-bp tandem repeat portion of the complex. Perhaps significantly, numerous additional copies of related elements of the type CAATATC CAATATGG, CATATATGG, CCTATTGA, CCATTGA, CCAATGG, TCATTGG, CTATTGC, TATATGG, TAAATGG, etc., occur throughout the entire 1,300-bp IE94 5′ upstream region and also in the proximal 500-bp portion of the first intron. These include the series I repeats within the enhancer region which represent 18-bp palindromes with the consensus sequence CCATTGACGTCAATGG (Fig. 5).

Cluster of at least 20 adjacent NF1-binding sites within the IE94 upstream region. To define more precisely the actual number and pattern of NF1-binding sites in the upstream SCMV IE94 promoter-regulatory region, we did DNase I footprinting analysis of those DNA sequences in the pKCMV plasmid that could be protected by binding to the NF1 protein. DNA fragments that were 3′ end labeled either at the *Nco*I site at −1020 or at the *Nde*I site at −440 were isolated and subjected to DNase I analysis in the presence of increasing amounts of NF1 protein. The set of autoradiographs given in Fig. 6 shows the patterns obtained with long and short gel electrophoretic analysis proceeding in the downstream direction from *Nco*I to *Sac*I on the coding strand or upstream from *Nco*I to *Ava*I on the noncoding strand. The results revealed a ladder of protected regions interspersed with hypersensitive sites stretching on both sides of *Nco*I all the way from position −700 to −1280. In all, we detected 20 separate binding sites with various affinities for NF1 (Table 1), including three sites at the far upstream portion of the cluster (positions labeled 21, 22, 23) for which there were no interspersed hypersensitive sites. In the area of the 30-bp tandem repeats the binding protected approximately 24 bp of each repeat unit, and this spacing pattern was preserved relatively well even beyond the confines of the previously recognized 11-copy, 30-bp tandem repeats. In each case, inspection of the DNA sequence involved revealed that the protected regions were centered almost directly over the conserved semipalindromic consensus $PyTGGA/CN_5GCCAA$ that was recognized originally in the DNA sequencing studies (see Fig. 9a). The proximal three copies of the 23-member family recognized earlier appear not to bind (or to do so only weakly), which apparently correlates with a divergence from the strict consensus binding sequence such that an A substitutes for the first Py position or a G or T replaces the final A (Table 1). The sites labeled 10, 12, and 14 also have substitutions of A for the initial Py in the consensus, and although all three showed clearly detectable protection, these sites bound to our preparation of NF1 protein relatively poorly. In additional footprinting studies no binding sites were observed over a 300-bp region downstream from the *Nde*I site at −440 (data not shown), although this portion of the enhancer region contains several inverted sequences that are closely related to the NF1 consensus (Table 1). The possibility of further 5′ upstream sites between approximately −1400 and −2100 and details of the filter binding locus detected on the far side of *Acc*I at −2100 have not been evaluated further.

Other NF1-binding sites surrounding the IE94 and IE68 promoter-regulatory regions. The filter binding studies presented in Fig. 3a also indicated that an additional NF1-binding locus must exist downstream of the IE94 promoter within the intron regions between the 5′ deletion boundary at +220 and the *Bam*HI site at +1440. In the DNA sequence of the first intron region of IE94 (Fig. 5), we again recognized a cluster of several highly diverged repeat elements spaced at irregular intervals over a 250-bp region from +320 to +570. One set of repeats contains at least seven copies of a sequence with recognizable similarity to the NF1 consensus spaced 16 to 30 bp apart. However, only those copies at +468 ($ATGGCN_5GCCAA$), +498 ($CTGGCN_5GCCAA$), +516 ($ATGGCN_5GCCAA$), +532 ($CTGGCN_5GCCAA$), and +557 ($TTGGCN_5GCCAA$) would be expected to retain


AvaI
NcoI
AccI
BglI
NdeI
NarI
SphI
mRNA Start
11 x 30 bp
Tandem
Repeats
Upstream
NF1
Cluster
Promoter/Enhancer
Region
(Series I, II, III & IV
Repeats)
'TATA'-
Homology
5'-Leader
(Exon-1)
Splice Donor
Intron-1
Intron
NF1
Cluster

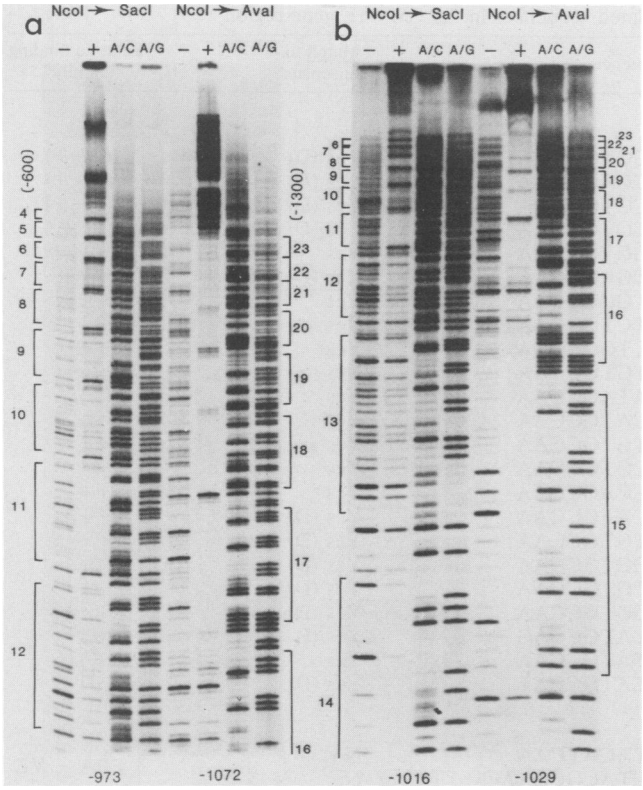


FIG. 6. DNase I footprint analysis of a tandemly repeated cluster of NF1-binding sites upstream from the IE94 gene. Isolated ^{32}P-labeled DNA fragments from pKCMV labeled at their 3′ termini were subjected to partial DNase I digestion in the presence (+) or absence (−) of NF1 protein. (a and b) Sets of autoradiographs from relatively long and short gel electrophoretic analyses, respectively. In each case, the left-hand panel (*Nco*I to *Sac*I) represents the coding strand proceeding from the *Nco*I site at −1020 in the middle of the cluster toward the promoter (Fig. 4 and 5). The right-hand panels (*Nco*I to *Ava*I) represent the noncoding strand in the direction away from the promoter. Two reference lanes of Maxam and Gilbert sequencing reactions on the same DNA samples are also shown (A/C; A/G). Individual NF1-binding sites are indicated and identified relative to the numerical order of tandemly repeated NF1 consensus elements in the DNA sequence (Fig. 5).

binding properties, and only that at +557 represents a strong consensus binding site (Table 1). Although they contain sequences that are closely related to the NF1 consensus (for example, TTGGCN$_6$GCCAA at −201), the enhancer region series II repeats of IE94 clearly did not bind to our preparation of NF1 protein according to either the competition filter binding or footprinting studies.

Consistent with our findings from the filter binding assays, a survey of existing DNA sequence data for HCMV IE68 (5, 38) revealed NF1 consensus sequences upstream of the enhancer complex in an equivalent position to the tandem-repeat cluster in IE94. Evidently, IE68 may contain only three copies of the far upstream binding sites spaced approximately 45 bp apart at −628, −684, and −729 and all in reverse orientation (Table 1). The first copy of this element (TTGGAN$_5$GCCAA) would be predicted to be a relatively strong binding site, but the other two (both ATGGCN$_5$GC CAA) probably represent relatively weak binding sites. There is no hint of any 30-bp tandemly repeated pattern in this region in HCMV DNA. Although direct footprinting studies have not been carried out, inspection of the DNA sequence data of Ackrigg et al. (1) within the first intron of the HCMV(AD169) IE68 gene revealed another consensus NF1-binding site at a location corresponding to the strong positive downstream signal detected in the filter binding experiments (Fig. 3b). A single perfect consensus sequence (TTGGCN$_5$GCCAA) occurs here at +350, although a second adjacent near-consensus version (ATGGCN$_5$GCCAC) at +324 would be predicted to lack binding activity.

The complex patterns of NF1-binding sites found both upstream and downstream of the human and simian versions of the major IE promoter are summarized in the diagram in Fig. 7. Known or predicted strong NF1-binding sites are indicated by solid squares, and known or predicted weak binding sites are denoted by solid circles. Note that the overall consensus NF1-binding site sequence observed in SCMV IE94 DNA matches precisely with that predicted previously from a comparison of the sites found in adenoviruses and immunoglobulin gene regions, etc. (7, 10, 14, 30).

Neither the SCMV upstream NF1 cluster nor a synthetic consensus NF1-binding site acts as a transcription enhancer. We have shown elsewhere that hybrid plasmids containing the intact IE94 promoter-regulatory region fused to either interferon or CAT coding regions give strong constitutive expression in microinjected oocytes or in DNA-transfected Vero cells (Jeang et al., submitted). To examine the effects of deleting the upstream NF1 site cluster, we compared two 5′ deletion plasmids (pTJ278 and pTJ279) containing either 990 or 340 bp of 5′ upstream SCMV DNA sequences fused to our standard CAT expression cassette at the *Sac*I site (position +30) in the IE94 sequence (Fig. 4c). Surprisingly, the CAT activity obtained with the 5′ deletion plasmid lacking the entire tandemly repeated NF1-binding site cluster (pTJ279) was not reduced significantly compared to that obtained with the plasmid retaining six of the upstream NF1 sites (pTJ278) (Fig. 8A). Similarly, we used two 3′ deletion derivatives of pTJ148 to ask whether the isolated upstream IE94 region containing the NF1 cluster could act as a *cis* activator or enhancer element when placed adjacent to the minimal SV40 promoter-CAT gene in pA10-CAT. The results showed that IE94 far upstream sequences derived from positions −3500 to −940 (pTJ273) gave almost no detectable enhancer activity when placed 3′ to A10-CAT and that sequences from positions −3500 to −216 (pTJ275) restored at most 10% of the level of activity obtained with pTJ278 (Fig. 8A). A similar pair of constructs with SCMV sequences

FIG. 5. Nucleotide sequence in the 5′ upstream promoter-regulatory region and proximal intron region of the IE94 gene. The sequence of the noncoding strand is given with key restriction enzyme cleavage sites identified. The 11-copy, 30-bp tandem repeats are indicated by brackets, and the 28 individual consensus NF1-binding site sequence elements are denoted by heavy underlining or overlining plus identification numbers for correlation with the data in Fig. 6 and Table 1. The TATAA homology and splice donor consensus sequences are denoted by boxing, and the positions of individual members of the series I palindromic dispersed repeats within the promoter-enhancer region are shown by overhead arrows. The broken underlines indicate additional elements throughout the region that are related to the conserved CCAATTCAATATGG sequences within the 30-bp tandem repeats. The numbering system used is based on the first nucleotide in the mRNA molecule representing position +1.

TABLE 1. Summary of NF1-binding sites and related sequences in the CMV IE gene region

Position	Site no.	Sequence	Match to consensus[a]	Relative binding affinity[b]
SCMV(Colburn)				
IE94 5′ tandem repeats				
−1279	23	TTGGCACGGTGCCAA	Yes (D)	(+++)
−1252	22	TTGGCACCGAGCCAA	Yes (D)	(+++)
−1231	21	TTGGCAACGTGCCAA	Yes (D)	(+++)
−1201	20	TTGGCACTGTGCCAA	Yes (D)	+++
−1172	19	CTGGCACGGTGCCAA	Yes	+++
−1143	18	TTGGCATTGAGCCAA	Yes (D)	+++
−1113	17	TTGGACTTGGGCCAA	Yes	+++
−1083	16	CTGGATTTCTGCCAA	Yes	+
−1053	15	TTGGATACTGGCCAA	Yes	++
−1023	14	ATGGATCTGTGCCAA	Yes	+
−993	13	CTGGACCATAGCCAA	Yes	(+)
−963	12	ATGGACTCATGCCAA	Yes	+
−933	11	CTGGACCTGTGCCAA	Yes	+++
−903	10	ATGGACTCGTGCCAA	Yes	+
−873	9	CTGGACCCCAGCCAA	Yes	++
−843	8	TTGGCACCATGCCAA	Yes (D)	+++
−813	7	TTGGCACTGTGCCAA	Yes (D)	++
−782	6	TTGGCACGGTGCCAA	Yes (D)	+++
−752	5	TTGGCCCTGTGCCAA	Yes (D)	+++
−722	4	TTGGCATGGTGCCAA	Yes (D)	+++
−693	3	TTGGCTATATGCCAG	Yes (I)	(−)
−653	2	ATGGCCCTATGCCAA	Yes	(−)
−607	1	TTGGCCCTATGCCAT	Yes (I)	(−)
IE94 5′ enhancer				
−251		TTGGCAGTACATCAA	No	FBA − VE
−201		TTGGCAAGTACGCCAA	No	FBA − VE
IE94 intron				
+407		TTGGCCCTATGCACA	No	
+438		CTGGCCCTATGCCCG	No	
+468	5	ATGGCCCCATGCCAA	Yes	
+498	4	TTGGCAACCCTCCAG	Yes (I)	FBA + VE
+516	3	ATGGCCCTATGCCAA	Yes	
+532	2	CTGGCTGTCCGCCAA	Yes	
+557	1	TTGGCACTATGCCAA	Yes (D)	
HCMV(Towne/AD169)				
IE68 5′ distal				
−729		TTGGCCATTAGCCAT	Yes (I)	
−684		TTGGCTATTGGCCAT	Yes (I)	FBA + VE
−628		TTGGCTCATGTCCAA	Yes (I)	
IE68 5′ enhancer				
−370		TTGGCAGTACATCAA	No	
−207		TTGGCAGTACATCAA	No	Unknown
−120		TTGGCACCAAAATCAA	No	
IE68 intron				
+318		ATGGCTCTTTGCCAC	No	
+344		TTGGCTATATGCCAA	Yes (D)	FBA + VE

[a] I, Match to consensus in inverted orientation; D, match to consensus in both orientations.
[b] −, +, ++, +++, (+++), Degree of protection in DNase I footprinting assay; FBA + VE or − VE, Positive or negative filter binding assay result with plasmid DNA containing the appropriate region of CMV DNA.

placed 5′ to the A10-CAT promoter (pTJ272 and pTJ274) gave even less activity (data not shown).

The overall interpretation of these results implies that sequences inside position −340 or even inside −216 contain the key elements in providing the strong, constitutive features of this promoter and that sequences beyond −340, including the 20-copy NF1-binding site domain, do not contribute significantly to the high transcriptional activity obtained in transient DNA transfection assays. However, the possibility that the 3′ deletion constructs included negative elements from beyond the NF1 cluster could affect these interpretations.

As a further test of the ability of NF1-binding sites to act as transcriptional enhancers, a single-copy, 26-bp synthetic oligonucleotide containing a consensus NF1-binding site was constructed and placed 5′ to the minimal SV40 promoter in pV21-CAT. The results of filter binding and DNA footprinting assays confirmed that this single consensus site retained strong binding affinity for purified NF1 protein. However, in transient CAT assays in CV-1 cells the presence of the

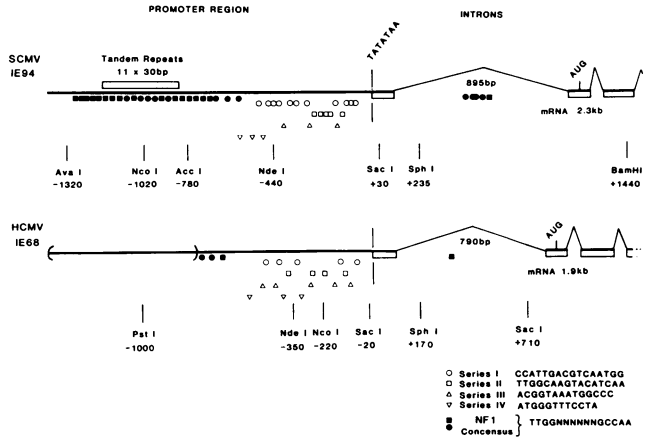


FIG. 7. Summary and comparison of the structure and arrangement of repeat elements and NF1-binding sites within the promoter-regulatory region and introns of IE94 and IE68. The 5′ upstream and NH_2-terminal portions of the two genes are drawn oriented with transcription proceeding 5′ to 3′ toward the right (although their orientation in the standardized physical map of the CMV viral genome is leftward). The open symbols (○, □, △, ▽) denote the four sets of interspersed repetitive elements within the enhancer region (−55 to −550). Solid symbols refer to known or predicted strong NF1-binding sites (■) and known or predicted weak NF1-binding sites or consensus sequences (●). Sequence data are not available for the region upstream of −760 in IE68.

upstream NF1-binding site failed to substitute for the SV40 two-copy, 72-bp enhancer, yielding no more CAT activity than pSV0-CAT or the parent pV21-CAT plasmid constructs (Fig. 8B). Although the results may differ in other cellular environments, these findings suggest that the NF1 binding on its own does not contribute directly to the stimulation of transcriptional activity. Instead, the cluster may function as an accessory region facilitating interactions with other transcription or replication factors and acting perhaps in a cell-type-specific manner.

DISCUSSION

We showed here that both the African green monkey CMV and HCMV genomes contain multiple binding sites for mammalian cellular NF1 protein within the vicinity of their major IE genes. In both the IE94 and IE68 promoter-regulatory regions authentic NF1-binding sites with standard consensus sequences occur at two distinct loci, one in a far upstream 5′ location beyond the enhancerlike sequences, and the other within the large first intron 3′ to the major IE mRNA start site of each gene. The summed relative NF1-binding affinity of DNA sequences within the SCMV(Colburn) IE94 and HCMV(Towne) IE68 genes was measured quantitatively to be 45-fold and 5-fold greater, respectively, than that for the single NF1 site near the left-hand end of Ad2 DNA. This extraordinarily high affinity was at least partly accounted for by the presence of clusters of tandemly repeated binding sites at both loci within the IE94 gene but not in the IE68 gene. After our studies were completed, Hennighausen and Fleckenstein (13) published the results of similar filter binding assays on the IE68 gene region of HCMV(AD169) DNA. Consistent with our results, they found that NF1 bound to at least one of the three upstream consensus elements equivalent to those predicted in Table 1 plus to a fourth site at −753. They also recognized a single strong binding site at +350 within the intron region and did not detect binding to the related sequences in the series II equivalent repeats within the HCMV(AD169) enhancer region.

In the 5′ upstream locus in IE94, at least 20 of the 23 adjacent copies of potential NF1 consensus DNA sequences, including all those in the 11-copy, 30-bp tandem repeats, interacted directly with partially purified NF1 protein by DNA footprinting analysis. The typical pattern of protection and hypersensitivity found in a prototype 30-bp repeat unit revealed a 24-bp protected region centered over the consensus sequence (Fig. 9A). However, even within the 30-bp tandem-repeat portion of the cluster there was considerable variability between individual sites in their affinity for the protein as judged by the DNase I protection studies. Sites with the sequence TTGG(A/C)N_5GCCAA showed the most consistently strong binding, followed by CTGG(A/C)N_5GCCAA and ATGG(A/C)N_5GCCAA. Furthermore, within the repeat cluster, both the internal nonrelevant six nucleotides and the flanking nucleotides outside the NF1 consensus signals showed considerable sequence divergence (summarized in Fig. 9B), implying that strong selective pressures for conservation of the NF1 consensus binding sites must have operated during the evolutionary drift of individual members of the cluster. This pattern of conservation also indicates that the NF1-binding sites must have some functional significance and implies that the tandemly repeated cluster was generated by a relatively ancient duplicative event rather than resulting from recent selection in

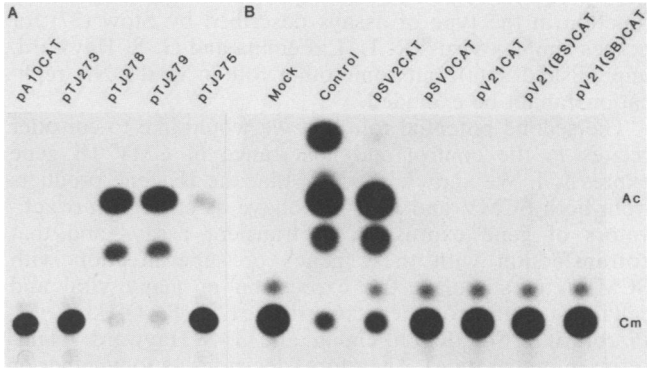


FIG. 8. Comparison of basal-level CAT activity produced by various IE94 constructs in transient DNA transfection assays. (a) Upstream NF1-binding sites do not contribute to the high transcriptional activity of the IE94 promoter. The diagram shows an autoradiograph of acetylated [^{14}C]chloramphenicol products obtained with extracts of Vero cells prepared 40 h after transfection with 2 μg of plasmid DNA. The structures of the IE94-CAT constructs used are shown in Fig. 4c. (b) Synthetic NF1-binding site does not exhibit transcriptional enhancer activity in transient assays. The autoradiograph shows [*acetyl*-^{14}C]chloramphenicol products produced with extracts of CV-1 cells that had been transfected with various SV40 promoter-derived CAT constructs. Control lanes show mock-transfected cell extracts with and without added CAT enzyme. pSV_2-CAT contains the complete SV40 early-region promoter-enhancer sequences. pA10-CAT and pV21-CAT are enhancer-minus constructs containing only the minimal SV40 early-region promoter with the 21-bp repeats and Sp1-binding sites. BS and SB refer to forward and backward orientations of a monomer synthetic NF1-binding site inserted at a *Sac*I linker site upstream from the Sp1-binding sites. Ac, [^{14}C]3-acetyl-chloramphenicol; Cm, [^{14}]chloramphenicol.

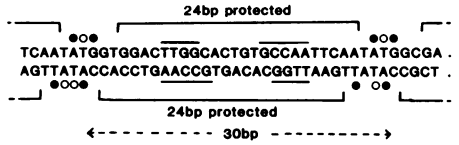
cell culture or from adaptation to human host cells. Note that all other Old World monkey SCMV isolates that we have tested also grow well in human diploid fibroblasts and therefore that this feature is not unique to SCMV(Colburn). The consensus IE94 30-bp tandem-repeat sequences have other interesting features (Fig. 9C), including a typical CAAT-box element (GCCAAT), which overlaps with the NF1 consensus and is found in the minimal promoter region of many eucaryotic promoters (3, 19), plus a short stretch of additional homology (CAATATG) with the Ad2 replication origin region to the right of the NF1 site.

What might be the functional significance of (i) the presence of two separate loci of NF1-binding sites surrounding the mRNA start sites and enhancer regions in the major IE genes of both types of primate CMV and (ii) the tandemly repeated nature of these sites in SCMV but not in HCMV? There are several obvious possibilities that are not necessarily mutually exclusive. First, the pattern in IE94 of two distinct binding loci within a relatively short distance from one another, including in one case a cluster of 30-bp tandemly repeated sites, closely resembles the pattern of Epstein-Barr virus EBNA-1-binding sites distributed within the *oriP* region of Epstein-Barr virus DNA that is involved in plasmid maintenance functions (29, 31, 40). Certainly, NF1 binding has a proven role in adenovirus DNA replication in vitro, and there is the precedent of replication origins being intricately associated with IE-type transcriptional control regions that occur in papovaviruses, adenoviruses, and herpes simplex virus. However, little is known as yet about the number or location of either lytic or possible latent plasmid DNA repllication origins in the CMV genomes. Our preliminary studies suggest that neither the IE68 nor the IE94 promoter region contains lytic replication origins that function in the type of assays described by Stow (37) for herpes simplex virus (R. L. LaFemina and G. S. Hayward, unpublished data), but some other role in viral DNA replication cannot be excluded.

The second potential role that we would like to consider relates to the control and host range of CMV IE gene expression. We show elsewhere that the IE gene products from both SCMV and HCMV behave as strong *trans* activators of gene expression in transient assays and that cotransfection with these genes or superinfection with SCMV can stimulate the expression of many viral and cellular target promoter constructs (28; P. O'Hare, M. Pizzorno, L. Sha, R. LaFemina, and G. S. Hayward, manuscript in preparation). Therefore, decisions as to whether or not the IE genes themselves are expressed would have a major impact on the nature and outcome of the virus-host interaction (e.g., permissive, nonpermissive, latency, etc.). The inclusion of a consensus CAAT-box element in many of the 30-bp tandem-repeat units suggests that there may be a transcriptional role for this feature (in IE94 in particular). Furthermore, NF1-binding loci have been observed near potential transcriptional control regions for several cellular genes (4, 14, 33). Both CMV IE promoters drive abundant constitutive expression of hybrid CAT genes in all cell types tested in both long-term and short-term DNA transfection assays (17; Jeang et al., submitted). Nevertheless, as we showed here, the upstream tandemly repeated sites apparently do not contribute significantly to the basal strength of the IE94 promoter in transient CAT expression assays, nor do they (or a synthetic consensus NF1 site) have the properties of transcriptional enhancers in such assays. Similarly, the 5′ upstream region containing the NF1-binding site(s) in HCMV(Towne) appeared not to contribute to the basal strength of the IE68 promoter in the studies reported by Stinski and Roehr (35).

On the other hand, the major IE promoter-regulatory regions of HCMV(Towne) and SCMV(Colburn) have rather different properties when expressed from the viral genomes in infected cells. Furthermore, despite the structural homologies between IE94 and IE68, HCMV and SCMV differ considerably in their ability to express these gene products after infection of several transformed or tumor-derived cell lines in culture (22). For example, the SCMV IE94K protein but not the HCMV IE68K protein is synthesized in human embryonic kidney fibroblast cell lines expressing either SV40 large-T antigen (NBE cells) or adenovirus E1A gene products (293 cells). Similarly, IE94K is expressed normally after infection of human teratocarcinoma stem cells, but the IE68K protein is not expressed in these cells unless they have been induced to differentiate with retinoic acid (8, 22, 27). Neither IE gene is expressed in infected mouse Ltk$^-$ cells or F9 teratocarcinoma cells, although treatment of F9 cell cultures with retinoic acid induces expression of IE94,

A. CMV(COLBURN) IE94
TANDEM REPEAT NFI BINDING SITES



B. CONSERVATION OF NFI CONCENSUS IN IE94 REPEATS

	(T)	T	G	G	C/A	A/C	(C)	N	(G)	(T)	G	C	C	A	A
A	4	-	-	-	9	10	3	1	5	4	-	-	-	23	21
C	5	-	-	-	14	9	14	7	2	-	-	23	23	-	-
G	-	-	23	23	-	-	-	4	14	2	23	-	-	-	1
T	14	23	-	-	-	4	6	11	2	17	-	-	-	-	1

C. HOMOLOGY WITH AD2 ORIGIN SEQUENCES

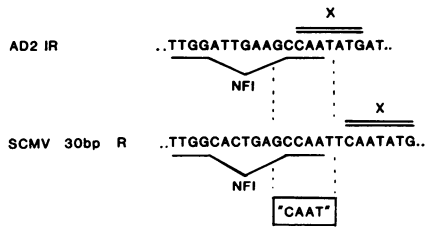


FIG. 9. Consensus and conservation of the SCMV IE94 5′ upstream NF1-binding sites. (A) Diagrammatic summary of a typical example of one of the 30-bp tandemly repeated NF1-binding sites showing the location and extent of DNase I protection on both DNA strands and the usual pattern of hypersensitive sites between adjacent protected regions. Solid and open circles denote relatively strong or weak hypersensitive sites, respectively. (B) Table illustrating the extraordinary conservation of the consensus NF1-binding sequence (T)TGGA/CN$_5$GCCAA within the otherwise highly diverged DNA sequences of the 23 copies of the upstream cluster. (C) Possible additional homologous elements (CAAT box and element X) between the IE94 30-bp tandem repeats and the Ad2 origin sequences surrounding the NF1-binding site.

but not IE68. Similarly, HCMV IE68, but not SCMV IE94, is repressed after infection of Vero cells. Synthesis of ^{35}S-labeled IE94K protein, but not IE68K, can also be demonstrated after infection of human peripheral blood leukocytes (R. L. LaFemina and G. S. Hayward, submitted for publication). We have shown elsewhere that input parental HCMV DNA enters the nuclei of nonpermissive transformed cells and that the block to IE expression occurs at the level of stable accumulation of mRNA (22). Therefore, the general impression is that both promoters possess unusually strong basal expression properties but that the presence or absence of specific cellular (and perhaps viral) transcription factors as well as the particular arrangement of *cis*-acting elements may have a major impact on their activity in infected cells. In nonpermissive transformed cells, expression from the IE68 promoter appears to be either much more sensitive than that from the IE94 promoter to the absence of positive factors or much more sensitive to the presence of negative factors. We postulate that the presence of the tandemly repeated clusters of NF1-binding sites surrounding the IE94 promoter-enhancer region may act indirectly by providing *cis*-acting elements that help to overcome the effects of repressors and allow transcription from the IE94 gene to be far more constitutive in stem cells and other nonpermissive transformed cells than is that from the IE68 gene.

ACKNOWLEDGMENTS

This work was funded by Public Health Service grants CA28473 and CA22130 awarded to G.S.H. and CA16579 awarded to T.J.K. from the National Institutes of Health. D.R.R. was supported by fellowship DRG-607 from the Damon Runyon-Walter Winchell Cancer Fund. K.-T.J. and P.J.R. were supported by the Medical Scientist Training Program at Johns Hopkins Medical School (grant GM07309 from the National Institutes of Health). G.S.H. was supported by a Faculty Research Award from the American Cancer Society (FRA 249).

We thank Robert LaFemina for mapping data for the HCMV IE68 region, Mark Stinski and Keith Peden for recombinant DNA clones, Mabel Chiu for technical assistance, and Judy DiStefano, Phyllis Broughton, and Pamela Wright for assistance in preparation of the manuscript.

LITERATURE CITED

1. **Ackrigg, A., G. W. G. Wilkinson, and J. D. Oram.** 1985. The structure of the major immediate early gene of human cytomegalovirus strain AD169. Virus Res. **2**:107–121.
2. **Bencini, D. A., G. A. O'Donovan, and J. R. Wild.** 1984. Rapid chemical degradation sequencing. Biotechniques **2**:4–5.
3. **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.
4. **Borgmeyer, U., J. Nowock, and A. E. Sippel.** 1984. The TG GCCA-binding protein: a eucaryotic nuclear protein recognizing a symmetrical sequence on double-stranded linear DNA. Nucleic Acids Res. **12**:4295–4311.
5. **Boshart, M., F. Weber, G. Jahn, K. Dorsch-Hasler, B. Fleckenstein, and W. Schaffner.** 1985. A very strong enhancer is located upstream of an immediate-early gene of human cytomegalovirus. Cell **41**:521–562.
6. **Challberg, M. D., and T. J. Kelly, Jr.** 1979. Adenovirus DNA replication in vitro. Proc. Natl. Acad. Sci. USA **76**:655–659.
7. **Challberg, M. D., and D. R. Rawlins.** 1984. Template requirements for the initiation of adenovirus DNA replication. Proc. Natl. Acad. Sci. USA **81**:100–104.
8. **Gonczol, E., P. W. Andrews, and S. A. Plotkin.** 1984. Cytomegalovirus replicates in differentiated but not in undifferentiated human embryonal carcinoma cells. Science **224**: 159–161.
9. **Gorman, C. M., L. F. Moffat, and B. H. Howard.** 1982. Recombinant genomes which express chloramphenical acetyltransferase in mammalian cells. Mol. Cell. Biol. **2**:1044–1051.
10. **Gronostajski, R. M., S. Adhya, K. Nagata, R. A. Guggenheimer, and J. Hurwitz.** 1985. Site-specific DNA binding of nuclear factor I: analyses of cellular binding sites. Mol. Cell. Biol. **5**:964–971.
11. **Guggenheimer, R. A., B. W. Stillman, K. Nagata, F. Tamanoi, and J. Hurwitz.** 1984. DNA sequences required for the *in vitro* replication of adenovirus DNA. Proc. Natl. Acad. Sci. USA **81**:3069–3073.
12. **Hayward, G. S., R. Ambinder, D. Ciufo, S. D. Hayward, and R. L. LaFemina.** 1984. Structural organization of human herpesvirus DNA molecules. J. Invest. Dermatol. **83**:29s–41s.
13. **Hennighausen, L., and B. Fleckenstein.** 1986. Nuclear factor I interacts with five DNA elements in the promoter region of the human cytomegalovirus major immediate-early gene. EMBO J. **5**:1367–1371.
14. **Hennighausen, L., U. Siebenlist, D. Danner, P. Leder, D. Rawlins, P. Rosenfeld, and T. J. Kelly, Jr.** 1985. High-affinity binding site for a specific nuclear protein in the human IgM gene. Nature (London) **314**:289–292.
15. **Huang, E.-S., and J. S. Pagano.** 1974. Human cytomegalovirus. II. Lack of relatedness to DNA of herpes simplex I and II, Epstein-Barr virus, and nonhuman strains of cytomegalovirus. J. Virol. **13**:642–645.
16. **Jeang, K.-T., G. Chin, and G. S. Hayward.** 1982. Characterization of cytomegalovirus immediate-early genes. I. Nonpermissive rodent cells overproduce the IE94K protein from CMV(Colburn). Virology **121**:393–403.
17. **Jeang, K.-T., M.-S. Cho, and G. S. Hayward.** 1984. Abundant constitutive expression of the immediate-early 94K protein from cytomegalovirus (Colburn) in a DNA-transfected mouse cell line. Mol. Cell. Biol. **4**:2214–2223.
18. **Jeang, K.-T., and W. Gibson.** 1980. A cycloheximide-enhanced protein in cytomegalovirus-infected cells. Virology **107**: 362–374.
19. **Jones, K. A., K. R. Yamamoto, and R. Tjian.** 1985. Two distinct transcription factors bind to the HSV thymidine kinase promoter *in vitro*. Cell **42**:559–572.
20. **LaFemina, R. L., and G. S. Hayward.** 1980. Structural organization of the DNA molecules from human cytomegalovirus. ICN-UCLA Symp. Mol. Biol. **18**:39–55.
21. **LaFemina, R. L., and G. S. Hayward.** 1983. Replicative forms of human cytomegalovirus DNA with joined termini are found in permissively infected human cells but not in non-permissive Balb/c-3T3 mouse cells. J. Gen. Virol. **64**:373–389.
22. **LaFemina, R. L., and G. S. Hayward.** 1986. Constitutive and retinoic acid-inducible expression of cytomegalovirus immediate-early genes in human teratocarcinoma cells. J. Virol. **58**:434–440.
23. **Laimins, L. A., G. Khoury, C. Gorman, B. Howard, and P. Gruss.** 1982. Host-specific activation of transcription by tandem repeats from simian virus 40 and Moloney murine sarcoma virus. Proc. Natl. Acad. Sci. USA **79**:6453–6457.
24. **Maxam, A. M., and W. Gilbert.** 1977. A new method for sequencing DNA. Proc. Natl. Acad. Sci. USA **74**:560–564.
25. **Nagata, K., R. A. Guggenheimer, T. Enomoto, J. H. Lichy, and J. Hurwitz.** 1982. Adenovirus DNA replication *in vitro*: identification of a host factor that stimulates synthesis of the preterminal protein-dCMP complex. Proc. Natl. Acad. Sci. USA **79**:6438–6442.
26. **Nagata, K., R. A. Guggenheimer, and J. Hurwitz.** 1983. Specific binding of a cellular DNA replication protein to the origin of replication of adenovirus DNA. Proc. Natl. Acad. Sci. USA **80**:6177–6181.
27. **Nelson, J. A., and M. Groudine.** 1986. Transcriptional regulation of the human cytomegalovirus major immediate-early gene is associated with induction of DNase I-hypersensitive sites. Mol. Cell. Biol. **6**:452–461.
28. **O'Hare, P., and G. S. Hayward.** 1984. Expression of recombinant genes containing herpes simplex virus delayed-early and

immediate-early regulatory regions and *trans* activation by herpesvirus infection. J. Virol. **52**:522–531.
29. **Rawlins, D. R., G. Milman, S. D. Hayward, and G. S. Hayward.** 1985. Sequence specific DNA binding of the Epstein-Barr virus nuclear antigen (EBNA) to clustered sites in the plasmid maintenance region. Cell **42**:859–868.
30. **Rawlins, D. R., P. J. Rosenfeld, R. J. Wides, M. D. Challberg, and T. J. Kelly, Jr.** 1984. Structure and function of the adenovirus origin of replication. Cell **37**:309–319.
31. **Reisman, D., J. Yates, and B. Sugden.** 1985. A putative origin of replication of plasmids derived from Epstein-Barr virus is composed of two *cis*-acting components. Mol. Cell. Biol. **5**:1822–1832.
32. **Rosenfeld, P., and T. J. Kelly, Jr.** 1986. Purification of nuclear factor I by DNA recognition site affinity chromatography. J. Biol. Chem. **261**:1398–1408.
33. **Siebenlist, U., L. Hennighausen, J. Battey, and P. Leder.** 1984. Chromatin structure and protein binding in the putative regulatory region of the c-myc gene in Burkitt lymphoma. Cell **37**:381–391.
34. **Stenberg, R. M., D. R. Thomsen, and M. F. Stinski.** 1984. Structural analysis of the major immediate early gene of human cytomegalovirus. J. Virol. **49**:190–199.
35. **Stinski, M. F., and T. J. Roehr.** 1985. Activation of the major immediate early gene of human cytomegalovirus by *cis*-acting elements in the promoter-regulatory sequence and by virus-specific *trans*-acting components. J. Virol. **55**:431–441.
36. **Stinski, M. F., D. R. Thomsen, R. M. Stenberg, and L. C. Goldstein.** 1983. Organization and expression of the immediate early genes of human cytomegalovirus. J. Virol. **46**:1–14.
37. **Stow, N. D.** 1982. Localization of an origin of DNA replication within the TR_S/IR_S repeated region of the herpes simplex virus type 1 genome. EMBO J. **1**:863–867.
38. **Thomsen, D. R., R. M. Stenberg, W. F. Goins, and M. F. Stinski.** 1984. Promoter-regulatory region of the major immediate-early gene of human cytomegalovirus. Proc. Natl. Acad. Sci. USA **81**:659–663.
39. **Thomsen, D. R., and M. F. Stinski.** 1981. Cloning of the human cytomegalovirus genome as endonuclease *Xba*I fragments. Gene **16**:207–216.
40. **Yates, J., N. Warren, D. Reisman, and B. Sugden.** 1984. A *cis*-acting element from the Epstein-Barr viral genome that permits stable replication of recombinant plasmids in latently infected cells. Proc. Natl. Acad. Sci. USA **81**:3806–3810.