Codons 262 to 490 from the herpes simplex virus ICP4 gene are sufficient to encode a sequence-specific DNA binding protein

Chin-Lee Wu and Kent W.Wilcox*
Department of Microbiology, Medical College of Wisconsin, Milwaukee, WI 53226, USA

Received October 13, 1989; Revised and Accepted December 18, 1989

ABSTRACT

The HSV-1 immediate early (IE) protein ICP4 (α4, IE175, Vmw175) is an oligomeric molecule which activates transcription of viral early genes, represses transcription of viral IE genes, and binds to specific sequences in certain viral promoters. The extent to which these functions are interrelated has not been fully established. We have expressed truncated portions of the ICP4 gene in *E. coli* as *trpE* fusion proteins. DNA-binding studies with these hybrid proteins revealed that ICP4 residues 262 to 490 are sufficient for sequence-specific DNA-binding. DNA-binding was not detected with polypeptides extending from residue 262 to 464 or from residue 306 to 490. Multiple bands of protein-DNA complexes observed in gel mobility shift assays indicate that residues 262 to 490 may also contribute to the oligomerization of ICP4.

INTRODUCTION

ICP4 (α4, IE175, Vmw175) is a 1298 amino acid polypeptide encoded by herpes simplex virus (1). This protein, which is expressed during the immediate early phase of productive infection (2,3), functions as both a positive regulator of early and late viral gene expression and as a negative regulator of immediate early viral gene expression (4–10). Previous reports have established that ICP4 is a sequence-specific DNA binding protein (11–17). High-affinity binding sites for ICP4 have been identified by DNase footprinting in the promoters for the HSV-1 genes encoding the immediate early proteins ICP0 (α0, IE110, Vmw110) and ICP4 (14,17–19). There is strong evidence that interactions between ICP4 and the binding sites in the ICP0 and ICP4 gene promoters provide one mechanism for repressing expression of these immediate early genes (8,20–24).

The mechanism whereby ICP4 stimulates transcription of viral genes during the early and late phases of infection has not been established. The failure to detect high-affinity ICP4 binding sites in some genes which are positively regulated by ICP4 prompted the proposal that the DNA-binding ability of ICP4 is primarily relevant for repression and autoregulation of IE gene transcription (10). The observation that some early and late genes contain weak binding sites for ICP4 (15,25) prompted the proposal that efficient transcription of early and late viral genes requires ICP4 to stabilize complexes between promoters for these genes and one or more cellular factors (26,27). The presence of three well-defined ICP4 binding sites in the HSV-1 gene encoding glycoprotein D (gD) raises the possibility that some viral genes may respond to ICP4 by more than one mechanism (28–30).

To determine if the functional activities of ICP4 (i.e., DNA binding, transactivation, and repression) are interrelated several laboratories have undertaken mutational analyses of ICP4. The initial finding that an ICP4 gene with a missense mutation in ICP4 codon 475 encoded a temperature-sensitive protein that failed to activate or repress viral gene transcription at the nonpermissive temperature suggested that there was some overlap in the domains required for the regulatory activities of ICP4 (31). The isolation of a non-defective HSV-1 deletion mutant lacking ICP4 codons 209 to 237 implied that this region was nonessential (32). Subsequent studies demonstrated that ICP4 codons 1 to 774 encoded a polypeptide with the three known activities of ICP4 and revealed that codons 1 to 89 were not essential (26,33,34). Three recent reports (27,35,36) have provided compelling evidence that mutations which disrupt the DNA-binding activity of ICP4 also impair both the positive and negative transcriptional regulatory activities of ICP4. In particular, the results demonstrated that in-frame insertions of two to four codons in the ICP4 gene at sites between codons 263 and 489 yielded polypeptides that had reduced abilities to bind to DNA and to regulate viral gene transcription. Although these reports clearly revealed that residues 263 to 489 contribute to the DNA binding ability of ICP4, the data did not indicate whether these residues alone were necessary and sufficient for DNA binding. Of particular interest in this regard was the finding by Paterson and Everett (36) that in-frame insertions adjacent to ICP4 codons 591, 681, 803, 843, 934, or 1066 resulted in production of altered proteins with reduced DNA binding ability.

We have undertaken a molecular dissection of ICP4 in order to identify functional domains in the protein. Our approach consists of expressing coding portions of the ICP4 gene in bacterial expression systems, followed by partial purification and analysis of the activities of the truncated polypeptides. In this report we provide evidence that ICP4 residues 262 to 490 are sufficient to yield a polypeptide with the same DNA binding specificity as native ICP4.

* To whom correspondence should be addressed

MATERIALS AND METHODS

Construction of expression vectors

The pATH expression vectors were generous gifts of Dr. Carol Dieckmann (37) and contain a portion of the *E. coli* tryptophan operon extending from several hundred nucleotides upstream of the promoter to codon 323 in the *trpE* gene followed immediately by derivatives of the pUC12 multiple cloning region (MCR). We prepared pATH2Nae and pATH2Sac by inserting an oligonucleotide containing a NaeI or SacII site, respectively, into the MCR in pATH2. The pATER expression vectors were constructed by digesting pATH vectors with NruI and EcoRI to remove *trpE* codons 19 to 323; single-stranded termini were filled in with polymerase and then ligated to yield plasmids in which *trpE* codon 18 was immediately adjacent to the EcoRI site in the MCR. Transcription terminators from the *rrnB* gene were inserted downstream from the MCR by substituting the HindIII to ScaI region in these plasmids with a HindIII to ScaI fragment from pKK233-2 (38). The resulting vectors were designated pATER1 and pATER2. To construct pATER3, the termini of EcoRI-digested pATER2 were filled in with polymerase and ligated. pATER2Nae and pATER2Sac were prepared by inserting the appropriate oligonucleotides into the MCR in pATER2. Figure 1 presents a schematic diagram of the pATER vectors and the nucleotide sequence of the MCR of each of the pATER vectors.

Construction of recombinant trpE:ICP4 genes

For the sake of brevity, some intermediate steps required for construction of the recombinant genes are not described. The wild-type ICP4 gene was obtained from pGX58 which contains the XhoI C fragment from HSV-1 strain 17 and was kindly sent to us by Dr. Nigel Stow (39). Plasmid pXK519 contains the entire ICP4 coding region except codons 186 to 256, which were deleted by digestion with PpuMI. The PpuMI termini were filled in with polymerase and religated, yielding a SalI site and an in-frame serine codon between residues 185 and 257. The 5′ end of the ICP4 gene was cleaved with SfaNI and filled in with polymerase, yielding a blunt-ended fragment that terminated with the second nucleotide of the ICP4 translation initiation codon. The 3′ end of the ICP4 gene was cleaved with Sau3AI at a site 453 bp downstream from the ICP4 translation termination codon. The SfaNI to Sau3AI fragment containing the entire ICP4 coding region except codons 186 to 256 was ligated to two fragments derived from pATER2. The first vector fragment consisted of the ScaI to SalI portion of pATER2 that includes the *trpE* promoter. The SalI end was filled in with polymerase and the adenine at the terminus was ligated to the filled-in SfaNI terminus at the 5′ end of the ICP4 gene to restore the ATG codon in the ICP4 gene. The second vector fragment consisted of the ScaI to BamHI portion of pATER2 that includes the transcription terminators. The Sau3AI sticky end at the 3′ end of the ICP4 gene was ligated to the complementary BamHI site from the vector. Finally, the ScaI sites at the other ends of the two vector fragments were ligated to restore the amp^R gene. pXK519 encodes a 1254 residue fusion protein designated FP519 that contains the first 18 amino acids of *trpE*, 8 amino acids (IRGDPLES) from the MCR, followed by ICP4 residues 1 to 185, a serine residue, and then ICP4 residues 257 to 1298.

Plasmid pXK518 contains the ICP4 gene from codons 84 to 185 and codons 257 to 1298 followed by the same 3′ end as for pXK519. Codons 186 to 256 were deleted as described for pXK519. The ICP4 gene was cut between codons 83 and 84 with PvuII and linked via an oligonucleotide to the EcoRI site in pATER2. pXK518 encodes a 1171 residue fusion protein designated FP518 that contains the first 18 amino acids of *trpE*, 8 amino acids (IRARPGIP) from the MCR, followed by ICP4 residues 84 to 185, a serine residue, and then ICP4 residues 257 to 1298.

Plasmid pXK445 was constructed by isolating the EcoRI to HindIII fragment that contains the ICP4 gene from pXK519 and ligating it to the complementary EcoRI and HindIII sites in pATH1. pXK445 encodes a 1564 residue fusion protein designated FP445 that contains the first 323 amino acids from *trpE*, 13 amino acids (PGRARIRGDPLES) from the MCR, followed by ICP4 residues 1 to 185, a serine residue, and then ICP4 residues 257 to 1298.

Plasmids pXK525 and pXK449 were constructed by isolating the SacII fragment that contains codons 262 to 490 from the ICP4 gene and ligating it to the SacII sites in pATER2Sac and pATH2Sac respectively. Both plasmids contain a translation termination codon immediately adjacent to ICP4 codon 490. pXK525 encodes a 257 residue fusion protein designated FP525 that contains the first 18 amino acids of *trpE*, 9 amino acids (IRGDPLEST) from the MCR, followed by ICP4 residues 262 to 490. pXK449 encodes a 561 residue fusion protein designated FP449 that contains the first 323 amino acids from the *trpE* gene, 8 amino acids (PGCPLES) from the MCR, followed by ICP4 residues 262 to 490.

Plasmids pXK527 and pXK452 were constructed by isolating the NaeI to HindIII fragment that contains ICP4 codons 275 to 490 from pXK525 and inserting it between the NaeI and HindIII sites in pATER2Nae and pATH2Nae, respectively. pXK527 encodes a 244 residue fusion protein designated FP527 that contains the first 18 amino acids from the *trpE* gene, 9 amino acids (IRGDPLEST) from the MCR, followed by ICP4 residues 275 to 490. pXK452 encodes a 548 residue fusion protein

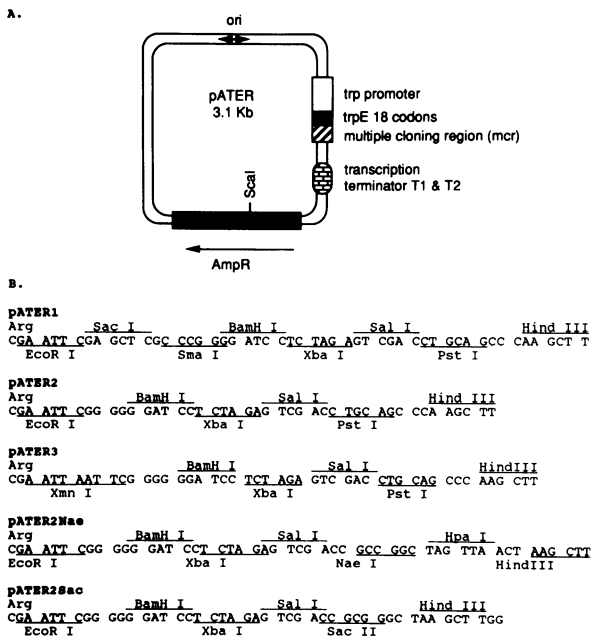

Figure 1. Structure of the pATER vectors. Panel A: Schematic diagram of the relevant components of all pATER vectors. Construction of these vectors is described in Materials and Methods. Panel B: Sequences of the multiple cloning regions in the pATER vectors. The first three nucleotides in each sequence correspond to the 18th codon in the *trpE* gene.

designated FP452 that contains the first 323 amino acids from *trpE*, 8 amino acids (PGDPLEST) from the MCR, followed by ICP4 residues 275 to 490.

Plasmid pXK529 was constructed by first isolating the SalI to HindIII fragment that contains ICP4 codons 262 to 490 from pXK525. The portion of the ICP4 gene in pXK519 from the NcoI site at ICP4 codon 90 to the 3′ end of the gene was deleted by digestion with NcoI and HindIII. The NcoI terminus was filled in with polymerase and ligated to the filled-in SalI terminus, which restored both the SalI site and ICP4 codon 90. Ligation of the complementary HindIII sites completed the construction. pXK529 encodes a 348 residue fusion protein designated FP529 that contains the first 18 amino acids of *trpE*, 8 amino acids (IRGDPLES) from the MCR, ICP4 residues 1 to 90, 2 amino acids (ST) encoded by the SalI site, and ICP4 residues 262 to 490.

Plasmid pXK530 was constructed by deleting the NruI fragment from pXK449 that spans codon 18 in the *trpE* gene to codon 307 in the ICP4 gene, inclusive. Ligation of the NruI termini maintained the correct reading frame and restored codons 306 and 307 of the ICP4 gene. pXK530 encodes a 202 residue fusion protein designated FP530 that contains the first 16 amino acids of *trpE*, followed by ICP4 residues 306 to 490.

Plasmid pXK442 was constructed by isolating the NruI to BamHI fragment that contains ICP4 codons 308 to 524. This fragment was inserted into BamHI-cleaved pATH3 (the NruI terminus was ligated to a filled-in BamHI site in the vector to maintain the correct reading frame). pXK442 encodes a 548 residue fusion protein designated FP442 that contains the first 323 amino acids from *trpE*, 7 amino acids (PPNSGGS) from the MCR, and ICP4 residues 308 to 524. There is a translation termination codon in the MCR immediately adjacent to codon 524.

Plasmids pXK526 and pXK453 were constructed by deleting from pXK525 or pXK449, respectively, the fragment that extends from the XmnI site at ICP4 codon 464 to the ScaI site in the amp^R gene and inserting in its place the fragment from pATER2Nae or pATH2Nae, respectively, that extends from the NaeI site in the MCR to the ScaI site in the amp^R gene. Plasmid pXK526 encodes a 234 residue fusion protein designated FP526 that contains the first 18 amino acids of *trpE*, 9 amino acids (IRGDPLEST) from the MCR, ICP4 residues 262 to 464, and 3 amino acids (LVN) from the MCR at the carboxy terminus. Plasmid pXK453 encodes a 538 residue fusion protein designated FP453 that contains the first 323 amino acids from *trpE*, 8 amino acids (PGDPLES) from the MCR, ICP4 residues 262 to 464, and 3 amino acids (LVN) from the MCR at the carboxy terminus.

Preparation of bacterial extracts

Overnight cultures grown in M9 medium supplemented with tryptophan (20 μg/ml) and ampicillin were diluted 1/20 into fresh M9 medium lacking tryptophan and shaken at 30°C for 3 hours. Indole acrylic acid was added (20 μg/ml final) and the culture was shaken at 30°C for an additional 5 hours. Cells were harvested by centrifugation, resuspended in ice-cold Buffer A (25 mM Tris-HCl pH 8.0, 10 mM EDTA), and disrupted by sonication at 4°C. Cell debris was removed by centrifugation at 12,000×g for 15 minutes. The *trpE*:ICP4 fusion proteins in the supernatant were precipitated at 4°C by addition of $(NH_4)_2SO_4$ to 40% of saturation. The proteins were collected by centrifugation and resuspended in Buffer B (10 mM Tris-HCl pH 8.0, 1 mM EDTA, 0.1% CHAPS, 10 mM mercaptoethanol) to a final concentration of 4 to 10 mg/ml. The solution was dialysed against Buffer B containing 0.01% CHAPS, clarified by centrifugation, mixed with an equal volume of glycerol, and stored at −20°. For preparation of the FP526 extract, an unbuffered solution of 0.1 mM TPCK in distilled water was used in place of Buffer A.

DNA mobility shift assays

Targets for DNA binding assays were isolated from the following plasmids which contain a fragment of HSV-1 DNA with a defined binding site for ICP4 (12−14,30) inserted between the HindIII and EcoRI sites in the MCR of either pUC13 or pBS (Stratagene). Plasmid pXK23 contains the AvaI fragment spanning the region from −128 to +57 relative to the mRNA cap site (+1) in the ICP0 gene. Plasmid pXK26 contains the EcoRI to SalI fragment spanning the region from −107 to +183 in the ICP4 gene. Plasmid pBetI contains the SstI to PvuII fragment spanning the region from −392 to −262 in the gene encoding glycoprotein D (gD). Plasmid pBetII contains the PvuII to HindIII fragment spanning the region from −262 to +11 in the gD gene. Plasmids pBetI and pBetII were generous gifts from Dr. Lewis Pizer. Plasmid pXK34 contains the HindIII to PvuII fragment spanning the region from +11 to +238 in the gD gene. Plasmids were purified from *E. coli* strain GM2163 (*dam*$^-$, *dcm*$^-$) by the procedure of Birnboim and Doly (40). The appropriate HindIII-EcoRI fragments were isolated from gels and end-labeled with [α-^{32}P]dATP (14). The standard DNA-binding reaction contained 4 ng of radiolabeled DNA fragments, 1 μg of sonicated calf thymus DNA as a non-specific competitor, and 0.2 to 4 μl of bacterial extract in 20 μl of DNA binding buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA, 10 mM mercaptoethanol, 0.1% CHAPS, 100 mM NaCl). After 30 minutes at 22°C, 5 μl of loading buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA, 10 mM mercaptoethanol, 0.1% CHAPS, 50% glycerol, 0.25 mg/ml bromophenol blue) was added and the sample was subjected to electrophoresis at 4°C for 3 hours at 200 volts on a 4% polyacrylamide gel (41). The location of the radiolabeled DNA was determined by autoradiography.

Preparation of antibody

FP449 was partially purified from bacterial extracts by DNA-affinity chromotography (14) and further purified by electroelution from SDS-polyacrylamide gels (42). The denatured

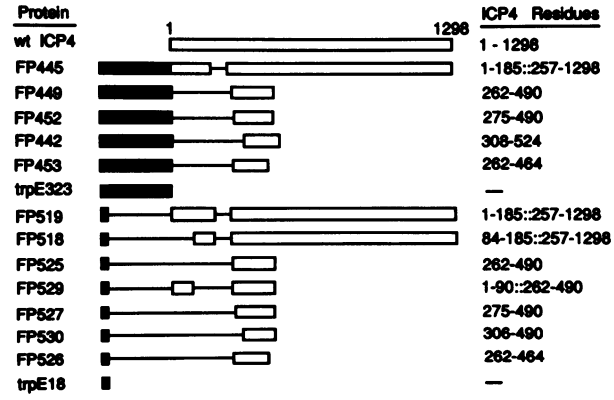

Figure 2. Schematic diagram of the regions encoded by each of the recombinant *trpE*:ICP4 genes. The filled box represents the *trpE* portion of the fusion protein. The open box represents the ICP4 portion of the fusion protein. The '400' series of fusion proteins includes the first 323 amino acids encoded by the *trpE* gene whereas the '500' series contains only the first 18 *trpE* residues. The double colon (::) indicates an internal deletion in the ICP4 gene. Details on construction of these recombinant genes are presented in Materials and Methods.

protein was mixed with Freund's adjuvant and injected into rabbits. IgG was purified from serum by $(NH_4)_2SO_4$ precipitation.

RESULTS

Construction and expression of recombinant trpE:ICP4 genes

The bacterial expression vectors used for this investigation contain the inducible promoter and part of the *trpE* gene from the *E. coli* tryptophan operon. Fragments obtained from the ICP4 gene were inserted into a multiple cloning region adjacent to either the 18th or the 323rd codon of the *trpE* gene. Details on construction of vectors and recombinant *trpE*:ICP4 genes are provided in Materials and Methods. Since the insertions were in-frame with *trpE* codons, the amino-terminal portion of each fusion protein includes either 18 (short leader) or 323 (long leader) residues encoded by the *trpE* gene. Schematic diagrams of the fusion proteins encoded by these recombinant plasmids are shown in Figure 2. Expression of recombinant *trpE*:ICP4 genes was induced by addition of IAA to log-phase cultures of *E. coli* maintained in tryptophan-free medium. Extracts that were enriched for soluble *trpE*:ICP4 fusion proteins were prepared as described in Materials and Methods and analysed by SDS-PAGE. In initial studies, we were unable to achieve detectable expression of any recombinant gene that contained ICP4 codons 186 to 256 (unpublished observations). An in-frame deletion of these codons from a full-length copy of the ICP4 gene that was fused to the long-leader version of the *trpE* gene resulted in the production of a 1564 residue protein designated FP445 that was readily-visible on a stained gel (Fig. 3A, lane 3). Faster-migrating bands that are presumably proteolytic cleavage products of FP445 were also observed. The stained gel revealed that expression of the recombinant genes encoding the long-leader fusion proteins designated FP449, FP452, FP442, and FP453 was also quite high and yielded polypeptides of the predicted molecular mass (Fig. 3A, lanes 4–7).

With the exception of FP530, the short-leader fusion proteins could not be detected on a stained gel (data not shown). An immununoblotting procedure using a polyclonal antibody generated against FP449 detected each of the short-leader fusion proteins and also revealed extensive proteolytic cleavage products from FP519 and FP518 (Fig. 3B). The yield of FP526 was consistently lower (Fig. 3B, lane 8) for unknown reasons.

DNA-binding properties of trpE:ICP4 fusion proteins

The primary goal of this investigation was to determine the miniumum region of the ICP4 gene that can encode a sequence-specific DNA binding protein. DNA mobility shift assays (43,44) were used to detect the formation of complexes between *trpE*:ICP4 fusion proteins and a radiolabeled target DNA that contains a well-defined ICP4 binding site (14). Calf thymus DNA was included in the binding reactions to minimize the formation of complexes between the radiolabeled target DNA and bacterial proteins in the extracts. DNA binding assays with recombinant genes expressing nearly full length ICP4 revealed at least two bands with FP445 and three bands with FP519 (Fig. 4). By comparison, a binding assay with native ICP4 purified from HSV-infected cells yielded at least three bands (Fig. 4B, lane 2), which we interpret as complexes between DNA and oligomeric forms of ICP4 (45). The results with FP445 and FP519 demonstrate that (i) fusion proteins containing either the long or the short *trpE* leader can bind to DNA, (ii) ICP4 residues 186 to 256 are not required for DNA binding, and (iii) both FP445 and FP519 apparently form oligomers which bind to DNA.

A series of additional deletions were made that yielded progressively smaller *trpE*:ICP4 fusion proteins. The results of

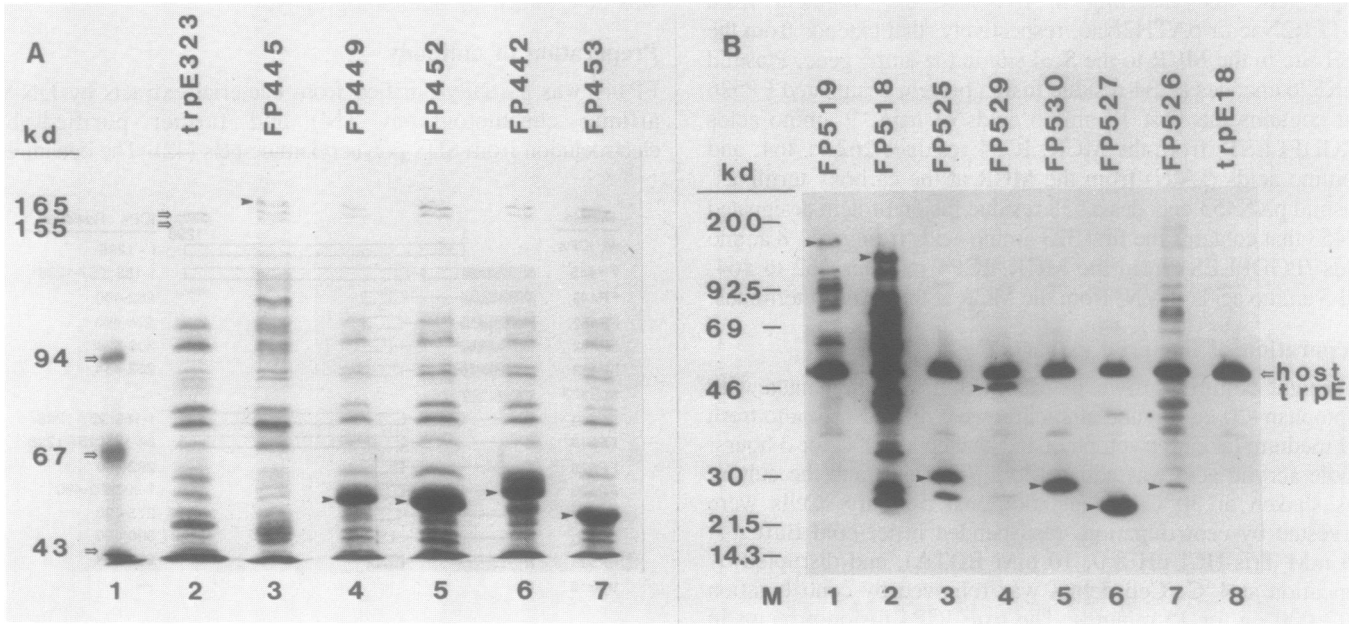

Figure 3. Analysis of fusion proteins present in bacterial extracts. The fusion proteins are indicated with solid arrows. Panel A: Extracts from cells expressing the indicated protein were electrophoresed through a 9% SDS-polyacrylamide gel (SDS-PAG) and stained with Coomassie Blue. Open arrows indicate molecular mass standards (Pharmacia) in Lane 1 and the 165 kd and 155 kd subunits of *E. coli* RNA polymerase in lane 2. The 36 kDa *trpE*323 protein encoded by pATH2 was not resolved from the tracking dye (lane 2). Panel B: Extracts from cells expressing the indicated protein were electrophoresed through a 15% SDS-PAG, transferred to nitrocellulose, and probed with antibody raised against FP449. The antibody-antigen complexes were tagged with ^{125}I-protein A and detected by autoradiography. The antibody also recognized the 57 kd host *trpE* protein.

DNA binding assays with the most informative of these fusion proteins are shown in Figure 4. FP449 and FP525 bound with high affinity to the target DNA and yielded a pattern with multiple bands. The fastest migrating FP449:DNA complex (Fig. 4A, lane 4) has nearly the same mobility as the major FP525:DNA complex; we suspect that the FP449 extract contains a protein that resembles FP525 due to proteolysis near the *trpE*:ICP4 junction in FP449. FP452 and FP527 were also capable of binding to the target DNA, although apparently with reduced affinity. In particular, we observed that although the molar amounts of FP449 and FP452 were nearly equal in the bacterial extracts (Fig. 3A, lanes 4 and 5), detection of equivalent DNA binding with FP452 required twenty-fold more extract than with FP449. This result suggests that ICP4 residues 262 to 274 contribute to formation of a high-affinity DNA-binding protein. No DNA binding was detected with FP442 or FP530, which suggests that a portion of ICP4 between residues 275 and 306 also contributes to the DNA binding capability of ICP4. Furthermore, no DNA binding was detected with FP453 or FP526, which suggests that a portion of ICP4 between residues 465 and 490 is important for DNA binding.

After we reported that native ICP4 forms homodimers (46), Shepard and DeLuca (47) proposed that the first 90 amino acid residues in native ICP4 may facilitate the formation of ICP4 oligomers that yield multiple electophoretic variants of ICP4:DNA complexes such as seen in Figure 4B, lane 2. To determine how the first 90 ICP4 residues affect the electrophoretic mobility of ICP4:DNA complexes, we compared the pattern of protein:DNA complexes obtained with FP519 and FP518 or with FP529 and FP525. FP519 yielded a well-resolved pair of bands that migrated significantly slower than the poorly-resolved pair of bands observed with FP518. FP529 yielded a minor band that co-migrated with the major FP525:DNA complex and two slower-migrating well-resolved bands of equal intensity (Fig.4B, lane 6), one of which migrated only slightly faster than the FP518:DNA complexes. Although interpretation of these results is complicated by the presence of proteolytic fragments in the extracts (Fig. 3B), it appears that the first 90 amino acid residues of ICP4 make a significant contribution to the mobility of ICP4:DNA complexes. However, because multiple bands were also observed with truncated forms of ICP4 that lack residues 1 to 90, these results do not reveal which amino acid residues are required to form the homodimers observed with wild-type ICP4 and imply that residues 262 to 490 may be sufficient for formation of oligomeric protein:DNA complexes.

DNA binding specificity of residues 262 to 490

Although the results shown in Figure 4 establish that ICP4 residues 262 to 490 are sufficient for DNA binding, it could be argued that the DNA binding specificity of the truncated polypeptide differs from that of native ICP4. To test this, DNase footprinting experiments were performed with extracts containing either FP449 or FP525. FP449 yielded a well-defined footprint on both strands (Fig. 5A) that coincided exactly with the footprint produced by wild-type ICP4 (14). The footprint with FP525 was slightly weaker (Figure 5B), but spanned the same nucleotides as seen with FP449 and wild-type ICP4. These results clearly demonstrate that the residues in FP525 are sufficient to form a polypeptide that binds in the same manner as native ICP4 to this particular target DNA.

At least six different ICP4 binding sites have been identified in the HSV-1 genome by DNase I footprinting and methylation interference assays (12,13,17,25,30). Among these binding sites there is some degeneracy from the consensus sequence proposed for an ICP4 binding site (13). Evidence for additional binding sites in the viral genome has been obtained from mobility shift

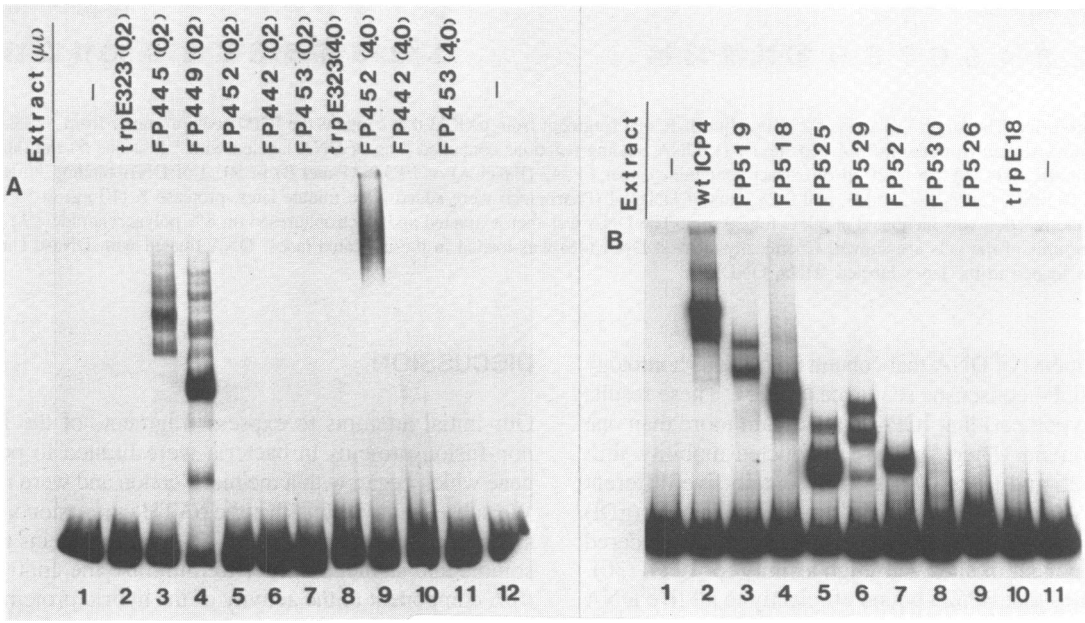

Figure 4. DNA mobility shift assays with *trpE*:ICP4 fusion proteins. Samples from DNA binding reactions containing the indicated protein and a radiolabeled 233 bp HindIII-EcoRI DNA fragment from pXK23 were subjected to gel electrophoresis. Autoradiograms of the gels are shown here. Panel A: Assays with long-leader *trpE*:ICP4 fusion proteins. Lanes 1 and 12: target DNA alone; Lanes 2–11: binding assays contained 0.2 (lanes 2–7) or 4.0 μl of extract (lanes 8–11) from bacteria expressing the indicated protein. Panel B: Assays with short-leader *trpE*:ICP4 fusion proteins. Lanes 1 and 11: target DNA alone; Lane 2: binding assay contained 300 fmoles of purified native ICP4 (14) and 1 μg of HeLa cell nuclear proteins to stabilize ICP4:DNA complexes (45); Lanes 3–10: binding assays contained 1 μl of extract from bacteria expressing the indicated protein.

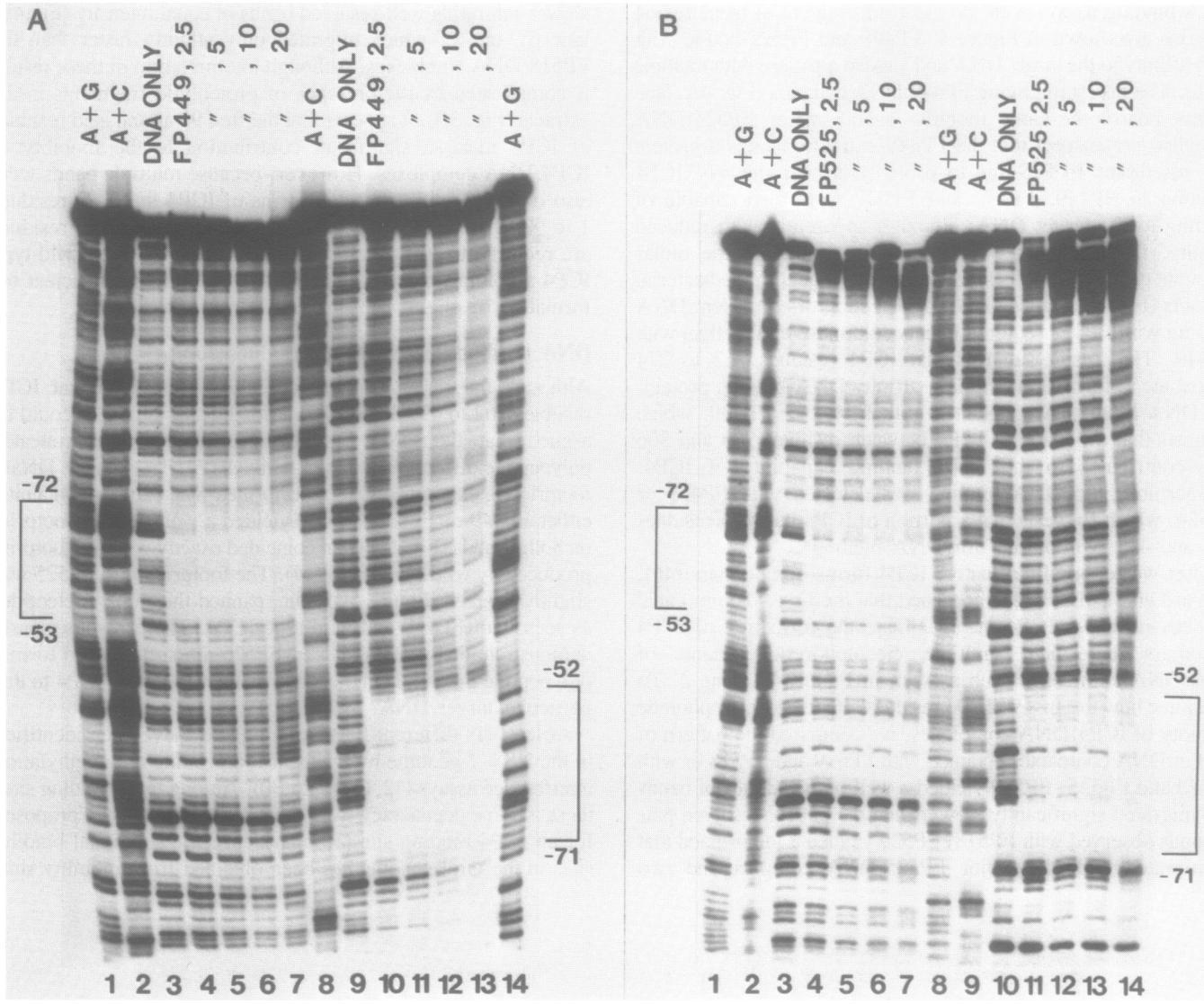

Figure 5. DNase I footprint. The target DNA was a 233 bp HindIII-EcoRI fragment from pXK23 that contains the ICP0 gene promoter from −128 to +57 oriented with the EcoRI site downstream from the mRNA cap site (+1). DNA binding reactions contained 4 ng of DNA labeled with ^{32}P at one 5′ end (lanes 1−7, EcoRI end; lanes 8−14, HindIII end) and 2.5 to 20 μl of extract containing either FP449 (Panel A) or FP525 (Panel B) in 50 μl of DNA binding buffer. After 30 min at 22°C, $MgCl_2$ (to 10 mM), $CaCl_2$ (to 2 mM), and 0.25 units of DNase I (Pharmacia) were added. One minute later, protease K (10 μg) in 55 μl of stop buffer (55) was added and the reaction was incubated at 50°C for 20 min. The DNA was then extracted and electrophoresed on 8% polyacrylamide/8 M urea sequencing gels (55). Autoradiograms of the gels are shown. Chemically-cleaved DNA (48) was loaded in the indicated lanes. DNA treated with DNase I in the absence of binding proteins was loaded in the lanes labeled 'DNA ONLY'.

assays using fragments of DNA that contain no obvious homology to the proposed ICP4 consensus sequence (15,25). These results have led to the hypothesis that ICP4 may contain more than one DNA binding domain. Therefore, we conducted mobility shift assays with FP525 and target DNAs that contain five different well-defined ICP4 binding sites. One of these target DNAs (gDI) lacks the hexanucleotide sequence (ATCGTC) that is considered to be a core component of sites with high affinity for ICP4 (30). The results revealed that FP525 bound efficiently to all five DNA probes (Fig. 6). No protein:DNA complexes were observed when control assays were performed with each of these target DNAs and an extract containing FP453 (data not shown). We conclude that the ICP4 domain from residue 262 to 490 contains sufficient information to recognize not only sites which differ slightly from the proposed consensus sequence but also at least one site which differs substantially from the proposed consensus sequence.

DISCUSSION

Our initial attempts to express fragments of the ICP4 gene as non-fusion proteins in bacteria were limited to portions of the gene which began with a methione codon and were not successful. We subsequently found that the pATH expression vectors yielded significant quantities of *trpE*:ICP4 hybrid proteins that were both soluble and stable. In order to minimize the contribution of the *trpE* component to the activity of the hybrid proteins, we derived the pATER series of vectors by deletion of *trpE* codons 19−323 inclusive from pATH. Both long-leader (*trpE* residues 1 to 323) and short-leader (*trpE* residues 1 to 18) variants of the *trpE*:ICP4 fusion proteins were produced in bacteria and tested for DNA binding activity. We consistently found that the long-leader variants gave much higher yields of protein, which implies that either the mRNAs or polypeptides encoded by the long-leader

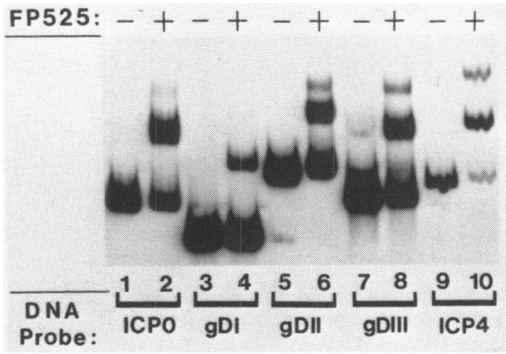

Figure 6. DNA mobility shift assays with different target DNAs. Samples from DNA binding reactions containing FP525 (1 μl of extract) and the indicated radiolabeled target DNA were subjected to gel electrophoresis. Autoradiograms of the gels are shown. The ICP0 probe is a 233 bp fragment from pXK23 that spans the ICP0 gene from −128 to +57. The gDI probe is a 174 bp fragment from pBetI that spans the gD gene from −392 to −262. The gDII probe is a 297 bp fragment from pBetII that spans the gD gene from −262 to +11. The gDIII probe is a 242 bp fragment from pXK34 that spans the gD gene from +11 to +238. The ICP4 probe is a 298 bp fragment from pXK26 that spans the ICP4 gene from −107 to +183.

recombinant genes are more stable than the short-leader versions. The recovery of most of these fusion proteins in the soluble component of bacterial extracts was unexpected, based on the experience of others with expression of some eucaryotic proteins in bacteria (49,50) and on our previous experience with native ICP4 (46).

The gel mobility shift assays clearly indicated which of the fusion proteins bound tightly to the target DNA. Although quantitative measurements were not attempted, we observed that, in general, for a given region from ICP4, both the long- and short-leader versions of the fusion proteins bound to DNA with similar affinity. Since the product of the *trpE* gene is a component of a tetrameric enzyme complex in *E. coli* (51), the presence of multiple protein:DNA bands in assays with the long-leader fusion proteins must be interpreted with caution. However, the multiple protein:DNA bands observed in assays with FP525 indicates the presence of oligomeric protein:DNA complexes that could result from (i) binding of FP525 independently to multiple sites along the target DNA, (ii) non-specific aggregation of FP525 with other bacterial proteins in the extract, or (iii) specific aggregation of FP525 with itself. The DNase protection experiment reveals that FP525 is bound to only 1 site on the target DNA, which rules out the first possibility. The presence of well-defined bands rather than broad smears argues against the second possibility. Thus we favor the third possibility. Although this implies that ICP4 residues 262 to 490 may facilitate formation of oligomers, there is no evidence presented here which indicates that the putative oligomerization of these truncated polypeptides is related to oligomerization of native ICP4.

Our conclusion that the portion of ICP4 spanning residues 262 to 490 contains a region that is required for sequence-specific DNA binding activity is consistent with analyses of ICP4 mutants conducted by other investigators (26,27,33,35,36), as described in the Introduction. The fact that bacterially-produced fusion proteins FP449 and FP525 bind to the ICP0 gene promoter in apparently the same manner as native ICP4, as evidenced by the equivalent footprints, indicates that ICP4 codons 262 to 490 are sufficient to encode a sequence-specific DNA binding protein. Although the results presented in Figure 4 suggest that residues 306 to 490 (FP530) or residues 262 to 464 (FP526) are insufficient to encode a sequence-specific DNA binding protein, there are two caveats that must be mentioned. First, the yield of FP526 was quite low, which may have reduced the sensitivity of the binding assays with this extract. Second, because FP530 and FP526 are quite small, the mobilities of the free and protein-bound DNAs might be similar. To circumvent these problems, DNA binding assays were performed with the long-leader versions of these hybrid proteins, which yielded similar negative results (Fig. 4A). Furthermore, DNA binding assays with a 34 bp oligonucleotide which migrates faster than the 233 bp target DNA revealed strong binding with FP525, moderate binding with FP527, very weak binding with FP530, and no detectable binding with FP526 (data not shown). The observation that FP527 binds to DNA with moderate affinity (Fig. 4B) and specificity (based on competition assays, data not shown) suggests that residues 262 to 274 are not directly involved in DNA binding. However, the observations that neither ICP4 variants with 2-amino-acid insertions adjacent to residue 263 or 274 (27) nor FP452 (long-leader equivalent of FP527) bind DNA efficiently (Fig. 4A) suggests that the tertiary structure immediately prior to residue 275 has a strong influence on the DNA-binding properties of ICP4.

The DNA-binding activity of FP525 suggests that the post-translational phosphorylation of ICP4 that occurs naturally in HSV-1 infected cells (42,52−54) is not required for DNA binding activity. Paterson and Everett (36) have shown that the DNA binding activity of ICP4 can be greatly reduced by alterations in the primary structure of ICP4 at locations distal to residue 490. Thus it could also be argued that post-translational modifications of ICP4 at residues distal to 490 may modify the DNA binding affinity or specificity of ICP4 (54). A bacterial expression system may provide a means of obtaining unmodified ICP4 for subsequent studies to determine how post-translational modifications affect DNA binding.

It should be noted that we have not shown that all residues from 262 to 490 are necessary to encode the sequence-specific DNA binding domain of ICP4, nor have we delineated which amino acids within this region make direct contact with the DNA. The suggestion that the portion of ICP4 from residue 445 to 487 resembles a helix-turn-helix motif that might be the DNA recognition element (27) should be testable if sufficient quantities of FP449 or FP525 can be obtained for structural studies.

ACKNOWLEDGMENTS

Many of the DNA clones and target DNAs were expertly prepared by Mary Jo Rademacher. Supervision for the oligonucleotide synthesis was provided by Liane Mende-Mueller. We are indebted to Bill Cashdollar for the suggestion to use the pATH vectors. Sequence analyses were performed using resources provided by BIONET. This research was supported by Public Health Service research grant AI 17246 from the National Institutes of Health; by a grant from the Cancer Research Foundation, Chicago, IL; and by a grant from the Cancer Center of the Medical College of Wisconsin.

REFERENCES

1. McGeoch,D.J., Dolan,A., Donald,S. and Brauer,D.H.K. (1986) Nucleic Acids Res. *14*,1727−1744.
2. Honess,R. and Roizman,B. (1974) J. Virol. *12*,1347−1365.
3. Watson,R.J., Preston,C.M. and Clements,J.B. (1979) J. Virol. *31*,42−52.
4. DeLuca,N. and Schaffer,P.A. (1985) Mol. Cell. Biol. *5*,1997−2008.
5. Dixon,R.A.F. and Schaffer,P.A. (1980) J. Virol. *36*,189−203.

6. Everett,R.D. (1988) J. Mol. Biol. *203*,739–751.
7. Gelman,I.H. and Silverstein,S. (1985) Proc. Natl. Acad. Sci. USA *82*,5265–5269.
8. O'Hare,P. and Hayward,G.S. (1987) J. Virol. *61*,190–199.
9. Watson,R. and Clements,J.B. (1980) Nature (London) *285*,329–330.
10. Everett,R.D. (1987) Anticancer Res. *7*,589–604.
11. Beard,P., Faber,S., Wilcox,K.W., and Pizer,L.I. (1986) Proc. Natl. Acad. Sci. USA *83*,4016–4020.
12. Faber,S.W. and Wilcox,K.W. (1986) Nucleic Acids Res. *14*,6067–6083.
13. Faber,S.W. and Wilcox,K.W. (1988) Nucleic Acids Res. *16*,555–570.
14. Kattar-Cooley,P. and Wilcox,K.W. (1989) J. Virol. *63*,696–704.
15. Kristie,T.M. and Roizman,B. (1986) Proc. Natl. Acad. Sci. USA *83*,3218–3222.
16. Michael,N., Spector,D., Mavromara-Nazos,P., Kristie,T.M. and Roizman,B. (1988) Science *239*,1531–1534.
17. Muller,M.T. (1987) J. Virol. *61*,858–865.
18. Kristie,T.M. and Roizman,B. (1986) Proc. Natl. Acad. Sci. USA *83*,4700–4704.
19. DiDonato,J.A. and Muller,M.T. (1989) J. Virol. *63*,3737–3747.
20. Gelman,I.H. and Silverstein,S. (1987) J. Virol. *61*,2286–2296.
21. Gelman,I.H. and Silverstein,S. (1987) J. Virol. *61*,3167–3172.
22. O'Hare,P. and Hayward,G.S. (1985) J. Virol. *56*,723–733.
23. Resnick,J., Boyd,B.A., and Haffey,M.L. (1989) J. Virol.*63*,2497–2503.
24. Roberts,M.S., Boundy,A., O'Hare,P., Pizzorno,M.C., Ciufo,D.M. and Hayward, G. S. (1988) J. Virol. *62*, 4307–4320.
25. Roizman,B., Kristie,T., McKnight,J.L., Michael,N., Mavromara-Nazos,P. and Spector,D. (1988) Biochimie *70*,1031–1043.
26. DeLuca,N.A. and Schaffer,P.A. (1988) J. Virol. *62*,732–743.
27. Shepard,A.A., Imbalzano,A.N. and DeLuca,N.A. (1989) J.Virol. *63*,3714–3728.
28. Arsenakis,M., Campadelli-Fiume,G. and Roizman,B. (1988) J. Virol. *62*,148–158.
29. Tedder,D.G. and Pizer,L.I. (1988) J. Virol. *62*,4661–4672.
30. Tedder,D.G., Everett,R.D., Wilcox,K.W., Beard,P. and Pizer,L.I. (1989) J. Virol. *63*,2510–2520.
31. Davison,M.J., Preston,V.G. and McGeoch,D.J. (1984) J. Gen. Virol. *65*,859–863.
32. Schroder,C.H., DeZazzo,J., Knopf,K.W., Kaerner,H.C., Levine,M. and Glorioso,J. (1985) J. Gen. Virol. *66*,1589–1593.
33. DeLuca,N.A. and Schaffer,P.A. (1987) Nucleic Acids Res. *11*,4491–4511.
34. Hubenthal-Voss,J., Houghten,R.A., Pereira,L. and Roizman,B. (1988) J. Virol. *62*,454–462.
35. Paterson,T. and Everett,R.D. (1988) Virol. *166*,186–196.
36. Paterson,T. and Everett,R.D. (1988) Nucleic Acids Res. *16*,11005–11025.
37. Dieckmann,C.L. and Tzagoloff,A. (1985) J. Biol. Chem. *260*,1513–1520.
38. Amann,E. and Brosius,J. (1985) Gene *40*,183–190.
39. Stow,N.D. and McMonagle,E.C. (1982) Eukaryotic Viral Vectors, Gluzman,Y., Ed., pp.199–204, Cold Spring Harbor Laboratory Publications, New York.
40. Birnboim,H.C. and Doly,J.(1979) Nucleic Acids Res.*7*,1513–1523.
41. Strauss,F. and Varshavsky,A. (1984) Cell *37*,889–901.
42. Faber,S.W. and Wilcox,K.W. (1986) Arch. Virol. *91*,297–312.
43. Fried,M. and Crothers,D. (1981) Nucleic Acids Res.*9*,6505–6525.
44. Revzin,A. (1989) BioTechniques *7*,346–355.
45. Kattar-Cooley,P. (1988) Ph.D. Dissertation, Medical College of Wisconsin, Milwaukee, WI.
46. Metzler,D.W. and Wilcox,K.W. (1985) J. Virol. *55*,329–337.
47. Shepard,A.A. and DeLuca,N.A. (1989) J. Virol. *63*,1203–1211.
48. Bencini,D.A., O'Donovan,G.A. and Wild,J.R. (1984) Biotechniques *2*,4–5.
49. Kadonaga,J.T., Carner,K.R., Masiarz,F.R. and Tjian,R. (1987) Cell *51*,1079–1090.
50. Roth,M.J., Tanese,N. and Goff,S.P. (1985) J. Biol. Chem. *260*,9326–9335.
51. Ito,J. and Yanofsky,C. (1969) J. Bacteriol. *97*,734–742.
52. Pereira,L., Wolff,M.H., Fenwick,M. and Roizman,B. (1977) Virol. *77*,733–749.
53. Preston,C.M. and Notarianni,E.L. (1983) Virol. *131*,492–501.
54. Wilcox,K.W., Kohn,A., Sklyanskaya,E. and Roizman,B. 1980. J. Virol. *33*,167–182.
55. Triezenberg,S.J., LaMarco,K.L. and McKnight,S.L. (1988) Genes and Develop. *2*,730–742.