The EMBO Journal vol.10 no.13 pp.4321 – 4329, 1991

Identification of the origin of replication of bovine papillomavirus and characterization of the viral origin recognition factor E1

Mart Ustav, Ene Ustav, Paul Szymanski and Arne Stenlund[1]

Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA

[1]Corresponding author

Communicated by U.Pettersson

Expression of the viral polypeptides E1 and E2 is necessary and sufficient for replication of BPV in mouse C127 cells. By providing these factors from heterologous expression vectors we have identified a minimal origin fragment from BPV that contains all the sequences required *in cis* for replication of BPV in short term replication assays. This same sequence is also required for stable replication in the context of the entire viral genome. The identified region is highly conserved between different papillomaviruses, and is unrelated to the previously identified plasmid maintenance sequences. The minimal *ori* sequence contains a binding site for the viral polypeptide E1, which we identify as a sequence specific DNA binding protein, but surprisingly, an intact binding site for the viral transactivator E2 at the *ori* is not required. The isolated origin shows an extended host region for replication and replicates efficiently in both rodent and primate cell lines.

Key words: DNA replication / origin of replication/ papillomavirus

Introduction

DNA replication in mammalian cells has been studied mainly using viral model systems (Challberg and Kelly, 1989; Stillman, 1989). In particular, studies using simian virus 40 (SV40) have yielded a wealth of information about the mammalian replication machinery and especially the processes of initiation of DNA replication and elongation have been studied in great detail using this system (Stillman, 1989; Borowiec *et al.*, 1990). Contributing reasons for the success of SV40 as a model system have been that the viral components of the replication machinery, the *ori*, which specifies the site of initiation, and T-antigen, which recognizes and targets the *ori* for the cellular proteins that perform the process of DNA synthesis, are very well characterized. This has allowed detailed analysis of SV40 DNA replication both *in vivo* and *in vitro* (Challberg and Kelly, 1989; Stillman, 1989). This has been of particular importance since cellular *cis*-acting sequences or *trans*-acting factors involved in initiation of DNA replication from cellular origins have so far not been identified.

Bovine papillomavirus, BPV, has been considered an interesting model for study of DNA replication because the viral DNA is maintained as a freely replicating plasmid with a constant, relatively low, copy number in BPV transformed cells (Mecsas and Sugden, 1987). BPV replication therefore offers some unique aspects for study that are different from replication in lytic viral systems and appear to be more similar to cellular DNA replication. However, detailed studies of the replication properties of this virus have been severely hampered by the lack of consistent information about the basic viral components that are required for viral DNA replication, including the necessary *cis*-elements and the identity and function of the viral *trans*-acting factors.

The current information about *cis*-acting elements that are required for replication presents a complicated picture. Two such elements were initially identified by introduction of fragments of viral DNA linked to a selectable marker into BPV transformed cells which provided the necessary viral *trans*-acting factors. These experiments resulted in the identification of two short regions in the BPV genome (termed plasmid maintenance sequences, PMSs), that could independently of each other maintain the *neo*-linked fragment as replicating plasmids (Lusky and Botchan, 1984). Subsequently one of these sequences, PMS1, was characterized further using a short term replication assay. In this assay a plasmid containing a subgenomic fragment of BPV was tested for replication in BPV transformed cells in the absence of selection. It was shown that linker insertions in the previously defined PMS1 region of this fragment failed to replicate, indicating that the PMS1 region corresponded to a *cis*-acting element that was also required for short term replication (Lusky and Botchan, 1986).

Very recently a different region has been proposed to be an important *cis*-acting element for replication of BPV. This suggestion was based on the observation that the E1 and E2 polypeptides, which are both required for replication, can form a complex when overexpressed together from baculovirus vectors (Mohr *et al.*, 1990; Blitz and Laiminis, 1991). This complex can bind to the cluster of strong E2 binding sites *in vitro*, and it was proposed that the cluster of E2 binding sites constituted the origin of replication (Mohr *et al.*, 1990).

Physical mapping of the site of initiation of replication has been carried out in two studies. One study used electron microscopy to map the origin in wild-type DNA to a sequence within the PMS1 region; however, only a very limited number of molecules were measured (Waldeck *et al.*, 1984). Another more recent study used two dimensional gel electrophoresis to map the site of initiation of replication to yet another region located in the E2 inducible enhancer ~0.4 kb away from the PMS1 region (Yang and Botchan, 1990). Neither of these studies detected any activity in the PMS2 region.

In a previous study we developed an efficient short term replication assay and determined for the first time that two viral polypeptides, encoded by the E1 and E2 open reading frames, were necessary and sufficient for replication of BPV in mouse C127 cells (Ustav and Stenlund, 1991). In this

report we have addressed which *cis*-acting sequences in the viral genome are required for viral DNA replication and also what role the two viral polypeptides play in the viral replication process. We have used the short term replication assay to identify a minimal *ori* fragment that contains all the sequences required for E1 and E2 dependent replication in C127 cells. This sequence is different from all of the previously suggested *cis*-acting elements for replication. Furthermore, we have determined that the E1 polypeptide has specific DNA binding activity and serves to recognize this origin, while intact binding sites for the known DNA binding transcriptional transactivator E2 (Spalholz *et al.*, 1985; Androphy *et al.*, 1987; Moskaluk and Bastia, 1987) are not required for replication.

Results

All restriction fragments that replicate have one region in common

We have previously shown that the viral E1 and E2 polypeptides are the only viral factors that are required for replication of BPV in C127 cells. In addition we have shown that a 3.2 kb fragment from the BPV genome contains the *cis*-acting sequences required for replication (Ustav and Stenlund, 1991). To delineate further the sequences required *in cis* for replication of BPV, we co-transfected mouse C127 cells with restriction fragment mixtures from various digests of the cloned viral DNA, plus vectors expressing E1 (pCGEag) and E2 (pCGE2) proteins to provide the required *trans*-acting factors. Following transfection, low molecular weight DNA was isolated at various time points, digested with *Dpn*I (which cuts unreplicated DNA) (Peden *et al.*, 1980) linearized with a single cutting enzyme for each fragment and analyzed by Southern blotting. Any fragment that was capable of recircularization and replication would be detected as a *Dpn*I resistant band. Our previous experiments had shown that restriction fragments could circularize and replicate in this assay regardless of whether the generated ends were compatible or not (data not shown). A drawback of this assay is that small fragments are difficult to assay for replication since the migration of small replicating fragments will coincide with the *Dpn* sensitive material at the bottom of the gel. Therefore we chose to map the *cis*-acting elements by generating larger fragments which partly overlapped, and to deduce from the overlap which sequences were important for replication.

The fragment mixtures were generated by first digesting the DNA with *Bam*HI, which released the viral DNA from the prokaryotic vector pML. In addition the DNA was digested with other enzymes that cleave the viral DNA at various positions (see Figure 1, bottom panel). To identify which of the fragments replicated, we used the same restriction fragment mixture as a marker which is shown to the left of each set of time points. The *Bam*HI + *Hin*dIII digest gives rise to three fragments (2.5, 2.6 and 5.5 kb), the largest of which replicated. The *Bam*HI + *Cla*I digest gives rise to four fragments (0.7, 1.5, 2.6 and 5.0 kb) where again the largest fragment replicated. These results indicated that the sequences upstream of the *Cla*I restriction site (nucleotide 7477) were dispensable for replication. This was an interesting result since PMS1 is not present in the replicating fragment. To generate fragments that lacked the downstream sequences we digested with *Bam*HI + *Kpn*I and

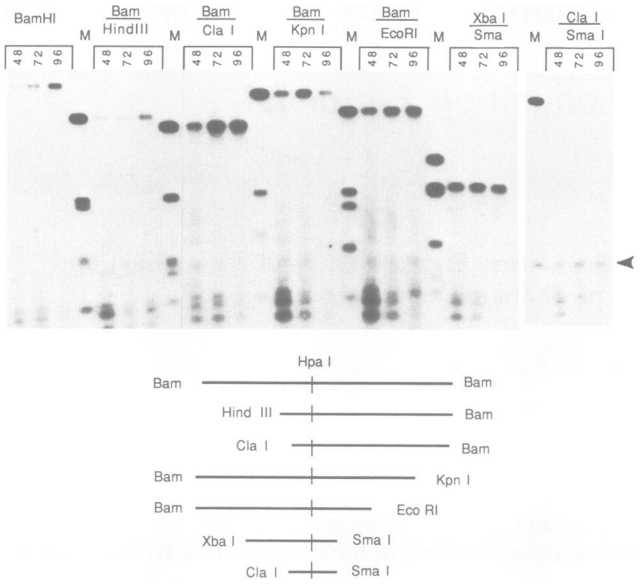

Fig. 1. All restriction fragments that replicate have one region in common. The viral genome was separated from the vector sequences by digestion of the plasmid pMLBPV with *Bam*HI. Various additional restriction enzymes were then used to generate subgenomic fragments of BPV, which were transfected together with expression vectors producing E1 and E2 polypeptides into mouse C127 cells. At the indicated time points low molecular weight DNA was harvested and digested with *Dpn*I and linearized with a single cutting enzyme for each generated fragment. The enzymes used for linearization were in each case *Bam*HI: *Xba*I, *Bam*−*Hin*dIII: *Xba*I + *Eco*RI, *Bam*−*Cla*I: *Eco*RI + *Xba*I + *Hin*dIII, *Bam*−*Kpn*I and *Bam*−*Eco*RI: *Hin*dIII + *Sma*I, *Xba*I−*Sma*I: *Hin*dIII + *Eco*RI, *Cla*I−*Sma*I: *Hpa*I + *Eco*RI. The lane M indicates the marker which in each case was the input restriction fragment mixture that was used in the transfection. The arrow in the right panel indicates the replicating 1.5 kb *Cla*I−*Sma*I fragment. Below is shown a diagram of the viral genome indicating the fragments that replicated in this assay.

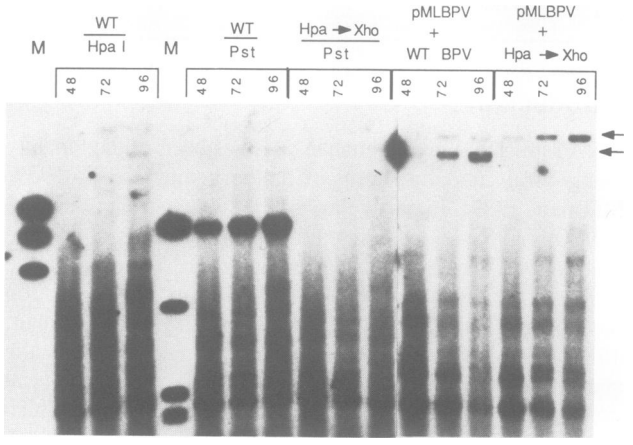

Fig. 2. Cleavage or mutation of the unique *Hpa*I site disrupts DNA replication. Left panel: pMLBPV was digested with *Bam*HI + *Hpa*I and tested for replication as described in Figure 1. The low molecular weight DNA was digested with *Dpn*I and *Xba*I + *Eco*RI for linearization. Middle panel: the wild-type genome and a linker insertion mutant carrying an *Xho*I linker at the *Hpa*I site were digested with *Pst*I and tested for replication as described in Figure 1. M represents the input digest. To the right is shown a complementation assay where pMLBPV was co-transfected with either the wild-type BPV genome which was linearized at the *Bam*HI site, or with the *Xho*I linker insertion mutant linearized at the *Bam*HI site.

*Bam*HI + *Eco*RI. In these digests again the large fragments replicated, indicating that the sequences downstream of the *Eco*RI site (nucleotide 2113) were dispensable for replication.

To further narrow down the elements required for replication we digested with *Bam*HI + *Xba*I + *Sma*I which gives rise to four fragments (3.5, 2.7, 2.6 and 1.6 kb). The fragment that replicated was a 2.7 kb fragment (nucleotides 6132−945) that spans the *Hpa*I restriction site, indicating that the sequences downstream of nucleotide 945 were not required for replication. These combined results indicated that the *cis*-acting element required for replication was located somewhere between the *Cla*I site at nucleotide 7477 and the *Sma*I site at nucleotide 945. To verify that this was the case we tested the digest *Cla*I + *Sma*I, which gives rise to three fragments (0.5, 1.5 and 8.5 kb). In this case we could detect replication of the expected 1.5 kb fragment migrating close to the *Dpn*I sensitive material (arrow). In summary, in all cases replication of one fragment from each digest could be detected (see bottom part of Figure 1). These fragments all had same sequence in common and overlapped the unique *Hpa*I restriction site. We concluded from these results that a single region of the BPV genome located between the *Cla*I and *Sma*I site was sufficient for E1 and E2 dependent replication in C127 cells.

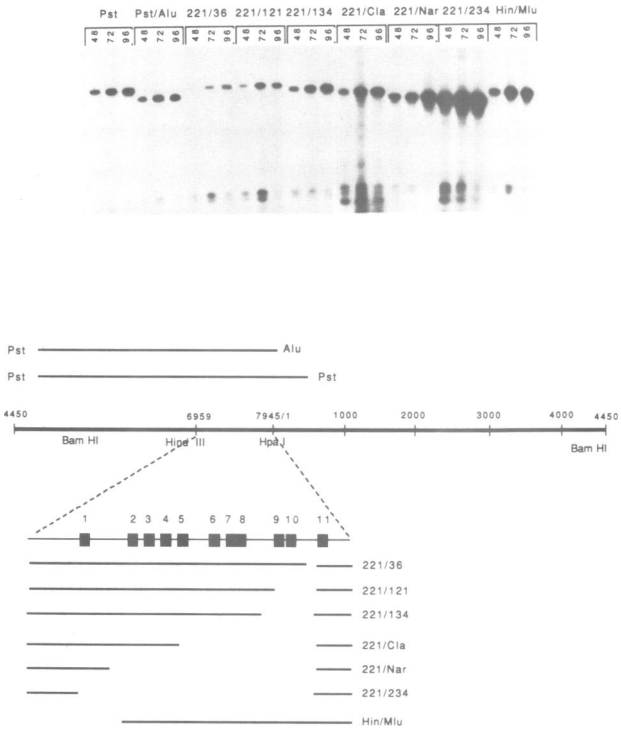

Fig. 3. The E2 dependent enhancer is not required for replication. A series of deletion mutants generated in the region around the *Hpa*I site were tested for replication in the short term replication assay. The deletions that are shown schematically in the lower part of the figure were assayed in the context of a subgenomic *Pst* I fragment as shown below. The filled boxes in the lower part of the figure represent the known E2 binding sites in the E2 enhancer.

*Cleavage of the unique Hpa*I *site in the viral genome disrupts DNA replication*

The sequence common to all replicating fragments in Figure 1 contained the the unique *Hpa*I site. We therefore tested the digest *Bam*HI + *Hpa*I for replication. As shown in Figure 2, left panel, this digest gives rise to three fragments (2.6, 3.5 and 4.5 kb) none of which replicated. This was the only digest that we tested where no replicating fragment could be detected. To determine if this result reflected the presence of important sequences at the *Hpa*I site, we inserted an *Xho*I linker into this site and assayed a larger *Pst*I fragment containing this insertion for replication. The linker insertion into the *Hpa*I site abolished the ability of the *Pst*I fragment to replicate as shown in Figure 2, while the wild-type *Pst*I fragment replicated efficiently. To determine the importance of this region in the context of the entire viral genome, we inserted the *Xho*I linker into the *Hpa*I site of the plasmid pMLBPV. This mutant was assayed for replication in the presence of the wild-type genome as a source of *trans*-acting factors. To able to distinguish between the wild-type genome and the *Xho*I linker insertion mutant,

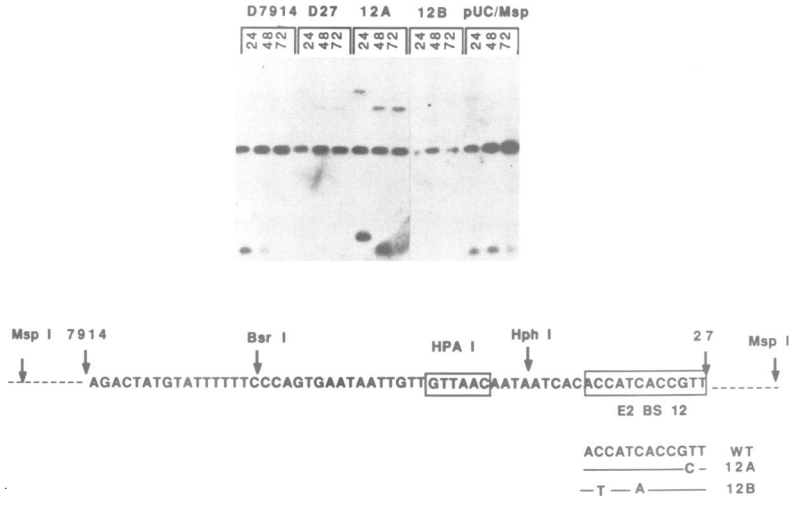

Fig. 4. E2 binding sites are not required for *ori* function. Deletions and point mutations were generated in the context of the *Msp* fragment and tested for replication in the short term replication assay. D7914 and D27 are deletions from the upstream and downstream sides, respectively, with endpoints at the indicated nucleotide. 12A and 12B are point mutants in E2 BS12 in the context of the *Msp* fragment as shown in the bottom part of the figure.

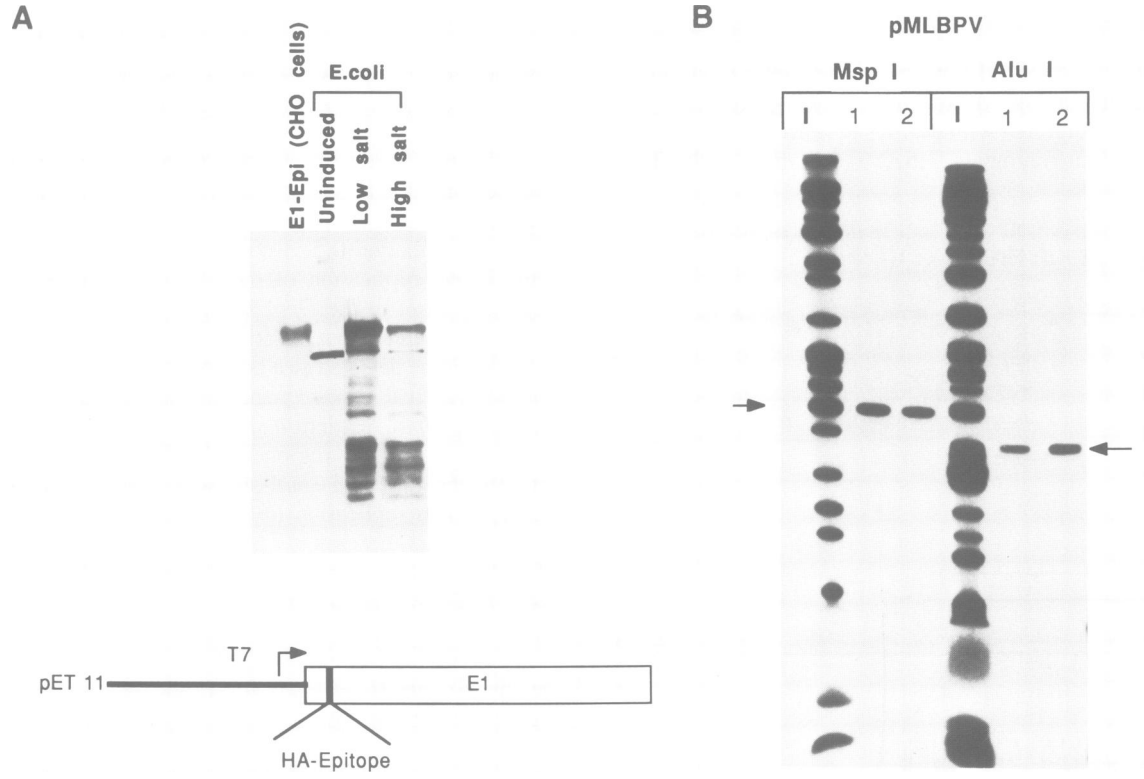

Fig. 5. (**A**) Immunoblot showing E1 protein expressed in CHO cells and in *E.coli*. Extracts from induced and uninduced cultures of *E.coli* harboring the plasmid pETE1 were separated on an SDS−PAGE gel and transferred to nitrocellulose. The blot was probed with the monoclonal antibody 12CA5 that recognizes the HA epitope inserted into the E1 coding sequence. As a marker, an extract from CHO cells transfected with an epitope containing E1 expression vector was included. The uninduced lane contains five times the amount of protein that was used in the low and high salt fractions from induced cultures. (**B**) E1 binds specifically to the *ori* fragment. pMLBPV was digested with *Msp*I or *Alu*I and end labeled. The fragment mixtures were incubated with E1 beads for 1 h at room temperature followed by several washes. Retained fragments were released from the beads by phenol extraction and analyzed on a 5% acrylamide gel. I represents the input fragment mixture. 1 and 2 represent the retained fragments in the presence and absence of carrier DNA, respectively.

the wild-type DNA was transfected still linked to the prokaryotic vector pML, while the mutant was digested with *Bam*HI before the transfection. This gives rise to a size difference between the two DNAs of ~2.5 kb. As shown in the right panel of Figure 2, the *Bam*HI digested wild-type DNA as well as undigested pMLBPV replicated together, whereas the mutant with the *Xho*I linker insertion failed to replicate even in the presence of a wild-type genome. This result indicated that the *Xho*I linker insertion at the *Hpa*I site affects a *cis*-acting element that is required for replication of BPV. This mutant was also defective for replication when assayed in a stable replication assay (data not shown).

The E2 binding sites in the E2 dependent enhancer are not required for replication of BPV

Because of the recently reported results that an origin of replication was located in the E2 dependent enhancer (Mohr *et al.*, 1990; Yang and Botchan, 1990), we decided to investigate the importance of this region in replication. We continued to define the *cis*-acting elements in the context of a large fragment. As shown in Figure 2 a 3.5 kb *Pst*I fragment (nucleotides 4780−471) replicated efficiently in the short term replication assay. Using this fragment as a base we next analyzed a series of internal deletions in the E2 enhancer region upstream of the *Hpa*I site. These deletions were generated from a series of linker insertion mutants (Lusky and Botchan, 1986). The deleted constructs and the results from the short term replication assays are shown in Figure 3. The first set of deletions have a fixed 3′ endpoint at nucleotide 7900 which is 46 nucleotides upstream of the *Hpa*I site. The 5′ endpoints of the deletions were chosen such that all the E2 binding sites (BSs) present in the URR (with the exception of BS11) were successively deleted. As shown in the top part of Figure 3 the *Pst*I fragment from all these constructs replicated, indicating that the sequences upstream of nucleotide 7900, including the E2 binding sites in the URR region (BS1−10), were dispensable for *ori* function. To narrow down further the important sequences downstream of the *Hpa*I site we generated a fragment which extended from the *Pst*I site at nucleotide 4780 to an *Alu*I site at nucleotide 52. This fragment replicated to similar levels to the larger *Pst*I−*Pst*I fragment indicating that the *Pst*I−*Alu*I fragment still retained all sequences important for replication. Finally, to confirm that the PMS1 region did not play a role in this assay we generated a deletion that removes the sequences between *Hin*dIII and *Mlu*I (nucleotides 6959−7353). This deletion mutant also replicated. Taken together, these results indicated that sequences upstream of nucleotide 7900 and downstream of nucleotide 52 were dispensable for replication.

The experiments above resulted in a deduced minimal *cis*-acting element for replication of ~100 bp (nucleotides 7900−52) in length. Consistent with a replication element at this position, our results indicated that the integrity of the *Hpa*I site, which is located in the center of this deduced sequence, was essential for replication. We therefore cloned

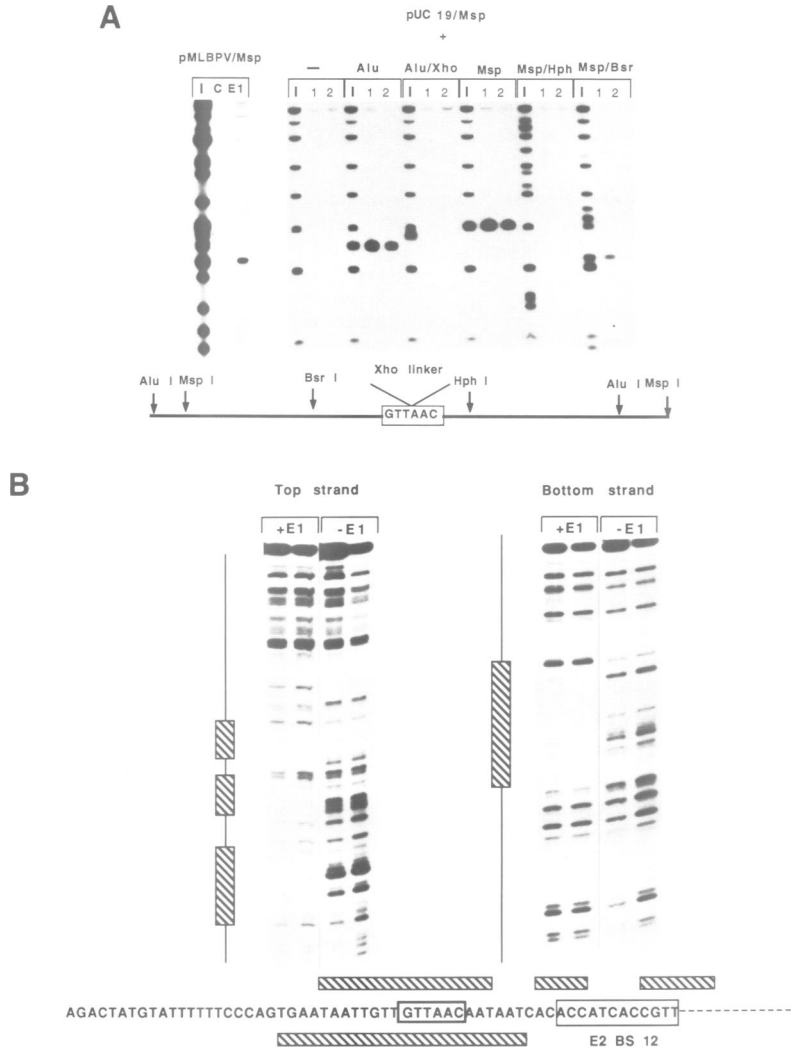

Fig. 6. (**A**) Mutations in the *ori* region destroy E1 binding. Left panel: an end-labeled *Msp*I digest of pMLBPV was precipitated with either beads generated with a control *E.coli* extract (C) or an E1-containing extract (E1). I represents the input fragment mixture. Right panel. DNA precipitation assays were carried out by mixing pUC19 digested with *Msp*I with fragments from the minimal *ori* plasmids. − indicates no added specific fragment. Alu indicates addition of the purified *Alu*I *ori* fragment, Alu/Xho indicates addition of the *Alu*I fragment with an *Xho*I linker inserted at the *Hpa*I site. Msp represents addition of the purified *Msp*I fragment, Msp/Hph and Msp/Bsr indicate additional digestion of the *Msp*I *ori* plasmid with *Hph*I or *Bsr*I. I represents the input mixture of fragments, 1 and 2 represent the retained fragments in the presence or absence of carrier DNA, respectively. (**B**) E1 footprints on the *ori* fragment. DNase footprint analysis was carried out by incubating the end-labeled *Msp*I fragment with E1 beads. After several washes the DNA retained on the beads was treated with DNase I at two different concentrations of DNase. The −E1 samples were treated with DNase in solution. The protected regions are shown schematically at the bottom of the figure.

two small restriction fragments *Alu*I−*Alu*I (nucleotides 7891−52) and *Msp*I−*Msp*I (nucleotides 7903−81) into the polylinker of the plasmid vector pUC19 to generate pUC/Alu and pUC/Msp, respectively. Both these plasmids replicated (data not shown).

Intact binding sites for E2 are not required at the origin

To delineate further the minimally required *ori* sequences we generated deletions from both ends of the *Alu*I and *Msp*I fragments in the context of pUC19. As shown in Figure 4, deletions from the upstream side to nucleotide 7914 and from the downstream side to nucleotide 27 still replicated while deletions to nucleotide 7932 (*Bsr*I) and nucleotide 13 (*Hph*I) failed to replicate to detectable levels (data not shown). This analysis resulted in a minimal *ori* fragment of only 60 nucleotides. This fragment contains one known E2 binding site (BS12). To determine if this E2 binding site was important for replication we introduced point mutations in positions in this site that were likely to lead to loss of E2 binding (Li *et al.*, 1989). These mutant sites were tested for binding to E2 in gel retardation assays. BS12 which is a very weak E2 binding site (Li *et al.*, 1989) was reduced to undetectable levels by both of these mutations (data not shown). The mutants were tested for replication in the context of pUC/Msp. As seen in Figure 4 the mutant 12A replicated to wild-type levels, and the mutant 12B also replicated, albeit at a slightly reduced level. These results indicated that an intact E2 binding site at the origin was not required for replication. The pUC19 vector contains two known E2 binding sites. These sites are unlikely to be involved in replication of the construct since the *Alu*I and *Msp*I fragments were removed from the plasmid vector and instead linked to a fragment from the late region of BPV that lacks known E2 binding sites, both fragments still replicated efficiently (data not shown).

These results clearly indicated that recognition of the origin was unlikely to reside with the E2 polypeptide. As a candidate for an *ori* recognition factor we instead turned our attention to the E1 polypeptide. This protein has previously been reported to bind DNA non-specifically (Santucci *et al.*, 1990; Mohr *et al.*, 1990; Blitz and Laiminis, 1991) and recently site specific binding by the E1 polypeptide has been demonstrated (Wilson and Ludes-Meyer, 1991). We chose to express the E1 polypeptide in *Escherichia coli* by inserting the E1 coding sequence into the pET11 vector (Studier *et al.*, 1990). The E1 coding sequence was modified by the insertion of an epitope tag 30 amino acids from the N-terminus (Field *et al.*, 1988). This insertion does not affect the function of the protein *in vivo* (Ustav and Stenlund, 1991). The E1 protein, extracted as described in Materials and methods, was largely soluble in low salt, low detergent conditions. As shown in Figure 5A, which shows an immunoblot probed with the monoclonal antibody 12CA5 that recognizes the epitope tag, the *E.coli* expressed E1 protein co-migrated with E1 protein expressed in transient transfection assays in CHO cells. The lower molecular weight bands that react with the antibody are not degradation products but appear to be C-terminal truncations. This conclusion is based on immunoblots with a C-terminal antibody. To assay for DNA binding activity of E1, the crude soluble fraction from the *E.coli* lysate was bound to an affinity resin consisting of the monoclonal antibody, which is directed against the inserted epitope, crosslinked to protein A Sepharose. These 'E1 beads' were then used for DNA precipitation assays.

The E1 polypeptide binds specifically to the origin fragment

The DNA precipitation assays were carried out by mixing the E1 beads with end-labeled restriction fragment mixtures generated by digesting the entire pMLBPV plasmid with *Msp*I or *Alu*I. After several washes, fragments that were associated with the E1 beads were identified by gel electrophoresis, using the fragment mixture as a marker. As can be seen in Figure 5B, in both digests one single specific fragment was retained on the E1 beads. In both cases, this fragment corresponded to the fragment containing the minimal *ori* region as determined by digestion with *Hpa*I.

To verify that the polypeptide responsible for binding to these fragments was the overexpressed E1 polypeptide we generated a control extract from an uninduced culture. This extract was not capable of precipitating any specific fragment from pMLBPV (Figure 6A, left panel). To further our analysis, binding of smaller fragments from the *ori* region was tested. We used the constructs that we had previously used for replication experiments where the *Alu*I and *Msp*I minimal *ori* fragments had been cloned into the polylinker of pUC19. As an internal control fragments from these constructs were combined with an end-labeled fragment mixture derived from pUC19 digested with *Msp*I. As shown in Figure 6A, none of the pUC19 fragments were retained on the E1 beads. When the purified *Alu*I fragment was included, as expected this fragment was retained (Alu). In contrast, the *Alu*I fragment containing an *Xho*I linker insertion at the *Hpa*I site (Alu/Xho), which is defective for replication, was not retained. To generate smaller labeled fragments from the *ori* region, pUC/Msp was digested with *Eco*RI and *Hin*dIII which cleave on either side of the *ori* fragment in the polylinker sequence. The sample was also

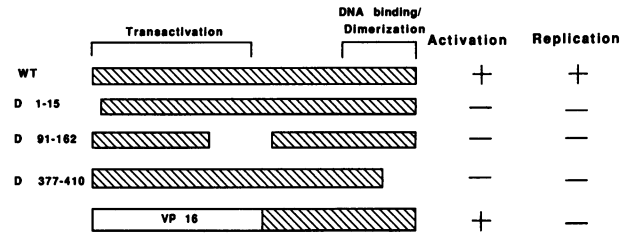

Fig. 7. Deletions in the E2 coding sequence abolish replication. Deletions were generated in the E2 coding sequence in the context of the pCGE2 expression vector. The deletion mutants were tested for their ability to support replication in the short term replication assay and also for their ability to transactivate gene expression from the BPV genome in transient expression assays. On the left the deleted amino acids are indicated, on the right the activity of the respective polypeptide in transactivation and replication assays is indicated.

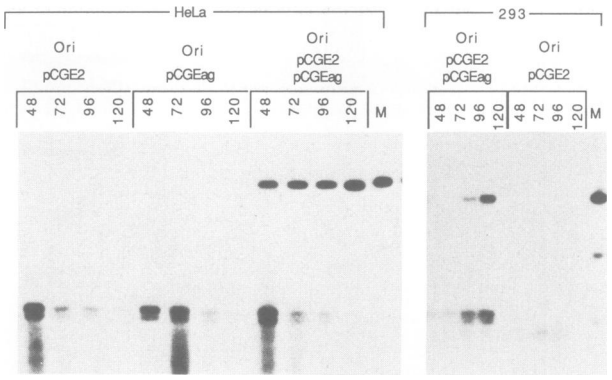

Fig. 8. The *ori* fragment replicates in primate cell lines. Short term replication assays were carried out in the human cell lines HeLa and 293 as described in Figure 1. Transfections of the *ori* fragment were carried out with or without the E1 (pCGEag) and E2 (pCGE2) expression vectors as indicated in the figure.

digested with *Hph*I or *Bsr*I, which cleaves the *ori* fragment immediately downstream and upstream of the *Hpa*I site, respectively, in addition to multiple cleavage sites in the vector. The full length *Msp*I fragment bound to the E1 beads, while no fragment was retained from the *Hph*I digest. The *Bsr*I fragment corresponding to the right hand portion of the *ori* fragment could still bind E1. These results indicated that the sequences important for E1 binding were located adjacent to the *Bsr*I site and included the *Hpa*I and the *Hph*I restriction sites.

To pinpoint the sequences involved in E1 binding we performed DNase footprint analysis of the origin region (Figure 6B). The *Msp*I fragment was end-labeled and incubated with E1 beads, washed several times, and the retained DNA was treated with DNase I. On the bottom strand a single prominent protection overlapping the *Hpa*I restriction site was apparent. On the top strand, the protection pattern was more complex and extended downstream over a large region with interruptions of short DNase sensitive sequences.

E2 mutants fail to support replication

The E2 open reading frame encodes several related proteins (for a review see Lambert *et al.*, 1988). The full length open reading frame encodes a 410 amino acid protein where the N-terminal half of the protein is important for transcriptional activation (Giri and Yaniv, 1988; McBride *et al.*, 1989). The C-terminal 100 amino acids are sufficient for DNA binding

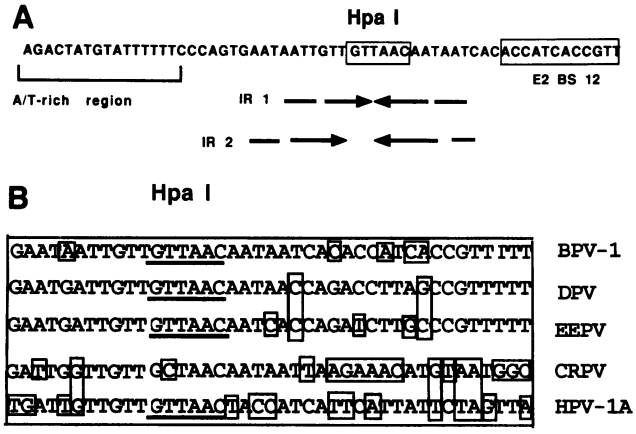

Fig. 9. (**A**) DNA sequence of the minimal *ori* region showing the A/T rich region, the conserved *Hpa*I site, and the inverted repeat regions. (**B**) Sequence comparison between the *ori* region of BPV1 and the corresponding region of four other papillomaviruses. The *Hpa*I recognition sequence is underlined.

and dimerization of the protein (McBride *et al.*, 1988, 1989). We have previously shown that the two repressor forms of E2 (E2C and E8/E2) that share the C-terminal domain with the E2 transactivator, including the dimerization and DNA binding domain, are incapable of supporting replication of the BPV origin (Ustav and Stenlund, 1991). To analyze the requirement for the E2 transactivator further we generated deletions within the E2 coding sequence affecting either the DNA binding/dimerization domain (deletion of 33 amino acids at the C-terminus of the protein), or deletions affecting the transactivation domain (deletion of the N-terminal 15 amino acids, or a deletion of amino acids 91−162) (Figure 7). These deletions all led to loss of transactivation activity consistent with previous reports (McBride *et al.*, 1989) and simultaneous loss of ability to support replication of an *ori* fragment. When tested for expression by immunoprecipitation of the E2 polypeptides after transient expression, these mutants were all expressed at comparable levels and appeared stable. In addition we tested a VP16/E2 fusion protein that retains the site specific transactivation properties of E2 (Li *et al.*, 1991). This polypeptide also failed to support replication indicating that transactivation function *per se* was not sufficient for the replication activity of E2. These results indicated that the N-terminal transactivation domain, and at the C-terminus, either DNA binding or dimerization, or both, were important for the replication function of E2.

The BPV ori replicates in primate cells

Replication of BPV1 DNA has been shown to take place in a number of rodent cell lines and in primary bovine fibroblasts, but primate cells are not permissive for viral DNA replication (Black *et al.*, 1963; Meischke, 1979; Roberts and Weintraub, 1986). To determine if this limited host range was due to restrictions in viral gene expression, or resided in the replication machinery *per se*, we tested several established primate cell lines for their ability to support BPV *ori* replication in a short term replication assay with the two *trans*-acting factors provided from expression vectors. As shown in Figure 8, replication of an *ori* fragment will take place both in HeLa and 293 cells when the E1 and E2 polypeptides are provided from expression vectors. Using the same conditions, a BPV *ori* fragment will replicate also in CV1 and COS7 cells (data not shown). However, complementation of the viral genome with mutations in E1 or in E2 with the corresponding expression vector does not result in replication, indicating that expression of both E1 and E2 is host cell restricted (data not shown). These results make an interesting distinction between on the one hand SV40 and polyoma virus which are largely restricted through specific interactions with components of the replication machinery, and the other hand BPV which appears to be restricted largely at the level of gene expression.

Discussion

The results presented in this paper unequivocally define a minimal *ori* fragment which contains all the sequences required for E1 and E2 dependent replication in a short term replication assay. These sequences are required *in cis* for both short term and stable replication since a linker insertion at the *Hpa*I site, which is located in the center of the fragment, abolishes replication in both types of assays in the context of the entire viral genome. This region has also been shown to be essential for initiation of replication *in vitro* (Yang *et al.*, 1991). The sequence of the *ori* fragment shown in Figure 9A shows some recognizable features. The central rergion of the fragment, which includes the *Hpa*I site, contains an inverted repeat sequence that can be arranged in two ways as indicated in the figure (IR1 and 2). On the upstream side of the *Hpa*I site is an A/T rich region. It should also be noted that the *ori* region is located immediately upstream of the P2 promoter, which is one of the major promoters in the BPV early region. The cap site for this promoter is located 89 nucleotides downstream of the *Hpa*I site which makes it likely that the minimal *ori* fragment contains some of the upstream elements of this promoter. In this regard the sequence arrangements appear reminiscent of the *ori* region of SV40 and polyomavirus (Tegtmeyer *et al.*, 1988; DePamphlis, 1988).

In the DNase footprints, E1 protects a sequence on the bottom strand that coincides almost exactly with the inverted repeat IR2. This sequence clearly constitutes a part of the binding site for the E1 polypeptide as shown by the reduction of E1 binding by point mutations in this region, and by methylation interference analysis (our unpublished data). However, sequences outside this inverted repeat region may be important for binding of E1 since cleavage with *Hph*I, which cuts outside the inverted repeat, destroys binding. The extended protection of the top strand by E1 supports this conclusion.

Comparison with sequences of the corresponding region from some other papillomaviruses shows that the region is highly conserved, including the *Hpa*I recognition sequence (Figure 9B). This sequence similarity seems significant since the BPV E1 protein can bind to this sequence in the related virus EEPV (M.Ustav and A.Stenlund, unpublished observation). A further indication that the three viral components E1, E2 and *ori*, which are highly conserved between different viruses serve the same function in replication for all papillomaviruses, is that the E2 polypeptide from the distantly related HPV16 can replace BPV E2 for BPV *ori* replication and thus appears functionally equivalent (M.Ustav and A.Stenlund, unpublished observation).

Based on the SV40 and polyomavirus model sytems, two main functions are provided by the viral *trans*-acting factors. These two functions are a specific *ori* recognition which targets the replication complex, including the cellular

replication factors, to the origin of replication, and an ATP dependent unwinding activity that serves to unwind the origin to allow initiation of DNA synthesis. The presence of an E1 binding sequence and the lack of a requirement for E2 binding sites at the origin clearly indicates that the origin recognition function resides primarily with E1. Binding of E1 to the *ori* appears necessary, since mutations in the *ori* that affect E1 binding are defective for replication. On the other hand, binding of E1 to the *ori* is not sufficient for replication, since some deletions in the *ori* region that do not affect binding of E1 are replication defective. Apart from its role as an *ori* recognition factor, E1 is likely to share other functional similarities with SV40 T-antigen. Several regions of homology have been described between E1 and T-antigen and at least some of these appear to have related functions in the two proteins, including the ATP binding domain (Clertant and Seif, 1984; Santucci *et al.*, 1990; Sun *et al.*, 1990).

The role of E2 in replication is still obscure. The lack of a requirement for intact E2 binding sites and the importance of the activation domain indicates a few possible functions. (i) E2 could serve to activate transcriptionally a cellular gene that is required for replication. (ii) The E2 transactivator could serve in its role as transcription factor at the origin, but is brought to the origin by interacting either with E1 or with a cellular factor that associates with the origin. (iii) The E2 polypeptide has dual functions and the correlation with transcriptional activity is either fortuitous or, more interestingly, a domain that is normally involved in transcriptional activation is also required for a replication function. For example, a domain that is used for interaction with other proteins could very well be required in two different biochemical functions.

Some data argue against the first two possibilities: for example, the VP16/E2 fusion that retains the site specific transcriptional activation function of E2 (Li *et al.*, 1991) does not support replication of BPV in short term assays. The idea that E2 is brought to the *ori* through interaction with another DNA binding factor is attractive: it has been reported that the E1 and E2 polypeptides can form a specific complex (Mohr *et al.*, 1990; Blitz and Laiminis, 1991). This complex could serve to bring E2 to the origin of replication in the absence of an intact E2 binding site. An apparent paradox still exists, however: although an intact E2 binding site is not required, point mutations in different parts of the C-terminal region of E2 that severely reduce the DNA binding activity of E2 without affecting dimerization are incapable of supporting replication of BPV (P.Szymanski and A.Stenlund, unpublished observation). This indicates that DNA binding activity of E2 is required for replication, and suggests that perhaps E2 in combination with E1 or some other factor has a relaxed or altered sequence specificity. This idea is supported by our recent mutational data which indicates that *ori* function requires retention of an intact half of the palindromic E2 BS12 (M.Ustav and A.Stenlund, unpublished observation).

The apparent lack of host cell specificity for replication of the BPV *ori* is very interesting in light of the mechanism that determines host range for the two other members of the papovavirus group. For both SV40 and polyomavirus the host cell specificity appears to reside in the interaction between the viral T-antigen and the host DNA polymerase α−primase that is required to initiate DNA replication (Murakami *et al.*, 1986a,b). One possibility is that the mechanism of initiation of replication for BPV could be fundamentally different from that of SV40 and polyomavirus. Alternatively, if BPV uses a similar mechanism for initiation of DNA synthesis, E1 or E2 or a combination of the two could interact with a different, more highly conserved region of the polymerase−primase complex.

As mentioned in the Introduction, various procedures have been used by several different groups to determine which *cis*-acting elements are required for replication and where initiation of replication occurs in the BPV genome. Interestingly, these studies show little consistency. The *cis*-acting sequences identified in this paper are completely unrelated to the initially defined PMS elements (Lusky and Botchan, 1984, 1986) by several criteria. (i) The PMS sequences can be deleted from the viral genome without adverse effects on replication in our short term replication assay. (ii) Fragments containing the PMS1 or 2 regions but lacking the minimal *Hpa*I region are devoid of replication activity in our short term replication assay. (iii) Mutants in the PMS1 region that were assayed in the context of a larger fragment and were determined to be replication defective in a transient replication assay (Lusky and Botchan, 1986), are in our hands replication competent.

The minimal *ori* fragment that we have defined also differs from another region that recently has been suggested to be important for *ori* function. Based on the association between E1 and E2, it has been suggested that a cluster of strong E2 binding sites in the E2 enhancer constituted the origin of replication (Mohr *et al.*, 1990). This hypothesis is inconsistent with our data since the cluster of E2 binding sites that the complex bound to lacks replication activity in our assay. Furthermore, deletions of this region do not appear to have deleterious effects on replication of *ori* fragments (see Figure 3).

Some of the conflicting results might be explained by differences in the assays that have been used. One possibility would be that short term and stable replication requires different *cis*-acting elements. However, this does not explain all the discrepancies, since the PMS1 element, which was initially shown to be active in stable assays, was subsequently shown to be active also in short term replication assays (Lusky and Botchan, 1986). In addition, the fact that a linker insertion at the *Hpa*I site in our hands renders the viral DNA replication defective in both stable and short term replication assays argues that some elements that reside well outside the PMS regions are required in both types of assay. We are at present trying to resolve some of these questions by determining if the minimal *ori* fragment can be maintained as a stably replicating plasmid.

Finally, in addition to the minimal *ori*, other sequences in the viral genome have effects on replication even though they are not absolutely required. For example, fragments that contain the *ori* but lack the E2 enhancer region are unable to compete efficiently with *ori* fragments that include the E2 enhancer, indicating an auxiliary role in replication for the E2 enhancer (M.Ustav and A.Stenlund, unpublished observation).

Materials and methods

Plasmid constructs and mutants

The BPV genome was used in the form of pMLBPV (Lusky and Botchan, 1985) which consists of the entire viral genome linearized at the unique *Bam*HI site and inserted into the *Bam*HI site of the plasmid pML. The linker insertion mutant *Hpa*I−*Xho*I was generated by insertion of a commercially

available *Xho*I linker into the unique *Hpa*I site of the BPV genome. The deletion mutants in the URR were constructed from a series of *Bam*HI linker insertion mutants described in Lusky and Botchan (1986). Briefly, the mutants were digested with *Hin*dIII and *Hpa*I which cleave on either side of the URR and *Bam*HI which cleaves at the linker insertion between these two sites. The *Bam*HI−*Hpa*I and the *Hin*dIII−*Bam*HI fragments from different mutants were then combined to generate the desired deletions. This deleted *Hin*dIII−*Hpa*I fragment was then used to replace the *Hin*dIII−*Hpa*I fragment of the wild-type genome.

The *ori* constructs pUC/Alu and pUC/Msp were generated by isolating the *Alu*I fragment (nucleotides 7891−52) and *Mps*I fragment (nucleotides 7903−81) and cloning these into the *Hin*cII site of pUC19. Deletions and point mutations in the minimal *ori* were geneated by PCR, using standard protocols.

The E1 and E2 expression vectors pCGEag and pCGE2 have been described previously (Ustav and Stenlund, 1991). Deletions in E2: the deletion of the N-terminal 15 amino acids was constructed by deleting the authentic ATG codon in E2 by digestion with pCGE2 with *Xba*I and *Sph*I. Initiation of translation now takes place at the first downstream ATG codon which is located at amino acid 16. The internal deletion between amino acids 91 and 162 was constructed by deletion of an internal *Nco*I fragment (nucleotides 2878−3089) followed by fill-in of the cohesive ends and ligation which generates an in-frame fusion. The deletion of the C-terminal 33 amino acids was generated by deleting an internal *Bcl*I fragment (nucleotides 3737−3838) followed by religation. In this construct the E2 coding sequence terminates after amino acids 378. All E2 expression vectors were tested for their ability to express the predicted polypeptide by transient expression in COS cells followed by *in vivo* labeling with [^{35}S]methionine and immunoprecipitation with a polyclonal antiserum specific for E2. Transactivation assays were carried out by co-transfection of the respective E2 expression vectors with the viral genome into C127 cells followed by assay of specific RNA levels by RNase protection.

Expression of E1 in E.coli

The E1 expression vector pETE1 was generated by PCR amplification of the E1 coding sequence using an upstream primer containing an *Nde*I restriction site and a downstream primer containing a *Bam*HI site. After digestion of the PCR product with *Nde*I and *Bam*HI the resulting fragment was cloned into the expression vector pET11c (Studier *et al.*, 1990). Growth and induction was carried out as described by Studier *et al.* (1990) with the exception that induction was done at room temperature (M.Tanaka, personal communication). The bacteria were harvested and resuspended in LS buffer (50 mM Tris−HCl, pH 7.5, 0.1 M NaCl, 2 mM EDTA, 0.5% NP40, 20% sucrose and 50 μg/ml PMSF) and treated with 100 μg/ml lysozyme for 10 min on ice. After sonication, insoluble material was pelleted by centrifugation at 10 000 *g* for 10 min and the pellet was re-extracted with HS buffer (50 mM Tris−HCl, pH 7.5, 1 M NaCl, 2 mM EDTA, 0.5% NP40, 20% sucrose and 50 μg/ml PMSF).

Generation of E1 beads and DNA precipitation assays

E1 beads were generated by first coupling the monoclonal antibody 12CA5 (Field *et al.*, 1988) to protein A Sepharose, followed by crosslinking with dimethylpimelimidate as described (Harlow and Lane, 1988). These beads were then incubated with the crude *E.coli* lysate for 2 h at 4°C in LS buffer. After five washes the beads were incubated with labeled DNA fragments in LS buffer and agitated for 1 h at room temperature. The beads were washed five times in LS buffer, resuspended in STOP buffer (0.2 M NaCl, 20 mM EDTA, 1% SDS, 250 μg/ml tRNA) and extracted with phenol/chloroform. After precipitation with ethanol the samples were analyzed on denaturing 5% acrylamide gels.

DNase footprints

DNase footprint analysis was carried out as the DNA precipitation assays, except that after the last wash, the beads were resuspended in 25 μl of a solution containing 20 mM HEPES, pH 7.5, 100 mM KCl, 1 mM EDTA, 0.1% NP40 and 1 μg of carrier DNA. 25 μl of a solution containing 10 mM $MgCl_2$ and 5 mM $CaCl_2$ was added and the sample was treated with DNase I for 1 min at room temperature. The reaction was stopped by addition of 90 μl STOP buffer (see above) and the sample was extracted with phenol/chloroform and precipitated with ethanol.

Gel retardation assays were performed as described in Li *et al.* (1989).

Immunoblots

Protein was transferred from SDS−PAGE gels using semi-dry blotting. The blots were blocked with 3% BSA and the monoclonal antibody 12CA5 was added at a 1:5000 dilution. After addition of the secondary antibody (alkaline phosphatase conjugated goat anti-mouse) the blot was processed and developed as described in Harlow and Lane (1988).

Cell culture

C127, CV1, COS7, HeLa and 293 cells were all maintained in DME + 10% fetal bovine serum. Replication assays in C127 cells were carried out as described (Ustav and Stenlund, 1991).

Acknowledgements

We thank Bruce Stillman for valuable comments and Masfumi Tanaka for helpful discussions throughout this work. We also thank Rong Li and M.Botchan for the VP16/E2 construct. M.U. is a recipient of an Andrew Seligson memorial fellowship. P.S. is supported by an NIH postdoctoral fellowship. This work was supported by a grant from the National Institutes of Health CA 13106-19.

References

Androphy,E.J., Lowy,D.R. and Schiller,J.T. (1987) *Nature*, **325**, 70−73.
Black,P.H., Hartley,J.W., Rowe,W.P. and Huebner,R.J. (1963) *Nature*, **199**, 1016−1018.
Blitz,I.L. and Laiminis,L.A. (1991) *J. Virol.*, **65**, 649−656.
Borowiec,J.A., Dean,F.D., Bullock,P.A. and Hurwitz,J. (1990) *Cell*, **60**, 181−184.
Challberg,M.D. and Kelly,T.J. (1989) *Annu. Rev. Biochem.*, **58**, 671−717.
Clertant,P. and Seif,I. (1984) *Nature*, **311**, 276−279.
DePamphlis,M.L. (1988) *Cell*, **52**, 635−638.
Field,J., Nikawa,J., Broek,D., MacDonald,B., Rodgers,L., Wilson,I.A., Lerner,R.A. and Wigler,M. (1988) *Mol. Cell. Biol.*, **8**, 2159−2165.
Giri,I. and Yaniv,M. (1988) *EMBO J.*, **7**, 2823−2829.
Harlow,E. and Lane,D. (1988) *Antibodies, A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Lambert,P.F., Baker,C.C. and Howley,P.M. (1988) *Annu. Rev. Genet.*, **22**, 235−258.
Li,R., Knight,J., Bream,G., Stenlund,A. and Botchan,M. (1989) *Genes Dev.*, **3**, 510−526.
Li,R., Knight,J., Jackson,S.P., Tjian,R. and Botchan,M. (1991) *Cell*, **65**, 493−505.
Lusky,M. and Botchan,M. (1984) *Cell*, **36**, 391−401.
Lusky,M. and Botchan,M. (1985) *J. Virol.*, **53**, 955−965.
Lusky,M. and Botchan,M. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 3609−3613.
McBride,A.A., Schlegel,R. and Howley,P.M. (1988) *EMBO J.*, **7**, 533−539.
McBride,A.A., Byrne,J.C. and Howley,P.M. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 510−514.
Mecsas,J. and Sugden,B. (1987) *Annu. Rev. Cell. Biol.*, **3**, 87−108.
Meischke,H.R.C. (1979) *J. Gen. Virol.*, **43**, 472−487.
Mohr,I.J., Clark,R., Sun,S., Androphy,E.J., MacPherson,P. and Botchan,M. (1990) *Science*, **250**, 1694−1699.
Moskaluk,K. and Bastia,D. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 1215−1218.
Murakami,Y., Eki,T., Yamada,M., Prives,C. and Hurwitz,J. (1986a) *Proc. Natl. Acad. Sci. USA*, **83**, 6347−6351.
Murakami,Y., Wobbe,C.R., Weissbach,L., Dean,F.B. and Hurwitz,J. (1986b) *Proc. Natl. Acad. Sci. USA*, **83**, 2869−2873.
Peden,K.W.C., Pipas,J.M., Pearson-White,S. and Nathans,D. (1980) *Science*, **209**, 1391−1396.
Roberts,J.M. and Weintraub,H. (1986) *Cell*, **46**, 741−752.
Santucci,S., Androphy,E.J., Bonne-Andrea,C. and Clertant,P. (1990) *J. Virol.*, **64**, 6027−6039.
Spalholz,B.A., Yang,Y. and Howley,P. (1985) *Cell*, **42**, 183−191.
Stillman,B. (1989) *Annu. Rev. Cell. Biol.*, **5**, 197−245.
Studier,F.W., Rosenberg,A.H., Dunn,J.J. and Dubendorff,J.W. (1990) *Methods Enzymol.*, **185**, 60−89.
Sun,S., Thorner,L., Lentz,M., MacPherson,P. and Botchan,M. (1990) *J. Virol.*, **64**, 5093−5105.
Tegtmeyer,P., Deb,S., DeLucia,A.L., Deb,S.P., Koff,A., Tsui,S., Parsons,R., Partin,K., Bour,C.P., Dean,F.D. and Hurwitz,J. (1988) *Cancer Cells 6*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Ustav,M. and Stenlund,A. (1991) *EMBO J.*, **10**, 449−457.
Waldeck,S., Rosl,F. and Zentgraf,H. (1984) *EMBO J.*, **3**, 2173−2178.
Wilson,V. and Ludes-Meyer,J. (1991) *J. Virol.*, **65**, 5314−5322.
Yang,L. and Botchan,M. (1990) *J. Virol.*, **64**, 5903.
Yang,L., Li,R., Mohr,I.J., Clark,R. and Botchan,M. (1991) *Nature*, **353**, 628−632.

Received on August 20, 1991; revised on October 2, 1991