JOURNAL OF VIROLOGY, Feb. 1994, p. 1094–1102
0022-538X/94/$04.00+0

Vol. 68, No. 2

DNA Binding Specificity of the Bovine Papillomavirus E1 Protein Is Determined by Sequences Contained within an 18-Base-Pair Inverted Repeat Element at the Origin of Replication

SHAWN E. HOLT, GENEVIEVE SCHULLER, AND VAN G. WILSON*

Department of Medical Microbiology and Immunology, Texas A&M University Health Science Center, College Station, Texas 77843

Received 20 May 1993/Accepted 15 November 1993

Bovine papillomavirus type 1 (BPV-1) DNA replicates episomally and requires two virally expressed proteins, E1 and E2, for this process. Both proteins bind to the BPV-1 genome in the region that functions as the origin of replication. The binding sequences for the E2 protein have been characterized previously, but little is known about critical sequence requirements for E1 binding. Using a bacterially expressed E1 fusion protein, we examined binding of the BPV-1 E1 protein to the origin region. E1 strongly protected a 28-bp segment of the origin (nucleotides 7932 to 15) from both DNase I and exonuclease III digestion. Additional exonuclease III protection was observed beyond the core region on both the 5′ and 3′ sides, suggesting that E1 interacted with more distal sequences as well. Within the 28-bp protected core, there were two overlapping imperfect inverted repeats (IR), one of 27 bp and one of 18 bp. We show that sequences within the smaller, 18-bp IR element were sufficient for specific recognition of DNA by E1 and that additional BPV-1 sequences beyond the 18-bp IR element did not significantly increase origin binding by E1 protein. While the 18-bp IR element contained sequences sufficient for specific binding by E1, E1 did not form a stable complex with just the isolated 18-bp element. Formation of a detectable E1-DNA complex required that the 18-bp IR be flanked by additional DNA sequences. Furthermore, binding of E1 to DNA containing the 18-bp IR increased as a function of overall increasing fragment length. We conclude that E1-DNA interactions outside the boundaries of the 18-bp IR are important for thermodynamic stabilization of the E1-DNA complex. However, since the flanking sequences need not be derived from BPV-1, these distal E1-DNA interactions are not sequence specific. Comparison of the 18-bp IR from BPV-1 with the corresponding region from other papillomaviruses revealed a symmetric conserved consensus sequence, T-RY--TTAA--RY-A, that may reflect the specific nucleotides critical for E1-DNA recognition.

Bovine papillomavirus type 1 (BPV-1) is a member of the papovavirus family and is the prototype virus for the papillomavirus subgroup. The BPV-1 genome can be taken from its natural host and will replicate, in both stable and transient replication assays (13–15, 27), in a variety of animal cell lines including C127 and NIH 3T3 cells (8, 12, 14, 27, 32). The 7,945-bp viral genomic DNA is preserved in these cells as an extrachromosomal entity at a constant copy number of 50 to 200 molecules per cell (8). Because of its limited genomic coding capacity, BPV-1 requires cellular host DNA replication factors and enzymes for its replication. This dependence on the host cell replication machinery makes BPV-1 a useful model for general studies of cellular replication mechanisms and control. However, a necessary first step will be the careful definition of viral *cis* and *trans* replication elements.

Two viral proteins are unequivocally required for transient replication of the viral DNA, the full-length E1 open reading frame protein and the full-length E2 open reading frame protein, E2TA (27). E1 was first shown to be a site-specific DNA-binding protein with a bacterially expressed E1-related fusion protein termed RecA-E1 (30). RecA-E1 binds specifically to a 219-bp BPV-1 fragment spanning nucleotides 7819 to 93. Cleavage of this fragment at the unique *Hpa*I site eliminates E1 binding, indicating that the E1 binding site is located at or near the *Hpa*I site (30). Subsequent studies showed that E1 protein expressed in either prokaryotic or eukaryotic cells protects the *Hpa*I region from DNase I digestion (28, 33). In addition, small deletions or insertions at the *Hpa*I site greatly reduced E1 binding (23, 34). Together, these studies establish that there is a prominent E1 binding site in the vicinity of the unique *Hpa*I site, but they do not specifically define critical sequences for binding.

Recently, the approximate boundaries of the functional BPV-1 origin of replication were defined both in vivo and in vitro (see Fig. 2) (28, 33). The functional origin encompasses the E1 binding region, which is consistent with a direct role for E1 in replication initiation via binding to origin DNA sequences. In addition to an E1 binding site, the BPV-1 origin of replication also contains an AT-rich region and an E2 DNA binding site (10, 26, 28). Ustav et al. have shown that an E2 binding site is strictly required for BPV-1 replication in vivo (26). The 48-kDa E2TA protein is a well-characterized DNA binding protein that is the major viral transcriptional activator (22, 24); however, the E2 transactivation function is not required for its function in DNA replication (1, 32). The role of E2 in replication appears to be related to its ability to form protein-protein complexes with E1 (2, 16, 17) and to enhance the binding of E1 to DNA (20, 33). Presumably, E1-E2 protein-protein interactions, along with E1-DNA interactions, are important for formation of the E1 replication complex at the origin.

The E1 protein shares both sequence and functional homology with simian virus 40 (SV40) large T antigen (4, 19); both

* Corresponding author. Mailing address: Department of Medical Microbiology and Immunology, Texas A&M University Health Science Center, College Station, TX 77843. Phone: (409) 845-5207. Fax: (409) 845-3479.

proteins are absolutely required for replication of their respective genomes, are located predominantly in the nucleus (6, 9), are origin-binding proteins (6, 21, 28, 30), have DNA helicase (6, 21) and origin-unwinding activity (6, 20, 21), and have DNA-dependent ATPase activity (21). T antigen binds a 27-bp palindromic sequence at the SV40 origin and assembles into a multimeric complex that orchestrates the initial unwinding of the SV40 origin (5, 6). Given the similarities between T antigen and E1, it is likely that E1 also functions in a complex fashion to promote initiation of BPV-1 replication. In this paper, we show that E1 binding specificity is contained within an 18-bp inverted repeat (IR) element that includes the unique *Hpa*I site at the origin of replication. However, while specificity resides within the 18-bp IR, formation of a stable E1-DNA complex requires additional nonspecific flanking sequences.

MATERIALS AND METHODS

Exonuclease III protection assay. RecA-E1 was immunoprecipitated with anti-$E1_{249}$ serum (31) and protein A-Sepharose (Pharmacia) as previously described (30). The washed immunocomplex was resuspended in 20 μl of 0.05 M TNE (10 mM Tris-HCl, 50 mM NaCl, 0.1 mM EDTA, pH 7.0) containing 50 ng of a 214-bp single end-labeled DNA. This double-stranded DNA substrate was produced by PCR amplification of pd-BPV-1 DNA with primers whose 5′ ends were at BPV-1 nucleotides 7830 and 99, respectively. DNA was labeled on the 5′ end of either the upper or the lower strand by phosphorylating the appropriate primer with [γ-^{32}P]ATP prior to the PCR. The labeled DNA was incubated with the immunocomplex for 30 min at 25°C, and then the complex was washed twice with exonuclease III buffer (66 mM Tris-HCl, 0.66 mM $MgCl_2$, 1 mM 2-mercaptoethanol, pH 7.6), resuspended in 10 μl of exonuclease III buffer containing from 0 to 10 U of exonuclease III (Boehringer Mannheim), and incubated for 5 min at 37°C. Control extracts lacking RecA-E1 were similarly immunoprecipitated and washed except that there was no incubation step with labeled DNA. For the control samples, the immunoprecipitates were supplemented with an amount of labeled DNA comparable to that bound in the RecA-E1 immunoprecipitates. For both RecA-E1 and control samples, exonuclease III digestion was stopped by addition of 4 μl of sequencing sample buffer (95% [vol/vol] formamide, 10 mM EDTA, 0.1% bromophenol blue, 0.1% xylene cyanol) followed immediately by 5 min at 95°C. Samples were analyzed directly on standard 6% sequencing gels.

Purification and labeling of oligos. Synthetic oligonucleotides (oligos) corresponding to the upper and lower strands of the two IR sequences in the BPV-1 origin of replication were constructed (see Fig. 3): 5′ATTGTTGTTAACAATAAT3′ (18-mer upper) and 5′ATTATTGTTAACAACAAT3′ (18-mer lower), and 5′CAGTGAATAATTGTTGTTAACAATA ATCACAC3′ (32-mer upper) and 5′GTGTGATTATTGTT AACAACAATTATTCACTG3′ (32-mer lower). Oligos were electrophoresed on a denaturing Tris-borate 12% polyacrylamide gel and were purified with the MERmaid kit (Bio 101). Purified oligos were quantitated by spectrophotometry.

Each oligo (100 ng) was end labeled with T4 polynucleotide kinase, kinase buffer (60 mM Tris-HCl [pH 7.8], 10 mM $MgCl_2$, 200 mM KCl), and [γ-^{32}P]ATP at 37°C for 1 h. This reaction was chased with 50 μM cold ATP for 30 min to ensure labeling of all oligo ends. Complementary labeled oligos (50 ng) were incubated at 68°C for 10 min and then slowly cooled to 4°C for 10 to 12 h to give efficient annealing. Annealed oligos (1 to 2 ng) were ligated with T4 DNA ligase at 4°C for 12 to 16 h. Labeled oligos (0.1 to 0.4 ng) were used in the DNA binding assay (below) in either single-stranded, double-stranded, or ligated forms.

Cloning oligos. The double-stranded 18-mer sequence was cloned into the *Sma*I site of the pUC18 polylinker. Potential positive clones were sequenced with the dsDNA Cycle Sequencing System as recommended by the manufacturer (BRL Life Technologies, Inc.). A positive clone containing a single copy of the double-stranded 18-mer sequence was designated pUC18/18-mer. The double-stranded 32-mer was cloned in the same manner and named pUC32-mer. A 105-bp *Alu*I fragment of BPV-1 DNA (nucleotides 7892 to 52) was cloned into the *Sma*I site of pUC18 and designated ORI-105.

Digestion and labeling of cloned oligos. The pUC18/18-mer clone was PCR amplified by standard methods to generate a 169-bp fragment. This 169-bp segment (0.5 to 1.0 μg) was digested with four combinations of restriction enzymes: *Eco*RI and *Sal*I, *Eco*RI and *Bam*HI, *Sal*I and *Acc*65I, or *Bam*HI and *Acc*65I. The 5′ overhangs of these DNA pieces were made blunt and radiolabeled by incubation at room temperature for 10 to 20 min with 1 U of Sequenase 2.0 (United States Biochemical Corporation); 7.5 μM (each) dATP, dCTP, and dGTP; 0.3 μM [α-^{32}P]TTP (ICN, Inc.); 40 mM Tris-HCl (pH 7.5); 20 mM $MgCl_2$; and 50 mM NaCl. The reaction was terminated by being heated to 70°C for 10 min. Each labeled fragment (10 to 20 ng) was used in the DNA binding assay described below.

For some experiments, the pUC clones containing the 18-mer sequence, the 32-mer sequence, or the ORI-105 sequence of BPV-1 were generated by incorporation of labeled [α-^{32}P]TTP during PCR amplification of the appropriate clones. The resultant fragments (256, 270, and 256 bp, respectively) had the same specific activity and were used directly for DNA binding assays.

DNA binding assays. Bacterial extracts with and without the RecA-E1 protein were prepared as previously described (30). DNA binding assays with the oligos were performed as previously described (30) with some modifications. For ligated oligomers, labeled DNA was incubated with the bacterial extract and sheared salmon sperm DNA in 10 mM TNE (10 mM Tris HCl [pH 7.0], 10 mM NaCl, 0.01 mM EDTA). After incubation, E1-DNA complexes were immunoprecipitated and washed as before. The washed immunoprecipitates were incubated with 10 μl of TBE sample buffer (89 mM Tris, 89 mM boric acid, 2.5 mM EDTA, 10% glycerol, 2.7% xylene cyanol, 2.4% bromophenol blue, 1.5% sodium dodecyl sulfate) for 1 h at 37°C to extract bound labeled DNA. Supernatants were electrophoresed on nondenaturing Tris-borate 15% polyacrylamide gels. Gels were dried and exposed to X-ray film for autoradiography with an intensifier screen.

The DNA binding assay for single-stranded oligos, monomeric double-stranded oligos, and PCR products of pUC18/18-mer, pUC32-mer, and ORI-105 was altered slightly. Briefly, 10 μl of bacterial extract was immunoprecipitated (without any DNA) with anti-E1 antibodies as described previously (30). These samples were then incubated at 25°C for 30 min with 1,500 ng of unlabeled sheared salmon sperm DNA, 150 mM TNE (10 mM Tris HCl [pH 7.0], 150 mM NaCl, 0.01 mM EDTA), and labeled cloned oligo fragments (quantities given above). The samples were washed three times with 1 ml of 10 mM TNE supplemented with 200 mM NaCl, 0.25% Nonidet P-40, and 5 μg of sheared salmon sperm DNA. A final wash was done with 1 ml of 10 mM TNE. Washed pellets were incubated with 15 or 60 μg of proteinase K for 30 min at 55°C, and then the labeled DNA was extracted with TBE sample buffer as described above. The extracted DNA was electrophoresed on nondenaturing Tris-borate 8, 10, or 15% poly-

acrylamide gels. Gels were dried and autoradiographed as described above. The relative amounts of bound DNA fragments were quantitated by densitometric analysis of the autoradiographs with the IS-1000 Digital Imaging System (Innotech Scientific Corp.).

Construction of deletion mutants. Purified pdBPV-1 (ATCC 37134) DNA was linearized at the unique *Hpa*I site and gel purified. Linearized DNA (5 μg) was treated with 0.5 U of *Bal* 31 (International Biotechnologies, Inc.) at 30°C. At 30-s intervals after addition of *Bal* 31, 1-μg samples were removed and *Bal* 31 digestion was terminated by addition of 1 volume of 0.2 M EDTA followed by phenol-chloroform extraction. Each 1-μg sample of *Bal* 31-treated DNA was concentrated with the Prep-A-Gene system (Bio-Rad) and self-ligated for 30 min at 16°C with 0.5 U of T4 DNA ligase (Boehringer Mannheim). After 30 min of ligation, the reactions were digested for 30 min at 37°C with 2 U of *Hpa*I (Boehringer Mannheim). The resultant DNA was used to transform *Escherichia coli* TB1 cells, and clones were screened for the absence of a *Hpa*I site. *Hpa*I site-minus clones were sequenced directly from colonies with the dsDNA Cycle Sequencing System (GIBCO BRL). A deletion mutant lacking 8 bp to the 5′ side of the *Hpa*I cleavage site was designated Δ5′, while a similar mutant lacking 8 bp on the 3′ side of the *Hpa*I site was designated Δ3′. Binding studies with the deletion mutant DNA were performed as previously described (30).

RESULTS

E1 protein protects an extended region of BPV-1 DNA from nucleases. Our previous study demonstrated that a bacterially expressed E1 protein (RecA-E1) is a site-specific DNA-binding protein. Strong binding of RecA-E1 to the BPV-1 genome was observed only in the vicinity of the unique *Hpa*I site on the BPV-1 genome (30). The *Hpa*I region is now known to be the functional origin of replication (28, 33), suggesting that the interaction of E1 with sequences in this region will be critical for the replication process. Subsequent DNase I footprinting with the RecA-E1 protein showed a clearly delimited 28-bp protected region from nucleotides 7932 to 15 on the lower strand, a region which encompassed the *Hpa*I site (29). Protection of this region was also observed on the upper strand, though the protection was less pronounced and extended further in both the 5′ and 3′ directions. Similar extended DNase I protection has been reported previously, confirming the general location of the E1 binding site on the BPV-1 genome (28, 33).

Since the boundaries of E1 protection on the upper strand were difficult to define precisely by DNase I footprinting, we performed exonuclease III footprinting of RecA-E1 bound to an origin-containing fragment (Fig. 1). The exonuclease III experiments were performed under the same conditions and at the same protein/DNA ratio used for McKay assays (30) (see Fig. 4 to 7) and presumably reflect similar E1-DNA interactions. On both the upper and lower strands, there were multiple clusters of exonuclease III stop sites, with some stop sites being more prominent than others. Only the exonuclease III stops that were consistently observed are marked in Fig. 1. The most distal stop sites were at nucleotide positions 25 to 28 on the upper strand and 7900 to 7903 on the lower strand. The boundaries of the protected region defined by the most distal stop sites spanned 74 bp and encompassed the region seen protected by DNase I. However, the most prominent lower- and upper-strand exonuclease III stops (positions 7932/33 and 15) corresponded exactly to the 28-bp DNase I-protected region seen on the lower strand (Fig. 2). This concordance

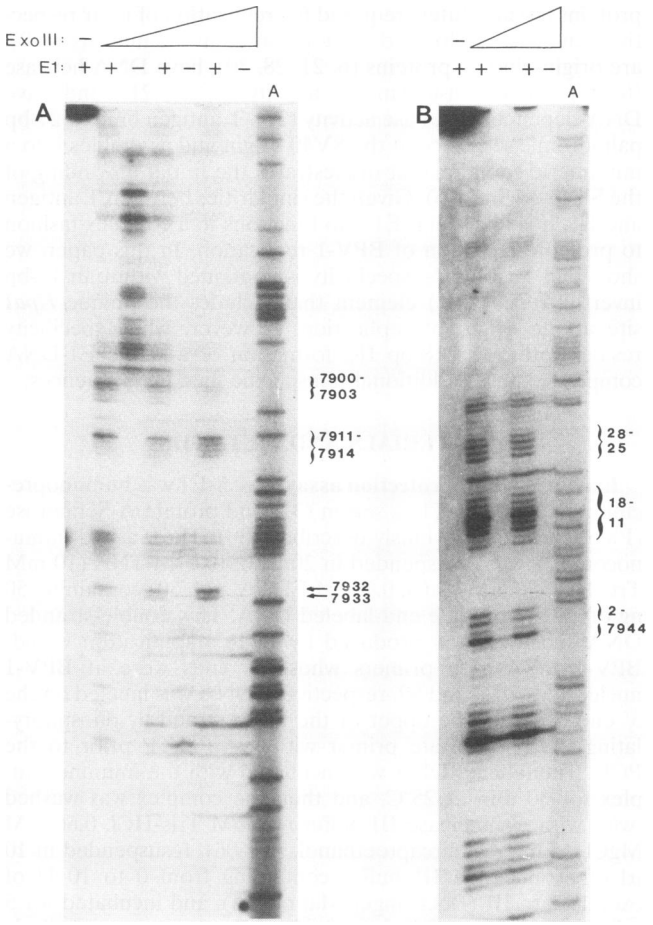

FIG. 1. Exonuclease III footprints of RecA-E1 bound to the BPV-1 origin region. Exonuclease III footprinting was performed on immunoprecipitated E1-DNA complexes as described in Materials and Methods. Footprinting reactions contained DNA labeled on the 5′ end of the lower strand (A) or the upper strand (B). Lanes without E1 contained immunoprecipitates of control extracts which were supplemented after precipitation with amounts of free labeled DNA comparable to the E1 precipitates. Lanes marked A contained an A sequencing reaction of the DNA used as the E1 binding substrate. Exonuclease stop sites that were reproducibly present in E1 bound samples are marked with braces or arrows, and the BPV-1 nucleotide numbers of the stop positions are given.

between the DNase I and exonuclease III footprints suggested that the primary binding site for E1 was located between positions 7932 and 15 on the genome map. The more distal exonuclease III stop sites were consistent with additional E1-DNA interactions that extended beyond the primary binding site at the protein/DNA ratio employed for these experiments. These more distal exonuclease III stop sites were asymmetric with respect to the central protected region; exonuclease III stops extended further to the 5′ side than to the 3′ side. The significance of this asymmetry is unknown.

As previously noted (28), the nucleotide sequence in the central E1-protected region (nucleotides 7932 to 15) contains alternative imperfect IR elements (Fig. 3). The smaller element is an 18-bp IR from nucleotides 7940 to 12, with a dyad axis between nucleotides 4 and 5. There is a single mismatched pair in this IR with a G at nucleotide 7943 in the 5′ half and a corresponding T at nucleotide 9 in the 3′ half. The larger IR

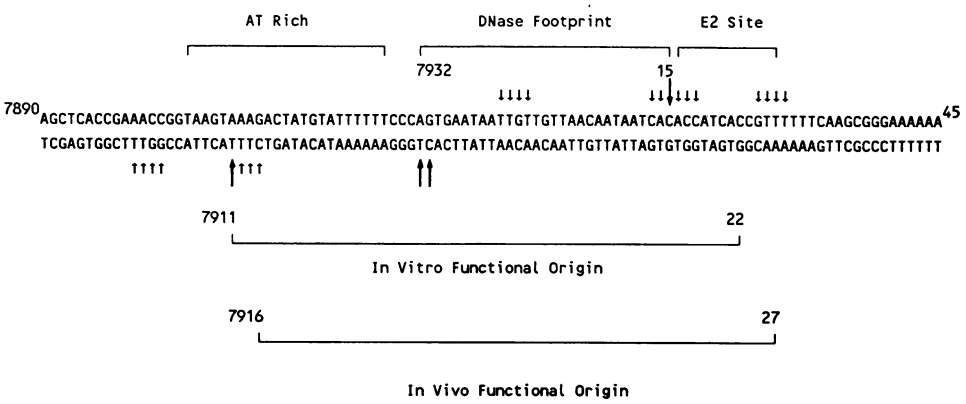

FIG. 2. Summary of DNase I and exonuclease III footprinting results. Shown are the upper and lower strands of the BPV-1 genome from nucleotides 7890 to 45. Brackets above the sequence indicate the positions of an AT-rich region, the region on the lower strand protected from DNase I digestion by RecA-E1, and a binding site for E2 protein. Brackets below the sequence indicate the location of DNA fragments that function as an origin of replication in vitro (33) and in vivo (28). Arrows indicate the positions of exonuclease III stop sites on the upper and lower strands. Larger arrows indicate the predominant stop sites.

element has a 13-bp 5′ segment and a 12-bp 3′ segment; the 5′ and 3′ portions are separated by a 3-bp sequence (GTT). As many DNA-binding proteins recognize sequences with dyad symmetry, the presence of the alternative IR elements suggested that one or both might be important for E1 binding. However, since both IRs were overlapping and entirely contained within the predominant E1-protected region, the significance of either IR for E1 binding specificity could not be deduced from the footprinting data.

E1 binding specificity resides within an 18-bp IR element. To further describe RecA-E1 binding site sequence requirements, two sets of oligos which corresponded to the two alternative IR elements within the RecA-E1-protected region were constructed. A set of 32-base oligos (upper and lower strands) was composed of BPV-1 sequences 7931 through 17, and contained the A′/B′ IR shown in Fig. 3. The second set consisted of 18-base oligos spanning the A/B IR. Each oligo was end labeled and used in single-stranded, double-stranded, and ligated forms as a substrate for RecA-E1 binding in the immunoprecipitation assay (Fig. 4). None of the single-stranded oligos was bound by RecA-E1 in this assay (data not shown). In addition, the monomeric, double-stranded forms of the 18-mer and 32-mer oligos were not efficiently bound by RecA-E1 under these conditions; no binding to the 18-mer was observed, and only minimal binding to the 32-mer was detected (Fig. 4A, lane 4; Fig. 4B, lane 2). However, the RecA-E1 protein effectively bound the ligated forms of both the 18-mer and the 32-mer (Fig. 4A, lane 6; Fig. 4B, lane 5). No binding to these ligated oligos was detected with extracts lacking RecA-E1 ("IP*" lanes) or when the immunoprecipitations were performed with preimmune serum ("pI" lanes), confirming that the precipitation of these ligated oligos was E1 dependent. Furthermore, binding to the ligated BPV-1 oligos was sequence specific as no E1-specific binding was detected to either of two control oligos (an 8-bp *Hin*dIII linker or an 18-bp AT-rich oligo) in monomeric, double-stranded form (not shown) or after self-ligation (Fig. 4C). Note that the AT-rich oligo consisted of an 18-bp IR sequence with an AT content similar to that of the BPV-1 origin 18-mer and yet still failed to bind E1. Consequently, the observed binding of RecA-E1 to the ligated BPV-1 oligos was not simply a function of nonspecific binding to larger-size or AT-rich DNA, but reflected a true sequence specificity. The ability of RecA-E1 to bind the ligated 18-mer indicated that sequences contained within the 18-bp IR were sufficient to confer specific recognition and binding by E1.

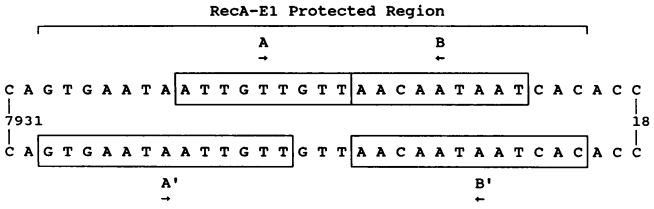

FIG. 3. Alternative IRs in the E1 binding region. The upper strand of the BPV-1 DNA sequence from nucleotides 7931 to 18 is shown twice. Boxes on these sequences indicate the locations of an 18-bp IR (AB) and a 13/12-bp IR (A′B′). The bracket above the sequence indicates the boundaries of the predominant region protected from DNase I digestion by bound RecA-E1.

To further confirm the specificity of the E1 fusion protein for these oligos, a competition assay was performed. Purified pdBPV-1 plasmid DNA was digested with *Ava*II and radiolabeled as previously described (30). Among the fragments generated was a 219-bp fragment that contained the E1 binding site and the BPV-1 origin of replication. Using the same immunoprecipitation conditions described above, bacterial extracts containing RecA-E1 were incubated with the radiolabeled pdBPV-1 fragments and increasing amounts of the unlabeled oligos in the double-stranded monomeric form or in the ligated form. The monomeric forms of each oligo showed little or no competition with the 219-bp fragment for E1 binding (data not shown). However, the ligated forms of both the 18-mer and the 32-mer competed effectively with the labeled 219-bp fragment for binding by RecA-E1 (data not shown). No competition was observed with the monomer or ligated form of the *Hin*dIII linker. The ability of the ligated 18-mer to compete with the authentic E1 binding site confirmed that sequences within the 18-bp IR were sufficient for DNA recognition by RecA-E1 protein.

The failure of the monomeric forms of either the 32-mer or 18-mer to compete with the 219-bp fragment suggested that while specificity was inherent in the 18-bp IR, the affinity of E1

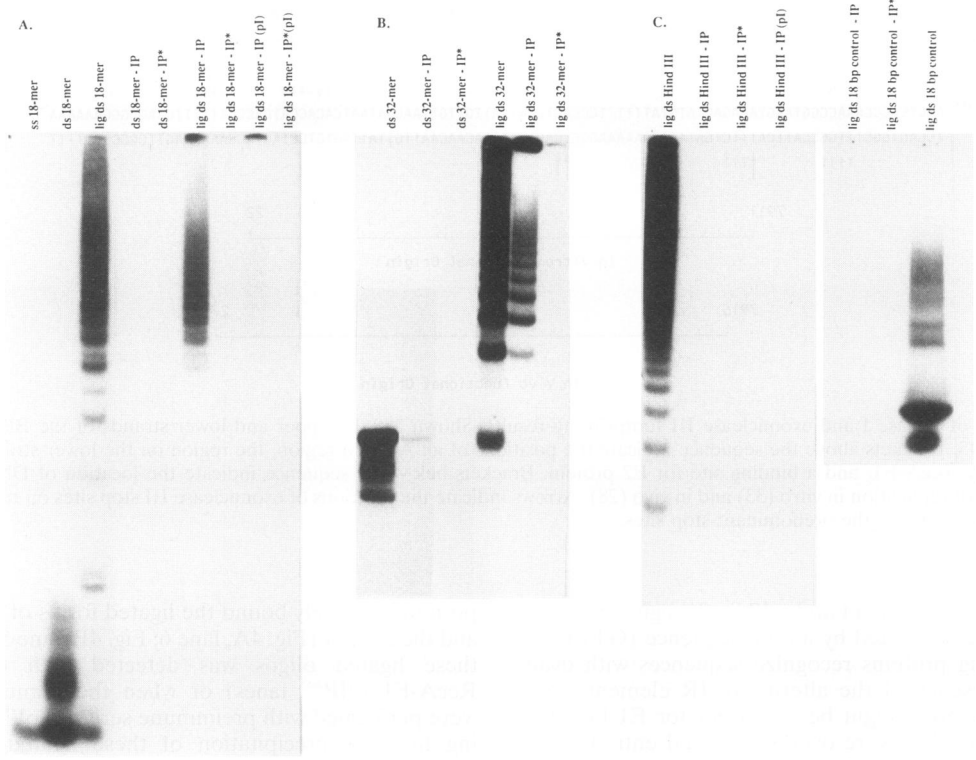

FIG. 4. Binding of RecA-E1 to oligos that correspond to alternative IR elements. The immunoprecipitation binding assay was performed with either the 18-bp BPV-1 IR oligo (A), the 32-bp IR oligo (B), a control *Hin*dIII linker (C), or a control AT-rich oligo (C) as the DNA substrate. Oligos were in either monomeric double-stranded ("ds" lanes) or ligated ("lig ds" lanes) forms as indicated. Precipitations were performed with extracts containing RecA-E1 ("IP" lanes) or control extracts lacking RecA-E1 ("IP*" lanes). All precipitations used anti-E1 serum except lanes marked "pI," which used preimmune serum. For monomeric double-stranded DNA samples, E1 protein was first immunoprecipitated and then incubated with the DNA substrates. Oligomeric substrates were mixed directly with the E1 extracts, and the E1-DNA complexes were coprecipitated. Lanes without the "IP" designation were marker lanes showing the oligos used as substrates for the binding reactions.

binding to short pieces of DNA was poor. One possible explanation for this was that stable binding of E1 required the 18-bp recognition element in the context of additional nonspecific DNA sequence. Presumably, these additional nonspecific sequences would contribute to the thermodynamic stabilization of the E1-DNA complex such as that observed for the bacterial catabolite gene activator protein (11). This also would be consistent with the footprinting results which indicated a significant E1-DNA interaction beyond the boundaries of the 18-bp IR.

To test this hypothesis, both the double-stranded 18-mer and the 32-mer were cloned into pUC18 and the resulting constructs were designated pUC18/18-mer and pUC32-mer, respectively. PCR-amplified pUC18/18-mer or pUC32-mer DNA was digested with four pairs of restriction endonucleases to generate fragments with BPV-1 sequences flanked by various lengths of nonspecific DNA. The digestions also released similar-size plasmid-derived fragments lacking BPV-1 sequence which served as internal specificity controls. The fragment mixtures were radiolabeled with a DNA polymerase fill-in reaction and then tested for binding to RecA-E1 by the immunoprecipitation procedure. The results for the cloned 18-mer are shown in Fig. 5, and similar results were obtained for the cloned 32-mer (not shown). In all cases, the fragments containing the 18-mer sequence were specifically bound by RecA-E1 (Fig. 5A). No binding was detected to plasmid-derived fragments lacking the 18-mer sequences, once again confirming that the 18-bp IR element was sufficient to confer E1 binding specificity. While no binding was observed for the monomeric 18-mer (Fig. 4), binding could be detected to the cloned 18-mer within a fragment of overall length of 31 bp (Fig. 5A). It was also observed that the relative binding by RecA-E1 increased with increasing fragment size (Fig. 5B). The E1 protein showed approximately threefold-greater binding to a 55-bp fragment than to the 31-bp fragment, while the two 43-bp fragments showed intermediate levels of E1 binding. As these four fragments differed only in the length of their non-BPV-1 flanking sequences, this was consistent with our hypothesis that E1-DNA contacts outside the 18-bp element helped stabilize the complex. In addition, comparison of E1 binding with the two 43-bp fragments versus the 31-bp fragment indicated that binding was improved by addition of nonspecific sequences onto either the 5′ or 3′ ends of the 18-bp IR. Together, the binding studies with the oligos and the cloned 18-mer suggest that a sequence length somewhere between 18 and 31 bp is necessary in order to detect binding of RecA-E1 in the immunoprecipitation assay.

Comparison of binding to the 18-mer and 32-mer in the absence of fragment length effects. The studies described above indicated that sequences within the 18-mer were sufficient to direct specific binding of the E1 protein. However, these studies did not rule out the possibility that additional BPV-1 sequences flanking the 18-mer region might contribute to an increased affinity of E1 for BPV-1 DNA. To address this

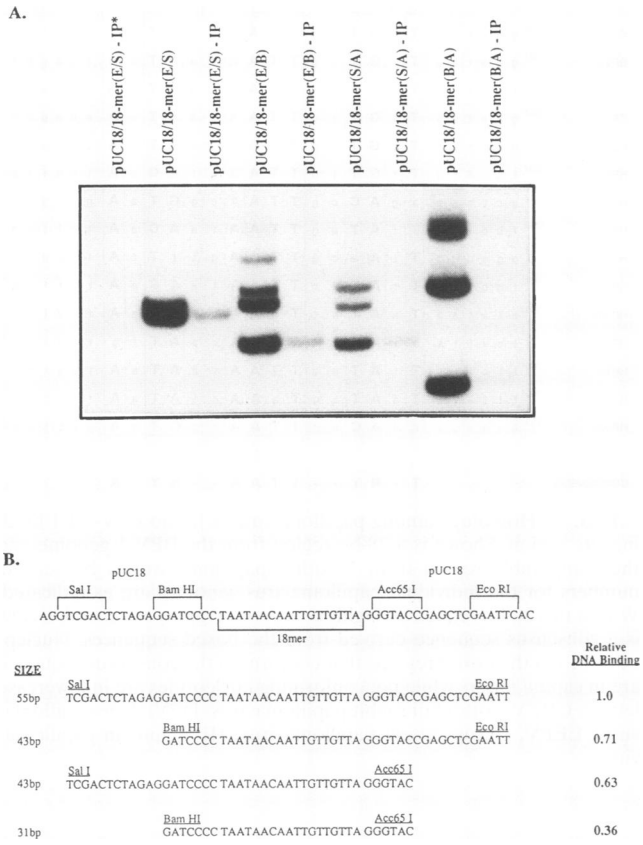

FIG. 5. Binding of the RecA-E1 protein to the cloned 18-bp IR. Panel A shows the binding of RecA-E1 to pUC18/18-mer cleaved with four pairs of restriction endonucleases. The region of pUC18/18-mer containing the cloned BPV-1 18-mer was PCR amplified, and the resultant 169-bp fragment was cleaved with *Eco*RI and *Sal*I (E/S), *Eco*RI and *Bam*HI (E/B), *Sal*I and *Acc*65I (S/A), or *Bam*HI and *Acc*65I (B/A). The fragments were labeled by filling in the overhanging ends and then used for the immunoprecipitation binding assay. Lanes marked "IP" indicate immunoprecipitations performed with extracts containing RecA-E1, while the "IP*" lane was immunoprecipitated with extracts lacking RecA-E1. Lanes without the "IP" designation were marker lanes showing the labeled fragment pattern from each digestion. For each digestion, the 18-mer-containing fragment was the smallest fragment in each lane. Other fragments were derived from vector sequences and served as internal specificity controls. Panel B summarizes the results found in panel A. Shown are the sizes and sequences of the pUC18/18-mer-derived fragments which contain the 18-bp IR and the relative binding of RecA-E1 protein to each fragment. Binding was normalized to the 55-bp fragment and represents an average of two experiments.

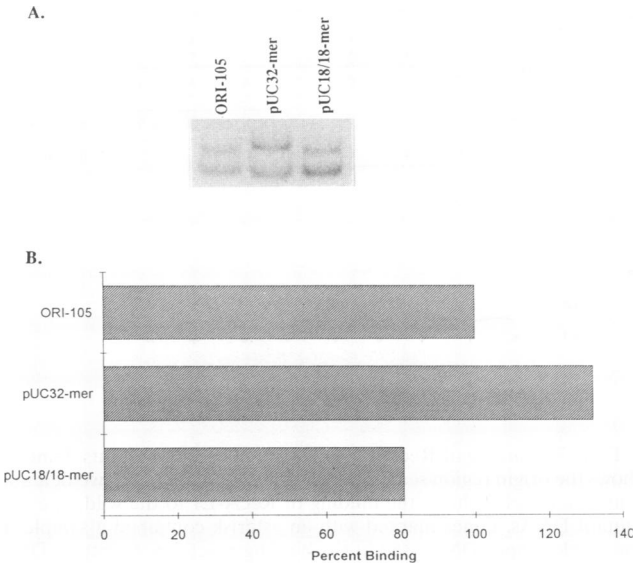

FIG. 6. Comparative binding of RecA-E1 to cloned 18-mer and 32-mer sequences in the context of a large DNA fragment. PCR-amplified fragments of pUC18/18-mer (256 bp), pUC32-mer (270 bp), or a cloned 105-bp BPV-1 origin fragment (ORI-105; 256 bp) were assayed for E1 binding (upper bands in panel A). A 211-bp PCR-amplified fragment containing the same BPV-1 sequences as ORI-105 was included in each binding reaction as an internal control for sample handling and recovery (lower bands in panel A). Binding was assayed as described in Materials and Methods, and panel A shows a typical binding experiment. To quantitate relative binding, the ratio of the amount of bound large and small fragments was determined for each sample. The ratio for the large to small fragments derived from the ORI-105 clone was set at 100, and the corresponding ratios for the pUC18/18-mer and pUC32-mer samples were normalized relative to this value. Panel B shows the average relative binding determined in three replicate experiments.

question in the absence of confounding effects due to differences in the overall length of flanking sequences, binding of E1 to the 18-mer and the 32-mer sequences was examined with each sequence embedded within a much larger fragment (256-bp overall length for the 18-mer and 270 bp for the 32-mer). E1 binding to the 18-mer- and 32-mer-containing fragments was compared relative to binding to a similar-size fragment, designated ORI-105. The ORI-105 fragment was 256 bp in length and contained 105 bp of BPV-1 DNA, including the entire functional origin region, along with additional adjacent BPV-1 sequences. As the ORI-105 fragment contained BPV-1 sequences including and beyond the boundaries of the most distal E1-protected regions detected by exonuclease III footprinting, it was presumed to contain all the necessary *cis* information for E1 origin binding and thus served as the wild-type substrate. Figure 6 shows a representative binding assay (panel A) and the average binding to the 18-mer and 32-mer sequences relative to ORI-105 (panel B). The 256-bp pUC18/18-mer fragment bound nearly as well as the 256-bp ORI-105 fragment, indicating that BPV-1 sequences outside the 18-bp IR did not dramatically influence E1-DNA interactions. In this series of experiments, the pUC32-mer fragment consistently bound somewhat better than the ORI-105 fragment. It is possible that the conjunction of the 32-mer sequences with vector sequences in this fragment creates a slightly more favorable environment for E1 binding than when the 32-mer sequences are in their normal BPV-1 context. Nonetheless, the results from this experiment indicate that sequences unique to the 32-mer do not greatly enhance E1 binding in comparison with the 18-mer.

Both halves of the 18-bp IR are critical for E1 binding. To examine the contribution of each half-site of the 18-bp IR to E1 binding, deletion mutations were constructed in either the 5′ or 3′ halves of the IR element (Fig. 7A). Removal of either half-site drastically reduced binding of RecA-E1 to the origin region (Fig. 7B). Densitometric quantitation of a longer exposure indicated that binding to the 5′ or 3′ deletion DNA was 1 to 3% of binding to the undeleted parental DNA. Identical results were obtained for binding of RecA-E1 to either the 5′ or the 3′ half-sites cloned into pUC18; the presence of either

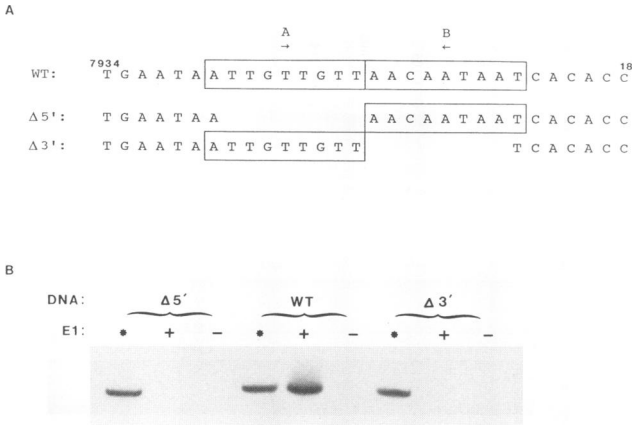

FIG. 7. Binding of RecA-E1 to half-site deletion mutants. Panel A shows the origin region sequences of wild-type BPV-1 and two deletion mutants. Panel B shows the binding of RecA-E1 to the wild-type and mutant DNAs. Lanes marked with an asterisk contained a sample of the labeled input DNA used in the binding reactions. Substrate DNA was produced by PCR amplification of the BPV-1 origin region from nucleotides 7830 to 99. The resultant fragment (214 bp for wild-type; 206 bp for the deletion mutants) was labeled by *Ava*II digestion followed by end repair with T7 DNA polymerase and [^{32}P]dCTP. One nanogram of labeled wild-type or mutant DNA was incubated with extracts containing (+ lanes) or lacking (− lanes) RecA-E1 and then immunoprecipitated with anti-E1 serum as described in Materials and Methods.

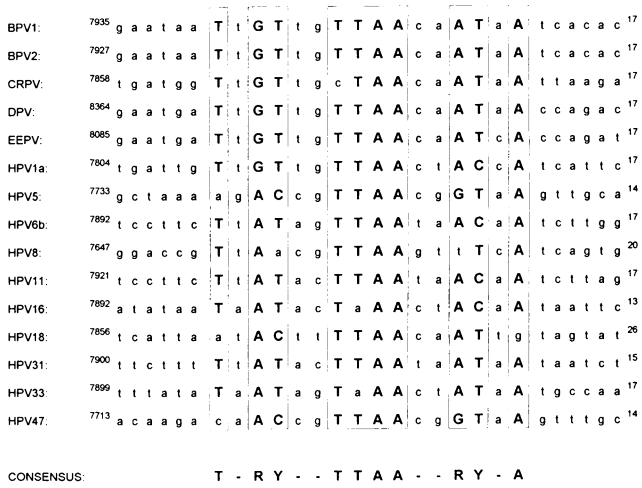

FIG. 8. Homology among papillomaviruses in the BPV-1 E1 binding site region. Shown is a 28-bp region from the BPV-1 genome and the comparable region from 14 other papillomaviruses. Nucleotide numbers for the individual papillomavirus genomes are as indicated. Within this 28-bp sequence, highly conserved regions are boxed. Below is a consensus sequence derived from the boxed sequences. Nucleotides within the boxed regions that conform to the consensus sequence are in capital letters while nonconforming nucleotides are in lowercase letters. CRPV, cottontail rabbit papillomavirus; DPV, deer papillomavirus; EEPV, European elk papillomavirus; HPV, human papillomavirus.

half-site in a pUC18 fragment was insufficient to direct specific binding by RecA-E1 (7). For both the deletion mutants and the cloned half-sites, the BPV-1-derived sequences were contained within a larger DNA fragment for the binding assay. Consequently, failure to bind was not due to a lack of nonspecific protein-DNA contacts. We conclude from these studies that the presence of sequences in both halves of the 18-bp IR was critical for the formation of a stable E1-DNA complex.

Identification of a consensus homology between the BPV-1 E1 binding site and comparable regions in other papillomaviruses. It has recently been shown that heterologous combinations of E1 and E2 proteins can support replication from papillomavirus origins of several species and subtypes (3). This functional interchangeability implies that origins from various papillomaviruses will show sequence and/or structural similarity. Figure 8 shows sequences from the presumed origin regions of 15 animal and human papillomaviruses. Each sequence was aligned for maximal homology with the E1 binding region on BPV-1. While many of the papillomaviruses had IR elements that were less perfect than the one in BPV-1 (i.e., had more mismatches), we were able to define a symmetric consensus sequence, T-RY--TTAA--RY-A, that was evident in all 15 of the papillomaviruses. This consensus sequence spanned 16 bp and resided entirely within the 18-bp IR region that was sufficient for BPV-1 E1 binding to BPV-1 DNA. Within the 16 bp spanned by the consensus, only 10 nucleotide positions were highly conserved, and these positions may represent the critical nucleotides for E1-DNA interactions. A point mutational analysis of this region is in progress to investigate this possibility.

DISCUSSION

Previous studies have shown that the BPV-1 E1 protein binds to the origin of replication but have not clearly defined the sequences responsible for binding specificity. Using DNase I and exonuclease III footprinting, we showed that E1 protection could extend over a 74-bp region of BPV-1 DNA. Within this overall protected region, there was a strongly protected core from nucleotides 7932 to 15. Examination of sequences within the core protected region revealed the presence of alternative, overlapping, IR elements. However, since both overlapping IR elements were located within the boundaries of the core nuclease-protected region, we could not distinguish the relative importance of either element by the footprinting studies. Subsequent experiments performed with double-stranded oligonucleotides corresponding to each IR demonstrated that sequences contained within the smaller, 18-bp IR element were sufficient for E1 binding specificity: (i) E1 bound efficiently to a self-ligated form of the 18-mer, but not to similar-size oligos with comparable AT content but unrelated sequence; (ii) ligated 18-mer effectively competed for E1 binding to a 219-bp origin-containing fragment; and (iii) the presence of a single copy of the 18-mer sequence in pUC18 fragments caused them to be bound by E1 while similar-size fragments lacking the 18-mer were not bound.

While the BPV-1 18 bp IR element was sufficient to confer E1-specific binding, the E1 protein was not able to form a stable complex with an isolated 18-mer sequence under our immunoprecipitation binding conditions. E1 binding could only be detected when the 18-bp sequence was contained within a larger DNA fragment generated either through self-ligation of the double-stranded oligonucleotide or by excision of the cloned 18-mer along with flanking sequences. As the flanking sequences for the cloned 18-mer were derived from the vector and not BPV-1, this indicated that there was no sequence specificity involved with this increased binding. From the studies with the oligonucleotides and the cloned 18-mer, we determined that the minimum sequence length required for

detectable E1 binding was between 18 and 31 bp. Furthermore, for fragments containing the 18-mer sequence, E1 binding increased with increasing fragment size over the range from 31 to 55 bp. This increased binding with larger DNA substrates was consistent with the footprinting experiments which detected E1-DNA interactions over an extended range. Together, these results suggest that E1 contacts the DNA for a significant distance beyond the 18-bp specificity element and that these nonspecific sequence contacts are important for the formation of stable E1-DNA complexes. Additional BPV-1 specific sequences outside the 18-bp IR element did not appear to contribute significantly to E1-DNA binding since the 18-mer embedded in a larger region of DNA was bound nearly as well as a comparable fragment containing the 32-mer or a 105-bp BPV-1 sequence.

How the E1 protein contacts the DNA over an extended region is not yet clear. The overall length of the contact region makes it unlikely that the distal interactions reflect binding of a single molecule of E1, unless there is significant deformation of the DNA structure. While such deformation cannot be excluded, a more likely possibility is that E1 is capable of forming multimeric complexes as is the case for SV40 T antigen. The symmetric nature of the binding region would be consistent with binding of dimeric or larger oligomeric forms of E1. If binding does involve formation of multimers, the failure to detect significant binding to either of the isolated half-sites of the 18-bp IR suggests that binding and assembly are highly cooperative.

Previous studies indicated that the overall origin regions from BPV-1 and other animal and human papillomaviruses are highly conserved (3, 28, 34). By focusing specifically on the 18-bp IR region, we identified a 16-bp symmetric consensus element (T-RY--TTAA--RY-A) that has not been previously described. This consensus element shows three degrees of conservation: (i) six positions, the outer two nucleotides (positions 1 and 16), and the central four nucleotides (positions 7 to 10) are highly conserved and consist of specific nucleotides; (ii) four positions are less stringent in that they are conserved for either purines (nucleotides 3 and 13) or pyrimidines (nucleotides 4 and 14) rather than specific nucleotides, and (iii) the remaining positions (no. 2, 5, 6, 11, 12, and 15) are fairly variable, with three different nucleotides being observed at those positions. This sequence organization suggests that the six highly conserved positions will be critical for binding of all E1 proteins from different papillomaviruses, while the other positions may contribute to papillomavirus type-specific E1-DNA interactions. We also noted that the 28-bp IR element in BPV-1 was not well conserved among other papillomaviruses (29). Lack of conservation in this larger IR is consistent with the absence of sequences critical for E1 binding outside the 18-bp IR element.

A prediction from this conservation of the E1 binding region is that the BPV-1 E1 protein might bind to heterologous papillomavirus origins and might functionally substitute in replication. Chiang et al. recently confirmed the second part of this prediction recently by demonstrating that BPV-1 E1 could substitute for the human papillomavirus type 11 (HPV-11) E1 in replication in vivo (3). We have shown that the BPV-1 E1 protein can bind to the homologous 18-bp IR in deer papillomavirus, through binding to HPV-1a DNA could not be detected (29). Deer papillomavirus has exactly the same sequence as BPV-1 for the 18-bp IR region, but diverges outside the boundaries of the 18-bp IR. Efficient binding of BPV-1 E1 to deer papillomavirus supports the contention that only sequences within the 18-bp IR are necessary for E1 binding specificity. The HPV-1a sequence, however, differs from BPV-1 at four positions within the 18-bp region, three in nonconserved positions and one at a conserved pyrimidine position. The failure to detect BPV-1 E1 binding to HPV-1a DNA is consistent with decreased binding affinity due to one or more of the nucleotide changes. Presumably, the cognate HPV-1a E1 would be more efficient at recognizing the HPV-1a sequences than is BPV-1 E1. The results with HPV-1a suggest that HPV-11 would be an even poorer substrate for BPV-1 E1 since it is more divergent with six nucleotide changes in the 18-bp region. If so, how could BPV-1 E1 substitute for HPV-11 E1 in the replication assay? We speculate that the presence of the E2 protein in vivo compensates for weak E1 binding. The E2 protein, which has been shown to enhance E1's ability to bind DNA (20, 23, 33), likely restored sufficient binding of BPV-1 E1 to the HPV-11 origin to allow formation of functional replication complexes. Such restoration is consistent with the origin mutation studies of Spalholz et al. in which origins that bound E1 poorly in the absence of E2 could still replicate in vivo (23).

From our current studies and those of other groups, the organization of the BPV-1 origin of replication is becoming more clearly defined. It is apparent that the BPV-1 origin has many features in common with the SV40 origin. Like SV40, the functional BPV-1 origin contains an initiation protein (E1) binding region and an adjacent AT-rich sequence. In both cases, the binding region contains a palindromic core element with extensive protein-DNA contacts occurring beyond the limits of the palindrome. Given the recent demonstration of E1 helicase activity (20, 21), these similarities in DNA organization and extended protein-DNA contact regions may reflect similar mechanisms for protein assembly and subsequent unwinding of origin DNA. However, while certain features are clearly similar, there are also significant differences. Unlike SV40, the BPV-1 origin also contains a binding site for a second viral protein, E2TA (26). E2 enhances binding of E1 to the origin, but it remains to be seen whether the enhancement activity is merely a quantitative effect or whether it in fact confers a qualitative difference on E1 function. In addition, the SV40 origin contains a functionally significant static DNA bend (25) while no such bend occurs in the BPV-1 origin (18). Thus, it is clear that the BPV-1 system will provide an alternative view of a eukaryotic origin. Identification of further differences and similarities between these two systems should continue to provide insight into fundamental aspects of replication initiation.

ACKNOWLEDGMENTS

This work was supported by grants from the National Science Foundation and the Texas Advanced Research Program and by PHS grant CA56699 from the National Institutes of Health.

REFERENCES

1. **Alderborn, A., N. Jareborg, and S. Burnett.** 1992. Evidence that the transcriptional trans-activating function of the bovine papillomavirus type 1 E2 gene is not required for viral DNA amplification in division-arrested cells. J. Gen. Virol. **73:**2639–2651.
2. **Blitz, I. L., and L. A. Laimins.** 1991. The 68-kilodalton E1 protein of bovine papillomavirus is a DNA binding phosphoprotein which associates with the E2 transcriptional activator in vitro. J. Virol. **65:**649–656.
3. **Chiang, C. M., M. Ustav, A. Stenlund, T. F. Ho, T. R. Broker, and L. T. Chow.** 1992. Viral E1 and E2 proteins support replication of homologous and heterologous papillomaviral origins. Proc. Natl. Acad. Sci. USA **89:**5799–5803.
4. **Clertant, P., and I. Seif.** 1984. A common function for polyomavirus large-T and papillomavirus E1 proteins? Nature (London) **311:**270–279.

5. **Dean, F. B., J. A. Borowiec, Y. Ishimi, S. Deb, P. Tegtmeyer, and J. Hurwitz.** 1987. Simian virus 40 large tumor antigen requires three core replication origin domains for DNA unwinding and replication in vitro. Proc. Natl. Acad. Sci. USA **84:**8267–8271.
6. **Fanning, E., and R. Knippers.** 1992. Structure and function of simian virus 40 large tumor antigen. Annu. Rev. Biochem. **61:**55–85.
7. **Holt, S. E.** Unpublished observations.
8. **Law, M. F., D. R. Lowy, I. Dvoretzky, and P. M. Howley.** 1981. Mouse cells transformed by bovine papillomavirus contain only extrachromosomal viral DNA sequences. Proc. Natl. Acad. Sci. USA **78:**2727–2731.
9. **Lentz, M. R., D. Pak, I. Mohr, and M. R. Botchan.** 1993. The E1 replication protein of bovine papillomavirus type 1 contains an extended nuclear localization signal that includes a $p34^{cdc2}$ phosphorylation site. J. Virol. **67:**1414–1423.
10. **Li, R., J. Knight, G. Bream, A. Stenlund, and M. Botchan.** 1989. Specific recognition nucleotides and their DNA context determine the affinity of E2 protein for 17 binding sites in the BPV-1 genome. Genes Dev. **3:**510–526.
11. **Liu-Johnson, H. N., M. R. Gartenberg, and D. M. Crothers.** 1986. The DNA binding domain and bending angle of the E. coli CAP protein. Cell **47:**995–1005.
12. **Lowy, D. R., I. Dvoretzky, R. Shober, M. F. Law, L. Engel, and P. M. Howley.** 1980. In vitro tumorigenic transformation by a defined sub-genomic fragment of bovine papilloma virus DNA. Nature (London) **287:**72–74.
13. **Lusky, M., and M. R. Botchan.** 1985. Genetic analysis of bovine papillomavirus type 1 *trans*-acting replication factors. J. Virol. **53:**955–965.
14. **Lusky, M., and M. R. Botchan.** 1986. A bovine papillomavirus type 1-encoded modulator function is dispensable for transient viral replication but is required for establishment of the stable plasmid state. J. Virol. **60:**729–742.
15. **Lusky, M., and M. R. Botchan.** 1986. Transient replication of bovine papilloma virus type 1 plasmids: cis and trans requirements. Proc. Natl. Acad. Sci. USA **83:**3609–3613.
16. **Lusky, M., and E. Fontane.** 1991. Formation of the complex of bovine papillomavirus E1 and E2 proteins is modulated by E2 phosphorylation and depends upon sequences within the carboxyl terminus of E1. Proc. Natl. Acad. Sci. USA **88:**6363–6367.
17. **Mohr, I. J., R. Clark, S. Sun, E. J. Androphy, P. MacPherson, and M. R. Botchan.** 1990. Targeting the E1 replication protein to the papillomavirus origin of replication by complex formation with the E2 transactivator. Science **250:**1694–1699.
18. **Schuller, G., S. E. Holt, J. Hsu, and V. G. Wilson.** 1994. The bovine papillomavirus type 1 genome contains multiple loci of static DNA bending, but bends are absent from the functional origin of replication. Virus Res. **31:**1–15.
19. **Seif, I.** 1984. Sequence homology between large tumor antigen of polyoma viruses and the putative E1 protein of papilloma virus. Virology **138:**347–352.
20. **Seo, Y.-S., F. Müller, M. Lusky, E. Gibbs, H.-Y. Kim, B. Phillips, and J. Hurwitz.** 1993. Bovine papilloma virus (BPV)-encoded E2 protein enhances binding of E1 protein to the BPV replication origin. Proc. Natl. Acad. Sci. USA **90:**2865–2869.
21. **Seo, Y. S., F. Muller, M. Lusky, and J. Hurwitz.** 1993. Bovine papilloma virus (BPV)-encoded E1 protein contains multiple activities required for BPV DNA replication. Proc. Natl. Acad. Sci. USA **90:**702–706.
22. **Spalholz, B. A., P. F. Lambert, C. L. Yee, and P. M. Howley.** 1987. Bovine papillomavirus transcriptional regulation: localization of the E2-responsive elements of the long control region. J. Virol. **61:**2128–2137.
23. **Spalholz, B. A., A. A. McBride, T. Sarafi, and J. Quintero.** 1993. Binding of bovine papillomavirus E1 to the origin is not sufficient for DNA replication. Virology **193:**201–212.
24. **Spalholz, B. A., Y. C. Yang, and P. M. Howley.** 1985. Transactivation of a bovine papilloma virus transcriptional regulatory element by the E2 gene product. Cell **42:**183–191.
25. **Tegtmeyer, P., S. Deb, A. L. DeLucia, S. P. Deb, A. Koff, S. Tsui, R. Parsons, K. Partin, C. P. Baur, F. B. Dean, and J. Hurwitz.** 1988. SV40 origin of replication, p. 123–132. *In* T. Kelly and B. Stillman (ed.), Cancer cells: eukaryotic DNA replication. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
26. **Ustav, E., M. Ustav, P. Szymanski, and A. Stenlund.** 1993. The bovine papillomavirus origin of replication requires a binding site for the E2 transcriptional activator. Proc. Natl. Acad. Sci. USA **90:**898–902.
27. **Ustav, M., and A. Stenlund.** 1991. Transient replication of BPV-1 requires two viral polypeptides encoded by the E1 and E2 open reading frames. EMBO J. **10:**449–457.
28. **Ustav, M., E. Ustav, P. Szymanski, and A. Stenlund.** 1991. Identification of the origin of replication of bovine papillomavirus and characterization of the viral origin recognition factor E1. EMBO J. **10:**4321–4329.
29. **Wilson, V. G.** Unpublished observations.
30. **Wilson, V. G., and J. Ludes-Meyers.** 1991. A bovine papillomavirus E1-related protein binds specifically to bovine papillomavirus DNA. J. Virol. **65:**5314–5322.
31. **Wilson, V. G., and J. Ludes-Meyers.** 1992. Partite expression of the bovine papillomavirus E1 open reading frame in Escherichia coli. Biochim. Biophys. Acta **1129:**215–218.
32. **Winokur, P. L., and A. A. McBride.** 1992. Separation of the transcriptional activation and replication functions of the bovine papillomavirus-1 E2 protein. EMBO J. **11:**4111–4118.
33. **Yang, L., R. Li, I. J. Mohr, R. Clark, and M. R. Botchan.** 1991. Activation of BPV-1 replication in vitro by the transcription factor E2. Nature (London) **353:**628–632.
34. **Yang, L., I. Mohr, R. Li, S. Nottoli, S. Sun, and M. Botchan.** 1991. Transcription factor E2 regulates BPV-1 DNA replication in vitro by direct protein-protein interaction. Cold Spring Harbor Symp. Quant. Biol. **56:**335–346.