

Orientation of a Novel DNA Binding Site Affects Human Papillomavirus-Mediated Transcription and Replication

CHRISTOPHER D. NEWHOUSE¹ AND SAUL J. SILVERSTEIN^{1,2*}

Integrated Program in Cellular, Molecular and Biophysical Studies¹ and Department of Microbiology,² Columbia University, New York, New York 10032

Received 18 October 2000/Accepted 22 November 2000

A consensus binding site for the human papillomavirus (HPV) E2 protein was determined from an unbiased set of degenerate oligonucleotides using cyclic amplification and selection of targets (CASTing). Detectable DNA-protein complexes were formed after six to nine cycles of CASTing. A population of selected binding sites was cloned, and a consensus was determined by statistical analysis of the DNA sequences of individual isolates. Starting from a pool with 20 random bases, a consensus binding site of ACAC-N₅-GGT was derived. CASTing and electrophoretic mobility shift analyses demonstrate that human but not bovine papillomavirus E2 proteins recognize this sequence. The presence of this sequence in papillomavirus genomes suggests a role for its function. We demonstrate that this site functionally substitutes for the canonical E2 binding site (ACCG-N₄-CGGT) in both transient-transcription and DNA replication assays. This sequence, in most instances, is interchangeable with the resident E2 binding sites in the context of the HPV type 16 long control region. Where the novel sequence does not support E2-mediated effects on gene expression or DNA replication, we demonstrate that changing the orientation of the novel sequence restores this effect.

Papillomaviruses (PVs) cause a benign hyperproliferation of epithelial cells that sometimes progress to form carcinomas. The model system for studying papillomaviruses has been bovine papillomavirus type 1 (BPV1) because of its ability to infect and transform a variety of rodent cells in tissue culture (41).

Papillomaviruses encode early and late proteins that are involved in regulation of virus gene expression and replication, and assembly of the virion, respectively. E2 is a dimeric multifunctional early protein that is intimately involved in the regulation of gene expression and viral genome replication. The primary structures of papillomavirus E2 proteins are highly conserved. They consist of an N-terminal domain that can act as a transcriptional activator and is also involved in viral DNA replication and interaction with the viral DNA replication protein E1; a central, poorly conserved hinge region; and the C-terminal DNA binding and protein dimerization domains (17). Analysis of the crystal structure of the C terminus of the BPV1 E2 protein bound to its DNA binding site revealed that two α -helices, one from each monomer of an E2 dimer, bind in the major groove of the DNA (8, 9). The original studies of the BPV1 E2 protein identified the high-affinity DNA binding sites within the BPV genome to be 12 bp long and to minimally contain the sequence ACC-N₆-GGT (canonical site), with the highest-affinity site being ACCG-N₄-CGGT (2, 15).

The occurrence of BPV1 E2 high-affinity DNA binding sites in the long control region (LCR) of the human papillomaviruses (HPVs) has led investigators to assume that the HPV E2 proteins preferentially bind this site (1, 33, 39, 40, 42).

Many E2 functions are dependent on the relative affinity of

the protein for its various DNA binding sites. HPVs that infect the genital mucosal epithelium contain multiple E2 binding sites within their LCRs, both proximal and distal to the transcription initiation site for early-gene expression. The distal, higher-affinity sites apparently act as enhancers, and the proximal, lower-affinity sites act as repressors of early gene expression (17). These sites are thought to be part of a switching mechanism that modulates the levels of early gene expression during the viral life cycle. In addition, it has been shown that E2 can compete for the binding of cellular transcription factors from their neighboring or overlapping sites within the LCR (5, 37, 38).

The papillomavirus proteins E1 and E2 are necessary for HPV genome replication (12, 32, 36). For HPVs, the minimal origin of replication is defined as an E1 (E1BS) and an E2 binding site (E2BS) in close proximity flanked by an A/T-rich region, with additional E2BSs facilitating replication (3, 16, 36). In the absence of an E1BS, two E2BSs near the A/T-rich region can support transient HPV replication (36). Deletion analyses around the BPV1 origin of replication have revealed that the location of E2BSs with respect to the E1BS and A/T-rich region is flexible. Moreover, the affinity of the E2 protein for a particular site directly correlates with its ability to stimulate DNA replication (6, 43, 45).

The primary objective of our studies was to elucidate the highest-affinity binding sites for HPV E2 proteins. To address this question, we employed the nonbiased cyclic amplification and selection of targets (CASTing) technique (48). We identified a unique set of sequences, ACAC-N₅-GGT, that HPV E2 proteins (but not the BPV1 E2 protein) bind with a relative affinity that is indistinguishable from the canonical high-affinity site. Our studies also suggest the existence of preferred nucleotides within the flanking and core (N₅) sequences. Comparisons of the relative affinities and binding complex half-lives were made for the HPV51, HPV-16, and BPV1 E2 proteins

* Corresponding author. Mailing address: Department of Microbiology Columbia University, 701 W. 168th St., New York, NY 10032. Phone: (212) 305-8149. Fax: (212) 305-5106. E-mail: sjs6@columbia.edu.

TABLE 1. Primers

Protein	Primer type	Primer
GST-HPV ₅₁ E2fl	5'	5'-GGGGGATCCAGAAAAATGGAGACCCTATGC-3'
	3'	5'-GGGAATTCACAATACATATATTACTAG-3'
GST-B ₁ E2fl	5'	5'-CGAGGAGGAGGATAGGGATCCGGATGGAG-3'
	3'	5'-GGCAATGGCAGAATTCAGAAGTCCAAG-3'
GST-HPV ₅₁ E2sct	5'	5'-TAAGACTGGATCCATATGGGAGGGCACCACAAAGTGCAAC-3'
	3'	5'-CAGTATATATGTAGAATTCATATATTACTAG-3'
GST-HPV ₁₆ E2sct	5'	5'-TAAGATCGGATCCATATGGGACGGATTAAGTGAATAGT-3'
	3'	5'-TAAGATCGAATTCAGCACGCCAGTAATGTTG-3'
GST-B ₁ E2sct	5'	5'-TAAGACTGGGATCCATATGGTACCCGGTGGACTTGGC-3'
	3'	5'-GGCAATGGCAGAATTCAGAAGTCCAAG-3'

with the different DNA binding sites. These novel sites are located within papillomavirus genomes at locations where E2BSs are typically found.

In order to assess how an E2 protein might utilize the novel E2BSs, we used it in place of the wild-type sites found in the early promoter and replication origin of the better-characterized HPV type 16. HPV16 infects the epithelial cells of the genital mucosa and, like all other high-risk HPVs, is strongly associated with cervical carcinoma. We have designed a single plasmid that allows the assay of both transient transcription and transient replication. This plasmid, pOri16L, was built with a portion of the HPV16 LCR that contains both the origin of replication and the early promoter driving the expression of the firefly luciferase reporter gene. There are three canonical E2BSs within this portion of the LCR (see Fig. 5). In this study, we mutated each of the wild-type E2BSs to either eliminate binding (BS-KO) or create new E2BSs with the novel binding site sequence ACACAAATCGGT. Here we demonstrate that the novel E2BS functionally substitutes for the native E2BSs within this portion of the LCR in both transient-transcription and replication assays. Because bp 3 of the novel E2BSs disrupts the canonical site's palindrome, we also addressed the influence of the orientation of this binding site on E2 function. The results of our experiments with these mutated LCRs demonstrate that the functional role for the ACAC-N₅-GGT sites in E2-mediated replication of and transcription from the HPV genome can be dependent on binding site orientation.

MATERIALS AND METHODS

DNA constructs for protein expression. DNA constructs used for expression of proteins in bacteria were made by PCR amplification of portions of the genes encoding the E2 proteins from HPV types 51 and 16 and BPV type 1. The resulting PCR products were digested with *Bam*HI and *Eco*RI, whose sites are in the primers used for amplification (underlined in the primer sequences listed in Table 1) and then ligated in frame with and C-terminal to the glutathione *S*-transferase (GST) gene of pGEX-3X (Amersham Pharmacia Biotech, Piscataway, N.J.). These constructs were designed to express GST fusion proteins with either the full-length (fl) or the short C terminus (sct) E2 proteins that contain only the DNA binding and dimerization domain of the respective E2 proteins. The fl constructs GST-₅₁E2fl, and GST-B₁E2fl contain the entire E2 coding sequences. The set constructs GST-₅₁E2sct, GST-₁₆E2sct, and GST-B₁E2sct contain papillomavirus nucleotide sequences 3536 to 3811 from HPV51, 3584 to 3892 from HPV16, and 3457 to 3840 from BPV1, respectively. Nucleotide numbering corresponds to papillomavirus genome sequences listed in the Human Papillomavirus Compendium, HPV database (22). The primers used to create these constructs are listed in Table 1.

pALEX, which expresses the GST protein fused to six histidines (GST-His₆) (23), and pCPC-XE2(51), which expresses the GST-₅₁E2fl protein, were provided by Christos Panagiotidis.

Recombinant baculoviruses used for expression of proteins were the HPV

₁₆E2 full length and short C terminus protein expressing viruses rvE2 and rvE2sct (34), BPV ₁E2 full length-protein expressing virus vE2 (20), and HPV ₅₁E2 full-length protein expressing virus v₅₁E2 that was constructed by cleaving the *Bam*HI-*Eco*RI fragment from pCPC-XE2(51) and ligating it into the pFastBac1 expression vector (Life Technologies, Grand Island, N.Y.), creating pBac₅₁E2. This construct was then used with the Bac-to-Bac expression system (Life Technologies) to create a baculovirus that expressed full-length HPV51 E2 protein. β-Glucuronidase-expressing baculovirus was provided as part of the Bac-to-Bac kit and was used as a control. All constructs were sequenced to confirm their identity.

Protein purification. *Escherichia coli* strain BL21/DE3 was used for bacterial expression of recombinant proteins. Proteins were extracted from cultures (500 to 1,000 ml) of bacteria grown in liquid overnight at 25°C without IPTG (isopropylthiogalactoside) induction because induction resulted in partitioning of most of the E2 proteins to the insoluble fraction upon extraction (data not shown). Total-cell extracts were made by sonication, and proteins were purified by their affinity for glutathione-agarose beads (Amersham Pharmacia Biotech). Cleavage of the GST from the E2 protein was performed by addition of factor Xa protease (Roche Molecular Biochemicals, Indianapolis, Ind.) to the glutathione-eluted fraction and incubation with gentle agitation at 4°C for 16 to 24 h in factor Xa cleavage buffer (50 mM Tris-HCl [pH 8.0], 100 mM NaCl, 10 mM MgSO₄, 1 mM CaCl₂, 5 mM dithiothreitol [DTT]). To further purify this protein, it was applied to an S-Sepharose (Amersham Pharmacia Biotech, Piscataway, N.J.) column that was equilibrated with S-Sepharose buffer (20 mM Tris-HCl [pH 8.5], 100 mM NaCl, 5 mM EDTA [pH 8.0]). Protein was eluted with a linear salt gradient (100 mM to 1 M NaCl) in S-Sepharose buffer, and 1-ml fractions were collected and screened for active E2 protein by electrophoretic mobility shift assay (EMSA) and Western blotting (data not shown).

Proteins expressed from Sf-21 cells infected with recombinant baculoviruses for 36 to 72 h postinfection (depending on virus and multiplicity of infection) were harvested, and nuclear extracts were made from them as described previously (24).

Cloning of PCR-amplified sequences. Sequences amplified by PCR were cloned using the TA cloning kit (Invitrogen Corp., Carlsbad, Calif.) with the pCRII and pCR2.1 vectors according to the manufacturer's instructions.

Sequencing reactions. DNA sequencing was performed using the Sequenase version 2.0 DNA sequencing kit as per the manufacturer's instructions (Amersham/USB) with M13 forward and M13 reverse primers end labeled with [γ -³²P] ATP using T4 polynucleotide kinase.

CASTing. The CASTing method was performed as previously described (48) with minor modifications. Each oligonucleotide in the degenerate oligonucleotide library (DOL) contains PCR primer binding sites that flank a core region of 20 randomly generated nucleotides. The oligonucleotides used were DOL, 5'-A GACGGATCCATTGCA-N₂₀-CTGTAGGAATTCGGA-3', and the primers used for PCR amplification were N₂₀-B (5'-AGACGGATCCATTGCA-3') and N₂₀-R (5'-TCCGAATTCCTACAG-3'). CASTing was performed by adding about 10 μg of double-stranded DOL to 20 μl of glutathione-agarose beads with approximately 15 to 20 μg of GST fusion protein (GST-HPV₅₁E2fl, GST-His₆, or GST-BPV₁E2fl) captured from dialyzed extracts in a total volume of 100 μl of binding buffer (10 mM Tris-HCl [pH 7.5] 50 mM NaCl, 1 mM EDTA, 4 mM DTT, 250 μg of bovine serum albumin [BSA] per ml, 5% glycerol). This mix was allowed to interact at room temperature for 3 h with gentle agitation followed by three washes with binding buffer. The complexes of glutathione-agarose beads/GST protein/double-stranded DOLs were then resuspended in the PCR buffer mix (30 μl of 10× *Taq* buffer [Promega], 18 μl of 25 mM MgCl₂, 3 μl of 10 mM deoxynucleotide triphosphate (dNTP) mix 3 μl of *Taq* DNA polymerase, 3 μl of

500-pmol/ μ l N₂₀B primer, 3 μ l of 500-pmol/ μ l N₂₀R primer, in a total volume of 300 μ l) and amplified according to the following protocol: 95°C for 5 min, then 10 PCR cycles of 94°C for 1 min, 40°C for 1 min, and 72°C for 30 sec, followed by cooling to 4°C, and a 100- μ l aliquot was removed and stored on ice. The remaining 200 μ l was cycled four more times (14 in total) and cooled to 4°C, and another 100- μ l aliquot was removed and stored on ice. The remaining 100 μ l was subjected to four more rounds of amplification (18 in total), followed by cooling to 4°C. A 10- μ l aliquot (plus 1 μ l of loading dye) from each of the PCR amplifications (10, 14, and 18 cycles) was electrophoresed on a 2.5% agarose-1 \times TBE (90 mM Tris-borate, 1 mM EDTA [pH 8.5]) gel and visualized by ethidium bromide staining and UV transillumination. A portion of the pool obtained from each round of CASTing was amplified using radiolabeled primers and tested for enrichment by EMSA. After nine rounds of CASTing with GST-HPV₅₁E2fl, a protein-probe complex was easily detected by EMSA, and the CASTing procedure was stopped. The amplified pools were cloned into the pCRII vector (Invitrogen), the resulting plasmid DNAs were isolated from individual clones, and their sequences were determined.

CASTing with the GST-BPV₁E2fl protein was performed as above with the exception that the alignment was based on 65 independent clones from two independent CASTing experiments.

Additional plasmid constructs. (i) E2 binding sites. The construct pBPV1E2BS (a gift from Eliot Androphy) is a pUC18 derivative containing the sequence tcgagaACCGAATTCGGTacc cloned into the polycloning sequence. This clone was used as a positive control in EMSAs that analyzed the products of the CASTing reactions and for screening TA clones (last two lanes of Fig. 1). Additional E2 binding sites used as probes for EMSAs were cloned by annealing two complementary oligonucleotides (H and B clones; see Table 2) that contain the same flanking sequences; +, 5'-CCAGAGTGAATTCAG- (12-bp binding site)-TCCCAAGCTTGGCG-3', and -, 5'-CGCCAAGCTTGGGA-(complement of 12-bp binding site)-TCTGGAATTCAGTCTGG-3'. These were then cleaved with *Eco*RI and *Hind*III and ligated into a pUC18 vector between its unique *Eco*RI and *Hind*III sites.

(ii) pOri16L. The pOri16L plasmid was constructed by building a portion of the HPV16 LCR and the firefly luciferase gene into a pUC19 plasmid backbone. The portion of the HPV16 LCR from nucleotide positions 7800 to 73 (nucleotide numbering as in reference 22) that includes the HPV16 origin of replication and the P97 early promoter, including its TATA box, was amplified by PCR. The primers used to amplify this portion of the LCR contain novel *Pst*I and *Bam*HI restriction sites to facilitate cloning: 16Ori-upper/*Pst*I primer, 5'-CATGAACTGTCTGCAGGTTAGTCATAC-3', and 16Ori-lower/*Bam*HI primer, 5'-GTGCATAAAGGATCCGCTTTTATAC-3'.

The plasmid backbone was provided by pUC19-EX. The pUC19-EX plasmid was constructed by digestion of pUC19 DNA at its *Eco*RI and *Xma*I sites in *Eco*RI buffer (New England Biolabs, Beverly, Mass.) and filling in the 5' overhangs with deoxyribonucleotides using T4 DNA polymerase in T4 DNA polymerase buffer (New England Biolabs) with 100 mM dNTPs. The blunt ends were then ligated using T4 DNA ligase in T4 DNA ligase buffer (New England Biolabs). The pUC19-EX and the HPV16 LCR PCR products were digested with *Pst*I and *Bam*HI, gel purified, and ligated to each other. The resulting plasmid is referred to as pOri16. The open reading frame (ORF) encoding firefly luciferase was purified after *Bam*HI digestion of the p19luc plasmid (46) and inserted into the pOri16 plasmid at the *Bam*HI site to yield pOri16L.

(iii) pOri16L mutants. The binding site knockouts and sequence substitutions were all created by site-directed mutagenesis of pOri16L using either the MORPH kit (5 Prime \rightarrow 3 Prime, Inc., Boulder, Colo.) or the QuickChange site-directed mutagenesis kit (Stratagene Cloning Systems, La Jolla, Calif.) as per the manufacturers' instructions. For plasmids with mutations in more than one E2BS, one site was altered using the pOri16L plasmid as the template and the second site was altered using the partially mutated plasmid as the template. All mutated plasmids were screened by DNA sequence analysis. Upon isolation of mutant clones, origin-containing fragments were removed by digestion with *Pst*I and *Eco*RI (the *Eco*RI used for this recloning step is found in the luciferase *Bam*HI cassette) and then ligated into a pOri16L plasmid from which the wild-type LCR was removed by digestion with the same enzymes.

(iv) pCMV Series. The pCMV-E2₁₆ and pCMV-E1₁₆ expression plasmids were kindly provided by Peter Howley (32). They express the full-length HPV16 E1 and HPV16 E2 proteins driven by the cytomegalovirus (CMV) promoter. The pCMV₄-XS plasmid was constructed by digesting pCMV-E1₁₆ with *Xba*I and *Sma*I, filling in the resulting overhangs with deoxyribonucleotides using T4 DNA polymerase, and ligating the resulting blunt ends as described above for the pUC19-EX plasmid construct. The identities of all of the above constructs were confirmed by DNA sequence analysis.

EMSA. M13rev primers were end labeled with [γ -³²P]ATP using T4 polynucleotide kinase. PCR was then performed using this radiolabeled primer plus an unlabeled M13for primer to create a single end-labeled, double-stranded DNA probe for EMSA. The constructs used as templates for making probes were either the CASTing TA clones in the pCRII vector or pH and pB clones (see above for cloning details and Table 2 for binding site sequences) in pUC18. The resulting PCR products were gel purified. EMSAs were performed in binding buffer with purified protein, 1 μ g of sonicated salmon sperm DNA (Sigma, St. Louis, Mo.), 3 μ g of BSA, and 10⁴ cpm of radiolabeled DNA probe. This binding reaction was incubated at room temperature for 30 min and then loaded directly onto 6 to 8% native polyacrylamide gels containing 0.25 \times TBE and 2.5% glycerol.

Relative affinities were determined by adding dilutions of unlabeled competitor DNAs prepared by PCR amplification of clones (see Table 2) using unlabeled M13 primers that were partially purified using the Concert PCR rapid purification system (Life Technologies, Rockville, Md.).

Off-rates ($T_{1/2}$) were determined by performing binding reactions as above, and after the 30-min incubation, a 500- to 1,000-fold excess of unlabeled competitor DNA (see Fig. 4 legend for details) was added and incubated for an additional 0, 1, 5, 10, 30, or 90 min before electrophoresis. Following polyacrylamide gel electrophoresis (PAGE), the gels were dried and exposed to PhosphorImager screens (Molecular Dynamics, Sunnyvale, Calif.), and the bands were quantified using ImageQuant software (Molecular Dynamics).

Tissue culture. J2-3T3 cells were cultured in Dulbecco's modified Eagle's medium with 10% bovine calf serum. SCC-13 cells (29) were grown on mitomycin C-treated J2-3T3 cell feeder layers in E medium as described previously (19).

Luciferase expression assays. SCC-13 cells were plated on mitomycin C-treated J2-3T3 feeder layers in 35-mm dishes 24 h before transfection. The plasmids pOri16L, pRL-TK, and pCMV-E2₁₆ or pCMV₄-XS (see the legends to Fig. 6 and 7 for details) were introduced into the SCC-13 cells using LipofectAmine, as per the manufacturer's instructions (Life Technologies). Cells were harvested and assayed at 36 h posttransfection using the dual luciferase kit (Promega, Madison, Wis.), and units of luciferase activity were determined using a Berthold Lumat LB9501 luminometer (Berthold Systems, Inc., Pittsburgh, Pa.). Expression levels were determined from duplicate transfections in three independent experiments.

Transient-replication assays. Transient-replication assays were performed with SCC-13 cells as previously described for HPV31 (12). Briefly, plasmid DNAs (quantities and identities detailed in the legend to Fig. 5) were electroporated into SCC-13 cells as detailed by Hubert et al. (12) and Ustav and Stenland (44). Replication was assayed after *Dpn*I digestion of low-molecular-weight DNA extracted by a modified Hirt protocol (10) followed by gel electrophoresis and Southern blot hybridization.

RESULTS

Cyclic amplification and selection of targets. High-affinity DNA binding sites for the HPV E2 protein were identified using the CASTing technique (48). A GST fusion with the full-length HPV type 51 E2 protein (GST-₅₁E2fl) was used to select specific DNA binding sites from a random pool of degenerate double-stranded oligonucleotides (see Materials and Methods for details).

DNAs from the degenerate oligonucleotide pool were amplified by PCR, cloned, and sequenced to determine if representation within the central 20-bp region was truly random. DNA sequence analysis shows that the starting DNA pool used for CASTing contained a stretch of 20 bp where the abundance of all four bases was the same (data not shown). This demonstrates the random nature of the starting degenerate oligonucleotide N₂₀S pool and that there was no apparent overrepresentation of any particular sequence.

Each round of the CASTing consisted of three steps: binding of the double-stranded DNA oligonucleotide pool (N₂₀S) to the GST-E2 protein, removal of unbound DNA, and PCR amplification of DNA that remained bound to the GST-E2 protein-glutathione-agarose bead complex. EMSAs were performed after each round of CASTing to assess the efficiency of binding site selection (Fig. 1). The enriched population pools

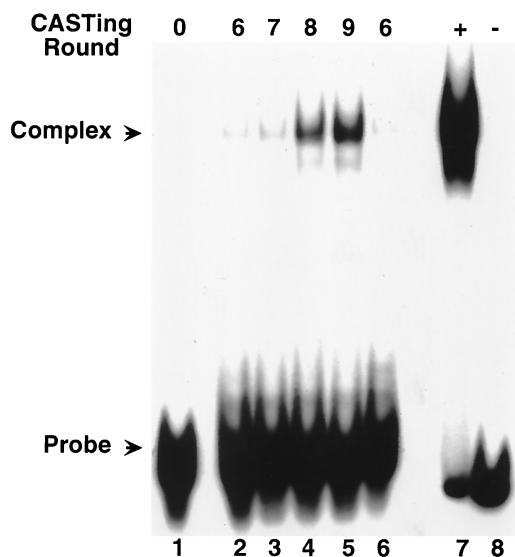


FIG. 1. EMSA analysis of CASTing-enriched probes. Oligonucleotide probes that formed DNA-protein complexes were amplified by PCR and radiolabeled by incorporation of [α - 32 P]dCTP in the reaction mixture. These DNA probes were used in EMSAs with the GST-HPV₅₁E2 protein to form complexes. The original degenerate oligonucleotide pool (N₂₀S) was used as the template for the lane labeled round 0. The next five lanes represent amplified probes from rounds 6, 7, 8, and 9, as labeled above the EMSA. The final two lanes were loaded with complex formed between radiolabeled probe amplified by PCR from the pBPV1-E2BS plasmid, which contains the binding site sequence gagaACCGAATTCGGTtagcc as the template with (+) or without (-) the GST-HPV₅₁E2 protein.

were cloned (see Materials and Methods), their DNA sequences were determined, and a consensus binding site was determined.

The first CASTing experiment was performed using the GST-HPV₅₁E2fl and GST-His₆ proteins in parallel to select for specific and nonspecific DNA binding sites, respectively. Probes made from the enriched population pools from rounds 6 through 9 of the GST-HPV₅₁E2fl CASTing were used in EMSAs to screen for enrichment of the population with high-affinity binding sites (Fig. 1). Abundant gel-shifted probes were obtained after eight and nine rounds of selection. The enriched populations of oligonucleotides were then cloned for sequence analysis from both the GST-HPV₅₁E2fl and GST-His₆ CASTings.

A second CASTing experiment was performed with the GST-HPV₅₁E2fl protein to ensure the reproducibility of our results. A total of 75 TA clones pooled from the GST-HPV₅₁E2fl CASTing experiments were sequenced and aligned to determine the presence of a consensus binding site sequence for the GST-HPV₅₁E2fl protein (not shown). The frequency with which a given nucleotide was located at each position is presented in Fig. 2A, and statistical analysis of the sequences identifies the nonpalindromic sequence RACACAAATTCGGTY (where R is a purine and Y is a pyrimidine) as the consensus binding site for the GST-HPV₅₁E2fl protein. This sequence is similar to the palindromic site previously identified for the BPV1 E2 protein (ACCG-N₄-CGGT) (15) in that it has an A/T-rich core and purine and pyrimidine nucleotides are present at the -1

flanking positions as described for the sites preferred by HPV16 E2 (40). However, two nucleotides from only one half of the 12-bp site (ACAC, also underlined above) differ from the canonical BPV1 E2 binding site.

We also examined the frequency with which particular intact sites are represented within the enriched population pool (Fig. 2B). Sequences that contain ACAC-N₄-CGGT at the first and last four positions of the 12-bp site represent 36% of the sequenced clones. Applying slightly less stringent criteria by searching for ACAC-N₅-GGT, we observed that 47% of the total pool match this site. In contrast, the canonical DNA binding site for BPV1 E2, ACCG-N₄-CGGT, was detected in only 3 of the 75 sequenced clones from the HPV51 E2 CASTing pools. A lower-affinity site (ACC-N₆-GGT) is recognized by HPV51E2, though it was identified in only 15% of the sequenced pool.

As a further control for specificity, a random set of the round 9 TA clones were subjected to analysis by EMSA. Radiolabeled probes were amplified by PCR using the TA-cloned DNAs as the template, and EMSA analyses were performed (data not shown). From differences in the abundance of the complexes, we concluded that there was a wide range of relative affinities between the GST-HPV₅₁E2fl protein and the various DNA sites. In contrast, when a similar analysis was performed using probes made from the clones isolated from the GST-His₆ CASTing experiment (data not shown), neither the GST-HPV₅₁E2fl nor GST-His₆ protein formed complexes with these DNAs. Therefore, neither the GST portion of the fusion protein nor glutathione-agarose beads were selecting specific DNA sequences.

Because the CASTing consensus sequence differed from the canonical BPV1 E2 site, we were concerned that the CASTing procedure was not correctly identifying the HPV E2 consensus binding site. To address this possibility, we repeated the CASTing using the GST-BPV1E2fl protein to ensure that its high-affinity, palindromic site was efficiently recognized within the pool of degenerate oligonucleotides. Under our conditions of selection and enrichment, the canonical, high-affinity, pal-

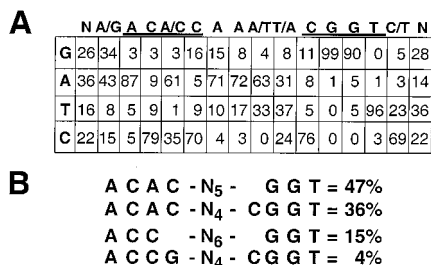


FIG. 2. Determination of a binding site consensus from the GST-HPV₅₁E2fl CASTing. PCR amplifiers from multiple rounds of the CASTing process were cloned, and the sequence of the insert was determined. Seventy-five sequences from the GST-HPV₅₁E2fl CASTing experiment were isolated, sequenced, and aligned based on homologies within the variable 20 bp of each clone. (A) The numbers represent the frequency with which a nucleotide was found at each position (expressed as a percentage of the total). Listed above the table is the consensus arrived at by determination of the most frequently occurring nucleotide at each position. (B) The frequency of specific sequences within the pool of sequenced clones expressed as a percentage of the total. Note that the first two and last two sequences listed are subsets of one another, respectively.

indromic BPV1 E2 DNA binding site (ACCGgatCGGT) but not the novel site (ACAC-N₄-CGGT) was identified (data not shown). These results confirm that, despite their poor representation among the TA clones isolated from the GST-HPV₅₁E2fl CASTing experiment, this procedure readily identifies the canonical, high-affinity DNA binding site of the BPV1 E2 protein.

HPV E2 proteins bind to the novel nonpalindromic site with higher relative affinities than to a canonical BPV E2 site. EMSAs were performed using E2 proteins derived from different mucosal HPVs to ask if binding to the nonpalindromic HPV site was a general characteristic of HPV E2 proteins (Fig. 3). The probes used were the CASTing consensus sequence and a high-affinity canonical E2 binding site. Equal amounts of extract were used for each EMSA reaction; however, because the expression levels of the different E2 proteins varied, meaningful conclusions can only be drawn from comparisons between the relative amounts of the probes shifted by the same E2 protein extract. In addition, the expression levels of the baculoviruses expressing BPV₁E2fl and HPV₁₆E2ct were very low, making the complexes in Fig. 3, lanes 6 and 10, difficult to see in this reproduction of the autoradiograph. Note also that the complexes formed with nuclear extracts may contain more proteins that interact with the probes than just the baculovirus-expressed E2. Thus, the relative mobilities of the various E2 complexes cannot be meaningfully compared. The GST-₅₁E2fl protein has a higher relative affinity for the CASTing consensus site than for the canonical E2 binding site (Fig. 3, lanes 1 and 2). The same result is found with the full-length HPV51 E2 protein purified from a baculovirus expression system (Fig. 3, lanes 3 and 4), demonstrating that recognition of the novel binding site was not a property of either the GST fusion or bacterial expression of the protein. Note that in the case of the GST fusions with full-length E2 proteins, because both the E2 proteins and the GST tag can dimerize independently, there is a pair of shifted complexes that likely represent dimers and tetramers of these proteins (Fig. 3, lanes 1 and 2). When a full-length BPV1 E2 protein was used in the EMSA, there was no detectable shift of the GST-₅₁E2 CASTing consensus site (Fig. 3, lanes 5 and 6), consistent with the fact that this site was not represented within the pool of sequenced GST-BPV₁E2fl CASTing clones (data not shown).

Further analysis of binding reactions performed with the HPV₁₆E2fl and the HPV₁₆E2ct proteins reveals that each of these proteins has a higher relative affinity for the novel site identified by CASTing than for the BPV1 E2 consensus site (Fig. 3, lanes 7 to 10). The proteins prepared from bacteria and baculovirus-infected insect cells had similar relative affinities (Fig. 3, lanes 1 to 4). Also, the full-length and C-terminal HPV16 E2 proteins had similar relative affinities for the two probes (Fig. 3, lanes 7 to 10); this result is consistent with that described for full-length and C-terminal BPV1 E2 proteins (18). Therefore, all further analyses of the E2 protein interactions with DNA binding sites used C-terminal proteins containing only the DNA binding and dimerization domains.

Analysis of DNA binding by highly purified E2 proteins. So that a more stringent determination of relative affinities and off-rates could be made, a purification scheme was designed to isolate highly purified ₁₆E2sct, ₅₁E2sct, and GST-BPV₁E2sct.

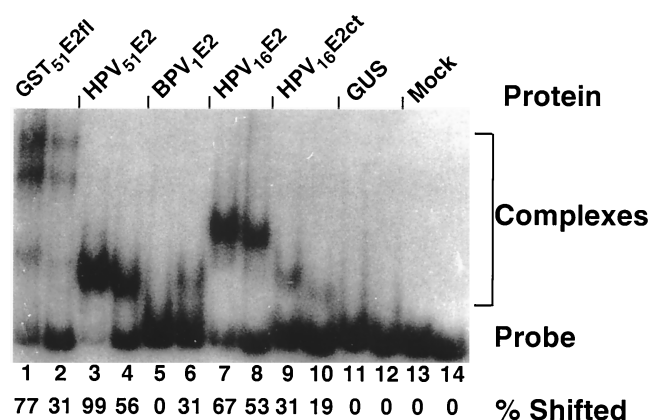


FIG. 3. EMSA of multiple E2 proteins. Two clones from the GST-HPV₅₁E2fl CASTing experiment were used to make radiolabeled probes. They represent the GST-HPV₅₁E2fl CASTing consensus sequence (ACACAAATCGGT, odd numbered lanes) and a high-affinity binding site similar to those previously described for other E2 proteins (ACCGAATATGGT, even numbered lanes). Below each lane number is the percentage of the total probe shifted by each E2 protein. These numbers were calculated from the scanned autoradiogram using ImageQuant software to calculate pixel densities for each band. Lanes 1 and 2 represent complexes formed between these probes and the bacterially expressed GST-HPV₅₁E2fl proteins used for the CASTing experiment. Lanes 3 through 12 all contained nuclear extracts from baculovirus-infected insect cells. Lanes 3 and 4 are binding reactions with the full-length E2 protein of HPV51. Lanes 5 and 6 are binding reactions with full-length E2 protein of BPV1. Lanes 7 and 8 are binding reactions with full-length E2 protein of HPV16. Lanes 9 and 10 are binding reactions with the C-terminal portion of the HPV₁₆E2ct protein. Lanes 11 and 12 contain extracts from insect cells infected with a baculovirus expressing β -glucuronidase. Lanes 13 and 14 contain the two probes loaded without exposure to any protein extracts.

Attempts to cleave the GST portion of the GST-BPV₁E2sct protein with factor Xa resulted in loss of all DNA binding activity (data not shown). Therefore, factor Xa cleavage was not performed on the GST-BPV₁E2sct protein. The GST-E2sct constructs used above were expressed in bacteria and purified as described in Materials and Methods. Fractions were screened for activity by EMSA and examined for the level of purity by sodium dodecyl sulfate (SDS)-PAGE (Fig. 4A). The active, highly purified fractions were pooled and used for all further work defining the relative affinities and off-rates for these proteins with selected DNA binding sites. The E2 proteins used in this study were of comparable length, and this choice was based on the results of Pepinsky et al., who demonstrated a correlation between the length of BPV1 E2 C-terminal constructs and their binding affinity (25).

Relative affinities and off-rates for E2BSs. To further confirm the CASTing results, we constructed a novel set of binding site clones with defined nucleotide differences based on the consensus E2BSs determined from the GST-HPV₅₁E2fl and GST-BPV₁E2fl CASTing experiments. Figures 4B and C are representative EMSAs used to define the relative affinities and off-rates for these protein-DNA complexes. Table 2 lists the binding sites utilized and the EMSA results for HPV₁₆E2sct, HPV₅₁E2sct, and GST-BPV₁E2sct. The relative affinities suggested by the frequency with which a sequence was identified in the GST-HPV₅₁E2fl CASTing experiment correlate with the relative affinities of both HPV E2sct proteins in the EMSAs.

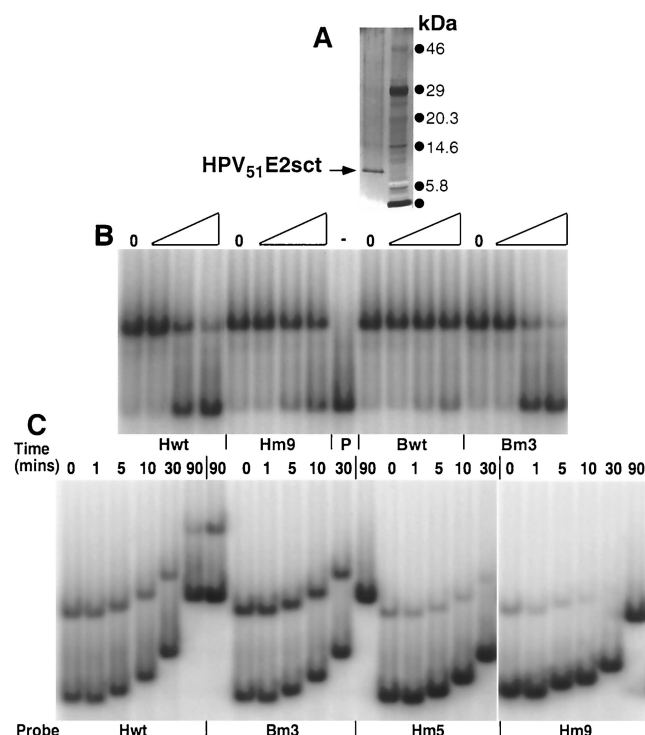


FIG. 4. Determination of the K_{rel} and $T_{1/2}$ for the HPV₅₁E2sct protein with various DNA binding sites. (A) Silver-stained 4 to 12% gradient SDS-PAGE of the highly purified HPV₅₁E2sct protein preparation (lane 1). Proteins in lane 2 are prestained molecular size markers. (B) EMSAs to determine the relative affinity of the E2 protein for various DNA probes. Reaction mixtures were made with the HPV₅₁E2sct protein and equal amounts of radiolabeled Bm#3 probe DNA with no competitor DNAs (lanes 0) or with sequential dilutions corresponding to a 2-, 100-, and 500-fold excess of the PCR-amplified unlabeled competitor DNA listed below the EMSAs. Binding was for 30 min before electrophoresis. Gels were dried, exposed to PhosphorImager screens, and analyzed using ImageQuant software. (C) EMSAs to determine the rate of E2 protein-DNA dissociation. The HPV₅₁E2sct protein was reacted for 30 min with the probe DNAs identified below each section of the EMSA. Then, a 1,000-fold excess of unlabeled competitor DNA, Bm#3, was added to each reaction mixture. Lanes 0 were loaded immediately, while others were allowed to incubate for 1, 5, 10, 30, or 90 min before electrophoresis. Gels were dried, exposed to PhosphorImager screens, and analyzed using ImageQuant software.

Changes in the most highly conserved nucleotides, the first and last two base pairs, result in abrogation of binding (Table 2, compare Hwt with Hm#7). HPV E2 proteins prefer A/T base pairs in the core of their recognition sites for (Table 2 compare Hwt with Hm#9 and Bm#3 with Bwt). The HPV E2sct proteins bound both the CASTing consensus and canonical E2BSs with very similar affinities (Table 2, compare Hwt with Bm#3). This result does not directly correlate with the GST-HPV₅₁E2f CASTing experiment because the canonical E2BS was identified with a much lower frequency than the novel E2BS (see Fig. 2B). A novel palindromic sequence, ACACAAATGTGT (Table 2, Hm#5), was also bound by the HPV E2sct proteins with relative affinities that were lower but still within the same order of magnitude as the CASTing consensus and canonical E2 binding sites (Table 2, compare Hm#5 with Hwt and Bm#3).

As expected, GST-BPV₁E2sct proteins bind with significantly lower K_{rel} s to the novel site (Table 2, Hwt, Hm#5, and Hm#9). The set of cloned E2BS variants described here confirms that binding to the novel E2BS sequence is a property common to the HPV16 and HPV51 E2 proteins but not to the BPV1 E2 protein.

As reported previously (42), the off-rate for BPV1 E2 is much faster than for HPV E2 proteins. The GST-BPV₁E2sct protein has $T_{1/2}$ values of less than 1 min for each of the binding sites analyzed here (Table 2, Bwt and Bm#3), whereas the HPV proteins formed more stable interactions with the sequences that they recognized.

Novel DNA binding sites are located throughout HPV genomes. As a first step in addressing if these novel E2BSs may have biological relevance, we searched the papillomavirus genome database (21) for the occurrence of ACAC-N₅-GGT, ACC-N₅-GTGT, and ACAC-N₄-GTGT. Most papillomavirus genomes contain one or more of these sites in regions where E2BSs are commonly found. Table 3 identifies the locations of these sites in selected papillomavirus genomes. The HPV51 genome has two such novel high-affinity sites within the LCR proximal to the putative E1BS. The site ACCGATTTGTGT (Table 3, column LCR [*ori/E*]) closest to the E1BS is identical to the HPV51 E2 CASTing consensus sequence (Fig. 3). For other HPVs that infect mucosal epithelium (e.g., HPV11 and HPV18), the analogous E2BS is involved in initiation of replication (E2BS 3) (3, 4, 31, 36), suggesting that this novel site may serve a similar function in replication of HPV51.

TABLE 2. Relative affinity (K_{rel}) and off-rates ($T_{1/2}$) for papillomavirus E2sct proteins^a

Clone	Sequence	HPV ₁₆ E2sct		HPV ₅₁ E2sct		GST-B ₁ E2sct	
		K_{rel}	$T_{(1/2)}$	K_{rel}	$T_{(1/2)}$	K_{rel}	$T_{(1/2)}$
Hwt	cagaACACAAATCGGTtccc	1.00	>90	1.00	61	<0.01	ND
Hm#5	cagaACACAAATGTGTtccc	0.20	<1	0.08	8	ND	ND
Hm#7	cagaGTACAAATCGA C tccc	<0.01	ND	<0.01	ND	ND	ND
Hm#9	cagaACACGGATCGGTtccc	<0.01	ND	0.03	2	0.04	ND
Bwt	cagaACCGGGATCGGTtccc	<0.01	ND	<0.01	ND	1.00	<1
Bm#3	cagaACCGAAATCGGTtccc	1.20	>90	1.60	87	1.54	<1

^a K_{rel} s were calculated based on the amount of unlabeled competitor clone DNA necessary to compete away 10,000 cpm (~1 nM) of radiolabeled probe. Radiolabeled Bm#3 was used as the probe for HPV E2s, and Bwt was used for BPV1E2. For each experiment, the results were normalized to the values of Hwt for the HPV E2s and to Bwt for GST-BPV₁E2sct, the binding affinities of which were defined as 1.00. Results listed represent the average of at least three experiments. ND not determined. $T_{1/2}$ is equal to the amount of time (in minutes) that a 500- to 1,000-fold excess of unlabeled competitor DNA requires to compete away 50% of a 10,000 cpm (~1 nM) radiolabeled clone probe (identified in the Clone column) in at least three experiments. Unlabeled competitors were PCR amplified from pBm#3 (for HPV E2s) and pBwt (for BPV1 E2) templates as detailed in Materials and Methods.

TABLE 3. Locations of E2BSs in papillomavirus genomes^a

Papillo- mavirus	Sequence type and position					
	LCR	LCR(<i>ori/E</i>)	E2	L2	L1	Other
HPV51	B, 7391 H ₃ , 7742	H ₂ , 7760 B, 30 B, 46	H ₂ , 3754	B, 4497	H ₁ , 5798	
HPV16	B, 7457	B, 7860 B, 35 B, 50			H ₃ , 6102 B*, 6106 H ₁ , 6522 B*, 6565	
HPV11	B, 7592	B, 7890 B, 35 B, 49		B, 4767		B, 585/E7 B, 2514/E1
HPV8	B, 7401	B, 7535 B, 22	B, 3048 B, 3509	B, 4455 H ₁ , 5337	H ₂ , 6186 B, 7367	
BPV1	B, 7203 B, 7365 B, 7408 B*, 7459 B, 7510 B, 7591 B, 7620 B, 7634 B, 7760	B, 7896 B*, 16	B, 3088 B*, 2921	H ₁ , 5324		B*, 855/E1/E7 B*, 1125/E1 B*, 2396/E1

^a Listed are all of the potential E2BSs from selected genomes based on the sequences identified previously (canonical E2BS, B = ACC-N₆-GGT; B* identifies binding sites that differ from the canonical site yet were still bound by BPV1 E2 [15] or in this study (novel human E2BSs, H₁ = ACAC-N₅-GGT, H₂ = ACC-N₅-GTGT, and H₃ = ACAC-N₄-GTGT). The numbers correspond to the first 5 dA of the BS sequence, with nucleotide numbering corresponding to that listed in the Papillomavirus Compendium (22). These viruses represent different papillomavirus classes. HPV 51, -16, and -11 are intermediate-risk, high-risk, and low-risk mucosal types, respectively. HPV8 infects human cutaneous epithelium, and BPV infects bovine cutaneous epithelium. The regions of the papillomavirus genome where the E2BSs are found are listed at the top of the table; LCR *ori/E* refers to the overlapping origin of replication and early promoter in the LCR, and the other designations identify regions within the rest of the LCR or specific ORFs.

Among HPVs, the novel E2BSs are most commonly found within the L1 ORF (Table 3, column L1). The proteins encoded by the two papillomavirus late genes L1 and L2 form the viral capsid. No studies published to date have directly defined a function for E2BSs within these late ORFs. The novel E2BS DNA sequences within L1, ACA CCT AGT GGT, encode the conserved amino acid sequence Thr-Pro-Ser-Gly (21). Therefore, we cannot rule out the possibility that these DNA sequences may also be conserved because of their participation in capsid structure or for receptor interactions.

Construction of the pOri16L reporter plasmids. We next asked if the novel E2BS could functionally substitute for the canonical sites. The E2 protein is involved in the regulation of both papillomavirus gene transcription and genome replication. By substituting the novel E2BS for the canonical site in a system that allows assay of E2-mediated effects, we can assess the functional role of the sites. The well-characterized HPV type 16 genome was used for these experiments. To this end, the pOri16L reporter was constructed; it contains the 3' portion of the HPV16 LCR that includes the DNA origin of replication and the overlapping early promoter structure that contains three canonical E2BSs fused to a luciferase reporter. All three sites are involved in the HPV16 E2 protein-mediated modulation of viral replication and transcription (4, 26, 32).

Figure 5A presents a schematic of the HPV16 LCR from which the HPV16 LCR portion of pOri16L originates. Below it is

another schematic representing the pOri16L construct, with the relevant features of the plasmid highlighted, including each of the wild-type E2BSs within the construct and the E1BS and A/T-rich sequence that are required for replication of virus DNA. In addition, two cellular transcription factor binding sites (Sp1 and the TATA box) that are known to facilitate the initiation of transcription from this promoter are present in this portion of the HPV16 LCR (27, 37).

To determine if the novel human E2BS has biological activity in the context of the LCR, each of the wild-type E2BSs (BS1, BS2, and BS3) in the pOri16L plasmid were knocked out (KO) or replaced with the novel E2BS by site-directed mutagenesis (Fig. 5B). In addition, single-nucleotide substitutions for the BS3 and BS1 E2BSs were also made to help delineate which nucleotide changes resulted in phenotypic changes in plasmid activity (Fig. 5B, BS3-t→a and BS1-c→t). Because the novel site is not a perfect palindrome, it was placed in the

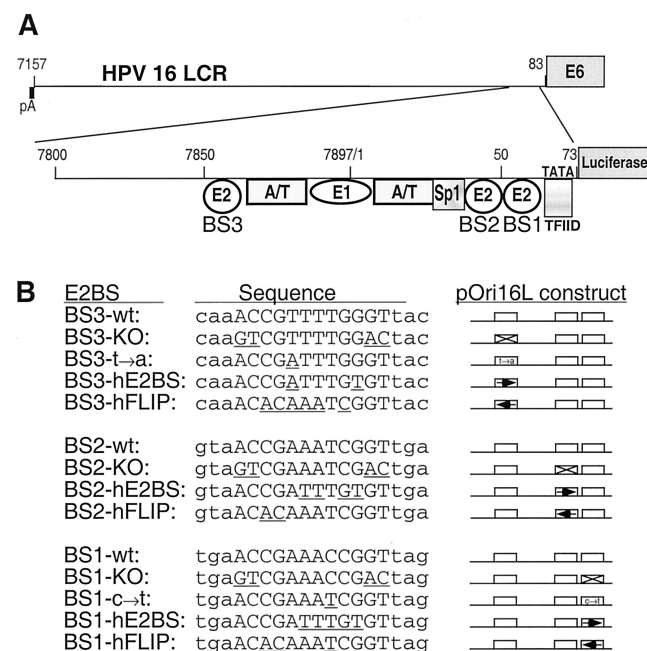


FIG. 5. pOri16L constructs. A fragment of the LCR that contains three of the four E2BSs from the HPV16 LCR (A) was fused to the luciferase gene. The three E2BSs are labeled and represented by the circles below the LCR (BS1 to BS3). Adjacent to BS2 is an Sp1 site, and adjacent to BS1 is the early-gene TATA box. Also contained within this portion of the LCR is the viral origin of replication (*ori*). The minimal viral *ori* is defined by the proximal E1BS and E2BS and their neighboring A/T-rich regions. All of these LCR E2BSs stimulate viral DNA replication. (B) The wild-type and modified sequences for each of the three E2BSs of the pOri16L constructs. The names correlate with either the single-nucleotide-transition mutations or the orientation of the novel E2BS that is created within each mutated pOri16L construct. Within the sequences, lowercase lettering corresponds to flanking nucleotides, uppercase letters correspond to the 12-bp E2BS, and underlined letters correspond to those nucleotides that differ from the wild type for each site. At the right is a schematic of each construct. Open boxes are wild-type E2BSs; an X through the box corresponds to a knocked-out binding site; transition mutations are specified within the E2BS box; and large arrows within a box correspond to the novel E2BS in either the forward or reverse orientation.

TABLE 4. Summary of relative affinities (K_{rel}) and off-rates ($T_{1/2}$) for the HPV E2sct proteins with DNA sites

Clone	Sequence	K_{rel}^a	$T_{1/2}$ (min) ^b
BS3-wt	cagaACCGTTTTGGGTtccc	0.09	>90
BS3-t→a	cagaACCGATTTTGGGTtccc	0.11	>90
BS3-g→t	cagaACCGTTTTGTGTtccc	0.13	>90
BS1-wt	cagaACCGAAACCGGTtccc	0.22	>90
BS2-wt/BS1-c→t	cagaACCGAAATCGGTtccc	1.00	>90
hE2BS	cagaACCGATTGTGTtccc	0.24	>90
hFLIP	cagaACACAAATCGGTtccc	0.78	>90
hE2BS-KO	cagaGTACAAATCGA <u>t</u> ccc	<0.01	N.D.

^a K_{rel} s were calculated based on the amount of unlabeled competitor DNA (Clone column) necessary to compete away 10,000 cpm (~1 nM) of radiolabeled BS2-wt probe in at least three experiments.

^b $T_{1/2}$ is equal to the amount of time (minutes) that a 1,000-fold excess of unlabeled BS2-wt competitor DNA requires to compete away 50% of a 10,000 cpm (~1 nM) radiolabeled clone probe (Clone column) in at least three experiments. N.D., not determined. Within the sequences, lowercase lettering corresponds to flanking nucleotides, uppercase letters correspond to the 12-bp E2BS, and the underlined letters correspond to those nucleotides which differ from the wild type for each site.

pOri16L plasmid in both orientations to determine if this would affect the function of E2 (Fig. 5B, hE2BS [the hE2BS orientation found at the HPV 51 *ori*; see Table 3] and hFLIP). The mutated pOri16L plasmids are all named for the mutation(s) made within their E2BS(s).

Relative affinities and off-rates of the pOri16L E2BSs. Relative affinity and stability studies with the HPV₁₆E2sct protein and the pOri16L construct E2BSs were performed to analyze the effects of the sequence changes on two important biochemical parameters. Table 4 lists the results of these HPV16 E2 binding affinity and complex stability analyses.

The single-nucleotide differences in the BS3-g→t and BS3-t→a clones do not appreciably alter the binding affinity of the HPV₁₆E2sct protein for these targets compared with the BS3-wt site (Table 4). When both of these nucleotides are altered, hE2BS, a member of the novel E2BS family, is created. The relative affinity of E2 for this binding site is 2.5-fold greater than it is for BS3-wt. Comparisons between the Rep-wt and hFLIP sites, which differ by 6 bp, show an approximately 8.5-fold-higher relative binding affinity for hFLIP. These two comparisons reveal that substitution of the novel E2BS in either orientation for the Rep-wt site results in an increase in the binding affinity of the E2 protein for that site.

The affinity of E2 for the BS1-wt site is increased by approximately 4.5-fold when a single-nucleotide change was made that converts it to the BS2 core sequence (BS2-wt/BS1-c→t clone). Changing two nucleotides to create hFLIP results in a 3.5-fold-higher affinity compared with BS1-wt (Table 4). There is no appreciable difference in the relative binding affinity of E2 for the BS1-wt and the hE2BS site despite the fact that 5 of the 12 bp differ between these two E2BSs.

E2 had the highest relative affinity for the BS2-wt/BS1-c→t site. Its relative affinity for the novel E2BSs hFLIP and hE2BS was 78 and 22% of that for the BS2-wt/BS1-c→t site respectively (Table 4). Alteration of the first two and last two nucleotides of the 12-bp E2BS, as in hE2BS-KO, eliminated any detectable binding in these assays.

There is an apparent threefold difference in the relative

binding affinities, depending on the binding site orientation (hE2BS versus hFLIP), that correlates with the differences in the flanking sequences. Despite this, except for the hE2BS-KO, all binding affinities for this collection of DNA sequences are within the same order of magnitude. Considering that K_d s for the HPV16 E2 protein binding to its recognition sites are in the range of 10^{-10} to 10^{-11} M (33), all of the interactions described above are very strong. Thain et al. (40) have determined the K_d s for each of the E2BSs within the HPV16 LCR in the context of their wild-type flanking sequences. The results reported here agree with both of those studies, as the binding sites ranked from lowest affinity to highest affinity are BS3-E2BS, BS1-E2BS, and BS2-E2BS (Table 4).

Off-rates of the protein-DNA complexes were also determined for each of the E2BSs (Table 4). The relative stabilities ($T_{1/2}$) of the complexes formed with the tested E2BS variations are indistinguishable in these assays.

Effects of site substitutions on basal promoter activity. This experiment was designed to determine if the novel E2BS can functionally substitute for the wild-type E2BSs within pOri16L in either orientation. Previous studies have shown that the HPV16 E2 protein affects expression from the early promoter at the level of transcription (26). Therefore, a luciferase cassette driven by the pOri16L constructs was used as a reporter to monitor the effects of the E2 protein on transcription from the mutated HPV16 promoters. These assays were first performed in the absence of the E2 protein to determine what effects the sequence alterations might have on the basal activity

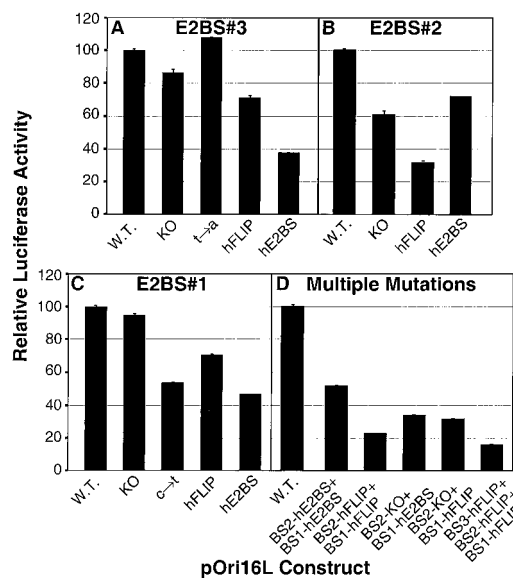


FIG. 6. Basal expression levels from the pOri16L plasmids. SCC-13 cells were transfected with 0.05 μ g of pRL-TK and 0.5 μ g of pCMV₄-XS plus 1 μ g of each pOri16L template. Dual luciferase assays were performed on cell extracts prepared at 36 h posttransfection. The luciferase activity for each cell lysate was expressed as the ratio between the firefly and *Renilla* luciferase activities and then normalized to the pOri16L wild-type (W.T.) levels (100%) for comparisons between experiments. The average results from three independent experiments are plotted for each set of templates. (A) E2BS#3; (B) E2BS#2; (C) E2BS#1; (D) constructs with alterations to more than one E2BS. Error bars correspond to the standard deviation for each data set.

of the HPV16 early promoter. The pOri16L wild-type and mutant plasmids were introduced into SCC-13 cells along with a reference plasmid and pRL-TK (see Materials and Methods and Fig. 6 legend for details).

Figure 6A shows the effects on promoter activity when changes were made in E2BS#3. Alteration of the first and last two nucleotides of the 12-bp palindrome in BS3-KO (Fig. 5B), which are critical for E2 protein-E2BS contact (8, 15), had little effect on the basal activity of the promoter (Fig. 6A). Similarly, changing a single nucleotide from within the core 4 bp of the E2BS, as in BS3-t→a (Fig. 5B), also had little effect on expression levels (Fig. 6A). However, the 2-bp substitution that creates BS3-hE2BS (Fig. 5B) decreased the accumulation of luciferase activity by 60% (Fig. 6A). Creation of BS3-hFLIP results in a more extensive sequence change from the wild type (involving 6 of the core 12 bp of the E2BS; Fig. 5B) but only reduces the promoter activity by 30% (Fig. 6A). There are no known cellular transcription factor binding sites that directly overlap E2BS#3, but the reductions in basal promoter activity suggest that the BS3-hE2BS and BS3-hFLIP sequence changes interfere with some aspect of the gene expression process.

Figure 6B details the effects of changes within E2BS#2 (Fig. 5B). The changes to the first and last 2 bp of the 12-bp E2BS found in clone BS2-KO reduced the level of luciferase expression by about 40%. The BS2-hE2BS mutant clone, which has two substitutions in the core and two in the 3' portion of the palindrome (Fig. 5B), has the least effect of the E2BS#2 mutants on basal-level expression (Fig. 6B). In contrast, the BS2-hFLIP mutant which has two substitutions in the 5' portion of the palindrome (Fig. 5B), has the lowest basal expression levels (Fig. 6B). These alterations to E2BS#2's sequence could influence binding of the cellular transcription factor Sp1 to its adjacent binding site. In the absence of E2, Sp1 activates expression from this HPV early-gene promoter by binding to its core recognition sequence, GGGCGT (7). It is known that sequences flanking the core recognition sequence of Sp1 affect its ability to bind DNA (7), but no Sp1 binding analysis has been performed with the mutants that we have made.

Figure 6C details the effects of changes within E2BS#1 (Fig. 5B). Substitution with the BS1-KO has very little effect on basal expression. However, when a single dC-to-dT transition is made within the core 4 bp of the 12-bp E2BS (Fig. 5B, BS1-c→t), luciferase expression is reduced by 50%. More extensive mutations, such as the 4-bp changes made to create BS1-hE2BS (Fig. 5B), result in a similarly reduced level of luciferase expression (Fig. 6C). These results suggest that the dC→dT transition affected the basal activity of the promoter. However, when this mutation is combined with alterations to the third and fourth nucleotides of this E2BS to create the BS1-hFLIP clone (Fig. 5B), basal promoter activity is partially restored (Fig. 6C). There are no known transcription factor binding sites that overlap E2BS#1 whose interruption or alteration could explain the effects of these sequence changes on basal promoter activity. The TATA box is spaced only 3 bp from E2BS#1 in the HPV16 LCR and pOri16L constructs (Fig. 5A). It would not be surprising if any or all of the mutations that we made to E2BS#1 had effects on the interaction of the transcription initiation complex with this region of the promoter.

Comparison of the results of mutations to E2BS#3 and

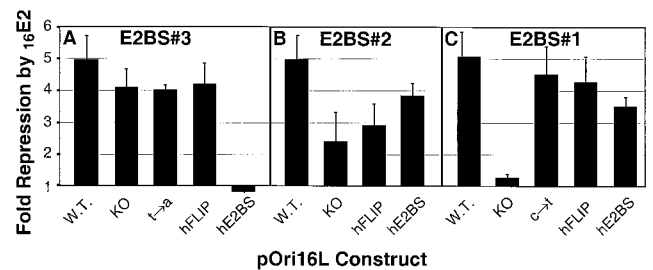


FIG. 7. pOri16L repression by the HPV16E2 protein. SCC-13 cells were transfected with 0.05 μ g of pRL-TK, 1 μ g of each pOri16L template, and either 0.5 μ g of pCMV₄-XS or 0.5 μ g of pCMV-E2₁₆. Dual luciferase assays were performed on cell extracts prepared at 36 h posttransfection. The luciferase activity for each cell lysate was expressed as the ratio between the firefly and *Renilla* luciferase activities. The levels of repression were calculated by dividing the basal promoter activity (transfections with the pCMV₄-XS plasmid) by the promoter activity in the presence of the E2 protein. The results from three independent experiments are plotted for each set of templates: (A) E2BS#3; (B) E2BS#2; and (C) E2BS#1. Error bars correspond to the standard deviation for each data set.

E2BS#2 reveals that substitution with the novel E2BS in a particular orientation does not cause an equivalent decrease in levels of basal promoter activity (Fig. 6A and B, compare BS3-hE2BS and BS3-hFLIP versus BS2-hE2BS and BS2-hFLIP).

We also intended to determine if the novel E2BSs might affect the way that E2 cooperatively mediates repression by making substitutions to more than one of the three E2BSs found in the pOri16L construct. However, changes to multiple E2BSs, in the combinations that we tested, resulted in severely reduced levels of basal expression compared to the wild type (Fig. 6D). Lewis et al. (14) described a similar phenomenon when making mutations to multiple E2BSs within plasmids containing the full HPV16 LCR driving a luciferase reporter. This suggests that cellular transcription factors interact with many sequences overlapping the E2BSs within the viral LCR even in the absence of viral proteins. Because of these results, clones with mutations to multiple E2BSs were not used for subsequent analyses.

Novel E2BS functionally substitutes for wild-type E2BSs in transient transcription assays. The E2 protein can repress transcription from the HPV early promoter (30). Therefore, an HPV16 E2 protein expression construct was cotransfected into SCC-13 cells along with each of the pOri16L mutants and the reference plasmid pRL-TK to determine whether the novel E2BS could functionally substitute for the wild-type sites in an E2-responsive manner.

Although many of the E2BS mutations resulted in reduced levels of basal expression from this cassette, the activities of all the promoters were still high enough to assay for repression by the HPV16 E2 protein.

The HPV16 E2 protein repressed luciferase expression from the BS3-KO, BS3-t→a, and BS3-hFLIP plasmids almost as efficiently as it did from the wild-type promoter (Fig. 7A). In stark contrast, expression from the BS3-hE2BS reporter, which had only 40% of the basal luciferase activity of the wild-type promoter (Fig. 6A), was stimulated slightly by E2 (Fig. 7A). There is no precedent for alteration of the E2 function from a repressor to an activator of early promoter activity when only

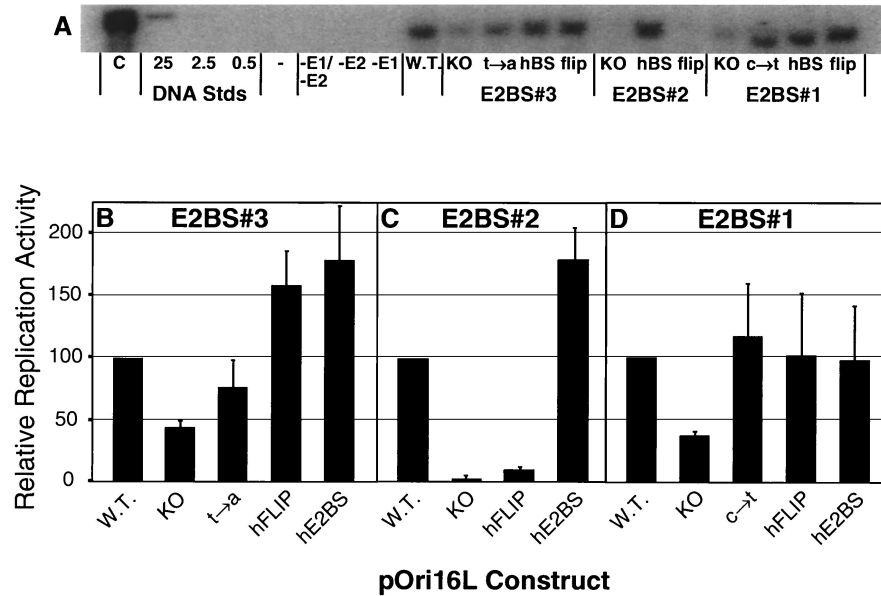


FIG. 8. Transient-replication assays with pOri16L mutants. (A) Representative Southern blot prepared by electrophoresis of Hirt DNA digested with both *AlwNI* and *DpnI*, transferred to a nylon membrane, and visualized with radiolabeled probe made using pOri16L as the template. In the left lane is one-eighth of the wild-type sample that was digested only with *AlwNI*. The next three lanes contain dilutions of DNA standards (in picograms) to control for hybridization efficiency between blots. (Because all cells are not efficiently transfected by the electroporation procedure, these standards cannot be used to determine pOri16L copy numbers replicated per cell) The lanes labeled -E1/-E2, -E2, and -E1 were transfected with wild-type (W.T.) pOri16L alone or with the addition of only the E1 or E2 expression plasmid respectively. The remaining lanes are labeled with the name of the pOri16L clone (3 μ g) that was transfected into SCC-13 cells along with the E1 and E2 expression plasmids (in equal molar amounts; 3 μ g of pOri16L plus 3.3 μ g of pCMV-E2₁₆ plus 3.8 μ g of pCMV-E1₁₆). Southern blots were exposed to PhosphorImager screens and analyzed using ImageQuant software. Shown are the relative replication activities from three independent transfections of pOri16L templates with mutations to their (B) E2BS#3, (C) E2BS#2, or (D) E2BS#1. Error bars correspond to the standard deviation for each data set.

a 2-bp change in an E2BS sequence is made. A similarly binding-site orientation-dependent effect was observed for E2 protein function in transient-replication assays (see below). We propose that E2's functional dependence on the orientation of the novel E2BS indicates that the E2 protein may asymmetrically bind to the nonpalindromic, novel E2BS and that this profoundly affects how it is able to interact with other proteins.

Knocking out the BS2 E2BS results in a greater than twofold reduction in the ability of the HPV16 E2 protein to repress luciferase expression from this promoter (Fig. 7B). Substitution of the BS2 E2BS with the novel E2BS in the forward or reverse orientation partially restores E2 mediated repression (Fig. 7B).

Substitution of E2BS#1 with the BS1-KO sequence virtually eliminates repression of this promoter by E2 (Fig. 7C). In contrast, the presence of the novel E2BS sequence in either orientation (Fig. 5B, BS1-hE2BS and BS1-hFLIP) restores E2 repression (Fig. 7C). Repression occurs even though the basal levels of expression from these mutated promoters were reduced compared to the wild type (Fig. 6C). These E2BS#1 mutants demonstrate that the novel E2BS can functionally substitute for the wild-type site in either orientation at this site (Fig. 7C).

These luciferase expression studies demonstrate that the novel E2BS can functionally substitute for the wild-type E2BSs in the context of transient transcription at each of the three HPV16 LCR E2BSs studied here. However, in the case of E2BS#3, functional substitution depends on the binding site orientation. Thus, the ability of the E2 protein to interact with

cellular proteins can be affected by the orientation of the novel E2BS.

Novel E2BS functionally substitutes for the wild-type E2BSs in transient-replication assays. Another major function of the E2 protein in the viral life cycle is the stimulation of DNA replication in conjunction with the papillomavirus E1 protein. To determine if the novel E2BS can functionally substitute for the canonical E2BSs in E2-mediated DNA replication, we used the pOri16L mutants in transient-replication assays. The pOri16L constructs were designed to be analogous to plasmids used in other studies of HPV replication (32, 36). Each of the E2BSs contained in the pOri16L plasmid is known to influence the efficiency of replication. Transient-replication assays were performed as described for HPV31 (12, 28, 32).

There is no detectable replication in the absence of E1-and/or E2-expressing plasmids (Fig. 8A, -E1/-E2, -E2, and -E1) (32). In addition, various amounts of E1 and E2 expression plasmids relative to the pOri16L plasmid were tested to determine if they had any effect on transient replication. Replication of pOri16L was more readily detected when the E1 and E2 expression plasmids were transfected in molar excess to the pOri16L target. However, regardless of what the ratios of pOri16L mutant to E1 and E2 expression plasmids were, the relative replication activities between the mutant pOri 16L targets remained unaffected. These results agree with those of Sakai et al. (32). Figure 8A is a representative Southern blot from one of the three experiments used to determine the replication activities summarized in Fig. 8B to D.

Figure 8B shows the effect of changes to sequences in

E2BS#3 on DNA replication. Elimination of the E2 binding activity at this site reduces replication levels by about 50% (BS3-KO). The substitution of a single dA for the wild-type dT within the central four nucleotides of E2BS#3 (Fig. 5B, BS3-t→a) has little effect on pOri16L replication (Fig. 8B, BS3-t→a). In contrast, substitution with the novel E2BS sequence (Fig. 5B, BS3-hE2BS) results in an almost twofold increase in DNA replication. A similar increase in replication efficiency is detected when the novel E2BS is placed in the opposite orientation (Fig. 8B, BS3-hFLIP). Therefore, the novel E2BS can substitute in either orientation for E2BS#3 and it enhances E2's stimulation of transient replication activity 1.5- to 2-fold. This increase in replication efficiency correlates with an increase in relative affinity as detected by EMSAs (Table 4).

Figure 8C shows the effect of changes to sequences in E2BS#2 (see Fig. 5B for sequence details). Elimination of the E2 binding activity at this site virtually eliminated detectable levels of transient replication (Fig. 8C, BS2-KO). Substitution of this site with the novel E2BS stimulates replication levels twofold (Fig. 8C, BS2-hE2BS), as it did for the BS3-wt to BS3-hE2BS and BS3-hFLIP changes (Fig. 8B). However, placement of the novel E2BS at E2BS#2 in the opposite orientation abrogates replication (Fig. 8B, BS2-hFLIP). This result contrasts with those from the expression assay, where the HPV16E2 protein repressed expression from the BS2-KO, BS2-hE2BS, and BS2-hFLIP plasmids to similar levels (Fig. 7B). Thus, at this position in the LCR, the orientation of the novel E2BS influences E2-mediated replication. This orientation dependence may reflect changes in the conformation of the E2 protein-DNA complexes that form at this site with respect to the replication machinery. We detected a threefold effect of orientation on binding affinities in our EMSAs. These binding affinity differences between the BS2-hE2BS and BS2-hFLIP site, in the context of the pOri16L plasmid, could explain the differences in their replication efficiencies.

Figure 8D shows the effect of changes to sequences in E2BS#1. The replication capacity of the BS1-KO mutant template is only 40% of the wild-type template (Fig. 8D, BS1-KO). In contrast, substitution for E2BS#1 with the BS1-c→t or the novel E2BS in either orientation (BS1-hE2BS and BS1-hFLIP) has little effect on the replication efficiency (Fig. 8D). Therefore, in contrast to its activity when used to replace E2BS#2, the novel E2BS can functionally substitute in either orientation for the wild-type E2BS#1.

These replication studies demonstrate that the novel E2BS can functionally substitute for the wild type E2BSs in the context of transient replication at each of the three HPV16LCR E2BSs studied here, but in the case of the E2BS#2, functional substitution depends on orientation. These results, taken together with the similar orientation dependence seen in our luciferase expression assays for the E2BS#3 site, suggest that the ability of the E2 protein to interact with cellular proteins can be affected by the orientation of the novel E2BS.

DISCUSSION

CASTing was used to identify a novel family of HPV E2BSs whose members are bound with affinities similar to that of the canonical E2BS. The CASTing results reveal the promiscuous nature of the HPV E2 protein with respect to DNA binding

site selection. Our results allude to the potential flexibility of the HPV E2 protein's conformation upon binding to the E2BSs, as these proteins can bind to either the CASTing consensus (ACAC-N₄-CGGT) or canonical E2BS (ACCG-N₄-CGGT) with similar relative affinities if their core and flanking nucleotides are conserved (Table 2, Hwt versus Bm#3). These novel sites are present in HPV genomes at locations where E2BSs are commonly found.

To show that the novel site also has biological significance, we substituted it for each of three of the wild-type canonical sites within the HPV16 LCR (Fig. 5). We demonstrated that in both transient-transcription assays (Fig. 7) and transient-replication assays (Fig. 8), substitution for the wild-type E2BSs with the novel E2BSs had, in some instances, very strong and unpredicted effects on the ability of the HPV16 E2 protein to modulate transcription and replication. In the case of BS3-hE2BS, reporter expression was activated rather than repressed. For BS2, replication was only supported when the site was replaced with the novel hE2BS in one orientation. Thus, binding to this novel, asymmetric site affects the HPV16 E2 protein's activities in an orientation-dependent manner.

Binding properties of HPV E2 proteins. The CASTing experiment (Fig. 2) and relative affinity studies (Tables 2 and 4) reveal that the HPV E2 proteins that we examined have a preference for an A/T-rich 4-bp core. There was a >100-fold difference in binding affinities of the HPV16 E2 protein for an A/T-rich core versus a core that contained only 2 G/C bp (e.g., Table 2, Hwt versus Hm#9).

In agreement with studies by Thain et al. (40), we noted a preference by the HPV E2 proteins for purine (R) and pyrimidine (Y) residues at the -1 positions flanking the binding site (Fig. 2A, 78% R adjacent to the ACAC and 92% Y adjacent to the CGGT). This property of the E2BSs is preserved throughout the HPV genomes. The LCRs of the mucosa-specific HPVs contain three E2BSs involved in the initiation of replication (4). In all of these sites, the flanking purine and pyrimidine residues are well conserved.

The novel E2BS was the most frequently isolated sequence in two independent CASTing experiments with GST-HPV₅₁E2fl. Yet, when the core and flanking sequences are conserved, the HPV E2 proteins bind the canonical palindromic sites (2) with a slightly higher affinity (e.g., Table 2, Hwt versus Bm#3). This paradox may reflect the sensitivity of the CASTing procedure to subtle differences in the binding affinities or stability of these protein-DNA complexes that were undetectable by EMSA.

Finally, EMSAs confirm that binding to the novel site is not a property shared by the BPV1 E2 protein (Table 2, Hwt and Hm#9).

Structural determinants for E2 binding site preferences. The molecular basis for the differences in the DNA binding activities of these proteins is not readily apparent. Cocystal structures have been published of only the BPV1 and HPV18 E2 proteins bound to canonical DNA binding sites (9, 13) (Fig. 9). If we make two assumptions, that the affinity for the novel hE2BS site and the specific protein-DNA contact points are conserved between the HPV16 and HPV18 E2 proteins, then the cocystal structures cannot explain the differences in binding properties between the HPV and BPV E2 proteins. The identity of all but one of the amino acids that contact the DNA is conserved among the E2 proteins used in this study. The

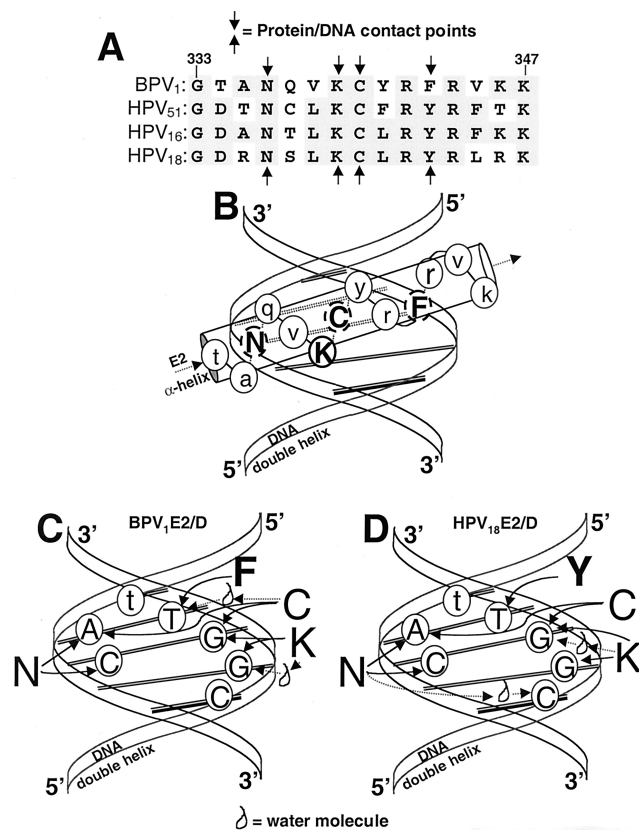


FIG. 9. BPV₁E2ct protein-DNA contacts. (A) The amino acid sequence of the α -helix from the BPV1E2 protein that contacts DNA is aligned with the homologous regions from the E2s of HPV51, -16, and -18. Numbers above the sequence corresponds to the BPV1 E2 protein. Amino acid residues conserved between these proteins are highlighted. Arrows above and below the alignment indicate amino acids that make contact with the bases in the DNA major groove of the E2BS and are presumed for HPV51 and HPV16 E2 (9, 13). (B) Schematic of the E2 protein α -helix contacting the DNA in the major groove. The amino acids that make contact with the DNA are capitalized and in boldface. The 5 and 3 ends of the DNA strands are indicated at the top and bottom of each strand. (C and D) DNA double helix with protein-contacting base pairs exposed for the BPV1E2 (C) and HPV18E2 (D) proteins. DNA-contacting amino acid residues are listed along the sides. Contact points are indicated by arrows between the letters representing the amino acids and nucleotides. Those contacts that are mediated by a water molecule are indicated by the waterdrop symbol.

single difference is a change from Phe343 in the BPV1 E2 protein to a Tyr in the corresponding site in the HPV E2 proteins (Fig. 9A). As both crystal structures indicate that these residues make comparable DNA contacts, this sequence difference cannot explain the differences in the binding properties of the BPV and HPV E2 proteins. In addition, this aromatic residue makes contact with the T at the 3' end of the DNA binding site (-GGT) that is absolutely conserved in all functional E2BSs examined to date.

The cocrystal structure of the E2 proteins bound to the canonical DNA binding site identified the amino acid-DNA contact points in the 16-bp sequence, caACCGAATTCGGTtg (9, 13) (arrows in Fig. 9A, C, and D). Figure 9B is a schematic representation of the results, showing the E2 protein α -helix oriented in the major groove of the DNA. Figures 9C and D

show the amino acids that make specific contacts with the DNA base pairs. The E2 proteins contain a conserved Lys residue that makes direct contact with the adjacent dG's in the DNA binding site (CGGT) (Fig. 9C and D). Because the Lys residue is conserved in all of the papillomavirus E2 proteins sequenced to date, binding to the novel sequences ACCG-N₄-GTGT and ACAC-N₄-GTGT by the HPV E2 proteins suggests a fundamental difference in the amino acid-DNA contacts and/or the positioning of the DNA-contacting α -helices.

The data available to us suggest that protein flexibility is an important determinant of sequence recognition. Nuclear magnetic resonance and X-ray crystallographic studies comparing the bound and free states of the homodimeric DNA binding domain of BPV1 E2 describe a conformational change upon binding to DNA. This change involves both the dimerization domain and the α -helix that contacts DNA (9, 47). It is plausible that HPV E2 proteins are more flexible than the BPV1 E2 protein. The fact that the HPV E2 proteins strongly prefer a flexible and/or prebent A/T-rich, 4-bp core within their 12-bp binding sites also supports the concept that the HPV E2 proteins can undergo a greater conformational change to achieve a stable complex with DNA.

Structural analyses of an HPV E2 protein complexed with both the canonical and novel HPV E2 DNA binding sites would help to address these issues.

Effect of E2 protein affinity for its binding sites. Steger et al. (35) used in vitro transcription assays to demonstrate that the amount of HPV18 E2 can determine whether it acts as an activator or repressor of transcription from the HPV18 early promoter. When template containing the HPV18 LCR (with four E2BSs) is in excess, addition of the HPV18 E2 protein results in stimulation of transcription. By contrast, when the HPV18 E2 protein is in excess, transcription is repressed. Thus, E2 abundance may act as a switch to differentially control early-gene expression at distinct stages of the viral life cycle.

Transient-transfection assays in which the E2 protein is expressed from a strong promoter, as used in this study, resemble the situation where the E2 protein is in excess of the available E2BSs. Therefore, we would predict that if the new sites had affinities similar to those of the wild-type E2BS, E2 should repress early-gene expression. The ₁₆E2sct protein used in this study had similar or elevated affinities for all of the sites studied here, with the exception of the KO sites (Table 4). Despite this, the substitution of the BS3-hE2BS site in pOri16L resulted in a promoter that was weakly stimulated rather than repressed by the HPV16 E2 protein (Fig. 7A). The BS3-hE2BS construct retains wild-type E2BS#1 and E2BS#2, which E2 utilizes to repress transcription from this promoter. Even if the E2 protein is not expressed in excess, it does not have a higher affinity for the BS3-hE2BS site than for the E2BS#1 and E2BS#2 E2BSs (Table 4). Thus, the stimulatory effect of E2 binding to BS3-hE2BS overrides the repressive effects of E2 binding to E2BS#1 and E2BS#2. These data suggest that binding affinity alone is not sufficient to explain the effect of E2 protein interactions with an E2BS.

Mechanisms of E2 protein-mediated transcription repression. All three of the E2BSs within pOri16L facilitate E2-mediated repression of gene expression from this promoter (17). The spacing and relative locations of the E2BSs are well conserved, with respect to the TATA box for the early pro-

moter and with respect to the binding sites of the known cellular transcription factors and the viral E1 protein, among the mucosal HPVs. At two E2BSs, E2BS#1 and E2BS#2, the E2 protein can compete for binding with the cellular transcription factors TFIID and Sp1, respectively (5, 37, 38). These two factors are involved in the stimulation of early-gene expression from this promoter in the absence of the E2 protein.

The TATA box utilized for the initiation of early-gene expression is adjacent to E2BS#1. If the E2 protein is bound to E2BS#1, it will interfere with the binding of the transcription initiation complex that forms around the TFIID complex (11). Therefore, direct interference with the formation of the transcription machinery complex is a mechanism for E2-mediated repression of transcription. All of the mutated E2BS#1 sites studied here, which were bound by HPV16 E2 (Fig. 5B and Table 4, BS1-c→t, BS1-hE2BS, and BS1-hFLIP), maintained their ability to repress luciferase expression (Fig. 7C), demonstrating that the novel E2BS can functionally substitute for the canonical E2BS.

The mechanism of E2-mediated repression via E2BS#3 has not yet been characterized. There are no known cellular transcription factor binding sites that overlap with E2BS#3, although Lewis et al. (14) have described a protein complex (CEF-2) from HeLa cell extracts that interacts with some of the nucleotides within this site. Alternatively, E2 binding to E2BS#3 may merely enhance the repressive effects of E2 binding to E2BS#1 and E2BS#2. The marked differences in the response to the E2 protein of the pOri16L constructs BS3-wt and BS3-hFLIP versus BS3-hE2BS will require further characterization to elucidate how the E2 protein utilizes E2BS#3 to repress transcription.

E2-stimulated replication. Transient-replication assays were used in this study because they combine the relative ease of genetic manipulation with the ability to perform the assays in an epithelial cell. However, only low-level replication, corresponding to the maintenance stage of the viral life cycle, is detected under these conditions.

Substitution of BS2 with the novel hE2BS (Fig. 5B, BS2-hE2BS) stimulated replication levels above wild type (Fig. 8C, BS2-hE2BS). In contrast, when the orientation of the novel hE2BS was reversed (Fig. 5B, BS2-hFLIP), DNA replication was not detected (Fig. 8B, BS2-hFLIP). This indicates that functional substitution for wild-type E2BS#2 by the novel hE2BS is orientation dependent.

Transient-replication studies with BPV1 have shown that BPV1 E2's relative affinity for an E2BS correlates with its ability to stimulate replication (43). From relative affinities alone, we presume that both the hE2BS and hFLIP E2BSs should support E2-mediated transient replication (Table 4). The caveat is that HPV16 transient replication is much less robust than that seen for BPV1 (32, 43) (Fig. 8A). Because replication levels are low to begin with, we cannot rule out the possibility that the lower relative affinity of BS2-hFLIP than of BS2-hE2BS is responsible for the reduced transient-replication levels (Fig. 8C). Alternatively, there may be an orientation dependence of the E2BS that cannot be explained solely by differences in relative affinities created by flanking sequences. Because the novel E2BS is not a palindrome, E2 may bind asymmetrically to the DNA and/or bend the DNA asymmetrically. Asymmetrical binding may influence the ability of the E2

protein to functionally interact with the replication machinery, whereas distortion of the DNA bend may influence how neighboring replication factors interact with the DNA.

ACKNOWLEDGMENTS

We thank Lou Laimins and Walter Hubert for help in getting the replication assays up and running and for many helpful discussions. We thank Ken Alexander for helpful suggestions for purification of the E2 proteins. We also thank Carey Waldberger, Rashmi Hedge, Richard Mann, and Hamish Young for reading the manuscript and for insightful comments.

REFERENCES

- Alexander, K. A., and W. C. Phelps. 1996. A fluorescence anisotropy study of DNA binding by HPV-11 E2C protein: a hierarchy of E2-binding sites. *Biochemistry* **35**:9864–9872.
- Androphy, E. J., D. R. Lowy, and J. T. Schiller. 1987. Bovine papillomavirus E2 trans-activating gene product binds to specific sites in papillomavirus DNA. *Nature* **325**:70–73.
- Chiang, C. M., G. Dong, T. R. Broker, and L. T. Chow. 1992. Control of human papillomavirus type 11 origin of replication by the E2 family of transcription regulatory proteins. *J. Virol.* **66**:5224–5231.
- Chow, L. T., and T. R. Broker. 1994. Papillomavirus DNA replication. *Intervirology* **37**:150–158.
- Dong, G., T. R. Broker, and L. T. Chow. 1994. Human papillomavirus type 11 E2 proteins repress the homologous E6 promoter by interfering with the binding of host transcription factors to adjacent elements. *J. Virol.* **68**:1115–1127.
- Gillitzer, E., G. Chen, and A. Stenlund. 2000. Separate domains in E1 and E2 proteins serve architectural and productive roles for cooperative DNA binding. *EMBO J.* **19**:3069–3079.
- Gloss, B., and H. U. Bernard. 1990. The E6/E7 promoter of human papillomavirus type 16 is activated in the absence of E2 proteins by a sequence-aberrant Sp1 distal element. *J. Virol.* **64**:5577–5584.
- Hegde, R. S., S. R. Grossman, L. A. Laimins, and P. B. Sigler. 1992. Crystal structure at 1.7 Å of the bovine papillomavirus-1 E2 DNA-binding domain bound to its DNA target. *Nature* **359**:505–512.
- Hegde, R. S., A. F. Wang, S. S. Kim, and M. Schapira. 1998. Subunit rearrangement accompanies sequence-specific DNA binding by the bovine papillomavirus-1 E2 protein. *J. Mol. Biol.* **276**:797–808.
- Hirt, B. 1967. Selective extraction of polyoma DNA from infected mouse cell cultures. *J. Mol. Biol.* **26**:365–369.
- Hou, S. Y., S. Y. Wu, T. Zhou, M. C. Thomas, and C. M. Chiang. 2000. Alleviation of human papillomavirus E2-mediated transcriptional repression via formation of a TATA binding protein (or TFIID)-TFIIB-RNA polymerase II-TFIIF preinitiation complex. *Mol. Cell. Biol.* **20**:113–125.
- Hubert, W. G., T. Kanaya, and L. A. Laimins. 1999. DNA replication of human papillomavirus type 31 is modulated by elements of the upstream regulatory region that lie 5' of the minimal origin. *J. Virol.* **73**:1835–1845.
- Kim, S. S., J. K. Tam, A. F. Wang, and R. S. Hegde. 2000. The structural basis of DNA target discrimination by papillomavirus E2 proteins. *J. Biol. Chem.* **275**:31245–31254.
- Lewis, H., K. Webster, A. M. Sanchez-Perez, and K. Gaston. 1999. Cellular transcription factors regulate human papillomavirus type 16 gene expression by binding to a subset of the DNA sequences recognized by the viral E2 protein. *J. Gen. Virol.* **80**:2087–2096.
- Li, R., J. Knight, G. Breame, 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.
- Lu, J. Z., Y. N. Sun, R. C. Rose, W. Bonnez, and D. J. McCance. 1993. Two E2 binding sites (E2BS) alone or one E2BS plus an A/T-rich region are minimal requirements for the replication of the human papillomavirus type 11 origin. *J. Virol.* **67**:7131–7139.
- McBride, A., and G. Myers. 1997. The E2 proteins, p. III-54–III-73. *In* G. Myers, C. Baker, K. Münger, et al. (ed.), *Human papillomaviruses: a compilation of analysis of nucleic acid and amino acid sequences*, vol. IV. Theoretical Biology and Biophysics Division, Los Alamos National Laboratory, Los Alamos, N. Mex.
- McBride, A. A., R. Schlegel, and P. M. Howley. 1988. The carboxy-terminal domain shared by the bovine papillomavirus E2 transactivator and repressor proteins contains a specific DNA binding activity. *EMBO J.* **7**:533–539.
- McCance, D. J., R. Kopan, E. Fuchs, and L. A. Laimins. 1988. Human papillomavirus type 16 alters human epithelial cell differentiation in vitro. *Proc. Natl. Acad. Sci. USA* **85**:7169–7173.
- 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.

21. Myers, G., C. Baker, K. Münger, F. Sverdrup, A. McBride, H.-U. Bernard, and J. Meissner (ed.). 1997. Human papillomaviruses: a compilation and analysis of nucleic acid and amino acid sequences. Theoretical Biology and Biophysics Division, Los Alamos National Laboratory, Los Alamos, N. Mex.
22. Myers, G., F. Sverdrup, C. Baker, A. McBride, K. Munger, H.-U. Bernard, and J. Meissner (ed.). 1994–1997. Human papillomaviruses: a compilation of analysis of nucleic acid and amino acid sequences, vol. I through IV. Theoretical Biology and Biophysics Division, Los Alamos National Laboratory, Los Alamos, N. Mex.
23. Panagiotidis, C. A., and S. J. Silverstein. 1995. pALEX, a dual-tag prokaryotic expression vector for the purification of full-length proteins. *Gene* **164**:45–47.
24. Papavassiliou, A. G., and S. J. Silverstein. 1990. Characterization of DNA-protein complex formation in nuclear extracts with a sequence from the herpes simplex virus thymidine kinase gene. *J. Biol. Chem.* **265**:1648–1657.
25. Pepinsky, R. B., S. S. Prakash, K. Corina, M. J. Grossel, J. Barsoum, and E. J. Androphy. 1997. Sequences flanking the core DNA-binding domain of bovine papillomavirus type 1 E2 contribute to DNA-binding function. *J. Virol.* **71**:828–831.
26. Phelps, W. C., and P. M. Howley. 1987. Transcriptional trans-activation by the human papillomavirus type 16 E2 gene product. *J. Virol.* **61**:1630–1638.
27. Rank, N. M., and P. F. Lambert. 1995. Bovine papillomavirus type 1 E2 transcriptional regulators directly bind two cellular transcription factors, TFIID and TFIIB. *J. Virol.* **69**:6323–6334.
28. Remm, M., R. Brain, and J. R. Jenkins. 1992. The E2 binding sites determine the efficiency of replication for the origin of human papillomavirus type 18. *Nucleic Acids Res.* **20**:6015–6021.
29. Rheinwald, J. G., and M. A. Beckett. 1981. Tumorigenic keratinocyte lines requiring anchorage and fibroblast support cultures from human squamous cell carcinomas. *Cancer Res.* **41**:1657–1663.
30. Romanczuk, H., F. Thierry, and P. M. Howley. 1990. Mutational analysis of cis elements involved in E2 modulation of human papillomavirus type 16 P97 and type 18 P105 promoters. *J. Virol.* **64**:2849–2859.
31. Russell, J., and M. R. Botchan. 1995. *cis*-Acting components of human papillomavirus (HPV) DNA replication: linker substitution analysis of the HPV type 11 origin. *J. Virol.* **69**:651–660.
32. Sakai, H., T. Yasugi, J. D. Benson, J. J. Dowhanick, and P. M. Howley. 1996. Targeted mutagenesis of the human papillomavirus type 16 E2 transactivation domain reveals separable transcriptional activation and DNA replication functions. *J. Virol.* **70**:1602–1611.
33. Sanders, C. M., and N. J. Maitland. 1994. Kinetic and equilibrium binding studies of the human papillomavirus type-16 transcription regulatory protein E2 interacting with core enhancer elements. *Nucleic Acids Res.* **22**:4890–4897.
34. Sanders, C. M., P. L. Stern, and N. J. Maitland. 1995. Characterization of human papillomavirus type 16 E2 protein and subdomains expressed in insect cells. *Virology* **211**:418–433.
35. Steger, G., and S. Corbach. 1997. Dose-dependent regulation of the early promoter of human papillomavirus type 18 by the viral E2 protein. *J. Virol.* **71**:50–58.
36. Sverdrup, F., and S. A. Khan. 1995. Two E2 binding sites alone are sufficient to function as the minimal origin of replication of human papillomavirus type 18 DNA. *J. Virol.* **69**:1319–1323.
37. Tan, S. H., B. Gloss, and H. U. Bernard. 1992. During negative regulation of the human papillomavirus-16 E6 promoter, the viral E2 protein can displace Sp1 from a proximal promoter element. *Nucleic Acids Res.* **20**:251–256.
38. Tan, S. H., L. E. Leong, P. A. Walker, and H. U. Bernard. 1994. The human papillomavirus type 16 E2 transcription factor binds with low cooperativity to two flanking sites and represses the E6 promoter through displacement of Sp1 and TFIID. *J. Virol.* **68**:6411–6420.
39. Thain, A., O. Jenkins, A. R. Clarke, and K. Gaston. 1996. CpG methylation directly inhibits binding of the human papillomavirus type 16 E2 protein to specific DNA sequences. *J. Virol.* **70**:7233–7235.
40. Thain, A., K. Webster, D. Emery, A. R. Clarke, and K. Gaston. 1997. DNA binding and bending by the human papillomavirus type 16 E2 protein. Recognition of an extended binding site. *J. Biol. Chem.* **272**:8236–8242.
41. Thomas, M., M. Boiron, J. Tanzer, J. P. Levy, and J. Bernard. 1964. In vitro transformation of mice cells by bovine papilloma virus. *Nature* **202**:709–710.
42. Ushikai, M., M. J. Lacey, Y. Yamakawa, M. Kono, J. Anson, T. Ishiji, S. Parkkinen, N. Wicker, M. E. Valentine, I. Davidson, L. P. Turek, and T. H. Haugen. 1994. trans activation by the full-length E2 proteins of human papillomavirus type 16 and bovine papillomavirus type 1 in vitro and in vivo: cooperation with activation domains of cellular transcription factors. *J. Virol.* **68**:6655–6666.
43. 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.
44. 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.
45. 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.
46. van Zonneveld, A. J., S. A. Curriden, and D. J. Loskutoff. 1988. Type 1 plasminogen activator inhibitor gene: functional analysis and glucocorticoid regulation of its promoter. *Proc. Natl. Acad. Sci. USA* **85**:5525–5529.
47. Veeraraghavan, S., C. C. Mello, E. J. Androphy, and J. D. Baleja. 1999. Structural correlates for enhanced stability in the E2 DNA-binding domain from bovine papillomavirus. *Biochemistry* **38**:16115–16124.
48. Wright, W. E., M. Binder, and W. Funk. 1991. Cyclic amplification and selection of targets (CASTing) for the myogenin consensus binding site. *Mol. Cell. Biol.* **11**:4104–4110.