Molecular Cell, Vol. 6, 149–158, July, 2000, Copyright ©2000 by Cell Press

Crystal Structure of the DNA Binding Domain of the Replication Initiation Protein E1 from Papillomavirus

Eric J. Enemark,*†‖ Grace Chen,†‡‖ Daniel E. Vaughn,*† Arne Stenlund,†§ and Leemor Joshua-Tor*†§
*W. M. Keck Structural Biology Center
†Cold Spring Harbor Laboratory
1 Bungtown Road
Cold Spring Harbor, New York 11724
‡Graduate Program in Genetics
State University of New York at Stony Brook
Stony Brook, New York 11794

Summary

Papillomaviral infection causes both benign and malignant lesions and is a necessary cause of cervical carcinoma. Replication of this virus requires the replication initiation proteins E1 and E2, which bind cooperatively at the origin of replication (*ori*) as an $(E1)_2$–$(E2)_2$–DNA complex. This is a precursor to larger E1 complexes that distort and unwind the *ori*. We present the crystal structure of the E1 DNA binding domain refined to 1.9 Å resolution. Residues critical for DNA binding are located on an extended loop and an α helix. We identify the E1 dimerization surface by selective mutations at an E1/E1 interface observed in the crystal and propose a model for the $(E1)_2$–DNA complex. These and other observations suggest how the E1 DNA binding domain orchestrates assembly of the hexameric helicase on the *ori*.

Introduction

Papillomaviruses are members of a large family of closely related DNA tumor viruses that give rise to warts in their hosts. Infection of the genital tract by a subset of the human viruses is the most common sexually transmitted disease with a cumulative 3-year incidence of infection among college-aged students that has escalated to 43% (Ho et al., 1998). One of the complications of prolonged and recurrent infection is the progression of these lesions to cervical cancer. HPV infection is a necessary cause of invasive cervical carcinoma (Walboomers et al., 1999) and thus represents one of the few firmly established links between viral infection and the development of cancer. No effective antiviral treatment currently exists. Two viral proteins, the products from the E1 and E2 open reading frames, are required for viral DNA replication (Ustav and Stenlund, 1991). E1 is a viral initiator protein that recognizes the origin of replication (*ori*). E2 is a site-specific DNA binding transcriptional activator that interacts with E1 to affect its DNA binding properties (for reviews, see McBride and Myers, 1997; Sverdrup and Myers, 1997).

E1 from bovine papillomavirus (BPV) is a 605 amino acid polypeptide that is highly conserved in all papillomaviruses (Figure 1A). In vivo experiments have demonstrated that primary sequence conservation also corresponds to functional conservation: E1 proteins from viruses as distantly related as BPV and the human genital papillomavirus type 11 (HPV 11) can function to replicate either genome (Chiang et al., 1992; Del Vecchio et al., 1992). Initiator proteins such as E1 carry out specific biochemical processes that prepare the template for replication. These processes include site-specific DNA binding at the *ori* (Ustav et al., 1991; Wilson and Ludes-Meyers, 1991; Yang et al., 1991; Lusky et al., 1994), melting of the DNA (Seo et al., 1993b; Gillette et al., 1994; Sanders and Stenlund, 1998), ATPase and a 3′ → 5′ DNA helicase activity generated by a hexameric form of the initiator protein (Seo et al., 1993b; Yang et al., 1993; Sedman and Stenlund, 1998; Fouts et al., 1999). Papillomavirus E1 also interacts with DNA polymerase α (MacPherson et al., 1994; Park et al., 1994; Bonne-Andrea et al., 1995; Conger et al., 1999). E1 is functionally similar to the initiator proteins SV40 and polyomavirus T antigens, although the sequence similarity is significant only in the helicase domain (Clertant and Seif, 1984; Thorner et al., 1993; Mansky et al., 1997).

A physical interaction between E1 and E2 is required for DNA replication in vivo (Mohr et al., 1990; Benson and Howley, 1995; Sedman and Stenlund, 1995; Ferguson and Botchan, 1996; Sakai et al., 1996; Masterson et al., 1998). Specific and efficient recognition of the *ori* is accomplished by cooperative binding of the two proteins (Yang et al., 1991; Seo et al., 1993a; Sedman and Stenlund, 1995). The resulting complex is a precursor to a larger multimeric E1 complex, which after removal of E2 can distort the *ori* and ultimately unwind the DNA double helix (Figure 1B) (Lusky et al., 1994; Sanders and Stenlund, 1998). Thus, the initial binding of E1 to DNA is a critical step in formation of a replication competent complex. The DNA binding domain of E1 has been shown to reside between residues 142 and 308 (Chen and Stenlund, 1998). Similar to the full-length protein, this fragment is monomeric in solution and binds to the *ori* initially as a dimer with each monomer binding to one half-site of the palindromic E1 binding site (Chen and Stenlund, 1998). This fragment is also capable of binding to the *ori* cooperatively with E2. Here we present the crystal structure of the E1 DNA binding domain (DBD) at 1.9 Å resolution. From the structure and from site-directed mutagenesis, we have identified the region of E1 involved in self-association when bound to DNA and suggest a model for interaction of a dimeric E1 with its binding site.

Results and Discussion

Overall Structure

A stable domain containing the E1-DBD was identified by limited proteolysis (residues 159–303), expressed in *Escherichia coli* as a GST fusion protein, cleaved from GST, and purified. This fragment is capable of binding

§To whom correspondence should be addressed (e-mail: stenlund@cshl.org [A. S.], leemor@cshl.org [L. J.]).
‖These authors contributed equally to this work.

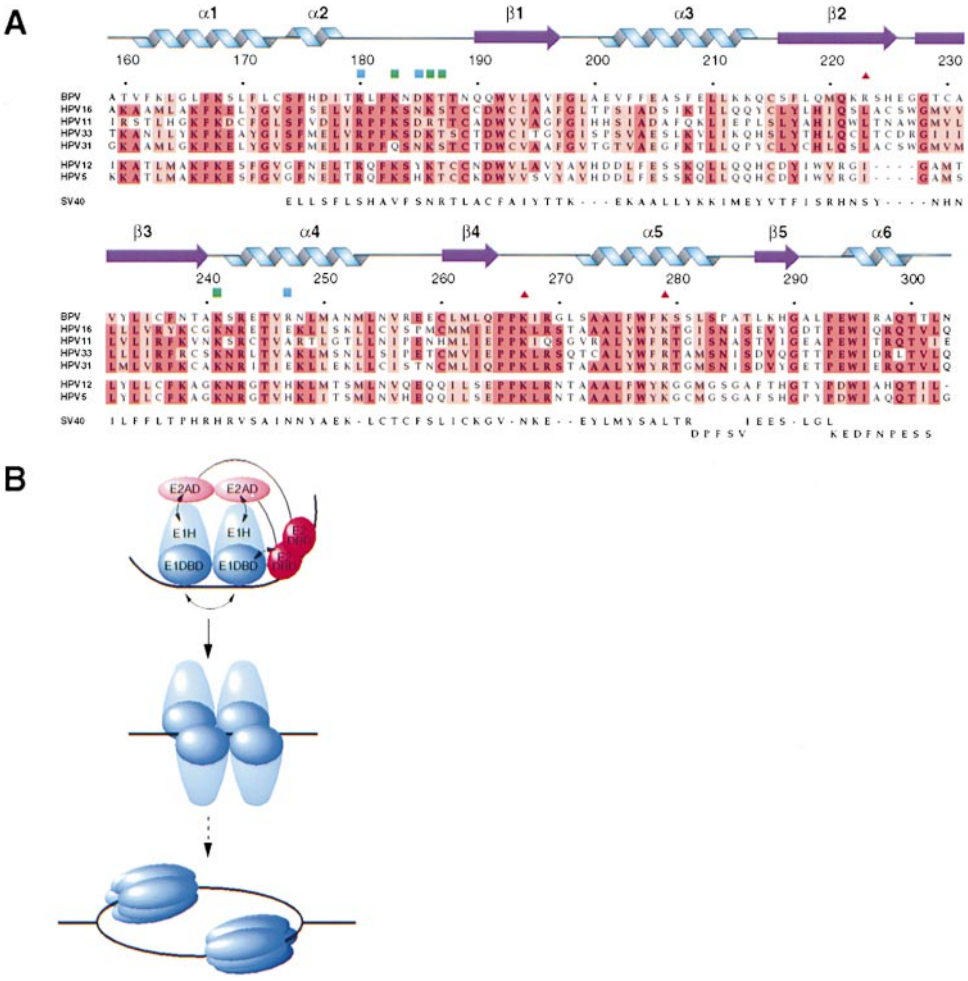

Figure 1. Alignment of Selected Papillomavirus E1 DNA Binding Domain Sequences and Pathway for Assembly of an Initiation-Competent Complex at the Origin of DNA Replication

(A) Alignment of selected papillomavirus E1 DNA binding domain (DBD) sequences with the crystallographically determined secondary structure. The sequence of SV40 T antigen is structurally aligned on the structure of the E1-DBD (sequence identity 6%). Squares: residues that are critical for DNA binding by mutational analysis. Green squares: residues proposed to be important for DNA contact. Blue squares: residues important for the structure of the binding surface. Red triangles: residues that have no apparent effect on DNA binding when mutated to alanine.

(B) Pathway for the assembly of an initiation-competent complex at the origin of DNA replication. The E1 initiator binds cooperatively with E2 to the *ori* forming a specific $(E1)_2$–$(E2)_2$–DNA complex. As a consequence of the interaction between the E1 and E2 DBDs, the origin DNA is sharply bent, allowing an interaction between the E2 transactivation domain (E2AD) and the E1 helicase (E1H) domain. The resulting highly sequence-specific complex serves to recognize the *ori*. In an ATP-dependent process, E2 is displaced from the *ori* and additional E1 molecules are added resulting in a complex where four E1 molecules are bound specifically to the *ori*. Subsequently, additional E1 molecules are bound. This complex can distort the DNA duplex and give rise to partially single stranded regions. In the final step, E1 is assembled onto the exposed single strands to form a hexameric ring-like structure as the replicative DNA helicase.

DNA and binds cooperatively with E2-DBD. The crystal structure, containing two molecules in the asymmetric unit, was solved by multiwavelength anomalous dispersion (MAD) (Table 1) and refined to 1.9 Å resolution ($R_{cryst} = 0.242$; $R_{free} = 0.272$). A portion of the experimental map is shown in Figure 2A. The E1-DBD structure consists of a central five-stranded antiparallel β sheet flanked by loosely packed α helices on one side (α1, α2, α5, and α6) and more tightly packed helices (α3 and α4) on the other (Figure 2B) (demonstrated using the program VOIDOO [Kleywegt and Jones, 1994]).

The structure of E1-DBD resembles that of the DNA binding domain of SV40 T antigen determined by NMR (Luo et al., 1996) (Figure 2C). The root-mean-square deviation (rmsd) between the two DBDs is 2.4 Å for 64 aligned α carbons (LSQMAN [Kleywegt and Jones, 1997]), despite the sequence identity being only 6%. The topology of the DBD of SV40 T antigen is similar to E1-DBD with a central five-stranded β sheet and flanking α helices on either side of the sheet (Figure 2C). However, the first (α1) and last (α6) helices of E1-DBD are not observed in the NMR structure. In addition, α5 of E1-DBD and the equivalent helix (the fourth helix) in SV40 T antigen are oriented differently with respect to the β sheet. Perhaps the most striking difference is the presence of an extended β hairpin in E1-DBD (a "finger")

Table 1. Crystallographic Statistics

A. Data Reduction Statistics

	λ (Å)	Resolution (Å)	Measured Reflections	Unique Reflections	Percent Complete	I/σ(I)	R_{sym}[a]	R_{ano}[b]
$E1_{Se}^{Cl}$ peak	0.9788	2.00 (2.07–2.00)	132236 (6279)	51043 (2771)	87.6 (48.8)	14.4 (3.9)	0.054 (0.288)	0.070 (0.298)
$E1_{Se}^{Cl}$ edge	0.9791	2.00 (2.07–2.00)	77686 (3810)	44283 (2599)	76.0 (45.8)	14.2 (3.6)	0.049 (0.280)	0.065 (0.311)
$E1_{Se}^{Cl}$ remote	0.9643	2.00 (2.07–2.00)	76119 (5484)	46616 (3687)	80.0 (64.9)	12.1 (2.1)	0.058 (0.396)	0.066 (0.426)
$E1_{Nat}^{Br}$	1.100	1.90 (1.97–1.90)	332339 (21175)	35804 (3500)	99.2 (96.4)	11.9 (2.2)	0.087 (0.418)	0.078 (0.334)
$E1_{Nat}^{Cl}$	1.100	2.40 (2.49–2.40)	134772 (13527)	17971 (1776)	98.7 (94.4)	8.6 (5.1)	0.065 (0.110)	0.054 (0.092)

B. Phasing Statistics

	Acentric Phasing Power[c]		Centric Phasing Power	
	50–1.9 Å	2.0–1.9 Å	50–1.9 Å	2.0–1.9 Å
Se peak anomalous	1.80	0.857		
Se edge isomorphous	0.179	0.083	0.162	0.0596
Se edge anomalous	1.80	0.827		
Se remote isomorphous	0.530	0.372	0.447	0.289
Se remote anomalous	1.13	0.501		

FOM[d]	50–1.9 Å	2.0–1.9 Å
Acentric reflections	0.476	0.341
Centric reflections	0.290	0.256

C. Refinement Statistics

Reflections used	25.0–1.9 Å (1.97–1.90 Å); all data
Number of atoms	2335 protein, 14 anions, 100 water
R factor/#ref	0.241/31937 (0.303/2621)
R_{free}/#ref	0.271/1668 (0.330/149)

D. Geometry

	Core[e]	Allowed
Ramachandran plot (%)	95.8	4.2
Bond length RMSD (Å)	0.014	
Bond angle RMSD (°)	1.61	

[a,b] R_{sym}; $R_{ano} = \Sigma\ |I - <I>|/\Sigma\ (I)$, calculated by retaining anomalous-mates and symmetry-mates as independent observations, respectively.
[c] Phasing power calculated as F_H(calc)/phase integrated lack of closure.
[d] FOM is weighted over F amplitude and phase as calculated by SHARP (de La Fortelle and Bricogne, 1997).
[e] Core as defined by Kleywegt and Jones (1996). The model was built using the program O (Jones et al., 1991).

that is formed by β2 and β3 and protrudes from the central β sheet. Interestingly, the presence or absence of this feature divides the papillomavirus E1s into two groups, those with and those without the extended β hairpin (see Figure 1A).

The DNA Binding Surface

Previous mutational studies by alanine substitutions (Thorner et al., 1993; Gonzalez et al., 2000) have revealed that R180, K183, K186, T187, and K241 are critical for DNA binding, while D185 and R247 affect binding to a lesser extent. These residues are located in two conserved regions in the sequence of the E1-DBD (see Figure 1A). The crystal structure demonstrates that these two regions form a continuous area on the surface of the protein. The first region is located on an extended loop between α2 and β1 (which we refer to as the "DNA binding loop"), and the second is on and N-terminal to helix α4 (which we refer to as the "DNA binding helix"). This area is positively charged as demonstrated by a GRASP representation (Nicholls et al., 1991) (data not shown), and several anions (chloride and/or bromide ions) are positioned near this surface in the crystal structure. Replacement of the conserved residues K267 and K279, as well as nonconserved K222 and K157, by alanine, located on other surfaces, does not affect DNA binding (see Figure 1A). The DNA binding loop and the DNA binding helix correspond to the equivalent structural regions in SV40 T antigen, termed A and B2, that were shown to be essential for sequence-specific DNA binding to the SV40 origin (Luo et al., 1996).

The DNA binding loop (residues 180–190), which contains several of the residues shown to be critical for DNA binding, is a very striking feature of the structure. Although this loop does not possess internal secondary structure, it is very well defined in the crystal and adopts a rather extended path. There are two major contributions to the structural integrity and precise positioning of this loop. One is the packing of three highly conserved hydrophobic residues: F182, W192, and P265. The first residue is either a phenylalanine or a tyrosine, while the other two are invariant in the 84 E1 sequences present in the database. The second major contribution to the structure of the DNA binding loop is provided by the invariant residue R243 located on α4, the DNA binding helix (Figure 3). The side chain of this arginine extends toward the middle of the loop, with its guanidinium group forming both a salt bridge with the side chain of D185

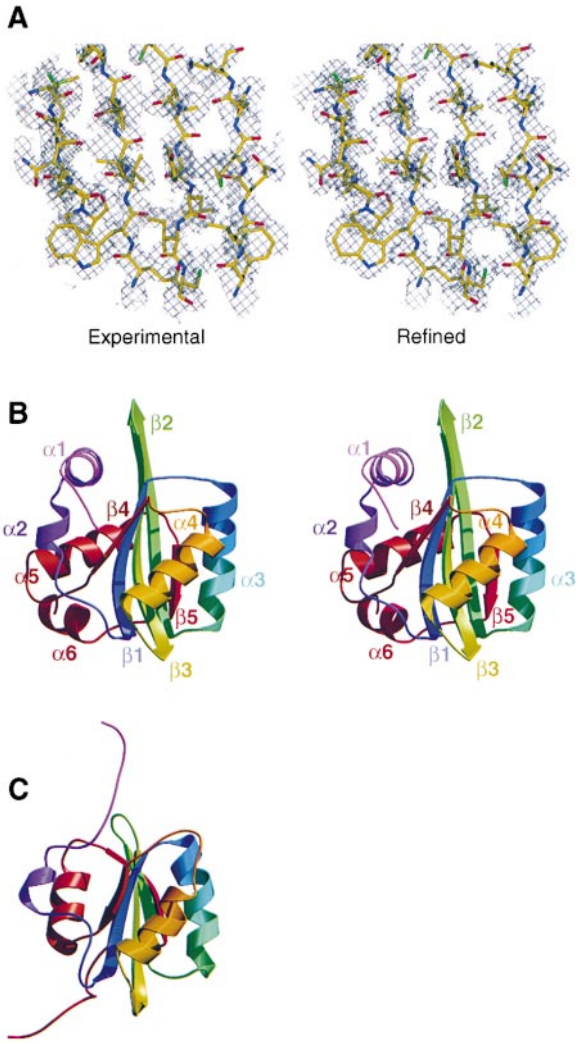

Figure 2. Structure of the E1-DBD

(A) Initial experimental (left) and final electron (right) density maps determined for the β sheet region of E1 DBD superimposed on the refined model of E1-DBD. This figure along with Figures 2B, 2C, 3, 4A, and 5 were prepared with Bobscript (Kraulis, 1991; Esnouf, 1997), and Raster3D (Bacon and Anderson, 1988; Merritt and Murphy, 1994).

(B) Stereo ribbon diagram of the three-dimensional structure of E1-DBD color-ramped from purple at the N terminus to red at the C terminus.

(C) Ribbon diagram of the structure of the SV40 T antigen DBD color-ramped and in the same view as in (C) showing the similarity of the fold.

and a hydrogen bond to the main chain carbonyl of F182 of the loop. In addition, it interacts with β4 by forming hydrogen bonds to the side chain of Q264 and the main chain carbonyl of L263. Thus, this arginine anchors three separate regions of primary structure of the protein, linking two residues of the DNA binding loop with regions of well-defined secondary structures, α4 and β4. The DNA binding loop wraps around the loop emerging from β4, and in particular around the two invariant consecutive prolines, P265 and P266. As mentioned earlier, P265 is part of a hydrophobic interaction within the loop.

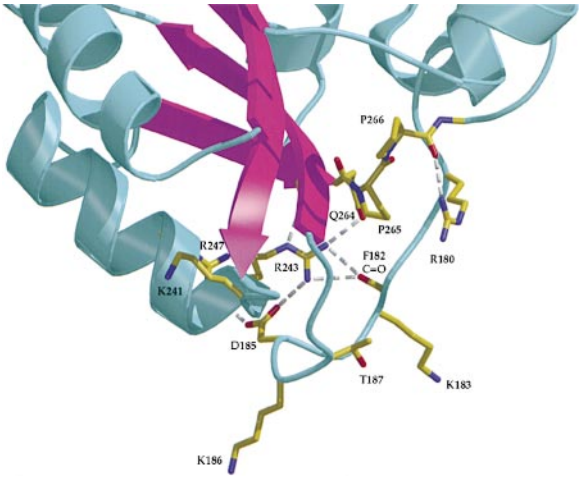

Figure 3. The DNA Binding Loop and Helix

Multiple hydrogen bonds formed by R243 within the DNA binding loop. Contacts are made to side chains of α4, β4, and the loop, anchoring three separate regions of the protein. The DNA binding loop also encircles consecutive proline residues of the loop following β4 shown in stick representation. This results in a well-defined loop presenting DNA binding residues, K183, K186, and T187 for interaction with DNA. K241, also shown to be critical for DNA binding, is N-terminal to α4.

The carbonyl group of P266, on the other hand, forms a hydrogen bond with loop residue R180.

Other interactions also affect the positioning of this loop. In addition to the interaction with R243, loop residue D185 interacts with R247 of the DNA binding helix, α4. These residues, D185, R243, and R247, help position the two DNA binding modules, loop and helix, close together to make a continuous surface. Therefore, mutations of these residues that were shown to have an effect on DNA binding probably affect the conformational integrity and positioning of the two DNA binding regions rather than direct contacts with the DNA. It is possible, of course, that some conformational changes would occur upon DNA binding, for example, R247 may swing out to make direct DNA contacts.

These interactions together result in a structured loop that appropriately displays the residues required for DNA binding on the surface of the protein. DNA binding residues in regulatory proteins are frequently displayed on secondary structural elements, either α helices or β strands. Here, in contrast, an extended, well-ordered, structured loop is employed as an element for DNA binding. Loops are used as elements for DNA recognition in p53 (Cho et al., 1994), STAT-1 (Chen et al., 1998), and NF-κB (Ghosh et al., 1995; Muller et al., 1995). However, in p53 and STAT-1 the major DNA recognition modules involve short loops, whereas in NF-κB, the loop does not appear to be fixed by other elements of the structure and may be more flexible. The main DNA binding module in papillomavirus E1, rather than having an inherent secondary structure, relies on interactions with other regions of the protein for its precise and stable configuration.

Dimer Interaction

As mentioned earlier, full-length E1 initially binds to the *ori* as a dimer together with a dimer of E2, as does the

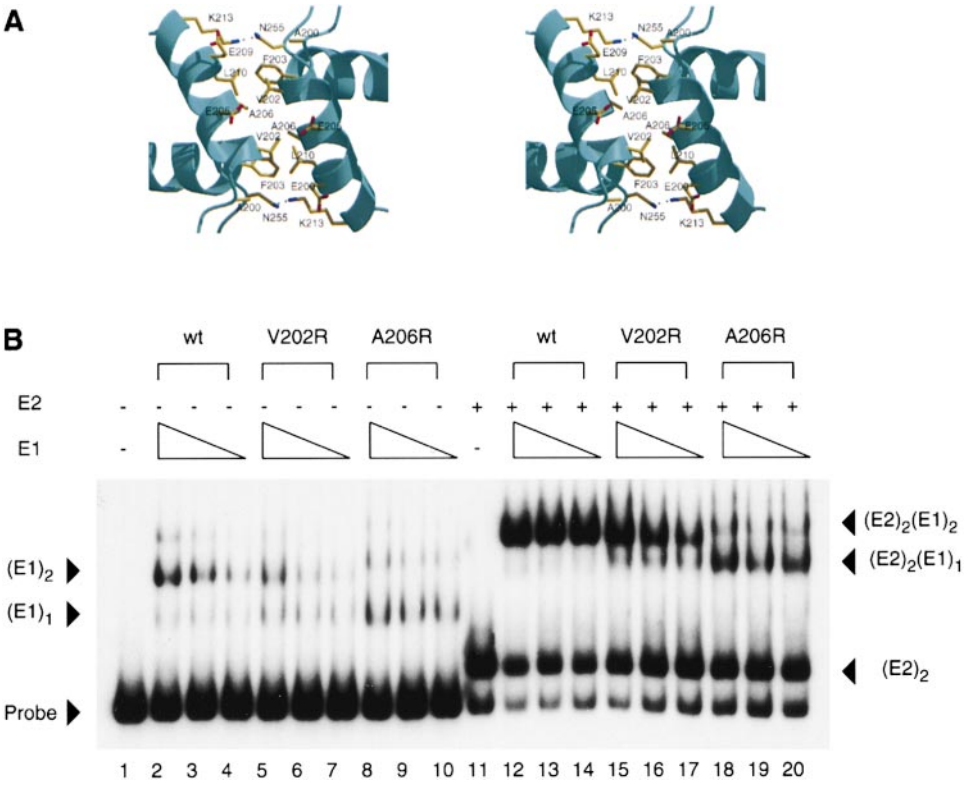

Figure 4. E1-DBD Dimerizes through an Interaction between α3 Helices

(A) Stereo ribbon diagram of the dimer interface viewed down the NCS two-fold axis showing the mostly hydrophobic nature of this interface. Side chains of residues at the interface are shown in stick representation and are labeled. A206 is in a particularly buried position.

(B) DNA binding activity of wild-type (wt) E1-DBD 142–308 (lanes 2–4 and 12–14) and the mutants V202R (lanes 5–7 and 15–17) and A206R (lanes 8–10 and 18–20) was measured in the absence (lanes 1–10) and presence of E2-DBD (lanes 11–20) using an origin probe with a high affinity E2 binding site. The migration of the different complexes is indicated by the arrows. Three two-fold dilutions of E1-DBD were used (1, 0.5, and 0.25 ng of wild-type E1-DBD; 1.6, 0.8, and 0.4 ng of V202R; 3, 1.5, and 0.75 ng of A206R). E2-DBD (20 pg) was used in lanes 11–20.

E1-DBD. Mutational studies, including inosine substitutions, diethylpyrocarbonate (DEPC) interference, and ethylation interference studies, demonstrate that the E1 dimer binds to two consecutive major grooves (Sedman et al., 1997; Chen and Stenlund, 1998; Sanders and Stenlund, 1998; Chen, 1999), one helical turn apart. It is likely that the natural E1/E1 dimeric interaction is maintained as a lattice contact in the crystal structure. There are three different interaction surfaces between E1-DBD monomers in the crystal. One of these E1/E1 interactions, where two monomers are related by a noncrystallographic (NCS) two-fold axis, places the two DNA binding surfaces of each monomer approximately 34 Å apart on the same face of this putative dimer. This distance corresponds to one helical turn of a B-DNA double helix. Since the E1 binding sites on DNA are positioned one helical turn apart, this dimeric form is particularly appealing. Moreover, the same contact is present in two other crystal forms of E1-DBD that were solved by molecular replacement (MR) (data not shown).

The E1/E1 interface in this putative dimer is mostly hydrophobic in nature and consists mainly of the N-terminal halves of symmetry-related α3 helices with residues V202, F203, A206, and L210 at the interface (Figure 4A). Additionally, K213 at the end of α3 of each monomer makes a hydrogen bond to N255 of the other monomer. This dimeric E1/E1 interface is not very extensive, burying approximately 500 Å^2 of the monomer surface area. E1-DBD is monomeric in solution (data not shown) consistent with a weakly interacting interface. To test whether this interaction is relevant for dimerization, we substituted two of the small hydrophobic residues at the interface in α3 with the large charged residue arginine (V202R and A206R). We have previously demonstrated that E1-DBD can bind to a half-site *ori* as a monomer in the presence and absence of the E2-DBD (Chen and Stenlund, 1998). E1 defective for dimerization, therefore, should retain DNA binding activity but bind DNA as a monomer. The mutant proteins were expressed and purified, and DNA binding activity was compared to the wild-type E1-DBD (Figure 4B). Wild-type E1 binds predominantly as a dimer, forming the slower migrating $(E1)_2$–DNA complex, but small quantities of E1 monomer also bind [$(E1)_1$–DNA complex] as shown previously (lanes 2–4) (Chen and Stenlund, 1998). The mutant V202R has reduced DNA binding activity and produces more monomer–DNA complex than wild-type E1 (lanes 5–7). The mutant A206R forms the E1 monomer complex predominantly, indicating that this mutation is more defective in dimerization than both the wild-type and the

Figure 5. A Model for DNA Binding and Dimerization by E1-DBD

A crystallographic interaction between E1 monomers reveals a potential model for E1/E1 interaction in which the distance between the two DNA binding surfaces corresponds to the separation between two major grooves of DNA. The resulting E1 dimer was docked onto a DNA double helix containing the *ori* sequence. The E1 binding sites occupied in the initial complex (sites E1–2 and 4) (Sedman et al., 1997; Chen and Stenlund, 1998) are colored green, and the phosphates, which result in binding interference when ethylated (Sanders and Stenlund, 1998), are colored red. The DNA binding loop and the N-terminal portion of the DNA binding helix of each monomer fit nicely into the major groove. Monomer–monomer contacts are made primarily through helix α3 (see Figure 4). Shown below is the sequence of the *ori* with the E1 (G. C. and A. S., unpublished data) and E2 binding sites indicated. Sites E1–5 and 6 are putative E1 binding sites (G. C. and A. S., unpublished data).

V202R mutant (lanes 8–10). In the presence of E2-DBD, the wild-type E1-DBD forms a cooperative complex containing a dimer of E2 and a dimer of E1 [$(E1)_2$-$(E2)_2$–DNA] (lanes 12–14). In contrast, with A206R (lanes 18–20) and to a lesser extent V202R (lanes 15–17), significant amounts of a faster migrating complex is apparent. This complex corresponds to a monomer of E1 and a dimer of E2 [$(E1)_1$-$(E2)_2$–DNA]. These results demonstrate that although the 202 and 206 mutations affect dimerization of the E1-DBD, the interaction between E1 and E2 DBDs remains intact. The more severe dimerization defect observed with the A206R mutant is consistent with the more sterically constrained and buried position of A206 versus V202 at the interaction interface (Figure 4A).

A Model for DNA Binding

We docked this dimer on a DNA double helix containing the E1 binding site (Figure 5). As seen in the figure, the two DNA binding surfaces in this model lie on top of the two half sites, consistent with interference by ethylation of the phosphates colored in red in Figure 5. It has been shown that binding of E1-DBD dimer to DNA causes a bend in the DNA of approximately 40° (Gillitzer et al., 2000). The surfaces of the protein on the outside of the proposed dimer are made up of a portion of the DNA binding loop and the loop between β4 and α5. These surfaces are positively charged and one can envision that when bent, the DNA wraps around these surfaces on both ends of the dimer. The initial *ori* recognition complex contains a dimer of E1 and a dimer of E2. The crystal structure of the E2-DBD bound to its site was determined previously (Hegde et al., 1992). This structure can be placed in our model by extending the DNA in our model to include the E2 binding site from the crystal structure (data not shown). The interaction between the E1 and E2 DBDs generates a considerable bend of greater than 90° (Gillitzer et al., 2000) and would be difficult to model precisely. Nevertheless, inclusion of E2 in our model would place the E2 dimer on the same face of the DNA as the E1 dimer with its E1 interaction surface (Chen and Stenlund, 2000) facing one of the E1 monomers.

Other E1/E1 Interfaces

While most DNA helicases require a region of single-stranded DNA (ssDNA) for entry, initiator proteins from small DNA tumor viruses such as E1 and SV40 T antigen can initiate unwinding from completely double-stranded DNA by causing helix melting and assembly of the DNA helicase onto a single-stranded region. A consequence of these three activities—sequence-specific DNA binding, helix melting, and DNA helicase activity—is that, unlike other helicases, the initiator proteins have the capability of sequence-specific unwinding of a circular, double-stranded template. The DBD of E1 is likely to play a role in all of these activities. Sequence-specific binding is required for *ori* recognition and efficient melting, and the minimal helicase domain in SV40 T antigen includes the DBD (Wun-Kim and Simmons, 1990).

As mentioned earlier, the E1–E2–DNA complex is a precursor to larger E1 multimeric complexes, eventually resulting in an E1 hexamer with helicase activity (see Figure 1B). The initial E1–E2–DNA complex causes substantial DNA bending (∼90°) (Gillizer et al., 2000). Upon ATP hydrolysis, larger multimeric E1 complexes are formed that bring about DNA distortion (i.e., base pair melting), as can be seen by $KMnO_4$ modification (Gillette et al., 1994; Sanders and Stenlund, 1998). This eventually leads to the formation of a hexameric form of E1 that has helicase activity (Sedman and Stenlund, 1998) and forms a ring-like structure, as shown by electron microscopy (EM) studies (Fouts et al., 1999). Based on these and EM studies of other hexameric helicases (Egelman et al., 1995), it was suggested that these hexameric helicases encircle ssDNA acting as a wedge between the two strands. In this manner, they would separate the strands and unwind the DNA (Lohman and Bjornson, 1996; Bird et al., 1998). The formation of these varied

E1 complexes is likely to require the use of a number of different E1/E1 interfaces. An interesting E1/E1 interaction observed in the crystal structure occurs through a pair of NCS-related α5 helices. Helix α5 is a well-conserved helix, and the interaction surface is primarily hydrophobic. This interaction could be important in the formation of the higher order E1 complexes. It is conceivable that the conversion from one E1 complex to another does not require dissociation and reassociation, but rather a rearrangement of subunits by changing their interaction surfaces and forming different and/or new interfaces.

Conclusions

The crystal structure of the E1-DBD reveals an unusual mode of DNA binding by an extended loop and α helix that together form a continuous DNA binding surface. Most DNA regulatory proteins use secondary structural elements as their DNA binding modules, which present an explicit DNA binding surface. In general, loops may be used for higher flexibility. In E1, however, the main DNA binding module, the DNA binding loop, is a well-structured loop that is stabilized by a variety of interactions with other elements of the structure.

As the full-length protein, the DBD is monomeric in solution but dimerizes when bound to DNA. The dimerization surface was identified in the crystal structure and verified by mutational analysis. This interface is rather limited in surface area and hydrophobic in nature. Since we have shown here that dimerization is important for binding of E1-DBD to the *ori*, this represents an interesting target for small molecule intervention of papillomaviral replication. The constraints imposed by the positions of the DNA binding and dimerization surfaces enabled us to construct a model for the $(E1)_2$–DNA complex, which is consistent with mutational analysis of the E1 binding site.

Although an important function of the E1-DBD is to recognize and mark the *ori*, this is not its sole function. E1-DBD plays an important role in double-stranded DNA melting and formation of the hexameric replicative helicase. As mentioned, initiator proteins, such as E1 and SV40 T antigen, are capable of unwinding a circular double-stranded template. A key question is how the initial *ori*–recognition complex is converted into the active helicase. From our structural and biochemical analyses, we now understand the binding of the dimer that represents the initial step in this process. Most of what is known about the final step, the helicase, stems from EM studies. E1 and SV40 T antigen form hexameric rings with ssDNA. From STEM analysis of SV40 T antigen, Mastrangelo et al. (1989) determined that a bilobed structure with MW corresponding to a double hexamer is bound to the *ori*. Recent EM and image reconstruction have given a more detailed view of this complex (Valle et al., 2000). It can be described as a dumbbell with two larger outer rings and two inner rings. Each of the outer rings contains six helicase domains as shown by antibody decoration. By the same method, each of the inner rings was shown to contain the DBDs. The double hexamer has two-fold symmetry perpendicular to the DNA helix axis and centers around the middle of the palindromic origin sequence. Given the positioning of the E1 binding sites, a simple model for how this type of complex forms from the E1 dimer can be envisioned. The presence of multiple binding sites (see Figure 5) results in encircling of DNA by E1 subunits. In this scenario, each monomer in the E1 dimer (Figure 5) "nucleates" a hexameric unit of the helicase. This indicates that the dimerization interface may play a crucial role in the assembly of the helicase. The weak interaction as seen in the crystal structure would allow disruption of this contact and subunit rearrangement, eventually leading to the active helicase.

Helicases bind ssDNA. ssDNA strongly stimulates hexamer formation only in the presence of the DBD in SV40 T antigen and E1 (Wun-Kim and Simmons, 1990; A. S., unpublished data), which is consistent with the DBD being the entity that binds ssDNA in these initiator proteins. Thus, the DBD appears to provide the DNA binding activity of the hexameric helicase. Although we do not know how ssDNA is bound, it is intriguing that the contacts made by the major DNA binding module, the DNA binding loop, seems to be with one strand only, which is rather unusual. Often, a secondary structure motif, such as an α helix lies in the major groove interacting with both strands of DNA, providing for structural (surface) complementarity. The use of a loop does not impose that kind of constraint. Because of the head to head arrangement of the DBDs, each of the DNA binding loops contacts separate strands. This may be the means by which the two helicases achieve strand specificity for unwinding in opposite directions.

Experimental Procedures

Protein Expression and Purification

BPV E1 residues 159–303 and its seleno-methionine derivative were obtained by expression of N-terminal GST fusion proteins with an intermediate thrombin cleavage site in strain BL21DE3 ($E1_{Nat}^{Br}$, $E1_{Nat}^{Cl}$) and in methionine auxotroph strain DL41 grown on LeMaster's seleno-methionine medium ($E1_{Se}^{Cl}$) (Hendrickson et al., 1990). GST-E1 DBD was purified and subsequently cleaved from the GST fragment. The resulting species was further purified by conventional chromatography.

Crystallization

E1-DBD was crystallized as plates from 8 mg/mL solutions by the hanging drop method in the presence of Cl^- ($E1_{Se}^{Cl}$, $E1_{Nat}^{Cl}$) or Br^- ($E1_{Nat}^{Br}$) precipitants.

Diffraction Data Collection and Processing

All data sets were collected at the NSLS at Brookhaven National Laboratory under cryogenic conditions (100 K). Multiwavelength data for $E1_{Se}^{Cl}$ were collected at beamline X-8C, and data for $E1_{Nat}^{Br}$ and $E1_{Nat}^{Cl}$ were collected at beamline X-26C. The $E1_{Se}^{Cl}$ and $E1_{Nat}^{Cl}$ crystals were cryoprotected by the presence of 10% glycerol in the crystallization buffer and were frozen in liquid propane. A crystal of $E1_{Nat}^{Br}$ was cryoprotected by brief immersion in a solution consisting of 75% reservoir solution and 25% commercial automotive antifreeze prior to freezing in a stream of 100 K nitrogen gas. The crystals of $E1_{Se}^{Cl}$ and $E1_{Nat}^{Cl}$ were relatively mosaic ($>1°$) and displayed lower diffraction quality in crystal orientations that aligned the crystal plate parallel to the beam. In these orientations, the diffraction limit decreased with concurrent spot elongation and streaking. In contrast, the crystal of $E1_{Nat}^{Br}$ had a low mosaicity ($<0.3°$) and displayed spot shapes and diffraction limits independent of crystal orientation.

Data for both crystals were integrated with DENZO (Otwinowski and Minor, 1997) and processed with SCALEPACK (Otwinowski and Minor, 1997). For $E1_{Se}^{Cl}$, crystal mosaicity was refined for each frame by DENZO, and each value was used directly in SCALEPACK and not

refined. The mosaicity of this crystal resulted in reflection overlap, particularly at high resolution, and the rejection of these reflections was the primary cause for incompleteness of the data. The data for a fast, low-resolution sweep and for a slow, high-resolution sweep of native $E1_{Nat}^{Br}$ were processed separately and then merged.

A Patterson map revealed a strong pseudo-I centering with an NCS translation of 0.5, 0.5, 0.44. The space group was therefore ambiguous because reflections corresponding to the 2_1 condition along each crystallographic axis were anticipated to be systematically weak due to this pseudo-I centering. While the data collected for the $E1_{Se}^{Cl}$ and $E1_{Nat}^{Cl}$ demonstrated ambiguous systematic absences, particularly with respect to the *a* axis, the data collected for the nonmosaic $E1_{Nat}^{Br}$ demonstrated clear absences for the 2_1 condition along each axis, and the space group was assigned as $P2_12_12_1$ within a pseudo-$I2_12_12_1$ lattice. Cell dimensions were $a = 42.20$, $b = 84.97$, $c = 124.29$ for $E1_{Nat}^{Br}$ and $a = 41.69$, $b = 84.60$, $c = 122.02$ for $E1_{Se}^{Cl}$.

Structure Solution

The anomalous data collected at the Se absorption peak were used to locate eight Se sites by the program *SnB* (Weeks and Miller, 1999) in space group $P2_12_12_1$. These correspond to two independent E1-DBD molecules (four Met/E1-DBD) related by the above NCS translation. The data collected at all three wavelengths for the Se-Met derivative were used in maximum likelihood refinement of Se site parameters by the program SHARP (de La Fortelle and Bricogne, 1997). Refinement of the coordinates determined by *SnB* followed by solvent flattening and NCS averaging by the program DM (CCP4, 1994) led to an encouraging electron density map (Figure 2A). A partial structure model was autobuilt by the program wARP (Perrakis et al., 1999) using the density-modified phase model. Additionally, the phase model was used to locate several bromides in $E1_{Nat}^{Br}$ by Fourier synthesis. The program SIGMAA (CCP4, 1994) was used to combine the partial structure model with a new experimental phase model of SHARP that was based upon the three-wavelength MAD data as well as the isomorphic replacement of Cl^- by Br^- (five total data sets). Solvent flattening and NCS averaging of the resulting phase model (DM) led to a fully traceable electron density map.

Structure Refinement

The model was refined against the $E1_{Nat}^{Br}$ data using the program CNS (Brünger et al., 1998) with iterative model building using the program O (Jones et al., 1991). The model contains 2335 protein atoms, 14 bromides, and 100 water molecules. Since both Br^- and Cl^- were present in the crystallization buffer, some of the Br^- sites may be partially occupied by Cl^-. This was modeled by refining the site occupancy of each Br^-.

DNA Binding Assays

DNA binding was examined using electrophoretic mobility shift assays (EMSA) as described previously (Chen and Stenlund, 1998). Probe (5000 cpm/reaction) was mixed with E1-DBD and/or the E2-DBD in the presence of 10 ng of nonspecific competitor DNA (pUC 119) in binding buffer (20 mM potassium phosphate [pH 7.4], 0.1 M NaCl, 1 mM EDTA, 0.1% NP-40, 3 mM DTT, 0.7 mg/ml bovine serum albumin, 5% glycerol). Binding reactions were performed in a total volume of 10 μl. After incubation for 30 min at room temperature, the samples were directly loaded on 6% 40:1 (acrylamide/bisacrylamide) polyacrylamide gels and subjected to electrophoresis (PAGE) in 0.5× Tris-borate-EDTA. The gels were dried and autoradiographed. The level of E1 binding was quantified using a Fuji imager system.

Acknowledgments

We thank Julie Rosenbaum for help with protein purification, Dr. Ryuji Kobayashi and Nora Poppito for mass spectrometry, Dr. Joel Berendzen (beamline X8C), Dr. Dieter Schneider (beamline X26C), and Dr. Craig Ogata (beamline X4A) for help and support with data collection at the National Synchrotron Light Source (NSLS) at Brookhaven National Laboratory (BNL). We also thank Dr. Alexander Gann for critical reading of the manuscript. The NSLS at BNL is funded by the U. S. Department of Energy Division of Material Sciences and Division of Chemical Sciences under contract number DE-AC02-98CH10886. This work was supported by a National Cancer Institute training grant to E. J. E., National Institutes of Health grant (C13106) to A. S., and a grant from the American Cancer Society (RPG-99-208-01-MBC) to L. J.

Received March 28, 2000; revised May 2, 2000.

References

Bacon, D.J., and Anderson, W.F. (1988). A fast algorithm for rendering space-filling molecule pictures. J. Mol. Graph. *6*, 219–220.

Benson, J.D., and Howley, P.M. (1995). Amino-terminal domains of the bovine papillomavirus type 1 E1 and E2 proteins participate in complex formation. J. Virol. *69*, 4364–4372.

Bird, L.E., Subramanya, H.S., and Wigley, D.B. (1998). Helicases: a unifying structural theme? Curr. Opin. Struct. Biol. *8*, 14–18.

Bonne-Andrea, C., Santucci, S., Clertant, P., and Tillier, F. (1995). Bovine papillomavirus E1 protein binds specifically DNA polymerase α but not replication protein A. J. Virol. *69*, 2341–2350.

Brünger, A.T., Adams, P.D., Clore, G.M., De Lano, W.L., Gros, P., Gross-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., et al. (1998). Crystallography and NMR system: a new software suite for macromoleclar structure determination. Acta Crystallogr. D *54*, 905–921.

Chen, G.G. (1999). Origin binding and recognition by the BPV-1 E1 initiator, PhD thesis, State University of New York at Stony Brook, Stony Brook, New York.

Chen, G., and Stenlund, A. (1998). Characterization of the DNA-binding domain of the bovine papillomavirus replication initiator E1. J. Virol. *72*, 2567–2576.

Chen, G., and Stenlund, A. (2000). Two patches of amino acids on the E2 DNA binding domain define the surface for interaction with E1. J. Virol. *74*, 1506–1512.

Chen, X., Vinkemeier, U., Zhao, Y., Jerusalmi, D., Darnell, J.E., and Kuriyan, J. (1998). Crystal structure of a tyrosine phosphorylated STAT-1 dimer bound to DNA. Cell *93*, 827–839.

Chiang, C.M., Ustav, M., Stenlund, A., Ho, T.F., Broker, T.R., and Chow, L.T. (1992). Viral E1 and E2 proteins support replication of homologous and heterologous papillomaviral origins. Proc. Natl. Acad. Sci. USA *89*, 5799–5803.

Cho, Y., Gorina, S., Jeffrey, P.D., and Pavletich, N.P. (1994). Crystal structure of a p53 tumor suppressor–DNA complex: understanding tumorigenic mutations. Science *265*, 346–355.

Clertant, P., and Seif, I. (1984). A common function for polyoma virus large-T and papillomavirus E1 proteins? Nature *311*, 276–279.

Collaborative Computational Project 4 (CCP4) (1994). The CCP4 suite: programs for protein crystallography. Acta Crystallogr. D *50*, 760–763.

Conger, K.L., Liu, J.S., Kuo, S.R., Chow, L.T., and Wang, T.S. (1999). Human papillomavirus DNA replication. Interactions between the viral E1 protein and two subunits of human DNA polymerase primase. J. Biol. Chem. *274*, 2696–2705.

de La Fortelle, E., and Bricogne, G. (1997). Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. Methods Enzymol. *276*, 472–494.

Del Vecchio, A., Romanczuk, H., Howley, P.M., and Baker, C.C. (1992). Transient replication of human papillomavirus DNAs. J. Virol. *66*, 5949–5958.

Egelman, E.H., Yu, X., Wild, R., Hingorani, M.M., and Patel, S.S. (1995). Bacteriophage T7 helicase/primase proteins form rings around single-stranded DNA that suggest a general structure for hexameric helicases. Proc. Natl. Acad. Sci. USA *92*, 3869–3873.

Esnouf, R.M. (1997). An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. J. Mol. Graph. *15*, 132–134.

Ferguson, M.K., and Botchan, M. (1996). Genetic analysis of the

activation domain of bovine papillomavirus protein E2: its role in transcription and replication. J. Virol. *70*, 4193–4199.

Fouts, E.T., Yu, X., Egelman, E.H., and Botchan, M.R. (1999). Biochemical and electron microscopic image analysis of the hexameric E1 helicase. J. Biol. Chem. *274*, 4447–4458.

Ghosh, G., van Duyne, G., Ghosh, S., and Sigler, P.B. (1995). Structure of NF-κB p50 homodimer bound to a κB site. Nature *373*, 303–310.

Gillette, T.G., Lusky, M., and Borowiec, J.A. (1994). Induction of structural changes in the bovine papillomavirus type 1 origin of replication by the viral E1 and E2 proteins. Proc. Natl. Acad. Sci. USA *91*, 8846–8850.

Gillitzer, E., Chen, G., and Stenlund, A. (2000). Separate domains in E1 and E2 proteins serve architectural and productive roles for cooperative DNA binding. EMBO J. *19*, 3069–3079.

Gonzalez, A., Bazaldua-Hernandez, C., West, M., Wotek, K., and Wilson, V.G. (2000). Identification of a short, hydrophilic amino acid sequence critical for origin recognition by the bovine papillomavirus E1 protein. J. Virol. *74*, 245–253.

Hegde, R.S., Grossman, S.R., Laimins, L.A., and Sigler, P.B. (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.

Hendrickson, W.A., Horton, J.R., and LeMaster, D.M. (1990). Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. EMBO J. *9*, 1665–1672.

Ho, G.Y.F., Bierman, R., Beardsley, L., Chang, C.J., and Burk, R.D. (1998). Natural history of cervicovaginal papilloma virus infection in young women. New Engl. J. Med. *338*, 423–428.

Jones, T.A., Zou, J.Y., Cowen, S.W., and Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallogr. A *47*, 110–119.

Kleywegt, G.J., and Jones, T.A. (1994). Detection, delineation, measurement and display of cavities in macromolecular structures. Acta Crystallogr. D *50*, 178–185.

Kleywegt, G.J., and Jones, T.A. (1996). Phi/Psi-chology: Ramachandran revisited. Structure *4*, 1395–1400.

Kleywegt, G.J., and Jones, T.A. (1997). Detecting folding motifs and similarities in protein structures. Methods Enzymol. *277*, 525–545.

Kraulis, P.J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. J. Appl. Crystallogr. *24*, 946–950.

Lohman, T.M., and Bjornson, K.P. (1996). Mechanisms of helicase-catalyzed DNA unwinding. Annu. Rev. Biochem. *65*, 169–214.

Luo, X., Sanford, D.G., Bullock, P.A., and Bachovchin, W.W. (1996). Solution structure of the origin DNA-binding domain of SV40 T-antigen. Nat. Struct. Biol. *3*, 1034–1039.

Lusky, M., Hurwitz, J., and Seo, Y.S. (1994). The bovine papillomavirus E2 protein modulates the assembly but is not stably maintained in a replication-competent multimeric E1-replication origin complex. Proc. Natl. Acad. Sci. USA *91*, 8895–8899.

MacPherson, P., Thorner, L., and Botchan, M. (1994). The bovine papillomavirus E1 protein has ATPase activity essential to viral DNA replication and efficient transformation in cells. Virology *204*, 403–408.

Mansky, K.C., Batiza, A., and Lambert, P.F. (1997). Bovine papillomavirus type 1 E1 and simian virus 40 large T-antigen share regions of sequence similarity required for multiple functions. J. Virol. *71*, 7600–7608.

Masterson, P.J., Stanley, M.A., Lewis, A.P., and Romanos, M.A. (1998). A C-terminal helicase domain of the human papillomavirus E1 protein binds E2 and the DNA polymerase α-primase p68 subunit. J. Virol. *72*, 7407–7419.

Mastrangelo, I.A., Hough, P.V. C., Wall, J.S., Dodson, M., Dean, F.B., and Hurwitz, J. (1989). ATP-dependent assembly of double hexamers of SV40 T antigen at the viral origin of DNA replication. Nature *338*, 658–662.

McBride, A., and Myers, G. (1997). The E2 proteins. In Human Papillomaviruses 1997, G. Myers, C. Baker, K. Munger, F. Sverdrup, A. McBride, and H.-U. Bernard, eds. (Los Alamos, NM: Los Alamos National Laboratory), pp. 37–53.

Merritt, E.A., and Murphy, M.E.P. (1994). Raster3D version 2.0—a program for photorealistic molecular graphics. Acta Crystallogr. D *50*, 869–873.

Mohr, I.J., Clark, R., Sun, S., Androphy, E.J., MacPherson, P., and Botchan, M. (1990). Targeting the E1 replication protein to the papillomavirus origin of replication by complex formation with the E2 transactivator. Science *250*, 1694–1699.

Muller, C.W., Rey, F.A., Sodeoka, M., Verdine, G.L., and Harrison, S.C. (1995). Structure of the NF-κB p50 homodimer bound to DNA. Nature *373*, 311–317.

Nicholls, A., Sharp, K.A., and Honig, B. (1991). Protein folding and association—insights from the interfacial and thermodynamic properties of hydrocarbons. Proteins *11*, 281–296.

Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. *276*, 307–326.

Park, P., Copeland, W., Yang, L., Wang, T., Botchan, M., and Mohr, I. (1994). The cellular DNA polymerase α-primase is required for papillomavirus DNA replication and associates with the viral E1 helicase. Proc. Natl. Acad. Sci. USA *91*, 8700–8704.

Perrakis, A., Morris, R., and Lamzin, V.S. (1999). Automated protein model building combined with iterative structure refinement. Nat. Struct. Biol. *6*, 458–463.

Sakai, H., Yasugi, T., Benson, J.D., Dowhanick, J.J., and Howley, P.M. (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.

Sanders, C.M., and Stenlund, A. (1998). Recruitment and loading of the E1 initiator protein: an ATP-dependent process catalyzed by a transcription factor. EMBO J. *17*, 7044–7055.

Sedman, J., and Stenlund, A. (1995). Co-operative interaction between the initiator E1 and the transcriptional activator E2 is required for replicator-specific DNA replication of bovine papillomavirus in vivo and in vitro. EMBO J. *14*, 6218–6228.

Sedman, J., and Stenlund, A. (1998). The papillomavirus E1 protein forms a DNA-dependent hexameric complex with ATPase and DNA helicase activities. J. Virol. *72*, 6893–6897.

Sedman, T., Sedman, J., and Stenlund, A. (1997). Binding of the E1 and E2 proteins to the origin of replication of bovine papillomavirus. J. Virol. *71*, 2887–2896.

Seo, Y.S., Muller, F., Lusky, M., Gibbs, E., Kim, H.Y., Phillips, B., and Hurwitz, J. (1993a). Bovine papillomavirus (BPV)-encoded E2 protein enhances binding of E1 protein to the BPV replication origin. Proc. Natl. Acad. Sci. USA *90*, 2865–2869.

Seo, Y.S., Muller, F., Lusky, M., and Hurwitz, J. (1993b). Bovine papillomavirus (BPV) encoded E1 protein contains multiple activities required for BPV DNA replication. Proc. Natl. Acad. Sci. USA *90*, 702–706.

Sverdrup, F., and Myers, G. (1997). The E1 proteins. In Human Papillomaviruses 1997, G. Myers, C. Baker, K. Munger, F. Sverdrup, A. McBride, and H.-U. Bernard, eds. (Los Alamos, NM: Los Alamos National Laboratory).

Thorner, L., Lim, D., and Botchan, M. (1993). DNA binding domain of bovine papillomavirus type 1 E1 helicase: structural and functional aspects. J. Virol. *67*, 6000–6014.

Ustav, M., and Stenlund, A. (1991). Transient replication of BPV-1 requires two viral polypeptides encoded by the E1 and E2 open reading frames. EMBO J. *10*, 449–457.

Ustav, M., Ustav, E., Szymanski, P., and Stenlund, A. (1991). Identification of the origin of replication of bovine papillomavirus and characterization of the viral origin recognition factor E1. EMBO J. *10*, 4321–4329.

Valle, M., Gruss, C., Halmer, L., Carazo, J.M., and Donate, L.E. (2000). Large T-antigen double hexamers imaged at the simian virus 40 origin of replication. Mol. Cell. Biol. *20*, 34–41.

Walboomers, J.M., Jacobs, M.V., Manos, M.M., Bosch, F.X., Kummer, J.A., Shah, K.V., Snijders, P.J., Peto, J., Meijer, C.J., and Muñoz, N. (1999). Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. J. Pathol. *189*, 12–19.

Weeks, C.M., and Miller, R. (1999). The design and implementation of *SnB* version 2.0. J. Appl. Crystallogr. *32*, 120–124.

Wilson, V.G., and Ludes-Meyers, J. (1991). A bovine papillomavirus E1-related protein binds specifically to bovine papillomavirus DNA. J. Virol. *65*, 5314–5322.

Wun-Kim, K., and Simmons, D.T. (1990). Mapping of helicase and helicase substrate-binding domains on simian virus 40 large T antigen. J. Virol. *64*, 2014–2020.

Yang, L., Li, R., Mohr, I., Clark, R., and Botchan, M.R. (1991). Activation of BPV-1 replication in vitro by the transcription factor E2. Nature *353*, 628–633.

Yang, L., Mohr, I., Fouts, E., Lim, D.A., Nohaile, M., and Botchan, M. (1993). The E1 protein of the papillomavirus BPV-1 is an ATP dependent DNA helicase. Proc. Natl. Acad. Sci. USA *90*, 5086–5090.

Protein Data Bank ID Code

The atomic coordinates have been deposited in the Protein Data Bank under the ID code 1F08.