



Published in final edited form as:
Front Biosci. ; 14: 900–917.

Evolutionary and biophysical relationships among the papillomavirus E2 proteins

Dukagjin M. Blakaj¹, Narcis Fernandez-Fuentes², Zigui Chen³, Rashmi Hegde⁴, Andras Fiser¹, Robert D. Burk^{3,5}, and Michael Brenowitz¹

¹Department of Biochemistry, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx NY 10461

²Leeds Institute of Molecular Medicine, Section of Experimental Therapeutics, St. James Hospital, Leeds LS9 7TF, United Kingdom

³Department of Microbiology and Immunology, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx NY 10461

⁴Division of Developmental Biology, Cincinnati Children's Hospital Medical Center, Department of Pediatrics, University of Cincinnati School of Medicine, 3333 Burnet Avenue, Cincinnati OH 45229

⁵Departments of Obstetrics and Gynecology and Women's Health and Epidemiology and Population Health, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx NY 10461

Abstract

Infection by a human papillomavirus (HPV) may result in a variety of clinical conditions ranging from benign warts to invasive cancer depending on the viral type. The HPV E2 protein represses transcription of the E6 and E7 genes in integrated papillomavirus genomes and together with the E1 protein is required for viral replication. E2 proteins bind with high affinity to palindromic DNA sequences consisting of two highly conserved four base pair sequences flanking a variable 'spacer' of identical length. The E2 proteins directly contact the conserved DNA but not the spacer DNA. However, variation in naturally occurring spacer sequences results in differential protein binding affinity. This discrimination in binding is dependent on their sensitivity to the unique conformational and/or dynamic properties of the spacer DNA in a process termed 'indirect readout'. This article explores the structure of the E2 proteins and their interaction with DNA and other proteins, the effects of ions on affinity and specificity, and the phylogenetic and biophysical nature of this core viral protein. We have analyzed the sequence conservation and electrostatic features of three-dimensional models of the DNA binding domains of 146 papillomavirus types and variants with the goal of identifying characteristics that associated with risk of virally caused malignancy. The amino acid sequence, three-dimensional structure, and the electrostatic features of E2 protein DNA binding domain showed high conservation among all papillomavirus types. This indicates that the specific interactions between the E2 protein and its binding sites on DNA have been conserved throughout PV evolution. Analysis of the E2 protein's transactivation domain showed that unlike the DNA binding domain, the transactivation domain does not have extensive surfaces of highly conserved residues. Rather, the regions of high conservation are localized to small surface patches. The invariance of the E2 DNA binding domain structure, electrostatics and sequence suggests that it may be a suitable target for the development of vaccines effective against a broad spectrum of HPV types.

Keywords

Papillomavirus; DNA; Protein-DNA interactions; Electrostatics; E2; Review

2. INTRODUCTION

Human papillomaviruses (HPVs) are small, double-stranded DNA viruses that infect cutaneous and mucosal epithelial tissues. Worldwide screenings for papillomaviruses have identified over 120 viral types, about one third of which infect the epithelium of the genital tract (1-4). The viral types associated with development of anogenital cancers, including those of the cervix, are denoted 'high risk' while 'low risk' viruses induce benign genital warts or cause minimal or no cytological effects (5). The taxonomy used to classify the relationship among the different papillomaviruses has been based on the L1 open reading frame (ORF) DNA sequence (1). A new papillomavirus type is recognized if the L1 ORF differs by more than 10% from its closest relative. DNA sequence differences of 2 - 10% and < 2% differences in sequence identity define a subtype and a variant, respectively (6). The papillomavirus types have recently been reclassified into species, groups and higher order taxonomy (1).

Of the viral types associated with cancer, HPV16 is associated with half of all cervical cancers. At least twenty four variant lineages of HPV16 have been identified; these variants are divided broadly into European and Non-European lineages. Studies investigating HPV16 variants and risk for cancer of the cervix and their precursor high grade lesions indicate an increased risk of disease associated with the non-European variants (7-9). For instance, an epidemiological study of 10,000 women in Costa Rica revealed that those infected with Non-European HPV16 variants were eleven times more likely than those infected with the prototypic HPV16 to be diagnosed with CIN3/cervical cancer (7).

The important role played by the E2 protein in papillomavirus life cycle and human infection is supported by epidemiological, evolutionary and clinical studies (1, 7, 10-13). The HPV E2 protein represses the transcription of the E6 and E7 genes in integrated papillomavirus genomes (14) and together with the E1 protein is required for viral replication. Whether the E2 protein activates or represses gene transcription is dependent on the composition of the E2 DNA binding site and its position within the Long Control Region (LCR) of the viral genome. E2 also participates in DNA replication by binding to, and recruiting the E1 helicase to the viral origin of replication (15). The regulation of the oncoproteins E6 and E7 expression by E2 has clinical significance; Loss of their E2-dependent repression through viral integration contributes to cancer progression (15, 16). The intracellular concentration, binding site affinity, cooperative interactions between E2 proteins bound to multiple sites and interaction with E1 are critical to control of viral life cycle (17-19). A goal of our comparative analysis is to extrapolate our understanding of E2 protein function from the few viral types and variants for which detailed structural, biophysical and biochemical studies have been conducted (20).

The solution of atomic resolution structures of DNA Binding Domain (E2/D) from several papillomavirus types, free and bound to DNA (20-25) together with detailed binding and thermodynamic analyses, allow nuanced inquiries into the molecular mechanisms of direct and indirect readout of DNA sequence affinity and specificity by the E2 proteins (26-30). More limited structural information about the transactivation domain provides comparable insight into the protein-protein interactions that also contribute to the biological function of the E2 protein (31-33).

This article explores E2 protein structure and function by comparative analysis to explore its role in the viral life cycle, virulence and contribution to the oncogenic potential of clinically important papillomavirus types. Our analysis of the amino acid sequence, three-dimensional structure, and the electrostatic features of the E2/D shows high conservation among all papillomavirus types, indicating that the specific interactions between the E2 protein and its binding sites on DNA have been highly conserved through evolution.

2.1. Overview of the structure of the E2 protein

The papillomavirus HPV16 E2 gene encodes a DNA binding protein of 360 amino acids in length that dimerizes to regulate viral gene expression and replication (Figure 1A). The E2 protein consists of a N-terminal transactivation domain (Figure 1B) and a C-terminal DNA binding domain (Figure 1C) connected by a flexible linker. The protein binds as a dimer to its cognate DNA sequence (20, 34). Structures of the DNA binding domain (abbreviated E2/D as noted above) from several viral types have been reported, free and in complex with different cognate DNA sequences. The solved E2/D crystal structures include the human high-risk cancer associated types HPV16, HPV18 and HPV31 (23-25, 32), the low risk cancer associated types include HPV6 (30), which cause benign genital warts in humans and the cow wart causing type BPV1. Solved NMR solution structures include the E2/D from HPV16 (35), HPV31 (24, 36) and BPV1 (37, 38).

The E2/D is part of a novel structural class forming a dimeric β -barrel, with each subunit contributing a 4-stranded β -Sheet “half-barrel” (Figure 1C). The α_1 helix is termed the ‘recognition helix’ and contains all the amino acids involved in direct readout. The dimer interface is made up of hydrogen bonds between subunits and a substantial hydrophobic β -barrel core (20). This topology is unusual since secondary, tertiary and quaternary structure is coupled. Unfolding experiments with either urea or acid suggest early dimerization as a step in the folding pathway (39). The monomer to dimer transition for this system is nM or less (34) indicating that the protein is likely a stable dimer in the cell. The sequence conservation among the numerous E2/Ds ranges from 80% identity among closely related viral types to 30% sequence identity among distant viral types (Figure 3).

The N-terminal transactivation domain activates gene transcription and viral replication (Figure 1). Crystal structures of the HPV16 (31), and HPV18 (32) transactivation domains have been solved. The HPV18 protein has also been solved in complex with the E1 helicase domain (33). These structures are very similar, the C-alpha backbone of 189 residues superpose with a root means square deviation (RMSD) of ~1.2 angstroms (33). The transactivation domain contains two sub-domains, a curved anti-parallel beta-sheet domain and a helical domain containing three anti-parallel helices arranged to give the module an overall L shaped appearance (Figure 1b).

A ‘linker’ of 40 - 200 amino acids depending on the viral type (Figure 1A) connects the E2/D and transactivation domains. This region is poorly conserved and is believed to be unstructured. Little functional information is available for it. Some evidence exists to suggest that it is important to E2 function including nuclear localization (40). Phosphorylation of serine residues in the BPV1 E2 linker is required for viral DNA replication (41) and the linker is necessary for regulation during transcription and viral DNA replication of HPV11 E2 (40).

2.2. Overview of E2 protein DNA binding

The E2 protein binds to specific DNA sequences in the viral long control regions (LCR), thereby regulating transcription of viral genes (Figure 2A). The consensus recognition sequence is ACCG *NNNN*CGGT where highly conserved four base pair sequences flank a

four base pair ‘spacer’; the E2/D homodimer binds these sequences with nM affinity (Figure 2; (20, 21, 23, 25, 26, 42, 43)). The backbone and side chains of the recognition helix mediate direct sequence-specific contacts with the DNA (Figure 2; (23)) while the variable nucleotides of the central spacer region are not contacted (20, 21, 25). The sequence of the spacer is variable and profoundly modulates the E2 protein binding affinity (26, 27, 44-46). As will be discussed in detail later in this article, unique conformational and/or dynamic properties of the spacer sequence modulates the relative affinity of E2 proteins for the binding sites present in the viral genomes (20, 27, 47-49).

3. CONSERVED RESIDUES YIELD CONSERVED STRUCTURE AND FUNCTION

Papillomaviruses can be grouped into three phylogenetic clusters designated α , β , and other. Figure 3 summarizes the differences in amino acid conservation among the three groups for the E2/D by mapping the degree of conservation onto the HPV16 structure¹ These representations show that significant conservation is only present at the DNA binding and dimerization interfaces (Fig 3A & 3C). The alpha genus is the most conserved. Inspection of atomic resolution structures of the E2/D reveals that absolutely conserved Gly293 is located in the loop connecting the recognition helix to the strands of the beta-barrel (Figure 4A). Glycine residues are well known to reside within tight turns where side chains larger than a hydrogen atom would sterically clash with adjacent side chains (50-53). Mutating Gly293 to Val and Phe *in silico* disrupts the predicted structure confirming that spatial constraints preclude insertion of a larger side chain into the E2 protein at this position. Molecular models of the G293V and G293F (the amino acid numbering is based on HPV16 E2 structure assignment (22)) proteins reveal a shift in the recognition helix that prevents it from contacting the DNA without distortion of one of the macromolecules (Figure 4B).

Other residues are also highly conserved among all the aligned papillomavirus sequences. Residues with > 90% conservation include: Asn296, Lys299, Cys300 and Gln349 that are located on the surface of the recognition helix and mediate direct contact with the completely conserved nucleotides of the palindromic E2 recognition sequence (Figure 2B). Only conservative substitutions Glu for Asn, Arg for Lys, Ser for Cys and Glu (predominantly) for Gln occur at these positions (Appendix 1). The residues forming the dimerization interface of the E2 DNA binding domain are also highly conserved. For example, Ser313 resides within a loop that contributes to the stabilization of the dimerization interface and is conserved in 88% of the analyzed sequences. Again, substitutions are conservative; Thr substituted for Ser in most cases except for Ile in one HPV type (ChPV; Appendix 1). Trp319, Trp321 and Pro353 participate in intersubunit contacts and are invariant in > 90% of the papillomavirus types. Val333 and Leu335 are located within the subunit interface and are characterized by ~ 75% conservation.

3.1. The electrostatic surface of E2/D types and variants

The high degree of sequence identity among the E2 DNA binding domains of 122 HPV types and 24 HPV16 variants analyzed suggests corresponding conservation of structure (54). This hypothesis is supported by the low variability among the amino acids that dock the recognition alpha-helix on the beta-barrel and form the dimer interface as discussed above. To confirm the expected conservation of structure, homology models (54) of each of 122 viral types were determined using crystal structures from the alpha genus and other papillomavirus genera as templates (Figure 5). The average root means square deviation (RMSD) of the modeled structures is 2 angstroms indicating that they are essentially the

¹Methods used in studies not previously published are described in the legends to the figures.

same. No deviations were observed from the overall fold of the E2 DNA binding domain (data not shown). Thus, the amino acid variations of the studied virus types do not compromise the integrity of the domain's overall fold that is crucial for virus viability.

A subtle, but potentially important variability among the papillomavirus types is the nature of the protein-DNA interface. The electrostatic potential of the DNA binding surfaces of HPV16 and HPV18 E2 proteins differ with regard to both net charge and charge distribution (20). Since electrostatic interactions contribute significantly to protein-DNA interactions in general, and in a unique way to the E2 protein in particular (27), we used molecular modeling to explore differences in the electrostatic potential. The goal of this analysis was to assess whether the E2 DNA binding surfaces from HPV types defined as oncogenic (*i.e.*, high risk) (5) have unique characteristics that might contribute to their ability to cause disease.

We utilized the E2/D homology models from α papillomavirus types to compare the electrostatic potential of the E2 DNA binding surface. An electrostatic potential map was generated for each model by centering the domain within a $65 \times 65 \times 65$ Å cubic lattice with 1.5 angstrom spacing. We limited our analysis to the DNA binding surfaces and the alpha papillomaviruses. Each surface is analyzed against all the others to generate a similarity index (SI); zero (black) and one (white) denote dissimilar and identical surfaces, respectively (Figure 6A). The SI is a composite measure of similarity in net charge and charge distribution among the analyzed DNA binding surfaces. It is not surprising given the high degree of amino acid sequence similarity that the overall electropositive nature of the surface is conserved; the smallest SI among the surface potentials of the DNA binding surfaces of about 0.65 (Figure 6A).

Figure 6B compares the electrostatic potential maps of the DNA binding surfaces of the HPV6 (low risk) and HPV16 (high risk) E2 proteins whose SI is 0.70 among the most diverse pairs of proteins (Figure 6A); blue denotes a positive and red denotes a negative potential. While the HPV16 surface contains an increased electropositive surface potential within the middle of the DNA binding surface and the surface containing the recognition helix, a clear distinction between the high and low risk human mucosal papillomavirus types surfaces is not observed. The overall differences are minor corresponding to the magnitude of the potentials not their distribution. The SI scores denote strong electrostatic conservation with no correlation with epidemiological classifications. The overall conclusion drawn from this analysis is that the electrostatic nature of the DNA binding surface of the E2 proteins is highly conserved and lacks the heterogeneity necessary to explain the oncogenic potential of the mucosal human papillomaviruses.

3.2. Transactivation domain sequence conservation at the E2 - E1 Interaction Interface

The assembly of the E1 and E2 proteins at the viral origin is required for the initiation of papillomavirus DNA replication (55, 56). The E1 protein is a helicase that on its own, binds with low affinity and specificity to the origin of replication; specific binding is accomplished by the cooperative binding of the E1 and the E2 proteins to adjacent sites. Once the complex is formed, E2 is displaced and additional E1 molecules are added to the origin in an ATP dependent step (33, 57). The interaction of the HPV E1 helicase and E2 transactivation domains is well defined crystallographically (Figure 8B; (33)). It is less certain whether the DNA binding domain of the human E2 protein interacts with E1. Three bases separate the E1 and E2 binding sites at the origin or replication in BPV1; this close proximity is highly suggestive that the DNA binding domains of the two proteins interact (15, 33, 58). In contrast, the alpha papillomavirus E1 and E2 binding sites are further separated; significant distortion of the DNA would be required to bring the bound DNA binding domains into direct contact (15)

An amino acid sequence alignment analysis comparable to that described above for the E2/D was performed for the E2 transactivation domain to explore conservation in relation to the domain's three-dimensional structure. The alignment revealed similarity between any two PV of 36 - 100% among the transactivation domains indicating structural conservation among the papillomavirus types and variants comparable to that seen for the E2/D. Unlike the E2/D, the transactivation domain does not have extensive surfaces of highly conserved residues. Rather, the regions of very high conservation are localized to small surface patches (Figure 8). The residues Pro60, Ile73 and Gly156 are absolutely conserved (Figure 7 & Figure 8, colored red). Pro60 is within an alpha-turn-alpha motif. Glycine 156 is within a beta-turn-beta motif. Both residues are likely required for the stability of their respective turns and thus, the maintenance of the functional conformation of the domain.

High conservation is observed for a number of residues including several implicated in the interaction between the HPV11 E1 and E2 proteins (59). Tyr19, Glu20, Trp33 and Lys93 cluster around a pocket that binds a peptide which inhibits the E1-E2 interaction (Figure 3D, surrounding the green peptide; (59)). Mutation of Tyr19 to Ala inhibited the E2-E1 interaction (59). Arg37, Ala69 and Ile73 form a second cluster of residues on the opposite surface of the transactivation domain (Figure 3D, asterisk). Mutagenesis of Ile73 to Ala inhibits the E2-inhibitor interaction suggesting E1-E2 interference and that the surface defined by this cluster of conserved residues participates in the inter-protein interaction (59). Arg37 and Ile73 residues were found to be crucial for interaction with Brd4 (60). It has been suggested that this conserved surface helps regulate viral gene transcription (60-62). Highly conserved residues for which no functional correlation is available are Val59, Trp134 and Phe170.

4. ANALYSIS OF HPV16 VARIANTS

Twenty four variants of HPV16 have been identified by clinical screening and fully sequenced (12, 63). These variants are closely related and group into five phylogenetic branches designated European (E), Asian (As), Asian American (AA), African-1 (Af1) and African-2 (Af2) (Figure 7; (12)). An increased risk of squamous cell cervical carcinomas and its precursor high grade lesions is associated with non-European (NE) variants of HPV16 (7-9, 64). Patients infected with the non-European variants were 11 times more likely to be diagnosed with cervical cancer relative to infection with the prototype European-related HPV16 variants (7). The correlation between viral variant and disease may be due to differences in transcriptional regulation, the biological activities of the proteins encoded by HPV16 variants or in the ability of the host to mount an immunological response to specific viral epitopes (65). The summary of this analysis shown in Figure 7 and includes the transactivation domain, the 'linker' region connecting the two domains as well as the DNA binding domain. Since there is > 90% identity among the HPV16 variants, only substitutions relative to the prototype HPV16 E2pro reference sequence are shown (66).

The HPV16 variants can be dichotomized into European and Non-European clinical isolates (Figure 7); some variants denoted in Figure 7 as transitional, show characteristics of both the European and Non-European groups. The shared characteristics may be due to a common ancestor followed by separate evolution. The DNA binding domain of the Non-European variants are characterized by T310K (also noted in two European variants), W341C, and D344E amino acid variations. The residues W341C and D344E map to the surface of this domain that interacts with the p53 protein and can thus affect apoptosis (Figure 8A; (67)). The T310K variation may increase the electropositive character of the recognition helix and is also noted in two European HPV16 variants. The transactivation domain substitutions H35Q, T135K, H136Y and R165Q that are characteristic of the Non-European variants map to surfaces implicated in the E2 - E1 interaction at the origin of replication (Figure 8B; (31,

33)). This analysis shows that while there is > 93% amino acid identity among the E2 genes of the 24 HPV16 variants analyzed, the 7% variation between the European and Non-European viruses predominantly maps to the E2 - E1 interaction surface necessary for the initiation of viral replication. This clustering of the variations to an interface critical to the viral life cycle suggests functional significance.

We also asked whether the HPV16 E1 protein also has amino acid substitutions unique to the European and Non-European variants and if so, whether they mapped to the proposed inter-protein surfaces. Our analysis of nine E1 protein sequences (available at the time), revealed the amino acid substitutions Q78E, C168S, I326M and E452D. A structural correlation can be drawn only for E452D as it is located within the E1 helicase domain for which structural information is available. This residue is on the E2 - E1 interaction surface consistent with the conclusion drawn from the analysis of the E2 protein variants (Figure 7 & 8B; (59)).

Lastly, we note a difference in the E2/D between the European and Non-European HPV16 variants. These amino acid substitutions are likely to affect the cooperative binding of E1 and E2 at the replication origin and promoters. It is possible that alteration of the balance between viral replication and expression of the E6 and E7 oncoproteins might play a role in the increased oncogenic potential of the non-European viruses towards the development of high grade squamous intraepithelial lesions of the cervix (68, 69) and life threatening malignancy (10).

5. ROLE OF THE E2 PROTEIN IN MALIGNANCY AND ITS INTERACTION WITH P53

In cervical cancer, the genomes of high risk HPV types are often integrated into the host genome disrupting the E2 open reading frame (15, 55, 70-72). Since the E6 and E7 open reading frames remain intact in the integrated genome, through loss of E2 expression, they can be de-repressed resulting in expression of the E6 and E7 oncoproteins. These events abrogate cell cycle control, favor cell proliferation and thus contribute to oncogenesis (15, 20, 55, 67). Recent studies suggest that the HPV16 E2 proteins might also regulate cell proliferation and cell death through a direct interaction with p53 that induces apoptotic cell death (15, 67, 73, 74). It has been suggested that the E2 protein of high risk HPV's may also function as a tumor suppressor protein (15, 74). In contrast, the E2 protein from low risk human papillomaviruses such as HPV6 and HPV11 do not bind p53 (67). The three residues implicated in the E2 - p53 interaction, by alanine mutagenesis, are W341, D344 and D338. Mutation of these residues in HPV16 eliminates the E2 - p53 interaction and the induction of apoptosis in non-HPV transformed cell lines (67). Thus, the amino acid variations W341C and D344E (Figure 8A) that distinguish the European and Non European variants may influence the balance of proapoptotic signals by altering their interaction with p53; direct biochemical studies will be necessary to validate this hypothesis.

6. E2/D- DNA AFFINITY AND SPECIFICITY

The ability of proteins to 'read' DNA sequence is the net result of noncovalent interactions that include formation of enthalpically favorable protein-DNA contacts, entropically favorable release of bound water and ions and conformational changes in either or both partners. Base-specific interactions between protein and DNA, such as hydrogen bonds inferred from atomic resolution structures, are typically referred to as 'direct readout'. It is not unusual for conformational changes in either or both macromolecules to improve the configuration of direct interactions. In some cases, the propensity of duplex DNA to assume or change conformation is dependent upon the properties of nucleotides that do not directly

contact the protein. These contributions to binding are typically referred to as ‘indirect readout’ (20, 27, 75-78). The E2 protein utilizes both direct and indirect readout to bind its recognition sequences (Figure 1) (20).

The high level of primary sequence conservation in the E2/D results in conservation of tertiary structure and the amino acid residues that mediate direct interactions as discussed in the preceding sections (Figure 3). This conservation extends to the DNA sequence that is bound by the E2 protein; the two-fold symmetric four base pair sequences of the palindromic binding site directly bound by the E2 protein (20, 21, 23, 25, 26, 42, 43) are virtually invariant among papillomavirus genomes. Significant variability is only observed for positions -3 and +3 among the viruses analyzed in this study (Figure 2b). A little over half of the binding sites have palindromic C and G, respectively, at these two positions. The frequency for a C in position +3 or a G in position -3 is 61% and 65%, respectively. Only 10% of the remaining possible combinations of nucleotides are palindromes. The frequency of T, A or C is no greater than 15% at either position; the directly contacted base pairs are highly conserved. Taken together, these observations show that the direct component of DNA sequence specific binding by the E2 protein is critical to PV biology and has thus been ‘locked in’ by evolution. ‘Fine tuning’ of the affinity, structure and dynamics of the protein-DNA interaction can be attributed to the ‘indirect’ component of the reaction (26, 27).

The nature of the E2-DNA complex interface has been studied using a number of point mutations designed to effect amino acids directly within the interface of this complex (28). According to this analysis, the sum of the individual amino acid contributions differs by about 1.0 kcal/mol, or roughly 10% of the interaction energy was due to indirect readout. This study suggests that more water molecules are present at the molecular interface than visualized crystallographically for HPV18-E2/D and BPV1-E2/D DNA complexes (20). Solvent is an important component of the E2-DNA complex (27). The relative contributions of direct and indirect interactions are likely to be dependent on solution conditions. As discussed below, the contribution of indirect readout to binding can be gleaned from careful studies conducted as a function of salt concentration and type (Figure 9; (27)).

Although the cognate ‘spacer’ sequence is not contacted by the protein, spacer sequence differences among the binding sites present in each HPV genome result in 30 - 100 fold changes in binding affinity (26, 27). For example, the E2 cognate binding sites containing spacer sequences AATT, TTAA and ACGT have distinct structural propensities (29, 47). An analysis combining gel electrophoretic mobility measurements with X-ray crystallographic analysis and theoretical structural prediction has shown that DNA sequences containing AATT are curved $\sim 17^\circ$ while those containing TTAA are curved by ~ 11 degrees (47)². Net curvature is not observed for the ACGT sequence.

The sequence dependent effects of complex formation by the E2/D-DNA interaction have also been studied by directly comparing E2/D binding from HPV-16 and BPV-1 types. Utilizing quantitative gel-mobility shift experiment as well as solution equilibrium experiments, it was shown that the BPV-1 E2/D has moderate sensitivity to the sequence of the spacer, while the HPV-16 E2/D has a clear preference (30 - 100 fold greater) to spacer sequences rich in A:T base pairing, especially in high monovalent cation concentration or in the presence of divalent cations (Figure 9; (26, 27)). Nicked and gapped DNA sequences in the spacer region are detrimental to HPV-16 E2 binding, whereas only minimal effects in BPV-1 E2 binding were detected (26). Extending the consensus binding site by adding an

²There is some ambiguity regarding the structure of DNA containing the TTAA sequence; cyclization kinetics analysis indicates that an E2 binding-site containing this spacer sequence is flexible but not curved (29).

AT or a GC base pair to each end results in tighter binding affinities for the HPV16 E2 DNA binding domain when compared to adding a CG or TA base pair to the identical sites (79).

The structure analysis of DNA sequences A_4T_4 and T_4A_4 show monovalent cation-dependant bending manifest as changes in NMR signal and electrophoretic mobility (80). Divalent cation dependence for A-tract sequences has been reported with Mg^{2+} increasing the angle of curvature by ~ 2 fold (81). These reports show that the structures of the E2 binding sites are likely to differ based on the spacer sequence, and that monovalent and divalent cations play an important role in determining the structural conformation of a given DNA sequence. Analyzing DNA structure, by utilizing the cyclization method, enabled the prediction variations in E2/D affinity for the cognate sites (29). The predictive ability proved to be correct in 15 of 16 sequences, with the sole exception being traced to differential magnesium ion binding. These results further highlight the importance of indirect readout with regard to both DNA structure as well as the role of ions in sequence specific affinity and specificity (27, 29).

Cations penetrate within the grooves of duplex DNA; high resolution crystal structures show K^+ , Rb^+ and Cs^+ ions within the DNA duplex minor groove's 'spine of hydration' (82, 83). Molecular dynamics simulations show fractional occupancy of cations within the minor groove of the Dickerson dodecamer duplex (84, 85). An analysis of the electrostatic potential of this sequence identified a highly negative electrostatic potential within the 'ApT pocket' of the minor groove (86, 87). Cations are observed within the minor grooves of DNA duplexes bearing either A_nT_n or T_nA_n by NMR although their localization within the two sequences differs (88-90). Additional experimental and theoretical studies show monovalent cations localized deep within and near the top of the minor grooves of AT rich sequences, especially A tracts (84, 91-94).

The cations localized within the minor groove are hypothesized to reduce repulsion between proximal phosphates and the electronegative O2 of thymine and N3 of adenine on the groove's floor, resulting in narrowing of the groove width and facilitating bending of the helical axis (86, 95). The minor groove widths of free E2 binding sites are 9.4 Å and 12.1 Å, respectively, for the AATT and ACGT spacers (48) (Figure 1C). The structural differences of AT rich sequences as well as distinct differences in spacer sequences of cognate binding sites allows for a thorough exploration of structural effects of DNA in E2 protein binding affinity and specificity.

Biochemical and computational studies provide compelling evidence that the structure, dynamics and flexibility of the spacer DNA are a critical determinant of E2 binding affinity (26, 29). The E2 proteins in general cause a bend ~42 degrees to the spacer region of the DNA and bind to their cognate DNA sequences as homo-dimers (20). The available evidence suggests that full length E2 proteins and their isolated DNA-binding domains display comparable specificity for DNA sequence and bind to the DNA in the same manner (96). Peptides derived from the alpha1 (recognition) helix of 18 amino acid bind to the cognate sequence and not with non-specific DNA (97). This analysis revealed the propensity for this derived peptide to bind to the ACCG half site, demonstrating a capacity for discrimination of nucleic acid sequences without the need for the entire protein architecture (97).

Cations play a key role in E2/D sequence specific binding and affinity and the sequence-specific uptake of cations into the DNA upon binding of the E2 proteins is a key contribution to this binding-site discrimination (Figure 9; (27)). Augmenting the cation concentration increases the affinity of the E2 DNA binding domain for pre-bent sequences containing AT rich spacers (27). Furthermore, divalent cations also revealed an increase in

affinity and specificity for human papillomavirus type 16 E2 DNA binding domain when bound to cognate binding sites containing AT rich spacers. Thus, divalent cations in the intracellular milieu are essential to the ability of HPV16 E2 protein to discriminate among binding sites with different spacer sequences.

The mechanism for DNA binding utilized by E2 is distinct from the generally observed displacement of the cations condensed around DNA, neutralizing its highly negative charge, upon the binding of proteins (98-100). The E2 protein thus utilizes a novel mechanism of indirect readout in which cations penetrate into the grooves of the bound DNA's minor groove (27). These cations neutralize the highly electronegative charge density within the minor groove of the spacer DNA resulting from its distortion from a canonical B-helix induced by E2 binding (20, 27, 87).

Of the multitude of mechanisms that proteins use to recognize specific sequences of DNA, indirect readout is particularly intriguing since it is based upon their ability to distinguish subtle aspects of nucleic acid structure and dynamics. The results highlight differences in the contribution of electrostatics to spacer sequence discrimination by E2 DNA binding domains. Since the levels of K^+ and Mg^{2+} are homeostatically regulated in mammalian cells, the cation dependence of binding is unlikely to be a direct regulator of the papillomavirus life-cycle. However, these dependences illuminate aspects of the underlying mechanism of DNA sequence discrimination by the E2 proteins that may differ among the various papillomavirus types.

6.1. Computational analysis OF E2 structure and DNA binding

Insight into the contribution of DNA deformation to formation of E2/D-DNA complexes has been obtained through molecular dynamic simulations of the free and E2/D-bound DNA (101-104). Simulations of DNA containing the ACCGAATTCGGT E2 binding sequence that is tightly bound by HPV E2 proteins were run from the uncomplexed or E2/D-complexed starting coordinates. Both simulations rapidly relaxed to the dynamical structure represented by the crystal structure of the free DNA. This result shows that the structure of the bound DNA sequence is dynamically unstable in the absence of protein and arises as a consequence of conformational changes induced by the E2/D (101). Comparison of these simulations with those of an ideal canonical B form structure of the same sequence indicates a propensity for DNA bending to occur in the direction of the protein induced conformational changes. Since the free DNA structure containing the AATT spacer sequence is bent in the direction of the E2/D-induced conformational change, complex formation to some sequences is a consequence of both intrinsic DNA structure as well as protein induced structural change. The indirect readout mechanism manifests itself through the intrinsic structure and the flexibility of the sequence (101).

Simulations have also been conducted to explore the relative flexibility of the spacer and conserved half-sites of the E2 binding sequence. The ACGT spacer is more flexible than AAAC spacers especially in the backbone dynamics of the CpG step (102). The higher affinity of BPV1 E2/D for sites with the ACGT spacer is thus, likely due to the lesser penalty incurred in deforming the sequence upon protein binding. It was noted that the conserved half-sites behave identically and adapt conformations close to those seen in the bound conformations regardless of the spacer sequences present (104). Thus, the E2 proteins may take advantage of the invariant flanking half sites to form an initial complex whose spacer sequence subsequently relaxes to its final conformation (104). The overall effect of E2/D binding is to diminish global DNA motion and to impose and lock base displacements and helix curvature.

6.2. Modulation of E2 protein binding

Cytosine methylation at CpG dinucleotides influences transcription and replication of DNA in eukaryotic organisms. DNA methylation is hypothesized to be involved in silencing gene expression; the pattern of methylation is thought to reflect the gene expression profile of a cell. The E2 protein of papillomaviruses contributes to viral transcription and viral DNA replication, all of which are dependent on its ability to bind the consensus sequences located within the long control region (LCR; Figure 2A). E2 binding sites are potential sites for DNA methylation in the mammalian host cell because they contain CpG dinucleotides.

In vitro studies have shown that methylation of the CpG dinucleotides contained within the binding site block binding of the HPV16 E2 protein. Methylation in all four sites of the E2 binding site abolished DNA binding; partially methylated sites decrease but do not abolish specific binding (105). More recent studies have indicated that the ability of E2 proteins to activate transcription is inhibited by global methylation of CpG dinucleotides (106). Furthermore, in studies that detected HPV16 LCR methylation, hypomethylation was present in well differentiated epithelial cells; in contrast, hypermethylation was present in poorly differentiated epithelial cells (106). This data suggests that DNA methylation may play an important role in modulating both transcription and replication and in turn the viral life cycle. Studies on HPV18 have shown that the methylation of CpG steps repress promoter activity (107). Since the methylation state of the viral genome within a mammalian cell may vary during its life cycle, DNA binding by the E2 protein and hence its function can be modulated by DNA methylation.

7. SIGNIFICANCE

Non-covalent associations between proteins and their specific target sites on DNA plays a pivotal role in replication, transcription and replication, where the underlying mechanism involves molecular processes of protein-nucleic acid recognition and protein-protein association. The overall specificity and affinity of proteins to their target DNA involves a balancing act of diverse competing free energy components, including electrostatics, hydrogen bonding, ion and water release, and van der Waals contributions (108). Formation of sequence-specific protein-DNA complexes can also be viewed as a melding of direct and indirect readout contributions that adds up to an overall favorable free energy of binding (109). The structure and flexibility of the target DNA (indirect readout) plays a crucial role in determining the binding specificity and affinity of the E2-protein (27). The E2 protein presents a novel class of DNA binding protein that utilizes a novel mechanism of indirect readout in which cations penetrate into the grooves of the bound DNA's minor groove for sequence specific binding and affinity. These cations neutralize the highly electronegative charge density within the minor groove of the spacer DNA resulting from its distortion from a canonical B-helix induced by E2 binding (20, 27, 87).

The DNA binding domain of the papillomavirus E2 protein is a prototype of a novel structural class of DNA-binding proteins (20). With the evolutionary and molecular modeling investigations we have been able to extend our investigation to 146 papillomaviruses and found that the 'direct' component of DNA sequence specific binding by the E2 protein appears to have been maintained throughout evolution. Thus, the fine tuning of the affinity, structure and dynamics of the protein-DNA interaction can be attributed to the 'indirect' component of the reaction (26, 27).

The functional properties of the E2 protein are crucial to the viral life cycle via its regulation of gene transcription and DNA replication. As such, the conservation of the E2/D could have therapeutic implications for HPV infection and disease. Papillomavirus vaccines recently approved by the FDA are very effective in lowering viral load and preventing

disease (110-113). These vaccines target the L1 protein of HPV16 and 18 that together account for ~ 70% of infections that lead to malignancies. Since the L1 targeted vaccines are type specific, they do not offer protection against the remaining viral infections that can cause malignancy. The E2 protein has been tested as a possible vaccine target in rabbits; this test vaccine lowers viral load and reduces tumor size (114). In one report, the immune responses to the E2, E6 and E7 proteins is impaired (115); these patients cannot mount a T-cell response to these antigens and are at a higher risk for disease progression (115). In contrast, specific antibodies have been raised against a HPV16 E2-DNA complex (116) showing that the human E2 protein is immunogenic. Since the structural properties of the 'high risk' human papillomavirus E2 proteins are highly conserved and immunogenic, the E2 protein should be evaluated as a vaccine candidate for prophylactic protection against a broad spectrum of HPV types.

References

1. de Villiers EM, Fauquet C, Broker TR, Bernard HU, zur Hausen H. Classification of papillomaviruses. *Virology*. 2004; 324(1):17–27. [PubMed: 15183049]
2. Kadish AS, Burk RD, Kress Y, Calderin S, Romney SL. Human papillomaviruses of different types in precancerous lesions of the uterine cervix: Histologic, immunocytochemical and ultrastructural studies. *Hum Pathol*. 1986; 17(4):384–92. [PubMed: 3007326]
3. Van Ranst M, Fuse A, Fiten P, Beuken E, Pfister H, Burk RD, Opendakker G. Human papillomavirus type 13 and pygmy chimpanzee papillomavirus type 1: Comparison of the genome organizations. *Virology*. 1992; 190:587–596. [PubMed: 1325697]
4. Narechania A, Chen Z, DeSalle R, Burk RD. Phylogenetic incongruence among oncogenic genital alpha human papillomaviruses. *J Virol*. 2005; 79(24):15503–10. [PubMed: 16306621]
5. Munoz N, Bosch FX, de Sanjose S, Herrero R, Castellsague X, Shah KV, Snijders PJ, Meijer CJ. Epidemiologic classification of human papillomavirus types associated with cervical cancer. *N Engl J Med*. 2003; 348(6):518–27. [PubMed: 12571259]
6. Munger K, Howley PM. Human papillomavirus immortalization and transformation functions. *Virus Res*. 2002; 89(2):213–28. [PubMed: 12445661]
7. Hildesheim A, Schiffman M, Bromley C, Wacholder S, Herrero R, Rodriguez A, Bratti MC, Sherman ME, Scarpidis U, Lin QQ, Terai M, Bromley RL, Buetow K, Apple RJ, Burk RD. Human papillomavirus type 16 variants and risk of cervical cancer. *J Natl Cancer Inst*. 2001; 93(4):315–8. [PubMed: 11181779]
8. Villa LL, Sichero L, Rahal P, Caballero O, Ferenczy A, Rohan T, Franco EL. Molecular variants of human papillomavirus types 16 and 18 preferentially associated with cervical neoplasia. *J Gen Virol*. 2000; 81(Pt 12):2959–68. [PubMed: 11086127]
9. Xi LF, Koutsky LA, Galloway DA, Kuypers J, Hughes JP, Wheeler CM, Holmes KK, Kiviat NB. Genomic variation of human papillomavirus type 16 and risk for high grade cervical intraepithelial neoplasia. *J Natl Cancer Inst*. 1997; 89(11):796–802. [PubMed: 9182978]
10. Burk RD, Terai M, Gravitt PE, Brinton LA, Kurman RJ, Barnes WA, Greenberg MD, Haddjmichael OC, Fu L, McGowan L, Mortel R, Schwartz PE, Hildesheim A. Distribution of human papillomavirus types 16 and 18 variants in squamous cell carcinomas and adenocarcinomas of the cervix. *Cancer Res*. 2003; 63(21):7215–20. [PubMed: 14612516]
11. Terai M, Burk RD. Characterization of a novel genital human papillomavirus by overlapping pcr: Candhpv86 identified in cervicovaginal cells of a woman with cervical neoplasia. *J Gen Virol*. 2001; 82(Pt 9):2035–40. [PubMed: 11514712]
12. Yamada T, Manos MM, Peto J, Greer CE, Munoz N, Bosch FX, Wheeler CM. Human papillomavirus type 16 sequence variation in cervical cancers: A worldwide perspective. *J Virol*. 1997; 71(3):2463–72. [PubMed: 9032384]
13. Eriksson A, Herron JR, Yamada T, Wheeler CM. Human papillomavirus type 16 variant lineages characterized by nucleotide sequence analysis of the e5 coding segment and the e2 hinge region. *J Gen Virol*. 1999; 80(Pt 3):595–600. [PubMed: 10091997]

14. Bechtold V, Beard P, Raj K. Human papillomavirus type 16 e2 protein has no effect on transcription from episomal viral DNA. *J Virol.* 2003; 77(3):2021–8. [PubMed: 12525636]
15. Desaintes C, Demeret C. Control of papillomavirus DNA replication and transcription. *Semin Cancer Biol.* 1996; 7(6):339–47. [PubMed: 9284526]
16. Danos O, Katinka M, Yaniv M. Human papillomavirus 1a complete DNA sequence: A novel type of genome organization among papovaviridae. *Embo J.* 1982; 1(2):231–6. [PubMed: 6325156]
17. von Knebel Doeberitz M. Papillomaviruses in human disease: Part ii. Molecular biology and immunology of papillomavirus infections and carcinogenesis. *Eur J Med.* 1992; 1(8):485–91. [PubMed: 1341208]
18. Laimins LA. The biology of human papillomaviruses: From warts to cancer. *Infect Agents Dis.* 1993; 2(2):74–86. [PubMed: 8162357]
19. McMurray HR, Nguyen D, Westbrook TF, McAnce DJ. Biology of human papillomaviruses. *Int J Exp Pathol.* 2001; 82(1):15–33. [PubMed: 11422538]
20. Hegde RS. The papillomavirus e2 proteins: Structure, function, and biology. *Annu Rev Biophys Biomol Struct.* 2002; 31:343–60. [PubMed: 11988474]
21. Kim SS, Tam JK, Wang AF, Hegde RS. The structural basis of DNA target discrimination by papillomavirus e2 proteins. *J Biol Chem.* 2000; 275(40):31245–54. [PubMed: 10906136]
22. Hegde RS, Androphy EJ. Crystal structure of the e2 DNA-binding domain from human papillomavirus type 16: Implications for its DNA binding-site selection mechanism. *J Mol Biol.* 1998; 284(5):1479–89. [PubMed: 9878365]
23. Hegde RS, Grossman SR, Laimins LA, Sigler PB. Crystal structure at 1.7 Å of the bovine papillomavirus-1 e2 DNA-binding domain bound to its DNA target. *Nature.* 1992; 359(6395): 505–12. [PubMed: 1328886]
24. Bussiere DE, Kong X, Egan DA, Walter K, Holzman TF, Lindh F, Robins T, Giranda VL. Structure of the e2 DNA-binding domain from human papillomavirus serotype 31 at 2.4 Å. *Acta Crystallogr D Biol Crystallogr.* 1998; 54(2 (Pt 6)):1367–76. [PubMed: 10089498]
25. Hegde RS. Structure of the bpv-1 e2 DNA-binding domain bound to its DNA target. *J Nucl Med.* 1995; 36(6 Suppl):25S–27S. [PubMed: 7769463]
26. Hines CS, Meghoo C, Shetty S, Biburger M, Brenowitz M, Hegde RS. DNA structure and flexibility in the sequence-specific binding of papillomavirus e2 proteins. *J Mol Biol.* 1998; 276(4):809–18. [PubMed: 9500925]
27. Blakaj DM, Kattamuri C, Khrapunov S, Hegde RS, Brenowitz M. Indirect readout of DNA sequence by papillomavirus e2 proteins depends upon net cation uptake. *J Mol Biol.* 2006; 358(1): 224–40. [PubMed: 16513133]
28. Ferreira DU, Dellarole M, Nadra AD, de Prat-Gay G. Free energy contributions to direct readout of a DNA sequence. *J Biol Chem.* 2005; 280(37):32480–4. [PubMed: 16000299]
29. Zhang Y, Xi Z, Hegde RS, Shakked Z, Crothers DM. Predicting indirect readout effects in protein-DNA interactions. *Proc Natl Acad Sci U S A.* 2004; 101(22):8337–41. [PubMed: 15148366]
30. Dell G, Wilkinson KW, Tranter R, Parish J, Leo Brady R, Gaston K. Comparison of the structure and DNA-binding properties of the e2 proteins from an oncogenic and a non-oncogenic human papillomavirus. *J Mol Biol.* 2003; 334(5):979–91. [PubMed: 14643661]
31. Antson AA, Burns JE, Moroz OV, Scott DJ, Sanders CM, Bronstein IB, Dodson GG, Wilson KS, Maitland NJ. Structure of the intact transactivation domain of the human papillomavirus e2 protein. *Nature.* 2000; 403(6771):805–9. [PubMed: 10693813]
32. Harris SF, Botchan MR. Crystal structure of the human papillomavirus type 18 e2 activation domain. *Science.* 1999; 284(5420):1673–7. [PubMed: 10356398]
33. Abbate EA, Berger JM, Botchan MR. The x-ray structure of the papillomavirus helicase in complex with its molecular matchmaker e2. *Genes Dev.* 2004; 18(16):1981–96. [PubMed: 15289463]
34. Mok YK, de Prat Gay G, Butler PJ, Bycroft M. Equilibrium dissociation and unfolding of the dimeric human papillomavirus strain-16 e2 DNA-binding domain. *Protein Sci.* 1996; 5(2):310–9. [PubMed: 8745409]

35. Nadra AD, Eliseo T, Mok YK, Almeida CL, Bycroft M, Paci M, de Prat-Gay G, Cicero DO. Solution structure of the hpv-16 e2 DNA binding domain, a transcriptional regulator with a dimeric beta-barrel fold. *J Biomol NMR*. 2004; 30(2):211–4. [PubMed: 15702528]
36. Liang H, Petros AM, Meadows RP, Yoon HS, Egan DA, Walter K, Holzman TF, Robins T, Fesik SW. Solution structure of the DNA-binding domain of a human papillomavirus e2 protein: Evidence for flexible DNA-binding regions. *Biochemistry*. 1996; 35(7):2095–103. [PubMed: 8652551]
37. Veeraraghavan S, Mello CC, Androphy EJ, Baleja JD. Structural correlates for enhanced stability in the e2 DNA-binding domain from bovine papillomavirus. *Biochemistry*. 1999; 38(49):16115–24. [PubMed: 10587434]
38. Veeraraghavan S, Mello CC, Lee KM, Androphy EJ, Baleja JD. 1h, 15n, and 13c nmr resonance assignments for the DNA-binding domain of the bpv-1 e2 protein. *J Biomol NMR*. 1998; 11(4): 457–8. [PubMed: 9691287]
39. de Prat-Gay G, Nadra AD, Corrales-Izquierdo FJ, Alonso LG, Ferreira DU, Mok YK. The folding mechanism of a dimeric beta-barrel domain. *J Mol Biol*. 2005; 351(3):672–82. [PubMed: 16023675]
40. Zou N, Lin BY, Duan F, Lee KY, Jin G, Guan R, Yao G, Lefkowitz EJ, Broker TR, Chow LT. The hinge of the human papillomavirus type 11 e2 protein contains major determinants for nuclear localization and nuclear matrix association. *J Virol*. 2000; 74(8):3761–70. [PubMed: 10729151]
41. McBride AA, Romanczuk H, Howley PM. The papillomavirus e2 regulatory proteins. *J Biol Chem*. 1991; 266(28):18411–4. [PubMed: 1655748]
42. Herrington CS. Human papillomaviruses (hpv) in gynaecological cytology: From molecular biology to clinical testing. *Cytopathology*. 1995; 6(3):176–89. [PubMed: 7669929]
43. Rozenberg H, Rabinovich D, Frolow F, Hegde RS, Shakked Z. Structural code for DNA recognition revealed in crystal structures of papillomavirus e2-DNA targets. *Proc Natl Acad Sci U S A*. 1998; 95(26):15194–9. [PubMed: 9860945]
44. Remm M, Brain R, Jenkins JR. The e2 binding sites determine the efficiency of replication for the origin of human papillomavirus type 18. *Nucleic Acids Res*. 1992; 20(22):6015–21. [PubMed: 1334259]
45. Lu JZ, Sun YN, Rose RC, Bonnez W, McCance DJ. 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*. 1993; 67(12):7131–9. [PubMed: 8230435]
46. Sverdrup F, Khan SA. Two e2 binding sites alone are sufficient to function as the minimal origin of replication of human papillomavirus type 18 DNA. *J Virol*. 1995; 69(2):1319–23. [PubMed: 7815514]
47. Zimmerman JM, Maher LJ 3rd. Solution measurement of DNA curvature in papillomavirus e2 binding sites. *Nucleic Acids Res*. 2003; 31(17):5134–9. [PubMed: 12930964]
48. Hizver J, Rozenberg H, Frolow F, Rabinovich D, Shakked Z. DNA bending by an adenine--thymine tract and its role in gene regulation. *Proc Natl Acad Sci U S A*. 2001; 98(15):8490–5. [PubMed: 11438706]
49. Lewis H, Gaston K. Magnesium ions enhance the transfer of human papillomavirus e2 protein from non-specific to specific binding sites. *J Mol Biol*. 1999; 294(4):885–96. [PubMed: 10588894]
50. Tanaka T, Yokoyama S, Kuroda Y. Improvement of domain linker prediction by incorporating loop length-dependent characteristics. *Biopolymers*. 2005
51. Swindells MB, MacArthur MW, Thornton JM. Intrinsic phi, psi propensities of amino acids, derived from the coil regions of known structures. *Nat Struct Biol*. 1995; 2(7):596–603. [PubMed: 7664128]
52. Wilmot CM, Thornton JM. Beta-turns and their distortions: A proposed new nomenclature. *Protein Eng*. 1990; 3(6):479–93. [PubMed: 2371257]
53. Gunasekaran K, Ramakrishnan C, Balam P. Beta-hairpins in proteins revisited: Lessons for de novo design. *Protein Eng*. 1997; 10(10):1131–41. [PubMed: 9488138]
54. Fiser A, Feig M, Brooks CL 3rd, Sali A. Evolution and physics in comparative protein structure modeling. *Acc Chem Res*. 2002; 35(6):413–21. [PubMed: 12069626]

55. Hebner CM, Laimins LA. Human papillomaviruses: Basic mechanisms of pathogenesis and oncogenicity. *Rev Med Virol.* 2006; 16(2):83–97. [PubMed: 16287204]
56. Dell G, Gaston K. Human papillomaviruses and their role in cervical cancer. *Cell Mol Life Sci.* 2001; 58(12-13):1923–42. [PubMed: 11766888]
57. Chen G, Stenlund A. Two patches of amino acids on the e2 DNA binding domain define the surface for interaction with e1. *J Virol.* 2000; 74(3):1506–12. [PubMed: 10627562]
58. Berg M, Stenlund A. Functional interactions between papillomavirus e1 and e2 proteins. *J Virol.* 1997; 71(5):3853–63. [PubMed: 9094661]
59. Wang Y, Coulombe R, Cameron DR, Thauvette L, Massariol MJ, Amon LM, Fink D, Titolo S, Welchner E, Yoakim C, Archambault J, White PW. Crystal structure of the e2 transactivation domain of human papillomavirus type 11 bound to a protein interaction inhibitor. *J Biol Chem.* 2004; 279(8):6976–85. [PubMed: 14634007]
60. Senechal H, Poirier GG, Coulombe B, Laimins LA, Archambault J. Amino acid substitutions that specifically impair the transcriptional activity of papillomavirus e2 affect binding to the long isoform of brd4. *Virology.* 2007; 358(1):10–7. [PubMed: 17023018]
61. Schweiger MR, You J, Howley PM. Bromodomain protein 4 mediates the papillomavirus e2 transcriptional activation function. *J Virol.* 2006; 80(9):4276–85. [PubMed: 16611886]
62. Baxter MK, McPhillips MG, Ozato K, McBride AA. The mitotic chromosome binding activity of the papillomavirus e2 protein correlates with interaction with the cellular chromosomal protein, brd4. *J Virol.* 2005; 79(8):4806–18. [PubMed: 15795266]
63. Bernard HU. Controls in the papillomavirus life cycle. *FEMS Microbiol Immunol.* 1990; 2(4):201–5. [PubMed: 2178353]
64. Chen Z, Fu L, Herrero R, Schiffman M, Burk RD. Identification of a novel human papillomavirus (hpv97) related to hpv18 and hpv45. *Int J Cancer.* 2007; 121(1):193–8. [PubMed: 17351898]
65. Ordonez RM, Espinosa AM, Sanchez-Gonzalez DJ, Armendariz-Borunda J, Berumen J. Enhanced oncogenicity of asian-american human papillomavirus 16 is associated with impaired e2 repression of e6/e7 oncogene transcription. *J Gen Virol.* 2004; 85(Pt 6):1433–44. [PubMed: 15166426]
66. Seedorf K, Krammer G, Durst M, Suhai S, Rowekamp WG. Human papillomavirus type 16 DNA sequence. *Virology.* 1985; 145(1):181–5. [PubMed: 2990099]
67. Parish JL, Kowalczyk A, Chen HT, Roeder GE, Sessions R, Buckle M, Gaston K. E2 proteins from high- and low-risk human papillomavirus types differ in their ability to bind p53 and induce apoptotic cell death. *J Virol.* 2006; 80(9):4580–90. [PubMed: 16611918]
68. Giannoudis A, Herrington CS. Human papillomavirus variants and squamous neoplasia of the cervix. *J Pathol.* 2001; 193(3):295–302. [PubMed: 11241407]
69. Giannoudis A, Duin M, Snijders PJ, Herrington CS. Variation in the e2-binding domain of hpv 16 is associated with high-grade squamous intraepithelial lesions of the cervix. *Br J Cancer.* 2001; 84(8):1058–63. [PubMed: 11308254]
70. Durst M, Kleinheinz A, Hotz M, Gissmann L. The physical state of human papillomavirus type 16 DNA in benign and malignant genital tumours. *J Gen Virol.* 1985; 66(Pt 7):1515–22. [PubMed: 2991428]
71. Schwarz E, Freese UK, Gissmann L, Mayer W, Roggenbuck B, Stremlau A, zur Hausen H. Structure and transcription of human papillomavirus sequences in cervical carcinoma cells. *Nature.* 1985; 314(6006):111–4. [PubMed: 2983228]
72. Yee C, Krishnan-Hewlett I, Baker CC, Schlegel R, Howley PM. Presence and expression of human papillomavirus sequences in human cervical carcinoma cell lines. *Am J Pathol.* 1985; 119(3):361–6. [PubMed: 2990217]
73. Frattini MG, Hurst SD, Lim HB, Swaminathan S, Laimins LA. Abrogation of a mitotic checkpoint by e2 proteins from oncogenic human papillomaviruses correlates with increased turnover of the p53 tumor suppressor protein. *Embo J.* 1997; 16(2):318–31. [PubMed: 9029152]
74. Desaintes C, Demeret C, Goyat S, Yaniv M, Thierry F. Expression of the papillomavirus e2 protein in hela cells leads to apoptosis. *Embo J.* 1997; 16(3):504–14. [PubMed: 9034333]
75. Koudelka GB, Harrison SC, Ptashne M. Effect of non-contacted bases on the affinity of 434 operator for 434 repressor and cro. *Nature.* 1987; 326(6116):886–8. [PubMed: 3553960]

76. Koudelka GB, Carlson P. DNA twisting and the effects of non-contacted bases on affinity of 434 operator for 434 repressor. *Nature*. 1992; 355(6355):89–91. [PubMed: 1731202]
77. Rachofsky EL, Ross JB, Osman R. Conformation and dynamics of normal and damaged DNA. *Comb Chem High Throughput Screen*. 2001; 4(8):675–706. [PubMed: 11812262]
78. Aggarwal AK, Rodgers DW, Drottar M, Ptashne M, Harrison SC. Recognition of a DNA operator by the repressor of phage 434: A view at high resolution. *Science*. 1988; 242(4880):899–907. [PubMed: 3187531]
79. Thain A, Webster K, Emery D, Clarke AR, Gaston K. DNA binding and bending by the human papillomavirus type 16 e2 protein. Recognition of an extended binding site. *J Biol Chem*. 1997; 272(13):8236–42. [PubMed: 9079642]
80. Stefl R, Wu H, Ravindranathan S, Sklenar V, Feigon J. DNA a-tract bending in three dimensions: Solving the da4t4 vs. Dt4a4 conundrum. *Proc Natl Acad Sci U S A*. 2004; 101(5):1177–82. [PubMed: 14739342]
81. Tchernachenko V, Halvorson HR, Lutter LC. Topological measurement of an a-tract bend angle: Effect of magnesium. *J Mol Biol*. 2004; 341(1):55–63. [PubMed: 15312762]
82. Shui X, McFail-Isom L, Hu GG, Williams LD. The b-DNA dodecamer at high resolution reveals a spine of water on sodium. *Biochemistry*. 1998; 37(23):8341–55. [PubMed: 9622486]
83. Tereshko V, Minasov G, Egli M. A “hydrat-ion” spine in a b-DNA minor groove. *J Am Chem Soc*. 1999; 121:3590–3595.
84. McConnell KJ, Beveridge DL. Molecular dynamics simulations of b'-DNA: Sequence effects on a-tract-induced bending and flexibility. *J Mol Biol*. 2001; 314(1):23–40. [PubMed: 11724529]
85. McConnell KJ, Beveridge DL. DNA structure: What's in charge? *J Mol Biol*. 2000; 304(5):803–20. [PubMed: 11124028]
86. Lavery R, Pullman B. The molecular electrostatic potential, steric accessibility and hydration of dickerson's b-DNA dodecamer d (cpgpcpgpapaptptcpgpcpg). *Nucleic Acids Res*. 1981; 9(15):3765–77. [PubMed: 7279672]
87. Hud NV, Plavec J. A unified model for the origin of DNA sequence-directed curvature. *Biopolymers*. 2003; 69(1):144–58. [PubMed: 12717729]
88. Hud NV, Feigon J. Characterization of divalent cation localization in the minor groove of the a (n)t (n) and t (n)a (n) DNA sequence elements by (1)h nmr spectroscopy and manganese (ii). *Biochemistry*. 2002; 41(31):9900–10. [PubMed: 12146955]
89. Hud NV, Schultze P, Sklenar V, Feigon J. Binding sites and dynamics of ammonium ions in a telomere repeat DNA quadruplex. *J Mol Biol*. 1999; 285(1):233–43. [PubMed: 9878402]
90. Hud NV, Sklenar V, Feigon J. Localization of ammonium ions in the minor groove of DNA duplexes in solution and the origin of DNA a-tract bending. *J Mol Biol*. 1999; 286(3):651–60. [PubMed: 10024440]
91. Feigon J, Butcher SE, Finger LD, Hud NV. Solution nuclear magnetic resonance probing of cation binding sites on nucleic acids. *Methods Enzymol*. 2001; 338:400–20. [PubMed: 11460560]
92. Young MA, Jayaram B, Beveridge DL. Intrusion of counterions into the spine of hydration in the minor groove of b-DNA: Fractional occupancy of electronegative pockets. *J Am Chem Soc*. 1997; 119(1):59–69.
93. Shui X, Sines CC, McFail-Isom L, VanDerveer D, Williams LD. Structure of the potassium form of cgcgaattcgcg: DNA deformation by electrostatic collapse around inorganic cations. *Biochemistry*. 1998; 37(48):16877–87. [PubMed: 9836580]
94. McFail-Isom L, Sines CC, Williams LD. DNA structure: Cations in charge? *Curr Opin Struct Biol*. 1999; 9(3):298–304. [PubMed: 10361089]
95. Bergqvist S, O'Brien R, Ladbury JE. Site-specific cation binding mediates tata binding protein-DNA interaction from a hyperthermophilic archaeon. *Biochemistry*. 2001; 40(8):2419–25. [PubMed: 11327862]
96. McBride AA, Schlegel R, Howley PM. The carboxy-terminal domain shared by the bovine papillomavirus e2 transactivator and repressor proteins contains a specific DNA binding activity. *Embo J*. 1988; 7(2):533–9. [PubMed: 2835232]
97. Faber-Barata J, Mohana-Borges R, Lima LM. Specificity in DNA recognition by a peptide from papillomavirus e2 protein. *FEBS Lett*. 2006; 580(8):1919–24. [PubMed: 16542651]

98. Record MT Jr, Anderson CF, Lohman TM. Thermodynamic analysis of ion effects on the binding and conformational equilibria of proteins and nucleic acids: The roles of ion association or release, screening, and ion effects on water activity. *Q Rev Biophys.* 1978; 11(2):103–78. [PubMed: 353875]
99. Record MT Jr, deHaseth PL, Lohman TM. Interpretation of monovalent and divalent cation effects on the lac repressor-operator interaction. *Biochemistry.* 1977; 16(22):4791–6. [PubMed: 911790]
100. Record MT Jr, Lohman ML, De Haseth P. Ion effects on ligand-nucleic acid interactions. *J Mol Biol.* 1976; 107(2):145–58. [PubMed: 1003464]
101. Byun KS, Beveridge DL. Molecular dynamics simulations of papilloma virus e2 DNA sequences: Dynamical models for oligonucleotide structures in solution. *Biopolymers.* 2004; 73(3):369–79. [PubMed: 14755573]
102. Djuranovic D, Hartmann B. DNA fine structure and dynamics in crystals and in solution: The impact of bi/bii backbone conformations. *Biopolymers.* 2004; 73(3):356–68. [PubMed: 14755572]
103. Djuranovic D, Hartmann B. Molecular dynamics studies on free and bound targets of the bovine papillomavirus type i e2 protein: The protein binding effect on DNA and the recognition mechanism. *Biophys J.* 2005; 89(4):2542–51. [PubMed: 16055534]
104. Djuranovic D, Oguey C, Hartmann B. The role of DNA structure and dynamics in the recognition of bovine papillomavirus e2 protein target sequences. *J Mol Biol.* 2004; 339(4):785–96. [PubMed: 15165850]
105. Thain A, Jenkins O, Clarke AR, Gaston K. CpG methylation directly inhibits binding of the human papillomavirus type 16 e2 protein to specific DNA sequences. *J Virol.* 1996; 70(10): 7233–5. [PubMed: 8794373]
106. Kim K, Garner-Hamrick PA, Fisher C, Lee D, Lambert PF. Methylation patterns of papillomavirus DNA, its influence on e2 function, and implications in viral infection. *J Virol.* 2003; 77(23):12450–9. [PubMed: 14610169]
107. Rosl F, Arab A, Klevenz B, zur Hausen H. The effect of DNA methylation on gene regulation of human papillomaviruses. *J Gen Virol.* 1993; 74(Pt 5):791–801. [PubMed: 8388016]
108. Jayaram B, McConnell K, Dixit SB, Das A, Beveridge DL. Free-energy component analysis of 40 protein-DNA complexes: A consensus view on the thermodynamics of binding at the molecular level. *J Comput Chem.* 2002; 23(1):1–14. [PubMed: 11913374]
109. Jen-Jacobson L. Protein-DNA recognition complexes: Conservation of structure and binding energy in the transition state. *Biopolymers.* 1997; 44(2):153–80. [PubMed: 9354759]
110. Villa LL, Ault KA, Giuliano AR, Costa RL, Petta CA, Andrade RP, Brown DR, Ferenczy A, Harper DM, Koutsky LA, Kurman RJ, Lehtinen M, Malm C, Olsson SE, Ronnett BM, Skjeldestad FE, Steinwall M, Stoler MH, Wheeler CM, Taddeo FJ, Yu J, Lupinacci L, Railkar R, Marchese R, Esser MT, Bryan J, Jansen KU, Sings HL, Tamms GM, Saah AJ, Barr E. Immunologic responses following administration of a vaccine targeting human papillomavirus types 6, 11, 16, and 18. *Vaccine.* 2006
111. Harper DM, Franco EL, Wheeler CM, Moscicki AB, Romanowski B, Roteli-Martins CM, Jenkins D, Schuind A, Costa Clemens SA, Dubin G. Sustained efficacy up to 4.5 years of a bivalent 11 virus-like particle vaccine against human papillomavirus types 16 and 18: Follow-up from a randomised control trial. *Lancet.* 2006; 367(9518):1247–55. [PubMed: 16631880]
112. Villa LL, Costa RL, Petta CA, Andrade RP, Ault KA, Giuliano AR, Wheeler CM, Koutsky LA, Malm C, Lehtinen M, Skjeldestad FE, Olsson SE, Steinwall M, Brown DR, Kurman RJ, Ronnett BM, Stoler MH, Ferenczy A, Harper DM, Tamms GM, Yu J, Lupinacci L, Railkar R, Taddeo FJ, Jansen KU, Esser MT, Sings HL, Saah AJ, Barr E. Prophylactic quadrivalent human papillomavirus (types 6, 11, 16, and 18) 11 virus-like particle vaccine in young women: A randomised double-blind placebo-controlled multicentre phase ii efficacy trial. *Lancet Oncol.* 2005; 6(5):271–8. [PubMed: 15863374]
113. Lowy DR, Schiller JT. Prophylactic human papillomavirus vaccines. *J Clin Invest.* 2006; 116(5): 1167–73. [PubMed: 16670757]

114. Brandsma JL, Shlyankevich M, Zhang L, Slade MD, Goodwin EC, Peh W, Deisseroth AB. Vaccination of rabbits with an adenovirus vector expressing the papillomavirus e2 protein leads to clearance of papillomas and infection. *J Virol.* 2004; 78(1):116–23. [PubMed: 14671093]
115. de Jong A, van Poelgeest MI, van der Hulst JM, Drijfhout JW, Fleuren GJ, Melief CJ, Kenter G, Offringa R, van der Burg SH. Human papillomavirus type 16-positive cervical cancer is associated with impaired cd4+ t-cell immunity against early antigens e2 and e6. *Cancer Res.* 2004; 64(15):5449–55. [PubMed: 15289354]
116. Cerutti ML, Zarebski LM, de Prat Gay G, Goldbaum FA. A viral DNA-binding domain elicits anti-DNA antibodies of different specificities. *Mol Immunol.* 2005; 42(3):327–33. [PubMed: 15589321]
117. Rice P, Longden I, Bleasby A. Emboss: The european molecular biology open software suite. *Trends Genet.* 2000; 16(6):276–7. [PubMed: 10827456]
118. Thompson JD, Higgins DG, Gibson TJ. Clustal w: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 1994; 22(22):4673–80. [PubMed: 7984417]
119. Livingstone CD, Barton GJ. Protein sequence alignments: A strategy for the hierarchical analysis of residue conservation. *Comput Appl Biosci.* 1993; 9(6):745–56. [PubMed: 8143162]
120. Fiser A, Sali A. Modeller: Generation and refinement of homology-based protein structure models. *Methods Enzymol.* 2003; 374:461–91. [PubMed: 14696385]
121. Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol.* 1993; 234(3):779–815. [PubMed: 8254673]
122. Sippl MJ. Recognition of errors in three-dimensional structures of proteins. *Proteins.* 1993; 17(4):355–62. [PubMed: 8108378]
123. Laskowski RA, Moss DS, Thornton JM. Main-chain bond lengths and bond angles in protein structures. *J Mol Biol.* 1993; 231(4):1049–67. [PubMed: 8515464]
124. Russell RB, Barton GJ. Multiple protein sequence alignment from tertiary structure comparison: Assignment of global and residue confidence levels. *Proteins.* 1992; 14(2):309–23. [PubMed: 1409577]
125. Schleinkofer K, Wiedemann U, Otte L, Wang T, Krause G, Oschkinat H, Wade RC. Comparative structural and energetic analysis of ww domain-peptide interactions. *J Mol Biol.* 2004; 344(3):865–81. [PubMed: 15533451]
126. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc Natl Acad Sci U S A.* 2001; 98(18):10037–41. [PubMed: 11517324]

Abbreviations

HPV	human papillomavirus
E2/D	HPV E2 DNA binding domain
ORF	open reading frame
LCR	long control region
RMSD	root means square deviation
E	European
As	Asian
AA	Asian American
Af1	African-1
Af2	African-2
NE	non European

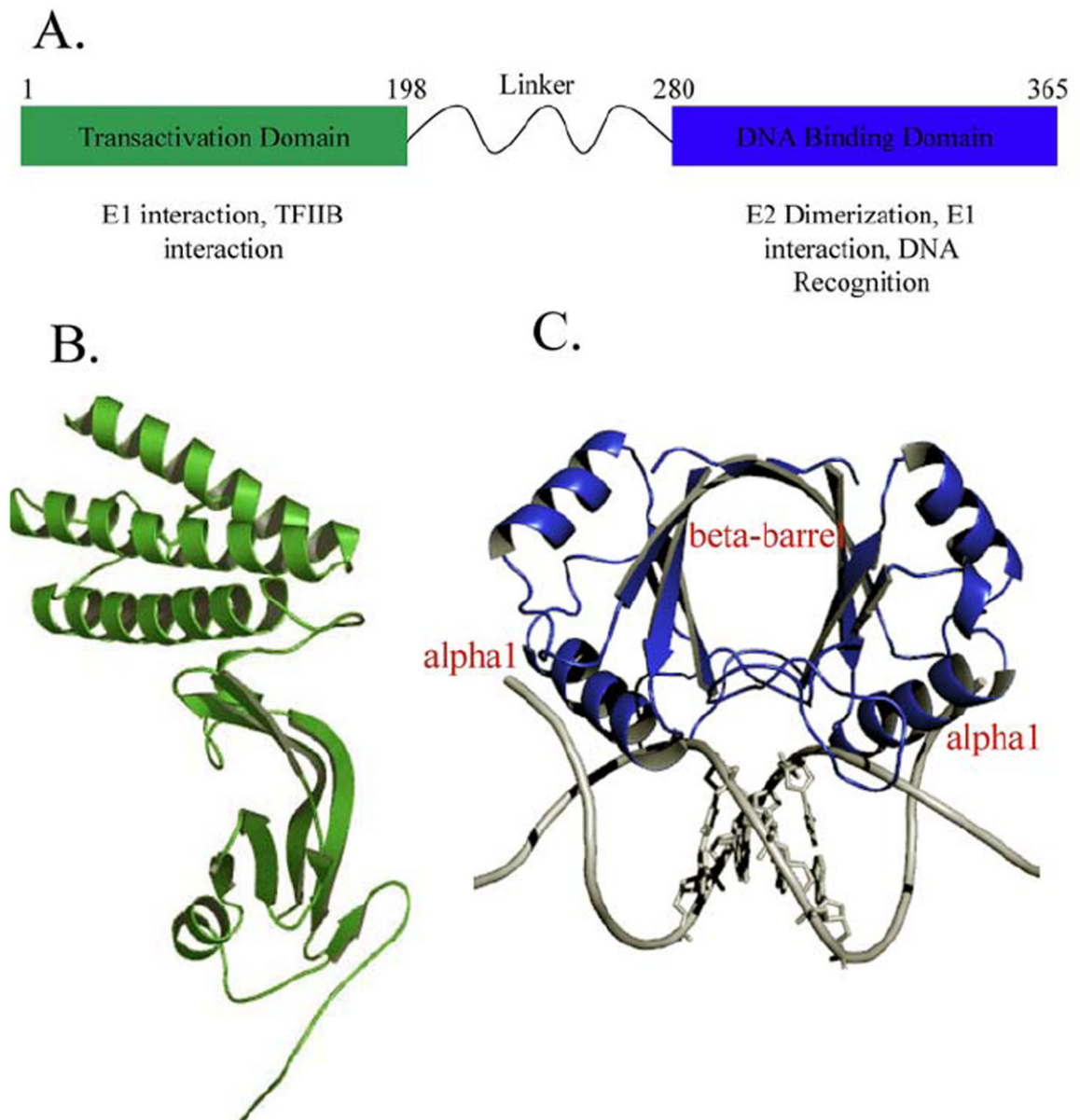


Figure 1.

A) Schematic of the HPV16 E2 gene and known proteins of interaction. B) Ribbon diagram of the structure of the HPV16 E2 transactivation domain (31). C) A ribbon representation of the structure of the HPV18 E2 DNA binding domain showing the dimeric protein bound to the DNA sequence ACCGAATTCGGT (PDB-ID: 1JJ4). The recognition helix alpha1 makes direct contact with the major groove of the DNA. The bases of the spacer region AATT are depicted.

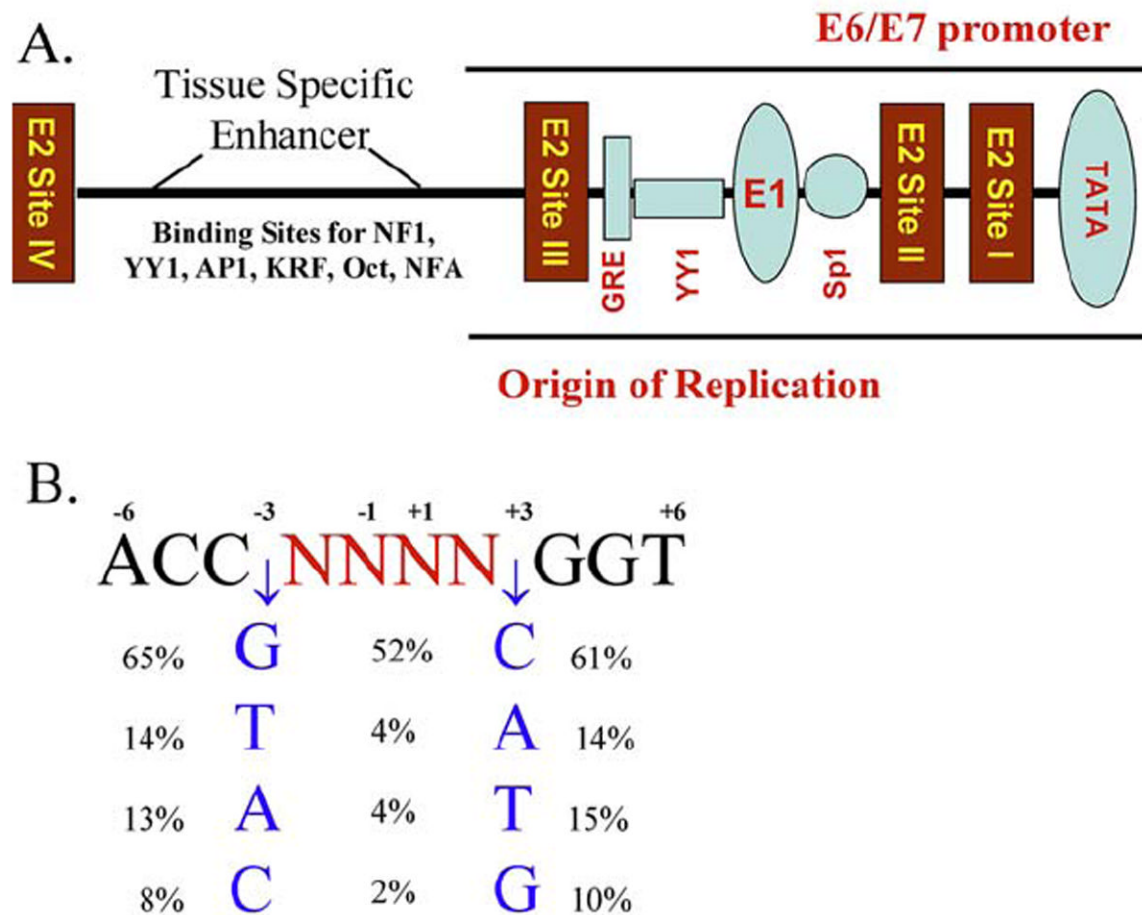


Figure 2.

A) Schematic of the papillomavirus Long Control Region (LCR) or Upstream Regulatory Region (URR) and E2 binding sites I-IV. B) The consensus E2 binding site derived from the 122 papillomavirus types were analyzed in this study. The frequency of the preferred base pairs at positions -3 and +3 (flanking percentages) and the occurrence of each pair at positions -3 and +3 is shown in the center (central percentages). DNA sequences within the upstream regulatory region (URR) of the papillomaviruses included in this study (Appendix 2) were analyzed for binding motifs containing the ACCN6GGT sequence. This template excludes binding sites in which there is a substitution in the highly conserved bases that participate in the direct interactions between the E2 protein and the DNA. The European Molecular Biology Open Software Suite package was used to locate motifs within a given sequence (117).

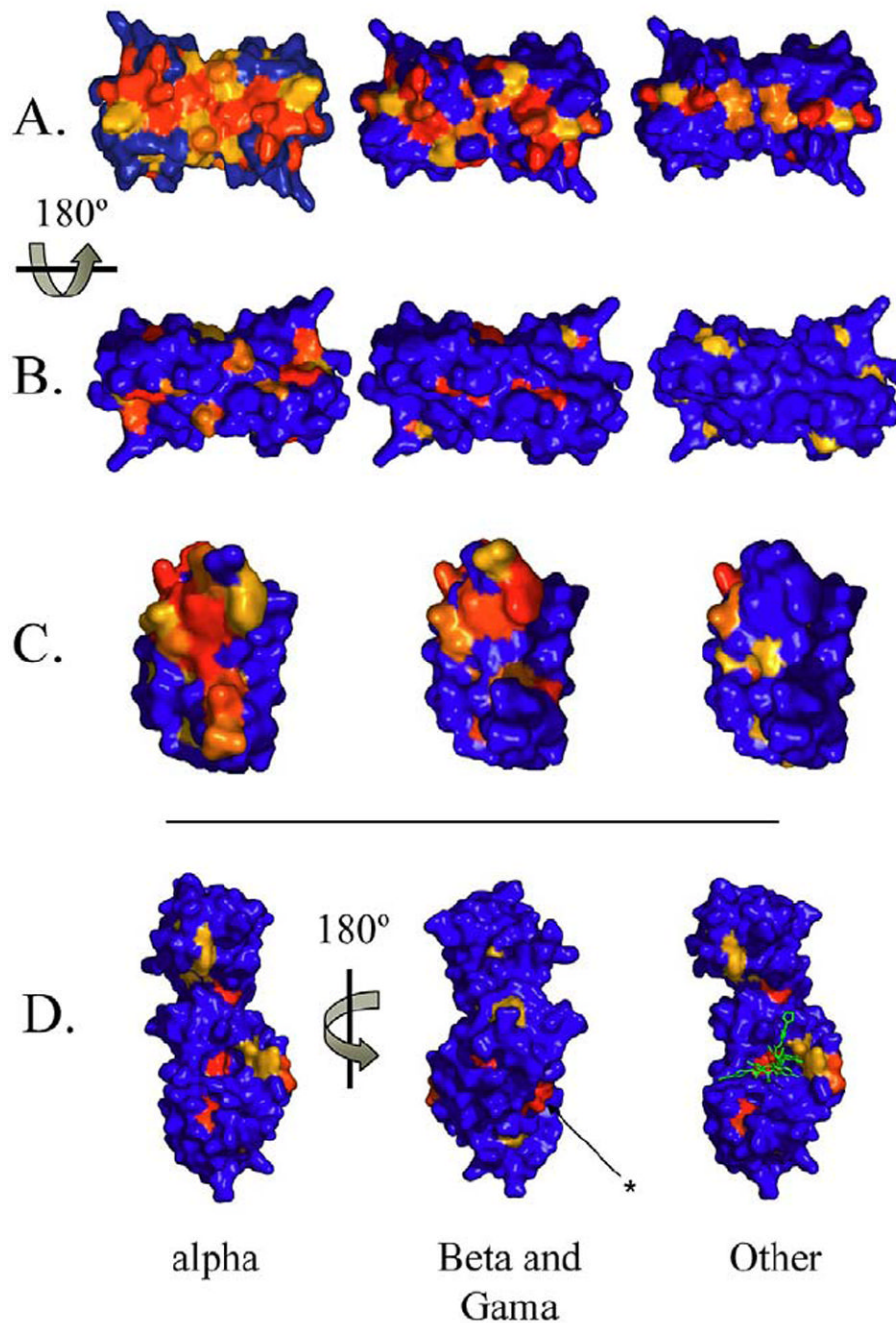


Figure 3. Structural surface representation of the amino acid sequence conservation using HPV16 E2 DNA binding domain as the template. Red denotes > 90% conservation, orange denotes a conservative substitution and blue denotes no conservation. A) The DNA binding surface of the DNA binding domain in surface fill view, B) 180 degree rotation showing the non-DNA contacted surface. The asterisk indicates a second conserved surface that is separate from the documented E2-E1 interface, and C) the dimer interface. The mucosal papillomaviruses are the most highly conserved in the DNA binding and dimerization surfaces. D) Row D shows the conservation of the E2 transactivation domain conservation using the structure of the domain of the HPV11 E2 protein in complex with a molecular inhibitor ((59); PDB ID:

1DTO). The inhibitor shown in green binds to pocket conserved throughout papillomavirus evolution. Note: The sequences used in this study were obtained from the sources summarized in Appendix 2. Grouping these sequences into (1) Other (animal E2 sequences), (2) beta genus and (3) and alpha genus was done as described in the text (1). Multiple sequence alignments were performed using the ClustalW program (118). The secondary structure assignments shown in Figure 5 were taken from structure of HPV16 E2/D (PDB ID: 1by9). Residue conservation calculations were performed using AMAS program (119). Briefly, AMAS is a program which performs a systematic characterization of the physical-chemical properties seen at each position in a multiple protein sequence alignment. A flexible set-based description of amino acid properties is used to define the conservation between any groups of amino acids.

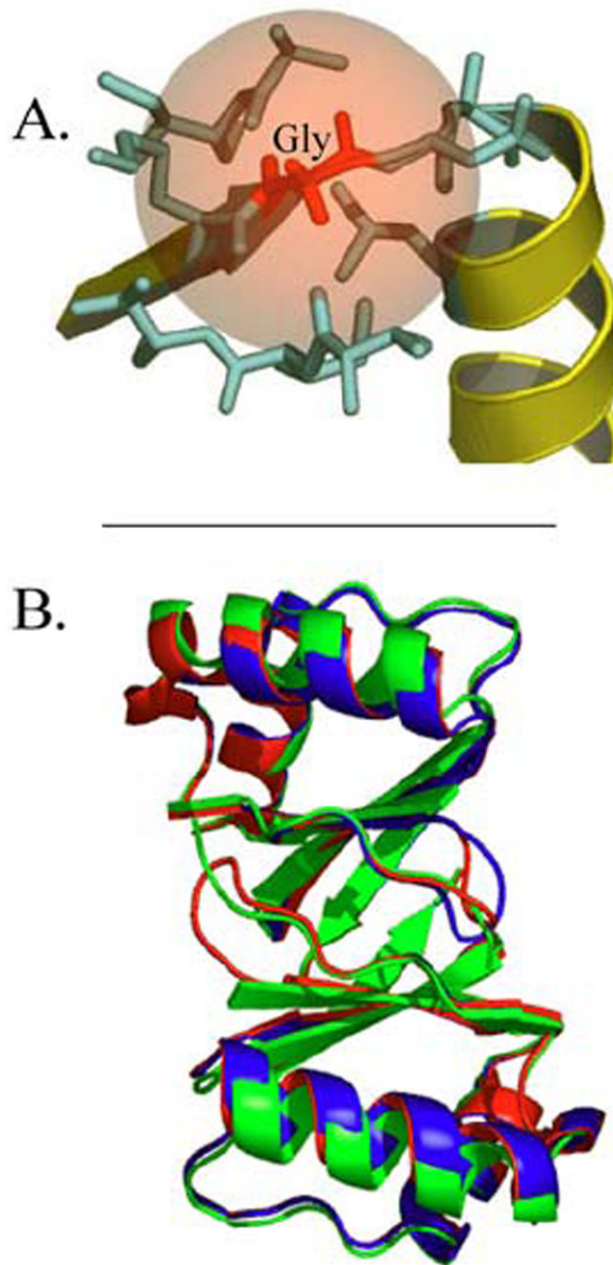


Figure 4.

A) Close-up of the structure of the HPV16 E2 DNA binding domain (PDB ID: 1BY9) highlighting the Gly residue in a loop preceding the DNA recognition helix that is absolutely conserved in all 122 papillomavirus types and 24 papillomavirus HPV16 variants that were analyzed. The picture shows the loop that contains Gly293 (red), the residues that are within 5 angstroms (Lys290, Asp292, Ala 293, Leu 296, Ala 329 and Ile 330) in magenta; B) Homology models that were generated after *in silico* mutagenesis of HPV 16 E2 Gly293 to Ala and Val, indicating the shift in the spatial position of the recognition helix. The wild type protein is colored green, the G293A mutant protein is colored red and the G293V mutant protein is colored blue. This view looks up at the recognition helices of the E2 protein dimer.

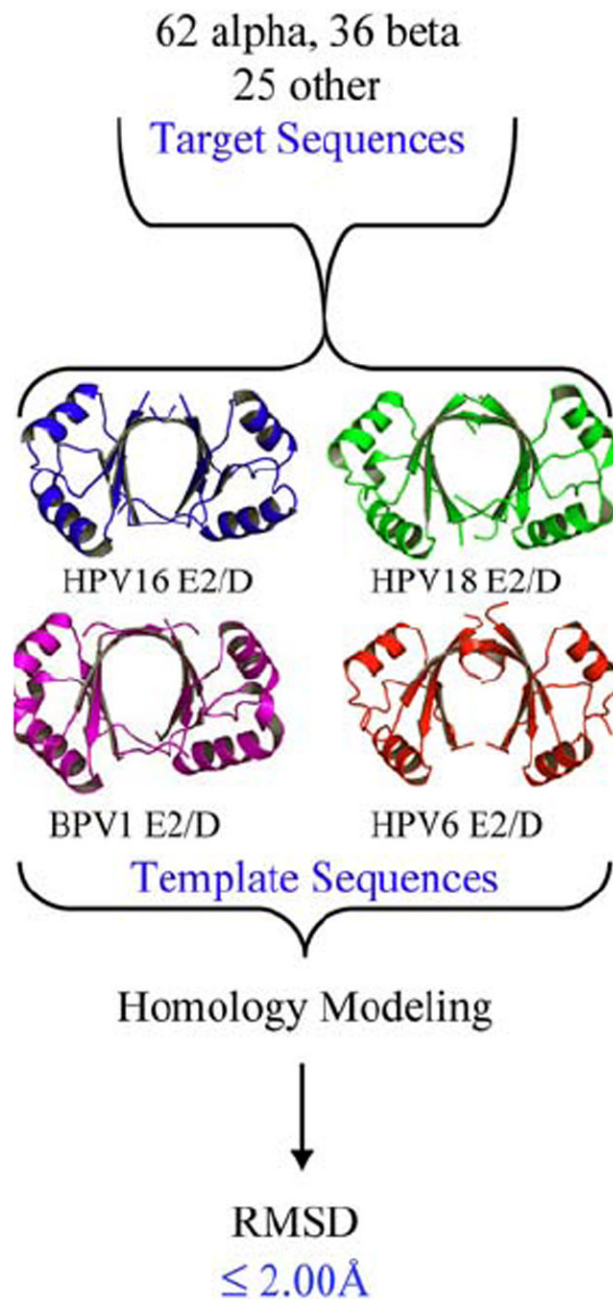


Figure 5.

Summary of the generation of homology models (54) by group from the indicated templates for 122 papillomavirus types and the RMSD values obtained from comparison of the models. Comparative protein structure modeling of E2 sequences were performed using MODELLER (120, 121). Briefly, MODELLER perform a comparative modeling of sequences by satisfaction of spatial restraints inherited from a protein (s) of know structure (s), also known as template (s). Three templates were used to generate the modeled structures, BPV1E2/D (PDB ID: 1jjh), HVP16 and 18E2/D (PDB IDs: 1by9 and 1f9f), and HPV6 E2/D (PDB ID: 1r8h). The sequence identities between target sequences and templates were in the range of 40 - 80%, which is considered a 'safe' range of sequence identity where accurate models can be obtained. Five models were constructed for each

sequence and the ones with best energy (according to MODELLER's energy function) were kept. In addition, models were inspected using PROSAA (122) and PROCHECK (123) to further analyze their quality. Structural superposition of models was calculated with STAMP using only main chain atom coordinates (124).

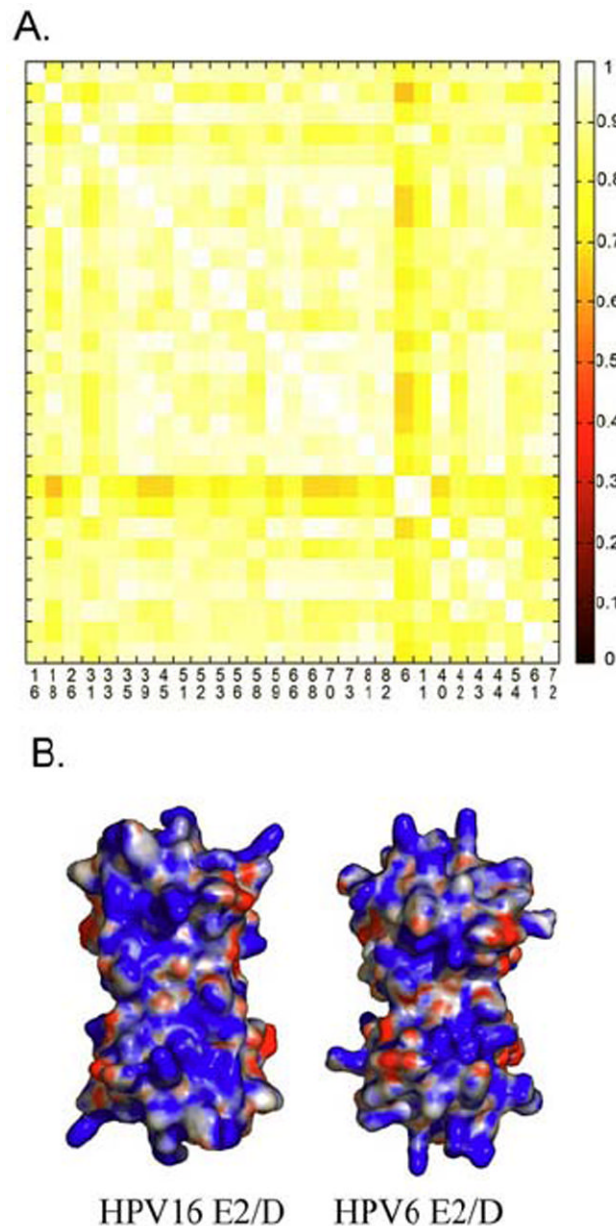


Figure 6.

A) A similarity index (125), with equivalent x and y axis, comparing the electrostatic potential of the DNA binding surface of E2 proteins from low and high HPVs (5). A value of one represents complete similarity while zero denotes no similarity; Protein interaction properties similarity analysis (PIPSA) (125) was used to compute the difference in electrostatic potential among structure models. PIPSA calculates the Hodgkin Similarity index (SI) that measures the similarity of two molecular potentials in sign, magnitude and spatial behavior. The SI index value ranges from -1 for anti-correlated index (i.e. opposite sign) to +1 for correlated potential. SI index were traditionally applied to measure electrostatic potential similarities between small molecules but it can be use also for protein using a grid approach. The electrostatic potentials were calculated using APBS program (126) and compared with PIPSA using the following parameters: regular cubic lattice with 65 angstrom dimensions, 1.5 angstrom spacing, two monovalent ionic species: -1 and +1 at

0.050 M, concentration, a dielectric constant of 78, and 298.15 K. Cubic lattice dimensions were manually inspected to ensure the complete immersion of proteins in the lattice. The cubic lattice was centered at the center of the models and only the DNA binding surface was used to compute the SI indexes. B) The surface potentials calculated for the HPV6 and HPV16 E2 proteins illustrating the high degree of similarity. These two surfaces only differ in a diffuse increase in the electropositive potential of the surface for the HPV16 E2 protein.

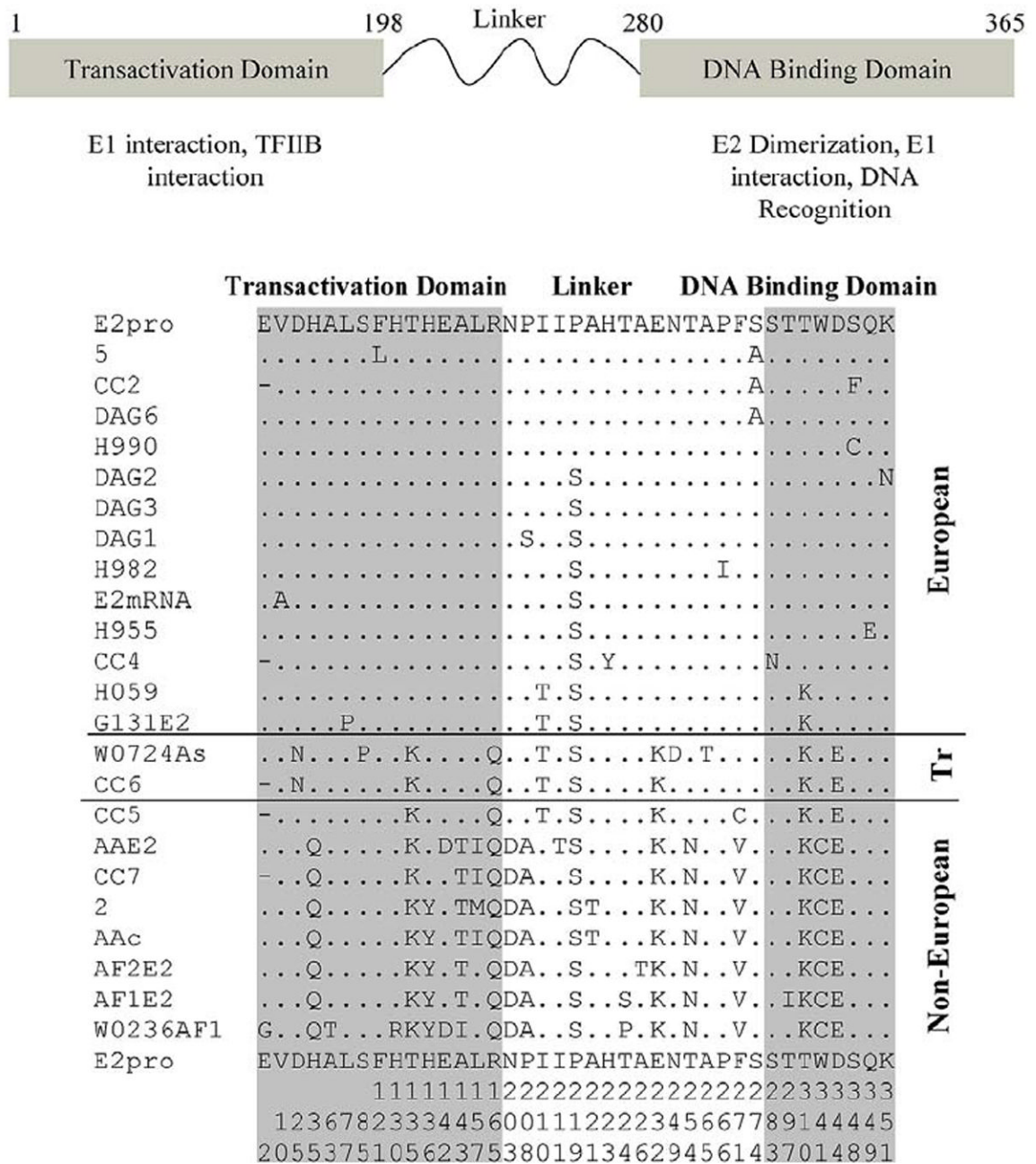


Figure 7.
 A) Annotation of the papillomavirus E2 gene including the N terminal transactivation domain and C terminal DNA binding domain that shows only the amino acid differences among the HPV16 variants. The overall the identity of HPV16 variant E2 genes is > 95%. The numbering corresponds to HPV16 E2 gene bank deposition. The numbering is not continuous and is represented along the bottom of the figure; B) The amino acid alignment of the HPV16 E2 prototype (E2pro) gene and 23 variants that depicts only the variable amino acid sequences. The amino acids that are uniquely conserved within the Non-European group are H35Q, T135K, H136Y, A143T, and R165Q in the transactivation domain and T310K, W341C and D344E in the DNA binding domain. *Tr* denotes a

transitional epidemiological classification 'European Asian'. Additional information about individual variants is in the Appendix 3. The numbers at the bottom of the figure represent the amino acid sequence number within the E2 gene (numbers are read from top to bottom).

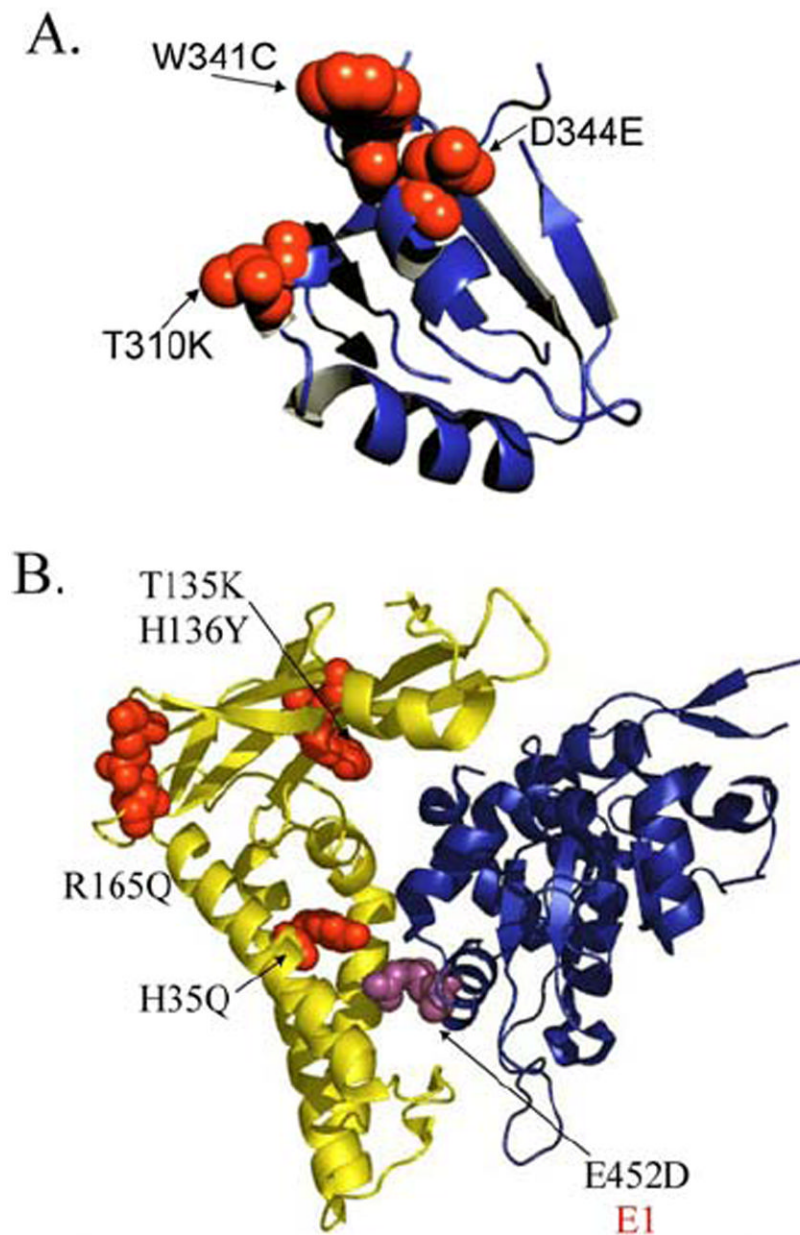


Figure 8.
 A) A 90° rotation of the E2/D protein structure (frame of reference is Figure 2C) showing in red the mutations present in the Non-European variants (Figure 6). These mutations are on the surface of the protein that interacts with the E1 replication protein (57); B) Ribbon diagram of the structure of the HPV18 E2 transactivation domain (yellow) in complex with the E1 helicase domain (blue) (33). The structures of the HPV18 and HPV16 E2 transactivation domains are highly similar with an RMSD of $\sim 1 \text{ \AA}$ (33). The mutations present in the proteins of the Non-European variants are shown in red (Figure 6). The E452D amino acid substitution present in Non-European HPV16 E1 protein variants is highlighted in magenta. No structural information is available for three other conserved amino acid variations within the E1 protein for the Non-European variants, Q78E, C168S and I326M, which are not present within this domain.

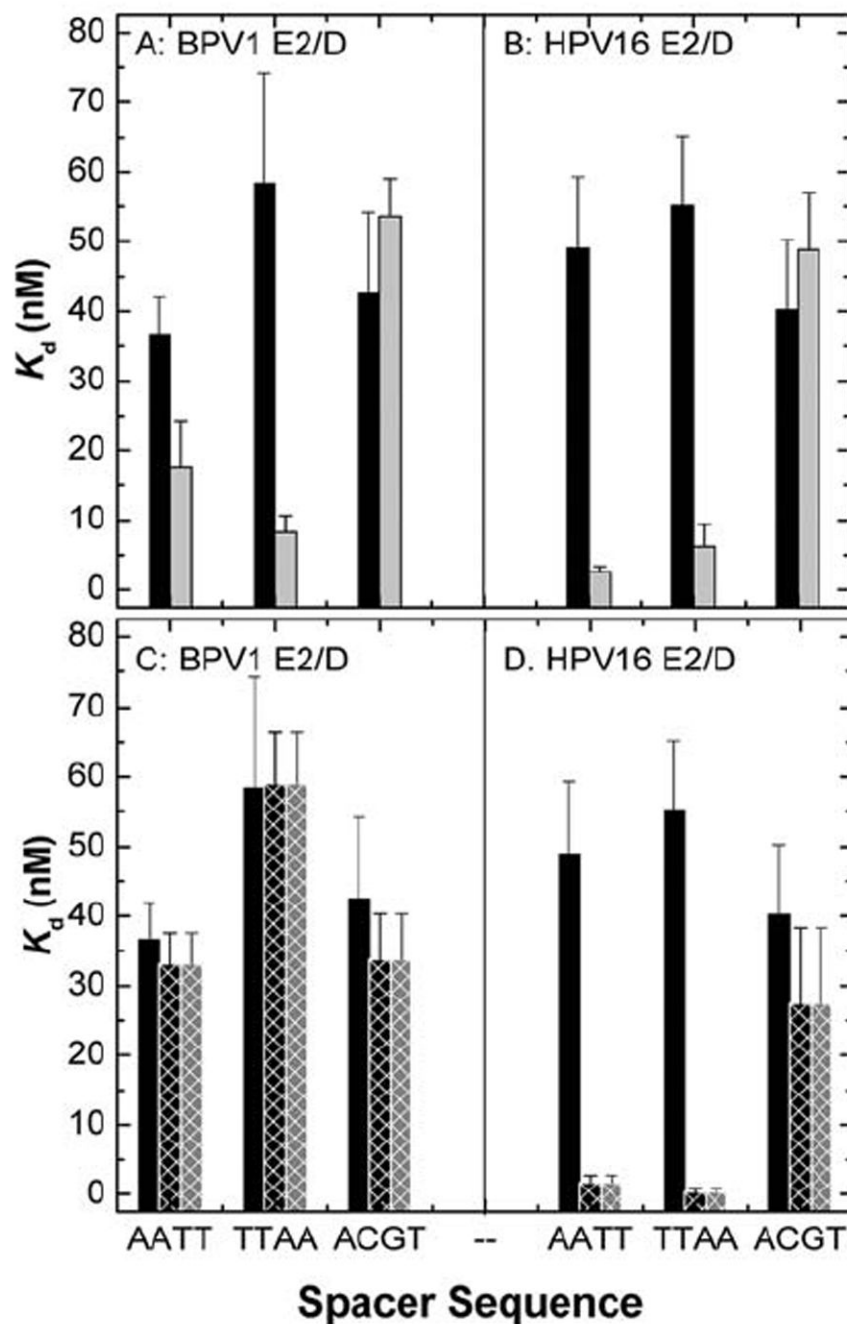


Figure 9.

Summary of the equilibrium binding constants (K_d , nM) determined for in BPV1-E2/D (A & C) HPV16-E2/D and (B & D) binding to DNA containing the AATT, TTAA and ACGT spacer sequences. Panels A and B summarize binding in buffer containing 150 (black bars) and 250 mM (grey bars). The ionic strength of these solutions is 0.15 and 0.25, respectively. Panels C and D summarize binding in the presence of Mg^{2+} : buffer containing 150 mM KCl (black bars), buffer containing 150 mM KCl and 10 mM $MgCl_2$ (black crosshatched bars; ionic strength = 0.19 M), buffer containing 110 mM KCl and 10 mM $MgCl_2$ (grey crosshatched bars; ionic strength = 0.15 M) (27). The data indicated the increased affinity

with AT rich sequences as KCL concentration increase, and in the presence of MgCl₂.
Figure reprinted from reference (27) with permission from Elsevier.