[Frontiers in Bioscience 13, 6006-6021, May 1, 2008]

# The papillomavirus E2 DNA binding domain

**Gonzalo de Prat-Gay[1], Kevin Gaston[2], Daniel O. Cicero[3]**

*[1] Instituto Leloir and IIBBA-CONICET, Av. Patricias Argentinas 435, (1405) Buenos Aires, Argentina, [2] Department of Biochemistry, School of Medical Sciences, University of Bristol, UK, [3] Department of Science and Chemical Technologies, University of Rome "Tor Vergata", Via della Ricerca Scientifica, 00133 Rome, Italy*

## TABLE OF CONTENTS



## 1. ABSTRACT

The DNA binding domain of the E2 master regulator from papillomaviruses is the primary effector for most the essential activities controlled by this protein. In this review we focus on the properties of the DNA binding domain of human papillomavirus strain 16 in solution, integrating structure, dynamics, folding, stability, conformational equilibria, and DNA binding mechanism. We discuss the relevance of these processes for the different biological activities, broadening the horizon for antiviral development. In addition, the particular fold of the DNA binding domain only shared with the Epstein-Barr nuclear antigen EBNA1, suggests a link between this unique architecture and the function of viral origin binding proteins of this kind. Finally, the E2 DNA binding domain proved to be an excellent model for addressing fundamental problems of DNA recognition mechanisms and folding of intertwined dimers.

## 2. INTRODUCTION

While mammalian cells have around 30,000 genes, viruses that cause diseases in the hosts that range from the mild or asymptomatic to the deadly, often contain only a handful of genes. Papillomaviruses have DNA genomes of eight thousand base pairs that encode only around 8 protein products depending on the viral strain. As is the case with many other DNA viruses, papillomaviruses must accomplish several key processes in order to complete their life cycle including; genome replication, viral gene transcription and episomal segregation of the viral genome with the host chromosomes. The latter is essential for latent and persistent viral infection. Some human papillomaviruses (HPV) are involved in cancer progression and this is thought to require long-term infection since the development of tumors may take several years. The papillomavirus E2 protein is directly involved in all three of these key processes. E2 was originally described as a

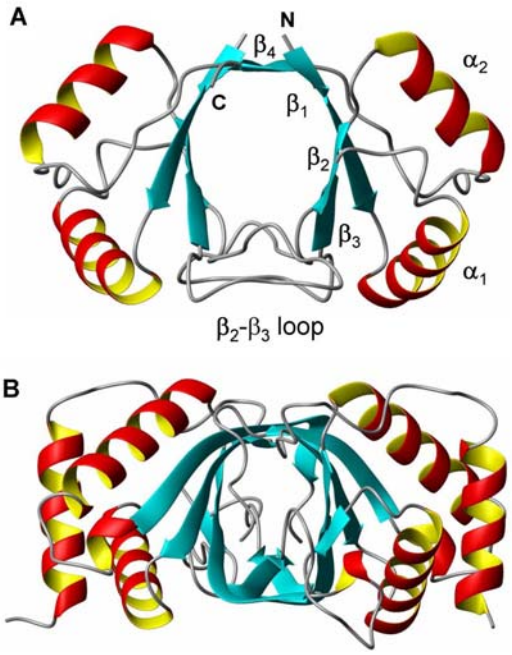


**Figure 1.** (A) Three-dimensional structure of the E2C protein belonging to HPV-16 (pdb entry 1zzf, (49)). The secondary structural elements are indicated. (B) Three-dimensional structure of the EBNA1 origin binding protein from the Epstein-Barr virus (pdb entry 1vhi, (4)).

transcriptional regulator and origin binding protein (1). The protein consists of ca 400 amino acids and is divided in three domains: an N-terminal transcription activation/replication domain, an intermediate flexible "hinge" domain and a C-terminal dimerization and DNA binding domain (E2C), of ca 80 amino acids, depending on the strain (2). The first E2C structure to be solved was that of the bovine papillomavirus (BPV) type 1 protein complexed with DNA. This revealed a novel fold; the dimeric beta-barrel domain, consisting of two half-barrels each composed of 4 beta-strands and two alpha-helices. One alpha-helix from each half-barrel acts as a DNA recognition site (3). To date, the only other protein known to contain this fold is the DNA binding domain of the EBNA1 origin binding protein from the Epstein-Barr virus (4). These viruses that belong to different families, share neither amino acid nor DNA site sequence homology, but both bind DNA at the viral replication origin (Figure 1).

This review is focused on the particular properties and structure-functional aspects of the E2C, in particular that of HPV E2C, its behavior in solution, and DNA binding mechanism. There are very comprehensive reviews on the biology of papillomaviruses (5), the biology of E2 (6,7) and the structures of E2C domains (8).

## 3. The DNA binding domain of the papillomavirus E2 master regulator

### 3.1. The E2 protein

E2 is required for efficient HPV replication and is generally acknowledged to play an important role in viral gene expression. However, it is important to note that despite years of intensive study, many of the biological functions of E2 are still poorly understood or indeed remain elusive. This is due in large part to difficulties associated with studying the HPV life cycle in vitro. Viral replication requires differentiation of the host epithelial cell and HPV DNA must therefore persist over an extensive period with many generations of host cell division (reviewed by Doorbar, (9)). During this period, HPV must evade the host immune response and in fact HPV infections can persist for several months or even years before viral clearance. The small size of the papillomavirus genome means that the virus is absolutely dependent on host cell proteins in order to replicate and complete its life cycle. Viral proteins recruit a wide variety of host proteins and subvert many host cell pathways in order to achieve these ends. The expression of viral proteins must be controlled and varied as the host cell differentiates, culminating in the large-scale production of infectious viral particles. The E2 protein plays crucial roles in all of these processes. Furthermore, although several host cell transcription factors are used by the virus to control viral gene expression, the E2 protein also regulates viral transcription. Thus, E2 is a multifunctional protein and that plays diverse roles in the HPV life cycle.

Mutations in the viral genome that prevent the expression of the E2 protein block viral replication in cells and E2 is essential for efficient HPV DNA replication in *in vitro* replication assays. E2 binds to the HPV origin of replication and acts to recruit the viral E1 helicase ((10) and references therein). Subsequent steps in viral replication depend on the host cell DNA replication machinery and E2 has been shown to interact with several host cell proteins involved in DNA replication including Topoisomerase I (11). E2 also binds to proteins involved in DNA repair including the tumor suppressor protein p53 and TopBP1 and E2 may play a role in directing the repair of damaged viral genome (12,13). Recent work has shown that E2 is also important for long-term persistence of viral DNA in infected cells (14,15). E2 proteins bind simultaneously to the cellular proteins Brd4 and ChlR1 that attach to mitotic chromosomes and to the viral origin of replication (14,15). The E2 protein is thus thought to act as a tether that links HPV genomes to the host chromosomes, thereby ensuring equal segregation of viral DNA during host cell division. The host proteins that enable E2 to perform this role appear to be specific to different viral strains and/or functionally redundant.

E2 functions as a transcription factor to regulate the expression of viral genes and very likely to regulate the expression of some host genes. The HPV genome contains four binding sites for E2 within a region of around 1kb known as the upstream regulatory region (URR) or long control region (LCR) (Figure 2). The LCR controls transcription of the viral genes and contains the origin of replication. When the LCR is placed upstream of a reporter gene E2 represses transcription and the mechanisms whereby E2 can repress transcription in this context have been described in elegant detail (16-19). The binding of E2 to two sites proximal to the transcription start site blocks the binding of cellular transcription factors to adjacent

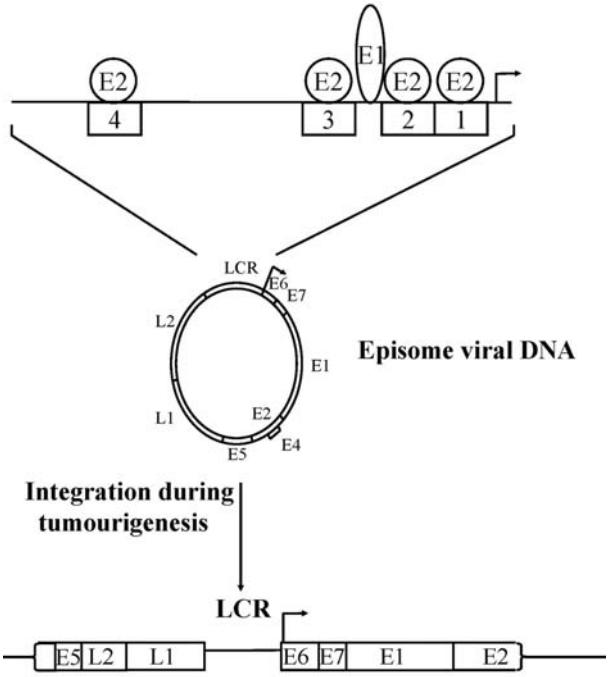


**Figure 2.** The HPV episomal genome and its chromosomal integration during tumourigenesis. The HPV genome usually exists as a circular double-stranded DNA episome. The Early (E) and Late (L) genes are indicated, as is the Long Control Region (LCR). The bent arrow represents the early transcription start point. The top line shows an expanded view of the LCR with the E2 binding sites as boxes and the locations of the E2 and E1 binding sites. The bottom line shows an example of a viral integration event found in tumour cells transformed by high-risk HPV types. The E2 gene is disrupted.

DNA sequences and thereby represses transcription initiation. More recent work has shown that the Brd4 protein also plays a role in transcriptional repression by E2 as well as viral genome segregation (20). However, it has been claimed that E2 has no effect on viral transcription in the context of the intact viral genome (21). When E2 sites are placed upstream of a reporter gene and a non-cognate promoter sequence E2 is generally reported to activate transcription, hence, the description of the N-terminal domain of E2 as a transcription activation domain. However, the ability to activate transcription is not essential for HPV replication (22). Interestingly, the full-length BPV 1 E2 protein binds co-operatively to DNA fragments with multiple E2 binding sites and can mediate the formation of DNA loops (23). The E2C domain does not form these DNA loops suggesting that the N-terminal transcription/replication domains of distantly bound E2 proteins interact to form DNA loops. Splice variants of E2 protein that lack the N-terminal domain have been characterized in BPV and HPV. These E2 splice variants can repress viral transcription or at least block the functions of the full-length E2 proteins (24,25). However, the role that these splice variants play in the HPV life cycle is not well understood.

The HPV E6 and E7 genes encode oncoproteins that interact with the cellular tumor suppressor proteins p53 and pRB, respectively (reviewed by zur Hausen, (26)). These interactions modulate the activities of these growth control proteins and allow completion of the HPV life cycle. During the process of viral tumorigenesis, HPV DNA from a sub-set of viral types including HPV 16 and 18 and collectively known as the high-risk HPV types, often become integrated into the host genome (27). Integration most often results in disruption of the E2 gene but leaves the E6 and E7 genes intact (Figure 2). The outcome of integration is therefore thought to be the derepression of E6 and E7 by removal of E2. This leads to uncontrolled cell proliferation (28), cell transformation and in due course, tumorigenesis. DNA from high-risk HPV types is found in virtually all human cervical carcinomas (29). In contrast, while low risk HPV types such as HPV 6 and 11 may well integrate into their host cell genomes at rates equivalent to those of high-risk HPV types, these viruses do not bring about tumorigenesis. Re-introduction of E2 can have dramatic effects on the proliferation and survival of HPV-transformed cells. BPV and HPV E2 proteins have been shown to repress transcription of the E6 and E7 oncogenes in HPV-transformed cells leading to the induction of cell senescence and apoptosis. E2 can also influence cell proliferation via direct interactions with p53 (15) and interactions with the E6 and E7 proteins (30,31). However, the roles that these interactions play in the viral life cycle have yet to be elucidated.

Finally, it is worth noting that E2 may also be important in viral clearance. Host immune responses to E2 are seen at the time of viral clearance (32). Furthermore, vaccination of rabbits with a recombinant adenovirus

expressing the cotton tail rabbit papillomavirus E2 protein has been shown to reduce the number and size of papillomas (33). This suggests that E2 proteins might be useful in vaccines that target HPV infections. More broadly, E2 also represents a promising target for drug development. Drugs that prevent E2 from interacting with E1 have the potential to block HPV replication (34). Similarly, drugs that prevent the DNA binding activity of E2 would also have the potential to block viral replication.

### 3.2. The DNA binding activity of E2

The E2 proteins bind with high affinity to specific DNA sequences found in the viral LCR. In early studies immunoprecipitation of protein-DNA complexes and DNase I foot printing were used to identify E2 binding sites within the BPV and HPV genomes (35,36). Later, electrophoretic mobility shift assays (EMSA) and DNAse I foot printing showed that the BPV E2 protein binds to many sites within the BPV genome (37). The consensus E2 binding site is an inverted repeat tract with the sequence 5' ACCG NNNN CGGT 3', where the four Ns represent a 4 base pair spacer of conserved length that varies in sequence (37). The HPV E2 protein binds to just 4 sites within the HPV genome and these have an extended version of this consensus sequence: 5 AACCG NNNN CGGTT 3' (38). Like the full-length E2 proteins, the E2C proteins recognize sequences that correspond to the consensus shown above with $K_D$s in the region of 1-10nM (39-41). The specificity of DNA binding thus appears to be unaffected by removal of the N-terminal domain and hinge region although the affinity for DNA is reduced (38,42,43). The nature of the bases in the spacer region has a profound effect on the binding of some E2 proteins (44,45) and this will be discussed in detail below.

Whether E2 represses or activates viral transcription is thought to depend on the affinity of E2 for each of its four sites within the LCR. Binding studies have shown that in HPV 16 the E2 protein binds tightly to the most distal promoter E2 site and less tightly to the promoter proximal E2 sites (38,40). This work and the earlier work described above, resulted in models in which at low protein expression levels E2 binds to the promoter distal site and activates viral transcription whereas at higher expression levels E2 binds to the promoter proximal sites to repress viral transcription. However, the hierarchy of binding site affinity is different between HPV 16 and HPV 11 and HPV 6 suggesting that different virus types may not share this regulatory mechanism (38,40,41). Furthermore, due to the cooperative binding of E2, the presence of diverse cellular transcription factors and the precise conditions within the cell, the occupancy of the E2 sites seen *in vivo* may differ from that predicted from the hierarchy of affinities measured *in vitro*.

### 3.3. Structures of E2C domains

Fifteen years have passed since the earliest work describing the structure of an E2C-DNA complex belonging to BPV-1 (3). The structure presented a novel fold, the dimeric β-barrel (Figure 1A), which has been extensively discussed and reviewed (8 and references therein). Remarkably, before the structure of an E2C from any other papillomaviruses was solved, a second example of the dimeric β-barrel was found in the DNA-binding domain of the Epstein-Barr virus origin-binding protein, EBNA1 (4). As Figure 1 shows, the two proteins present a high degree of structural similarity. Unexpectedly, the crystal structure of the EBNA1-DNA complex showed significant differences in the way the two proteins interact with the DNA (45). In fact, according to the crystal structure, the homologues of the E2C recognition helices are not used by EBNA1 to establish contacts with the DNA. Instead, sequence-specific contacts are made by an extended chain inserted into the minor groove and a helix reaching into the major groove (45).

Since then, both x-ray crystallography and NMR have contributed to unraveling the structural features of E2C and details of the DNA interaction mechanism. To date, structures of E2C proteins belonging to BPV-1 (46,47), high risk strains HPV-16 (48,49), HPV-31 (50,51), HPV-18 (52) and the low risk HPV-6 (41) have been determined. In addition to the already cited BPV-1-DNA complex (3), crystal structures of HPV-18 (52) and HPV-6 (53) complexes are also available. Also, an NMR study was conducted on the interaction of HPV-16 with one of the four DNA binding sites from the HPV 16 genome (54), although no high-resolution structure is yet available in this case.

A common feature of all structures determined so far is that the core of the barrel is intricate, suggesting that any rearrangement of the subunits upon interaction with DNA must be subject to a large energetic penalty. A number of hydrophobic residues occupy almost completely the inner space of the barrel, including a couple of tryptophan residues that are stacked in parallel. In the case of HPV-16 and HPV-31 a molecule of water or sulfate, respectively, was observed in a crystal structure, making a bridge between the two histidine residues (48,51). HPV-6 shows a longer C-terminus after ß4, with the inclusion of an extra hydrophobic residue, Leu365, participating in the formation of the core (see Figure 1A for the structural elements).

A region of the protein that shows variable behavior is the ß2-ß3 loop. It was not observed in the crystal structures of BPV-1, HPV-16 and HPV-18 E2C free proteins, suggesting that this region is flexible, whereas it was observed for HPV-31 and HPV-6. Evidences from NMR studies indicate that the loop is indeed highly flexible in solution for BPV-1, HPV-31 and HPV-16. For HPV-6, two proline residues and a charge interaction from Lys323 in the loop to residues Asp311 and His336 in the other subunit confer a particular rigidity to the loop, and its presence in the crystal structure reflects an ordered conformation that is probably maintained in solution.

One possible criterion to classify the different structures available so far is to superimpose one of the two monomers and observe the relative position of the other monomer (8). The five known crystal structures of E2C proteins were superimposed in this way and are shown in Figure 3A. By inspection of the non-superimposed

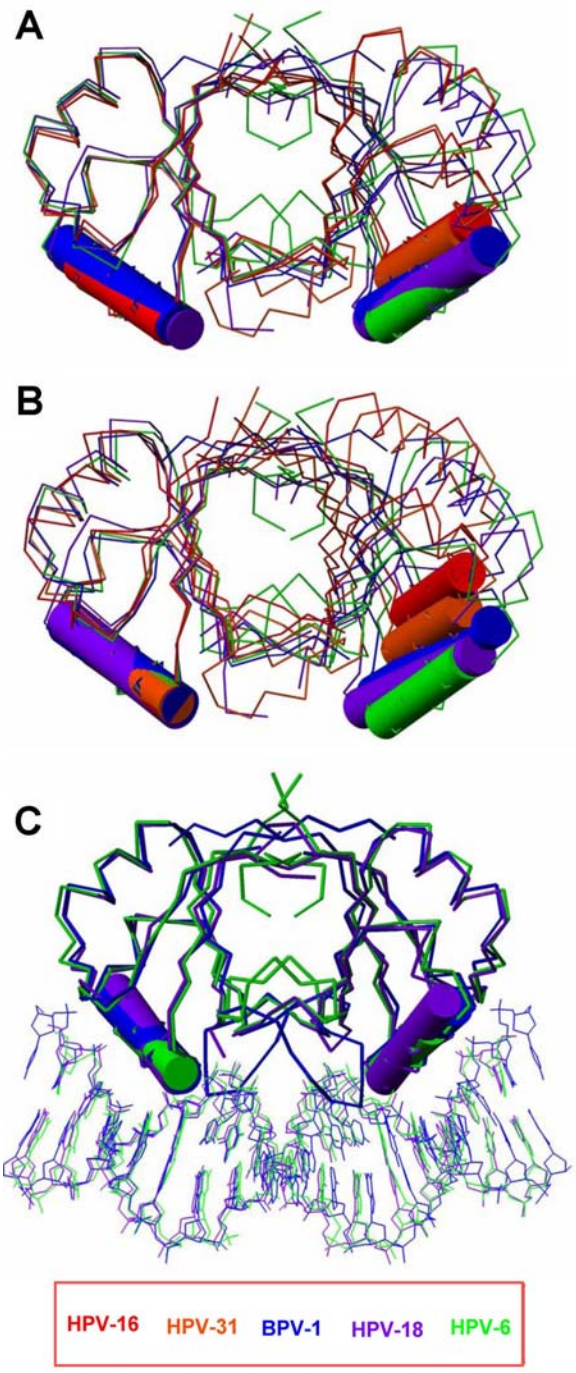


**Figure 3.** Superposition of E2C protein structures belonging to different viral strains: HPV-16 (red), HPV-31 (orange), BPV-1 (blue), HPV-18 (magenta) and HPV-6 (green). (A) Superposition of free E2C proteins, obtained by superimposing the left monomer. (B) Superposition of the left recognition helix of free E2C proteins. (C) Superposition of the left recognition helix of E2C-DNA complexes.

subunits, the E2 proteins can be divided into two classes: HPV-16 and HPV-31 belong to one group and HPV-18, BPV-1 and HPV-6 belong to another. There is a difference of about 4 Å in the relative position of the recognition helix of the non-superimposed subunit of HPV-16 E2 and that of BPV-1. The reason behind this differentiation is a shift in the β4-β4 hydrogen bond register, which creates the observed difference in quaternary structure (50). It is important to note that the E2Cs from high-risk strains HPV-16 and HPV-18 are classified as members of two distinct structural groups, suggesting that there is no correlation between the quaternary structure of E2C and the risk classification of HPV strains.

An alternative way of classifying the available structures of E2C proteins is to look at the relative orientation of the two recognition helices, rather than the two monomers. This is a sensitive way to detect changes in the relative position of the two helices, and is performed by superimposing a single recognition helix and looking for changes in the position of the other helix (48,53). This is particularly relevant because this helix contains the only residues that make contacts with the DNA bases, and which are conserved in all known E2 proteins. According to this alignment, a much larger spread in helix position is observed for the five known free E2C proteins (Figure 3B), with a continuum disposition of the non-aligned recognition helix ranging form the most internal helix of HPV-16 until the most external HPV-18. Table 1 summarizes the differences among the structures, all referring to HPV-18 E2C. The monomers show very similar conformations, as the RMSD for backbone atoms lies between 0.8 and 1.2 Å. When comparing the overall dimer and monomer RMSDs, the division into two families becomes clear, as HPV-16 and HPV-31 E2C proteins show a significantly higher value for the entire dimer as compared to the single monomer. When looking at the position of the non-aligned recognition helix, while superimposing the other, it is clear that we have significantly different situations, with RMSDs for Cα atoms ranging form 4.7 to 11.0 A, and with interhelical angles from 8.9 to 13.9°, all referred to the non-aligned helix of the HPV-18 E2 protein. The principal conclusion from this alignment is that the proteins appear to display the two recognition helices in ways that are specific for each viral strain.

Together with crystallography, NMR has contributed to the structure determination of E2C domains. Although the description of a structure like that of a dimeric β-barrel presents several technical difficulties for the application of NMR spectroscopy, the first evidence of a shift in the hydrogen bond register in the ß4-ß4 sheet and the consequent variation of the quaternary structure came from the solution structure of HPV-31 (50). In addition to the structure description, this work also showed that both the recognition helix and the ß2-ß3 loop presented a dynamical character in solution. Even though the flexibility of the latter was later corroborated by the absence of electron density in the crystal structures of HPV-18 and HPV-16, the observation of a dynamical behavior for the recognition helix remains an exclusive contribution of NMR, as the helix shows a very defined conformation in all crystal structures.

The E2C protein of BPV-1 was also studied by NMR (47). The difference with the crystal structure of the same protein lies in the inclusion of a stretch of 16 additional N-terminal residues. This extended version of the E2C protein is significantly more stable than the minimal domain comprising the C-terminal 85 amino acid residues. The extra N-terminal 16 residues were found to form a flap that covers a cavity at the dimer interface and may play a role in DNA binding.

In recent years, the potentiality of NMR to obtain more accurate structures was significantly improved by the introduction of residual dipolar couplings (RDCs) (55), in which informational content from NOEs is not local but rather can be considered long-range structural constraints. They were first used in the E2C structural field for the calculation of the HPV-16 solution structure (49). The introduction of these additional constraints improved the agreement between solution and crystal structures, with an observed RMSD of 1.2 Å for the superposition of the entire dimer. This number is substantially lower than those obtained for the same comparison of HPV-31 (1.8 Å) and BPV-1 (2.2 Å).

### 3.4. The structure of the DNA binding sites

The viral genome contains a number of DNA binding sites (E2-BS) for E2, ranging from four in the human strains, up to seventeen in the bovine counterpart (Figure 2) (8). All of the E2-BS have two features in common: an inverted repeat consensus sequence of the form 5' ACCG NNNN CGGT 3', and the presence of a central "spacer" region of conserved length but variable nucleotide composition. The discrimination between binding sites is specific for the different virus strains; spacers rich in A/T are preferred by all the human strains (41,44,52,56). On the other hand, BPV-1 E2 protein displays no apparent spacer sequence preference (44). This difference is reflected in the corresponding viral genomes: the HPV genomes have E2-BS with A/T-rich spacers, whereas the genomes of non-primate animal viruses (including BPV-1) have no such predominance of A/T-rich spacers (8,48).

The affinity of the protein for a given DNA sequence can be modulated by direct contacts between amino acids and DNA base pairs (direct readout), and/or by the recognition of the intrinsic three-dimensional shape or flexibility of the DNA binding sequence (indirect readout). In order to establish the importance of the latter, it is necessary to know the starting conformational state(s) that different DNA targets adopt in solution in the absence the protein. In this respect, X-ray crystallography was used to study the structures of dodecamers representing a bovine E2-BS (57) and a human E2-BS (58) revealing significant differences between the two. In the first case, DNA bearing an ACGT central spacer shows an α-helix continuously bent toward the major groove, which for the central spacer differs markedly with the situation in the protein-bound state (3) that shows bending towards the minor groove. On the contrary, the E2-BS with an AATT spacer was found to be already bent towards the central minor groove by 9° (58), constituting a favorable prebending of the E2-BS in the direction of the necessary deformation to bind the protein. Additional intrinsic curvature in the flanking major grooves gives rise to an overall helix axis deflection of ~10°. In contrast, the spacers ACGT and GTAC are straight (57). These features were recently reproduced by Monte Carlo simulations (59) and suggest a model in which the intrinsic DNA shape and/or flexibility of each spacer creates a distinct energy cost for converting the intrinsic DNA conformation to the protein-bound conformation, and thus modulates the E2 binding affinity.

Experimental data derived from high-resolution crystal structures provide detailed structural information only on static conformations of E2-BS. Other techniques that estimate the curvature of the DNA in solution, like gel electrophoretic phasing (60,61) and cyclization kinetics measurements (62) can provide measurements on averaged helical parameters. Using an electrophoretic phasing assay, it was possible to estimate that in solution an E2-BS with a central AATT spacer shows 18° of net curvature, whereas a sequence with the sequence ACGT shows only 1° (60). With respect to the crystal structure, this technique agrees in the direction of the curvature, which goes in the expected direction to form a high-affinity complex, but the estimated value is almost doubled. As the authors stressed, however, this technique cannot distinguish between static curvature and anisotropic flexibility, which can partially accounts for the difference with the quantitative estimation of the curvature between their results and the static X-ray structure.

Another attempt to correlate the intrinsic shape and mechanical properties of E2-BS with the observed affinity was carried out using a DNA cyclization method (63), coupled with a statistical mechanical theory (64). Using this approach, a number of structural parameters like roll, twist, bending flexibility and twisting flexibility were used to estimate the binding affinity of 16 sites containing different spacer sequences (62). For 15 of these sites, the variation in affinity was predicted within a factor of 3.

In principle, determination of the curvature by classical NMR methods is difficult (65). The DNA oligomer best characterized in solution by NMR is, however, very similar to that of a high affinity E2-BS: the so-called Dickerson dodecamer $(CGCGAATTCGCG)_2$ (66), which differs only by the two first and two last nucleotides with one of the HPV E2-BS (ACCGAATTCGGT) studied by X-ray crystallography (58). By an extensive use of residual dipolar couplings and $^{31}P$ chemical shift anisotropy, the final structure obtained represented a rather regular B-form helix with a global bending of ~10° (66), coincident to that observed for the E2-BS analogue (58).

### 3.5. The structure of E2C-DNA complexes

When comparing the available structures of E2C-DNA complexes using the criterion of superimposing just one of the two recognition helices, the observed spread in helix disposition is reduced with respect to the situation observed for the free proteins (Figure 3C and Table 1). Using the HPV-18 E2C complex as a reference, the non-

**Table 1.** Structural comparison of available crystal structures of E2C proteins, in the free and DNA bound states

| Free proteins (relative to HPV-18 E2C) | | | | |
|---|---|---|---|---|
| Viral Strain | RMSD | | Interhelical orientation[a] | |
| | Monomer[b] [Å] | Dimer[c] [Å] | Displacement[d] [Å] | Angle[e] [°] |
| HPV-16 | 1.1 | 2.0 | 11.0 | 14 |
| HPV-31 | 1.0 | 1.8 | 8.6 | 11 |
| HPV-6 | 0.8 | 1.1 | 5.1 | 12 |
| BPV-1 | 1.2 | 1.2 | 4.7 | 9 |
| DNA complexes (relative to HPV-18 E2C complex) | | | | |
| Viral Strain | RMSD | | Interhelical orientation | |
| | Monomer [Å] | Dimer [Å] | Displacement [Å] | Angle [°] |
| HPV-6(16mer) | 0.6 | 0.7 | 0.8 | 5 |
| HPV-6(18mer) | 0.7 | 0.7 | 1.5 | 2 |
| BPV-1 | 1.1 | 1.0 | 2.1 | 4 |

[a]Only one recognition helix was superimposed, whereas displacements and angles were calculated on the non-superimposed helix. Residues used for the alignment were: 296-307 for HPV-18, 293-304 for HPV-16 and HPV-6, 302-313 for HPV-31 and 335-346 for BPV-1. [b]Residues used for the alignment were 289-307,313-322,332-364 for HPV-18; 286-304,310-319,330-362 for HPV-16 and HPV-6; 295-313,319-328,339-371 for HPV-31 and 328-346,353-362,375-407 for BPV-1. [b,c]RMSD for backbone atoms is displayed. [d]RMSD for Ca positions of not-aligned helices. [e]Angle between not-aligned helices.

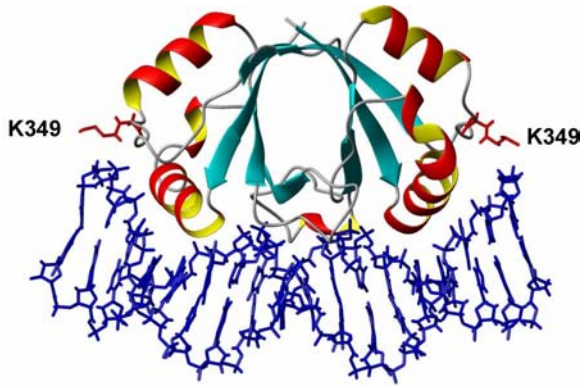


**Figure 4.** Model for HPV 16 E2C-DNA interaction based on the conformation showed by free E2C. The close proximity between K349 and the end of the 18bp DNA, which was shown to be partly responsible for the observed affinity (54) is also shown.

aligned helices are disposed in a similar way (Figure 3B). This suggests that, at least for proteins belonging to the HPV-18 family, binding to DNA makes the conformations more uniform. By analyzing the differences between free and bound states, it has been concluded that HPV-6 E2C experiences only slight changes upon binding a 16mer DNA fragment (with Cα displacements and interhelical angle changes of 2.5 Å and 5.6 °, respectively). This reduced adaptability of the protein can be related to the higher selectivity of this protein (41). BPV-1 and HPV-18 E2C proteins experience slightly larger changes upon DNA binding (46,52) (3.4 Å and 9.3° for HP18 E2C and 2.9 Å and 3.2° for BPV-1 E2C). Interestingly, HPV-6 bound to a 18mer DNA experiences a larger change (4.4 Å and 6.1°), leading to the highest displacement of the not-aligned recognition helix between free and bound states.

The global bending induced in the DNA is very different for these complexes, ranging from 51° for BPV-1 (3), 43° for HPV-18 (52) and 28° for HPV-6 (53). The latter being in remarkable agreement with the DNA bend angle estimated for the 6 E2C-DNA complex in solution (27°±3) using circular permutation assays (41). The minor groove width is also different, as a wider groove was observed for BPV-1 (4.0 Å) compared to HPV-18 (2.8 Å) and HPV-6 (2.7 Å). These data indicate that slight differences have a large effect on the bending and minor groove dimensions of the DNA. One of these differences is the fact that the β2-β3 loop becomes ordered in the BPV-1 complex, establishing electrostatic interactions with the phosphate groups in the minor groove (3). The HPV-6 E2C complexes also show an ordered β2-β3 loop, but its conformation is very different to that observed for the BPV-1 structure, and no interaction with the DNA was detected (53). HPV-18 E2C shows a disordered loop in the bound state. NMR evidence was also presented about the flexibility of the β2-β3 loop of the HPV-16 E2C in the DNA complex (54), however, mutagenesis analysis suggests at least temporary contacts between positively charged residues of this loop and the DNA (68) (see below).

To date, no high-resolution structure of the complex is available for HPV-16 E2C. However, a study was presented showing that HPV-16 displays differences in the interaction with DNA of varying lengths (54). A large enthalpy difference (ΔΔH) of 10.0 kcal $mol^{-1}$ was observed for the interaction of E2C and a 18mer or a 14mer site. This large difference may reflect a conformational change in the DNA in the two complexes, resembling the diversity found for HPV-6 E2C complexed to 16mer and 18mer DNAs (53). In addition, a charged residue located outside the recognition helix, K349, shows chemical shift perturbation only when a 18mer DNA is used for the interaction, but not with a 14mer duplex. The mutant K349A shows also a decreased affinity to the 18mer duplex by 1.4 kcal $mol^{-1}$ per symmetric interaction. This value is similar to the effect of shortening the duplex to 14 bases, which uncovers an additional contact between the E2C protein and the DNA outside the recognition helix and the ß2-ß3 loop, not previously observed. Figure 4 shows a model for the complex between HPV-16 E2C and an 18mer DNA, and the position of K349 very close to the C5' of the sugar in the 5' end of the nucleic acid. Clearly, a shorter DNA cannot exploit this contact and this will translate into lower affinity. K349 is replaced by an alanine in E2C-18, a mutation that weakens the DNA binding of E2C-16 domain, and by a proline in E2C-BPV1, but is conserved in HPV-6 and in the more frequently found high- and low-risk viral strains. This finding adds to the above-mentioned model of electrostatic aid that E2C uses to bend the DNA

molecule. As well as the electrostatic contribution to DNA bending, additional protein-base contacts may also facilitate DNA bending. EMSA suggests that the HPV 6 E2C protein makes additional base-specific contacts with the base pairs flanking the core recognition site when the central spacer region is AATT as opposed to CCGG (41).

The main question that still remains after all these studies is the adaptability of the E2C proteins of HPV-16 and HPV-31 strains during the interaction with their E2BS, for which there is no high resolution structure of the DNA complex. In a first analysis, it was suggested that a necessary event to change significantly the orientation of the recognition helices was a modification in the β4-β4 interaction. There are, however, indications against the occurrence of such an event during the interaction with DNA for HPV-16 E2C: chemical shift of residues belonging to the β4 strand do not change significantly between the free and bound proteins (54). Moreover, RDCs were used to calculate a model for the bound conformation, and although the pattern of hydrogen bond were not included during the calculations, two independent models starting either from free HPV-16 E2C or HPV-18 E2C converged to a similar structure, presenting the same features of the starting HPV-16 E2C protein (54).

Measurements of DNA conformation when bound to HPV-16 E2C gave contrasting results. On one hand, experimental data on the bending induced by HPV-6 and HPV-16 E2 proteins showed very similar results (41), indicating that the two proteins induce a similar distortion to the DNA. On the other hand, CD measurements showed that the conformation of the same DNA oligomer bound to HPV-16 E2C and BPV-1 displays a very different conformation (67). Moreover, the data were interpreted in terms of partial unwinding and base unstacking of the E2-BS when bound to HPV-16 E2 as distinct from the change in winding angle and base pair twist seen in the CD spectrum of a BPV-1 E2/E2-BS complex. A high-resolution structure of the complex between HPV-16 or HPV-31 E2C and DNA is still required to answer these remaining structural questions.

### 3.6. Folding mechanism of HPV16 E2C: a prototypic dimeric ß-barrel domain

The minimal DNA binding of E2 (E2C) was defined based on sequence alignments and this was confirmed by later structures (see above). The HPV16 E2C 80 amino acid (residues 286-365) domain was recombinantly expressed and initially characterized in solution, before a structure of this HPV type was available (40,69). The existing BPV1 E2C structure anticipated an intertwined domain (3), where the dimeric interface clearly appears to stabilize the whole fold. A two-state transition is observed at pH 5.6, where tertiary, secondary and quaternary structure change concomitantly to yield unfolded monomers. Chemical unfolding spectroscopic and sedimentation analysis coincide in unfolding constants ($K_u$) of 3.0-6.5 x$10^{-8}$ M, where $K_u$ is indistinguishable from the dissociation constant of the dimer ($K_D$), since both processes are coupled at equilibrium, with a ~10.0 kcal $mol^{-1}$ free energy of unfolding/dissociation. At pH 7.0, the $\Delta\Delta G_{U/D}$ is increased by 3.0 kcal $mol^{-1}$ and the use of a phosphate buffer further increases the stability by 6.2 kcal $mol^{-1}$. With $K_D$s of 0.5 nM and 1.4 pM, the latter in phosphate, in the *in vitro* binding analyses, E2C can be considered as a highly stable dimeric species. Efforts to uncouple dissociation of the dimer from unfolding have failed so far, indicating how cooperative secondary, tertiary and quaternary structure are.

Dissociation without unfolding may eventually be observed in experiments at extreme dilutions, but these are out of the reach of standard methods, including fluorescence, due to the sensitivity required to quantify monomeric and dimeric species in equilibrium. Besides, a close inspection of the structure suggests that any folded monomer most likely will not display a structure as observed in the intertwined native dimer. Non covalent interactions can be mildly and reversibly perturbed using high hydrostatic pressure, a physical as opposed to chemical denaturation method, that leaves no residues of chemical denaturants, and is non-invasive and often reversible (70). This approach allows the thermodynamic characterization of protein interactions in oligomers and protein folding. The pressure induced dissociation of HPV16 E2C is a fully reversible process with a $K_D$ of 60 nM at pH 5.5, 10-fold lower than that obtained from urea denaturation (71). The pressure denaturation yields an at least partly folded and rather compact monomer, not accessible to detailed structural studies, but not fully extended or unfolded. Thus, at pH 5.5 and in the absence of salts, the $K_D$ obtained from high pressure may well be a better estimation for an eventual "folded" monomer in solution. At pH 7.0, however, the dissociation process is incomplete (i.e., tighter $K_D$), indicating a similar pH stabilization to that observed in urea unfolding experiments (69), most likely related to a yet unidentified histidine residue.

Sodium chloride and phosphate show a substantial stabilization of the domain to urea denaturation, of 5.0 and 4.5 kcal $mol^{-1}$, respectively. Heparin and DNA showed a stabilization too large to be measured accurately, which suggests that the complex, once formed is extremely stable and may required other proteins or degradation to be disassembled (72).

The kinetic folding mechanism of HPV16 E2C reveals the formation of a monomeric intermediate species that precedes a concomitant dimerization-folding reaction leading to the final folded dimer. This species was shown to involve substantial secondary structure and the ability to bind an increased amount of the hydrophobic patch mapping dye ANS, with a negative heat capacity change (ΔCp) component in its transition state, and was proposed to be of non-native nature precisely because of the fact that half of the barrel cannot remained exposed to the solvent (73). In any case the intermediate converts rapidly to the dimeric natively folded form. Further analysis of the kinetic folding and unfolding pathway revealed that the monomeric intermediate is compact and cooperative indicating tertiary structure with a ΔG of unfolding of 3.5 kcal $mol^{-1}$, compared to 11.0 kcal $mol^{-1}$ for the overall

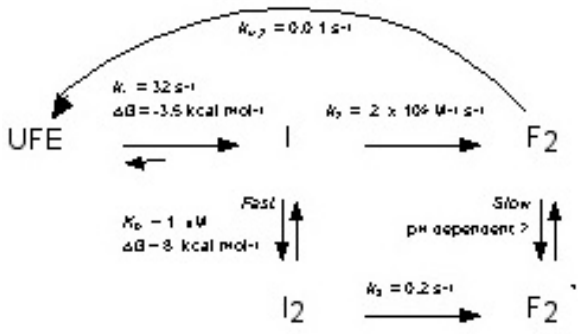


**Figure 5.** Model for the folding-unfolding mechanism of HPV16 E2C (Taken from *J Mol Biol* 351, 672-682 (2005)). Reproduced with permission from Journal of Molecular Biology, Elsevier Limited.

dimer unfolding transition, that represents 31% of the stability of the native dimer in identical buffer conditions (74). Reconstruction of its fluorescence spectrum at 100 ms by stopped flow experiments shows its tryptophan residues fairly exposed in the monomeric intermediate, but becoming buried as the native dimer interface is formed. Most of the burial of surface area takes place in the last rearrangement step leading to the consolidated native dimer (for details see discussion in (74)). It is tempting to suggest that the transient monomeric intermediate could be similar to a monomer, product of spontaneous dissociation of E2C in non-denaturaing conditions.

The unfolding reaction consists of a major phase with a half-life of ~1 min, with secondary and tertiary (and thus, quaternary) structures disappearing in parallel, indicative of no intermediated being populated. Using the $k_{off}/k_{on}$ approximation ($k_{uf2}/k_2$, see (74)) the equilibrium dissociation constant obtained is 0.5 nM, identical to that obtained by pressure denaturation (71). All the accumulated evidence strongly suggests that a spontaneous monomer will have to be obtained by a very sensitive dilution method, with the impossibility of structural characterization, mutagenesis or solvent modification. In addition, the bimolecular folding rate constant of HPV 16 E2C is 105 M-1 s-1, substantially lower than the theoretical rate expected from the collision of two spheres in solution, and much slower than that of trp or arc repressors (75,76). If the monomeric intermediate was native-like, the reaction would not present such a barrier and would be much faster. This species, however, has compact tertiary structure that must undergo a rearrangement or unfolding, coupled to the concomitant dimerization and native folding step. We believe that the slower folding association process is related to the complexity of formation of the interface barrel, in particular because ß-sheet structure is more dependent on long range interactions than □-helix formation, as in the case of the repressors. A similar picture emerges from the comparison of the folding of fragments of two paradigmatic proteins, CI2 and Barnase (77). Further studies with other HPV types will establish the possible existence of an isolatable monomeric species.

Refolding from urea or acid unfolded states yield identical pathways involving the monomeric intermediate and a subsequent dimerization-folding step, making the model more robust (74). This indicates equivalent pathways and that the unfolded state ensembles are similar in structure and/or energetically. NMR characterization of the urea unfolded state ensemble (UFSE) revealed two regions with clusters of residual structures, at the DNA binding helix and in the second ß-strand, although there appear to be no persistent long-range interactions (78). These regions most likely act as nuclei for the formation of early events in the folding pathway, i.e., the formation of the monomeric intermediate. Both equilibrium and kinetic folding depends on protein concentration, but this dependence is lost at 20 μM, which supports the two parallel routes, at high and low protein concentration, respectively (78). This is because of the formation of a weak dimer product either by a weak association ($K_D$) of the UFSE or the monomeric intermediate (see model in Figure 5).

### 3.7. Mechanism of DNA site recognition in solution

The first approach to *in vitro* DNA binding studies with purified components and spectroscopic methods in solution using a fluorophore was described for the HPV 11 E2C protein (39). The dissociation constant for the four sites in the HPV 11 genome was determined spectroscopically and an approximate hierarchy of affinity for these sites was established, where the closer to the replication origin, the higher the affinity. Modification of the central spacer sequence, for sequences such as CGCG led to much lower affinity (44).

The equilibrium DNA binding analysis of HPV 16 E2C, showed that at low ionic strength, the domain has a tendency to form aggregates in excess of protein over DNA, using a 18 bp duplex containing one of the E2 binding sites (site 35, based on the nucleotide position) (67). The domain has a 2:1 protein:DNA stoichiometry at pH 7.0 and a stoichiometry of 1:1 at pH 5.5, confirmed by different methods. A $K_D$ of 0.2 nM was determined for the high affinity 1:1 binding event, and a 10-fold lower equilibrium binding constant for the lower affinity binding event. The homologous BPV1 E2C domain binds to the same site with 350-fold lower affinity, and binds 7-fold less tightly even for a cognate bovine E2 DNA site. The HPV 16 E2C has 50-fold higher capacity than the bovine counterpart for discriminating cognate from non-cognate DNA, and is 180-fold better at discriminating specific from non-specific sequences. In addition, both domains impose a different conformation to the bound DNA, as indicated by circular dichroism, a highly sensitive technique (67), which could be related to the more pronounced bent or differences in base stacking.

The kinetic binding mechanism of HPV 16 E2C to the specific site 35 DNA duplex was investigated using stopped-flow techniques (79). Two parallel routes were observed. One, a multi-step pathway, initiated by the diffusion controlled formation of an encounter complex ($k_{on}$ $10^9$ $M^{-1}$ $s^{-1}$), which completely lacks sequence specificity and is weakly affected by a single charge mutation at the

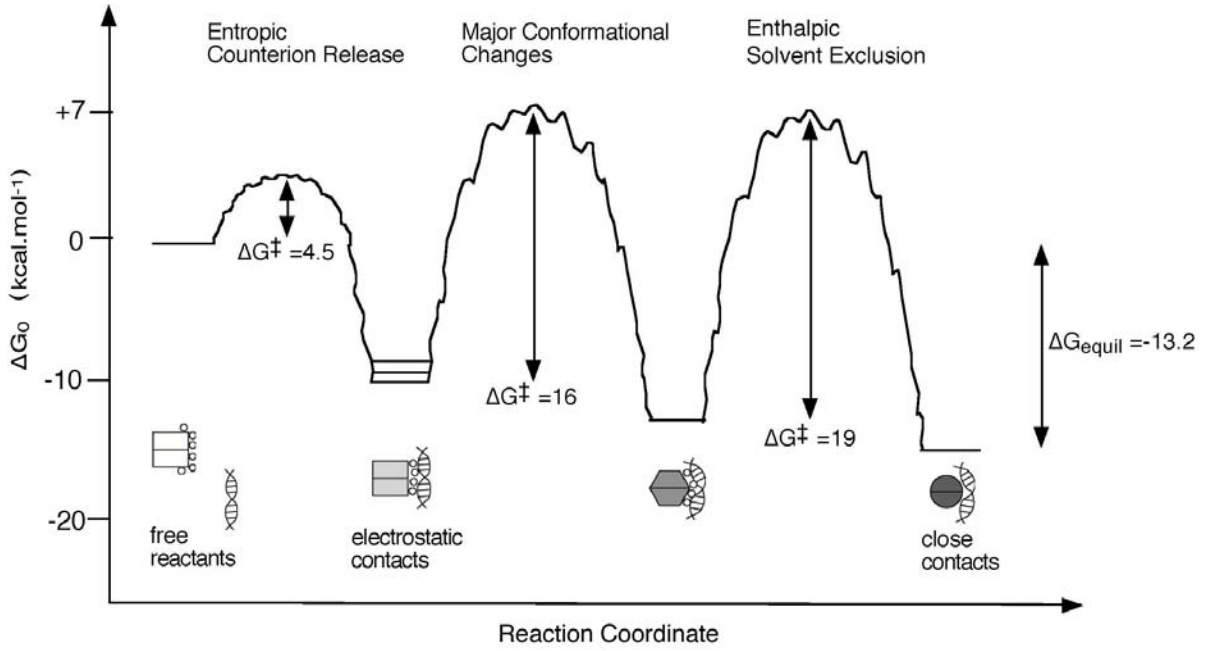


**Figure 6.** Energetic model for the multi-step DNA binding of HPV16 E2C (Modified from *J Mol Biol* 331, 89-99 (2003)). Reproduced with permission from Journal of Molecular Biology, Elsevier Limited.

DNA binding helix. This step is followed by a conformational first-order rearrangement ($k_2$ 8.0 $s^{-1}$), and finally, a slow solvent exclusion event ($k_3$ 0.04 $s^{-1}$) where the consolidated complex, including all the precise protein-DNA interactions, is formed. This last phase involves the major burial of surface area from the protein-DNA interface, leading to the consolidated direct readout of the DNA bases. The parallel pathway consists of a "fast-track" where the final complex is formed at the rate of collision, strongly suggesting a highly favorable conformational state of the interacting partners. Presumably this is means a protein conformation favorable for interaction and possibly a pre-bent DNA partner. The $k_{off}/k_{on}$ ratio of the fast route yields a $K_D$ of 0.15 nM, in astonishing agreement with that from equilibrium in identical conditions (0.2 nM). The ratio of constants of the multi-step pathway, the product of three forward and three reverse rate constants, yields a $K_D$ of 0.04 nM, still in excellent agreement with equilibrium data, considering the number of steps involved (79).

In the multi-step pathway, the encounter complex is stabilized by electrostatic interactions and even "accelerated" because of electrostatic steering, and no difference in rate was observed for any of the specific and non-specific DNA duplexes tested. Thus at this stage, "sliding" along the DNA could be occurring over this isopotential surface until the specific site is found, where conformational rearrangement and finally solvent exclusion takes place, strengthening the binding, and giving place of the precise direct readout of bases and side chains. The burial of surface, indicated by changes in heat capacity change (ΔCp) in the transition state, appears minimal in the initial stages and most of the burial takes place during the final solvent exclusion, something not unexpected but measured for the first time for a protein-DNA complex (79). Initial studies on the kinetic binding of HV6 E2C suggest minor rearrangements upon DNA binding, similar on rates and faster off-rates (Figure 6) (53). This is consistent with the proposed lack of conformational freedom in the ß2-ß3 loop of the HPV 6 E2C protein.

The DNA binding and dimerization domain of the EBNA1 protein, where the only connection to HPV E2C is their dimeric ß-barrel fold and a shared function as DNA replication origin binding proteins, displays very similar binding affinity, a multi-step DNA binding mechanism, and parallel routes originated by conformers, except that a fast one-step binding route was not found (80).

### 3.8. Energetic analysis of DNA recognition by E2C using rational site-directed mutagenesis

In order to dissect the energetics of binding of the HPV 16 E2C domain to its specific target DNA, site-directed mutations of all the DNA binding helix residues were produced in a conservative manner so as not to disturb the local or global folding or stability of the domain and to allow assignment of the energy changes to the binding to the replaced group of each protein side chain (68). The individual contribution of each contact from each symmetrical monomer is small, on average less than ~ 1.0 kcal $mol^{-1}$, independent of the physical chemical nature of the interaction. The sum of the individual contributions of each contacting side chain (11.1 ± 0.08 kcal $mol^{-1}$) differs by only 8% of the binding energy of the wild-type protein obtained from the $K_D$ (12.1 ± 0.08 kcal $mol^{-1}$), indicating that the interactions are additive and no "hot spots" are found at the interface as it often happens in protein-protein interactions (81). One mutant, T295A, shows a +0.8 kcal $mol^{-1}$ stabilization of the complex and was subtracted

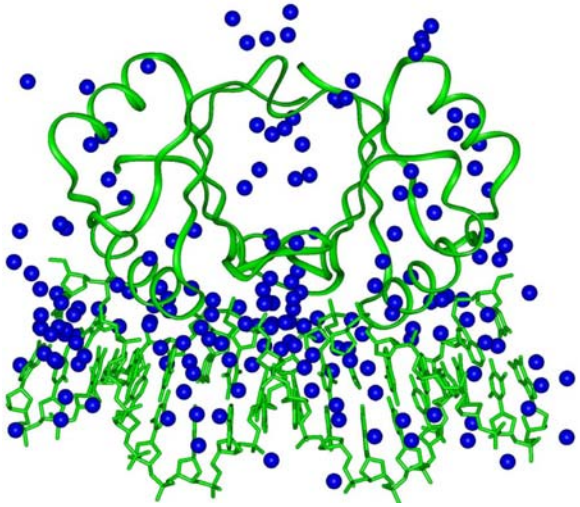

**Figure 7.** Structure of the HPV 18 E2C-DNA complex (pdb entry 1jj4, (52)) and the water molecules observed in the crystal structure (blue spheres). This figure shows the high number of water molecules in the interface between the protein and the nuclei acid.

instead of added. If we add the newly described interaction found out of the DNA binding helix in the second helix, K349, contributing with 1.0 kcal $mol^{-1}$ (54), the additivity is total (see Table I in (68)). Isothermal titration calorimetry showed that the effect of mutating the K349 residue to alanine was identical to that of shortening the DNA duplex from 18 to 14 base pairs, confirming not only the interaction with the base, but confirming that the base is out of the canonical E2 site ACCG NNNN CGGT, and most likely non-specific, i.e., a ionic interaction with a backbone phosphate (54). The complete additive as opposed to cooperative nature of the interactions was further confirmed by five double-mutant cycles (82) where the energy of the double mutations is the exact sum of the individual ones (68).

The site-directed mutagenesis approach also showed that the positively charged K325 and K327 residues in the HPV 16 E2C ß2-ß3 loop (Figure 1a) participate in coulombic interactions with the DNA, since the decrease in binding energy on mutation to alanine (0.3 and 0.4 kcal $mol^{-1}$ per monomer for K325 and K327, respectively) is restored by mutations to arginine (68), and suggest electrostatic contacts with the backbone. Residue V324 of the ß2-ß3 loop does not contribute to binding energy

The thermal energy at room temperature (25 °C) is ~0.6 kcal $mol^{-1}$, and the small effects of individual mutations indicate that at the higher physiological temperatures (37°C) the interface will present an even highly dynamic character with individual interactions being constantly formed and broken. The absolute free energy assigned to each contact is small, and the resulting picture is a robust yet flexible highly solvated interface, that can tolerate point mutations. Water mediated interactions can be as important as direct hydrogen bonds for specificity in molecular recognition (reviewed by Levy and Onuchic (83)) and to a lesser extent, for stability, mainly governed by ionic interactions. There are estimations of the number of interface waters from crystallographic (84) and solvent perturbation studies (85), but these are probably underestimates. The complex between E2 and DNA appears as a paradigmatic example of a "wet" protein interface as opposed to the "dry" interface found in TATA box binding protein (84) or in protein-protein interfaces. For instance, the BPV1-E2C DNA was shown to have 42 water molecules mediating protein-DNA interactions (83). A very similar situation is found for the HPV-18 E2C DNA complex (Figure 7), where a large number of water molecules are found in the proximity of the region where the non-observed ß2-ß3 loop is expected to be. It is very likely that interface water molecules are constantly filling gaps that may arise, for instance, upon mutations, giving a highly dynamic nature and complementarities of hydration, with a compensatory effect on mutation.

The residues at the ß2-ß3 interface (K325 and K327) appear to contact DNA only temporarily, and the free energy change on mutation is restored by an arginine, something that does not happen with the lysines in the DNA recognition helix (68). The highly dynamic nature of this loop is also supported by a free energy change near or below the thermal energy, and the lysine-arginine tolerance also indicates that there are no steric constrains, once again supporting flexibility. It is possible that this region is active in the initial stages of DNA recognition by steering the electrostatic interactions in the encounter complex, and that it remains flexible in the consolidated complex to allow bending of the central spacers in the E2 DNA sites (44).

HPV 16 E2C-DNA interaction is very sensitive to ionic strength, in fact, concentrations of 0.2M NaCl were required to avoid aggregation in titrations for the analysis of binding in solution at equilibrium where excess of protein is added (67). In addition, ionic strength and phosphate stabilize the E2C dimer by 5.0 kcal $mol^{-1}$ to urea denaturation-dissociation (72). A recent work reported that E2 sites with different spacer sequences can incorporate cations at low ionic strength producing different degrees of bending depending on the sequence, which in turn, affects specificity (86). These cations would be located in the minor groove of the spacer sequences. In contrast, at high $K^+$ concentrations, the typical cation displacement efffect seen in protein-DNA complexes is observed. This study suggests that monovalent and divalent cations can be an additional factor modulating specificity in DNA recognition in the different processes mediated by the E2C domains (86). Cation release at high ionic strength was also shown to take place in the HPV16 E2C (85). Divalent cations have been reported to facilitate the transfer of E2C from non-specific to specific DNA sequences although the mechanism responsible has yet to be determined (87).

### 3.9. HPV 16 E2C can enter a quasi-spontaneous amyloid route

Although the recombinant HPV 16 E2C is a stable and readily purified domain, along the years, we

have observed that the sample storing conditions, repetitive freeze and thaw cycles, or heating, lead to different amounts of aggregated conformers that affected some determinations and had to be eliminated, and we even observed gelified species in NMR samples (unpublished results from GPG and DC). These conformers show an increase in ß-sheet content by circular dichroism. The addition of small amounts of the solvent trifluoroethanol (12% TFE) at pH 5.6, and most importantly, at low protein concentrations (1 μM, and probably lower extending the elapsed time) and at room temperature, lead to the formation of an oligomeric species with increased ß-sheet content, and an expanded nature (88). The oligomer is slowly transformed into insoluble short "curly type" amyloid fibres as visualized by electron microscopy, with typical dye binding properties of amyloid structures.

Addition of stoichiometric amounts of specific DNA completely prevents the formation of the oligomeric ß-sheet species and the amyloid route, suggesting a role for the local unfolding of the major DNA binding helix, something also suggested by the slowing of the reaction at pH 7.0 over pH 5.6. The mild perturbation required to initiate this change is indicative of a pre-existing equilibrium, which is shifted by addition of small amounts of TFE. However, addition of over 30% TFE leads to species with non-native increased helical content. An important observation is that the formation of neither the ß-sheet oligomer intermediate nor that of the fibres can be achieved from unfolded protein, suggesting that partly folded/structured species are required, in line with general observations on amyloid forming proteins (89).

Although there is so far no evidence for a biological implication of the amyloid route, a functional significance or effect, either directly or as a side reaction, cannot be discarded due to the existing equilibrium and the mild perturbation required to enter the amyloid route. In a recent study (90), E2 was shown to be associated along the mitotic spindle fibres throughout the different phases of mitosis, which implies some sort of polymerization. Furthermore, the region mapped for this association was a β-strand of the C-terminal domain, the object of the present review. Oligomerization processes are required for the assembly of transcription initiation and DNA replication machineries, where proteins from diverse viruses must come together with host cell proteins. The E2 protein is a virus encoded multifunctional master regulator that may exert at least one of its multiple functions through its ability to oligomerize. Since the E2 proteins have multiple putative interaction partners one possible role of oligomerization might be to allow the integration of inputs from multiple partners by an E2 assembly.

### 3.10. Concluding remarks

The available structures of E2C proteins are starting to reveal some analogies and differences among proteins belonging to different viral strains. The overall architecture is almost invariant, but slight changes, particularly in the relative orientation of the two recognition helices, are particular features of each protein. Upon DNA binding, this characteristic becomes more uniform, revealing the capacity of the domain to undergo conformational changes, along with the large necessary deformation of the DNA, as determinants in the final complex formation. The HPV 16 E2C is a stable dimeric protein with a free energy of chemical unfolding ranging from ~10 to 16 kcal $mol^{-1}$, depending on the solvent conditions and in particular, pH (69). However, the exchange rates for most of the backbone NHs in the HPV 31 E2C domain (50), and the HPV 16 E2C (91) is high, which challenges the general correlation between compaction, stability, and low backbone amide hydrogen exchange rates in proteins. This could be explained by a constant breathing of the dimer where fast dissociation and association takes place. We hypothesize that the breathing is at the level of partial unfolding and dynamics, not dissociation. In particular, the DNA binding helix exchanges its amidic protons faster than any other region (50,91), implying an equilibrium between a well formed helix and a more disordered conformation.

So far, E2 has been difficult to detect in natural early infections; it can only be visualized in transfected cells. We speculate that the equilibrium monomer-dimer is clearly possible and can be shifted by changes in solution, mutagenesis, or by yet unknown *in vivo* conditions or agents. Finally, we should keep in mind that the N-terminal transactivation domain also has the ability to dimerize, which would make the full-length E2 dimer even tighter (91).

The mutagenesis analysis contributed to elucidate three different crucial aspects of the HPV16 domain. First, the protein-DNA interface, which showed that all interactions made by contacts pre-defined from crystal structure of the homologous HPV18 E2C-DNA, made additive and individually small (less than 1.0 kcal $mol^{-1}$) contributions to binding energy, and no individual "hot spot" contacts could be observed (68). Despite the high affinity of the complex, the small non-cooperative energies are indicative of a dynamic and "wet" interface, where water mediated interactions play a significant role and can replace those eliminated by individual mutations. Second, mutations in the flexible ß2-ß3 loop, indicate that this region does contact DNA, although dynamically (68). Finally, a new contact, lysine 349, the first contact identified outside of the DNA binding helix, actually located in the loop connecting α2 with β4 (54). The contact made by lysine 349, not present in several relevant strains, may be important in the enhanced transcriptional control of E2 from HPV16 (43).

Indirect readout components in protein-DNA acid recognition reflect the conformational and energetic effects on DNA sequences such as twisting, bending or various distortions. Indirect readout was assigned a considerable contributor to the energetics of the E2C-DNA interaction (62,86). However in these studies, the variable is the sequence and conformation of the DNA, with the "wild-type" HPV16 E2C being invariable. In addition, other players such as cations, DNA length or the complex biological environment can further influence positively or negatively the contribution of indirect readout, depending

on how these affect the conformation of DNA and what sort of distortion the specific binding protein must impose. Systematic mutational analysis of HPV16 E2C has revealed that most of the energetic component comes from direct readout and has provided a value of each of the interactions which coincide with the global binding free energy and was corroborated by the double mutant cycles (68).

### 3.11. Future prospects

The current amount of available strains to study, in particular those statistically relevant to cancer or common benign lesions, provide an extensive sample for investigating protein structure and function by various methods. This will allow classification of structures, affinities, species recognition, with biological events such as tropism, family of virus, cancer progression, persistence or gene silencing (20). NMR will provide insights into the overall dynamics of this unique dimeric ß-barrel fold and in particular the DNA binding helix. It will be essential to be able to analyze these structural and dynamic properties in solution with DNA present. Of course a major goal will be high-resolution structures of the overall complexes, and comparison with the conformation of free DNA in solution will be of continuing interest for fundamental scientific reasons.

A detailed structural analysis of the region corresponding to the DNA binding helix will be interesting not only for the ability to bind DNA, but also to investigate the structural tendencies in solution and the possible dynamic structures that may mediate the amyloid route. A recent report claimed that an isolated recognition helix can bind and discriminate specific DNA, but in the same conditions where the full dimeric domain binds (0.2M salt) the affinity is 1000-fold higher for the full domain (93), and it is expected that the full protein architecture positions the helices in a particular strained conformation for binding and regulatory reasons.

The investigation of the folding mechanism will have to be focused on mutants and domains from different HPV strains. In particular it would be desirable to isolate a folded monomeric species in amounts that will allow a functional and structural characterization of an intriguing link in the folding pathway of an intertwined dimeric ß-barrel. Finally, investigation of the protein-DNA interface should be extended with more mutants, other relevant strains and thermodynamic analysis and comparisons of the structures that will enlarge the existing list. Overall, the E2C domain constitutes an ideal model for testing the evolution of structure and function in proteins of a similar origin.

Furthermore, as more data is gathered and we learn about these structures and their connection with mechanism, the availability of structures and computational and chemical/combinatorial methods will allow the development of different compounds that can aim at the many different key species and interactions of E2, such as protein-DNA interfaces, E2-E1 interactions, apoptosis, chromosome tethering, and monomeric species or oligomers, in a search for new antivirals. The existence of effective vaccines will not affect the course of infection of millions of women worldwide, in particular those in developing countries. The tremendous prophylactic effect of the vaccine is counterbalanced by the lack of accessibility by most of the population.

## 4. ACKNOWLEDGEMENTS

The authors are indebted with Prof. Mattia Falconi and Prof. Alessandro Desideri for the DNA complex model shown in Figure 4 . DOC is grateful to Prof. Maurizio Paci for his constant stimulation and help. The work was supported by Wellcome Trust Grant GR077355AIA. GPG is a Career Investigator from CONICET.

## 5. REFERENCES

1. Spalholtz B.A., Y. C. Yang & P. M. Howley: Transactivation of a bovine papilloma virus transcriptional regulatory element by the E2 gene product. *Cell*, 42, 183-191 (1985)
2. McBride, A. A., H. Romanczuk & P. M. Howley: The papillomavirus E2 regulatory proteins. *J. Biol. Chem*., 266, 18411-18414 (1991)
3. Hegde, R. S. , S. R. Grossman, L. A. Laimins & P. B. Sigler: Crystal structure at 1.7 Å of the bovine papillomavirus-1 E2 DNA-binding domain bound to its DNA target. *Nature*, 359, 505-511 (1992)
4. Bochkarev, A. , J. A. Barwell, R. A. Pfuetzner, W. Furey Jr, A. M. Edwards & L. Frappier: Crystal structure of the DNA-binding domain of the Epstein-Barr virus origin-binding protein EBNA 1. *Cell*, 83, 39-46 (1995)
5. Lowy, D. R. & P. M. Howley in *Fields Virology* (B. N. Fields, P. M. Howley, D. E. Griffin, R. A. Lamb, M. A. Martin, B. Roizman, S. E. Straus and D. M. Knipe, Eds.) pp 2213-2219, Lippincott Williams and Wilkins, Philadelphia (2001)
6. I. M. Morgan and M. M. in *Papillomavirus Research* (M. S. Campo, Ed.) pp 73-82, Caister Academic Press, Wymondham, Norfolk, U.K. (2006)
7. McBride, A. A. & G. Myers: The E2 proteins: an update. (G. Myers, C. Baker, K. Munger, F. Sverdrup, A. McBride and H.-U. Bernard, Eds.), *Human Papillomaviruses*. Los Alamos National Laboratory, Los Alamos (1997)
8. R. S. Hegde: The papillomavirus E2 proteins: structure, function, and biology. *Annu. Rev. Biophys. Biomol. Struct*., 31, 343-60 (2002)
9. Doorbar, J.: Molecular biology of human papillomavirus infection and cervical cancer. Clin. Sci. (Lond), 110, 525-41 (2006)
10. Abbate, E. A., J. M. Berger & M. R. Botchan: The X-ray structure of the papillomavirus helicase in complex with its molecular matchmaker E2. *Genes Dev*., 18, 1981-1996 (2004)
11. Clower, R. V., Y. Hu & T. Melend: Papillomavirus E2 protein interacts with and stimulates human topoisomerase I. *Virology*, 348, 13-18 (2006)
12. Massimi, P., D. Pim, C. Bertoli, V. Bouvard & L. Banks: Interaction between the HPV-16 E2 transcriptional activator and p53. *Oncogene*, 18, 7748-7754 (1999)

13. Boner, W., E. R. Taylor, E. Tsirimonaki, K. Yamane, M. S. Campo & I. M. Morgan: A functional interaction between the human papillomavirus 16 transcription/replication factor E2 and the DNA damage response protein TopBP1. *J. Biol. Chem*., 277, 22297-22303 (2002)
14. You, J., J. L. Croyle, A. Nishimura, K. Ozato & P. M. Howley: Interaction of the bovine papillomavirus E2 protein with Brd4 tethers the viral DNA to host mitotic chromosomes. *Cell*, 117, 349-360 (2004)
15. Parish, J. L., A. M. Bean, R. B. Park & E. J. Androphy: ChlR1 is required for loading papillomavirus E2 onto mitotic chromosomes and viral genome maintenance. *Mol. Cell*, 24, 867-876 (2006)
16. Cripe, T. P., T. H. Haugen, J. P. Turk, F. Tabatabai, P. G. Schmid 3rd, M. Durst, L. Gissmann, A. Roman & L. P. Turek: Transcriptional regulation of the human papillomavirus-16 E6-E7 promoter by a keratinocyte-dependent enhancer, and by viral E2 trans-activator and repressor gene products: implications for cervical carcinogenesis. *EMBO J*., 6, 3745-3753 (1987)
17. Tan, S. H., L. E. Leong, P. A. Walker & H. U. Bernard: The human papillomavirus type 16 E2 transcription factor binds with low cooperativity to two flanking sites and represses the E6 promoter through displacement of Sp1 and TFIID. *J. Virol*., 68, 6411-6420 (1994)
18. Bouvard, V., A. Storey, D. Pim & L. Banks: Characterization of the human papillomavirus E2 protein: evidence of trans-activation and trans-repression in cervical keratinocytes. *EMBO J.*, 13, 5451-5459 (1994)
19. Thierry, F. & M. Yaniv: The BPV1-E2 trans-acting protein can be either an activator or a repressor of the HPV18 regulatory region. EMBO J., 6 ,3391-3397 (1987)
20. Wu, S. Y., A. Y. Lee, S. Y. Hou, J. K. Kemper, H. Erdjument-Bromage, P. Tempst & C. M. Chiang: Brd4 links chromatin targeting to HPV transcriptional silencing. *Genes Dev*., 20, 2383-2396 (2006)
21. Bechtold, V., P. Beard & K. Raj: Human papillomavirus type 16 E2 protein has no effect on transcription from episomal viral DNA. *J. Virol*., 77, 2021-2028 (2003)
22. Stubenrauch, F., A. M. Colbert & L. A. Laimins: Transactivation by the E2 protein of oncogenic human papillomavirus type 31 is not essential for early and late viral functions. *J. Virol*., 72, 8115-8123 (1998)
23. Knight, J. D., R. Li & M. Botchan: The activation domain of the bovine papillomavirus E2 protein mediates association of DNA-bound dimers to form DNA loops. *Proc. Natl. Acad. Sci. U S A*, 88, 3204-3208 (1991)
24. Barsoum, J., S. S. Prakash, P. Han & E. J. Androphy: Mechanism of action of the papillomavirus E2 repressor: repression in the absence of DNA binding. *J. Virol*., 66, 3941-3945 (1992)
25. Stubenrauch, F., T. Zobel & T. Iftner: The E8 domain confers a novel long-distance transcriptional repression activity on the E8E2C protein of high-risk human papillomavirus type 31. *J. Virol*., 75, 4139-4149 (2001)
26. zur Hausen, H.: Papillomaviruses and cancer: from basic studies to clinical application. *Nat Rev Cancer* 2, 342-350 (2002)
27. Baker, C. C., W. C. Phelps, V. Lindgren, M. J. Braun, M. A. Gonda & P. M. Howley: Structural and transcriptional analysis of human papillomavirus type 16 sequences in cervical carcinoma cell lines . *J. Virol*., 61, 962-971 (1987)
28. Jeon, S., B. L. Allen-Hoffmann & P. F. Lambert: Integration of human papillomavirus type 16 into the human genome correlates with a selective growth advantage of cells. *J. Virol*., 69, 2989-2997 (1995)
29. Walboomers, J. M., M. V. Jacobs, M. M. Manos, F. X. Bosch, J. A. Kummer, K. V. Shah, P. J. Snijders, J. Peto, C. J. Meijer & N. Munoz: Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. *J. Pathol*., 189, 12-19 (1999)
30. Gammoh, N., H. S. Grm, P. Massimi & L. Banks: Regulation of human papillomavirus type 16 E7 activity through direct protein interaction with the E2 transcriptional activator. *J. Virol*., 80, 1787-1797 (2006)
31. Grm, H. S., P. Massimi, N. Gammoh & L. Banks: Crosstalk between the human papillomavirus E2 transcriptional activator and the E6 oncoprotein. *Oncogene*, 24, 5149-5164 (2005)
32. Bontkes, H. J., T. D. de Gruijl, A. Bijl, R. H. Verheijen, C. J. Meijer, R. J. Scheper, P. L. Stern, J. E. Burns, N. J. Maitland and J. M. Walboomers: Human papillomavirus type 16 E2-specific T-helper lymphocyte responses in patients with cervical intraepithelial neoplasia. *J. Gen. Virol*., 80, 2453-2459 (1999)
33. Brandsma, J. L., M. Shlyankevich, L. Zhang, M. D. Slade, E. C. Goodwin, W. Peh & A. B. Deisseroth: Vaccination of rabbits with an adenovirus vector expressing the papillomavirus E2 protein leads to clearance of papillomas and infection. *J. Virol*., 78, 116-123 (2004)
34. Kasukawa, H., P. M. Howley & J. D. Benson: A fifteen-amino-acid peptide inhibits human papillomavirus E1-E2 interaction and human papillomavirus DNA replication *in vitro*. *J. Virol.*, 72, 8166-8173 (1998)
35. Androphy, E. J., D. R. Lowy & J. T. Schiller: Bovine papillomavirus E2 trans-activating gene product binds to specific sites in papillomavirus DNA. *Nature*, 325, 70-73 (1987)
36. Moskaluk, C. & D. Bastia: The E2 “gene” of bovine papillomavirus encodes an enhancer-binding protein. *Proc. Natl. Acad. Sci. U S A*, 84, 1215-1218 (1987)
37. Li, R., J. Knight, G. Bream, A. Stenlund & M. Botchan: Specific recognition nucleotides and their DNA context determine the affinity of E2 protein for 17 binding sites in the BPV-1 genome. *Genes Dev*., 3, 510-526 (1989)
38. Thain, A., K. Webster, D. Emery, A. R. Clarke & K. Gaston: DNA binding and bending by the human papillomavirus type 16 E2 protein. Recognition of an extended binding site. *J. Biol. Chem*., 272, 8236-8242 (1997)
39. Alexander, K. A. & W. C. Phelps: A fluorescence anisotropy study of DNA binding by HPV-11 E2C protein: a hierarchy of E2-binding sites. *Biochemistry*, 35, 9864–9872 (1996)
40. Sanders, C. M. & N. J. Maitland: Kinetic and equilibrium binding studies of the human papillomavirus type-16 transcription regulatory protein E2 interacting with core enhancer elements. *Nucleic Acids Res*., 22, 4890-4897 (1994)
41. Dell, G., K. W. Wilkinson, R. Tranter, J. Parish, R. L. Brady & K. Gaston: Comparison of the structure and DNA-

binding properties of the E2 proteins from an oncogenic and a non-oncogenic human papillomavirus. *J. Mol. Biol*., 334, 979-991 (2003)
42. Pepinsky, R. B., S. S. Prakash, K. Corina, M. J. Grossel, J. Barsoum & E. J. Androphy: Sequences flanking the core DNA-binding domain of bovine papillomavirus type 1 E2 contribute to DNA-binding function. *J. Virol*., 71, 828-831 (1997)
43. Hou, S. Y., S. Y. Wu & C. M. Chiang: Transcriptional activity among high and low risk human papillomavirus E2 proteins correlates with E2 DNA binding. *J. Biol. Chem*., 277, 45619-45629 (2002)
44. Hines, C. S., C. Meghoo, S. Shetty, M. Biburger, M. Brenowitz & R. S. Hegde: DNA structure and flexibility in the sequence-specific binding of papillomavirus E2 proteins. *J. Mol. Biol*., 276, 809-818 (1998)
45. Bochkarev, A., J. A. Barwell, R. A. Pfuetzner, E. Bochkareva, L. Frappier & A. M. Edwards: Crystal structure of the DNA-binding domain of the Epstein-Barr virus origin-binding protein, EBNA1, bound to DNA. *Cell*, 84, 791-800 (1996)
46. Hegde, R. S., A. F. Wang, S. S. Kim & M. Schapira: Subunit rearrangement accompanies sequence-specific DNA binding by the bovine papillomavirus-1 E2 protein. *J. Mol. Biol*., 276, 797–808 (1998)
47. Veeraraghavan, S., C. C. Mello, E. J. Androphy & J. D. Baleja: Structural correlates for enhanced stability in the E2 DNA-binding domain from bovine papillomavirus. *Biochemistry*, 38, 16115–16124 (1999)
48. Hegde, R. S. & E. J. Androphy: Crystal structure of the E2 DNA-binding domain from human papillomavirus type 16: implications for its DNA binding-site selection mechanism. J. Mol. Biol., 284, 1479–1489 (1998)
49. Nadra, A. D., T. Eliseo, Y. K. Mok, C. L. Almeida, M. Bycroft, M. Paci, G. de Prat-Gay & D. O. Cicero: Solution structure of the HPV-16 E2 DNA binding domain, a transcriptional regulator with a dimeric beta-barrel fold. *J. Biomol. NMR*, 30, 211–214 (2004)
50. Liang, H., A. M. Petros, R. P. Meadows, H. S. Yoon, D. A. Egan, K. Walter, T. F. Holzman, T. Robins & S. W. Fesik: Solution structure of the DNA-binding domain of a human papillomavirus E2 protein: evidence for flexible DNA-binding regions. *Biochemistry*, 35, 2095–2103 (1996)
51. Bussiere, D. E., X. Kong, D. A. Egan, K. Walter, T. F. Holzman, F. Lindh, T. Robins & V. L. Giranda: Structure of the E2 DNA-binding domain from human papillomavirus serotype 31 at 2.4 Å. *Acta Crystallogr. D. Biol. Crystallogr*., 54, 1367–1376 (1998)
52. Kim, S. S., J. K. Tam, A. F. Wang & R. S. Hegde: The structural basis of DNA target discrimination by papillomavirus E2 proteins. *J. Biol. Chem*., 275, 31245–31254 (2000)
53. Hooley, E., V. Fairweather, A. R. Clark, K. Gaston & L. Brady: The recognition of local DNA conformation by the human papillomavirus type 6 E2 protein. *Nucleic Acids Res*., 24, 3897-3908 (2006)
54. Cicero, D. O., A. D. Nadra, T. Eliseo, M. Dellarole, M. Paci & G. de Prat-Gay: Structural and thermodynamic basis for the enhanced transcriptional control by the human papillomavirus strain-16 E2 protein. Biochemistry, 45, 6551–6560 (2006)
55. Bax, A. & A. Grishaev: Weak alignment NMR: a hawk-eyed view of biomolecular structure. *Curr. Opin. Struct. Biol*., 15, 563-570 (2005)
56. Bedrosian, C. L. & D. Bastia: The DNA binding domain of HPV-16 E2 protein interaction with the viral enhancer: protein induced DNA bending and role of the nonconserved core sequence in binding site affinity. *Virology*, 174, 557–575 (1990)
57. Rozenberg, H., D. Rabinovich, F. Frolow, R. S. Hegde & Z. Shakked: Structural code for DNA recognition revealed in crystal structures of papillomavirus E2-DNA targets. Proc. Natl. Acad. Sci. U S A, 95, 15194–15199 (1998)
58. Hizver, J., H. Rozenberg, F. Frolow, D. Rabinovich & Z. Shakked: DNA bending by an adenine–thymine tract and its role in gene regulation. *Proc. Natl. Acad. Sci. U S A*, 98, 8490–8495 (2001)
59. Rhos, R., H. Sklenar & Z. Shakked: Structural and energetic origins of sequence-specific DNA bending: Monte Carlo simulations of papillomavirus E2-DNA binding sites. *Structure*, 13, 1499-1509 (2005)
60. Zimmerman, J. M. & L. J. Maher III: Solution measurement of DNA curvature in papillomavirus E2 binding sites. *Nucleic Acids Res.*, 31, 5134–5139 (2003)
61. Zinkel, S. S. & D. M. Crothers: DNA bend direction by phase sensitive detection. *Nature*, 328, 178-181 (1987)
62. Zhang, Y. L., Z. Q. Xi, R. S. Hegde, Z. Shakked & D. M. Crothers: Predicting indirect readout effects in protein-DNA interactions. *Proc. Natl. Acad. Sci. U S A*, 101, 8337–8341 (2004)
63. Zhang, Y. L. & D. M. Crothers: High-throughput approach for detection of DNA bending and flexibility based on cyclization. *Proc. Natl. Acad. Sci. U S A*, 100, 3161–3166 (2003)
64. Zhang, Y. L. & D. M. Crothers: Statistical Mechanics of Sequence-Dependent Circular DNA and Its Application For DNA Cyclization. *Biophys. J*., 84, 136–153 (2003)
65. Vermeulen, A., H. Zhou & A. Pardi: Determining DNA Global structure and DNA bending by application of NMR residual dipolar couplings. *J. Am. Chem. Soc*., 122, 9638-9647 (2000)
66. Wu, Z., F. Delaglio, N. Tjandra, V. B. Zhurkin & A. Bax: Overall structure and sugar dynamics of a DNA dodecamer from homo- and heteronuclear dipolar couplings and 31P chemical shift anisotropy. *J. Biomol. NMR*, 26, 297-315 (2003)
67. Ferreiro, D. U., M. T. R. Lima, A. D. Nadra, L. G. Alonso, F. A. Goldbaum & G. de Prat-Gay: Distinctive cognate sequence discrimination, bound DNA conformation, and binding modes in the E2 C-terminal domains from prototype human and bovine papillomaviruses. *Biochemistry*, 39, 14692-14701 (2000)
68. Ferreiro, D. U., M. Dellarole, A. D. Nadra & G. de Prat-Gay: Free energy contributions to direct readout of a DNA sequence. *J. Biol. Chem*., 280, 32480-32484 (2005)
69. Mok, Y. K., G. de Prat-Gay, P. J. Butler & M. Bycroft: Equilibrium dissociation and unfolding of the dimeric human papillomavirus strain-16 E2 DNA-binding domain. *Protein Sci*., 5, 310-319 (1996)
70. Silva, J. L. & G. Weber: Pressure stability of proteins. *Annu. Rev. Phys. Chem*., 44, 89-113 (1993)

71. Foguel, D., J. L. Silva & G. de Prat-Gay: Characterization of a partially folded monomer of the DNA-binding domain of human papillomavirus E2 protein obtained at high pressure. *J. Biol. Chem*., 273, 9050-9057 (1998)
72. Lima, L.M. & G. de Prat-Gay: Conformational changes and stabilization induced by ligand binding in the DNA-binding domain of the E2 protein from human papillomavirus. *J. Biol. Chem*., 272, 19295-19303 (1997)
73. Mok, Y. K., M. Bycroft & G. de Prat-Gay: The dimeric DNA binding domain of the human papillomavirus E2 protein folds through a monomeric intermediate which cannot be native-like. *Nat. Struct. Biol*., 3, 711-717 (1996)
74. Prat-Gay, G. de, A. D. Nadra, F. J. Corrales-Izquierdo, L. G. Alonso, D. U. Ferreiro & Y. K. Mok: The folding mechanism of a dimeric beta-barrel domain. *J. Mol. Biol*., 351, 672-682 (2005)
75. Waldburger, C. D., T. Jonsson & R. T. Sauer: Barriers to protein folding: formation of buried polar interactions is a slow step in acquisition of structure. *Proc. Natl. Acad. Sci. U S A,* 93, 2629–2634 (1996)
76. Gloss, L. M. & C. R. Matthews: Mechanism of folding of the dimeric core domain of Escherichia coli trp repressor: a nearly diffusion-limited reaction leads to the formation of an on-pathway dimeric intermediate. *Biochemistry*, 37, 15990–15999 (1998)
77. Prat-Gay, G. de: Association of complementary fragments and the elucidation of protein folding pathways. *Protein Eng*., 9, 843-847 (1996)
78. Mok, Y. K., L. G. Alonso, L. M. Lima, M. Bycroft & G. de Prat-Gay: Folding of a dimeric beta barrel: residual structure in the urea denatured state of the human papillomavirus E2 DNA binding domain. Protein Sci., 9, 799–811 (2000)
79. Ferreiro, D. U. & G. de Prat-Gay: A protein-DNA binding mechanism proceeds through multi-state or two-state parallel pathways. *J. Mol. Biol.*, 331, 89-99 (2003)
80. Oddo, C., E. Freire, L. Frappier & G. de Prat-Gay: Mechanism of DNA recognition at a viral replication origin. *J. Biol*, *Chem*., 281, 26893-26903 (2006)
81. Wells, J. A.: Systematic mutational analyses of protein-protein interfaces. *Methods Enzymol*., 202, 390-411 (1991)
82. Fersht, A.R. in Structure and Mechanism in Protein Science, Chapter 3, pp 103-131, W. H. Freeman and Company Ed (1999)
83. Levy, Y. & J. N. Onuchic: Water mediation in protein folding and molecular recognition. *Annu. Rev. Biophys. Biomol*. *Struct*., 35, 389-415 (2006)
84. Janin, J.: Wet and dry interfaces: the role of solvent in protein-protein and protein-DNA recognition. *Structure Fold. Des*., 7, R277-R279 (1999)
85. Lima, L. M. & J. L. Silva: Positive contribution of hydration on DNA binding by E2c protein from papillomavirus. *J Biol Chem* 279, 47968-47974 (2004)
86. Blakaj, D. M., C. Kattamuri, S. Khrapunov, R. S. Hegde & M. Brenowitz: Indirect readout of DNA sequence by papillomavirus E2 proteins depends upon net cation uptake. *J. Mol. Biol*., 358, 224-240 (2006)
87. Lewis, H. & K. Gaston: Magnesium ions enhance the transfer of human papillomavirus E2 protein from non-specific to specific binding sites. *J. Mol. Biol*., 294, 885-896 (1999)
88. Wetzler, D., E. Castaño & G. de Prat-Gay: A quasi-spontaneous amyloid route in a DNA binding gene regulatory domain of the papillomavirus HPV16 E2 protein. *Prot. Sci*. in press (2007)
89. Dobson, C.M.: An Overview of Protein Misfolding Diseases. In *Protein Folding Handbook*. (eds. J. Buchner, and T. Kiefhaber), pp. 1093-1113. Wiley-VCH Verlag GMBH & Co. KgaA, Weinheim (2005)
90. Dao, L. D., A. Duffy, B. A. Van Tine, S. Y. Wu, C. M. Chiang, T. R. Broker & L. T. Chow: Dynamic localization of the human papillomavirus type 11 origin binding protein E2 through mitosis while in association with the spindle apparatus. *J. Virol*., 80, 4792-4800 (2006)
91. Nadra, A. D., : "Estudios estructurales en solución del factor de transcripción E2C de HPV-16 y su interacción con ADN", Ph.D. Thesis, University of Buenos Aires (2005)
92. Antson, A. A., J. E. Burns, O. V. Moroz, D. J. Scott, C. M. Sanders, I. B. Bronstein, G. G. Dodson, K. S. Wilson & N. J. Maitland: Structure of the intact transactivation domain of the human papillomavirus E2 protein. *Nature*, 403, 805-809 (2000)
93. Faber-Barata, J., R. Mohana-Borges, L. M. Lima: Specificity in DNA recognition by a peptide from papillomavirus E2 protein. *FEBS Lett*., 580, 1919-1924 (2006)

**Key Words:** HPV, E2, DNA Replication, Transcription, Dna Binding, Anti-Viral, Review

**Send correspondence to:** Dr Gonzalo Prat Gay, Instituto Leloir and IIBBA-CONICET, Av. Patricias Argentinas 435, (1405) Buenos Aires, Argentina, Tel: 54-11-5238-7500, Fax: 54-11-5238-7501, E-mail: gpratgay@leloir.org.ar

**http://www.bioscience.org/current/vol13.htm**