

**Universita' degli Studi di Roma** *Tor Vergata*

**Dottor RICCARDO MELIS**

**Dottorato in Biochimica e Biologia Molecolare**



**UNIVERSITÀ DEGLI STUDI DI ROMA  
"TOR VERGATA"**

FACOLTA' DI MEDICINA E CHIRURGIA

DOTTORATO DI RICERCA IN  
BIOCHIMICA E BIOLOGIA  
MOLECOLARE

XXI CICLO

STRUCTURAL ASPECTS OF DYNAMIC AND DNA RECOGNITION  
OF HPV-16 E2C PROTEIN

Dr. Riccardo Melis

A.A. 2008/2009

Docente Guida/Tutor: Prof. Daniel Oscar Cicero

Coordinatore: Prof. Daniel Oscar Cicero

<b><u>ABSTRACT</u></b>	<b>4</b>
<b><u>OBJECTIVES OF THE RESEARCH</u></b>	<b>5</b>
<b><u>INTRODUCTION</u></b>	<b>6</b>
<b><u>NUCLEAR MAGNETIC RESONANCE IN STRUCTURAL BIOLOGY</u></b>	<b>31</b>
<b><u>MATERIALS AND METHODS</u></b>	<b>38</b>
<b><u>RESULTS</u></b>	<b>43</b>
<b><u>DISCUSSION</u></b>	<b>66</b>
<b><u>CONCLUSIONS</u></b>	<b>73</b>
<b><u>BIBLIOGRAPHY</u></b>	<b>76</b>
<b><u>CURRICULUMVITAE</u></b>	
<b>85</b>	
<b><u>LIST OF PUBLICATIONS</u></b>	
<b>87</b>	

## **ABSTRACT**

The Human Papillomavirus (HPV) infection is linked to cervical cancer and represents a serious problem for women health worldwide .The existence of effective vaccines will not affect the course of infection , in particular those in developing countries and the tremendous prophylactic effect of the vaccine is counterbalanced by the lack of accessibility to most of the population. In view of these considerations, there is a need to develop specific antiviral drugs to prevent HPV infections. My PhD work was focused on the particular properties and structure-functional aspects of the E2 HPV16 DNA binding domain (E2C) E2C is the only transcription factor encoded by the viral genome and plays a central role in controlling the expression of all Papilomavirus genes and in regulating the virus life cycle .

## **OBJECTIVES OF THE RESEARCH**

Our objective was to learn about the structure and its connection with the function of the E2C protein using NMR spectroscopy . In particular we focused our research on

- Structural characterization of the recognition helix of E2C to explore the differences with the whole protein.
- Study about the protein dynamics and the connection with the DNA recognition mechanism.
- Characterization of the role that key residues play in determining the interaction of the protein with DNA
- The study of a monomerized version of the protein to infer the role of dimer-monomer equilibrium in the structural properties of the protein.

## INTRODUCTION

Human papillomavirus (HPV) is the name of a large group of viruses that includes more than 100 different strains or types. More than 30 strains of these viruses are responsible of a sexually transmitted disease (STD) that infect the skin and mucosal areas of the body of men and women. Today approximately 20 million people are affected by genital Papilloma virus in the world. At least 50% of sexually active men and women acquire genital HPV infection at some point in their lives and, by age 50, at least 80% of women will have acquired genital HPV infection. In the U.S. country, for example, about 6.2 million Americans get a new genital HPV infection each year. The mucosal HPV types are recognized as the major cause of cervical cancer., the second most common cancer in women, and are also a causative agent of vaginal, anal, penile and head and neck cancer (Bosch et al. 2006; D'Souza et al. 2007). Moreover, the mucosal HPV types can infect also the genital areas of men, including the skin on and around the penis or anus. According to epidemiological evidence and oncogenic potential, the mucosal HPV are classified as “high-risk” and “low-risk” types (Munoz et al. 2003). To the “high-risk” group belong all oncogenic or carcinogenic HPVs types that are involved in malign tumors insorgents while HPVs belong to “low-risk” group may cause warts, or skin papillomas, which are benign (noncancerous) tumors. Actually are knowed 19 HPVs types classified as “high-risk” (types 16, 18, 26, 31, 33, 35, 39, 45, 51, 52, 53, 56, 58, 59, 66, 68a, 73, 82, 82subtype) and 13 as “low-risk” (types 6, 6a, 6b, 11, 40, 42, 43, 44, 54, 61, 70, 72 and 81) .

### **HPV diagnosis**

Unfortunately, most HPV infections have a problematic diagnostic because the virus that lives in the skin or mucous membranes usually causes no syntomatic evidence except in some cause of visible genital warts, or insurgence of visible genital pre-cancerous changes . Therefore, most infected persons are unaware they are infected, yet they can transmit the virus to a sex partner. Rarely, a pregnant woman can also pass HPV to her baby during vaginal delivery, furthermore a baby that is exposed to HPV very rarely develops warts in the throat or voice box. Some people can present genital warts which usually appear as soft, moist, pink, or flesh-colored swellings, especially in the genital area. They can be raised or flat, single or multiple, small or large, and sometimes cauliflower shaped. They can appear on the vulva, in or around the vagina or anus, on the cervix, and on the penis, scrotum, groin, or thigh. After sexual contact with an infected person, warts may appear within weeks or months, or not at all. In these case Genital warts are diagnosed by visual inspection. Visible genital warts can be removed by medications the patient applies, or by treatments performed by a health care provider. Some individuals choose to forego treatment to see if the warts will disappear on their own. No

treatment regimen for genital warts is better than another, and no one treatment regimen is ideal for all cases. Most women are diagnosed with HPV on the basis of abnormal Pap tests. The Pap test is the primary cancer-screening tool for cervical cancer or pre-cancerous changes in the cervix, many of which are related to HPV. Also, a specific test is available to detect HPV DNA in women. The Pap test involves the collection of cells from the cervix for examination under the microscope. and may eventually lead to cancer if left untreated. Pap test results may also be described using an older set of categories called the “dysplasia scale.” Dysplasia is a term used to describe abnormal cells. Although dysplasia is not cancer, it may develop into very early cancer of the cervix. The cells look abnormal under the microscope, but they do not invade nearby healthy tissue. There are four degrees of dysplasia: mild, moderate, severe, and carcinoma in situ. Carcinoma in situ is a precancerous condition that involves only the layer of cells on the surface of the cervix, and has not spread to nearby tissues. Cervical intraepithelial neoplasia (CIN) is another term that is sometimes used to describe abnormal tissue findings. Neoplasia means an abnormal growth of cells. The term CIN along with a number (1, 2, or 3) describes how much of the thickness of the lining of the cervix contains abnormal cells. CIN-3 is considered to be a precancerous condition that includes carcinoma in situ. The results of HPV DNA testing can help health care providers decide if further tests or treatment are necessary. No HPV tests are available for men. Actually there is no "cure" for HPV infection, although in most women the infection goes away on its own. The treatments provided are directed to the changes in the skin or mucous membrane caused by HPV infection, such as warts and pre-cancerous changes in the cervix. All types of HPV can cause mild Pap test abnormalities which do not have serious consequences. Approximately 10 of the 30 identified genital HPV types can lead, in rare cases, to development of cervical cancer. Research has shown that for about 90% of women, cervical HPV infection becomes undetectable within two years. Although only a small proportion of women have persistent infection, persistent infection with "high-risk" types of HPV is the main risk factor for cervical cancer. A Pap test can detect pre-cancerous and cancerous cells on the cervix. Regular Pap testing and careful medical follow-up, with treatment if necessary, can help ensure that pre-cancerous changes in the cervix caused by HPV infection do not develop into life threatening cervical cancer. The Pap test used in U.S. cervical cancer screening programs is responsible for greatly reducing deaths from cervical cancer. For 2004, the American Cancer Society estimates that about 10,520 women will develop invasive cervical cancer and about 3,900 women will die from this disease. Most women who develop invasive cervical cancer have not had regular cervical cancer screening. The surest way to eliminate risk for genital HPV infection is to refrain from any genital contact with another individual. For those who choose to be sexually active, a long-term, mutually monogamous relationship with an uninfected partner is the strategy most likely to prevent future genital HPV infections. However, it is difficult to determine whether a partner who has been sexually active in the past is currently



infected. For those choosing to be sexually active and who are not in long-term mutually monogamous relationships, reducing the number of sexual partners and choosing a partner less likely to be infected may reduce the risk of genital HPV infection. Partners less likely to be infected include those who have had no or few prior sex partners. HPV infection can occur in both male and female genital areas that are covered or protected by a latex condom, as well as in areas that are not covered. While the effect of condoms in preventing HPV infection is unknown, condom use has been associated with a lower rate of cervical cancer, an HPV-associated disease.

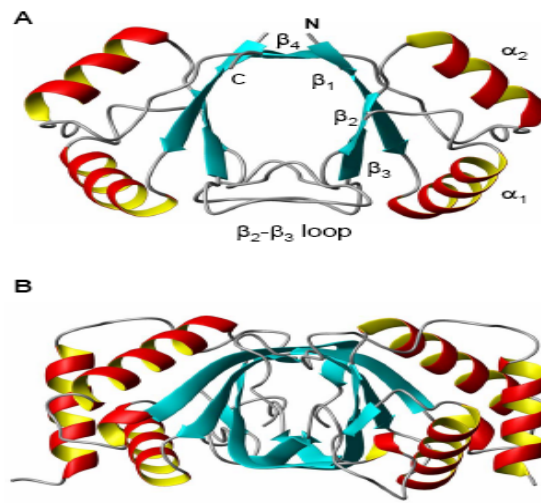
### **HPV vaccine prospective**

No HPV vaccines has been proven to provide complete protection against persistent infection with other HPV types. Therefore, about 30 % of cervical cancers and 10 % of genital warts will not be prevented by these vaccines. In addition because the vaccines will not protect against all infections that cause cervical cancer, it is important for vaccinated women to continue to undergo cervical cancer screening as is recommended for women who have not been vaccinated. Briefly the HPV vaccines work like other immunizations that guard against viral infection. The advanced hypothesis is that the HPVs unique surface components might create an antibody response that is capable of protecting the body against infection, and these components could be used to form the basis of a vaccine. These surface components can interact with one another to form virus-like particles (VLP) that are noninfectious and stimulate the immune system to produce antibodies that can prevent the complete papillomavirus from infecting cells. They are thought to protect primarily by causing the production of antibodies that prevent infection and the development of those cervical cell changes seen on Pap tests that may lead to cancer. Although these vaccines can help prevent HPV infection, it is not known if they can help eliminate existing cervical cell changes due to HPVs. Very recently, a recombinant vaccine produced by Merck & Co., Inc. (Merck), the Gardasil, was approved as prophylactic HPV vaccine. It is called quadrivalent vaccine because it protects against four HPV types (6, 11, 16, and 18). Another promising vaccine, Cervarix™, which targets HPV types 16 and 18, produced by GlaxoSmithKline (GSK), was approved by the FDA in 2007. However, they are unlikely to be introduced in the short term in developing countries, which account for 80% of the deaths due to cervical cancer (Cohen 2008). Moreover, they do not protect against infection with all high-risk types and cannot cure the millions of people that are already infected. Therefore, there is still a need for understanding the oncogenicity of papillomaviruses in more details

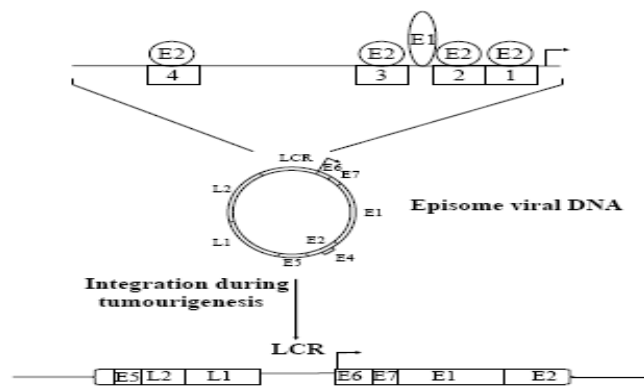
### ***The E2 protein***

The E2 protein is one of the eight proteins expressed by the HPV genome (Figure 1). 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 (Doorbar et al, 2006). 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 a multitude of 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 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 (Abbate et al, 2004). 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 (Clower et al 2006). E2 also binds to proteins involved in DNA repair including the tumour suppressor protein p53 and TopBP1 and E2 may play a role in directing the repair of damaged viral genome (Massimi et al, 1999; Boner et al, 2002). Recent work has shown that E2 is also important for long-term persistence of viral DNA in infected cells (You et al, 2004; Parish et al, 2006). E2 proteins bind simultaneously to the cellular proteins Brd4 and ChlR1 that attach to mitotic chromosomes and to the viral origin of replication (You et al, 2004; Parish et al, 2006). 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).

**Figure 1**



**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 ( Cripe et al, 1987; Tan et al, 1994; Bouvard et al, 1994; Thierry et al, 1987) the binding of E2 to two sites proximal to the transcription start site blocks the binding of cellular transcription factors to adjacent 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 ( Wu et al, 2006). However, it has been claimed that E2 has no effect on viral transcription in the context of the intact viral genome (Bechtold et al, 2003). 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 (Stubenrauch et al, 1998). 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 (Knight et al, 1991). 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 characterised in BPV and HPV. These E2 splice variants can repress viral transcription or at least block the functions of the full-length E2 proteins ( Barsoum et al, 1992; Stubenrauch et al, 1998 ). 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 tumour suppressor proteins p53 and pRB, respectively (Hausen et al, 2002). 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 ( Baker et al, 1987). 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 (Jeon et al, 1995), cell transformation and in due course, tumorigenesis. DNA from highrisk HPV types is found in virtually all human cervical carcinomas (Walboomers et al, 1999). 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 (Parish et al , 2006) and interactions with the E6

and E7 proteins (Gammoh et al, 2006; Grm et al 2005). 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 (Bontkes et al, 1999). 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 (Brandsma et al, 2004). 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 (Kasukawa et al , 1998). Similarly, drugs that prevent the DNA binding activity of E2 would also have the potential to block viral replication.

### **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 (Androphy et al, 1987; Moskaluk et al, 1987). 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 ( Knight et al. 1989). 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 (Knight et al, 1989). 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' (Thain et al, 1997). Like the full-length E2 proteins, the E2C proteins recognise sequences that correspond to the consensus shown above with KDs in the region of 1-10nM (Alexander et al, 1996; Sanders et al 1994). 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 (Thain et al , 1997; Pepinsky et al 1994; Hou et al 2002). The nature of the bases in the spacer region has a profound effect on the binding of some E2 proteins (Hines et al, 1998; Bochkarev et al, 1996) 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 (Thain et al, 1997; Sanders et al, 1994 ). 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 (Thain et al, 1997; Sanders et al, 1994, Dell et al, 2003). 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*.

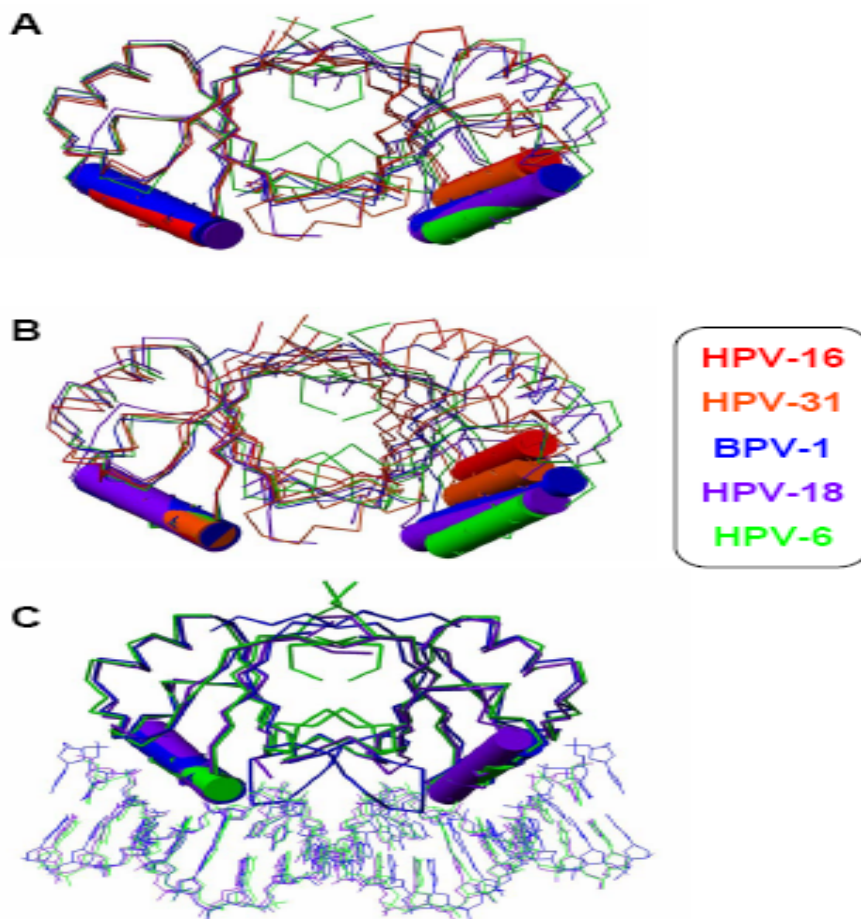
### **The structure of E2C: defining a unique fold**

Fifteen years have passed since the earliest work describing the structure of an E2C-DNA complex belonging to BPV-1 (Hegde et al, 1998). In their pioneering paper, Hegde, Sigler and coworkers described a previously unobserved dimeric antiparallel  $\beta$ -barrel. Contacts with the DNA bases were exclusively observed for two symmetrically disposed  $\alpha$ -helices, the so-called recognition helices, positioned inside two successive major grooves of the DNA. The DNA double helix is severely bent in the structure as a consequence of these interactions. Each monomer displays a  $\beta 1-\alpha 1-\beta 2-\beta 3-\alpha 2-\beta 4$  topology, with a long loop connecting the two repeated motifs (Figure 1A). The dimeric beta-barrel is closed by intermonomer  $\beta 2-\beta 2$  and  $\beta 4-\beta 4$  hydrogen bonds. Remarkably, before the structure of an E2C from any other papillomaviruses was solved, a second example of the dimeric  $\beta$ -barrel was found in the DNA-binding domain of the Epstein-Barr virus origin-binding protein, EBNA1 (Bochkarev et al, 1996). 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 protein interact with the DNA (Bochkarev et al, 1996). In fact, the homologues of the E2C recognition helices appear not to be 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 (Bochkarev et al, 1996). Since then, both x-ray crystallography and NMR have contributed to unraveling the structural features of E2C and details of the DNA interaction mechanism. Analysis of the data contained in some of these structures has already been accomplished exhaustively in a recent review (Hegde et al, 2002). To date, structures of E2C proteins belonging to BPV-1 (Hegde et al, 1998) high risk strains HPV-16 (Hegde et al, 1998; Nadra et al 2004), HPV-31 (Liang et al, 1996; Bussiere et al 1998), HPV-18 (Kim et al, 2000) and the low risk HPV-6 (Dell et al, 2003) have been determined. In addition to the already cited BPV-1-DNA complex (Hegde et al 1992), crystal structures of HPV-18 (Kim et al, 2000) and HPV-6 (Hooley et al 2006) 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 (Cicero et al, 2006), although no high-resolution structure is yet available in this case.

## Crystal structures of E2C domains

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 (Hegde et al, 1998; Bussiere et al 1998)]. HPV-6 shows a longer C-terminus after  $\beta_4$ , 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 behaviour is the  $\beta_2$ - $\beta_3$  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 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 situation that probably is 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 (Hegde et al 2002). 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 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  $\beta_4$ - $\beta_4$  hydrogen bond register, which creates the observed difference in quaternary structure. 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 direct connection between the quaternary structure of E2C. 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 (Hegde et al, 1998; Hooley et al 2006). This is particularly relevant because this helix contains the only residues that make contacts with the DNA bases, and which is 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 from the most internal helix of HPV-16

**Figure 3**





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 C $\alpha$  atoms ranging from 4.7 to 11.0 Å, 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.

### **Solution structures of E2C domains**

Together with crystallography, NMR has contributed to the structure determination of E2C domains. However, from a technical point of view, the description of a structure like that of the dimeric  $\beta$ -barrel is a major challenge for NMR. This is so mainly because solution structure by NMR was classically derived using only short-range interproton distances, along with dihedral angles from coupling constants. Small deviations in the relative position of residues in the central core can be translated into large deviations for the two recognition helices that are positioned at the two sides of the barrel. Nevertheless, the first evidence of a shift in the hydrogen bond register in the  $\beta_4$ - $\beta_4$  sheet and the consequent variation of the quaternary structure came from the solution structure of HPV-31 (Liang et al, 1996). In addition to the structure description, this work also showed that both the recognition helix and the  $\beta_2$ - $\beta_3$  loop presented a dynamical character in solution. Although 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 (Veeraraghavan, et al, 1999). 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) ( Bryce DI et al, 2005) , 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 (Nadra et al 2004). 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).

### **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). All of the E2-BS have two features in common: an inverted repeat consensus sequence of the form 5' ACCG NNNN CCGT 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. On the other hand, BPV-1 E2 protein displays no apparent spacer sequence preference (Hines et al, 1998) . 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 ( Hedge et al 1998-2002). 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 (Rozenberg et al, 1998) and a human E2-BS (Hizver et al, 2001) revealing significant differences between the two. In the first case, DNA bearing an ACGT central spacer shows an  $\alpha$ -helix continuously bent toward the major groove, which for the central spacer differs markedly with the situation in the protein bound state that shows bending towards the minor groove. On the contrary, the E2-BS with an AATT spacer was found to be already bent toward the central minor groove by 9° (Hizver et al, 2001), 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 ( Rozenberg et al, 1998). These features were recently reproduced by Monte Carlo simulations (Rhos et al, 2005) 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

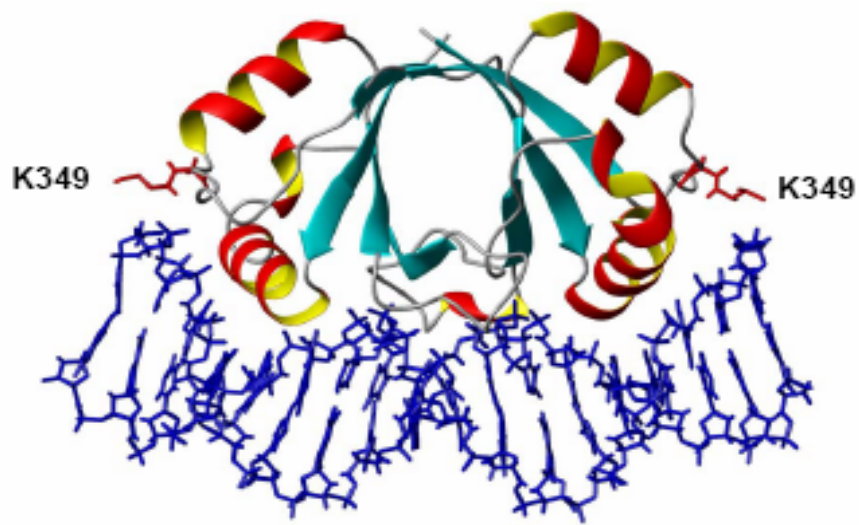
conformation to the protein-bound conformation, and thus modulates the E2 binding affinity. Experimental data derived from high-resolution crystal structures provide detailed structural information only on static conformations of E2-BS. Other techniques that estimate the curvature of the DNA in solution, like gel electrophoretic phasing ( Zimmerman et al, 2003) and cyclization kinetics measurements ( Zhang et al 2004) 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  $18^\circ$  of net curvature, whereas a sequence with the sequence ACGT shows only  $1^\circ$  (Zimmerman et al, 2003). 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, coupled with a statistical mechanical theory ( Zhang et al, 2004). 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 (Zhang et al, 2004). For 15 of these sites, the variation in affinity was predicted within a factor of 3. No attempt has been made so far to obtain a high-resolution structure of one of the E2-BS by NMR. In principle, determination of the curvature by classical NMR methods, as was already discussed for the intermonomer orientation, is difficult, and theoretical studies have demonstrated that without the aid of residual dipolar couplings, the local information content of NOEs and dihedral angles are not sufficient to accurately define the curvature of a short DNA sequence ( Vermeulen et al, 2000). 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)<sub>2</sub> (Wu et al 2003), which differs only by the two first and two last nucleotides with one of the HPV E2-BS (ACCGAATTCGGT) studied by X-ray crystallography . By an extensive use of residual dipolar couplings and <sup>31</sup>P chemical shift anisotropy, the final structure obtained represented a quite regular B-form helix with a global bending of  $\sim 10^\circ$ , coincident to that observed for the E2-BS analogue. This work demonstrated that with this new approach DNA curvature is now accessible to high resolution NMR, and also pointed to the existence of flexibility in certain regions of the molecule, particularly the pyrimidines, by detecting the presence of rapid equilibria between C2'-endo and C3'-endo deoxyribose conformations.

## The structure of E2C-DNA complexes

As already mentioned, HPV and BPV E2C proteins show distinctive cognate sequence discrimination (Ferreiro et al, 2000). One of the main sources for this differential selectivity appears to be of electrostatic nature. BPV-1 E2 displays positive charges located near the C-terminus of the recognition helices, which can help bending even of non-predeformed DNA, explaining the lower selectivity. On the other hand, HPV16 lacks this accumulation of positive charges, and this explains the greater affinity toward sequences with A/T-rich spacers, considered as prebent DNA molecules (Shatzky-Schwartz et al 1997; Stefl et al, 2004). In the case of the HPV-18 E2C protein, there is an accumulation of positive charges in the center of its DNA-interaction surface, which results in electrostatic complementarity to the negative minor groove of A/T-rich sequences, contributing to its enhanced selectivity. However, this is less apparent in the HPV 6 E2C DNA-interaction surface, despite the fact that this protein is more selective towards A/T-rich spacers than the HPV 16 E2C protein. 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-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  $C\alpha$  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. BPV-1 and HPV-18 E2C proteins experience slightly larger changes upon DNA binding (3.4 Å and 9.3° for HPV-18 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 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  $\beta$ 2- $\beta$ 3 loop becomes ordered in the BPV-1 complex, establishing electrostatic interactions with the phosphate groups in the minor groove. The HPV-6 E2C complexes also show an ordered  $\beta$ 2- $\beta$ 3 loop, but its conformation is very different to that observed for the BPV-1 structure, and no interaction with the DNA was detected. HPV-18 E2C shows a disordered loop in the bound state. NMR evidence was also presented about the flexibility of the  $\beta$ 2- $\beta$ 3 loop of the HPV-16 E2C in the DNA complex, however, mutagenesis analysis suggests at least temporary contacts between positively charged residues of this loop

and the DNA (Ferreiro et al, 2005). 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. A large enthalpy difference ( $\Delta\Delta H$ ) of 10.0 kcal mol<sup>-1</sup> 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. 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<sup>-1</sup> per symmetric interaction. This value is similar to the effect of shortening the duplex to 14 bases, which indicates the uncovering of an additional contact between the E2C protein and the DNA outside the recognition helix and the  $\beta$ 2- $\beta$ 3 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. 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  $\beta$ 4- $\beta$ 4 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  $\beta$ 4 strand do not change significantly between the free and bound proteins. 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 (Cicero et al, 2006). 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, 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

**Figure 4**



and BPV-1 displays a very different conformation. 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.

### **The E2 protein folding of a dimeric $\beta$ -barrel domain**

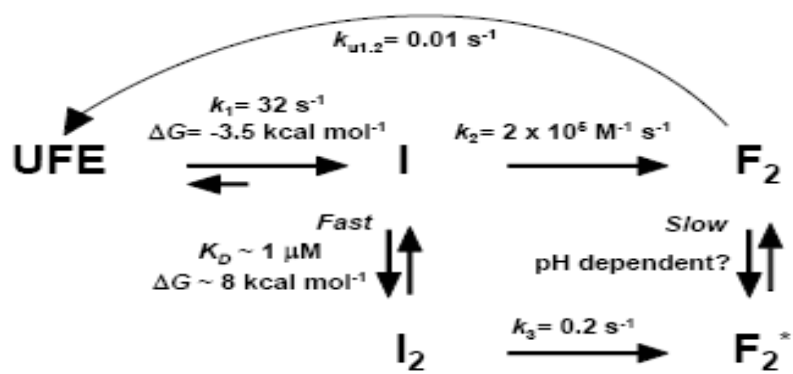
The minimal DNA binding of E2 (E2C) was defined based on sequence alignments and this was confirmed by later structures (see above). The HPV16 E2C 80 amino acid (residues 286-365) domain was recombinantly expressed and initially characterised in solution, before a structure of this HPV type was available (Mok et al, 1996). The existing BPV1 E2C structure anticipated an intertwined domain, where the dimeric interface clearly appears to stabilise 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  $3.0\text{-}6.5 \times 10^{-8}$  M, where  $K_u$  is indistinguishable from the dissociation constant of the dimer ( $K_D$ ), since both processes are coupled at equilibrium, with a  $\sim 10.0$  kcal mol<sup>-1</sup> free energy of unfolding/dissociation. At pH 7.0, the  $\Delta\Delta G_{U/D}$  is increased by 3.0 kcal mol<sup>-1</sup> and the use of a phosphate buffer further increases the stability by 6.2 kcal mol<sup>-1</sup>. 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 noninvasive and often reversible (Silva et al, 2003). 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 (Foguel et al, 1998). 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 KD), indicating a similar pH stabilisation to that observed in urea unfolding experiments, most likely related to a yet unidentified histidine residue. Pressure dissociation of HPV16 E2C is accompanied by a volume change of 76 ml/mol corresponding to an estimated increase in exposed area upon dissociation of 2775 Å<sup>2</sup>, in agreement with the expected exposed surface from the crystal structures. Nevertheless, it is difficult to imagine how the halfbarrel interface will remain unaltered in a monomer, and the partial burial of tryptophans in the pressure induced monomer suggest that a native monomer is not viable. In addition, the stability of this species is marginal and it is converted into dimers immediately upon release of pressure (Foguel et al, 1998). Sodium chloride and phosphate show a substantial stabilisation of the domain to urea denaturation, of 5.0 and 4.5 kcal mol<sup>-1</sup>, respectively. Heparin and DNA showed a stabilisation too large to be measured accurately, which suggests that the complex, once formed is extremely stable and may require other proteins or degradation to be disassembled (Lima et al, 1997). The kinetic folding mechanism of HPV16 E2C reveals the formation of a monomeric intermediate species that precedes a concomitant dimerisation-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 ( $\Delta C_p$ ) 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 remain exposed to the solvent (Mok et al, 1996). 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  $\Delta G$  of unfolding of 3.5 kcal mol<sup>-1</sup>, compared to 11.0 kcal mol<sup>-1</sup> for the overall dimer unfolding transition, that represents 31% of the stability of the native dimer in identical buffer conditions (Prat-Gay et al, 2005). 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. It is tempting to suggest that the transient monomeric intermediate could be similar to a monomer, product of spontaneous dissociation of E2C in non-denaturing conditions. However, as the protein concentration increases there is a dimerisation process of the intermediate that leads to a parallel unimolecular folding route (Prat-Gay et al, 2005). The KD of this association is estimated between 1 and 5  $\mu$ M, much weaker than the minimal expected for a spontaneous association-dissociation in the absence of denaturants. The unfolding reaction consists of a major phase with a half-life of  $\sim$ 1 min, with secondary and tertiary (and thus, quaternary) structures disappearing in parallel, indicative of no intermediates being populated.



Using the  $k_{off}/k_{on}$  approximation the equilibrium dissociation constant obtained is 0.5 nM, identical to that obtained by pressure denaturation. 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 characterisation, mutagenesis or solvent modification. In addition, the bimolecular folding rate constant of HPV 16 E2C is  $105 \text{ M}^{-1} \text{ s}^{-1}$ , 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 (Waldburger et al 1996). 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 dimerisation 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  $\beta$ -sheet structure is more dependent on long range interactions than  $\alpha$ -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 (Prat-Gay et al, 2005). 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 dimerisation-folding step. This indicates equivalent pathways and that the unfolded state ensembles are similar in structure and/or energetically. NMR characterisation of the urea unfolded state ensemble (UFSE) revealed two regions with clusters of residual structures, at the DNA binding helix and in the second  $\beta$ -strand, although there appear to be no persistent long-range interactions (Mok et al 2000). 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  $20 \mu\text{M}$ , which supports the two parallel routes, at high and low protein concentration, respectively. This is because of the formation of a weak dimer product either by a weak association (KD) of the UFSE or the monomeric intermediate (see model in Figure 5).

Figure 5

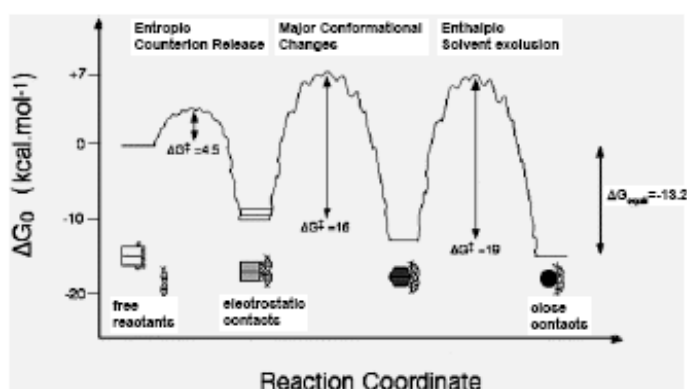


## Detailed mechanism of DNA site recognition in solution

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 ( Alexander et al 1996). 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. 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) . 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 noncognate 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, 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 (Ferreiro et al 2003) . 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<sup>-1</sup> s<sup>-1</sup>), which completely lacks sequence specificity and is weakly affected by a single charge mutation at the DNA binding helix. This step is followed by a conformational first-order rearrangement ( $k_2$   $8.0$  s<sup>-1</sup>), and finally, a slow solvent exclusion event ( $k_3$   $0.04$  s<sup>-1</sup>) 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 favourable conformational state of the interacting partners. Presumably this means a protein conformation favourable 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 multistep 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 . 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 ( $\Delta C_p$ ) 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. Initial studies on the kinetic binding of HV6 E2C suggest minor rearrangements upon DNA binding, similar on rates and faster off-rates (Figure 6) (Hooley et al 2006). This is consistent with the proposed lack of conformational freedom in the  $\beta 2$ - $\beta 3$  loop of the HPV 6 E2C protein. The DNA binding and dimerisation domain of the EBNA1 protein, where the only connection to HPV E2C is their dimeric  $\beta$ -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 ( Oddo et al 2006).

**Figure 6**



## **HPV 16 E2C and 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  $\beta$ -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  $\mu$ M, and probably lower extending the elapsed time) and at room temperature, lead to the formation of an oligomeric species with increased  $\beta$ -sheet content, and an expanded nature (Wetzler et al, 2007). The oligomer is slowly transformed into insoluble short “curly type” amyloid fibres as visualised 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  $\beta$ -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  $\beta$ -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 .

## **The single-chain E2 variant**

The E2C binding to its target DNA takes place in its native, dimeric state. Thus, stability and function of this domain are coupled, as in many transcription factors (Beckett et al 2001). Several groups have engineered monomerized variants of dimeric transcription factors in order to decouple stability and DNA binding and, in most cases, was demonstrated that the stability at physiological protein concentrations was improved, while retaining wild type binding affinity (Jana et al, 1998; Moran et al, 1999; Liang et al, 1993; Robinson et al, 1996; Sieber et al, 1998). In order to better understand and characterize the folding and DNA binding mechanisms of the E2 protein model domain, three single-chain variants of HPV16 scE2C were constructed. with engineered linkers of 6, 9, and 12 residues, respectively (Dellarolle et al, 2008). The length of the linker was constrained by the need to join approximately 17 Å, the distance between the N-

and C-termini of two different subunits in the E2C dimer. A glycine-rich linker sequences (GGGTGGGSGGGS, SGGSGGTGG, and GGTGGS) were chosen in an attempt to maintain maximum flexibility and reasonable solubility, on the basis of previous studies on monomerized transcription factors (Jana et al, 1998; Moran et al, 1999; Liang et al, 1993; Robinson et al, 1996; Sieber et al, 1998). The scE2C-12 variant showed the best expression level and purity and so was chosen as monomeric model of native of E2C. Preliminary spectroscopic investigation on this single chain variants of E2C suggest that monomerization of dimeric transcription factors using flexible linkers does not lead to substantial changes in the structure. The affinity of scE2C-12 for binding to its target DNA increases upon monomerization evenly along the tested phosphate concentration range, while the affinity for nonspecific DNA decreases to a larger extent, resulting in a 10-fold increase in specificity. Specific binding of E2C to DNA induces changes in the stability and dynamics of the protein, and nonspecific binding is known to induce larger changes, at least in one protein. We interpret that monomerization of E2C restricts the structural and dynamic changes that take place upon nonspecific DNA binding but favors the changes that take place upon specific DNA binding. Because the native E2C is known to be highly dynamic in solution as seen in molecular dynamic simulations (Falconi et al, 2007), NMR spectroscopy (Cicero et al., 2006), and binding to DNA (Hegde et al, 2002; Cicero et al, 2006) an ulterior purpose of this work is the characterization of the solution structure and dynamics of the scE2C-12 variant in its free form and in the complex with its target site.

**NUCLEAR MAGNETIC RESONANCE IN STRUCTURAL  
BIOLOGY**

NMR spectroscopy application in structural biology is today a field of great interest in basic research and pharmaceutical applications. Since the first experimental observation of a protein NMR spectrum was achieved in 1957), tremendous technological developments and fundamental methodological advances occurred, both in spectrometer development, sample preparation, specific pulse sequence and computational analysis. Proteins are the molecules for which the greatest effort has been spent in biomolecular NMR, followed by nucleic acids and their complexes with binding proteins (oligosaccharides conformational characterization and membrane structures are nowadays promising subjects of investigation). With respect to X-Ray crystallography NMR spectroscopy presents advantages and drawbacks. The major advantage is the study of the system in solution (and in membranes), that is in an environment much more similar to the native state of biological molecules. The major drawback is the limit in molecular size imposed to structural investigation by NMR. Transverse magnetization constitutes the observable signal in NMR experiments. The profound effect of the molecular dimensions on the transverse magnetization relaxation rates precludes the high-resolution determination of macromolecular systems larger than 25-30 kD, although recent advances have brought further this frontier (Kay et al, 1997). Moreover NMR is very well suitable for rapid study of interactions of small and intermediate biologically active molecules with proteins and membranes and is the unique technique capable to obtain exhaustive information on the dynamics properties at the atomic level.

### **The assignment problem**

The assignment is the first, necessary phase that constitutes a prerequisite in every NMR structure study. It consists in labeling the atomic sites of the molecule by the corresponding resonance frequency and in the successive extraction of structural information by quantifying the interatomic magnetic interactions. A protein contains a remarkable number of protons with similar magnetic characteristics, that is protons situated in residues of the same type present in multiple copy within the sequence. The traditional assignment method, developed by Wuthrich and coworkers, is based on homonuclear two-dimensional spectroscopy (Wuthrich, 1986; Roberts, 1993). The first step in this assignment strategy is to identify proton spins that belong to a particular amino acid using through-bond correlation spectroscopy (prevalently COSY and TOCSY experiments). This result is achieved by correlating the J-coupling network of aliphatic side chain protons to the respective amide proton which normally display greater C.S. (chemical shift) dispersion. The resonance pattern is then examined in comparison with the residue-dependent chemical shifts. The identification of the residue location within the sequence



is accomplished by analyzing NOESY spectra to individuate inter-residue cross-peaks generated by amide protons belonging to sequential adjacent amino acids. When the molecular size increases, the spectra appear too crowded of signals. Moreover, homonuclear J-scalar correlation experiments for proton spins fails for systems with large molecular dimensions. The time needed to build-up the scalar correlation is not adequate because of the large relaxation rate of proton spins that dramatically reduces the efficiency of magnetization transfer. Resolution limits prevent the possibility to carry out a correct and unambiguous assignment. The heteronuclear one-bond couplings,  $^1J_{CH}$  (125-160 Hz),  $^1J_{CN}$  (12 Hz) and  $^1J_{NH}$  (about 92 Hz), are sufficiently large and relatively uniform, depending weakly on conformation. On the other hand, transverse relaxation rates for heteronuclei in the globular proteins below the 25-30 kDa size still permit to transfer magnetization from  $^{15}N$  and  $^{13}C$  nuclei to scalar-coupled nuclei with high efficiency. Thus, it's possible and convenient to resolve very close proton resonances using an additional heteronuclear dimension which doesn't suffer in serious relaxation limits. In addition, from the late 1980 sensitive three-dimensional experiments were developed specifically for uniformly or fractionally  $^{13}C$ - and  $^{15}N$ -labeled proteins (Bax et al., 1993). Three-dimensional NMR spectroscopy is conceptually identical to two-dimensional one and makes use of the same pulse schemes, that are concatenated in such a way to extend the dimensionality of the experiment. The success of 3D and 4D spectra rapidly led to revolutionate the assignment procedure for proteins. In the traditional two-dimensional approach the critical step in the assignment is the individuation of NOE correlations between sequential residues. Errors in assignment of NOE connectivities at this stage propagate into the successive phase of the NMR structure determination. On the other hand, in a dipeptide unit of  $^{15}N$ - and  $^{13}C$ - labeled protein the  $^{13}C$  carbonyl atom of the first amino acid is scalar-coupled to the  $^{15}N$  amide atom of the second amino acid. This one-bond and other two-bond correlations between two sequential residues offer the opportunity to establish the connectivity avoiding the use of NOE crosspeaks, which could introduce ambiguities in the assignment process. Transverse magnetization is generated for the proton nucleus, then is transferred to  $^{15}N$  nucleus. In this way only magnetization of protons that are correlated to nitrogen atoms survives at the end of the pulse scheme. Successively, magnetization can be transferred to  $^{13}C$  nucleus of the same residue (residue i) or to the  $^{13}C$  nucleus of the preceding residue (residue i-1) because scalar interaction magnitude for the two nuclei is similar (11 and 7 Hz). The uniqueness of amino acid identification is achieved by comparison of this sequential connectivity with the chemical shift values that are characteristic of amino acid type. Once the resonances are assigned to the corresponding amino acid type, these sequential stretches of chemical shifts are mapped onto the protein sequence to find the aminoacidic segment that matches those chemical shifts. Side chain spins are successively correlated to backbone spins of their specific residue by  $^{15}N$ -edited and  $^{13}C$ -edited correlation spectra (HBHANH, HBHA(CO)NH, COSY, TOCSY, etc.). The

combined use of three-dimensional and, if needed, four-dimensional heteronuclear experiments allows to obtain the sequential connectivity and to internally verify the assignment relying only on J-scalar pathways (Ikura et al., 1990; Bax and Grzesiek, 1993). Advantages arise from the heteronuclear multidimensional techniques also in the assignment process of nucleic acids. A difference exists with respect to proteins: there is no way to date to connect by scalar couplings sequential nucleotide residues. Anyway, isotopic enrichment can be useful to connect the base spin system to the sugar spin system and resolve such ambiguities in the assignment.

### NMR determination of histidine protonation state

Histidine is frequently involved in the function of proteins, mainly because of the chemical versatility of its imidazole ring, which includes protonated and deprotonated forms as well as the tautomeric states. To determine the protonated and deprotonated forms and the tautomeric states of histidine residues in solution, NMR spectroscopy has been extensively used with good success. Traditionally, the  $pK_a$  of histidines has been determined by measuring chemical shift changes during pH titration experiments. The protonated or low pH form of the imidazole side chain **1** exists according to acid-base equilibrium described by constant  $K_a$  with two unprotonated species. These unprotonated or high pH forms of imidazole, the more common  $N^{\delta\epsilon}$ -<sup>2</sup>H tautomer **2** and the rare  $N^{\delta\epsilon}$ -<sup>1</sup>H tautomer **3** exist in equilibrium described by constant  $K_T$ . NMR, particularly <sup>15</sup>N NMR, has made important contributions to the elucidation of the chemical behavior of histidine in  $\alpha$ -lytic protease, a member of the trypsin superfamily (Bachovchin et al., 2001). <sup>15</sup>N NMR is exquisitely sensitive to the local magnetic environment, and thus capable of revealing the presence of H-bonding on both sides of the imidazole ring as well as  $pK_a$  values and tautomeric states. <sup>1</sup>H NMR, when correctly applied, can also reveal the presence of histidine C <sup>$\delta\epsilon$</sup> -H-donated H-bonds in serine protease active sites, as we first presented in 1996 (E.L. Ash, M.P. Vincent, J.L. Sudmeier, and W.W. Bachovchin, unpubl.; Ash et al. 2000), in addition to the often reported determination of histidine  $pK_a$  values in proteins by <sup>1</sup>H NMR. If the ionization of a monobasic acid HA is rapid on the NMR time scale, the observable chemical shift  $\delta_{obs}$  of an adjacent nucleus is averaged over the individual chemical shifts of the species HA ( $\delta_{HA}$ ) and A<sup>-</sup> ( $\delta_{A^-}$ ), the weighting coefficients ( $x_{HA}$  and  $x_{A^-}$ ) being the pH dependent molar fractions:

$$\delta_{obs} = x_{HA}\delta_{HA} + x_{A^-}\delta_{A^-} = \frac{\delta_{A^-} + 10^{pK - pH}\delta_{HA}}{1 + 10^{pK - pH}}$$

By measuring  $\delta_{obs}$  for samples of varying pH, acidity constant can be obtained as follows:

$$pK = pH + \log \frac{\delta_{A^-} - \delta_{obs}}{\delta_{obs} - \delta_{HA}}$$

However, this method cannot be applied to proteins that are unstable over a wide pH range, and it does not provide information about the tautomeric states. To circumvent this problem, several NMR techniques have been developed, based on HMQC (Pelton et al , 1993; Van Dijk et al 1992 ) HMBC (Bax et al, 1986; Schmidt et al 1991), and HSMQC (Zuiderweg, et al 1990; Xia et al 1995) experiments, which discriminate each state from the ratio of two- and three-bond remote couplings, such as  $^2J_{N\delta_2-H\delta_2}$  and  $^3J_{N\delta_1-H\delta_2}$  in the imidazole ring. Unfortunately, the small value of these couplings necessitates long coherence transfer times, making this method not applicable to large proteins. The method can, however, be extended to proteins of higher molecular weight if larger one-bond couplings centered at the  $C^{\epsilon_1}$  and  $C^{\delta_2}$  carbons in the imidazole ring are used .

. However, so far the C-N coupling constants of the histidine residues have only been used for a qualitative analysis to identify the state with the highest population. Today NMR titration has become a routine means of pK determination , even for individual groups of large biopolymers combining the advantages of the potentiometric and NMR spectroscopic approach over potentiometry alone.

### Nuclear spin relaxation and protein dynamics

It's well known that proteins are not static systems in solution but undergo extensive fluctuations and motions on a broad range of time scales. Distinct NMR techniques exist that can report on these motional properties. Each technique displays characteristic sensitivity to a particular range of motional frequencies. Nuclear spin relaxation rates can accurately describe the global protein mobility due to the overall molecular tumbling and can individuate local flexibility on the picoseconds time scale. In addition, relaxation data analysis allows characterizing slower conformational motions such as exchange processes. A great advantage of NMR studies of protein mobility is its high-resolution nature that is the ability to examine the local motions with site-resolved specificity. After a spin system is perturbed by one or more radiofrequency pulses (spectroscopists say that it is excited), it relaxes to the original equilibrium state following different pathways that depend on the particular excited state, on the magnetic environment surrounding the nuclei and on the molecular geometry. In this way, it's possible to extract structural and dynamical information by appropriate perturbation of the equilibrium state and by monitoring the rate at which the spin system recovers to the original state. For the relaxation of diamagnetic protein spins, two mechanisms of interaction with the

environment determine the relaxation rate: the dipole-dipole interaction and the chemical shift anisotropy (C.S.A.) interaction. These interactions are generated by the sources of local magnetic fields in the environment surrounding the nuclei that is the nucleus itself and the electron density distribution. These local magnetic fields are modulated by the global and specific movements of the protein. Relaxation requires that the energy of these two interaction terms fluctuates with frequencies corresponding to the transition frequencies between the fundamental and excited nuclear states. The frequency distribution of dipolar and C.S.A. interactions span a broad range and are described by the spectral density function  $J(\omega)$  which measures the magnitude of the interaction at each particular frequency. The spectral density function derives from the autocorrelation (for simplicity we exclude cross-correlation terms) of the Hamiltonian  $H$  relevant for the magnetic interaction (dipole-dipole and C.S.A.):

$$J(\omega) = \int_0^{\infty} \cos(\omega t) H(0)H(t) dt$$

In this way the spectral density function represents the decomposition of the power of interaction energies with respect to the frequencies. The relaxation rates are directly related to the specific spectral density values at the frequencies corresponding to the transitions between nuclear states. Relaxation experiments for the study of protein dynamics usually measure the relaxation times of heteronuclei ( $^{13}\text{C}$  and  $^{15}\text{N}$ ) because in this case the predominant contribution is given by the dipolar interaction with the attached proton and other contributions are negligible or can be taken into account in a simple manner. The first approach to calculate the spectral density and hence the relaxation times  $T_1$ ,  $T_2$  and the heteronuclear NOE of  $^{13}\text{C}$  and  $^{15}\text{N}$  nuclei rely on the model-free formalism of Lipari and Szabo (1982) which provides the analysis of protein mobility in terms of the minimum number of motional parameters. These authors proposed that the dynamics properties are determined by two terms: the overall motion of the macromolecule as a whole and the internal fast motions of the internuclear vectors (N-H or C-H vectors). The method is termed free because no specific model for the internal motions is required. In fact, the autocorrelation function is considered simply as an exponentially decaying function of time characterized by a specific time constant. The model-free formalism is based on the assumption that global and internal fast motions, which display very different rates, are independent. As a consequence, the autocorrelation function can be decomposed into a product of two single autocorrelation functions (two exponentially decaying time functions with distinct time constants), one for the global motions and the other for the local motions. From the equation relating the autocorrelation function to the spectral density it mathematically derives that the spectral density functions can be decomposed as the sum of two single spectral density functions that respond to distinct time constants. The nuclear transitions involved are typically on the order of 30-200 MHz for heteronuclei (it depends on the nucleus magnetogyric ratio and

on the magnetic static field magnitude). Therefore, the primary source of relaxation is the global molecular tumbling for proteins, because the tumbling rates of these macromolecules adopt this range of values and match the nuclear transition frequencies. As a consequence, the weight of the internal motion density function is limited except when large amplitude fast internal motions characterize the molecular site where the heteronucleus is located. In this case the nucleus experiences an effective motion that is much faster than the overall tumbling and behaves as if the molecule was quite smaller and mobile. S values typically found for structured regions in global proteins (such as  $\alpha$ -helices and  $\beta$ -strands) are comprised in the range 0.9-0.95.

Explicit equations relate the relaxation rates for  $^{15}\text{N}$  and  $^{13}\text{C}$  nuclei to both these separated density functions (Kay et al., 1990). The best-fit of calculated to experimental relaxation data provide the important parameters that describe the molecular motions in the model-free formalism, that is the global tumbling frequency, the weight of internal fast motions and the rate of these internal motions. Additional parameters were added in successive models, the most important of which is the quantity  $R_{\text{ex}}$  that accounts for the contribution to the relaxation rates given by slower exchange processes (in the  $\mu\text{s}$ -ms range). The term  $R_{\text{ex}}$  measures the effect on the transverse  $T_2$  relaxation time of chemical and conformational exchange. The continuous exchange of a nucleus between two resonance frequencies (two conformers or two distinct sites that experiment a different chemical shifts) accelerates the defocusing of the coherent transverse magnetization (which is the observable signal in NMR). Also this parameter is extracted by the regression analysis of experimental and calculated data. The treatment above described for relaxation data is valid for molecules possessing an isotropic rotational diffusion tensor. Anisotropy causes large deviations of the relaxation rates with respect to the values predicted by the isotropic model (Pawley et al., 2001). Anisotropic effects can be simply taken into account calculating the components of the diffusion tensor. The result in introducing the anisotropy contribution is a differentiation of the global molecular tumbling time constant into distinct "directional" tumbling time constants which describe the average rate of rotational motion about the axes of the anisotropic diffusion tensor. Consequently, a different set of equations relating  $T_1$ ,  $T_2$  and heteronuclear NOE to the motional parameters are employed in the best-fitting procedure.

## **MATERIALS AND METHODS**

### **Peptide Synthesis and Purification.**

The  $\alpha$ 1-E2C peptide contains residues 289-307 of HPV16 E2 (LKGDANTLKCLRYRFKKHA). The last residue in the sequence was mutated from Cys to Ala to avoid redox side reactions. The peptide was synthesized at the Yale University Keck facility, with N-terminus acetylated and C-terminus amidated, and purified by reverse phase HPLC. We confirmed its molecular mass by mass spectrometry. Quantification of peptide concentration was carried out by absorbance at 274 nm and 220 nm.

### **DNA Synthesis**

Fluorescein-labeled single-stranded oligonucleotides containing the complete target site 35 of the HPV16 genome (5' GTAACCGAAATCGGTTGA 3'), a hemisite (5' GTAGCACAAATCGGTTGA 3') and a randomized site (5' ACATGGACCTGTCAAAGTA 3') were purchased, HPLC purified, from Integrated DNA technologies (Coralville, IA). Double-stranded oligonucleotides were annealed by mixing equal amounts of each strand in 10 mM Bis-Tris-HCl buffer, pH 7 and 50 mM NaCl, incubating 5 minutes at 95 °C, and slowly cooling to 25 °C for 16 hours. Completeness was checked by native polyacrylamide gel electrophoresis.

### **NMR Experiments.**

$\alpha$ 1-E2C (3 mM) was dissolved in aqueous buffer (20 mM phosphate buffer pH 6.5, 2 mM DTT, 0.01% NaN<sub>3</sub> and 5% D<sub>2</sub>O), TFE/water mixture (40% TFE-d<sub>3</sub>, 60% 20 mM phosphate pH 6.5, 2 mM DTT and 0.01% NaN<sub>3</sub>) and urea/water mixture (8M urea, 20 mM phosphate pH 6.5, 2 mM DTT, 0.01% NaN<sub>3</sub> and 5% D<sub>2</sub>O). We used urea instead of GdmCl for NMR experiments because it was not possible to tune the probe in presence of high concentrations of GdmCl. The NMR experiments were performed at 25 °C or 5 °C using a Bruker Avance 700 and Avance 400 spectrometers equipped with triple resonance probes incorporating self-shielded gradient coils. All the heteronuclear correlation experiments were carried out at natural abundance. Pulsed field gradients were appropriately employed to achieve suppression of the solvent signal and spectral artifacts. Quadrature detection in the indirectly detected dimensions was obtained using the States-TPPI method or the echo-antiecho method. The spectra were processed on Silicon Graphics workstations by the NMRPipe software and analyzed using NMR View.

Assignment of backbone resonances of  $\alpha$ 1-E2C was performed using a combination of the following 2D NMR experiments: TOCSY (mixing times of 50 and 80 ms), [ $^1\text{H}$ - $^{13}\text{C}$ ] and [ $^1\text{H}$ - $^{15}\text{N}$ ] HMQC, and [ $^1\text{H}$ - $^{13}\text{C}$ ] HMQC-TOCSY (mixing times of 40 and 80 ms) the latter mentioned being very helpful due to the overlap of the signals. The sequential connectivity across the peptide was established using the NOESY and ROESY spectra (mixing times of 0.15 s and 0.20 s), according to the sequential assignment method. The  $^1J_{\text{C}\alpha\text{H}\alpha}$  coupling constants were measured from the in-phase splitting patterns of the  $\text{C}_\alpha\text{-H}_\alpha$  cross-peaks in the HMQC spectra without carbon decoupling. The  $^3J_{\text{HNH}\alpha}$  values were measured from the in-phase splitting patterns of the NH-C cross peaks in the [ $^1\text{H}$ - $^{13}\text{C}$ ] HSQC-TOCSY spectrum collected with 8k complex data points in F2. Uniformly  $^{15}\text{N}$ -labeled HPV-16 E2 DNA-binding domain was expressed and purified as previously described. The protein concentration was 0.4 mM in 50 mM sodium phosphate, 5 mM DTT, pH 6.5. NMR experiments of protein samples were performed at 30 °C on a Bruker Avance400. A series of two-dimensional  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra were recorded at distinct pH values, namely pH 5.0, 5.5, 6.0, 6.5, 7.0, 7.5, 8.0, 8.5, 9.0. At pH higher than 9.0 and lower than 5.0, irreversible protein precipitation occurred. A total of 4096 complex points in  $t_2$  with 256  $t_1$  increments were acquired. The spectral widths were set to 6410 Hz in  $F_2$  and 1120 Hz in  $F_1$ , placing the carrier frequency at 4.70 ppm in  $F_2$  and at 118.6 ppm in  $F_1$ . At every pH a two-dimensional  $^1\text{H}$ - $^{15}\text{N}$  long range HSQC (LR-HSQC) spectrum was acquired to observe two-bond correlations between nitrogen and proton resonances belonging the histidine side-chain spin system. The delay for the INEPT-type magnetization transfer in the LR-HSQC was set to 11 ms. A total of 2048 complex points in  $t_2$  with 64  $t_1$  increments were acquired. Spectral width was 6410 Hz in  $F_2$  and 4258 Hz in  $F_1$ , placing the carrier frequency at 4.70 ppm in the proton dimension and at 208.4 ppm in the nitrogen dimension. The  $\text{pK}_a$  values of titration curves were determined by analyzing the pH dependencies of amide N and  $\text{H}^{\text{N}}$  resonances in the HSQC spectra for backbone signals and of  $\text{N}^{\square 1}/\text{N}^{\square 2}$  and  $\text{H}^{\square 2}/\text{H}^{\square 1}$  resonances in the LR-HSQC spectra for histidine side chains. The pH titration curves were fitted to a modified Henderson-Hasselbalch equation by nonlinear least-squares analysis:

$$\delta_{\text{obs}} = \frac{\delta_{\text{pr}} + \delta_{\text{nonpr}} \cdot (pH - pK_a)}{1 + 10^{(pH - pK_a)}} \quad [6]$$

in which  $\delta_{\text{obs}}$  is the chemical shift observed at each pH value and  $\delta_{\text{pr}}$  and  $\delta_{\text{nonpr}}$  are the chemical shifts for the protonated and deprotonated histidines, respectively. Curve fits were performed using the KaleidaGraph3.5 Software (Synergy Software).



### **E2C protein preparation.**

Uniformly  $^{15}\text{N}$ -labeled and  $^{15}\text{N}$ - $^{13}\text{C}$ -labeled proteins were expressed using the previously described plasmid ptz18U-E2 (Mok et al. 1996a), which contains a gene fragment encoding the C-Terminal 80 residues of the HPV-16 E2 plus a methionine residue in the first position as result of the cloning operation. This expression vector was transformed into *Escherichia coli* JM109 strain cells. Overnight cultures of cells in M9 minimal medium were inoculated (1%) into M9 medium containing  $^{15}\text{NH}_4\text{Cl}$  and  $^{13}\text{C}$  glucose as the sole nitrogen and carbon sources. The cells were grown at 37 °C up to 0.35 O.D.<sub>600</sub> units. Protein expression was induced by adding phage M13/T7 (Invitrogen, San Diego, CA) at a multiplicity of infection of 5 and 0.3 mM IPTG (isopropyl- $\beta$ -thio-D-galactoside). The cells were grown overnight at 37 °C before harvesting. After overnight incubation, cells were spun down and sonicated twice at 0°C in extraction buffer (100 mM Tris-HCl, pH 6.8, 0.6 M NaCl, 5 mM 2-mercaptoethanol, 1 mM EDTA). The supernatant was subjected to 80% ammonium sulfate precipitation, and the precipitate was resuspended and dialyzed in a medium of 50 mM Tris-HCl, pH 6.8, 0.6 M NaCl, and 5 mM 2-mercaptoethanol. A two-step chromatographic procedure was efficient to yield high purity protein as following. The dialyzate was loaded onto a Heparin Hyper D affinity column (BioSeptra, Villeneuve la Garenne, France) equilibrated with 50 mM Tris, pH 8.0, 0.6 M NaCl, 5 mM 2-mercaptoethanol, and 1 mM EDTA. The column was washed with five column volumes of buffer, and the bound protein was eluted with a 0.6-2.0 M NaCl gradient. Fractions that were >90% pure were pooled and dialyzed against 50 mM sodium acetate buffer, pH 5.6, 0.2 M NaCl, and 5 mM 2-mercaptoethanol before loading onto a preparative Superdex G75 gel filtration column (Pharmacia Biotech, Uppsala, Sweden). The main peak retention time is consistent with a globular dimeric protein possessing the calculated molecular mass of E2C. The fractions were collected, dialyzed against 50 mM sodium acetate buffer, pH 5.6, with 200 mM NaCl and 5 mM 2-mercaptoethanol, and then concentrated using a YM10 Centriprep concentrator (Amicon, Bedford, MA). This procedure yielded around 15 mg/L and 8 mg/L of >98% pure  $^{15}\text{N}$  - and  $^{15}\text{N}$  -  $^{13}\text{C}$  E2C, respectively. Protein concentration was determined using an extinction coefficient of 41,920  $\text{M}^{-1} \text{cm}^{-1}$  for the dimer (Mok et al., 1996). The protein was aliquoted and frozen at - 80 ° C for conservation.

### **Construction, Expression, and Purification of Single Chain E2C**

The starting point for the construction was the previously described E2C ptzU18-based expression vector . In this plasmid, the open reading frame of E2C is flanked at the 5' end by a restriction site for *EcoRI*, a ribosome binding site and a restriction site for *ClaI* at the ATG

codon, and at the 3' end by a restriction site for *Bam*HI downstream of the Stop codon. We constructed a monomeric version of E2C by inserting a sequence coding for E2C, followed in tandem by a sequence coding for a linker at the restriction site for *Cl*aI. The inserted sequence thus precedes the original ORF for E2C. We synthesized the insert by a PCR reaction from the *Eco*RI to the *Bam*HI restriction sites of ptzU18-E2C with the following primers: N-term AGGGAATTCA AAGAGGAGAA ATTACATATG ACACCCATAG TA, C-term scE2C-6 CATCGATCCA CCGGTACCGC CTATAGACAT AAATCCAGT, C-term E2Csc-9 CATCGATCCG CCTGAGCCAC CGGTACCGCC TATAGACATA AATCCAGT, C-term scE2C-12 CATCGATCCG CCACCTGAAC CGCCACCGGT ACCGCCTCCT ATAGACATAA ATCCAGT. The PCR product yields an E2C ORF without the stop codon and with a linker sequence and a restriction site for *Cl*aI at the 3' end. At the 5' end we mutated the original *Cl*aI restriction site to an *Nde*I restriction site. The PCR product was then cloned in tandem with the original gene in the same T7 expression vector ptzU18, between the *Eco*RI and *Cl*aI sites, generating the E2C-linker-E2C gene. Finally, the ligation product was sequenced, stored, and transformed into the B121(DE3) plys *E. coli* expression strain. Finally scE2C-12 were recombinantly expressed, purified, and stored as previously described for E2C.

## **RESULTS**

### Determination of $\alpha$ -helix and PII structures in $\alpha$ 1-e2c by NMR

We investigated the solution structure of  $\alpha$ 1-E2C by NMR spectroscopy, which provides information about the conformation of individual residues in the sequence. We analyzed the conformation of the peptide in aqueous solution, in a water-TFE mixture and in an 8 M urea solution. Figure 7 shows the overlapping of the region of crosspeaks in the [ $^{13}\text{C}$ ,  $^1\text{H}$ ]-HMQC spectrum for  $\alpha$ 1-E2C in the three solvent conditions. The analysis was carried out using coupling constants and chemical shifts in addition to the usual NOE observations, to minimize biased conclusions that can be obtained in highly flexible systems. The analysis of the secondary chemical shifts allows determining the elements of secondary structure stabilized in the peptide in the different environments here studied. Figure 4A shows the chemical shift index (CSI) of the  $\text{H}_\alpha$ ,  $\text{C}_\alpha$ , and  $\text{C}_\beta$  resonances of the  $\alpha$ 1-E2C peptide in the aqueous buffer and in the water-TFE and water-urea solutions. CSI values of -1, +1 and -1 for  $\text{H}_\alpha$ ,  $\text{C}_\alpha$ , and  $\text{C}_\beta$ , respectively, are indicative of  $\alpha$ -helical conformations (depicted as upward red arrows in the figure), while the opposite values are observed for  $\beta$ -strands (downward blue arrows in the figure). In contrast, 0 indexes (gray bars in Figure 8) suggest that no secondary structure elements are stabilized for the corresponding residues. In the TFE solution the  $\alpha$ -helical conformation spans from A293 to H306 as evidenced from the CSI analysis (Figure 4A). This result is consistent with the NOESY pattern, where  $\alpha$ -helix characteristic crosspeaks are observed for this region (data not shown).  $^1\text{J}_{\text{C}\alpha\text{H}\alpha}$  values larger than random coil values ( Wishart et al, 1994) and  $^3\text{J}_{\text{H}\text{N}\text{H}\alpha}$  smaller than random coil values, in general smaller than 5 Hz, are typical for  $\alpha$ -helices ( Vuister et al 1992 ). The values of  $^1\text{J}_{\text{C}\alpha\text{H}\alpha}$  and  $^3\text{J}_{\text{H}\text{N}\text{H}\alpha}$  coupling constants for this tract are indeed indicative of an  $\alpha$ -helix (Figure 8B, black,). In order to investigate regions involving PII conformation in  $\alpha$ 1-E2C, we studied its behavior in a solution containing 8 M urea. As expected for high denaturant concentrations, the chemical shifts appear mainly clustered according to amino acid type (Figure 7). Correspondingly, most of the CSI observed are 0, except for the C-terminal region of the peptide (Figure 8A). This region displays some  $\text{C}_\alpha$ , and  $\text{C}_\beta$  indexes compatible with a  $\beta$ -strand conformation, while the  $\text{H}_\alpha$  indexes are zero. Overall, from the CSI analysis, in the urea-water mixture the peptide seems to adopt a disordered structure with the stabilization of an extended conformation in the C-terminal region. The PII is an extended left-handed helix, often lacking characteristic proton or carbon chemical shift deviations from random coil values as found in  $\alpha$ -helix and  $\beta$ -sheet conformations (Shi, et al, 2002; Lam et al, 2003 ).

FIGURE 7

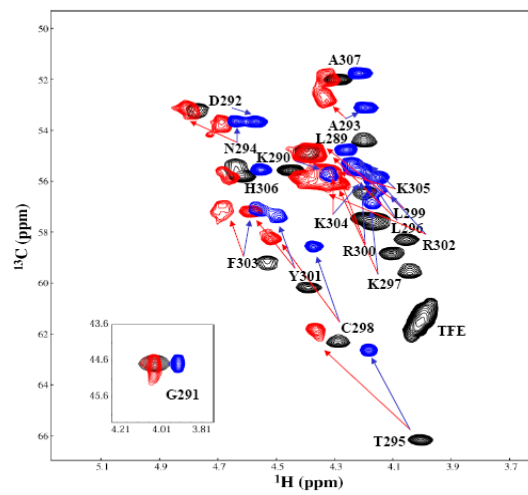
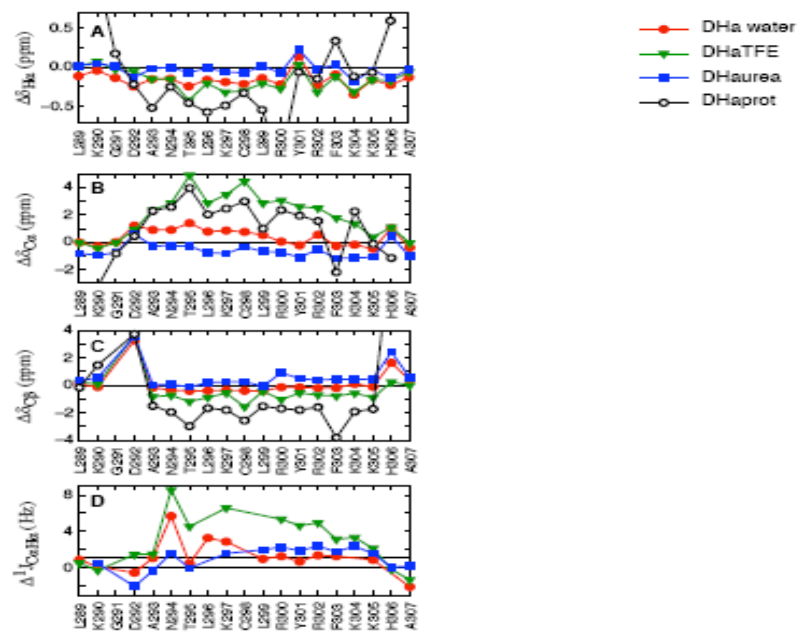


FIGURE 8



This absence of chemical shift dispersion leads to a strong overlap of signals in the NOESY spectrum and precludes an accurate determination of PII structure solely from chemical shifts and from the pattern of NOESY crosspeaks. As a consequence, PII is often misclassified as random coil conformation by NMR techniques (Lam et al 2003; Bachovchin et al, 1986) . Although the C-terminal region of the peptide displays some CSI coherent with a  $\beta$ -strand conformation, inspection of the  $^1J_{C\alpha H\alpha}$  values shows values larger than those observed in a random coil conformation for this tract (Figure 8B, red), while  $\beta$ -strands display the  $^1J_{C\alpha H\alpha}$  smaller than random coil ( Vuister et al, 1992). Likewise, the difference between measured  $^1J_{C\alpha H\alpha}$  and random coil values characteristic of PII helices is reported to be larger than 1.1 Hz<sup>60</sup>. In the case of  $\alpha$ 1-E2C in urea,  $^1J_{C\alpha H\alpha}$  larger than random coil values are observed systematically for residues from L299 to K305 (Figure 8B). The  $^1J_{C\alpha H\alpha}$  values of these residues are larger in the urea solution at 5°C than at 25°C. Since the large  $^1J_{C\alpha H\alpha}$  values increase at 5°C and are incompatible with a  $\beta$ -strand, we conclude that residues L299 to K305 adopt PII conformation.

### **Two distinct structural regions for $\alpha$ 1-E2C in aqueous solution**

Through NMR experiments we can assign conformations to specific residues of the peptide. From A293 to L299, the crosspeak pattern in the NOESY spectrum and the CSI show a tendency of the peptide to adopt  $\alpha$ -helical structure (Figure 8A). This tendency is weaker than in TFE, as judged from the magnitude of the chemical shifts and the very low intensity of the NH-NH (i, i+1) crosspeaks characteristic of  $\alpha$ -helices (data not shown).  $C_\beta$  indexes are 0 for this region in water, in contrast to -1 observed in TFE. In addition, the mean difference with random coil values is -0.18 and 0.9 ppm for  $H_\alpha$  and  $C_\alpha$  nuclei, respectively, for residues 293-299 in water, versus -0.25 and 3.4 ppm for the same nuclei in the TFE solution. From residue R300 to the C-terminal end of the peptide, the chemical shifts are more similar to random coil values. Interestingly, these residues have  $^1J_{C\alpha H\alpha}$  constants larger than those observed for random coil conformation (Figure 8B,). The  $^1J_{C\alpha H\alpha}$  coupling constants for this tract are even larger in the presence of urea (Figure 8B), a clear evidence that the PII conformation spans these residues. In aqueous solution,  $\alpha$ 1-E2C shows a dynamic behavior, displaying a conformational equilibrium between disordered conformations and a nascent PII and disordered conformations and an  $\alpha$ -helix structure in two well-differentiated regions. The PII region that goes from R300 to K305 corresponds very well with the PII detected in the presence of urea. The  $\alpha$ -helix, however, only spans from A293 to L299 compared to the 293-305 region detected in the presence of TFE.

### **Conformation of the DNA binding helix in $\alpha$ 1-E2C versus in the context the full-length E2C domain**

We examined whether the conformational space explored by  $\alpha$ 1-E2C resembles that of the DNA binding helix in the context of the full-length domain. The determined three dimensional structure of HPV16 E2C defined residues 292-305 as the DNA binding helix (Hegde et al, 1998), matching the region of the peptide that forms an  $\alpha$ -helix in the presence of TFE. In the light of combining the results in solution presented here, together with NMR data indicating non-canonical structures in solution, we carefully re-examined the DNA binding helix of the crystal structure. The observed dihedral angles are compatible with a  $\alpha$ -helix only for the tract 293-303, whereas the region 304-309 displays characteristics of  $3_{10}$  helix. Figure 8C shows the CSI for  $H_{\alpha}$ ,  $C_{\alpha}$ ,  $C_{\beta}$  and CO nuclei for residues 289-307 in the protein ( Tiffany et al , 1972). Chemical shifts, the NOE pattern (not shown), and the  $^3J_{\text{HNH}}$  observed for this region of the protein are compatible with a very stable  $\alpha$ -helix limited to the region A293-R302 ( Nadra et al, 2004) Thus, in the N-terminal region, the tertiary structure of the domain can consolidate the local conformational tendencies of  $\alpha$ 1-E2C. The DNA binding helix kinks at residue F303, which presents unusual chemical shift values, and residues 303-305 show deviations from canonical  $\alpha$ -helix ( Nadra et al, 2004; Cicero et al 2006; Falconi et al, 2007) (Figure 8C). In the C-terminal region, the tertiary contacts do not completely override the mixed helix/PII local tendencies of the sequence, leading to a distorted and kinked DNA binding helix after F303. This confirms the reclassification of the DNA binding helix of the crystal structure as pure  $\alpha$ -helix for residues 293-303, and  $3_{10}$  helix for residues 304-309.

## The pH dependence of specific E2C-DNA binding

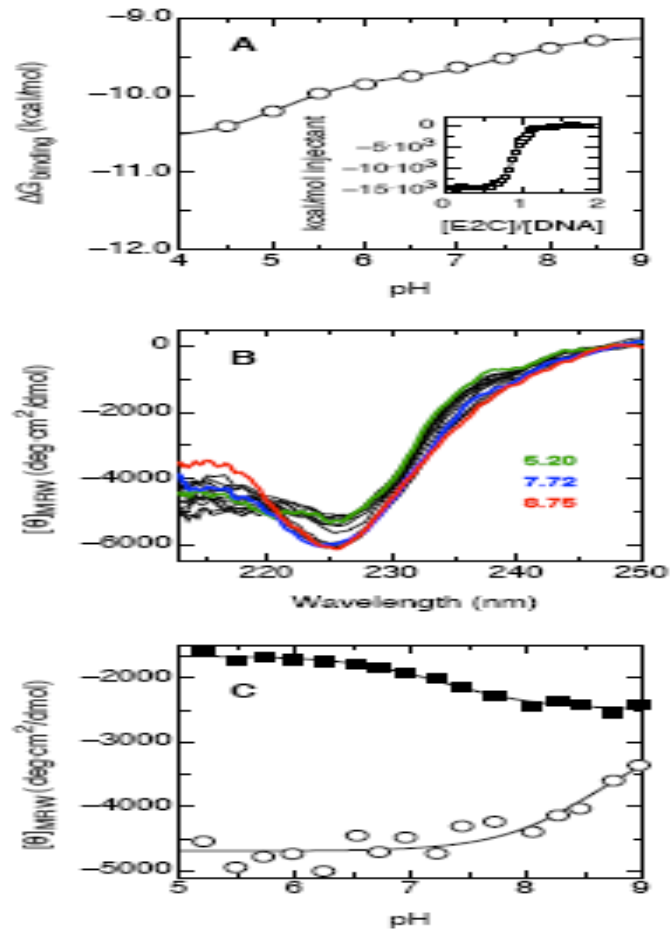
We have characterized binding of HPV16 E2C to a cognate double-stranded nucleotide (site35, sequence GTAACCGAAATCGGTTGA) using isothermal titration calorimetry (ITC) (Table 1 and Figure 1A). We carried out the experiments at 25 °C in a three-component buffer (100 mM Tris, 50 mM MES, 50 mM sodium acetate, 200 mM NaCl, 0.2 mM DTT) that ensures a constant ionic strength of 300 mM over the pH range studied. In the pH range between 4.5 and 8.5, the observed stoichiometry is 0.8 or higher and the ratio of the dissociation constant and the E2C concentration, is above 40 (Table 1), validating the fitting of the data for obtaining a dissociation constant for the E2C-DNA interaction. At higher or lower pH values E2C was not stable in solution, leading to the observation of abnormal stoichiometries (data not shown).

The free energy for specific E2C-DNA binding shows two transitions as a function of pH (Figure 1A), indicating that at least two functional groups change their pK-values upon binding<sup>46</sup>. Binding is stronger at lower pH for both transitions, showing that the pK of both groups is higher in the complex than in the free reagents<sup>46</sup>. We have fitted the data in Figure 1A to a model with two independently titrating residues in order to obtain the relevant pK-values<sup>46</sup>:

$$\Delta G_{binding}(pH) = \Delta G_{binding}(pH = 14) + 2 \cdot RT \cdot \ln \left( \frac{1 + 10^{pK_{free}^A - pH}}{1 + 10^{pK_{complex}^A - pH}} \right) + 2 \cdot RT \cdot \ln \left( \frac{1 + 10^{pK_{free}^B - pH}}{1 + 10^{pK_{complex}^B - pH}} \right) \quad [1]$$

where A and B are the two titrating groups and the factor 2 comes from the homodimeric nature of E2C. According to the fit, one of the titrating groups has a pK of  $4.95 \pm 0.08$  in the free state and  $5.24 \pm 0.07$  in the complex, while the other one has a pK of  $7.42 \pm 0.09$  in the free state and  $7.61 \pm 0.09$  in the complex. These pK-values are compatible with the ionization of two of the five histidine residues in E2C<sup>47</sup>.





## The pH dependence of E2C native state structure

The ITC data show a coupling between the ionization of two E2C residues and the free energy of specific E2C-DNA binding. This coupling may be due to titration of residues involved in direct protein-DNA contacts or through a modulation of the protein structure and/or dynamics caused by protonation. Circular dichroism spectroscopy (CD) in the far-UV region showed that heparin, sodium chloride and sodium phosphate can induce changes in the structure of E2C, indicative of a plastic native state ensemble. DNA binding induces similar changes, showing that conformational plasticity is coupled to binding. We used far-UV CD to follow the influence of protonation on the secondary structure of E2C at pH values between 5 and 9, in the three-component buffer used in the ITC experiments (Figure 1B). While E2C shows a far-UV CD spectrum typical for a fully native protein between pH 5 and 9, we could also observe two minor transitions in the spectrum. The region between 220 and 250 nm changes between pH 5 and 8, while the region between 213 and 220 nm changes between pH 7 and 9 (Figure 1B). We fitted the data at each wavelength using the equation for titration of a single chemical group:

$$\theta_{MRW} = \frac{\theta_{MRW}^{deprotonated} + \theta_{MRW}^{protonated} \cdot 10^{(pK-pH)}}{1 + 10^{(pK-pH)}} \quad [2]$$

All wavelengths within each of the abovementioned regions can be fit with the same pK (data not shown), suggesting that both changes in secondary structure are cooperative and coupled to single protonation events. Figure 1C shows titration data at 235 nm (squares), which can be fit with a pK value of  $7.3 \pm 0.1$ . This pK for changes in the secondary structure of the domain corresponds to one of the pK values for the pH dependence of DNA binding (Figure 1A). The data at 215 nm (circles) are an incomplete titration, from which we can only estimate a lower limit of 8.5 for the pK. This second pK value for changes in secondary structure does not have a counterpart pK value in the pH dependence of DNA binding (Figure 1A).

## Conformational plasticity of the DNA binding helix in the context of the full-length E2C domain

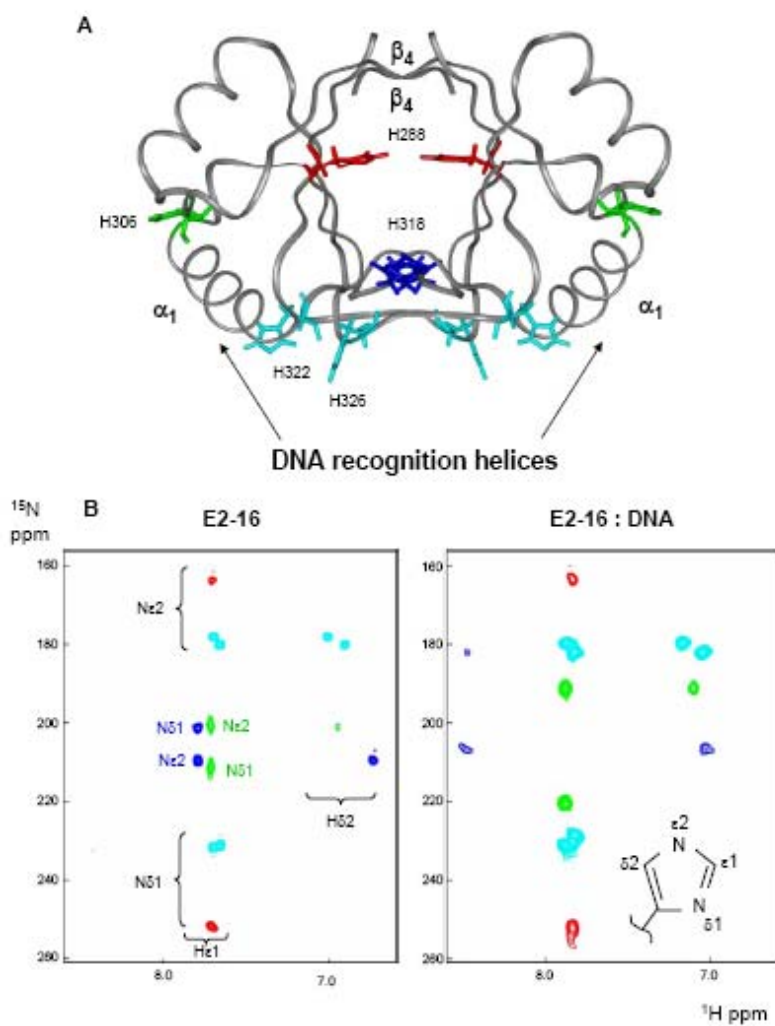
The protein contains five histidine residues per monomer, and the assignment of the  $^1\text{H}$  and  $^{15}\text{N}$  nuclei of the imidazolic ring for all histidines were already presented. The correlations between the  $\text{H}^{\square 1}$  and  $\text{H}^{\square 2}$  with the two nitrogens ( $\text{N}^{\square 2}$  and  $\text{N}^{\square 1}$ ) were sufficiently well resolved in the long-range  $^1\text{H}$ - $^{15}\text{N}$  HSQC for the five histidine residues and assigned using NOE contacts. By following their chemical shift as a function of pH we could estimate the apparent  $\text{pK}_a$  for each histidine. We have further explored the conformational plasticity of the C-terminal region of the DNA binding helix in full-length E2C by measuring the influence of H306 protonation on its conformation. The observed  $\text{pK}_a$  for H306 following the chemical shift of  $\text{H}^{\square 1}$  is  $5.65 \pm 0.05$  (a complete structural and dynamic analysis of the  $\text{pK}_a$  of all histidine side chains will be published elsewhere). The resulting  $\text{pK}_a$  is about 0.6 units lower than that expected for a

solvent exposed histidine, indicating that the protein tertiary structure stabilizes the neutral form of H306. The observed chemical shift for the two nitrogens in the non-protonated state (211.4 and 201.9 ppm for  $\text{N}^{\square 1}$  and  $\text{N}^{\square 2}$ , respectively), indicate that this histidine exist as a mixture of ~6:4 of the  $\text{N}^{\square 2}$ -H and the  $\text{N}^{\square 1}$ -H tautomers. Taking into account that for a solvent exposed histidine the expected population of the two tautomers is 4:1, the interaction of H306 with both the DNA binding helix and the hydrophobic core of the monomer stabilizes the  $\text{N}^{\square 1}$ -H tautomer. The interconversion between the two tautomers is fast, leading to a rapid movement of the hydrogen between the two nitrogens. In order to estimate the influence of the protonation of H306 on the conformation of the protein, the chemical shift perturbation of amide nitrogens was followed by recording a series of  $^1\text{H}$ - $^{15}\text{N}$  HSQC in the same range of pH (5.0-9.0). In this way, it was possible to detect those residues that titrate with the same apparent  $\text{pK}_a$  as H306 and the associated differences in  $^{15}\text{N}$  chemical shift between the conformers with protonated and non protonated H306 (see methods). As an example, Figure 9B shows the titration curve for the amide nitrogen of F303. It displays an apparent  $\text{pK}_a$  of  $5.75 \pm 0.08$ , in excellent agreement with the value found following the  $\text{H}^{\square 1}$  signal of H306. Overall, eleven residues were found to be perturbed by H306 protonation (Figure 9C). These residues include the C-terminus of the  $\alpha$ -helix (R300-R302), F303, the entire  $3_{10}$  helix (K304-L309) and three residues close in space (T316, Y335 and Q347). This result indicates that there is a conformational transition in the DNA binding helix triggered by H306 protonation, involving both the C-terminus of the canonical  $\alpha$ -helix and the  $3_{10}$  helix (residues 300-309).

## **Molecular environment of E2C histidine residues and coupling between histidine protonation and specific E2C-DNA binding**

Based on thermodynamic and spectroscopic results indicating that titration of at least three E2C residues between pH 5 and 9 is coupled with changes in the free energy for specific DNA binding or with conformational changes in the native state of the domain, we have characterized in detail the molecular environment of the E2C the five histidine residues in the E2C sequence ( Figure 9A ) in the free and DNA-bound states of the protein using nuclear magnetic resonance (NMR) in order to identify the residues whose protonation is coupled with changes in protein conformation and binding. We performed all NMR experiments at pH values between 5 and 9 and in 50 mM sodium phosphate, 5 mM DTT. Temperature was 30 °C for the free protein and 45 °C for the E2C-DNA complex. These conditions optimized spectral resolution and protein stability and solubility as observed in  $^1\text{H}$ - $^{15}\text{N}$  long-range HSQC spectra (data not shown). On the other hand, ionic strength varies for this buffer in the pH range used and the pK-values obtained should be regarded as approximations. First, we recorded a series of  $^1\text{H}$ - $^{15}\text{N}$  HSQC and long range  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra of free E2C as a function of pH to observe the correlations between the two carbon-attached hydrogens ( $\text{H}^{\delta 1}$  and  $\text{H}^{\delta 2}$ ) and the two nitrogens ( $\text{N}^{\delta 1}$  and  $\text{N}^{\delta 2}$ ) of the imidazole ring. The relative disposition of the cross-peaks in the  $^1\text{H}$ - $^{15}\text{N}$  long range HSQC spectrum can be used to determine which of the two nitrogens is attached to the hydrogen in the neutral form.( FIGURE 9B E 9C) We used the chemical shift for both nuclei of the histidine imidazole ring and for the amide group of the protein backbone to determine the pK-values for the histidine residues. Overall, the free E2C histidine pK-values presented here are in good agreement with a previous study carried out under different solvent conditions ( Bose et al, 2007). An additional residue, Cys298, titrates within this pH range with a pK of 8.8 in free E2C This value is close to the pK of 8.6 found for one of the conformational transitions monitored by CD (Welzer at al,). The imidazol ring of a deprotonated histidine can exist in two tautomeric forms, depending on whether the hydrogen is bound to the  $\text{N}^{\delta 2}$  or  $\text{N}^{\delta 1}$  atoms (Bachovchin et al , 1986). The relative populations of the  $\text{N}^{\delta 2}$ -H and  $\text{N}^{\delta 1}$ -H tautomers are shown in Table 2. We can combine the tautomer populations and pK-values to draw a picture of the molecular environment of the five E2C histidines in the two forms of the protein. For a solvent-exposed histidine, the expected pK is approximately 6.3 and the populations of the  $\text{N}^{\delta 2}$ -H and  $\text{N}^{\delta 1}$ -H tautomers are expected to be in a 80:20 ratio (Bachovchin et al , 1986)

FIGURE 9



### **His322 and His326**

These two histidine residues are located in the solvent-exposed and flexible  $\beta$ 2- $\beta$ 3 loop (Figure 9A). In free E2C they both present the canonical pK and tautomer ratio, which indicates that they are fully exposed to the solvent. The  $\beta$ 2- $\beta$ 3 loop is known to remain highly flexible in the DNA complex (Cicero et al, 2006 ) and forms only long-range contacts with the DNA. The loop is close in space to the negatively charged phosphate backbone (Cicero et al, 2006 ) and mutation of residues Lys325 and Lys327 to alanine decreases the affinity for DNA moderately (Ferreiro et al , 2005). From these data, we hypothesized that the two chargeable histidine residues could establish electrostatic interactions with the target DNA. However, the chemical shifts, pK and tautomer populations of His322 and His326 do not change upon DNA binding (data not show). Thus, His322 and His326 do not take part in DNA binding and remain largely solvent exposed in the DNA complex. This can be explained by the presence of two positively charged lysines in the  $\beta$ 2- $\beta$ 3 loop, which may hinder histidine protonation.

### **His306**

This histidine residue is in the solvent-exposed side of the C-terminal half of the DNA binding helix (Figure 9A). In free E2C, His306 shows a decreased pK of 5.7 and a modest decrease in the population of the N<sup>ε</sup>-H tautomer. This indicates that this residue establishes weak interactions with its tertiary environment, which disfavor the N<sup>ε</sup>-H tautomer and proton binding. The NMR parameters of His306 change little upon binding, indicating that the side chain is not directly involved in binding. The role of this residue in E2C function may be to stabilize the strained conformation of the C-terminus of the DNA binding helix (Wetzler et al., submitted).

### **His318**

This histidine residue belongs to the solvent-exposed side of the  $\beta$ 2 strand, on the DNA binding surface of the domain (Figure 3A). His318 shows an increased pK of 6.7 in free E2C, implying that the native E2C structure favors protonation. In agreement with this, in the neutral form of His318 the two possible tautomers are almost equally populated, with both nitrogen atoms oscillating rapidly between non-protonated and protonated states, causing the side chain to resemble a charged histidine. DNA binding shifts the pK of His318 to 7.8, the strongest perturbation of all E2C histidines ( Figure 4A). The tautomer population of the neutral His318 in complex with DNA could not be determined due to the shift in pK. We interpret that His318 interacts with the negatively charged phosphate backbone of the DNA via water molecules, as in some (Hooley et al , 2006) but not all (Hegde et al 2002; Hooley et al , 2006 ) homologous

complexes. Thus, we observe that His318 protonation is coupled to both specific DNA binding and the modulation of the E2C native state ensemble.

### **His288**

His288 belongs to the  $\beta 1$  strand and is buried in the hydrophobic core of the  $\beta$ -barrel, far from the DNA binding surface (Figure 9A). In the crystal structure of HPV16 E2C there is a clear interaction between the His288 from each monomer and a water molecule acting as a bridge between the two  $H^{\epsilon 2}$  atoms (Hegde et al, 1998) This water molecule was observed also in the crystal structure of HPV31 E2C (Bussiere et al , 1998). We determined that neutral His288 presents a 100% of the  $N^{\epsilon 2}$ -H tautomer in free E2C, indicating that the histidine residues act as hydrogen bond donors and the central oxygen of the water molecule as a double hydrogen bond acceptor. Strong hydrogen bonding may cause a downfield shift of the  $N^{\epsilon 2}$  nucleus of up to 10 ppm. However, the chemical shifts of the  $N^{\epsilon 2}$  and  $N^{\delta 1}$  atoms show that they are not involved in strong hydrogen bond contacts, suggesting that the hydrogen bond between the water molecule and the  $N^{\epsilon 2}$  atom is a consequence of the dimer interface structure rather than a strong stabilizing factor of the overall structure (Campos et al, 2005). His288 also presents a sharply decreased pK in the range of 4-5 (Figure 4B). The exact value could not be accurately determined by using the  $^1H$ - $^{15}N$  cross peaks due to both line broadening and overall protein instability below pH 5. We interpret that partial protonation of His288 renders a more dynamic E2C, while full protonation leads to global unfolding of the domain. The changes in chemical shift reported for E2C NH groups between free and complexed E2C are small ( Cicero et al, 2006 ). As for the free form of the protein, we could not determine the pK of His288 in the E2C-DNA complex due to incomplete titration (Figure 4B). Interestingly, the curve for  $N^{\delta 1}$  shows a double sigmoidal shape that can be fitted with pK values of 4.4 and 7.7 (see methods). Thus, protonation of His288 is linked to both DNA binding and the dynamics and global unfolding of the domain. Overall, our results show that the five E2C histidines depart to different degrees from the canonical pK values and tautomer populations, these deviations being the largest for His288 and His318. Comparison of the NMR data for free E2C and the E2C-DNA complex identifies His288 and His318 as two key residues whose protonation is coupled to both specific DNA binding and the structure and dynamics of the domain.

## Coupling between histidine protonation and E2C structure

Next, we determined which regions of E2C change conformation upon protonation of histidine residues. The  $^{15}\text{N}$  chemical shifts of many residues in the free form of the domain change in the pH range from 5 to 9. Fitting the curves of chemical shift versus pH to equation [2] allowed to extract the pK-values and well as the differences in chemical shift between the structures with a given histidine protonated or non-protonated. Some of the titrations were biphasic, indicating that some  $^{15}\text{N}$  chemical shifts change upon titration of more than one residue (see Figure 10 for an example). In these cases, the data were fitted using a double sigmoidal curve, obtaining two pK-values and the corresponding changes in chemical shift. Most residues titrate with apparent pK-values that are almost identical to those determined for one of the five histidine residues or Cys298. We could not measure the influence of His322 and His326 protonation on the structure of the  $\beta$ 2- $\beta$ 3 loop because most residues in this region cannot be monitored above pH 5. Only a few residues from the rest of the domain co-titrate with these two histidines, indicating that protonation of His322 and His326 does not influence residues that are far away in the structure. Similarly, only residues in the DNA binding helix or close in space co-titrate with His306 (11 residues, average pK  $5.6 \pm 0.1$ ) or Cys298 (6 residues, average pK  $8.7 \pm 0.1$ ). In contrast with other histidines, protonation of His318 leads to  $^{15}\text{N}$  chemical shift changes in residues belonging to the four strands of the dimeric  $\beta$ -barrel (Table 1). Interestingly, titration of His318 extends its effect far away from its position in the structure, including residues 360-363 in the  $\beta$ 4 strand on the opposite side of the barrel that are on average 16.5 Å apart ( $C_\alpha$ - $C_\alpha$  distance) and residues Ala329, Val331 and Thr332, which are 4.9, 7.6 and 10.4 Å apart, respectively. Most of the co-titrating residues are involved in contacts between the two monomers. On the other hand, the alpha helices of E2C do not show widespread changes in  $^{15}\text{N}$  chemical shift upon protonation of His318. His288 co-titrates with 12 residues, with an average pK-value of  $4.6 \pm 0.1$  (Table 1). This value is in the range anticipated from the tautomer populations and chemical shifts of this side chain. As for His318, protonation of His288 causes extensive  $^{15}\text{N}$  chemical shift changes in the four strands of the dimeric  $\beta$ -barrel and the dimerization interface, but not in the alpha helices (Figure 10C). The perturbation reaches the  $\beta$ 2 strand in the DNA binding surface, providing an explanation for the coupling between His288 protonation and DNA binding. Altogether, the perturbation of  $^{15}\text{N}$  chemical shifts by protonation of E2C side chains confirms the key role of His288 and His318 protonation in both E2C native state structure and DNA binding. The similarity of the effects of His288 and His318 protonation shows that for the dimeric  $\beta$ -barrel behaves as a cooperative unit that undergoes global changes in conformation, regardless of where it is perturbed. The alpha helices of E2C, including the DNA binding helix, seem to behave independently from the  $\beta$ -barrel.



FIGURE 10

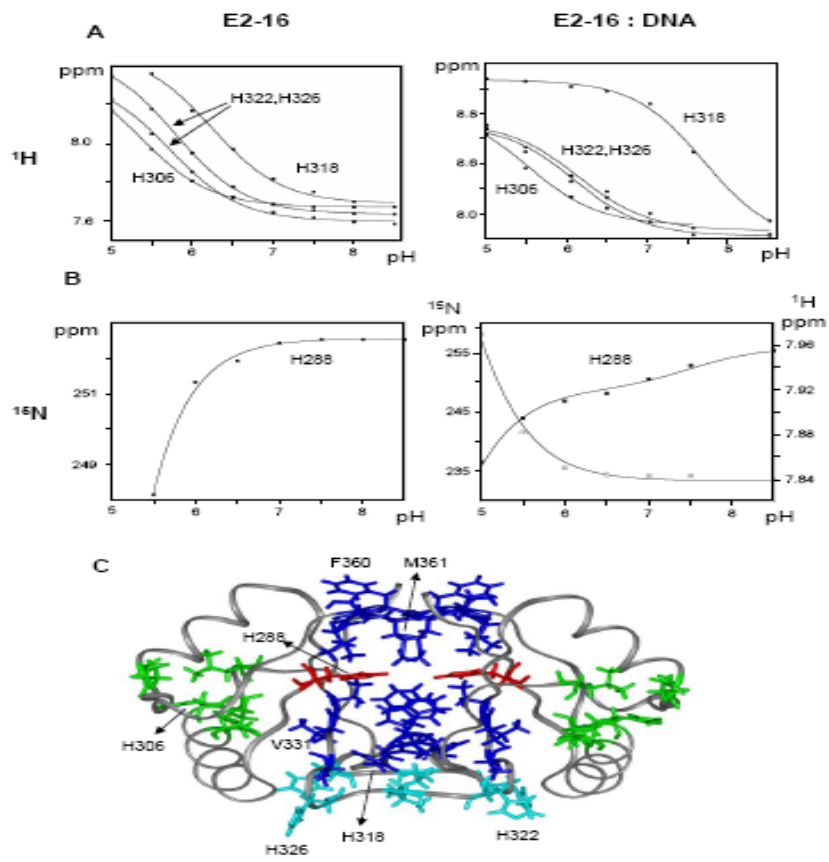


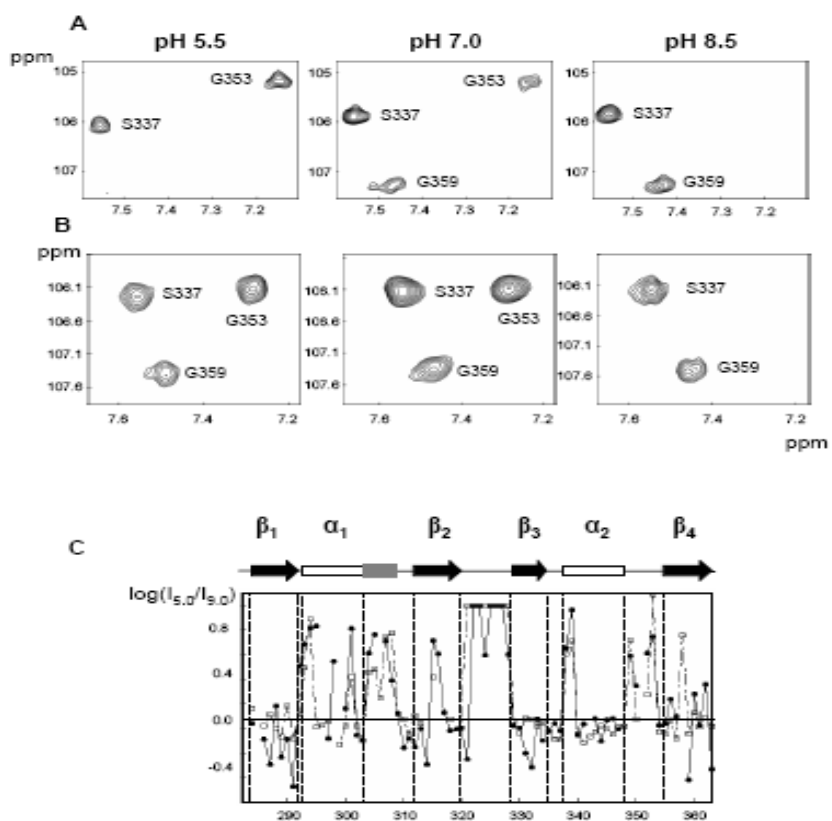
TABLE 1

Residue	pK	$\Delta\text{ppm}^{15}\text{N}$
T284		0.91
H288		
L289		0.22
K290	$4.7 \pm 0.2$	0.78
H318	$4.5 \pm 1.0$	1.21
W319	$4.5 \pm 0.2$	1.45
I330	$4.5 \pm 0.2$	0.43
T332	$4.7 \pm 0.2$	0.41
D336		1.15
S337	$4.5 \pm 1.0$	0.74
S357	$4.7 \pm 0.2$	0.83
M360	$4.5 \pm 0.2$	0.69
F361	$4.5 \pm 0.2$	0.57
T284	$6.84 \pm 0.03$	0.96
L289	$6.51 \pm 0.05$	0.27
K290	$6.7 \pm 0.1$	0.26
T311	$6.65 \pm 0.05$	0.52
H318	$6.73 \pm 0.05$	1.46
W319	$6.81 \pm 0.05$	0.93
I330	$6.5 \pm 0.1$	0.14
T332	$6.73 \pm 0.05$	0.65
T334	$6.72 \pm 0.09$	0.28
S337	$6.7 \pm 0.2$	0.18
Q347	$\pm$	0.1
S357	$6.5 \pm 0.1$	0.24
M360	$6.76 \pm 0.03$	0.44
F361	$6.70 \pm 0.07$	0.95
S362	$6.72 \pm 0.01$	0.62
I363	$6.5 \pm 0.2$	0.86
C298	$8.78 \pm 0.04$	3.92
R300	$8.6 \pm 0.2$	0.27
Y301	$8.5 \pm 0.2$	0.73
R302	$8.8 \pm 0.2$	0.26
H306	$8.8 \pm 0.2$	1.18
L309	$8.6 \pm 0.2$	0.21
I350		0.17
G291	$6.28 \pm 0.08$	
W317	$6.3 \pm$	
T320	$6.3 \pm$	

## Coupling between histidine protonation and the global conformational dynamics of E2C

The pH dependence of the amide  $^{15}\text{N}$  chemical shifts shows that the  $\beta$ -barrel of E2C undergoes a small but cooperative change in conformation upon protonation of His318 and His288. We will now examine the effect of histidine protonation on the dynamics of the domain using the pH dependence of the cross-peaks in the  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectrum (Figure 11). The intensity of a peak is in principle independent of pH for residues not exchanging with the solvent. This is the case of Ser337 (Figure 11A). In general, peaks for residues exchanging with the solvent will diminish their intensity when the pH is incremented. This is the case for Gly 353 (Figure 5A), located in the solvent exposed  $\alpha 2$ - $\beta 4$  loop. The intensity of a peak may also decrease because of an increase in protein motions in the  $\mu\text{s}$ - $\text{ms}$  time scale. Figure 5C shows the logarithm of the ratio between the peak intensities at pH 5 and 9 (filled circles) for all characterized E2C amides. In this graph residues exchanging with the solvent will present a positive value for  $\log(I_{5.0}/I_{9.0})$ . Among the regions with regular secondary structure, helix  $\alpha 2$  and strands  $\beta 1$  and  $\beta 3$  show little exchange of the amide proton with water. Strands  $\beta 2$  and  $\beta 4$ , both closing the dimeric  $\beta$ -barrel, and the recognition helix  $\alpha 1$ , show a high degree of water exchange. It is worth noting the differences between the two helices: for  $\alpha 2$  only the first two residues, not involved in hydrogen bonding, show fast water exchange kinetics. In contrast, the amides of several residues belonging to the DNA binding helix show fast exchange with water. They include the N-cap region (Ala293, Asn294, Thr295), and residues with their amide groups close to the side of the DNA interaction (Cys298, Lys304 and Lys305). Last, the exchange effect is strongest for the solvent exposed  $\beta 2$ - $\beta 3$  loop, for which no signal was detected at pH 7, but that could be assigned at pH 5.5 using a combination of HNCO, CBCAcoNH,  $^{15}\text{N}$ -edited TOCSY and  $^{15}\text{N}$ -edited NOESY 3D spectra. On the other hand, a number of residues show a large decrease in intensity upon going to acidic pH, shown as a negative  $\log(I_{5.0}/I_{9.0})$  in Figure 11C. Gly359 is shown as an example in Figure 5A. The residues are spread over the four strands of the barrel: Val287, Leu 289 and Gly291 in  $\beta 1$ , Ser314 in  $\beta 2$ , Val331 and Thr332 in  $\beta 3$  and Gly359 and Ile363 in  $\beta 4$ . None of these residues gave a signal in the HSQC recorded at pH lower than 5.5. For these residues, the decrease in intensity cannot be due to solvent exchange, pointing at an increase in motions in the  $\mu\text{s}$ - $\text{ms}$  time scale in the  $\beta$ -barrel at acidic pH. We conclude that protonation of His288 and His318 not only induces changes of conformation in the  $\beta$ -barrel but also increases its slow dynamics.

FIGURE 11

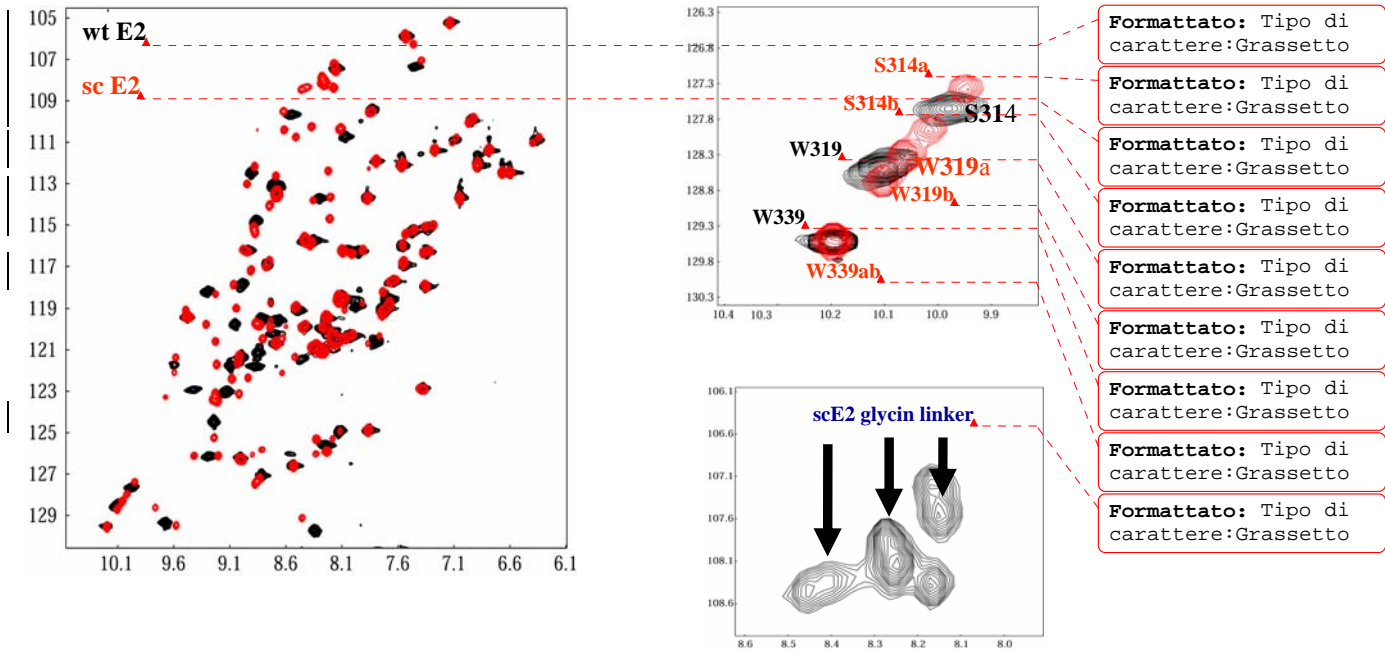


### **Backbone assignments and dimer asymmetry of HPV-16 sc-E2**

The assignment of the sc-E2 backbone resonances was obtained following standard procedures based on triple-resonance and double-resonance heteronuclear experiments.

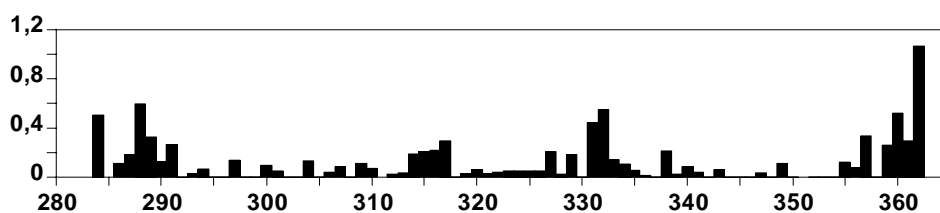
All HPV-16 scE2C backbone  $^{15}\text{N}$ ,  $^{13}\text{C}$  and  $^1\text{H}$  resonances were sequentially assigned on the basis of the following standard set of double- and triple-resonance heteronuclear spectra:  $^1\text{H}$ - $^{15}\text{N}$  HSQC (Bodenhausen and Ruben, 1980; Kay et al., 1992), CT (constant time)  $^1\text{H}$ - $^{13}\text{C}$  HSQC (Vuister and Bax, 1992), HNCA (Ikura et al., 1990), HN(CO)CA (Ikura et al., 1990), HNCO (Muhandiram and Kay, 1994), CBCA(CO)NH (Muhandiram and Kay, 1994), HNHA (Vuister and Bax, 1993) and HACACO (Bazzo et al., 1995). Following the well-established methodology of heteronuclear sequential connectivity in three-dimensional spectra (Ikura et al., 1990; Bax and Grzesiek, 1993) the combined information extracted from these spectra permitted to obtain correct backbone assignment. Figure 12 shows the  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectrum of the scE2 protein ( red ) superimposed with the wild-type E2 ( black ). While the wild-type protein shows a single set of resonances, indicating a symmetric structure for the dimer in solution, several residues in the E2 single-chain version have distinct pairs of peaks which come from the two monomer units . Thus it follows the idea that the wt E2 symmetry is disrupted because of the scE2 monomerization and that the two scE2 monomers can potentially assume a different behaviour in solution.

**FIGURE 12**

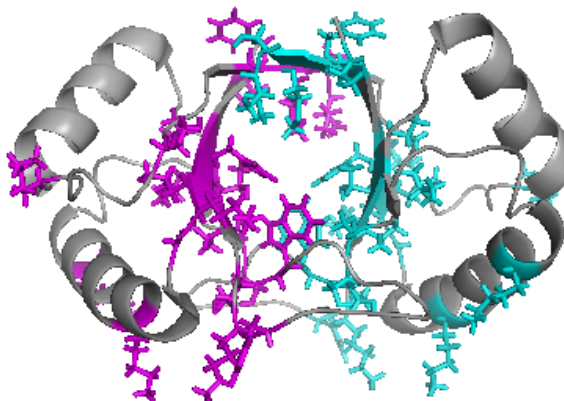


### The Sc-E2 Asymmetry in Chemical Shift

The NH resonances are exquisitely sensitive to modest modifications in the disposition of surrounding nuclei and become very dispersed for protein folded conformations in a non-predictable fashion. Nevertheless, nuclei of the backbone are mainly dependent on the local stereochemical and stereoelectronic arrangement. Fig 13 shows the differences in chemical shift of amide  $^1\text{H}$ - $^{15}\text{N}$  resonances between the two scE2 monomers along the protein aminoacidic sequence at pH 6.61. Several regions shows significant chemical shift inter-monomers differences in particular in the region be I286-G291, V313-W317, V331-T332 and T358-I363. This indicates that the chemical shift equivalence present in wild-type E2 is lost in scE2 and that the perturbation of the symmetry produced by the glycine linker extends up to the opposite face, the DNA-binding surface ( Fig.11).



Monomer A



Monomer B

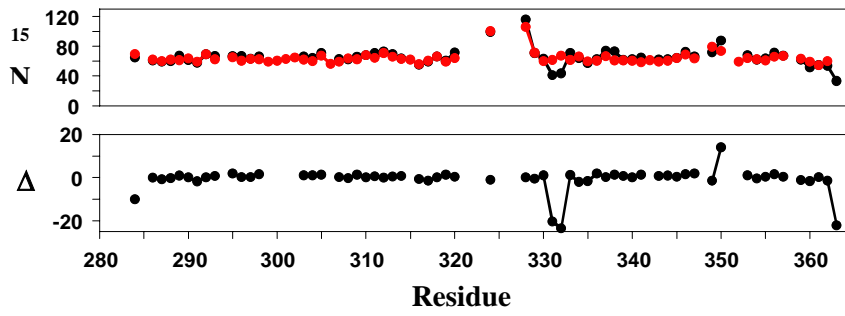
## **E2 single-chain Dynamics**

In order to study the influence that the monomerization have on the overall structure of the scE2 we have measured the pH titration by monitoring the chemical shift changes of backbone NH groups. As performed for the wt E2 , we have measured the logarithm of the ratio between the peak intensities at pH 5 and 8 for all characterized amides of scE2C ( red) and we have compared with the results for wtE2C ( black) and E2C+DNA complex ( green ) ( Fig 14A). Residues that show no internal dynamics and do not exchange significantly with the solvent show values around zero. We see in the logarithmic graph that residues showing a large decrease in intensity upon going to acidic pH ( i.e with negative  $\log(I_{5.0}/I_{9.0})$  in the wtE2 become pH independent both in E2C:DNA and in the scE2. Gly359 is a representative example because it doesn't change its intensity at acidic pH as in wtE2. This suggests that the scE2 glycine-linker, as the DNA binding, stabilizes the dimer and masks the pH driven process in the E2 single chain variant. This results was confirmed by the  $^{15}\text{N}$  relaxation data at pH 6.5 for scES by comparing  $^{15}\text{N}$   $T_2$  with the wtE2 profile ( fig 14 A). The slow and fast motions that confer plasticity and adaptability in the DNA target search process in the wt protein are quenched in the single chain variant because the linker changes the protein backbone mobility

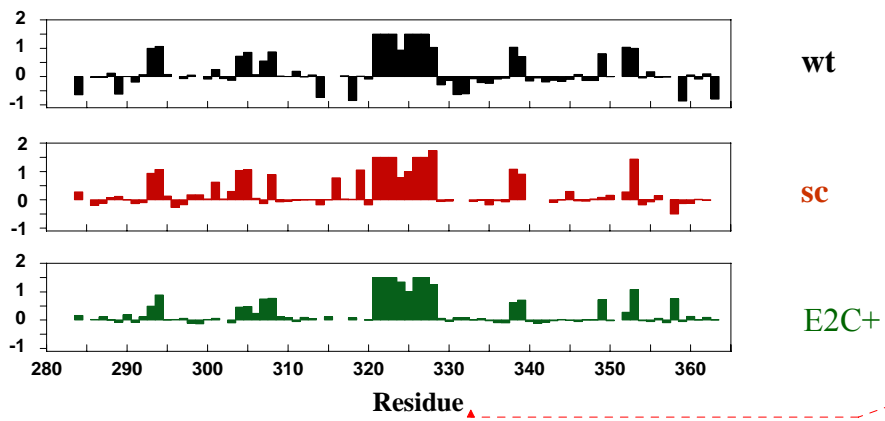


FIGURE 14

A



B



## E2sc histidine Investigation Reveals Asymmetry in the pH Profile

As for the wild-type E2 protein, the histidine side chains of E2sc were investigated and their pKa values were determined in order to evaluate in detail the scE2 molecular environment around each histidine residue. Fig 15 show the  $^1\text{H}$ - $^{15}\text{N}$  long-range HSQC spectra of scE2 histidine residues. We see that each histidine give rise to two sets of signals, confirming that two distinct monomers exist in solution. As for the wild-type E2, we determined which regions of scE2 change conformation upon protonation of histidine residues by fitting the curves of chemical shift versus pH. In this way, the pKa-values were determined. The results were compared with the corresponding values for wtE2 ( see table 2). It can be seen that the monomerization doesn't change the pH profile for E2sc histidine residues with the exception of His288, situated in the hydrophobic core of the  $\beta$ -barrel and far from the DNA binding surface. It exhibits a different pKa between the two monomers and is another evidence of the break of the structural symmetry by monomerization.

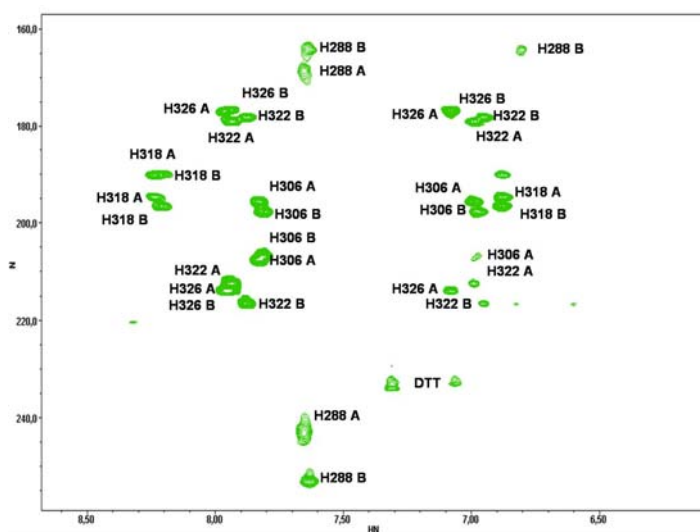
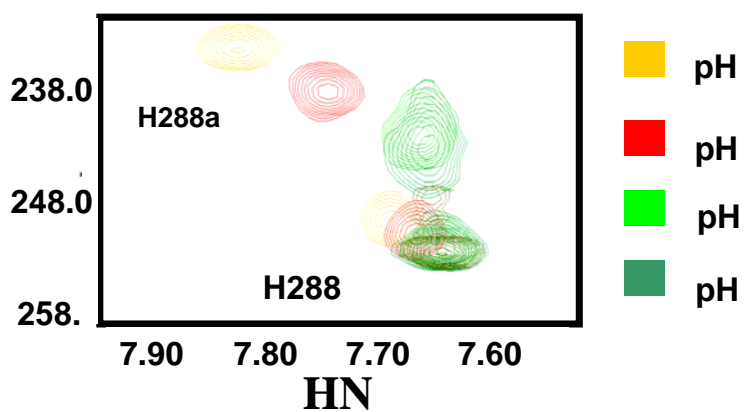


TABLE 2

Res	pK <sub>a</sub>	
	wt E2	sc E2a
His288	4-5	>3.5
His306	5.7	5.8
His318	6.7	6.7
His322	6.3	6.3
His326	6.2	6.2

H-<sup>15</sup>N long range HSQC





## **DISCUSSION**

The HPV 16 E2C is a stable dimeric protein. 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. In fact, the recognition helix itself, which appears in the crystal structures in a canonical conformation, displays in solution elements of flexibility as revealed by NMR analysis. The exchange rates for most of the backbone NHs in the HPV 31 E2C domain, and the HPV 16 E2C is high which apparently contradicts the general correlation between compaction, stability, and low backbone amide hydrogen exchange rates in proteins. This can be explained by a constant breathing of the dimer where fast dissociation and association occur but this remains to be determined. We hypothesize that the breathing is at the level of partial unfolding and dynamics, not dissociation. This domain presents a two-state dissociation at equilibrium, and a structured monomeric intermediate species in the kinetic pathway that cannot be native-like. At this stage, it is not possible to know how related this monomeric species is to the putative monomer that could result from extreme dilution of the dimeric domain. So far, no report is available on the isolation of a “native” monomer, which, due to the impossibility of a solvent exposed half-barrel, cannot be native and remain monomeric. The exception is one report that claims that a cooperative binding isotherm at high glycerol content is due to the formation of a monomer-DNA intermediate that binds the second monomer subsequently. The same report showed a stabilisation of the dimer to pressure dissociation by addition of the same amounts of glycerol, making the results difficult to interpret. As mentioned above, a monomer with residual structure was obtained at high hydrostatic pressure, but so far, no monomer has been isolated in standard native conditions, because of the large stability of the dimer. In addition, the early binding experiments carried out with the purified HPV 16 E2C domain at 0.1 nM detected no signs of cooperativity. The existence of a transient monomeric species seems plausible *in vivo*, but it is difficult to picture it as an isolated entity. In addition, it may bind DNA, but it would be very surprising if it binds with higher affinity than the “divalent” dimer recognising a palindromic sequence with two symmetric helices. So far, E2 has been difficult to detect in natural early infections; it can only be visualised 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, but the formation of a monomer-

DNA species on the typical E2 site is not clear. Finally, we should keep in mind that the N-terminal transactivation domain also has the ability to dimerise, which would make the full-length E2 dimer even tighter. As we have been discussing, E2C domain is stable but fairly dynamic, in particular, the DNA binding helix exchanges its amide protons faster than any other region, implying an equilibrium between a well formed helix and a more disordered conformation. This equilibrium may not involve a native monomer since the NMR experiments are carried out in millimolar concentrations. On the other hand, circular dichroism experiments show that no global unfolding is required for entering the amyloid route: the large increase in  $\beta$ -sheet signal not only implies an increase in such structural motif, but also the disappearance of  $\alpha$ -helix, which always presents a comparative larger molar ellipticity. In addition, DNA prevents E2C entering the amyloid route, and the same goes for higher amounts of TFE, both stabilising the main helix. We hypothesise that the local destabilisation of the DNA binding helix is essential for entering the amyloid route. 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. E2 was recently found associated along the mitotic spindle fibres throughout the different phases of mitosis. Moreover, the region mapped for this possible oligomerisation was mapped to a  $\beta$ -strand in the E2C. 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 oligomerise. Since the E2 proteins have multiple putative interaction partners one possible role of oligomerisation might be to allow the integration of inputs from multiple partners by an E2 assembly. 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<sup>-1</sup>) contributions to binding energy, and no individual “hot spot” contacts could be observed. 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  $\beta$ 2- $\beta$ 3 loop, indicate that this region does contact DNA, although dynamically. This suggests a reevaluation of the condition of “non-contacting” loop and “non-contacting” spacer DNA sequence. Finally, a new contact, lysine 349, the first contact identified outside of the DNA binding helix, actually located in the loop connecting  $\beta$ 2 with  $\beta$ 4. The contact made by lysine 349, not present in several relevant strains, may be important in the enhanced transcriptional activity (positive and negative) of E2 from HPV16. 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. 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. This apparent contradiction is not such, because in the latter approach, a single short duplex (18 bp site 35) was used, and the DNA sequence was not a variable. The combination of mutagenesis and variations in DNA conformation, domains from different strains, or the addition of cations and other molecules, will likely provide a measure of proportionally different contributions from direct and indirect readout modes. Our NMR measurements of the regions corresponding to helix1 and the  $\beta$ 2 strand in the native state of the HPV-16 E2 is in good agreement with the nascent secondary structure of  $\alpha$ 1-E2C. This result shows that residual structure in E2C helix1 in the unfolded state is of local nature and is independent of residual structure in the adjacent  $\beta$ 2 strand. It also establishes  $\alpha$ 1-E2C as a model for the unfolded state of E2C and should facilitate future studies on the role of unfolded state structure in the folding of E2. The coincidence between the residual structure in the unfolded state and the regions of  $\alpha$  and  $3_{10}$  helix in the folded domain is also remarkable and underlines the importance of the local conformational tendencies of this sequence, even in the context of a folded protein. Molecular dynamics simulations suggested that the C-terminal region of alpha helix1 spends a significant amount of time in a polyproline type II conformation. We have now established experimentally that this  $\alpha$ -helix to polyproline II transition is caused by the conformational preferences of this sequence stretch and of local nature. It has been suggested that polyproline type II conformations can act as nucleation sites for the formation of oligomeric  $\beta$ -sheet structure. Since helix1 unfolds in the initial step of the E2C amyloid route, we speculate that nascent polyproline type II structure in its C-terminal region may act as a nucleation site for misfolding of the whole domain. Another possible effect of partial unfolding of helix1 could be the acceleration of DNA binding via a fly-casting mechanism. The particular dimeric  $\beta$ -barrel topology of E2C holds the DNA binding helix in a “ready for binding” conformation. Moreover we have extended the study of the HPV16 E2C domain DNA binding surface to His306 and His318 by measuring the coupling between DNA binding and histidine protonation. The direct contacts between His318 and the DNA backbone are important for binding, as anticipated from structural studies. Regarding the other histidine



residues in E2C, we initially expected that the extent of coupling between protonation and binding would correlate with the proximity to the DNA in the complex, with His306 in the DNA binding helix and His318 in  $\beta 2$  being the most important; His322 and His326 from the flexible  $\beta 2$ - $\beta 3$  loop playing a secondary role and His288 in the core of the  $\beta$ -barrel being decoupled from function. Surprisingly, we found that coupling is strongest for His288 and His318, weak for His 306 and absent for His322 and His326. This hierarchy does not correlate with direct protein-DNA contacts in the complex but with the molecular environment of each histidine in the free state of E2C. His288 and His318, which deviate most from a canonical, solvent exposed side chain, are the ones with the strongest influence on binding. His288 exerts its influence in spite of being far from the binding surface and His318 contacts the DNA directly, but its protonation is clearly coupled with the structure and dynamics of E2C through long-range effects. This strongly suggests that the network of intramolecular interactions in the domain plays a role in function through an indirect readout mechanism. The different conservation levels of the five E2C histidines in human genital papillomaviruses also correlate with their environment in the free protein rather than with direct protein-DNA contacts. Histidines 288 and 318 are conserved in most human genital papillomavirus types, His306 is present in around half of the viral types and histidines 322 and 326 appear only in a minority of types. Moreover, naturally occurring substitutions for His288 and His318 parallel the dominant protonation and tautomeric states of the side chains in free HPV16 E2C. The only other residue observed at position 288 is a glutamine, which is the closest mimic possible for the  $N^{\epsilon 2}$ -H tautomer of a neutral histidine thanks to the  $\epsilon$  amide nitrogen. Similarly, in nature His318 is only substituted by arginine, which is structurally similar to a protonated histidine due to the neutral  $N^{\delta}$  and the partially charged  $N^{\gamma}$ . We interpret that the conservation/substitution patterns observed in nature support the functional significance of our *in vitro* pK values and tautomer populations and suggest that histidine protonation may actually play a regulatory role *in vivo*. The long-range effects of histidine protonation on E2C structure make the domain more dynamic upon lowering the pH from 9 to 5, which leads to 8-fold stronger DNA binding. The simplest explanation would be that at pH 9 the protein needs to change conformation to bind its target DNA. This conformational transition, even if subtle, would impose a free energy penalty. In this model, weakening of the structure of free E2C at more acidic pH lowers the free energy cost of the rearrangement and leads to higher affinity. It will be interesting to test whether the same mechanism modulates the kinetics of binding along one or both of the parallel reaction pathways. On the other hand, upon going from pH 5 to 9, the dissociation constant for the E2C homodimer increases by about 500-fold, unfolding becomes 5-fold faster and the rate-limiting refolding reactions become about 100-fold slower. Thus, there is a trade-off between function and the folding thermodynamics and kinetics of the globular E2C domain, similar to other DNA

binding protein. This trade-off limits how far the dynamics of E2C can be modulated for function without becoming a partially or fully disordered protein that folds upon binding. The HPV16 E2C  $\beta$ -barrel responds to histidine protonation in a highly concerted manner, with protonation of histidines at different locations leading to similar effects on both the structure and the dynamics of the domain at different timescales. Moreover, His318 and DNA binding clearly influence sites far away from the binding surface. The behavior of His318 shows that two of the previously proposed mechanisms for indirect sequence readout, namely, barrel flexibility and the electrostatic potential of the DNA binding surface, are coupled and need to be considered simultaneously. The changes in chemical shift upon histidine titration are small, as observed for DNA binding, and point at an overall robust structure of the barrel. Unfortunately, this subtle response precludes a detailed modeling of the conformational changes at this point. Here, we would like to propose that the quaternary interface between strands  $\beta$ 4- $\beta$ 4' and  $\beta$ 2- $\beta$ 2' is pivotal in the malleability of the  $\beta$ -barrel. In fact, the dynamics of the barrel in the  $\mu$ s-ms and hours timescales shows that the  $\beta$ 4 strands are the most dynamic in the barrel, followed by the  $\beta$ 2 strands. This observation of a “breathing” interface connecting two more stable monomers may be related to the monomeric intermediate observed in the refolding kinetics and the amyloid route of the domain. In agreement with our view, quenching of  $\beta$ -barrel breathing is the largest change in the dynamics of the HPV16 E2C domain upon DNA binding. It will be interesting to compare our results with the behavior of homologous E2C domains showing changes in sequence discrimination capacity and quaternary structure, such as a different hydrogen bond register in the  $\beta$ 4- $\beta$ 4' quaternary interface or the presence of N- or C-terminal extensions packing against the core of the barrel. The dynamics of the HPV16 E2C domain show a remarkable partition into a few cooperative units. The most prominent feature is that changes in the highly cooperative  $\beta$ -barrel do not propagate to the  $\beta$ 2- $\beta$ 3 loop or either helix. Additionally, the  $\beta$ 2- $\beta$ 3 loop and the DNA binding helix show only local changes upon protonation of His306, Cys298, His322 and His326, showing that they are independent cooperative units dominated by local interactions. These data are in agreement with the view of the E2C  $\beta$ -barrel as a sensitive hinge that modulates the relative position of the p53 and DNA binding helices, while the  $\beta$ 2- $\beta$ 3 loop behaves independently. We would like to add that such a role for the  $\beta$ -barrel requires that the two monomers should behave as rigid flaps, and in particular that the DNA binding helices are firmly anchored to the  $\beta$ -sheet of each monomer. Indeed, the lack of coupling between the barrel and the helices corroborates the rigidity of the monomer “flaps”. In summary, we have uncovered that the dimeric  $\beta$ -barrel of the HPV16 E2C domain is a cooperative functional unit that takes part in DNA sequence recognition through indirect readout effects. It is known that the global dynamics of a protein are mainly determined by the overall arrangement of secondary structure elements within its tertiary and quaternary

structure. We propose that this topology-activity relationship is one of the reasons behind the rare fold of the DNA-binding domain of the E2 human papillomavirus master regulator and its homodimeric nature. Finally, using an engineered monomeric version of this protein inserting a short aminoacidic linker concatenates the two wild type monomers ( the single chain E2 protein ) in order to probe the link of the  $\beta$ -barrel fold to stability, dimerization and DNA-binding properties we demonstrated that monomerization of E2 increases the protein stability and the DNA binding specificity and quenches the slow motion flexibility of the native structure. The linker breaks the protein symmetry and the asymmetry is transmitted up to distant sites in the protein structure. Interestingly, the two opposite faces of the  $\beta$ -barrel, the linker region and DNA-binding site, communicated by a network of structural and dynamical interactions.

## **CONCLUSIONS**

- NMR shows that the peptide corresponding to the DNA-binding helix of the C-terminal domain of the HPV16 E2 master regulator has high alpha helix propensity in the N-terminal region of the peptide while the C-terminal region can form both alpha helix and polyproline type II. In the full length domain these two regions form alpha helix and  $3_{10}$  helix, respectively. Titration of a histidine residue in the  $3_{10}$  helix of the full-length domain induces local structural changes, similar to those observed between homologous domains and upon DNA binding. The peptide does not fold into an alpha helix upon binding a target DNA, and a mutation of the full-length domain that perturbs the tertiary interactions of the  $3_{10}$  helix decreases binding. Thus, the tertiary structure of the domain is crucial in modulating the intrinsic tendencies of this short sequence so that it meets its folding and binding requirements.
- The NMR dynamic data shows that the dimeric  $\beta$ -barrel of the HPV16 E2C domain is a cooperative functional unit that takes part in DNA sequence recognition through indirect readout effects. We identifies His318 in the complex interface and His288 in the core of the domain as the two residues whose titration is coupled with binding. The patterns of residue substitution in genital papillomaviruses show that titration of His288 and His318 may regulate DNA binding in vivo. Upon protonation of the same side chains, the dimeric beta-barrel core of E2C undergoes a highly cooperative change to a more dynamic conformation, pointing at a trade-off between structural stability and function. Specific DNA binding quenches the slow dynamics and amide hydrogen exchange in the beta-barrel, reaching residues far apart from the DNA recognition elements but leaving the two helices of each monomer unperturbed. Thus, the cooperative dynamics of the barrel are coupled to sequence recognition in an indirect readout mechanism.
- The changes in the E2C beta-barrel upon DNA binding are largest in the dimerization interface, suggesting that the barrel acts as a hinge that regulates the relative position of the DNA recognition helices. The most prominent feature is that changes in the highly cooperative  $\beta$ -barrel do not propagate to the  $\beta 2$ - $\beta 3$  loop or either helix. Additionally, the  $\beta 2$ - $\beta 3$  loop and the DNA binding helix show only local changes upon protonation of His306, Cys298, His322 and His326, showing that they are independent cooperative

units dominated by local interactions. We would like to add that such a role for the  $\beta$ -barrel requires that the two monomers should behave as rigid flaps, and in particular that the DNA binding helices are firmly anchored to the  $\beta$ -sheet of each monomer. Indeed, the lack of coupling between the barrel and the helices corroborates the rigidity of the monomer “flaps”.

- NMR study of the single chain E2 protein (sc-E2) reveals that the symmetry of the dimer is disrupted also at sites that are distant from the linker. In particular, amino acids at the DNA binding face, which is opposite in the  $\beta$ -barrel to the linker position, feel the asymmetric character of the protein. In addition, a different pH profile between the two monomers is observed for some residue and this behavior is structurally correlated to the presence of key histidines that modulate the stability of this protein and are sensitive to perturbations of the protein symmetry.

## **BIBLIOGRAPHY**

Abbate, E. A., J. M. Berger and M. R. Botchan: The X-ray structure of the papillomavirus helicase in complex with its molecular matchmaker E2 *Genes Dev* 2004 18, 1981-1996 (2004).

Alexander K. A. and 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).

D'Amelio, N., Bonvin, A.M., Czisch, M., Barker, P., Kaptein, R.: "The C terminus of apocytochrome b562 undergoes fast motions and slow exchange among ordered conformations resembling the folded state". *Biochemistry* (2002), 41:5505-5514.

Androphy E. J., D. R. Lowy and J. T. Schiller: Bovine papillomavirus E2 trans-activating gene product binds to specific sites in papillomavirus DNA. *Nature* 325, 70-73 (1987).

Bachovchin W.W. 15N NMR spectroscopy of hydrogen-bonding interactions in the active site of serine proteases: evidence for a moving histidine mechanism. *Biochemistry*. 1986 Nov 18;25(23):7751–7759

Bax, A., Griffey, R. H., and Hawkins, B. L. (1983) *J. Magn. Reson.* 55, 301-315

Bax, A., and Summers, M. F. J. (1986) *J. Am. Chem. Soc.* 108, 2093-2094.

Baker C. C., W. C. Phelps, V. Lindgren, M. J. Braun, M. A. Gonda and P. M. Howley: Structural and transcriptional analysis of human papillomavirus type 16 sequences in cervical carcinoma cell lines. *J Virol* 61, 962-971 (1987).

Baker, B. M. & Murphy, K. P. (1996). Evaluation of linked protonation effects in protein binding reactions using isothermal titration calorimetry. *Biophys J* 71, 2049-55.

Barsoum J., S. S. Prakash, P. Han and E. J. Androphy: Mechanism of action of the papillomavirus E2 repressor: repression in the absence of DNA binding. *J Virol* 66, 3941-3945 (1992).

Bazzo, R., Cicero, D.O., Barbato, G.: "A new 3D HCACO pulse sequence with optimized resolution and sensitivity. Application to the 21 kDa protein human interleukin-6". *J Magn Reson B* (1995), 107:189-191.

Beckett, D. (2001) Regulated assembly of transcription factors and control of transcription initiation, *J. Mol. Biol.* 314, 335-352.

Bechtold, P. Beard and K. Raj: Human papillomavirus type 16 E2 protein has no effect on transcription from episomal viral DNA. *J Virol* 77, 2021-2028 (2003).



- Bedrosian, C.L., Bastia, D.: "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* (1990), 174:557-575.
- Bochkarev, J. A. Barwell, R. A. Pfuetzner, E. Bochkareva, L. Frappier and 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).
- Bodenhausen, G.a.R., D.J.: "Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy". *Chem. Phys. Lett.* (1980), 69:185-189.
- Burd, E.M. "Human Papillomavirus and cervical cancer". *Clin. Microbiol. Rev.* (2003), 16:1-17.
- Bose, K., Yoder, N. C., Kumar, K. & Baleja, J. D. (2007). The role of conserved histidines in the structure and stability of human papillomavirus type 16 E2 DNA-binding domain. *Biochemistry* 46, 402-11.
- Boner, E. R. Taylor, E. Tsirimonaki, K. Yamane, M. S. Campo and 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).
- H. J. Bontkes, T. D. de Gruijl, A. Bijl, R. H. Verheijen, C. J. Meijer, R. J. Scheper, P. L. Stern, J. E. Bouvard V. A. Storey, D. Pim and 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).
- Brandsma, M. Shlyankevich, L. Zhang, M. D. Slade, E. C. Goodwin, W. Peh and 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).
- Buck, M. 1998. Trifluoroethanol and colleagues: cosolvents come of age. Recent studies with peptides and proteins. *Q Rev Biophys* 31: 297-355.
- 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).
- Bussiere, D. E., Kong, X., Egan, D. A., Walter, K., Holzman, T. F., Lindh, F., Robins, T. & Giranda, Structure of the E2 DNA-binding domain from human papillomavirus serotype 31 at 2.4 Å. *Acta Crystallogr D Biol Crystallogr* 54, 1367-76 (1998).
- Campos, L. A., Cuesta-Lopez, S., Lopez-Llano, J., Falo, F. & Sancho, J. (2005). A double-deletion method to quantifying incremental binding energies in proteins from experiment: example of a destabilizing hydrogen bonding pair. *Biophys J* 88, 1311-21.

- Cicero, D. O., Nadra, A. D., Eliseo, T., Dellarole, M., Paci, M., and Prat-Gay, G. d. (2006) Structural and thermodynamic basis for the enhanced transcriptional control by the human papillomavirus strain-16 E2 protein, *Biochemistry* 45, 6551-6560.
- Clower R. V., Y. Hu and T. Melend: Papillomavirus E2 protein interacts with and stimulates human topoisomerase I. *Virology* 348, 13-18 (2006).
- Clore, G.M., Driscoll, P.C., Wingfield, P.T., Gronenborn, A.M.: "Analysis of the backbone dynamics of interleukin-1 beta using two-dimensional inverse detected heteronuclear <sup>15</sup>N-<sup>1</sup>H NMR spectroscopy". *Biochemistry* (1990), 29:7387-7401.
- Clore, G.M. and, Gronenborn, A.M.: "New methods of structure refinement for macromolecular structure determination by NMR". *Proc Natl Acad Sci U S A* (1998), 95:5891-5898.
- Cripe TP, T. H. Haugen, J. P. Turk, F. Tabatabai, P. G. Schmid 3rd, M. Durst, L. Gissmann, A. Roman and 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).
- Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., Bax, A.: "NMRPipe: a multidimensional spectral processing system based on UNIX pipes". *J Biomol NMR* (1995), 6:277-293.
- Dell, G., Wilkinson, K.W., 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), 334:979-991.
- Doorbar, J. : Molecular biology of human papillomavirus infection and cervical cancer *Clin Sci (Lond)* 110, 525-41 (2006).
- Dotsch, V., Wagner, G.: "New approaches to structure determination by NMR spectroscopy". *Curr Opin Struct Biol* (1998), 8:619-623.
- 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), 339:785-796.
- Ellis, K. J. & Morrison, J. F. (1982). Buffers of constant ionic strength for studying pH-dependent processes. *Methods Enzymol* 87, 405-26.
- Englander, S.W., Sosnick, T.R., Englander, J.J., Mayne, L.: "Mechanisms and uses of hydrogen exchange". *Curr Opin Struct Biol* (1996), 6:18-23.
- Faber-Barata, J., Mohana-Borges, R. & Lima, L. M. (2006). Specificity in DNA recognition by a peptide from papillomavirus E2 protein. *FEBS Lett* 580, 1919-24.

Falconi, M., Santolamazza, A., Eliseo, T., de Prat-Gay, G., Cicero, D. O., and Desideri, A. (2007) Molecular dynamics of the DNA-binding domain of the papillomavirus E2 transcriptional regulator uncover differential properties for DNA target accommodation, *FEBS J.* 274, 2385-2395.

Farrow, N.A., Muhandiram, R., Singer, A.U., Pascal, S.M., Kay, C.M., Gish, G., Shoelson, S.E., Pawson, T., Forman-Kay, J.D., Kay, L.E.: "Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by <sup>15</sup>N NMR relaxation". *Biochemistry* (1994), 33:5984-6003.

Ferentz, A.E., Wagner, G.: "NMR spectroscopy: a multifaceted approach to macromolecular structure". *Q Rev Biophys* (2000), 33:29-65.

Ferreiro, D. U., Lima, L. M., Nadra, A. D., Alonso, L. G., Goldbaum, F. A. & Prat Gay, G. d. (2000). Distinctive cognate sequence discrimination, bound DNA conformation, and binding modes in the E2 C-terminal domains from prototype human and bovine papillomaviruses. *Biochemistry* 39, 14692-701.

Ferreiro, D. U., Dellarole, M., Nadra, A. D. & Prat Gay, G. d. (2005). Free energy contributions to direct readout of a DNA sequence. *J Biol Chem* 280, 32480-4.

Finucane, M.D., Jardetzky, O.: "The pH dependence of hydrogen-deuterium exchange in trp repressor: the exchange rate of amide protons in proteins reflects tertiary interactions, not only secondary structure". *Protein Sci* (1996), 5:653-662.

Gammoh N, H. S. Grm, P. Massimi and 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).

Grm H. S., P. Massimi, N. Gammoh and L. Banks: Crosstalk between the human papillomavirus E2 transcriptional activator and the E6 oncoprotein. *Oncogene* 24, 5149-5164 (2005).

Griesinger, C., Sorensen, O.W. and Ernst, R.R.: "Two-Dimensional correlation of connected NMR transitions". *J. Am. chem. soc.* (1985), 107:6394-6396.

Grzesiek, S., Bax, A.: "Measurement of amide proton exchange rates and NOEs with water in <sup>13</sup>C/<sup>15</sup>N-enriched calcineurin B". *J Biomol NMR* (1993), 3:627-638.

Grzesiek, S.a.B., A.: "An efficient experiment for sequential backbone assignment of medium-sized isotopically enriched proteins". *J. Magn. Reson.* (1992), 99:201-207.

Grzesiek, S.a.B., A.: "The importance of not saturating H<sub>2</sub>O in protein NMR-Application to sensitivity enhancement and NOE measurements". *J. Am. chem. Soc.* (1993), 115:12593-12594.

Hajduk, P.J., Dinges, J., Miknis, G.F., Merlock, M., Middleton, T., Kempf, D.J., Egan, D.A., Walter, K.A., Robins, T.S., Shuker, S.B., et al.: "NMR-based discovery of lead inhibitors that block DNA binding of the human papillomavirus E2 protein". *J Med Chem* (1997), 40:3144-3150.

- Hard, T.: "NMR studies of protein-nucleic acid complexes: structures, solvation, dynamics and coupled protein folding". *Q Rev Biophys* (1999), 32:57-98.
- H, zur Hausen: Papillomaviruses and cancer: from basic studies to clinical application. *Nat Rev Cancer* 2, 342-350 (2002).
- Hegde, R. S., and Androphy, E. J. (1998) 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.
- Hegde, R. S. (2002) The papillomavirus E2 proteins: structure, function, and biology, *Annu. Rev. Biophys. Biomol. Struct.* 31, 343-360.
- Hines, C. Meghoo, S. Shetty, M. Biburger, M. Brenowitz and R. S. Hegde: DNA structure and flexibility in the sequence-specific binding of papillomavirus E2 proteins. *J Mol Biol* 276, 809-818 (1998).
- Hooley, E., Fairweather, V., Clarke, A. R., Gaston, K. & Brady, R. L. (2006). The recognition of local DNA conformation by the human papillomavirus type 6 E2 protein. *Nucleic Acids Res* 34, 3897-908.
- Hou S. Y. , S. Y. Wu and 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).
- Kasukawa, P. M. Howley and 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).
- Li, J. Knight, G. Bream, A. Stenlund and 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).
- Knight J. D., R. Li and 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 USA* 88, 3204-3208 (1991).
- Jana, R., Hazbun, T. R., Fields, J. D., and Mossing, M. C. (1998) Single-chain lambda Cro repressors confirm high intrinsic dimer-DNA affinity, *Biochemistry* 37, 6446-64523
- Lam, S. L. & Hsu, V. L. (2003). NMR identification of left-handed polyproline type II helices. *Biopolymers* 69, 270-81.
- Jeon S., B. L. Allen-Hoffmann and 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).

- Liang, H., Sandberg, W. S., and Terwilliger, T. C. (1993) Genetic fusion of subunits of a dimeric protein substantially enhances its stability and rate of folding, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 7010-7014.
- Liang, H., Petros, A. M., Meadows, R. P., Yoon, H. S., Egan, D. A., Walter, K., Holzman, T. F., Robins, T., and Fesik, S. W. (1996) Solution structure of the DNA-binding domain of a human papillomavirus E2 protein: evidence for flexible DNA-binding regions, *Biochemistry* **35**, 2095-2103
- Lima, L. M. & Prat Gay, G. d. (1997). Conformational changes and stabilization induced by ligand binding in the DNA-binding domain of the E2 protein from human papillomavirus. *J Biol Chem* **272**, 19295-303. *Oncogene*
- Massimi, D. Pim, C. Bertoli, V. Bouvard and L. Banks: Interaction between the HPV-16 E2 transcriptional activator and p53. *Oncogene* **18**, 7748-7754 (1999).
- Moran, L. B., Schneider, J. P., Kentsis, A., Reddy, G. A., and Sosnick, T. R. (1999) Transition state heterogeneity in GCN4 coiled coil folding studied by using multisite mutations and crosslinking, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 10699-10704.
- Mok, Y.-K., Prat-Gay, G. D., Butler, P. J., and Bycroft, M. (1996) Equilibrium dissociation and unfolding of the dimeric human papillomavirus strain-16 E2 DNA-binding domain, *Protein Sci.* **5**, 310-319.
- Mok, Y. K., Bycroft, M., and Prat-Gay, G. d. (1996) 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.
- Mok, Y. K., Alonso, L. G., Lima, L. M., Bycroft, M., and de Prat-Gay, G. (2000) 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.
- Moskaluk C. and D. Bastia: The E2 “gene” of bovine papillomavirus encodes an enhancer-binding protein. *Proc Natl Acad Sci U S A* **84**, 1215-1218 (1987).
- Nadra, A. D., Eliseo, T., Mok, Y. K., Almeida, C. L., Bycroft, M., Paci, M., Prat-Gay, G. d., and Cicero, D. O. (2004) 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
- Nelson, J.W., and Kallenbach, N.R. 1986. Stabilization of the ribonuclease S-peptide alpha-helix by trifluoroethanol. *Proteins* **1**: 211-217.
- Parish J. L. , A. M. Bean, R. B. Park and E. J. Androphy: ChlR1 is required for loading papillomavirus E2 onto mitotic chromosomes and viral genome maintenance *Mol Cell* **24**, 867-876 (2006).
- Pelton, J. G., Torchia, D. A., Meadow, N. D., and Roseman, S. (1993) *Protein Sci.* **2**, 543-558.

- Pepinsky, R. B., Androphy, E. J., Corina, K., Brown, R., and Barsoum, J. (1994) Specific inhibition of a human papillomavirus E2 trans-activator by intracellular delivery of its repressor, *DNA Cell Biol.* 13, 1011-1019.
- Prat-Gay, G., Nadra, A. D., Corrales-Izquierdo, F. J., Alonso, L. G., Ferreiro, D. U., and Mok, Y. K. (2005) The folding mechanism of a dimeric beta-barrel domain, *J. Mol. Biol.* 351, 672-682
- Robinson, C. R., and Sauer, R. T. (1996) Covalent attachment of Arc repressor subunits by a peptide linker enhances affinity for operator DNA, *Biochemistry* 35, 109-116.
- Robinson, C. R., and Sauer, R. T. (1996) Equilibrium stability and sub-millisecond refolding of a designed single-chain Arc repressor, *Biochemistry* 35, 13878-13884.
- C. M. Sanders and 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).
- Schmidt, J. M., Thuring, H., Werner, A., Ruterjans, H., Quaas, R., and Hahn, U. (1991) *Eur. J. Biochem.* 197, 643-653.
- Shi, Z., Woody, R. W. & Kallenbach, N. R. (2002). Is polyproline II a major backbone conformation in unfolded proteins? *Adv Protein Chem* 62, 163-240.
- Shimba, N., Takahashi, H., Sakakura, M., Fujii, I., and Shimada, I. (1998) *J. Am. Chem. Soc.* 120, 10988-10989
- Sieber, M., and Allemann, R. K. (1998) Single chain dimers of MASH-1 bind DNA with enhanced affinity, *Nucleic Acids Res.* 26, 1408-1413.
- Spolar, R. S. & Record, M. T., Jr. (1994). Coupling of local folding to site-specific binding of proteins to DNA. *Science* 263, 777-84.
- Stubenrauch F., A. M. Colbert and 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).
- Sturtevant, J. M. (1977). Heat capacity and entropy changes in processes involving proteins. *Proc Natl Acad Sci U S A* 74, 2236-40.
- Sudmeier, J. L., Ash, E. L., Günther, U. L., Luo, X., Bullock, P. A., and Bachovchin, W. W. (1996) *J. Magn. Reson. Ser. B* 113, 236-247.
- Tan S. H., L. E. Leong, P. A. Walker and 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).

- Thain, K. Webster, D. Emery, A. R. Clarke and 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).
- Thierry F. and 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).
- Tiffany, M.L., and Krimm, S. 1972. Effect of temperature on the circular dichroism spectra of polypeptides in the extended state. *Biopolymers* 11: 2309-2316.
- You, J. L. Croyle, A. Nishimura, K. Ozato and P. M. Howley: Interaction of the bovine papillomavirus *Cell* 117, 349-360 (2004).
- Xia, B., Cheng, H., Skjeldahl, L., Coghlan, V. M., Vickery, L. E., and Markley, J. L. (1995) *Biochemistry* 34, 180-187 E2 protein with Brd4 tethers the viral DNA to host mitotic chromosomes
- Van Dijk, A. A., Scheek, R. M., Dijkstra, K., Wolters, G. K., and Robillard, G. T. (1992) *Biochemistry* 31, 9063-9072.
- 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 and N. Munoz: Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. *J Pathol.* 189, 12-19 (1999).
- Wishart, D. S. & Sykes, B. D. (1994). Chemical shifts as a tool for structure determination. *Methods Enzymol* 239, 363-92.
- Wu S. Y. , A. Y. Lee, S. Y. Hou, J. K. Kemper, H. Erdjument-Bromage, P. Tempst and C. M. Chiang: Brd4 links chromatin targeting to HPV transcriptional silencing. *Genes Dev* 20, 2383-2396 (2006).
- Vuister, G. W., Delaglio, F. & Bax, A. (1992). An empirical correlation between <sup>1</sup>JCaHa and protein backbone conformation. *J. Am. Chem. Soc.* 114, 9674-5.
- Wiseman, T., Williston, S., Brandts, J. F. & Lin, L. N. (1989). Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Anal Biochem* 179, 131-7.
- Zuiderweg, E. R. P. (1990) *J. Magn. Reson.* 86, 3

## **CURRICULUM VITAE**

**Name:** Riccardo  
**Surname:** Melis  
**Age :** 28 years old  
**Born in :** Milan (MI) on April 14, 1980  
**Nationality:** Italian  
**Social Condition:** Single  
**Gender :** M  
**Adress1 :** Via Cagliari 76,  
Cap 09019  
Teulada  
(CA)  
**cell :** 3334898391  
**email :** [rmriccardomelis@gmail.com](mailto:rmriccardomelis@gmail.com)

### **EDUCATION AND RESEARCH EXPERIENCE:**

#### **-July 2007- January 2008**

Stage as visiting PhD in the laboratory of Estructura-Función Ingeniería de Proteínas of Dr. Gonzalo De Prat Gay of Instituto Leloir Fundacion of Buenos Aries (Ar) ( <http://web.leloir.org.ar/> )

Project title: "Biochemistry and Molecular Biology Caraterization of Human Papilloma Virus Proteins to discover potential antiviral targets "

#### **-January 2007**

Partecipation to Second EuCheMs School on Protein Chemistry to Alba di Canazei (IT) organized by University of Verona.

#### **-January 2005- Settember 2006**

One year fellow in the laboratory of Prof. Maurizio Paci, at Chemical Science Department of University of Tor Vergata (RM)

#### **-Settember 2004**

Winner of 4 years Post-Degree Fellowship inded by P.C, B.C., I, S.&S Assessorate of Autonomy Sardegna Region. (<http://www.regione.sardegna.it/regione/assessorati/istruzione>)



**-Settembre1999- May2004**

Degree in Biological Sciences awarded on May 24, 2004 with the final valuation of 110/110 and mention of lode c/o the Faculty of Mathematics and Natural Physics of the University of Cagliari (Ca)  
Tesis Title: “ Purification and Characterization by <sup>129</sup>Xenon NMR Spectroscopy of myoglobins “

**Computer Skills:**

- Operative systems: Windows (95, 98, XP), GNU Unix, Linux and Machintosh.
- Languages:  
(Dos, HTML), Explorer (up to v6.0
- Programs:  
MS Office (Word, Excel, PowerPoint, And Outlook), Adobe( Reader, Photoshop, Illustrator)
- Scientific Programs:  
Mathematica, Kaleidagraph, Profit, InsightII, Pymol, MOLMOL.

**SPOKEN LANGUAGES:**

- Italian (mother tongue)
- English (medium)
- Spanish (medium)

**PUBLICATIONS LIST:**

1)

J Biomol NMR. (2006) ;**36** Suppl 1:73.Aug 24.

*"Backbone NMR assignment of the 29.6 kDa rhodanese protein from Azotobacter vinelandii"*.

Gallo M., Melino S., Melis R., Paci M. and Cicero D.O.

2)

Spectroscopy (2006) Volume **20**, Number 4, pp 153 - 167

*"Single scan TROSY and E.COSY suite of experiments for the measurement of residual dipolar couplings in proteins"*

Tommaso Eliseo, Mariana Gallo, Riccardo Melis, Maurizio Paci, Renzo Bazzo and Daniel O. Cicero

**Formattato:** Italiano  
(Italia)