La Sapienza
Università degli Studi di Roma

TOMMASO ELISEO
Dipartimento di Chimica

Ph.D. Thesis, 2004

The role of conformational plasticity in the DNA-protein recognition mechanism: NMR study of the Human Papillomavirus E2 system

Supervisor: Prof. Pasquale De Santis
Dipartimento di Chimica

ABBREVIATIONS

HPV	Human Papillomavirus
HPV-16	Human Papillomavirus type 16
BPV	Bovine Papillomavirus
LCR	Long Control Region
E2C	E2 Carboxy-Terminal DNA-binding domain
E2-BS	E2 binding site
BS(ACGT)	dodecameric E2 binding site with sequence ACCGACGTCGGT
BS(ACGT)	dodecameric E2 binding site with sequence ACCGAATTCGGT
BS(AAAT)	octadecameric E2 binding site with sequence GTAACCGAAATCGGTTGA
NMR	Nuclear Magnetic Resonance
C.S.	Chemical Shift
NOE	Nuclear Overhauser effect
ROE	Rotating-frame Overhauser effect
NOESY	Nuclear Overhauser effect spectroscopy
ROESY	Rotating-frame Overhauser effect spectroscopy
RDC	Residual Dipolar Coupling
SA	Simulate annealing
T_1	Longitudinal relaxation time constant
T_2	Transverse relaxation time constant
PFGLEDSTE	Pulse Field Gradient Longitudinal Stimulated Echo

INTRODUCTION

The rapid and successful conclusion of the genoma project led to knowledge of the gene map in the human specie. This enormous amount of information requires further theoretical and experimental investigation in order to elucidate the mechanisms by which the linear genetic code is selectively expressed in response to intrinsic, specific cell life cycle and to internal and external signals. The molecular biology of the cell is mediated by an intricate network of intermolecular interactions which rely on the dynamic recognition of proteins and nucleic acids. A promising area of investigation is the conversion of gene sequence information into three-dimensional structural information of specific DNA elements and their cognate binding proteins. The structural biology of DNA-protein recognition (in conjunction to proteomics, the structural investigation of protein-protein interactions) could give fundamental contributions to the comprehension of the fine details of basic biological processes.

DNA-binding proteins represent efficient machines exquisitely sensitive to differences in target sequence while the thermodynamic and kinetic features of the interaction are markedly affected by changes in the environment, allowing a fine modulation and a tight control of their activity. Theoretical models and experimental works are providing new insights into the nature of protein-DNA binding in order to obtain a structural code able to explain and predict the behaviour of the biomolecules involved in the recognition. Binding of DNA to a protein is often characterized by conformational rearrangements in the nucleic acid, in the protein or in both. With this regard, flexibility and adaptability are fundamental structural properties which facilitate the process of recognition and increase the specificity of interaction. The specificity of protein-DNA reactions derives from a balance of several factors. The lack of a simple correspondence

between amino acids and DNA bases in direct readout (direct contacts of specific functional groups) makes a general prediction of protein-DNA affinity very problematic. On the other side, variations due to changes in sequence-dependent conformation appear partially predictable by analysing the high-resolution structures of DNA oligomers and their global and local mechanical properties. Different models were proposed to account for the contribution of the indirect effects of DNA structure to the binding (indirect readout). The biological importance of these investigations is underlined by the fact that small differences in the affinity or kinetics can profoundly affect critical regulatory decisions.

Papillomaviruses infect epithelial tissues and some strains are cancer-associated. Besides of their healthy relevance, Papillomaviruses are adopted as a good model system to describe the fine modulation of gene expression. The E2 protein is the unique transcription factor of Papillomavirus genome and is also necessary for its replication. The E2 protein occupies few specific sites on the viral genome in a defined order to activate or repress transcription depending on the virus life cycle and on the cellular context. It's remarkable the ability of this protein to discriminate between binding sites on the basis of the intrinsic curvature and the flexibility of its target DNA sequences. The E2 protein is an optimal system to study which stereochemical, mechanical and dynamical factors prevalently affect protein-DNA recognition. In addition, Papillomaviruses coevolved together with their vertebrate hosts and a structural analysis of E2 proteins from different strains provides a good approach to understand structural and functional coevolution.

NMR techniques and methodologies can play an important role in unravelling the contributions to the structural code of DNA recognition, enlightening the dynamical aspects of this kind of process. Since NMR applications were used to determine the solution three-dimensional structures of proteins at atomic level, a continuous and

impressive development of instrumentation and techniques extended the range of applicability of this discipline in structural biology. Furthermore, the greater progresses and results arose from the use of new methodologies based on multidimensional, heteronuclear experiments that measure magnetic interactions between nuclei of isotope-enriched macromolecules. The theoretical treatment of NMR phenomena established strong correlations between experimental parameters and the dynamic properties of the macromolecule with a degree of accuracy and spatial resolution rarely accessible by other techniques. This development enhanced the possibility to conduct detailed studies on the stereochemical determinants in protein-DNA interactions and to assess the contributions of the dynamic nature of the biomolecular conformations.

E2 PROTEINS

Papillomaviruses

Paillomaviruses are a family of small, double-stranded DNA viruses that infect with high specificity the epithelial and mucosal tissues of a wide variety of host species (for good and brief reviews on Papillomaviruses and E2 proteins see Burd, 2003; Dell and Gaston, 2001; Hegde, 2002). Over 100 strains that infect humans were identified to date. Human Papillomavirus (HPV) strains can be subdivided into low-risk types and high-risk types. The first class (the vast majority) causes benign hyperproliferative lesions commonly known as warts, while the high-risk types are associated with tumours of the ano-genital tract. Particularly important is the incidence of cervical cancer in women, the first cause of female death in Developing Countries (where little or no screening of viral infection is carried out) and the most diffuse cause of women cancer-associated death in Western countries. HPV DNA is detected in virtually all cervical carcinomas and the most common high-risk viral types are HPV-16 (50 %), HPV-18 (15 %), HPV-45 (8%) and HPV-31 (5%). HPV infection is widely spread in women and alone is not sufficient for the development of tumours but it's demonstrated to be prerequisite. Genital HPVs infect the basal layer of the cervical epithelium but for efficient viral replication the virus needs to migrate down through the different epithelial layers, so infecting the terminally differentiated cells of the inner cervical tissue (figure 1).

In cervical warts HPV DNA is normally present as episomal DNA (episomes are supercoiled, closed DNA molecules) and is not integrated into the host cell genome. When these benign tumours are transformed into malignant carcinomas, HPV DNA is frequently integrated into chromosomal DNA.

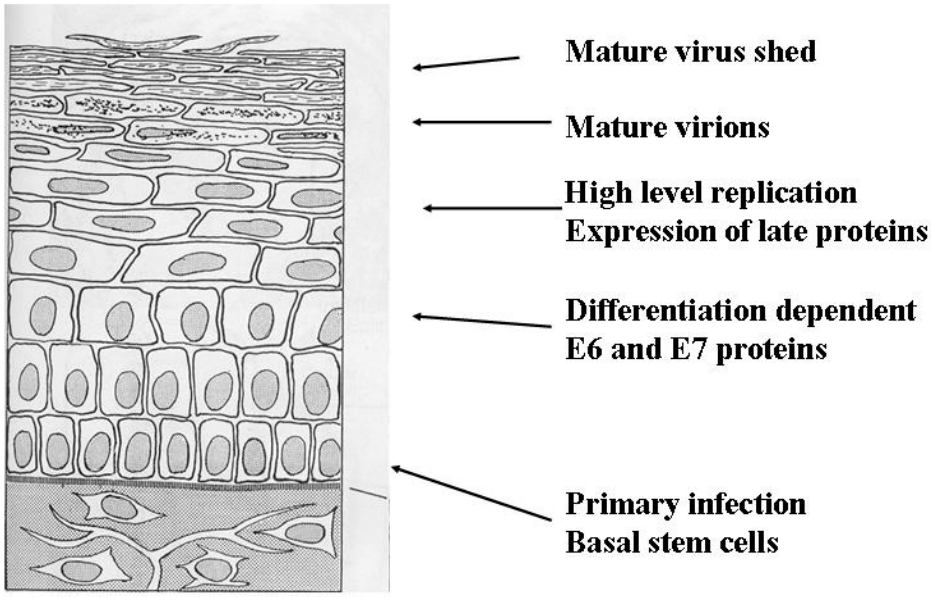

Figure 1. Scheme of the relationship between Papillomaviruses life cycle and cellular differentiation state of the infected epidermal tissue.

The viral genome is a 8-kb circular, double stranded DNA. The Open Reading Frame (the region that codes for proteins) encodes for at least eight early (E1, E2, etc.) and two late proteins (L1 and L2). As in many other viruses, the early proteins are expressed in the early stages of the infection and have the function of promoting infection, regulating expression and replication and activating defensive mechanism to escape the host organism control. The late proteins are expressed in a successive phase of the infection and are assembled to constitute the viral capside. Different variants of these gene products (proteins arising from splicing or fusion of viral RNAs) were also individuated. It's noteworthy but not rare that transcription and replication of the viral DNA are performed by host cell machinery but are directed and modulated by viral proteins. In addition to cellular factors, Papillomavirus DNA replication requires the viral E1 protein and the E2 protein as activator. A region of the viral genome of about 800 bp, known as the Long Control Region (LCR), harbours the regulatory elements of the viral transcription and modulates the expression in response to physiological and tissue-specific signals. The

regulatory viral E2 protein controls the transcription of all viral genes by binding to specific sequences in the LCR. The viral genes responsible for the malignant transformation are E6 and E7, which encode proteins that are found at high levels in cervical tumours and are supposed to destabilize the control system of the cell. Specifically, E6 blocks the defensive activity of the cellular tumour suppressor p63 (a protein that activates apoptotic pathways and is implicated in over 50% of all human cancers!) by complexing and targeting it to degradation. The oncoprotein E7, interacting with various cellular partners, promotes unchecked cell differentiation and proliferation.

Because link between HPV infection and cervical cancer is well established, cheap and simple diagnostic kits for the screening of the viral presence are now available and periodic checks are encouraged in women population. Although there is no virus-specific treatment currently available, the screening programs and the advances in the treatment of pre-cancerous warts have considerably reduced the risk of cancer in Western Countries. HPV infection remains a serious health problem for women in Developing Countries. A vaccine could be an ideal solution but HPV infection by one strain is often accompanied by infection of other strains that are not sensitive to the same vaccine. All the viral proteins are subjected to extensive studies in order to find specific antiviral agents.

E2 protein biology

Transcriptional regulation. The Papillomavirus E2 protein is a multifunctional protein which plays a central role in controlling key processes of the viral life cycle and has important implications in the progression of HPV-induced tumours (Bedrosian and Bastia, 1990; Burd, 2003; Dell and Gaston, 2001; Hegde, 2002). This protein is the unique viral transcription factor. Interacting with DNA regulatory elements of the Long Control

Region and with various cellular factors, E2 regulates the transcription of all the genes encoded in the Open Reading Frame of the Papillomavirus genome.

In Bovine Papillomavirus type 1 (BPV-1) the full-length E2 gene product is a transcriptional transactivator which binds to specific enhancer elements. These E2 Binding Sites (E2-BSs) are specific target sequences located proximal to important viral promoters in the LCR. By binding to the E2-BSs, E2 activates the transcription of early genes. BPV-1 also produces two shorter forms that lack the E2 N-terminal domain (see below) and which antagonize the function of the full-length transactivator. These truncated E2 proteins, called E2-TR and E8^E2C, can repress viral transcription by binding to the E2-BS in the BPV-1 genome regulatory regions and by formation of heterodimers with the full-length E2 transactivator.

There is no experimental evidence that human Papillomaviruses express truncated E2 repressor proteins. Therefore, there is another way by which E2 proteins exert a fine regulation of important biological processes, as evidenced by numerous and different studies (see references in: Dell and Gaston, 2001; Hegde, 2002). The LCR in high-risk HPVs is located upstream to the major promoter of viral early genes and contains four binding sites for the E2 protein as well as numerous binding sites for cellular transcription factors. In the absence of viral gene products, HPV gene transcription is activated by the binding of these host cell transcription factors to specific sequences near the major promoter in the LCR. The basal activity of this promoter can be further modulated by binding of viral E2 protein to specific target sequences. The number and disposition of the four E2 binding sites locations are conserved between all the high-risk HPV types, as illustrated in figure 2, suggesting that their precise arrangement is fundamental for viral expression and replication. In facts, transcriptional regulation depends on the position of E2-BSs with respect to the proximal promoter elements.

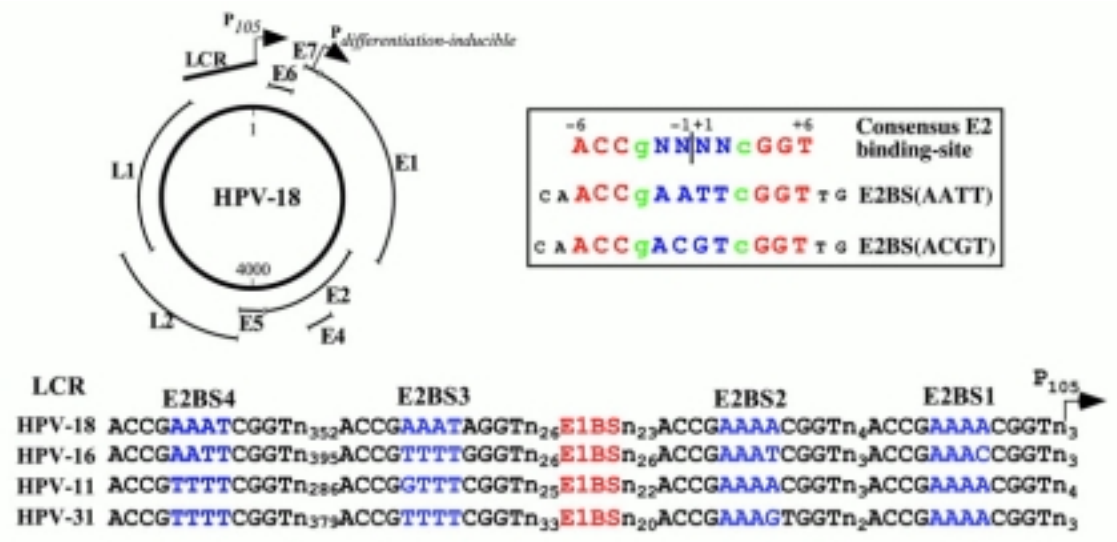

Figure 2. The location of Early and Late genes on Human Papillomaviruses circular genome are displayed in the upper-left part of the figure. The consensus sequence for DNA-binding E2 protein is shown in the upper-right part of the figure. In the lower part the arrangement in the regulatory Long Control Region of the four E2 binding sites is presented for human cancer-associated strains. The E2-BSs are always found proximal to the major promoter P_{105}.

Whether E2 activates or represses promoter activity depends on the relative affinity of E2 for each of its four binding sites and the amount of E2 within the cell. Cosequently, it's the relative occupancy of E2 binding sites as the concentration of E2 protein varies during the virus life cycle that dictates which regulatory events take place. In vitro binding experiments have shown that E2 proteins from human strains bind with the highest affinity to BS4, with lower affinity to BS1 and BS2, and much less to BS3 (Bedrosian and Bastia, 1990; Hines et al., 1998; Kim et al., 2000; Zhang et al., 2004).

Deregulated expression of E6 and E7 oncogenes leads to cell transformation and eventually to cancer. Thus a fine balance between activation and repression of the E6/E7 promoter is necessary. Chromosomal integration of the HPV genome commonly results in the disruption of the viral E2 Origin Reading Frame and, consequently, loss of the E2 protein activity. Inactivation of E2 protein causes the deregulated expression of the oncogenes. An optimal temporal regulation of viral expression, relying on the ability of E2 to select among the available binding sites, is required to maintain the best conditions for

an efficient and unperturbed viral replication. In addition, a direct role of E2 in tumourigenesis, independent of regulation of E6 and E7 activity, is also documented and implies interactions with various cellular transcription factors.

Activation of replication. Although HPV DNA is copied by the host cell machinery, replication requires the E1 and E2 proteins. The E1 protein has several replication-associated activities (act as an Origin-Binding Protein and as an helicase) but binds poorly to the viral Origin of Replication in absence of E2. In facts, the HPV origin of replication contains a binding site for E1 flanked by the mentioned E2-BS1 on one side and by E2-BS2 on the other side. The relative positions of the specific binding sites in the replication origin element are highly conserved. The function of E2 in replication is to enhance the affinity of E1 for its binding site by means of cooperative binding. Once the E1-E2-DNA complex is formed, E1 is ready to recruit proteins of the host cell machinery to initiate replication. The interaction of E2 with E1 is mediated by its N-terminal transactivation domain.

Other functions. The DNA binding activity of E2 protein can be modulated by cellular transcription factors that interact with the E2 N-terminal transactivation domain. In this way the transcriptional regulation can respond to changes in the cellular environment. The N-terminal domain is also supposed to mediate the cooperative binding of E2 to DNA molecules containing multiple E2 binding sites and to promote the formation of stable DNA loops, a mechanism by which E2 proteins bound to distant DNA elements can strongly interact. In addition, E2 protein is able to associate with chromosomal DNA in processes linked to DNA architecture organization. A role in the induction of apoptotic signals was also found for E2 in HPV-transformed cells and this subject is investigated for therapeutic purposes .

E2 protein structure

The full-length E2 protein is a nuclear phosphoprotein which possesses a modular and dimeric nature like many families of DNA-binding regulatory proteins (Dell and Gaston, 2001; Hegde, 2002). The modules represent independent, structural domains associated with specific functions that were selectively mixed and assembled during the structural-functional genetic evolution. The modular architecture in transcription factors is thought to confer large functional flexibility, necessary for such "promiscuous" molecules to be promptly sensitive to environmental changes. The dimeric association, which is often related to the palindromic or pseudo-palindromic sequence of DNA recognition sites, increases the affinity and the specificity of interaction and contributes to modulate DNA-binding activity. E2 exists in solution and binds to DNA as a stable dimer. The K_D value measurements fall within picomolar range (Mok et al., 1996), well above the binding constants for E2-BSs, indicating that E2 is dimeric also in absence of DNA (many transcription factors exist in monomeric form in solution and dimerize only upon DNA binding). Each monomer is composed of two conserved, structured domains (a N-terminal domain of about 200 amino acids and a C-terminal domain of about 100 amino acids) linked together by a central, flexible region whose length varies in the different Papillomavirus strains. A representation of E2 modular organization is shown in figure 3.

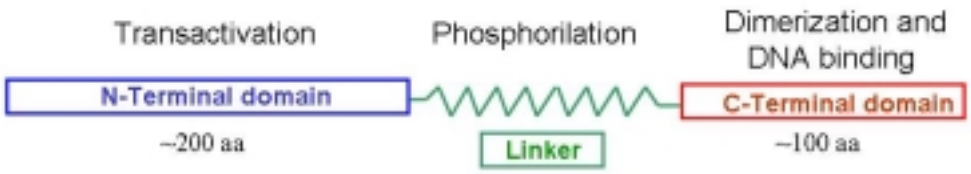

Figure 3. Schematic representation of the modular architecture of the full-length E2 protein.

The transactivation domain. The N-terminal module mediates E2 intaction with viral

E1 protein and several cellular factors. The transactivation domain is dimeric although its micromolar K_D doesn't account for the high dimerization energy of the entire protein. Unlike many other transactivation modules which are commonly disordered, the highly constrained conformation of the N-terminal E2 module permitted to determine its high-resolution structure by X-ray crystallography (Antson et al., 2000). The protein exhibits a global L shape and consists of two domains, an antiparallel β-sheet and an helicoidal domain composed of three anti-parallel α-helices forming a twisted plane. A systematic mutational analysis shows that some conserved amino acids located at the dimerization interface and on the outer surface have importance in the transactivation function and in the formation of hetero- and homodimers.

The linker region. Little structural information is available about the region of the full-length polypeptide chain that connects the ordered N-terminal and C-terminal domains. This segment, which varies both in length (40-200 amino acids) and in composition among different strains, possesses high content of proline residues. It was supposed that this central region doesn't adopt a structured conformation and acts as a flexible "hinge" allowing to transmit structural information between the other two domains. Biochemical data suggest additional roles for this region in phosphorilation-dependent modulation of E2 activity and in conferring greater stability to the protein.

The C-terminal DNA-binding domain. The specific and non-specific DNA-binding activity of E2 protein is localized in the C-terminal module, which is also responsible for the dimeric state. The minimal domain required for DNA-binding comprises about 80 amino acids and will be thereafter termed as E2C. The high-resolution structure of E2C from BPV-1, determined in 1992 by Hedge and co-workers using X-ray crystallography (Hegde et al., 1992)), reveals the peculiar fold that characterizes all E2C structures solved so far (listed in table I together with method of determination, corresponding references

and Protein Data Bank codes). The E2C structure is the prototype for the *dimeric β-barrel* structural class, a kind of fold shared to date only by EBNA1 (Epstein-Barr Nuclear Antigen 1, which is the Epstein-Barr virus replication protein), which is another viral DNA-binding protein (Bochkarev et al., 1995).

Structure	***PDB code***	***Method***	***Resolution*** (Å)	***Reference***
BPV-1	1JJH	RX	2,5	Hegde et al., 1992
BPV-1	1DBD	NMR	n/a	Veerararghavan et al., 1999
BPV-1 complex	2BOP	RX	1,7	Hegde et al., 1998
HPV-18	1F9F	RX	1,9	Kim et al., 2000
HPV-18 complex	1JJ4	RX	2,4	Kim et al., 2000
HPV-31	1A7G	RX	2,4	Bussiere et al., 1998
HPV-31	1DHM	NMR	n/a	Liang et al., 1996
HPV16	1BY9	RX	2,2	Hegde and Androphy, 1998
HPV-6	1R8H	RX	1,9	Dell et al., 2003
Site ACCGAATTCGGT	1ILC	RX	2,2	Hizver et al., 2001
Site ACCGACGTCGGT	423D	RX	1,6	Rozenberg et al., 1998

Table I. The various structures solved at atomic resolution for E2C protein and E2 binding sites are listed together with corresponding PDB codes, method of determination and reference.

Figure 4 represents the topology and quaternary arrangement found in E2C fold. The structure of E2C dimer is an eight-stranded antiparallel β-barrel composed of two four-stranded half–barrels from each monomer which oppose in such a way to form a continuous β–sandwich. The topology of the monomer is β1–α1–β2–loop–β3–α2–β4. The two helices lie on the outer surface of the β-sheet. The first and longer helix is termed the "recognition helix" because houses the amino acids that make direct contacts with the DNA target. The segment inserted between β2 and β3 strands appears as an elongated, disordered and flexible loop. Greatest conservation between strains is found in the recognition helix and at dimer interface.

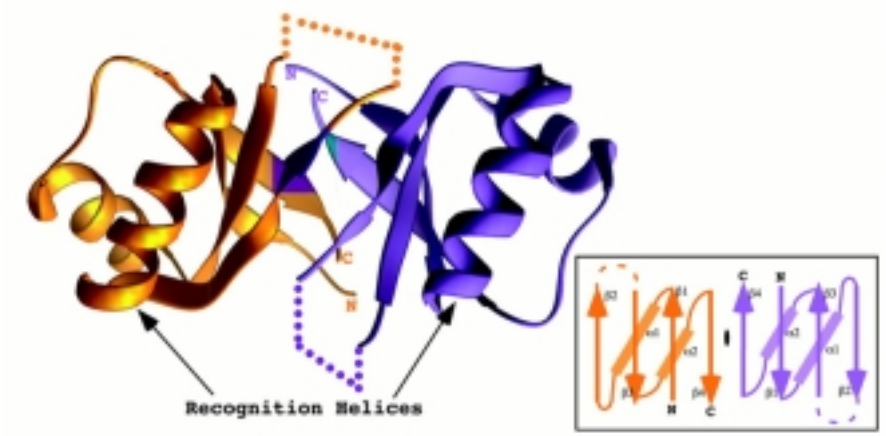

Figure 4. Ribbon representation of E2C dimer. The offset shows the monomer topology.

The two subunits are held together by a continuous hydrogen-bonding network between strands β2 and β4 in one monomer with their symmetric counterparts in the other monomer and by numerous hydrophobic interactions in the core of the dimeric barrel, resulting in significant stability for the dimer (10^{-13} M). The strict interdependence between this extensive network of interactions at dimer interface is reflected into the high cooperativity degree revealed in the chemical and acid unfolding experiments at equilibrium (Mok et al., 1996a; Mok et al., 1996b; Mok et al., 2000), suggesting that dimerization is coupled to the folding mechanism.

The monomers of E2C proteins whose structure has been solved possess identical topology and the same tertiary structural elements, resulting in very similar overall architecture. There is anyway difference in the quaternary structure as evidenced by a little but important variation in the orientation between the two subunits. E2C proteins can be divided into two structural subfamilies depending on the relative orientation of the subunits (Liang et al., 1996; Kim et al., 2000; Hegde, 2000). The first group includes BPV-1, HPV-18 and non-oncogenic HPV-6 and thereafter will be called the BPV-1 family, whilst

HPV-16 and HPV-31 constitute the second group (called the HPV-16 family). As it is evidenced in figure 5, when only one monomer is superimposed between the various structures, the orientation of the non-superimposed monomer in BPV-1 family differs from the orientation in the HPV-16 family.

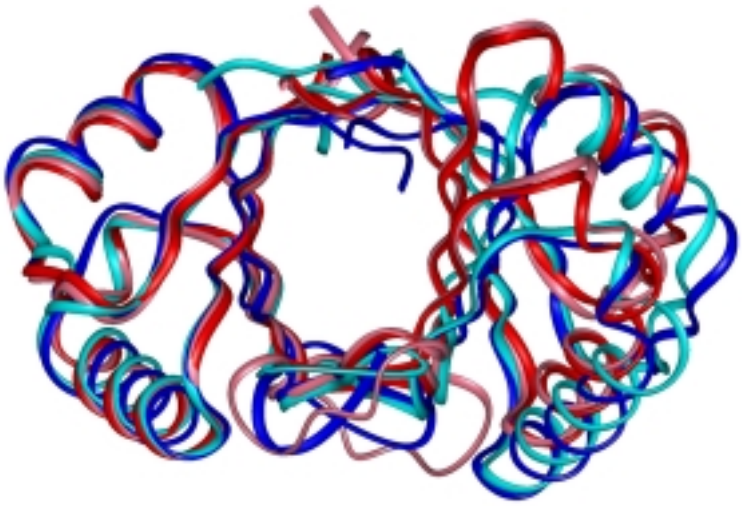

Figure 5. Superimposition of HPV-16, HPV-31, HPV-18 and BPV-1 dimeric structures (see table I for PDB codes). Only the backbone atoms of one subunit (subunit on the left) are superimposed. The coil representations are differentially coloured for the distinct strains.

This difference in quaternary disposition of subunits is thought to have important implications for DNA recognition and for the characteristic discriminatory ability displayed by E2 proteins, because it is reflected in the relative orientation of the two symmetrically disposed recognition helices. The recognition helix of the second monomer in BPV1 family, when only the first monomer is superimposed, is shifted away by approximately 4 Å and tilted by 25° relative to the recognition helix in the HPV-16 family. As discussed below, the precise placement of recognition helices determines the extent of DNA deformation required to optimise protein-DNA contacts in the various complexes formed by different E2C proteins with E2 binding sites.

The difference in quaternary structure of the two classes is attributed to a shift in the alignment of the opposing and dyad-symmetric β4 strands. The hydrogen-bonded residues of these C-Terminal strands in BPV-1 and HPV-18 are out of register relative to the hydrogen-bonded residues in HPV-16 and HPV-31, as shown by figure 6. The change in the β4 strands alignment propagates up to the recognition helices, which result displaced and tilted when comparing the two structural classes of E2C proteins.

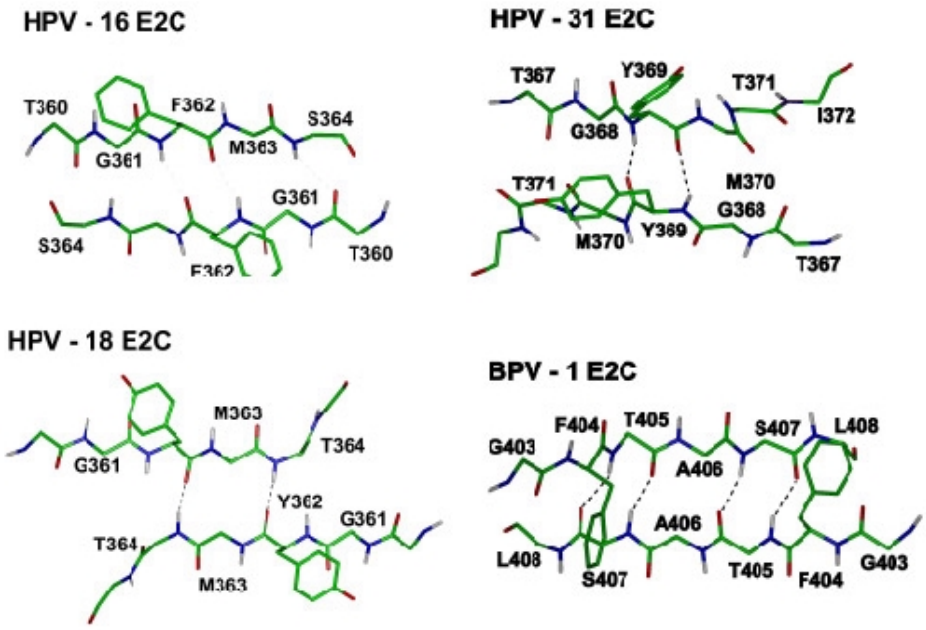

Figure 6. Detail of the β4 strands of the four structures. In HPV-16 and HPV-31 E2C, the conserved aromatic residue in one subunit aligns with the same residue in the other. On the other hand, in HPV-18 and BPV-1 the corresponding aromatic residues are shifted. Hydrogen bonds are depicted as broken lines.

Solvent exchange rate measurements performed by NMR on HPV-31 and BPV-1 E2C proteins suggested a local "breathing" of recognition helices, not individuated in the rest of the molecule. These regions are supposed to undergo dynamic equilibrium between hydrogen-bonded helicoidal conformation and locally solvent unprotected structure (Liang

at al., 1996; Veeraraghavan et al., 1999: Hedge, 2002).

E2 binding sites

E2 proteins from all viral strains recognize and bind to the dodecameric consensus sequence ACCGgN_4cGGT (upper case indicate absolutely conserved bases and lower case indicate highly preferred bases), which is called the E2 binding site or E2-BS (Bedrosian and Bastia, 1990; Dell and Gaston, 2001; Hegde, 2002). The central N_4 tract represents an element with absolutely conserved length but with variable composition and sequence and is referenced as the "spacer". E2 binding sites are present at several positions within the viral genome and their exact number and location varies in the different virus types. For example, BPV-1 genome contains seventeen E2-BSs and high-risk Human Papillomavirus genomes house four E2-BSs. These highly conserved regulatory elements are found proximal to the main promoters and to the Origin of Replication (see figure 2): specific binding of E2 protein is a critical event for transcriptional regulation and for initiation of replication of viral genome. E2 protein presents high specificity and recognizes its binding sites in a vast excess of non-specific DNA. In addition, E2 protein displays the discriminatory ability to detect between multiple binding sites that differ only in the central N_4 spacer. The preferential binding of E2 to its target sequences in the viral genome results in differential occupancy of E2-BSs during the virus life-cycle. The E2 target sequences vary in their affinity toward the protein depending on the identity of the base pairs in the spacer region.

The high-resolution structure of dodecameric ACCGACGTCGGT, hereafter referred as BS(ACGT), and ACCGAATTCGGT, hereafter referred as BS(AATT), E2 binding sites were determined by crystallographic studies (Rozenberg et al., 1998; Hizver et al., 2001) and a detailed comparison between them was carried out by authors. The structure of BS(ACGT) is characterized by a conformational variant of B-DNA in which the overall

double helix axis appears straight, but local axes are not parallel and continuously precess around the global axis, so that their cumulative effects cancel out with regard to global bending of the molecule. The prominent features of this conformation relative to B-DNA are small positive deviations of roll along all the sequence and modest variations of major and minor groove dimensions.

In contrast, BS(AATT) is shown by its crystal structure to possess overall static curvature, partitioned along the entire sequence and quantified in about 10° per helical repeat. The bending points towards the minor groove in the spacer region and towards the major groove in the flanking symmetric conserved half-sites. The duplex presents important deviations of helical parameters and of groove dimensions with respect to standard B-DNA. The bend is achieved by compressing the minor groove at the central spacer region (by negative rolling) and, on the contrary, compressing the major groove at the flanking conserved elements (by positive rolling). Therefore the correlated deformations of the central and the flanking tracts of the duplex accumulate to produce global curvature. The authors, by examining the two structures, conclude that there is a strict relationship between sequence-dependent deformation or deformability and the differential affinity displayed by HPV E2 protein. They add that mechanical and conformational features inherent to DNA specific sequence may constitute a structural code for indirect readout in protein recognition (Rozenberg et al., 1998; Hizver et al., 2001).

DNA-E2C Complexes

HPV16 E2 exhibits very high distinctive sequence preference and its affinities for E2-BSs vary over a 300 fold range (Bedrosian and Bastia, 1990; Hines et al., 1998; Hegde,

2002; Zhang et al., 2004). BPV-1 protein displays little differential affinity and is less sensitive to differences in the spacer sequence. As descrbed above, HPV-16 and BPV-1 represent two distinct structural types of the *dimeric β-barrel* fold, differing only in the relative subunit orientation.

In order to collect structural information about the bound-conformation of both E2C proteins and DNA targets, the three-dimensional high-resolution structures of the complex were determined by X-ray crystallography for BPV-1 and HPV-18 E2C bound to two distinct DNA targets of palindromic sequence ACCGACGTCGGT and ACCGAATTCGGT (corresponding references and PDB codes are inserted in table I). A high-resolution structure of a complex formed by E2C protein from HPV-16 family is not still available. Figure 7 presents the crystallographic structure of BPV-1 E2C protein (bound to the dodecamer ACCGACGTCGGT) superimposed to the DNA-free conformation.

Figure 7. Superimposition of the DNA-bound and free E2C from BPV-1. Only the left subunit is superimposed. The binding site in the complex is the 12-mer ACCGACGTCGGT. The recognition helix of the non-superimposed monomer is displayed in cyan for the free protein and in gold for the complexed form. From Hegde et al., 1998.

A detailed comparative analysis of these crystallographic studies, also in light of the

reported structures of the same two free E2 proteins and the same two free binding sites, permits to deduce common important features. While the protein undergoes only a small adjustment upon DNA binding, a large deformation of DNA target occurs and conformational changes are substantially localized in the central spacer zone (Hegde et al., 1998; Kim et al., 2000; Hegde, 2002). The protein inserts the recognition helices into two successive major grooves and their α-helical axes project straight and aligned along the groove. Only four highly conserved amino acids make direct contacts with the DNA target and only the strictly conserved and symmetrically disposed ACCG/CGGT identity elements of the binding sites establish interactions with the protein, acting as recognition half-sites. Thus the specificity of E2-BSs recognition is preserved by few conserved DNA and protein residues that are involved in discriminating interactions. It's worth to note that neither direct nor indirect protein-DNA contacts are found in the central spacer region. Nevertheless, it's the identity of this tetranucleotide sequence not contacted by the protein that determines the differential affinity toward E2-BSs. This is a typical example of "indirect readout" mechanism which means sequence-dependent DNA recognition of specific sequences without direct interactions (Crothers, 1998; Hegde, 2002; Zhang, 2004). The base sequence-specific interactions appear to be inter-digitizated in the crystal structures, that is each amino acids forms contacts with more than a nucleotide and each base pair is involved in interactions with more than one amino acid. The comparison of the bound-DNA conformation in the four complex structures solved so far reveals that not only the protein conformation, but also global DNA trajectory and direct protein-DNA contacts are nearly the same in all the complexes. The only remarkable difference in protein bound-conformation is a disorder-to-order transition of the central β2/β3 loop, which makes few direct contacts with DNA phosphate backbone in BPV-1 but remains unstructured and non-reactive in HPV-18. No significant difference is found between low-

affinity complex (for example HPV-18 E2C bound to ACCGACGTCGGT) and high-affinity complex (the same HPV-18 E2C bound to ACCGAATTCGGT) as observed by Kim and Hegde (Hegde et al., 1998; Kim et al., 2000; Hegde, 2002). Large deformations in DNA structure are required to achieve optimal and symmetrical interactions between the recognition helices and the consensus ACCG/CGGT half-sites. The DNA encompasses the protein surface tracing a smooth curve. By a more detailed analysis, the identity elements exhibit little alteration upon binding and appear relatively straight. As a consequence, the DNA axis results severely bend (by 45°-51°) toward the protein at the central non-contacted region. The protein scaffold induces a significant bend obtained by compression of DNA grooves that face the protein (at the concave side of the molecule). The minor groove is compressed at the central zone and the major groove is compressed at the flanking conserved regions. The major difference in DNA forms of BPV-1 and HPV-18 complexes concerns with the distinct characteristic groove width that the protein imposes to the central spacer minor groove. The minor grooves in the HPV-18 complexed form results narrower by ca. 3 Å in this region with respect to the regular width found in standard B conformation for DNA.

The conclusions that can be drawn from inspection of E2C complex structures reported to date are that the DNA adapts its conformation to stereochemically complement the convex surface of the protein scaffold, which in contrast undergoes very modest modifications. The bound state of E2-BSs appear relatively independent of sequence and is essentially determined by the precise placement of the recognition helices in order to optimise interaction with the consensus recognition half-sites. E2C proteins from BPV-1 and HPV-18 are comprised in the same structural family, as described above. The HPV-16 family mainly differs in the relative orientation of the recognition helices, which are though to be the structural determinants of local and global conformational properties of

the complex. No high-resolution structural data exist for a complex of this E2C type. It's interesting to investigate if important quaternary rearrangement occurs in HPV-16 protein upon DNA binding, if its bound state resembles the one of the other E2C class or if structural changes in the recognition process are limited to DNA. In the last case, E2-BSs would adopt a distinct conformation relative to the DNAs in BPV-1 family complexes.

The detailed examination of E2C complexes shows that protein-DNA direct contacts, which are the same in the different structures, don't account for the difference in the affinity for distinct binding sites. The differential affinity is predominantly dependent on the identity of the non-contacted base pairs sequence.

It has been suggested that a direct and important role in the preferential binding mechanism could be played by the electrostatic potential energy surface, which substantially differs between E2 proteins at the DNA-protein contact region (Bussiere et al., 1998; Hegde et al., 1998; Kim et al., 2000; Hegde, 2002; Dell et al., 2003). An electropositive patch, generated by a cluster of basic amino acids, is located on the interaction surface next to the recognition helices in BPV-1 and HPV-18, while HPV-16 shows poor accumulation of positive charges near the recognition α1 helices. The distribution of positive charges in this region could be implicated in the energetic and kinetic pathway of DNA binding by E2 proteins.

Relationship between structure and affinity in E2 DNA-binding

A great deal of experimental biochemical and structural data has been accumulated in order to understand the basis of E2 differential binding. A central question to be issued concerns with the different behaviour between HPV-16 and BPV-1. First of all, E2

proteins from human Papillomavirus strain display a clear preference for binding sites containing AT-rich spacer sequences, while BPV-1 protein displays poor discriminatory capability (Bedrosian and Bastia, 1990; Hines et al., 1998; Hegde, 2002; Zhang et al., 2004). The viral genome appears parallel to these trends: E2 binding sites of HPVs genome tend to contain AT-rich spacer regions when compared to E2-BSs of BPV-1. The relative binding affinity of HPV-16 and BPV-1 toward the target sequence containing AATT at the central spacer is 300-fold relative to the reference ACGT central spacer sequence. Whereas HPV-16 binding constants vary by more than three orders of magnitude, BPV-1 affinity spans a narrow range of values.

Secondly, as evidenced by Hines and co-workers (Hines et al., 1998), the introduction of elements that increase flexibility in the DNA molecule, as the use of oligonucleotides with nicks or gaps in the central variable region, considerably impacts the DNA-binding affinity and the sequence discrimination ability of HPV-16 protein. BPV-1 and HPV-16 E2 proteins present very different sensitivity not only to changes in the sequence of the spacer bases, but also to variations in the local flexibility of this zone. BPV-1 is moderately favoured by axial flexibility. It's reasonable that an increase in flexibility could help the DNA target to assume the distorted and bend conformation required to achieve symmetric and optimal contacts of the conserved ACCG/CGGT half-sites with the recognition helices. Several studies have documented a positive correlation between increased flexibility and enhanced affinity in DNA-binding proteins that bend the DNA target that they recognize (Hines et al., 1998; Zhang et al., 2004). In contrast, HPV-16 E2C exhibits highly reduced affinity toward very flexible targets, like nicked or gapped binding sites. A high degree of flexibility in the spacer region is largely detrimental to HPV-16 E2 DNA-binding, especially toward AT-containing spacer sequences. These observations are not consistent with a simple correspondence between increased flexibility

and enhanced affinity, complicating the analysis of the HPV E2 sequence-specific binding.

As concerning the high-resolution structural comparison of bound and free E2 binding sites, Rozenberg and co-workers (Rozenberg et al., 1998) suggest that the central spacer sequence of BS(ACGT) and, in particular, its CG step (positioned at the middle of the spacer) exhibits interesting deformability because adopts opposite values of roll in distinct complexes (such as in met operator complexed to met repressor and in ATF\CREB site complexed to transcription factor GCN4). On the other hand, the conformational features displayed by BS(AATT) are shared by other sequences incorporating the same tetranucleotide motif, such as the extensively studied Dickerson-Drew dodecamer (Hizver et al., 2001). The authors conclude that sequence-specific static curvature is an inherent property of AT-rich tracts inserted in a GC-rich context. They propose that the correct spacing of elements with appropriate conformational characteristics represent a structural code in indirect readout and other recognition mechanisms.

It's a plausible hypothesis that HPV E2 proteins exhibit discriminatory preference for binding sites characterized by specific conformation. In facts. HPV E2 proteins display great preference for AT-rich sequences, which are demonstrated to possess not only higher bending and twisting flexibilities, but also sequence-specific axial static curvature (Zhang et al., 2004). Experimental biochemical and structural data and theoretical models based on statistical mechanics indicate that double helices incorporating AT-rich tracts present significant bending characterized by negative roll angles associated with compression of minor groove. These specific features are also apparent in the high-resolution structure reported for the free E2-BS of dodecameric sequence ACCGTAATTCGGT, a high-affinity site for HPV E2 proteins. In addition, the severe bend imposed by the protein in the E2C-

DNA complexes presents the same sequence-specific deformations which are amplified because of the stringent requirements on bound-DNA conformation. The protein appears to enforce the local sequence-dependent conformational propensities of the free target molecule. In a recent study Zhang have measured the magnitude and direction of bending in a series of E2 binding sites by cyclization kinetics methods together with binding affinities in the same experimental conditions (Zhang et al., 2004). A correlation was proven between differential affinity and DNA intrinsic curvature. The authors suggest that HPV-16 recognizes and preferentially binds to intrinsically curved DNA that presents bending toward the minor groove at the central base pairs, in the same direction found for the bound-conformation in the reported structures of the complex. It's plausible to hypothesize that HPV E2 displays higher affinity toward less flexible but prebent DNA that predispose the helical axis to be more easily deformed into the required direction.

NUCLEAR MAGNETIC RESONANCE IN STRUCTURAL BIOLOGY

An overview

NMR spectroscopy represents an area that presents continuous and enormous progress, whose capability of application ranges from material science studies to metabolic pathways elucidation, from the investigation of physical laws governing the properties of matter to common use in diagnostics and medicine. In particular, structural biology is one of the fields from which important benefits were gained in NMR spectroscopy due to the great interest in basic research and pharmaceutical applications. Since the first experimental observation of a protein NMR spectrum was achieved (Saunders et al., 1957), tremendous technological developments and fundamental methodological advances occurred. Innovation in magnet and probe characteristics, sophisticated electronics design, efficient isotopic enrichment procedures, and consequent ideation of complicated and specific pulse schemes have pushed further the limits in sensitivity and resolution, considerably improving the spectral quality. Automation in data analysis, computational methods and the exploration of the physical principles determinimg magnetic interactions in biomolecules have permitted the investigation of ever more complex systems. 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 kδ, although recent advances have brought further this frontier (Kay and Gardner, 1997). However, NMR is very well suitable for rapid study of interactions of small and intermediate biologically active molecules with proteins and membranes (independent of the assembly size). Moreover, NMR is the unique technique capable to obtain exhaustive information on the dynamics properties at the atomic level.

The assignment problem

For the luck of chemists, NMR is an high-resolution spectroscopy, due to the long life times in the nuclear excited state. The essence of NMR in structure determination 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. The spread use of NMR in conformational studies relies on the possibility to assign spectral-resolved signals to the atoms that generated those signals. The assignment is a delicate and necessary phase that constitutes a prerequisite in every NMR structure study. Only when it was possible to adopt systematic assignment procedures for proteins and nucleic acids, NMR became a powerful tool in structural biology.

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 wich 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, $^{1}J_{CH}$ (125-160 Hz), $^{1}J_{CN}$ (12 Hz) and $^{1}J_{NH}$ (about 92 Hz), are sufficiently large and relatively uniform, depending weakly on conformation. On the other hand, transverse relaxation rates for heteronuclei are much smaller than proton rates for globular proteins below the 25-30 kδ size. These features permit to efficiently transfer magnetization to ^{15}N and ^{13}C nuclei. 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. For example, the HNCO experiment correlates H^N and ^{15}N chemical shifts of one residue with the ^{13}C carbonyl chemical shift of the preceding residue. The HNCA experiment presents a correlation peak of H^N and ^{15}N chemical shifts with the $^{13}C^{\alpha}$ chemical shift of the same residue and with the $^{13}C^{\alpha}$ chemical shift of the preceding residue (see sheme in figure 8).

In the names of three-dimensional experiments brackets indicate a nucleus through which magnetization transfer occurs but whose chemical shift is not measured neither correlates with other nuclei. For example, in the HN(CO)CA experiment a correlation peak is seen between H^N and ^{15}N chemical shifts of one residue and $^{13}C^{\alpha}$ chemical shift of the preceding residue (but not $^{13}C^{\alpha}$ chemical shift of the same residue). The carbonyl nucleus is used only to transfer magnetization between the NH group and $^{13}C^{\alpha}$ nucleus of the preceding residue by one-bond transfer steps.

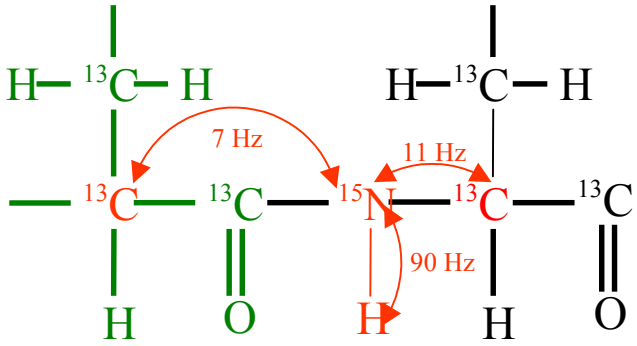

Figure 8. A dipeptide fragment is represented. The i-1 residue is drawn in green, the i residue is drawn in black. Nuclei whose chemical shift evolves and is measured in HNCA experiment are displayed in red. The scalar interactions are expressed in Hertz and are depicted as red curved arrows. Transverse 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 $^{13}C^{\alpha}$ nucleus of the same residue (residue i) or to the C^{α} nucleus of the preceding residue (residue i-1) because scalar interaction magnitude for the two nuclei is similar (11 and 7 Hz).

Stretches of sequentially correlated resonances are found for protein backbone atoms resonances. 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.

The structure determination process

Once the chemical shifts of the protein resonances are correctly assigned, site-specific structural information can be derived from different NMR measurements. The main source of geometric information is contained in the Nuclear Overhaiser Effect (NOE). Heteronuclear editing of NOESY spectra (^{13}C- and ^{15}N-separated NOESY) provides distance restraints. For an isolated two spin system and at short mixing times, the NOE depends on the inverse of the sixth power of the distance between the two nuclei involved in the dipolar interaction. This relationship is complicated and masked by spin diffusion (indirect magnetization transfer in multispin systems) and by internal dynamics effects. As a consequence, NOE intensities are converted into approximate interproton distances which are provided with an upper and lower limit value. Another common source of geometric information is represented by the vicinal scalar coupling and therefore three-bond coupling constants measurements are widely performed. The relationship between ^{3}J scalar coupling and the dihedral angle between the two atoms is described by the Karplus equation which depends on the nature of the atoms involved. The Karplus relationship is parameterized for different fragments in proteins and nucleic acids. The experience in NMR structure determination from many groups teaches that the major determinant in obtaining high-resolution and good quality structures is the number of accurate restraints, which results in efficient conformational restriction. Therefore, in the last years interest was brought to a series of other NMR experimental quantities (Clore and Gronenborg,

1998). Residual dipolar couplings are now routinely employed in many laboratories. Secondary structure elements are identified by established empirical correlations between chemical shifts of some kind of nuclei and secondary structure in addition to the individuation of slowly exchanging amide protons. Chemical shift stretches along the sequence can be searched in structure database and used directly in the refinement process. Paramagnetic effects on chemical shift and relaxation rates provide useful long-range information in some kinds of metal-containing proteins.

The structural information extracted from NMR experimental measurements are employed in structure calculations based on distance geometry methods or in restrained simulated annealing protocols. These processes use the geometric restraints deduced from experimental NMR data to explore the conformational space of the macromolecule minimizing a penalty global energy function. This energy function consists of two parts: the first describing the energy due to the covalent terms of structure (bonds, angles, planarity and chirality) together with non-bonded contacts (interactions between atoms not covalently bonded which simulate Van der Waals interactions); the second part includes the contributions from all NMR restraints. These energy functions are expressed as harmonic potentials or flat-bottom potentials in function of stereochemical and experimental parameters, which are varied during the calculation. A fine balance between the "chemical" and experimental part is achieved by modulation of the force constants relative to the distinct contributions. The main contribution to obtain structures that satisfy experimental data arises from NOE. Due to the degeneracy of proton chemical shifts typically occurring in proteins, not all NOEs can be unambiguously assigned using the sequential assignment procedure. Many NOEs can be correctly assigned only during the structure calculation phase using structure-based knowledge in iterative manner. The initial structures are examined to confirm the assigned NOEs and to attribute unassigned

crosspeaks of the NOESY spectrum. Subsequent rounds of calculation are run with ever increasing number of correctly assigned distances. When the global minimum region is reached by the conformational search, structures with no violations of stereochemical and experimental parameters are generated. The structures with the lowest energy are selected to represent the protein conformation. In fact, the NMR solution structure of a protein can be graphically represented by an ensemble of very close conformers.

Residual Dipolar Couplings

Traditional NMR methods in structure determination rely on experimental data that are sensitive to local structure, such as NOEs and scalar J couplings. Dipolar couplings were ignored for a long time in structure calculation of biomolecules untill Tolman and coworkers were able to measure residual dipolar interactions inducing the weak alignment of paramagnetic myoglobin in the spectrometer magnetic field by using the relatively large susceptivity of this protein (Tolman at al., 1995). Macromolecules can be aligned in an extrnal magnetic static field if they possess large magnetic susceptivity but it's not the case of diamagnetic proteins. Alternatively, sufficient degree of molecular alignment is achieved using slightly anisotropic media (Bax et al., 2001). In the extreme situation of ineffective orientation averaging, such as in solids, dipolar couplings are very strong and make intractable the spectrum due to the complexity of signals introduced by these magnetic interactions. On the contrary, in solution the continuous tumbling averages all dipolar interactions to null. Liotropic nematic liquid crystalline phases become oriented in a strong magnetic field and induce residual alignment in proteins, conferring them preferential orientation by steric interactions. A time- and population-averaged preferential alignment derives because molecular reorientation is reduced. The degree of residual

alignment can be modulated by varying the liquid crystalline particles concentration, allowing to achieve values for the dipolar couplings that are readily measurable by NMR methods. A wide range of orienting media was investigated and used in this kind of measurements for proteins (Bax et al., 2001).

RDCs have been shown to significantly improve the accuracy and precision of NMR structures and can be directly employed in structure refinement (Clore et al., 1999). On the contrary to other commonly used NMR data, RDCs contain valuable global information independent of distance.

In order to calculate the dipolar interaction D^{AB} between two nuclei A and B for a system aligned in a orienting medium with respect to the z axis (the static magnetic field axis) this equation holds (Bax et al., 2001):

$$D^{AB} = -\mu_0 (h/2\pi) \gamma_A\gamma_B S/(4\pi^2 r^3_{AB}) <\cos^2(\theta) -1>$$

where μ_0 is the magnetic permittivity in vacuum, h is the Panck's constant, γ_A and γ_B are the magnetogiric ratios for the two nuclei, S is the order parameter that measures the degree of internal mobility (derived from nuclear relaxation data in the order of 0.8-0.9 for rigid structured segments of a protein), r_{AB} is the internuclear distance between A and B, θ is the angle between the z axis and the internuclear vector (figure 9). The brackets < > refer to time and ensemble averaging.

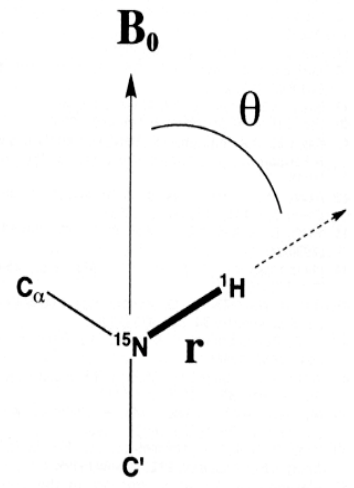

Figure 9. Schematic representation of the geometric relationship between the internuclear vector NH of the amide group and the external magnetic field B_0.

It's remarkable the fact that the dipolar coupling doesn't directly depend on the external field in this kind of alignment system (however, the field magnitude determines the overall alignment degree of the orienting media). To circumvent the distance dependence of this relation, residual dipolar couplings are typically measured for nuclei disposed at a fixed distance, for example covalently bonded atoms. In this way, the RDC depends only on the angle between the internuclear vector and the static magnetic field axis. The various bond vectors are differentially oriented in a molecule and hence the RDC measurements enable one to inspect the relative orientation between internuclear axes. This represents a stringent and useful restraint that is distance independent.

The angular dependence of RDC, the quantity $<\cos^2(\theta) -1>$, can be decomposed in a time-independent part that describes the relative orientation of the internuclear vector relative to an arbitrary molecular reference frame and in a time-dependent part. This second term measures the average degree and the average orientation of the molecules aligned in the magnetic field and is described by a tensorial quantity, the Saupe order matrix. The order matrix is symmetric, real and traceless and can be reduced to two independent components which contain all the information about the average orientation of the tensor itself with respect to the z axis and about the population of molecules that are on average aligned along the z axis. The two Saupe order tensor components are usually expressed as the axial component A_a and the rhombic component A_r. The rhombicity measures the tensor asymmetry (the xy component of the tensor) and the axial component measures the degree of residual molecular alignment .

In the case of complete alignment (all molecules aligned) an internuclear vector oriented perfectly parallel to z axis exhibits the maximum value for the residual dipolar coupling. The calculated value for RDC for NH is ca. 21.7 kHz, too high to perform measurements. In contrast, little residual alignment is observed for systems commonly

employed in high-resolution protein studies. The residual dipolar couplings usually exhibited by partially oriented proteins present values in the range of 2-50 Hertz. The dipolar coupling adds to the scalar coupling and causes a variation (positive or negative) of the characteristic splitting observed in the spectra of two nuclei that are not decoupled with the appropriate pulses. Acquisition of 2D correlation spectra without decoupling carried out on samples in orienting media allows to perform the measurement of accurate and precise RDCs.

Nuclear spin relaxation and protein dynamics

Despite the close packing in their structure, proteins 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 to characterize 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 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 nuclei itselves and the electron density distribution. These local magnetic fields are modulated by the global and specific movements of the protein. Relaxation requires that the interaction energy of these two 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 nuclear 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}C and ^{15}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 rates T_1, T_2 and the hetronuclear NOE of ^{13}C and ^{15}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 a 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 Fourier transformation of a time exponential function is a frequency Lorentzian function:

$$\exp(-t/\tau) \Rightarrow \frac{1}{1+\omega^2\tau^2}$$

Therefore, the total spectral density function can be written as:

$$J(\omega) = S^2 \frac{1}{1+\omega^2\tau_m{}^2} + (1\text{-}S^2) \frac{1}{1+\omega^2\tau_e{}^2}$$

where τ_m is global tumbling time constant, τ_e is the effective time constant experienced because of fast internal motions and S measures the weight of these internal motions. The nuclear transitions involved are typically on the order of 30-200 MHz for heteronuclei (it depends on the nucleus magnetogiric 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 α-helices and β-strands) are comprised in the range 0.9-0.95.

Explicit equations relate the relaxation rates for ^{15}N and ^{13}C nuclei to both these separated density functions (Kay et l., 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_{ex} that accounts for the contribution to the relaxation rates given by slower exchange processes (in the μs-ms range). The term R_{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 cab 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

Sample preparation

E2C protein preparation. Uniformly ^{15}N-labeled and ^{15}N–^{13}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 $^{15}NH_4Cl$ and ^{13}C glucose as the sole nitrogen and carbon sources. The cells were grown at 37 °C up to 0.35 $O.D._{600}$ 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-β-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 (BioSepra, 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}N - and ^{15}N – ^{13}C E2C, respectively. Protein concentration was determined using an extinction coefficient of 41,920 M^{-1} cm^{-1} for the dimmer (Mok et al., 1996). The protein was aliquoted and freezed at – 80 ° C for conservation.

E2 binding sites preparation. The palindromic dodecameric DNA oligonucleotides of sequence ACCGACGTCGGT and ACCGAATTCGGT were purchased from Bioneer (Bioneer Corporation, North Corea) as purified lyophilized powders. The self-complementary duplexes were dissolved and subjected to an annealing cycle consisting of gradually increasing the temperature to 96 °C and then to slowly cooling during several hours to final root temperature. The 18-mer single strand DNA oligonucleotide of sequence GTAACCGAAATCGGTTGA and its complementary strand were purchased from MWG (MWG Biotech Inc., Germany) as lyophilized and purified powders. Additional steps of purification were carried out using several cycles of ethanol precipitation in presence of high salt concentration to remove contaminants and impurities. Samples were then desalted by filtration on Sephadex G-15 column. Single strand oligo concentration was evaluated by calculating UV extinction coefficient as commonly reported. Each oligonucleotide was mixed with its complementary strand in the proper ratio and subjected to annealing treatment as above described. Purity and correct annealing of all oligonucleotides used in our study were checked by PAGE electrophoresis. NMR experiments confirmed that there was no detectable presence of single strand oligomer. The duplexes were lyophilized and conserved at –20 °C.

Complex formation. Different conditions for the formation of a stable complex between E2C and its 18-bp binding site of sequence GTAACCGAAATCGGTTGA, hereafter referred as BS(AAAT), were tested to obtain good-quality NMR spectra. Buffer identity and concentration, ionic strength, pH, temperature and reactant concentration were varied. The best conditions for NMR studies were found to include 50 mM sodium acetate, 10 mM sodium phosphate buffer, pH 6.5, 0.25 M NaCl, 5 mM DTT and a small excess (10-30 %) of target DNA. Temperature was demonstrated to disfavour aggregation. Appropriate aliquots of concentrated DNA provided with the calculated quantity of NaCl were added to the salt-free protein in Eppendorf tube. Turbidity was observed during the mixing, which disappeared by gently shaking the solution. Anyway, none of the samples remained stable for more than a few days and precipitation occurred in the course of time. Samples were conserved at room temperature.

NMR Spectroscopy

NMR experiments were performed on Bruker Avance700 and Avance400 spectrometers equipped with triple resonance probes incorporating shielded z-axis gradient coils. Sequence schemes employing pulsed field gradients were appropriately employed to achieve suppression of the solvent signal and spectral artefacts (Parella, 1998; Kay, 1995). Selective pulses to cover aliphatic or carbonyl ^{13}C nuclei spectral zone were obtained by three-gaussian pulse cascade modulation for selective excitation and four-gaussian pulse cascade modulation for selective inversion. Quadrature detection in the indirectly detected dimensions was obtained using the hybrid States-TPPI (Time proportional phase increment) method (Marion and Wuthrich, 1983). Multidimensional NMR experiments conducted on samples in water were provided with appropriate sequence blocks for

efficient solvent suppression such as watergate (Piotto et al., 1992) and flip-back (Grzesiek and Bax, 1993) pulse schemes. The NMR data were processed on Silicon Graphics workstations using NMRPipe (Delaglio et al., 1996) and analyzed using NMRView (Johnson and Blevins, 1994) software. Linear prediction was applied to extend the indirect ^{13}C-detected dimension. Direct and indirect dimension were normally zero filled prior to Fourier transformation to achieve higher digital resolution. Apodization 90°-shifted squared sine-bell functions were typically applied before Fourier transformation to improve resolution in ^{13}C- and ^{15}N-edited dimensions, while resolution enhancement by Lorentzian-to-Gaussian function was used in the proton dimensions. Proton chemical shifts were referenced to TMS using the relation

$$\delta(H_2O) = 3.18\ (177.6 - (T\ °C)/\ 96.9)$$

as reported by Orbons (Orbons et al., 1987). The ^{15}N and ^{13}C chemical shifts were referenced indirectly using the $^{1}H/X$ frequency ratio of 0.101329118 and 0.251449530 (Wishart et al., 1995).

Protein samples for NMR experiments on free E2C were exchanged into 50 mM sodium phosphate, pH 6.5, 5 mM DTT (1,4-dithio-DL-threitol) and 0,01% sodium azide. Concentration of the ^{15}N and $^{15}N/^{13}C$ uniformly labeled protein were 1.8 mM and 0.9 mM, respectively. Protein samples prepared in these conditions were stable for months when conserved at 4 °C, revealing no indications of ongoing degradation or aggregation as judged by NMR intensity peak analysis and PAGE electrophoresis. All NMR experiments on free E2C were conducted at 303 K.

Lyophilized DNA samples of protein-free E2-BSs for NMR experiments were dissolved in D_2O or in 95% H_2O and 5% D_2O containing 10 mM sodium phosphate buffer, pH 7.0. The concentration range for the various samples varied between 0.5 and 1.5 mM in

duplex. Temperature typically used in NMR experiments was 298 K, unless otherwise specified.

Samples of E2C-BS(AAAT) complex in 10 mM sodium phosphate buffer, pH 6.5, 0.25 M NaCl, 5 mM DTT were employed with concentration ranging from 0.1 to 0.5 mM at a temperature of 318 K.

DNA-free HPV-16 E2C resonance assignment and structure determination

Resonance Assignment. Nearly all HPV-16 E2C backbone ^{15}N, ^{13}C and ^{1}H resonances were sequentially assigned on the basis of the following standard set of double- and triple-resonance heteronuclear spectra: ^{1}H-^{15}N HSQC (Bodenhausen and Ruben, 1980; Kay et al., 1992), CT (constant time) ^{1}H-^{13}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. Simultaneously, resonance assignment of most side chain nuclei was achieved from analysis of ^{15}N-separated TOCSY (Grzesiek and Bax, 1992) with 25 ms spin locking time, HCCH COSY (Ikura et al., 1991) and HCCH TOCSY (Kay et al., 1993) with 16 and 24 ms spin locking time (Bax et al., 1990). Stereospecific assignments of valine and leucine residues were obtained by means of a biosynthetic approach (Neri at al., 1989), recording ^{1}H -^{13}C HSQC using selective decoupling on ^{13}CO or ^{15}N atoms which are scalar coupled to C^{α} on a 10 %-^{13}C-enriched and 100%-^{15}N-

enriched sample of E2C. Stereospecific assignment of other protons was achieved, where possible, from the inspection of the NOE patterns and J couplings followed by the analysis of the structures generated by the iterative structure calculation process (Wuthrich, 1986; Roberts, 1993).

Conversion of NMR data into structural restraints. In order to detect NOEs, 3D ^{15}N-edited NOESY and ^{13}C-edited NOESY were acquired using a mixing time of 100 ms. Cross-peak volumes were integrated using the NMRView tools. Approximate interproton distances were derived and the respective restraints were subdivided into three groups: 1.8-2.8 Å for strong NOEs, 1.8-4.0 Å for medium NOEs and 1.8-5.5 Å for weak NOEs. A set of $^{3}J_{HNH\alpha}$ was determined by measuring cross peak intensities in the HNHA spectrum (Vuister and Bax, 1993). Consequently, backbone ϕ torsional angle was restrained to -60° ±40° for residues exhibiting $^{3}J_{HNHa}$ < 6.0 Hz (indicating α-helix character) and to -120° ±40° for residues exhibiting $^{3}J_{HNHa}$ > 8.0 Hz (indicating β-strand character), while for all remaining residues except glycine residues ϕ angle was restrained to negative values (-90° ±90°). Only in the later stages of the structural calculation, when it was possible to identify defined secondary structure elements, ψ angles were restrained to -40° ±40° for α-helical regions and to 120° ±40° for β strands.

One-bond ^{1}H-^{15}N Residual Dipolar Couplings were measured by recording (F1) ^{1}H-coupled ^{1}H-^{15}N HSQC spectra on protein samples in isotropic medium and in 6% polyacrylamide gel matrix to induce the molecular alignment (Sass et al., 2000; Ishii et al., 2001). The measured separation of the upfield and downfield components of the doublet (expressed in Hz) directly yields the values of the dipolar coupling for that N-H spin system.

Hydrogen bonds were recognized by evaluating the spatial relationships of the slow exchangeable amide protons with potential acceptors in the initial structures calculated

without the use of hydrogen bond restraints. Two distance restraints were defined for each hydrogen bond: 1.8-2.8 Å for the H-O distance and 2.3-3.4 Å for the N-O distance in order to constrain both the distance and the orientation between hydrogen-bonded nuclei.

Structure Calculation. Structure calculation and refinement were performed with the updated NIH version of XPLOR (Schwieters et al., 2003) which is provided with new tools for the use of orientational and novel constraints such as Residual Dipolar Couplings (kindly supplied by G.M. Clore). The Simulated Annealing (SA) protocol for dimer calculation that we employed was ideated by Nilges and O'Donoghue (Nilges, 1993; O'Donoghue et al., 1996) and is available on the Web at Nilges group website. The scripts were modified in order to incorporate RDCs, a direct potential for J scalar couplings and a refinement against a statistical-derived database of Ramachandran plot dihedral angles.

The PARALLHDG force field of Xplor was employed for calculations. First, a series of monomers were generated using all constraints except intermonomer distances by means of a conventional SA protocol. Observed NOEs that were not consistent with generated monomers were attributed to intermonomer contacts and not included in the monomer generation. Monomeric structures with the least number of restraint violations and good stereochemistry were selected, duplicated and roto-translated to constitute the starting assemblies for the dimer structure calculation. Annealing of the two subunits was then driven by inclusion of intermonomer distance restraints in addition to intramonomer constraints. The two-fold symmetry was guaranteed by the non-crystallographic symmetry (NCS) and the distance symmetry (DSYM) energy terms included in XPLOR. NCS makes the first subunit superimposable to the second one, while DSYM makes the intermonomer distance between atom i in subunit A and atom j in subunit B identical to the symmetry-related distance between atom j in subunit A and atom i in subunit B (see Xplor manual). Dimers displaying violations of the restraints less than a defined threshold (0.3 Å for

distances, 5° for dihedral angles, 2 Hz for $^{3}J_{HNH\alpha}$ and 1 Hz for RDCs) were further subjected to a restrained energy minimization**.** The ensemble of the twenty lowest-energy target function structures not displaying violations of experimental constrains above the mentioned threshods was chosen to represent the solution structure of the dimer. The force constants employed in the SA protocol we used are listed in Table II. The programs AQUA and PROCHECK (Laskowski et al., 1996) were used to analyze the structures.

Cross validation for the dipolar coupling refinement was performed by excluding different sets of 8 RDCs (4 per monomer) from the calculation (Clore and Garrett, 1999). The resulting structures were used to back-calculate the missing RDCs with an average Q free quality factor of 22%. In addition, a set of 200 structures were generated without the use of RDCs to investigate the effect of RDC inclusion on E2C structure calculation. The twenty lowest-energy structures with no restraint violations were selected.

Backbone resonance Assignment for E2C-DNA Complex

Almost complete assignment of E2C-DNA complex backbone resonances was achieved acquiring ^{1}H-^{15}N HSQC, CT ^{1}H-^{13}C HSQC, HNCA, HN(CO)CA, HNCO, CBCA(CO)NH, HNHA and HACACO spectra, as done for the free protein. The C.S.I. for the protein, based on H^{α}, $^{13}C^{\alpha}$ and ^{13}CO atoms was inspected, yielding the same distribution of secondary structure found for the free protein.

Relaxation Data and Backbone Dynamics Analysis

Backbone dynamics analysis for free E2C. Relaxation experiments on E2C in its DNA-free form were carried out at 30 °C for E2C on both Bruker Avance700 and Avance400 spectrometers mentioned above. Measurements of ^{15}N T_1, ^{15}N T_2 and heteronuclear ^{1}H-^{15}N NOE were made by performing established ^{1}H-detected pulse schemes (Stone et al., 1992; Kay et al., 1989) in an interleaved manner. To sample the magnetization decay six points were collected with delays of 16, 240, 465, 689, 913, 1123 ms for T_1 at 40.5 MHz and of 14, 210, 420, 700, 1191, 1542 ms for T_1 at 70.9 MHz , and six points with delays of 8.3, 33.1, 82.2, 99.4, 132.5, 165.6 ms for T_2 at 40.5 MHz and of 8.2, 24.5, 40.8, 57.1, 73.4, 97.9 ms for T_2 at 70.9 MHz. Integrated cross-peak volumes of non-overlapped resonances were fitted to two-parameter mono-exponential decays. The uncertainties of peak intensities were evaluated as the standard deviations of the spectral noise measured in a region free of correlation peaks. The heteronuclear ^{1}H-^{15}N NOE values were determined by the ratio of peak volumes of spectra recorded with and without ^{1}H saturation, employing a net recycle delay of 3 seconds for each scan in both experiments to enable recovery of equilibrium for ^{1}H nuclei.

^{15}N relaxation data were analysed making use of the program DASHA (Orekhov et al., 1995). The first step of the analysis was the evaluation of a suitable model for the global tumbling of this molecule.. The simplest representation of molecular motional properties implies the assumption of an isotropic tumbling, which can be described by a single global correlation time τ_m. As a first estimation of the isotropic τ_m, the T_1/T_2 ratio approach was used (Farrow et al., 1994). The experimental values for T_1/T_2 at both field strengths were fitted to those calculated using the corresponding analytical expression, varying τ_m to minimize the error-weighted difference of the ratio T_1/T_2 . Only those

residues showing neither significant shortening of T_2 nor a ^1H-^{15}N NOE value less than 0.6 at 70.9 MHz ^{15}N frequency were used to derive τ_m (Tjandra et al., 1995). In this way 14 residues out of 64 showed T_1/T_2 outside one standard deviation of the mean value, giving a first indication that HPV-16 E2C can be treated as an isotropically tumbling molecule. This calculation yielded a τ_m of 10.3 ± 0.5 ns.

As a second step, the six relaxation quantities measured for each HN were analysed using the simple model-free formalism developed by Lipari and Szabo (Lipari and Szabo, 1982; Kay et al., 1989). In this approach a global τ_m (unique for all residue) and three residue-specific parameters describing dynamic motions, S^2, τ_e and R_{ex}, are extracted from the equations relating the T_1, T_2 and ^1H-^{15}N NOE to the spectral density function. S^2 is a generalised order parameter measuring the degree of spatial restriction of rapid internal motions of the N-H bond vector, τ_e is the rate of these rapid internal motions and R_{ex} is the term accounting for chemical exchange. Four different combinations of parameters were used to fit the experimental data: (1) S^2; (2) S^2, R_{ex}; (3) S^2, τ_e; (4) S^2, R_{ex}, τ_e. Statistical considerations regarding the fitting improvement and the number of freedom degrees were based on the F-value test. This statistical treatment enabled us to assign the individual ^{15}N data to either one of these four classes (Pawley et al., 2001). Seven residues (11%) present statistically significant deviations between measured and calculated relaxation parameters in all the four models and could not be fitted. These residues were excluded from the following calculations. Finally, the statistically best fitting model for each individual residue was used to optimise τ_m by minimizing a single penalty function measuring the deviation between calculated and measured relaxation rates. In this way, a global correlation time of 10.8 ± 0.1 ns could be estimated.

Backbone dynamics analysis for E2C-DNA complex. Relaxation data on the E2C-BS(AAAT) complex were acquired at 45 °C for E2C on Avance700. The same set of

experiments carried out for the free protein was conducted for the complex (measurements of ^{15}N T_1, ^{15}N T_2 and ^{1}H-^{15}N NOE). Six points were collected with delays of 14, 423, 844, 1265, 1686, 2107 ms for T_1 at 70.9 MHz , and six points with delays of 8.3, 24.9, 41.5, 58.1, 74.7, 99.6 ms for T_2 at 70.9 MHz. Identical data analysis and statistical treatment used for the free protein were applied to relaxation parameters extracted for the complex.

Amide Exchange measurements

Exchange rates measurements of free E2C. A series of ^{1}H-^{15}N HSQC were acquired at 0.5 h, 2 h, 4 h, 12 h and 24 h (midpoints of the experiments) after dissolving the lyophilized free E2C protein in D_2O to monitor the disappearance of signals of amide exchangeable protons. Water-NH magnetization transfer rates were investigated following the approach of Grzesiek and Bax (Grzesiek and Bax, 1993) by performing a water-selected NOE HSQC (60 ms mixing time) and a water-selected ROE HSQC (25 ms mixing time). The starting point for both experiments is the selective excitation of the water resonance, guaranteed by the purging of underlying H^{α} resonances which are coupled to ^{13}C nuclei, followed by NOESY- or ROESY-type transfer to amide protons. A 2D HSQC experiment was then performed to yield the necessary resolution to measure individual peak intensities for each NH group. Two separate 2D NOESY and two 2D ROESY experiments were performed, with the water magnetization parallel or antiparallel to the amide proton magnetization during the corresponding mixing periods. The difference in cross-peak intensities divided by the sum of intensities from the two experiments yields ζ_{NOE} and ζ_{ROE}. Using known relationships, it is possible to calculate two magnetization exchange rates for the NOESY and ROESY experiments, k_N and k_R respectively, from the following formulas:

$$k_N = \frac{2\zeta_{NOE}(\rho_1 + k_N)\{M_z(0) + [\exp((\rho_1 + k_N)t) - 1]M_z^0\}}{[f^+(1-\zeta_{NOE}) - f^-(1+\zeta_{NOE}) + 2\zeta_{NOE}][\exp((\rho_1 + k_N)t) - 1]M_z^0}$$

$$k_R = \frac{2\zeta_{ROE}(\rho_2 + k_R)M_y(0)}{[f^+(1-\zeta_{ROE}) - f^-(1+\zeta_{ROE})][\exp((\rho_2 + k_R)t) - 1]M_z^0}$$

where ρ_1 is the amide proton-spin flip rate, ρ_2 equals $1/T_{1\rho}$ in absence of magnetization exchange, $M_z(0)$ and $M_y(0)$ are the longitudinal or transverse water magnetization at the beginning of the NOE or ROE mixing period, M_z^0 is the amide proton z-magnetization at thermal equilibrium, t is the mixing time, and f^+ and f^- are the fraction of the thermal equilibrium water magnetization which ends up parallel or antiparallel to the amide proton magnetization at the beginning of the mixing period. We have estimated that in our conditions $M_z(0)$ and $M_y(0)$ are ca. $0.78\,M_z^0$, $f^+=f^-=0.85$. In the presence of magnetization exchange, the quantities (ρ_1+k_N) and (ρ_2+k_R) can be obtained from the following relationships:

$$1/T_{1\rho} = \rho_2 + k_R$$
$$1/T_{1ZZ} = 1/T_1^N + \rho_1 + k_N$$

We had calculated the two rates from the measured relaxation times, proton amide $T_{1\rho}$, the T_1 relaxation time of the $H_Z^N N_Z$ magnetization and the ^{15}N T_1. Finally, using the mentioned equations we had calculated k_N and k_R for all well-resolved NHs of HPV-16 E2C using the NOESY and ROESY experiments.

Resonance assignment for ACCGACGTCGGT and ACCGAATTCGGT binding sites

NMR studies on duplex ACCGAATTCGGT, hereafter referenced as BS(AATT), and on duplex ACCGACGTCGGT, hereafter referenced as BS(ACGT), were conducted

performing one-dimensional and two-dimensional experiments at 25 °C. TOCSY spectra employing MLEV-17 isotropic mixing scheme with a mixing period of 60 ms and NOESY spectra at mixing times of 50, 80, 120 and 150 ms were recorded on samples dissolved in 100% D_2O. In addition, heteronuclear 2D ^{1}H-^{13}C HSQC and 2D ^{1}H-^{13}C HMQC-TOCSY were carried out at natural abundance. NOESY spectra at mixing times of 50, 80, 120 and 150 ms were acquired on samples dissolved in 95% H_2O and 5% D_2O employing the "jump & return" scheme to suppress water signal (Plateau and Guéron, 1982). Also heteronuclear ^{1}H-^{15}N HSQC was recorded, using a delay for INEPT-type transfer steps of 2.9 ms which approximately matches the scalar one-bond coupling for imino and amino groups in nucleic acids (circa 85 Hz: see Zidek t al., 2001).

Proton assignment for both dodecamers was achieved with the standard methodology used for B-DNA double helix (Wuthrich, 1986; Roberts, 1993). Heteronuclear correlation spectra were analysed to assist with the proton assignment and to obtain partial assignment of ^{13}C and ^{15}N resonances.

E2 Binding Sites diffusion measurements

Diffusion experiments were carried out for both BS(AATT) and BS(ACGT) using the same concentration (0.4 mM in duplex) and at three distinct temperature points (25, 15 and 5 °C) in presence and absence of $MgCl_2$ 10 mM. The PFG-LED-STE (Pulsed Field Gradient Longitudinal Eddy-current Stimulated Echo) sequence (Gibbs and Johnson, 1991) was employed using a duration of gradient pulse δ=2 ms and a delay between defocusing and refocusing pulsed gradients Δ= 230 ms. The length of all pulses and delays were held constant while the gradient strength was varied from 5 % to 100 % of its maximum value (50 G/cm). For each diffusion measurement 15 gradient points were used to sample the

intensity decay at intervals dependent on the particular experimental conditions in order to obtain a ratio of about 10 between the most intense and the least intense point. Free induction decays were processed using a sensivity-enhancement exponential function prior to Fourier transformation. Two spectral regions were chosen to measure the diffusion constants: the aromatic region (from 6.5 to 8 ppm) and the H2'/H2''/Me region (from 1.5 to 3 ppm). These zones were preferred because don't present overlap with other resonances. Phase and base-line correction were applied to the spectral regions of interest. Peak convolutions in the selected regions were integrated. The intensity of the signal at a specific gradient strength G is given by the equation (Tanner, 1970).:

$$I(G) = I(0)\exp[-D\gamma^2_H \delta^2 G^2(\Delta-\delta/3)]$$

where I(0) is the intensity in absence of gradients, D is the diffusion coefficient, γ_H is the magnetogiric ratio for the proton, δ the duration of gradient pulse and Δ the delay between the defocusing and the refocusing pulsed gradients. Kaleidegraph (version 3.02) software was used to carry out the non-linear least-squares regression and to extract the diffusion constant and the Pearson correlation coefficient from the fit. Only diffusion rates obtained from fits with correlation coefficient greater than 0.99 were accepted. The results given by the two regions are very close.

RESULTS AND DISCUSSION

Target of the study.

As described in a previous chapter, the DNA recognition by E2 proteins is an archetypical case of "indirect readout", that is a mechanism by which selective DNA recognition is not mediated by direct contacts between the protein and the DNA sequence that is discriminated. The Papillomavirus E2 proteins are DNA-binding proteins that tightly control viral transcription and replication in a context-dependent manner and are implicated in HPV-induced tumour development. To efficiently exert their function, HPVs E2 proteins require to have differential affinity toward E2 binding sites (E2-BSs). This exquisite distinctive sequence preference has been extensively studied in order to investigate the relationship between sequence-dependent DNA conformational properties and preferential binding that occurs in fundamental biological processes. The three-dimensional structure of the C-terminal DNA-binding domain of E2 proteins in the DNA-free form, thereafter referred as E2C, is reported for different viral strains (listed in table I). Four structures of E2C protein complexes from HPV-18 and BPV-1 strains with low-affinity and high-affinity binding sites were solved by X-Ray crystallography (see table I). On the contrary, a detailed high-resolution analysis of HPV-16 complex is not available. To this purpose, we aimed at a structural characterization of this kind of complex. We decided to use Nuclear Magnetic Resonance techniques because we are interested not only in the elucidation of the stereochemical aspects of the interaction, but we are also focusing on the dynamic properties of E2C-DNA which can have important implications in the sequence-specific DNA recognition mechanism. The application of NMR spectroscopy in structural biology as increasingly diffused and consolidated methodologies are now largely spread to gain information at the atomic level not only about the solution conformation of

biomolecules, but especially about intermolecular interactions. NMR data can give precise indications on conformational and dynamical changes involved in the recognition process.

The first step in our work was the determination of the solution structure of HPV-16 E2C. This result permits us to examine eventual structural modifications that accompany the DNA binding by accurately comparing characteristic NMR parameters in the free and in the bound conformation of the protein. The lack of a HPV-16 E2C-DNA complex with sufficient stability prevented us to obtain a complete high-resolution structure because of the long time-consuming nature of NMR experiments required for such a determination. Nevertheless, interesting structural data were collected that could contribute to a detailed analysis of the binding process based on HPV-16 E2C conformational models derived from the DNA-free form and from the previously reported high-resolution structures of complexes from other strains. Chemical shifts are sensitive and precise indicators of minimal conformational changes in a molecule and allow to map the regions that are perturbed upon binding.

NMR spectroscopy is the most powerful technique for the characterization of biomolecular dynamics. Translational and internal motions profoundly affect specific NMR parameters and quantitative measurements at the atomic level are performed to examine global and local mobility. Relaxation data for backbone ^{15}N nuclei and water-exchange rates were indeed measured for the free and complexed state of HPV-16 E2C and were used to analyse the dynamics of the system in the two conformations.

Structural and biochemical data about E2C DNA-binding were published indicating that the major conformational alterations are situated in the DNA, while the protein exhibits modest changes, at least in the case of complexes of the BPV-1 structural type (Kim et al., 2000; Hegde, 2002). Theoretical and experimental models suggest that HPV-16 E2C displays great preference for E2-BSs with AT-rich spacer sequences, probably due

to their sequence-dependent structural propensity to adopt specific bend conformations that facilitate DNA binding. In order to investigate whether this hypothesis is correct, we conducted NMR experiments on the free form of two palindromic dodecameric E2 binding sites: the low-affinity ACCGACGTCGGT sequence and the high-affinity ACCGAATTCGGT sequence. Assignment of the free DNA resonances were obtained for both the molecules and a comparative analysis of NMR-derived local and global parameters, namely diffusion rates, is performed for the two oligomers.

NMR solution structure of HPV-16 E2C protein

The HPV-16 E2 DNA-binding C-Terminal domain (the last 80 residues plus a methionine in the first position) was expressed as a soluble recombinant protein. Numeration was chosen to be consistent with the full-length protein (residues 283-363). The sequence of HPV-16 E2C used in our studies is shown in figure 10.

283 MET THR PRO ILE VAL HIS LEU LYS GLY ASP **292**
293 ALA ASN THR LEU LYS CYS LEU ARG TYR ARG **302**
303 PHE LYS LYS HIS CYS THR LEU TYR THR ALA **312**
313 VAL SER SER THR TRP HIS TRP THR GLY HIS **322**
323 ASN VAL LYS HIS LYS SER ALA ILE VAL THR **332**
333 LEU THR TYR ASP SER GLU TRP GLN ARG ASP **342**
343 GLN PHE LEU SER GLN VAL LYS ILE PRO LYS **352**
353 THR ILE THR VAL SER THR GLY PHE MET SER **362**
363 ILE

Figure 10. Residue sequence of recombinant HPV-16E2C employed in our study

Chemical shift assignment and structure calculation. The analysis of a set of double- and triple-resonance NMR experiments (Ikura et al., 1990; Bax and Grzesiek, 1993)

provided the assignment of ^{1}H, ^{13}C and ^{15}N resonances of HPV-16 E2C. The ^{1}H-^{15}N HSQC spectrum (Bodenhausen and Ruben, 1980; Kay et al., 1992) is widely used to inspect the spectral quality of a sample and gives useful information about the reliability of NMR study of the protein which is examined. Virtually, this spectrum displays correlation peaks for all NH groups. In particular, there should be present ^{1}H-^{15}N correlations for all non-proline residue amide groups, except the first residue in which the basic nitrogen is protonated. Some peaks are usually missing. They belong to unstructured or flexible regions of the protein and experiment very fast exchange with the solvent. Other peaks could be missing if the corresponding protein sites are involved in relatively slow conformational exchange which causes broadening and disappearance of the signal. The ^{1}H-^{15}N correlation spectrum of HPV-16 E2C is presented in figure 11 with the observed peaks labeled by their assignment.

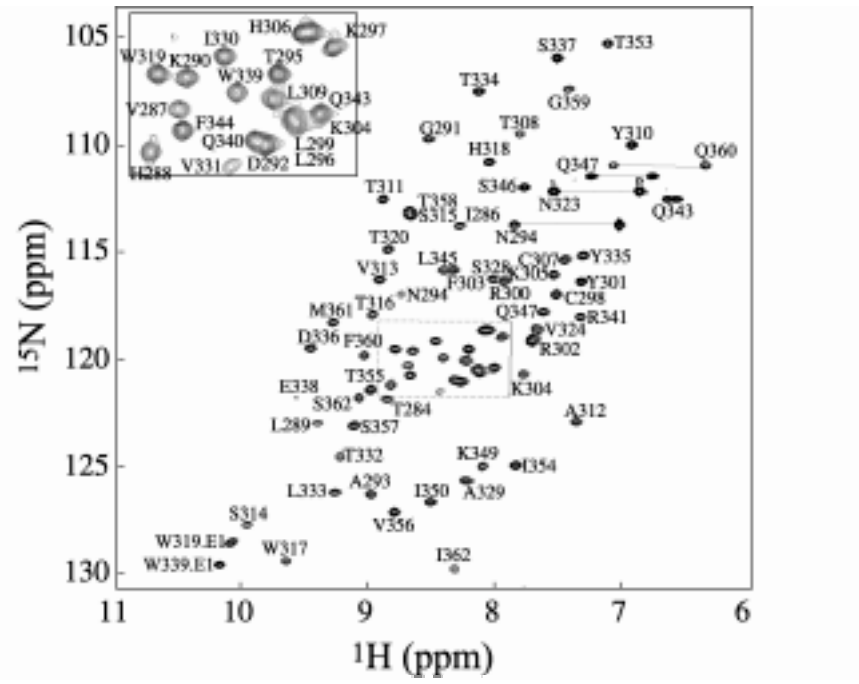

Figure 11. 700 MHz ^{1}H-^{15}N HSQC spectrum of HPV-16 E2C in 50 mM sodium phosphate, 5 mM DTT, 0.05% sodium azide and 95% H2O, 5% D2O at pH 6.5, collected at 303 K, reporting backbone resonance assignments (labelled peaks). Side chain amide protons of Asn and Gln residues are indicated by solid horizontal lines. The offset shows an enlarged view of the central crowded region of the spectrum.

Only a single set of resonances is observed for the E2C dimer, indicating that an averaged dimeric symmetric structure exists in solution. Amides of residues 321-323 and 325-327 were not detected in the ^{1}H-^{15}N HSQC spectrum and consequently could not be assigned. They all belong to a disordered loop (320-327) and exchange very fast with the solvent. Side chains ^{1}H and ^{13}C resonances were nearly all assigned on the basis of HCCH-TOCSY and ^{13}C-edited NOESY spectra. The C.S. assignment for HPV-16 E2C was deposited in the BMRB (BioMagneticResBank) database with accession number 5952.

A total of 1199 intramonomer NOEs were obtained from 3D ^{15}N- and ^{13}C-edited NOESY spectra. Additionally, 66 NOEs were assigned as unambiguous intermonomer distance constraints considering that the close proximity of the related protons is not compatible with the monomeric structure.

Chemical shifts, scalar $J_{HNH\alpha}$ couplings, characteristic NOE patterns and water amide exchange rates were used to identify HPV-16 E2C regular secondary structure elements, as evidenced by Chemical Shift Index (CSI) for C^{α}, H^{α} and CO atoms (Wishart and Sykes, 1994) illustrated in figure 12. The C.S.I. is a simple method to identify secondary structure propensities. It is based on the characteristic shits displayed by backbone proton and carbon atoms for residues situated in α–helices or β-strands relative to average value presented for the same residue in random coil structure. The delineation of E2C topology was confirmed in the final set of calculated structures.

Thus, 244 dihedral constraints for ϕ and φ were included in the calculation, based on both the inspection of the CSI for C^{α}, H^{α} and CO, and the values of the observed three-bond coupling constants ($^{3}J_{HNH\alpha}$) from the HNHA spectrum. These scalar couplings were used as constraints by the application of the corresponding Karplus equation parameterized by Vuister and Bax for the H^{N}-H^{α} spin system (Vuister and Bax, 1993). We have also

included 76 distance restraints corresponding to 38 hydrogen bonds. Hydrogen-bonded amides were identified by the application of two criteria: they exchange slowly with water and they appear to be spatially close to an oxygen atom in the initial NMR structures calculated without the hydrogen bond restraints.

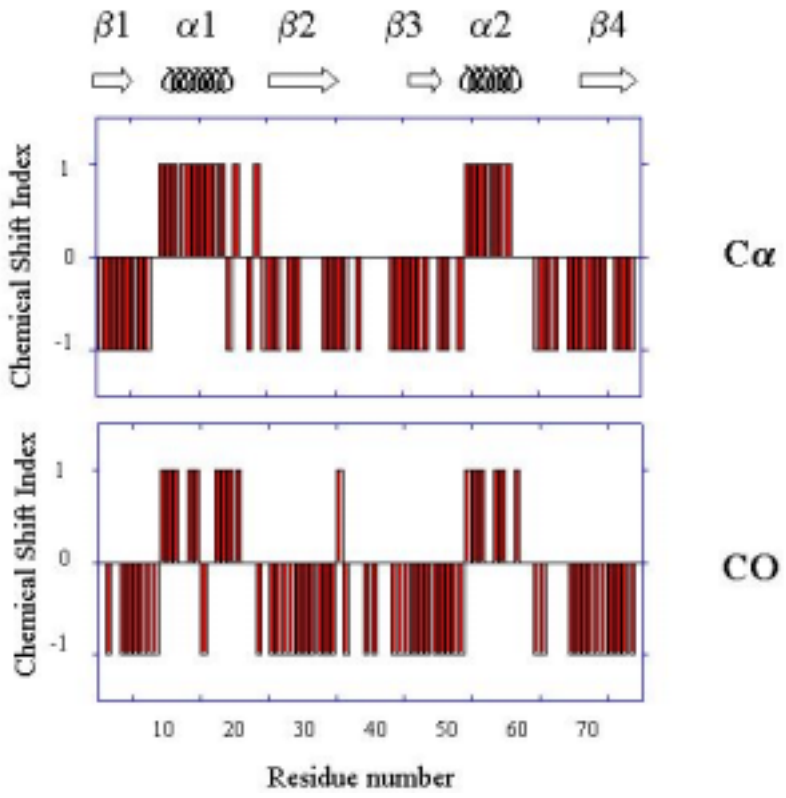

Figure 12. Chemical Shift Index of C^{α} and CO atoms for HPV-16 E2C. The C.S.I. is presented in "digital" fashion. In the case of C^{α} atom, a value of +1 is attributed to residues presenting C.S. higher than 0.8 ppm and a value of -1 for residues presenting C.S. lower than 0.5 ppm relative to random coil average value (a value of 0 is given for intermediate values). In the case of C^{α} atom, a value of +1 is attributed to residues presenting C.S. higher than 0.5 ppm and a value of -1 for residues presenting C.S. lower than 0.5 ppm relative to random coil average value (Wishart et al., 1994).

The nature of NMR derived restraints commonly employed in structure determination (NOEs, scalar J couplings, Chemical Shift values) is strictly local. On the contrary, the use of RDCs in structure calculations provides unique long-range orientational restraints (Bax et al., 2001; Clore et al., 1998). The residual dipolar coupling (RDC) between proton and

nitrogen nuclei of the amide group provides information about the angle between the vector connecting the two nuclei and the molecular reference frame. When dealing with molecules presenting a C2 symmetry like HPV-16 E2C dimer, one of the axis of the alignment tensor must be parallel and the other two orthogonal to the molecular 2-fold symmetry axis (Bewley and Clore, 2000; Al-Hashimi et al., 2001). As a consequence, the dipolar couplings for both halves of the dimer must be identical because the subunits share a unique single alignment tensor. This fact constitutes a stringent symmetry restraint for monomer orientation. The resulting structures exhibit higher accuracy with respect to structures generated without RDCs, especially at dimer interface where short interproton distances from NOE alone are less suitable to define this region (see below).

To this purpose, Residual Dipolar Couplings (RDCs) of 56 amide groups (28 for each monomer) were measured using polyacrylamide gel to produce a residual alignment (Sass et al., 2000; Ishii et al., 2001). As already observed for other proteins, only a limited decrease in amide ^{15}N T_2 was detected, allowing a sufficiently accurate measurement of the H-N splittings. Figure 13 shows examples of the observed couplings.

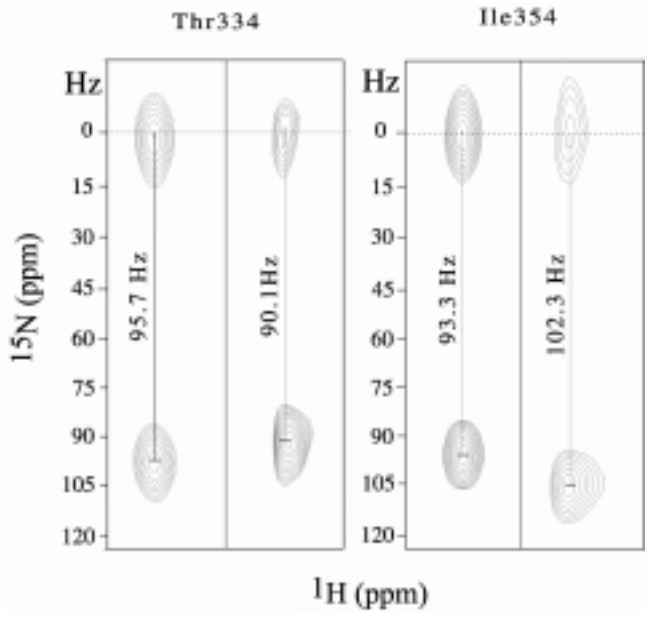

Figure13. Representative spectra of ^{1}H-^{15}N splittings, acquired at 700 MHz in the isotropic sample (left part of each panel) and in the polyacrilamide gel anisotropic sample (right part of each panel). Sections for Thr334 and Ile354 H-N nuclei are presented.

We used the program PALES (Zweckstetter et al., 2000) to extract values of D_a and R, which are the axial and rhombic components of the molecular alignment tensor and are needed for structure calculation. The resulting alignment is described by D_a = 4.7 Hz, R = 0.13, which were employed in the Xplor Simulated Annealing protocol.

Structure description. The family of the 20 lowest energy structures is presented in Figure 14a. The corresponding statistics are presented in Table II at the end of the Results and Discussion chapter. Good values were obtained for the general agreement with the NMR experimental restraints, the overall energy and the covalent bond geometry. The average atomic root mean square standard deviation (rmsd) of the 20 structures of dimer from the mean coordinates is 0.67 Å for backbone heavy atoms (N, C^{α}, C and CO) and 1.07 Å for all non-hydrogen atoms. The disordered loop spanning from residues 320 to 327 was excluded from these calculations. The statistical analysis of the structures shows that 99% of the residues present converged dihedral angles in the most favored region of the Ramachandran diagram. The ribbon model of the energy minimized average structure is shown in Figure 14b.

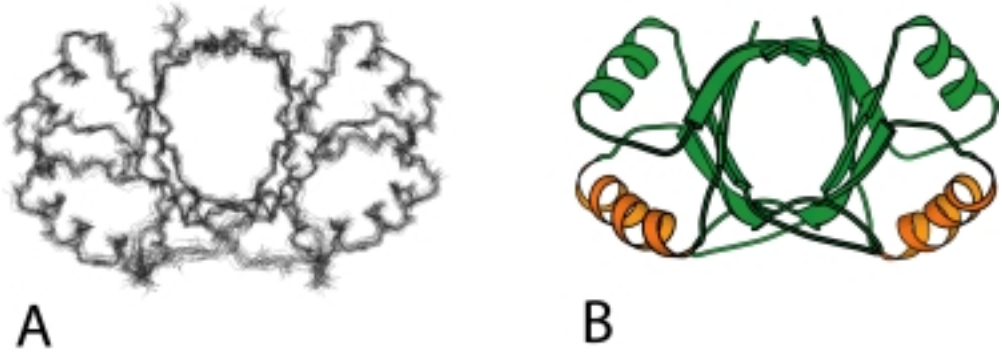

Figure 14. (a) Backbone superimposition of the 20 lowest-energy structures calculated for HPV-16 E2C dimer (rmsd is 0.67 Å for backbone heavy atoms and 1.07 Å for all non-hydrogen atoms). (b) Ribbon representation of the energy minimized average structure. The recognition helices are coloured in orange.

The homodimer is constituted by an eight-stranded dome-shaped β-barrel composed by two symmetric four-stranded antiparallel β-sheets, each from one monomer. The monomer presents a β_1–α_1–β_2–loop–β_3–α_2–β_4 secondary structure topology, where the first helix (α_1) is the "recognition helix" (figure 15a). The segment comprising residues 320-327 displays high conformational disorder and appears as a flexible loop. The helices lie on the barrel surface. The hydrophobic core of the monomer is formed by residues located on the outer face of β-sheets and on the buried hydrophobic side of the amphipathic α-helices. Residues Leu296, Leu299 and Phe303 from helix α_1 are packed against residues Leu289, Val313, Val331, Val348, Ile350 and Ile354. The second helix, α_2, presents residues Gln340, Arg341, Phe344 and Val348, buried against residues Val287, Leu289, Phe303, Leu309, Leu333, Tyr335 and Val356. Different types of interactions contribute to hold together the two subunits. At the edges of the barrel, β_2 and β_4 strands establish a hydrogen bonding network with their corresponding symmetry-related β strands on the opposite subunit. Numerous interactions are also made between side chains of these strands. A series of hydrophobic residues packed inside the subunit interface constitutes the protein core, as displayed in Figure 15b.

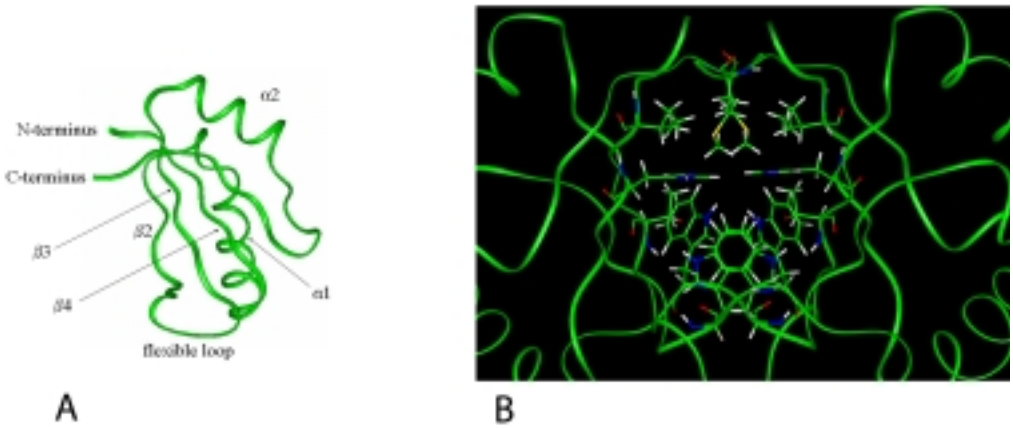

Figure 15. (a). The monomer is represented using a coil model. Secondary structure elements are indicated. (b) Detailed view of the dimer interface showing the hydophobic residues packing.

Two pairs of tryptophan residues (Trp317 and Trp319) are symmetrically disposed at the dimer interface in a characteristic fashion, which is conserved in E2C proteins. Trp317 from the two monomers are stacked in parallel, whereas Trp319 appears perpendicular to Trp317 in the same subunit. Other side chains making hydrophobic interactions at the interface are Ile286, His288, Thr332, Thr334 and Met361.

This extensive network of interactions is responsible for dimer stability, which was measured in the subpicomolar range (Mok et al., 1996a). The inspection of the structure explains also the results obtained in folding studies conducted on HPV-16 E2C (Mok et al., 1996a; Mok et al., 1996b; Mok et al., 2000), from which authors have postulated that folding and dimerization are strictly coupled. Native-like monomeric forms cannot survive at equilibrium (although non-native-like monomeric states are individuated in kinetic experiments) because the extensive hydrophobic surface forming the dimer interface would be exposed to the water.

The hydrogen bonding network between the two β_2 strands is invariant in all the known structures of E2C domains, and involves hydrogen bonding of residues Ser315 with His318 in HPV-16 E2C. Hydrogen bonds that connect the two β_4 strands are formed by interaction between the NH group with the opposing CO group of the two Phe360 and between the NH group of Ser362 with the CO of Thr358 (figure 5, in a preceding chapter). This register coincides with that observed for the crystal structure of HPV-16 and HPV-31 E2C and differs from that of HPV-18 and BPV-1, which constitute a distinct structural family of E2C fold (Kim et al., 1999; Hegde, 2002). The atomic coordinates of NMR solution structure of HPV-16 E2C together with restraints files were deposited in the PDB (Protein Data Bank) database with ID code 1R8P.

Comparison with the X-ray structure of HPV-16 E2C. When superimposed to its crystal structure (Table I for reference), the average NMR structure of HPV-16 E2C shows a rmsd

of 0.81 Å between the monomers and 1.15 Å between the entire dimers for backbone atoms when the β_2-β_3 loop is excluded (Figure 16).

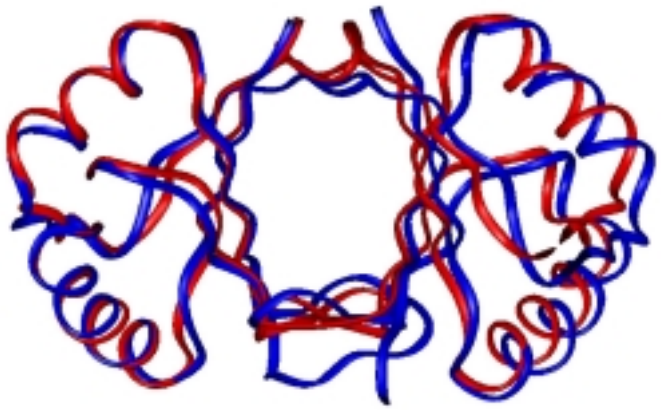

Figure16. Superimposition of HPV-16 E2C structure obtained by NMR (blue) and by X-Ray (red). Residues 284-319, 329-362 of both subunits were aligned (rmsd 0.81 for the C^{α} between monomers and 1.15 for the entire dimer).

Looking at other E2C structures which present both crystallographic and solution structure (HPV-31 and BPV-1), the backbone atoms pairwise rmsd between NMR and X-Ray structures is 1.10 Å and 1.54 Å for the monomers and the entire dimers, respectively, in the case of HPV-31 (residues 293-328, 338-371). The rmsd is 1.52 Å and 2.18 Å for the monomers and the entire dimers, respectively, in the case of BPV-1 (residues 327-363, 374-407). The closer agreement between the X-Ray and NMR structures in HPV-16 could be the consequence of including the RDCs in defining the quaternary structure (see below). In facts, this is the first time that residual dipolar coupling are employed in structural characterization of this kind of fold.

The average distance of the two $N^{\varepsilon 2}$ of His288 located at the dimer interface are 7.4 Å apart, somewhat longer from that observed in the crystal structure of HPV-16 E2C (5.3 Å). In the latter case, a molecule modeled as water makes a bridge between the two rings. It

was also proposed that the observed electron density can be due to a heavier atom like La^{3+} or $SO4^{2-}$, or another metal ion fortuitously carried through the purification process (Hegde and Androphy, 1998). The long-range ^{1}H-^{15}N HSQC spectrum (Bachovchin et al., 1986) correlates the aromatic non-labile protons of the histidine residue ($H^{\delta 2}$ and $H^{\varepsilon 1}$) with the nitrogen atoms ($N^{\delta 1}$ and $N^{\varepsilon 2}$) by means of ^{2}J scalar coupling. The chemical shift for the $N^{\delta 1}$ od His288 as measured in the long-range ^{1}H-^{15}N HMQC is around 251.3 ppm (figure 17) suggesting that there is no strong interaction involving the nitrogen lone pair.

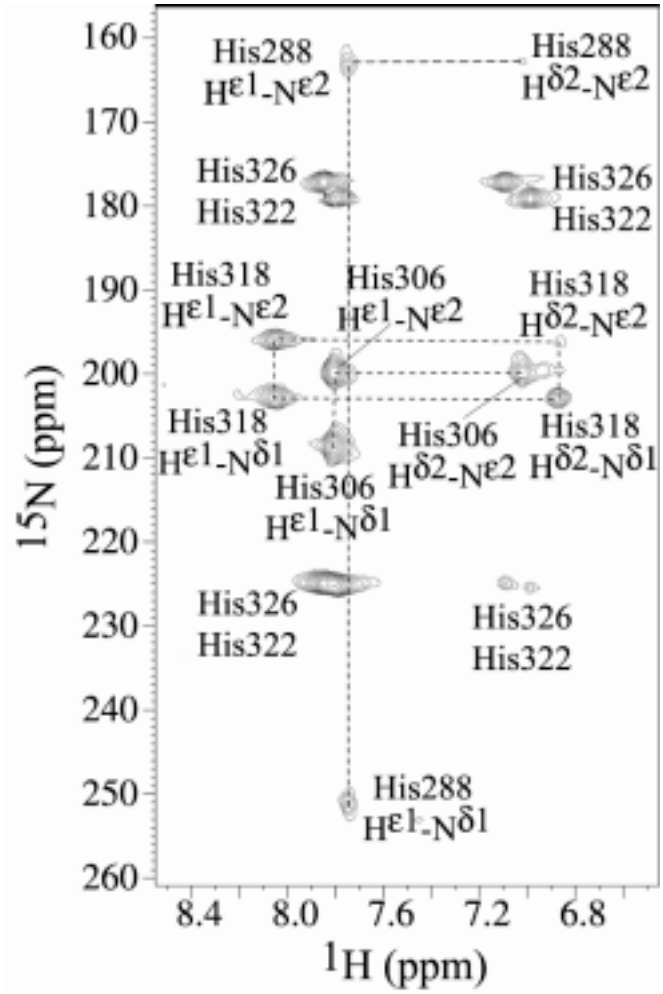

Figure 17. Section of the long-range ^{1}H-^{15}N HMQC showing the typical pattern of Histine. The crosspeaks relative to the five Histidine residues of E2C are labelled.

An interaction of the type of hydrogen bonding or metal interaction would shift the non-

protonated nitrogen resonance 10 to 25 ppm upfield respectively, from the mean value reported (249.5 ppm). In solution, the "triangle" crosspeak pattern for His288 is characteristic of the β tautomeric form (Van Dijk, 1992), in which $N^{\varepsilon 2}$ is protonated while $N^{\delta 1}$ is non-protonated. An alternative bridge could involve the ε2 NH of His288 as hydrogen bond donor. However, such an interaction is not compatible with the conformation of His288 that points $N^{\varepsilon 2}$ in the reverse direction. Additionally, the low chemical shift value observed for $N^{\varepsilon 2}$ (163 ppm) is also a good evidence for the absence of such an interaction (Bachovchin, 1986).

Comparison with solved structures from other strains. As described above, E2C proteins can be classified into two families on the base of relative subunit orientation (Kim et al., 1998; Hegde, 2002). Superposing HPV-16 E2C solution structure onto the other structures solved so far (listed in Table I) confirms that our structure resembles much more HPV-31 (which is comprised in the same class), whilst differs from HPV-18, BPV-1 and HPV-6 type. This structural comparison is graphically apparent in figure 4 (in a preceding chapter) and summarized in Table III.

Table III. Superimposition of E2C structures from different strains with the NMR E2C structure from sHPV-16[a].

Structure	*Monomer* [e]	*Dimer* [e]
HPV-31[b]	0.88	1.36
HPV-18[c]	1.75	2.02
BPV-1[d]	1.56	1.87

[a] PDB code: 1R8P (residues 284-319,330-358,360-361)
[b] PDB code: 1A7G (residues 293-328,339-367,369-370)
[c] PDB code: 1F9F (residues 287-322,332-360, 362-363)
[d] PDB code: 1JJH (residues 326-347,349-362,375-405)
[e] Only backbone atoms were superimposed: rmsd values in Å

Impact of RDCs on the accuracy and precision. Residual dipolar couplings (RDCs) contain all the information about the orientation of internuclear vectors and so represent a means to explore the geometric relationships of distinct fragments in the polypeptide chain.

In fact RDCs were suggested to be a powerful tool to validate the accuracy of NMR structures (Bax et al., 2001) and to rapidly assess the structural resemblance between structural folds of homologous proteins (Annila et al., 1999; Fowler et al., 2000). The best measure of similarity of the three-dimensional structure for two proteins is the rmsd of atomic coordinates. In absence of complete structure determination, it's possible to obtain RDCs for a protein and compare experimental values with RDCs calculated using a structural model, for example a protein which exhibits high sequence homology. A good fitting in the back-calculation of experimental and calculated dipolar coupling means that there is good agreement between the two structures and, consequently, fold similarity. RDCs depend not only on relative orientation of internuclear vectors (a characteristic intrinsic of a specific fold) but also on the degree and the mode of alignment of protein molecules in the orienting medium. The dependence on the medium, which is described by the components of the molecular alignment tensor, can be extracted using the fitting procedure itself. By varying the magnitude and orientation of the molecular alignment tensor (namely the axial and rhombic components D_a and R), the best agreement between experimental and calculated RDCs will be reached when the correct tensor components are included in the calculation. For this purpose, we chose the program PALES (Zweckstetter et al., 2000) which performs the best-fitting of RDCs giving as input a coordinate file and the list of experimental RDCs. When changing the structure file, for example using a homologous molecule, the results of the fitting operation indicate if the new structure could satisfy the same set of RDCs with respect to the original structure, that is if the new structure possesses a similar fold. A bad fit, in terms of high rmsd of RDCs values, Pearson correlation coefficient and high Q quality factor (Bax et al., 2001), implicates that the fold is quite different. The source of diversity could arise from modest deviations partitioned along the entire sequence or from large distortion in specific regions of the protein while

the rest of the molecule may remain similar to the reference molecule. To this end, it's convenient to perform a selective analysis excluding RDCs that are supposed to show large deviations or including only RDCs in specific regions. Obviously, if one compares small individual secondary structural elements (for example a single α-helix), a good fit generally is obtained. RDCs are especially useful when dealing with the relative disposition of distinct secondary structure elements, for example the orientation of the two recognition helices in E2C protein.

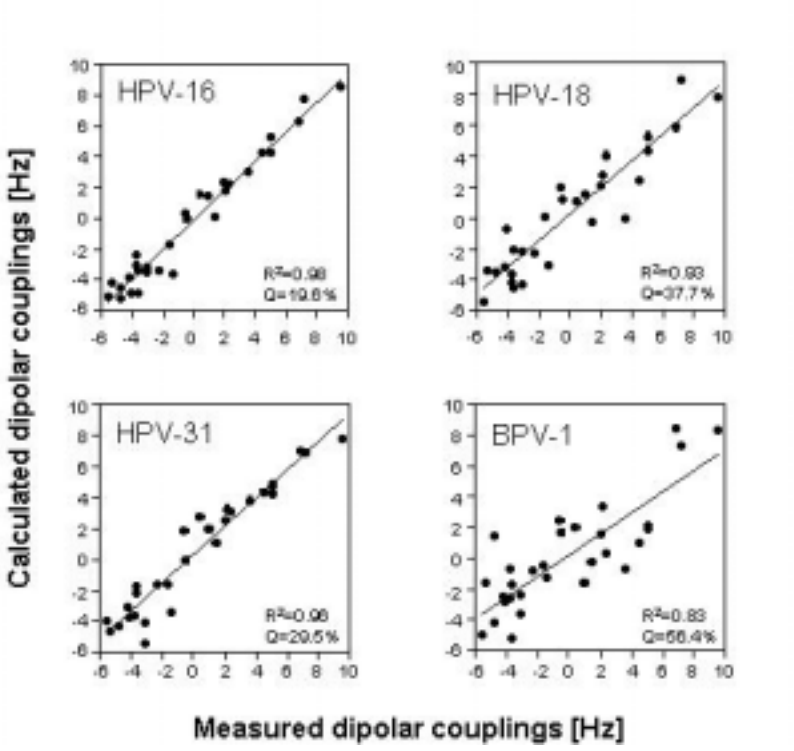

Figure 18. Fitting of calculated to experimental RDCs measured for HPV-16 E2C in polyacrylamide gel using as model the structure of HPV-16 E2C (PDB code 1R8P) solved by us, and the structures of HPV-18 (1F9F), HPV-31 (1DHM) and BPV-1 (1JJH) E2C. R^2 is the Pearson correlation coefficient for the fit while the Q quality factor is a normalized root mean square deviation which measures the predictive power of the structure with respect to satisfying experimental RDC (Bax et al., 2001). Lower the Q factor, better the fit.

In the case of our set of RDCs measured for HPV-16 E2C using polyacrylamide gel to align the protein, we carried out a comparison with E2C structures solved for other strains. Figure 18 shows the corresponding fits for the dimer using 28 RDCs distributed

uniformly along the polypeptide chain (the distribution ensures that we are not sampling a limited region of the structure).

The visual examination itself of the fits evidences how HPV-16 E2C resembles HPV-31 while relatively differs from BPV-1 and HPV-18. Interestingly, when we compare only monomers, much better fit is achieved also for HPV-18 and BPV-1. This result derives from the fact that the main source of structural differentiation between E2C from distinct strains is the relative subunit orientation, while the single monomer is structurally highly preserved. Our example confirms that RDCs can be confidently used to investigate structural relationships between distinct folds without achieving the high-demanding complete structure determination by NMR or X-Ray.

In order to assess the impact of RDCs in the final structure, we have generated 200 structures excluding the 56 RDCs constraints. The twenty lowest-energy structures show a good convergence (rmds of 0.85 Å and 1.25 Å were obtained for backbone and heavy atoms, respectively, when excluding the β2-β3 loop). At the single monomer level, the averaged structures calculated with and without RDC's show a reasonable agreement (rmsd 0.77 Å for backbone atoms, excluding the β2-β3 loop), but significative deviation is found at the level of the relative subunit orientation in the dimeric structure (rmsd 1.50 Å) as apparent in figure 19.

The averaged structure generated without RDC's also presents higher rmsd values with respect to the crystal structure: 1.08 Å (monomer) and 1.53 Å (dimer). When comparing the distances between the two recognition helices, it's notable that the structure calculated without RDC's place them circa 3 Å more distant. This helix shift makes a large difference in the predicted degree of bending for bound DNAs. The RDC effect on structure shows that in such a case NOEs alone are not able to accurately determine structural features at the dimer interface. Errors in this region propagate significantly along

the entire dimer. RDCs can considerably improve the accuracy of NMR structures of multimeric assemblies.

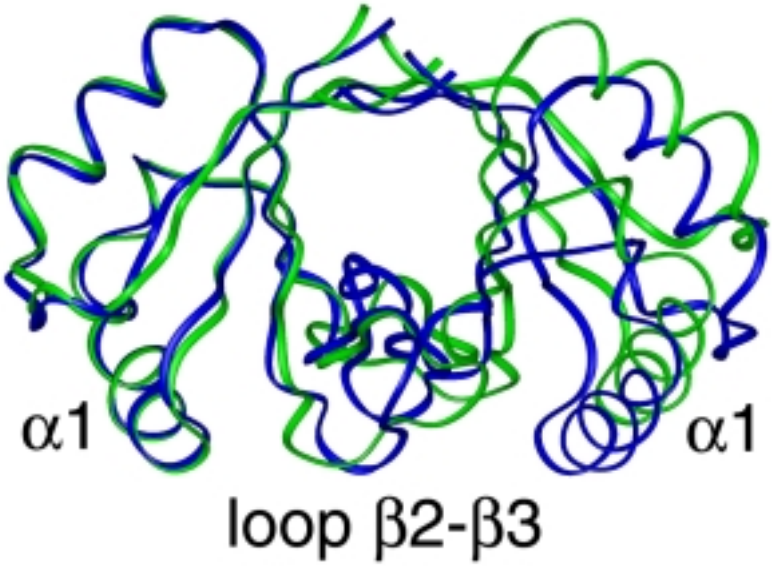

Figure 19. Superimposition of the averaged structure of HPV-16 E2C calculated with (blue) and without (green) RDC's. Only the left monomer is superimposed

E2C-DNA backbone resonance assignment and Chemical Shift perturbation

The appearance of visible and irreversible aggregation upon DNA and E2C mixing is the characterizing feature of HPV-16 E2C complex preparation. We searched different experimental conditions to achieve solubility and stability for the complex using its 18-bp binding site of sequence GTAACCGAAATCGGTTGA, hereafter referred as BS(AAAT). Buffer identity and concentration, ionic strength, pH, temperature and reactants concentration were varied. The best conditions for NMR studies were found to include 50 mM sodium acetate and 10 mM sodium phosphate buffer, pH 6.5, 0.25 M NaCl, 5 mM DTT and a small excess (10-30 %) of target DNA. Moreover, experiments were collected at 45 °C to obtain good quality spectra. Slight modifications of these conditions don't induce dramatic changes in the resonance patterns of correlation spectra but profoundly impact the aggregation state and the stability of the system. In particular, temperature was demonstrated to disfavour aggregation. Some examples of experiments employing ^{15}N-editing conducted at distinct increasing temperatures and different reactant concentrations are displayed in figure 20. The remarkable improvement of spectral signals in terms of sensitivity and resolution are apparent from this comparison. Thermal stability of E2C protein was investigated by Mok and co-workers (Mok et al., 1996) who found a relative high resistance to temperature-induced unfolding. Because DNA binding confers much greater stability to the protein (Ferreiro et al., 2000), the temperature we chose is supposed not to disrupt the dimeric fold of the bound-protein. Ionic strength is a fundamental factor in stabilizing E2C complex and preventing aggregation (a common feature to many DNA-binding proteins complexes). Noteworthy, high ionic strength causes aggregation for free E2C whereas increases solubility for the complex. High concentration of reactants promotes aggregation, especially at lower temperature.

The BS(AAAT) double strand was titrated against increasing concentrations of HPV-16 E2C protein. Comparison of one-dimensional "jump & return" spectra (Plateau and Guéron, 1982) was carried out for the double helix imino protons spectral region (which is free of protein signals because of the strong down-field shift of these hydrogen-bonded protons) at different E2C concentrations. The presence of two distinct sets of resonances during the titration shows that the complex is in slow-exchange with respect to chemical shift differences. This kinetics of binding, as revealed also by analysis of multidimensional correlation spectra of the bound protein signals, is expected because of the very high affinity of E2C complex (nM). Life-times measured in reported binding studies of HPV-16 (Sanders et al., 1994) were found to be relatevilely long (from minutes to hours).

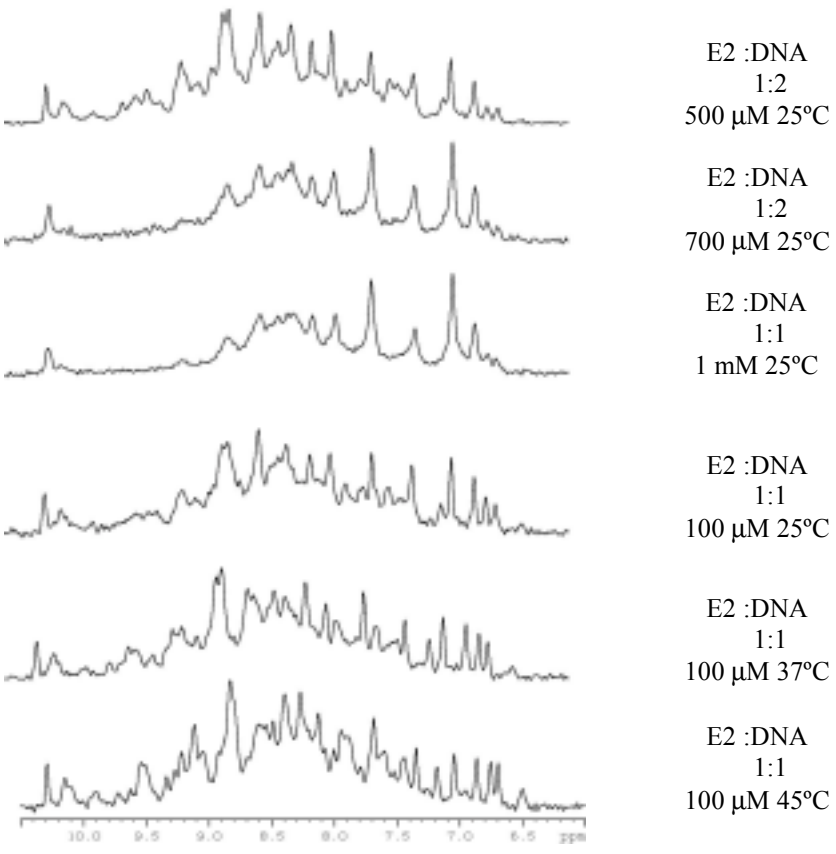

Figure 20. One-dimensional ^{15}N- edited spectra of E2C-DNA complex recorded at 25°, 37° and 45° C at distinct concentrations. The amide NH region is shown.

Nearly all backbone ^{1}H,^{15}N and ^{13}C resonances were unambiguously assigned by means of the same conventional backbone assignment procedure used for free E2C (Ikura et.al, 1990; Bax et Grzesiek, 1993). Because this assignment is achieved by rigorous inspection of through-bond scalar connectivities, backbone chemical shifts can be confidently used to examine any significant change in the magnetic environment of the polypeptide chain.

Figure 21 shows the ^{1}H-^{15}N HSQC spectrum of bound E2C. Only a single set of resonances is observed for ^{1}H-^{15}N correlations which suggests that the two subunits are equivalent in the complex as well as in the free protein.

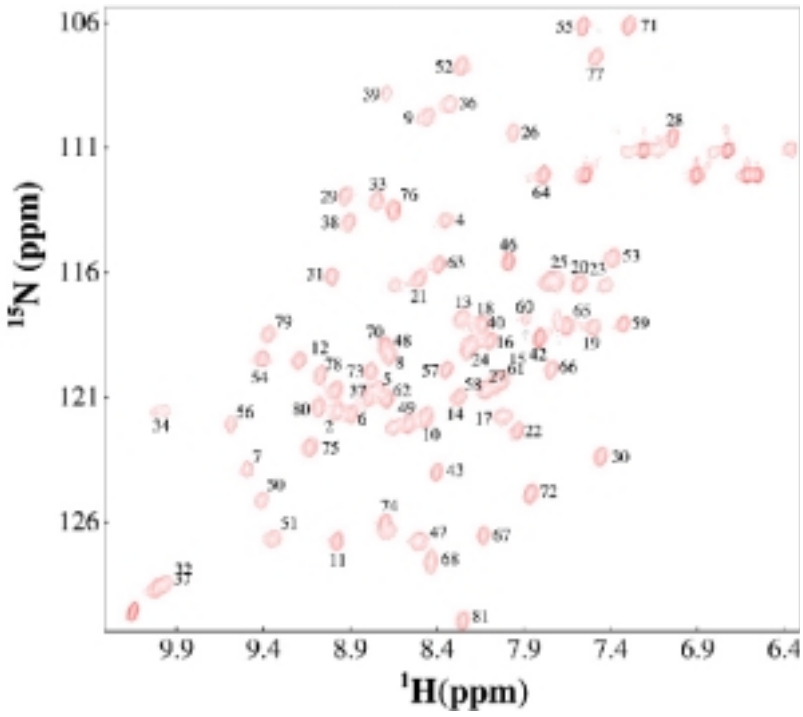

Figure 21. HSQC spectrum of E2C complexed to BS(AAAT) in 10 mM sodium phosphate, 25 mM sodium acetate, pH 6.5, 250 mM NaCl, 5mM DTT at 45 °C. The peaks that were assigned are labelled by their assignment

The comparison with the free protin spectrum evidences that most of ^{1}H-^{15}N correlations don't exhibit significant changes upon binding. The close similarity of spectral

signals is even more striking if we consider that experiments on free E2C and on the complex were acquired in different conditions (50 mM phosphate salt-free buffer for free E2C; 50 mM acetate, 10 mM phosphate buffer and 250 mM NaCl in the case of the complex).

NMR correlation spectroscopy is widely used to investigate the structural alterations that accompany protein-ligand binding and many examples are published demonstrating the capability of Chemical Shift Perturbation to map interaction regions onto protein structure (Zuiderweg, 2002). The tight correspondence between chemical shift and local and global structure implicates that only nuclei that experience a perturbation of their magnetic environment upon protein-DNA interaction display important resonance shifts. Extensive studies were carried out about the prediction of chemical shifts and spectral signal in NMR spectra of model compounds starting from *ab initio* principles (Sitkoff and Case, 1994). Unfortunately, nowadays these methods fail in the case of very complex systems such as proteins and nucleic acids. The dependence of magnetic shielding effects on the chemical structure cannot be quantified in a straightforward manner for biomolecules in which multiple sources of magnetic perturbation introduce too many parameters in the C.S. calculation. For example, 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, NH resonances constitute a formidable tool to characterize and identify a specific protein, representing a finger-print of the molecule of interest. Upon binding of a ligand, NH groups may undergo important changes of their magnetic environment. On the other hand, other nuclei of the backbone are mainly dependent on the local stereochemical and stereoelectronic arrangement. Wishart demonstrated the strict correlation of secondary structure and $^{13}C^{\alpha}$, $^{13}C^{\beta}$, ^{13}CO chemical shifts (with less confidence also of H^{α} and H^{β}

chemical shifts). This property is widely used in the prediction of secondary structural elements (Wishart et al., 1994). When the protein binds to a ligand, chemical shift variations generally accompany the recognition. Resonance shifts can be caused by spatial proximity to ligand moieties or by structural changes within the protein itself. However, proximity effects are relatively short-ranged: for example ring current-induced changes become negligible beyond approximately 5-6 Å. Only if conformational alterations are transmitted along the structure the chemical shift perturbation is observed for protein sites that are distant from the region of direct interaction with the ligand. Therefore, nuclei that don't present C.S. deviations upon binding experience a magnetic environment which is not modified. This indicates that the corresponding regions don't undergo structural transitions in response to ligand binding.

With this regard, analysis of E2C complex correlation spectra provides a sensitive tool to map local interactions between protein and DNA and to obtain detailed and confident description of conformational modifications that eventually occur in the bound-conformation. Figure 22a represents the Chemical Shift Perturbation of ^{1}H-^{15}N resonances along the protein aminoacidic sequence. A similar trend is seen when monitoring C.S. deviations C^{α} nuclei (figure 22b).

Quantitative analysis of structural effects that cause shielding are not confident in such a complex system. Anyway an estimate of statistical significance of C.S. perturbation can be assessed, for example defining a threshold that indicates how important is the change with respect to the distribution of C.S. deviations. Analyzing the dispersion of C.S. differences, the majority of residues appear to exhibit only minimal changes. The average difference is 0.097 ppm for NH groups and 0.11 for C^{α} nuclei, indicating very small overall perturbation. Values above one standard deviation with respect to the mean (which is represented by the line depicted in figures 22a and 22b) are found only for three

regions: the recognition helix and adjacent residues (residues 292- 303); the C-Terminal part of the β2 strand (residues 315-319) and the N-terminal tract of β4 strand (352-355).

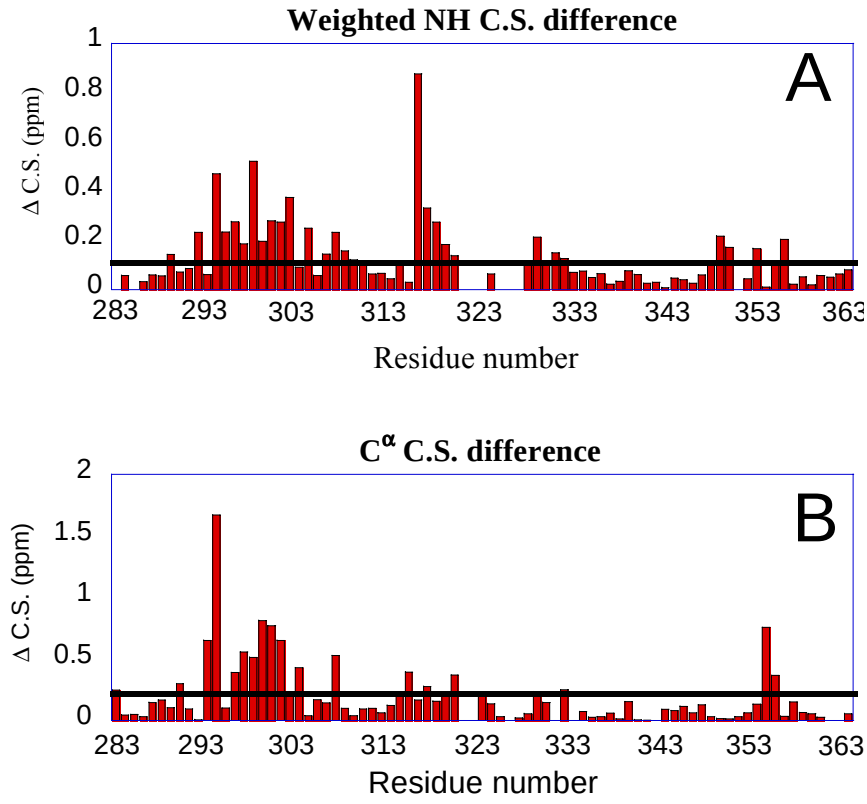

Figure 22. Chemical Shift Perturbation of E2C complexed to BS(AAAT) for (a) NH groups and (b) C^{α} atoms. Lines are drawn in both panel to indicate the mean value of the C.S. difference.

Even a visual inspection evidences that the largest C.S. perturbation is localized in the recognition helices. Both the reported high-resolution structures of E2C complexes (BPV-1: Hegde et al., 1992; HPV-18: Kim et al., 1998) show that the only significant modification upon DNA binding is a small adjustment in the recognition helices disposition, while the overall fold acts as a quite rigid scaffold. The other two regions affected by slight changes in the magnetic environment are β2 and β4 strands. These elements are critical for dimer formation and stability in the free protein.. The rest of the protein displays an interesting resistance to binding-induced structural modifications. The

trend is more evident for residues occupying the protein core. This result suggests that there is minimal binding-induced movement in the complex structure. The protein is predisposed to recognize its DNA target with minimal energetic variations because of the high enthalpic cost implicated in perturbing the densely packed subunit interface.

Only four amino acids were found to establish direct contacts with DNA bases. These residues are absolutely conserved: Asn293, Lys296, Cys297 and Arg344 (Kim et al., 2000; Hegde, 2002). In the structures of complexes determined by X-Ray the side chains of these amino acids make direct contacts with DNA bases in a interdigitated way: each amino acid establishes hydrogen bonds with two bases and each base contacted by the protein interacts with two amino acids. A rough structural model of HPV-16 complex may be built by docking the DNA target onto the interaction surface that is mapped by C.S. perturbation of our DNA-free protein solution structure. A detailed inspection of this model and comparison with crystallographic bound-conformations could be a promising way to gain information about regions directly involved in DNA binding.

The accurate examination of correlation spectra lineshapes may reveal the presence of protein regions characterized by very different motion properties with respect to the rest of the molecule. For example, residues involved in fast and large internal motions exhibit more intense and narrow peaks. When rising the visualization contour level, these signals clearly emerge with respect to the remaining peaks that disappear from the screen display. On the contrary, residues situated in regions subjected to conformational exchange experience broadening and weakening of signal. Some of these features are found in E2C complex correlation spectra. Cross-peaks of the segment comprising residues 320-327 have much greater intensity and small line-width relative the other peaks and suggest that the disordered central loop in the free form maintains high degree of mobility in the bound state. Relaxation data analysis seems to confirm this hypothesis (see below).

Backbone analysis dynamics for free and bound E2C

^{15}N Relaxation rates and dynamics of free HPV-16 E2C. The ^{15}N relaxation rates T_1, T_2 and the steady-state ^1H-^{15}N NOE of the backbone amides measured at 40.5 (Figure 23a) and 70.9 MHz (Figure 23b) were analyzed to describe the global and local motions of the HPV-16 E2C dimer in solution. Relaxation data appear to draw patterns of complementary shape. In the unrealistic case in which the isotropic overall tumbling is the unique source of spin relaxation, ^{15}N T_1, T_2 and ^1H-^{15}N NOE should exhibit uniform values without any deviation from the mean. The presence of internal motion affects locally the spectral density function and induces residue-specific changes in the relaxation behaviour of NH groups. As a consequence, correlated variations occur in relaxation data. In particular, ^1H-^{15}N NOE is the quantity more sensitive to fast internal motions for the range of relaxation rates found in proteins of this molecular size, presenting reduced or negative values for fast motions of large amplitude. In contrast, ^{15}N T_2 are considerably influenced by motions in the μs-ms range which could accelerate the defocusing of transverse magnetization.

Data were interpreted on the basis of statistically different best-fitting models derived from the generalized model-free formalism of Lipari and Szabo, modified to eventually include the contributions from anisotropic tumbling and exchange processes. Motional properties of proteins in solution can be described in the simplest case by the use of Lipari and Szabo model (1982) which provides analytical expressions for relaxation rates of ^{15}N nuclear spins. This model predicts that the spectral densities at the frequencies of interest for ^{15}N relaxation are determined by two distinct and independent terms: the overall molecular tumbling measured by the global correlation time τ_m and fast local internal motions, described by the two site-specific parameters S^2 and τ_e. S^2 is a generalised order parameter measuring the degree of spatial restriction of rapid internal motions of the fluctuating N-H bond vector, while τ_e is the rate of these rapid internal motions. An

extended version of this approach includes the calculation of the quantity R_{ex}, which accounts for the contributions of exchange processes (chemical or conformational exchange) to relaxation data (namely to T_2).

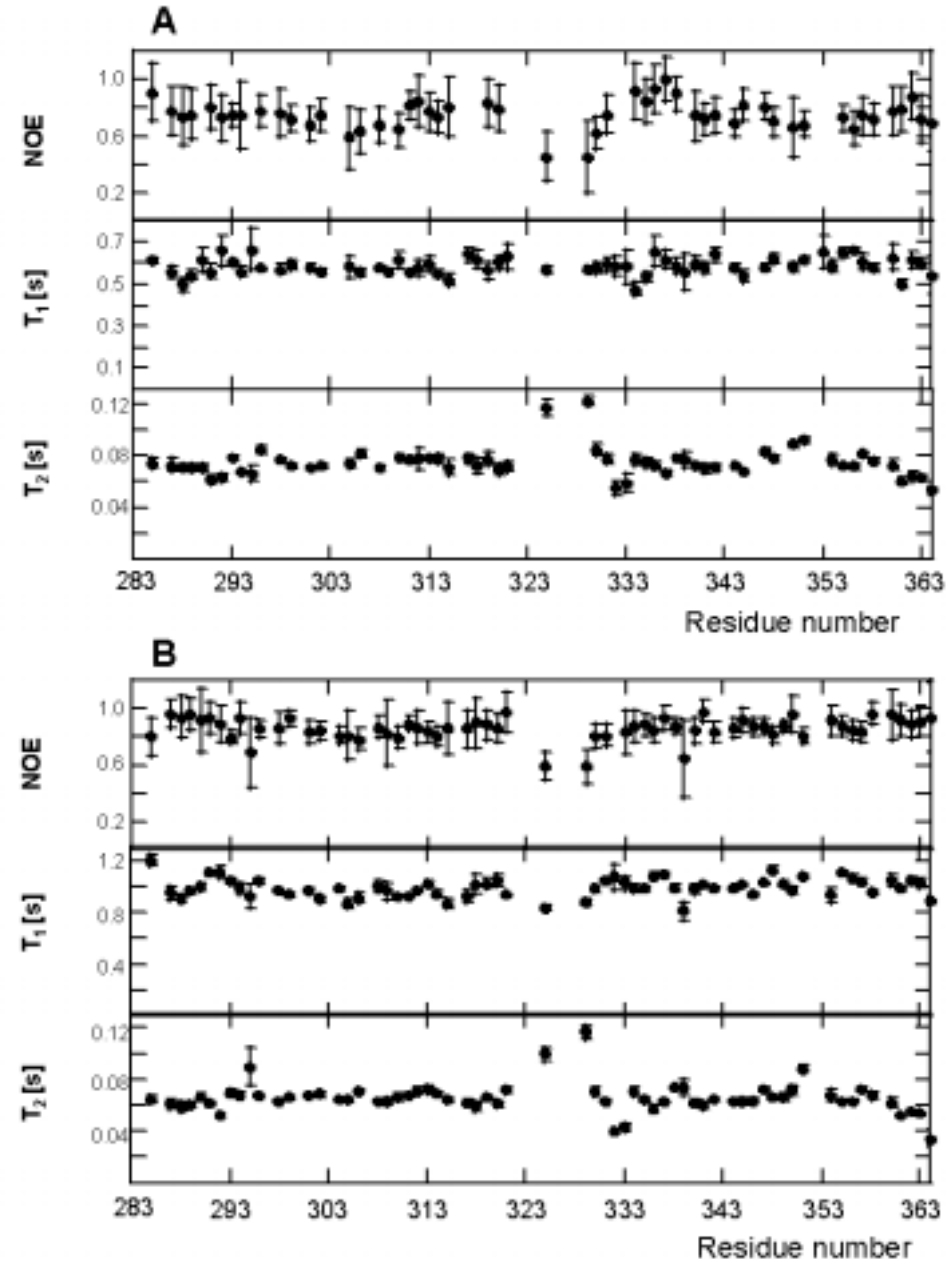

Figure 23. HPV-16 E2C ^{15}N relaxation data vs residue number. Longitudinal T_1, transversal T_2 and steady-state heteronuclear ^{1}H-^{15}N NOE were measured at 40.5 MHz (panel A) and 70.9 MHz (panel B) at 30° C.

Prior to assign to exchange processes the observed shortening of T_2, we need to determine the contribution of anisotropic tumbling to the observed relaxation rates, as it is

known that anisotropy effects in certain cases can be difficult to distinguish from slow motions (Pawley et al., 2001). In facts, ^{15}N relaxation rates in a molecule characterized by anisotropic diffusion tensor are determined also by the specific orientation of the NH vector with respect the tensor axis frame. Using the average solution structure of HPV-16 E2C reported here, we calculated the three components of the inertia tensor, obtaining values of 1.00:1.05:1.42 for D_{xx}, D_{yy} and D_{zz}, respectively. This result suggests that HPV-16 E2C can be represented by a slightly prolate ellipsoid presenting an axially symmetric rotational diffusion tensor with an anisotropy, $D_{\parallel}/D_{\perp}$, of 1.38. For this kind of molecular shape, it has been shown that NHs making an angle smaller than the magic angle (54.7˚) present a lowered T_2 with respect to the average value, and that the deviation is significant for very small angles (Pawley et al., 2001). In addition to this deviation, an increase in T_1 is also expected in such cases, leaving the T_1/T_2 ratio close to the average value. On the contrary, in a prolate ellipsoid NHs making an angle larger than the magic angle (54.7˚) exhibit T_2 values higher and T_1 values lower than the mean. In this case, the T_1/T_2 ratio displays significant deviation and reveals the presence of anisotropic motions. The average value for T_1/T_2 ratio considering all detectable residues is 15.4 ± 3.3 for E2C, showing reduced dispersion. Only the last four residues of the protein (359-362) presents values ranging from 19.0 to 25.6, above one standard deviation. The analysis indicates that chemical exchange is the main contributor to the observed relaxation rates. This observation was further corroborated by applying a variable τ_m in the Lipari-Szabo model (Phan et al., 1996) and calculating the anisotropic contributions using the average solution structure of HPV-16 E2C as a model. With this method a different τ_m is calculated for each residue and anisotropy effects are readily recognized if there is large dispersion in sequence-specific τ_m values. Both approaches show that the isotropic model is better suited to account for the global tumbling of HPV-16 E2C in solution.

Using the generalized Lipari and Szabo model under the assumption of isotropic tumbling we extracted a global correlation time of 10.8 ns from the best-fitting of calculated to experimental ^{15}N T_1, T_2 and ^{1}H-^{15}N NOE. At the same time residue-specific values for S^2, τ_e and R_{ex} were obtained, which are shown in Figure 24a plotted as function of residue number.

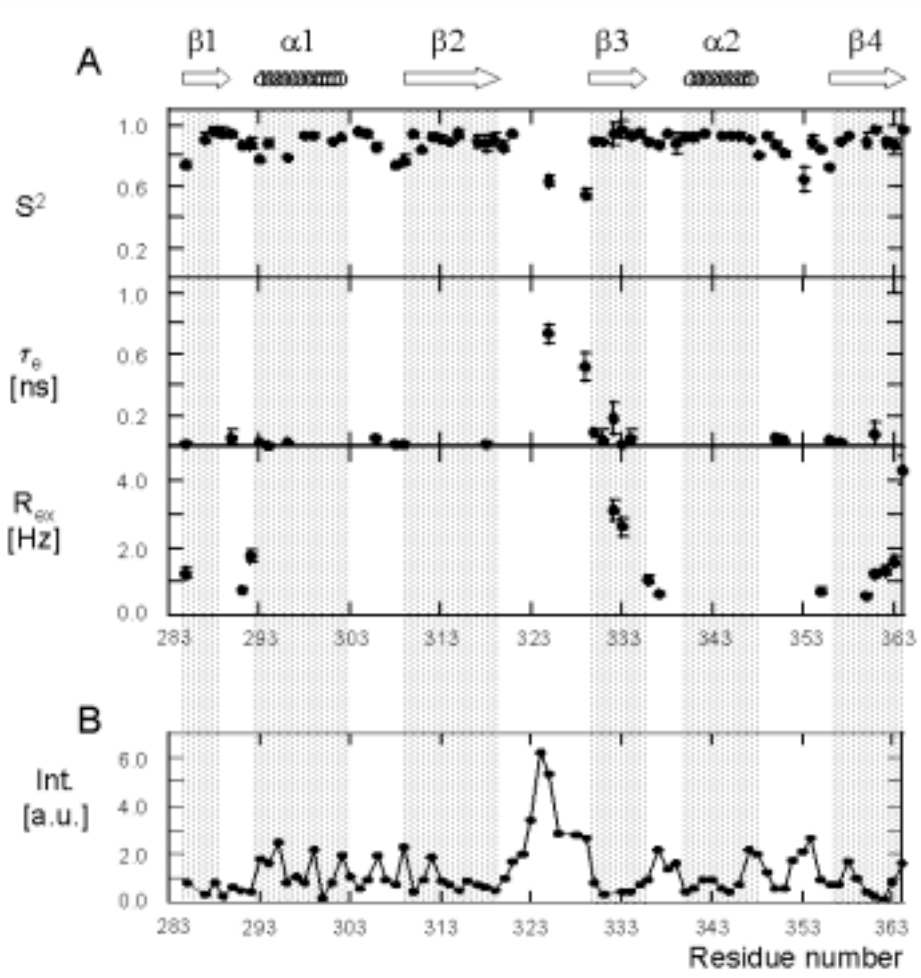

Figure 24. (a) Model-free parameters S^2, τ_e and R_{ex} of backbone amides derived from analysis of the HPV-16 E2C relaxation data. The secondary structure pattern for HPV-16 E2C is indicated in the upper portion of the figure. (b) Intensity of ^{1}H-^{13}C correlation peaks of H^{α}-C^{α} groups in the constant-time HSQC experiment. Intensities are given in arbitrary units.

S^2 could be considered as a measure of the importance of internal motions on the relaxation properties with respect to the global molecular tumbling. More S^2 values deviate

from unity, more influent is internal mobility. Moreover, we must add that R_{ex} magnitude doesn't represent a precise measure of slow exchange processes. An accurate quantitative estimate of the exchange effects requires different approaches. It should be considered as a qualitative indication of the presence and of the degree of such a kind of mobility. The model used herein indicates the existence of fast motions with relatively large amplitude in the long loop connecting the β2 and β3 strands. Only two NHs belonging to this loop can be detected for HPV-16 E2C at 30˚ because of large exchange rates with the solvent displayed by the other amide groups comprised in this segment. This flexibility can be further confirmed by measuring the intensity of ^{1}H-^{13}C correlation peaks of H^{α}-C^{α} groups in a constant-time HSQC experiment (figure 24b). Strictly speaking, the intensity of the correlation peak does not depend only on the ^{13}C T_2 and therefore cannot be used to analyze the motions quantitatively. Although it depends also on the ^{1}H T_1, T_2 and ^{13}C T_1, their influences on the final intensity of the signal is less important and this intensity can be used as a qualitative indication about the ^{13}C T_2. Figure 24b clearly shows that H^{α}-C^{α} signals of residues belonging to the central loop present intensities that are at least two times larger than the average values for other residues. The flexibility of the central loop was also observed for HPV-31 and BPV-1 E2C in solution (Liang et al., 1996; Veeraraghavan et al., 1999) and suggested by a low electronic density in the diffraction data of HPV-18 (Kim et al., 2000) and HPV-16 E2C (Hegde et al., 1998).

Other fast motions, but with relatively smaller amplitudes, can be deduced from the ^{15}N relaxation data profile in the turns connecting β1 with α1, α1 with β2 and α2 with β4. Interestingly, some degree of mobility was also observed for the first three residues of the recognition helix.

In addition to these fast motions, the application of the isotropic model suggests μs-ms motions that increase the ^{15}N linewidth for some residues located in the loops

connecting β1 with αι and α2 with β4. More significant are the relatively large values for R_{ex} observed for residues belonging to β2 (Val331, Thr332) and almost the whole β4 (Ser357-Ile362).

^{15}N Relaxation rates and dynamics of DNA-complexed HPV-16 E2C. Dynamics of the complex was investigated acquiring ^{15}N T_1, T_2 and ^{1}H-^{15}N NOE at 70.9 MHz. Relaxation rates of HPV-16 E2C-DNA are displayed in figure 25.

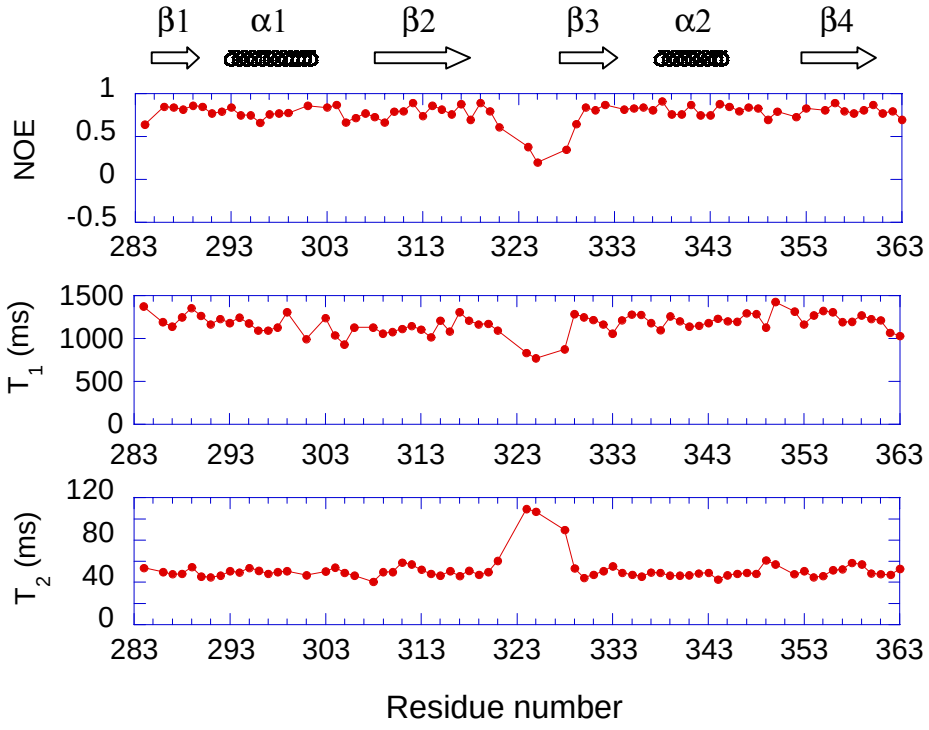

Figure 25. HPV-16 E2C-DNA ^{15}N relaxation data vs residue number. Longitudinal T_1, transversal T_2 and steady-state heteronuclear ^{1}H-^{15}N NOE were measured at 70.9 MHz for E2C complexed to BS(AAAT). The secondary structure elements as predicted by the Chemical Shift Index for HPV-16 E2C are displayed in the upper portion of the figure.

The extended model-free formalism could be applied to the complex but caution is needed when using this approach. In facts ^{15}N relaxation data for the free protein were collected at two distinct Larmor frequencies (40.5 and 70.9 MHz) and six relaxation quantities for residue were yielded in this way. Relaxation rates for the complex were measured only at the higher frequency (70.9 MHz) and only three independent data are

available for each residue. Because in the model a set of three parameters for each single NH site are virtually given by the back-calculation process and because large errors are associated to these measurements, the analysis could lead to misleading results if the fitted parameters are not overdetermined by the experimental data. All the information regarding the molecular mobility of the complex can be derived by simple direct inspection of T_1, T_2 and ^{1}H-^{15}N NOE patterns. The relaxation rates are only slightly dispersed about mean values along the sequence except for the central loop. The very high values for T_1 and the very low values for heteronuclear NOE at the middle of the profile show that this region experiences relaxation rates that reflect fast internal motions. This is a further and stronger indication that the loop maintains high mobility in the complex. Structures of E2C complexes determined by NMR methods are not available for any Papillomavirus strain. X-Ray studies reveal a difference between the features of the loop segment in HPV-18 with respect to BPV-1 complex (Kim et al., 2000; Hegde et al., 2002). In the bovine complex the loop is anchored to the DNA making few direct and solvent-mediated contacts with the phosphate backbone of the binding site. In contrast, electron density is clearly absent for the corresponding residues of HPV-18 DNA-bound protein. It should be added that DNA oligomers used for the complexes possess distinct sequence in the two cases. Our study is the first direct structural investigation of the "E2C loop question" in HPV-16 complexed form. The binding site employed in our experiments has a sequence different from the other two. If this conformational characteristic of the central loop depends on the specific protein or on the identity of the binding DNA is subject of debate.

When carrying out direct comparison of relaxation data between free and bound form it must be taken into account that the variations in relaxation rates mainly reflect the different molecular size of the two systems. Moreover, in this study measurements were performed at different temperatures (30 and 45 °C) for the two forms. Temperature is intuitively

expected to enhance molecular motions and flexibility. On the other hand, the bound state is expected to be more rigid then the non-complexed state. Nevertheless, many examples of increased flexibility of some regions upon protein-ligand binding were found and have been explained as a means to compensate unfavourable entropy loss. The mobility displayed by the unbound protein in the μs-ms range, which causes defocusing of transverse magnetization, appears not to be very influent in the complex. In facts, the T_2 values don't exhibit large dispersion and seem to follow a uniform trend when the central loop is excluded from examination. Also those regions that are markedly affected by exchange processes in the absence of DNA show relaxation rates predominantly governed by the global molecular tumbling. In particular, the recognition helices present a relaxation behaviour quite different with respect to the free form and their characteristic flexibility is quenched in the bound conformation. These observations suggest that the α1 helices and in general the entire structure becomes considerably more rigid upon binding, resulting highly constrained in the bound conformation.

Amide exchange rates

Amide proton exchange rates and NOEs with water of HPV-16 free E2C. Hydrogen exchange rates can be used to gain structural information and to monitor internal motions of folded and unfolded proteins (Raschke and Marqusee, 1998). The exchangeable hydrogens in proteins comprise the main-chain amide group and the side chain protons covalently bound to N,O and S atoms of polar groups. Amide NH exchange with the solvent is commonly studied by NMR methods, which offer residue-resolved specificity. Hydrogen exchange rates depend on environmental factors such as temperature, isotope involved, solvent, buffer salts and, in particular, pH. In addition, protein structure has an enormous influence on exchange mechanisms: amide groups can be dramatically protected from solvent exchange. Hydrogen bonding, low solvent accessibility and steric blocking can reduce exchange rates even by 6-7 orders of magnitude relative to exchangeable protons in unstructured peptides. Because of the physical protection of the amide site, the exchangeable proton is not accessible to solvent and to catalyst molecules (OH^- in the pH range employed for native states of proteins). Removal of this physical blocking can be achieved by an "opening" event. Different mechanisms are supposed to promote the opening step that precedes the chemical proton exchange reaction in highly protected sites: local fluctuations, transient breakage of hydrogen bonds, penetration of the catalyst molecule, unfolding reaction (Englander et al., 1996; Raschke and Marqusee, 1998).

Chemical exchange and NOE-mediated magnetization transfer are the two ways that mediate magnetization exchange between backbone amide protons and water. Magnetic and proton chemical exchange rates are correlated and can be used to indirectly assess hydrogen-bond stability and solvent accessibility of different regions of a protein. The amide exchange rates are typically measured for individual residues by monitoring the

isotope- proton exchange over time using NMR one-dimensional and two-dimensional spectroscopy. A vast range of experimental exchange rates is observed for a folded stable protein. NHs situated in solvent-accessible and flexible regions usually exchange too quickly with the solvent to be detected on the NMR time scales. NHs located in structured regions exhibit exchange rates spanning a broad range of values that can be measured by NMR techniques. A set of exchange experiments was performed on HPV-16 E2C by collecting ^{1}H-^{15}N HSQC 0.5 h, 2 h, 4 h, 12 h and 24 h after dissolving the sample in D_2O at pH* 6.5 and 30˚C. A summary of the remaining peaks in each experiment is shown in Figure 26a.

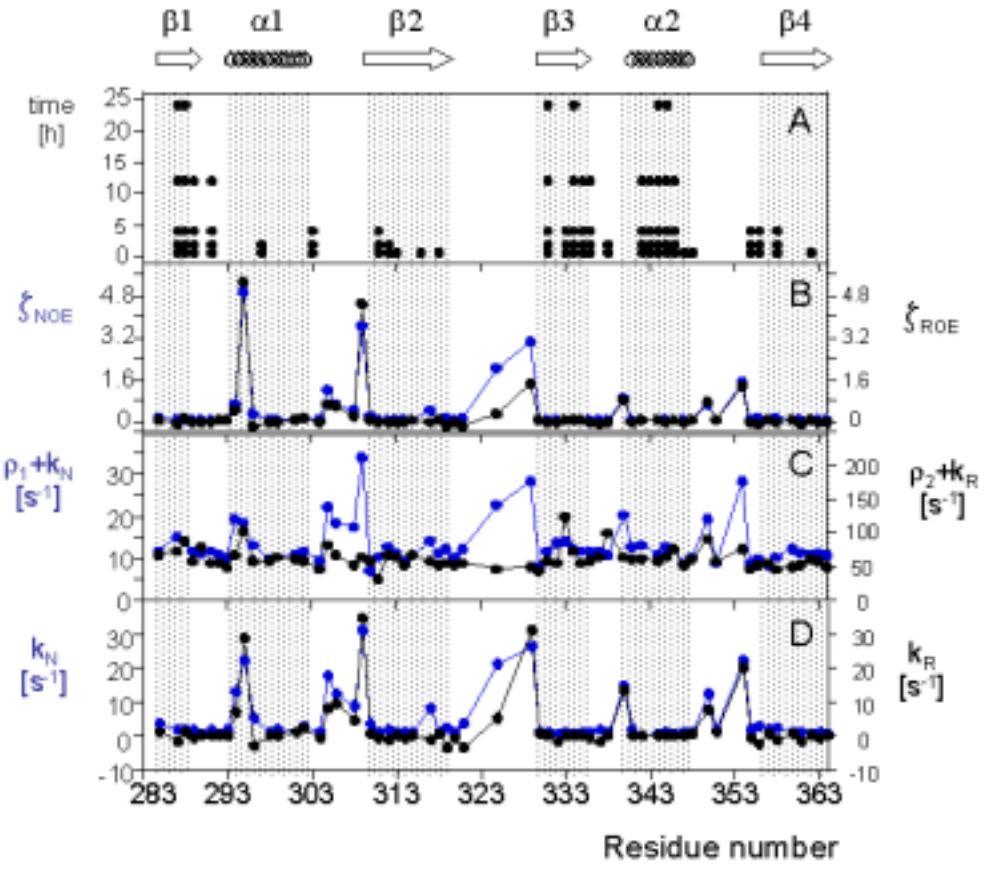

Figure 26. (a) H/D exchanging amide protons of HPV-16 E2C protein. Filled circles represent peaks that are detectable at 0.5 h, 2 h, 4 h, 12 h and 24 h after dissolution of the liophilized protein in D_2O. The secondary structure pattern for HPV-16 E2C is indicated in the upper portion of the figure. (b, c, d) Amide-water magnetization

transfer rates: (b) ζ_{NOE} and ζ_{ROE}; (c) $\rho 1 + K_N$ and $\rho 2 + K_R$; (d) k_N and k_R

The first observation from these data is that after 12 h, only NH peaks corresponding to residues 286-290 (strand β_1), 333-335 (strand β_3) and 341-345 (α_2) are still visible and only six NH peaks are detectable after 24 h. These residues, I286 and V287 of β-1, I330 and L333 of β-3 and residues N343 and F344 of α-2, are part of the hydrophobic core of each monomer, which results from the packing of the two helices against the outer face of the β-barrel. The slowly exchangeable protons are located in the segments that define the most stable hydrophobic region of the protein in solution. In contrast, literature is rich in examples of very slow exchange rates assayed in similar conditions in other proteins. These very slow exchangeable sites are usually found in the core of the protein structure and are often associated with cooperative folding units individuated in equilibrium and kinetics unfolding experiments (Kim et al., 1993). The low number of slow exchangeable amide groups suggests that HPV-16 E2C protein is characterized by a generalized high solvent accessibility.

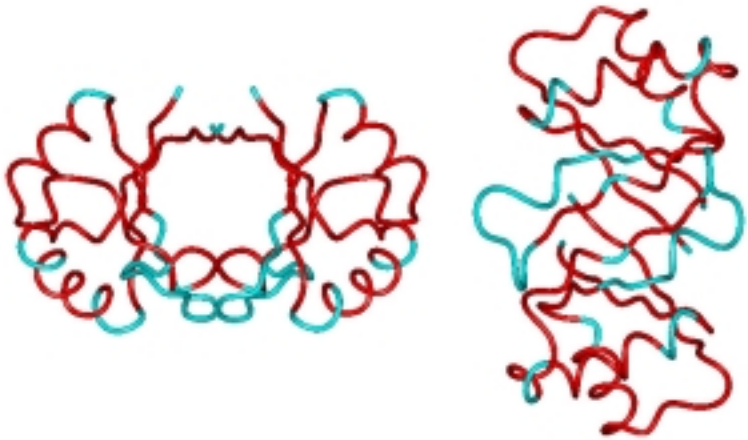

Figure 27. Backbone representation of HPV-16 E2C showing in cyan the residues displaying the highest exchange rates. Two views from orthogonal directions are presented.

Secondly, from the six regions showing a defined secondary structure, the two most exposed β-strands, β_2 and β_4, and the highly exposed helix α_1 show no protection against solvent exchange after a few hours. Figure 27 shows the location of these exchangeable protons. Strands β_2 and β_4 are directly involved in the dimer interface formation, and their relatively high water exposure suggests that the whole dimer could breath. The remaining structured regions, β_1, β_3 and α_2 show a higher degree of protection.

To gain more information about exchange rates of the faster NHs (in the range 0.1-100 s^{-1}), we used a second strategy that involves the measurement of magnetization exchange rates between amide protons and water. Magnetization can be transferred from water protons to protein amide protons by chemical exchange between the two sites or by dipolar interaction due to the vicinity of the two protons. We followed the method developed by Grzesiek and Bax (1993) which combines information obtained from different ^{15}N and ^{1}H relaxation times (T_1, T_{1zz}, $T_{1\rho}$), and two quantities, ζ_{NOE} and ζ_{ROE}, measured using 2D experiments on a $^{13}C/^{15}N$ labelled protein sample. The experimental quantities ζ_{NOE} and ζ_{ROE} measure the efficiency of magnetization transfer from H_2O to amide protons in these experiments. They have to be corrected for the fraction of water magnetization which effectively establishes NOE interactions with amides groups. In addition, they are corrected for water proton relaxation effects described by the amide proton-spin flip rate ρ1 and spin-locking relaxation rate ρ2 (see Material and Methods). In this way, values of magnetization exchange rates k_N and k_R are extracted. These new quantities constitute a measure of the rate at which magnetization is transferred from H_2O to amide protons. The values of ζ_{NOE}, ζ_{ROE}, $\rho1 + K_N$, $\rho2 + K_R$, k_N and k_R are presented in figure 26 (b, c and d). The magnetization exchange rates k_N and k_R are evaluated to distinguish between the two main mechanisms of magnetization exchange: if

magnetization exchange is due exclusively to hydrogen exchange**,** k_N equals k_R, while if all magnetization transfer is due to NOE interaction, k_R approaches $-2k_N$ (Grzesiek and Bax, 1993). The graphic of Figure 27d shows that all NHs located in the loops connecting regions with secondary structure present high and similar values of both k_N and k_R, indicating that hydrogen exchange is the main source of magnetization transfer to water. For these protons, the absence of hydrogen bonding and a high solvent accessibility explain this result. Six NHs from the long and flexible loop connecting β_2 to β_3, residues 321-323 and 325-327, are not detectable in the ^1H-^{15}N HSQC indicating a high exchange rate with the solvent.

On the other hand, magnetization exchange rates of NHs belonging to regions with well defined secondary structure fall into the 1-3 s^{-1} range. Exception to this behavior are Ala293 (13 s^{-1}), Asn294 (22 s^{-1}), Thr295 (5 s^{-1}), and Lys304 (18 s^{-1}) in α_1, Thr316 (8 s^{-1}) in β_2 and Trp339 (15 s^{-1}) in α_2. Residues 293-295 are the first three residues of the α_1 helix, and hence are not hydrogen-bonded. Although this is common to all α-helices, their exchange behaviour is in marked contrast with that observed for the first three residues of α_2. Even though the first residue (Trp339) shows a high magnetization transfer rate, the second and third residues show low values. The generalized order parameter values, S^2, for these three residues are lower with respect to other residues within the helix (see Figure 24), indicating some degree of flexibility. Lys304 is the last residue of α_1 and its NH deviates somewhat with respect to the other NH of the helix. As a consequence of this, the amide proton of Lys304 is not pointing directly to the carbonyl oxygen of Arg300 as it is expected in a regular α-helix. No evidence of flexibility was observed for this residue from the ^{15}N relaxation analysis. Thr316 is located in a bulge of β_2 with its NH group pointing almost perpendicular to the plane of the β-sheet formed by the intermonomer β_2-β_2 contacts, and therefore is markedly accessible to exchange with water (figure 28). This

accessibility is functional for the interaction of HPV-16 E2C with DNA, as it was shown that the NH of the corresponding residue of Thr316 makes contacts to phosphate in the crystalline structure of BPV1-DNA complex (Hegde et al., 1992).

Figure 28. Enlarged view of the dimer interface showing Thr316 residues belonging to the two opposing β2 strands. For clarity, only backbone atoms are displayed. Carbon atoms are coloured in green, nitrogen in blue, oxygen in red and protons in gray. Hydrogen bonded atoms of the flanking residues are joined by a dotted line.

Some amide protons involved in secondary structure elements display relatively high exchange rates. Secondary structure elements are defined by the geometric restraints required to optimise their characteristic hydrogen bonding network and are observed to exchange typically slower than structured turns. Slow hydrogen exchange is not simply related to intramolecular hydrogen bonds formation. The exchange rate depends on the dynamic accessibility of catalyst and water molecules to the NH site. Local and global fluctuations can break the restrained geometry of an H-bonded amide proton and transiently expose it to exchange with solvent. Exchange is promoted by transient high-energy structural fluctuations that remove the physical blocking of amide protons (Englander et al., 1996).

Two main mechanisms are supposed to explain the exchange behaviour of amide protons in native proteins (Kim et al., 1993). Exchange may occur from the folded state,

that is from a local conformation in which amide sites positions are only slightly perturbed with respect to the native conformation. In contrast, exchange may also occur from transient unfolding events which can involve only local segments (partial transient unfolding) or can involve the entire structure (transient global unfolding). In the first case protons that are very close in the molecule could exhibit quite different exchange rates because the conformational alteration required for the exchange reaction is strictly local. In the second case protons that are located in the same exchange unfolding unit should display similar protection factor. Strictly speaking, to assess the presence of such exchange unfolding units pulse-labeling experiments should be conducted for kinetics intermediates in the folding pathway (Rashke and Marqusee, 1998). Alternatively, mutational plasticity should be assayed and correlated to exchange rates in specific mutants (Woodward et al., 1993).

Amide protons inside HPV-16 E2C recognition helix exhibit distinct exchange rates depending on their location in the solvent-exposed surface or in the interior hydrophobic side, which is densely packed against the β-barrel protein core. Similar features are shown by α2 helix. These observations suggest that the two amphoteric helices are packed and not transiently unfolded when exchange occurs. The model-free analysis of ^{15}N relaxation data indicate that internal motions only of the recognition helix and of turns connecting α-helices and β-strands present large amplitude and fast frequency (figure 24). On the other hand, in our opinion the local folded state fluctuation model cannot account for the generalized high solvent accessibility of HPV-16 E2C. Taking into account the high hydrophobic nature of the protein interior it's difficult that water molecules can penetrate to reach the inner exchange sites so frequently. The conformational mobility manifested by α1, β3 and β4 strands in the μs-ms time scale can be correlated and reflect large motions of the protein representing a sort of breathing. These slow movements probably involve the

relative subunit displacement and open transient channels between single structured regions that facilitate water diffusion and hence solvent accessibility of amide protons.

High flexibility was individuated for the recognition helices in the NMR studies conducted on HPV-16 (Liang et al., 1996) and BPV-1 (Veeraraghavan et al., 1999) E2C proteins. Using only exchange rates and ^{1}H-^{15}N NOE, they found a difference in the dynamic properties of the two α1 helices with respect to the rest of the molecule which appeared more compact and protected (only the central loop showed high mobility). They related this difference to a possible functional role of recognition helix flexibility during the DNA-binding reaction. We also observed high flexibility for the recognition helix of HPV-16. These motional properties are now deduced from an extensive set of nuclear relaxation data combined to accurate measurement of solvent exchange rates. If important flexibility is confirmed for α1 helices, an extended generalized mobility is apparent for all the molecule. The high-resolution structure is well defined and ill-determined regions are not evident. The protein behaves as a highly compact protein, with a dense hydrophobic core and relevant stability. Internal fast motions of significant amplitude characterize the structure turns which are critical "hinges" for the protein fold. Local fluctuations when are combined together could lead to transient opening of the structure in some regions, such as the dimer interface, promoting larger motions that involve a little relative displacement between folded compact units. Exchange rates were used to individuate submolecular motional domains (Kim et al., 1993), which are protein regions displaying distinct motional properties. In HPV-16 structure the β2/β3 loop is an independent motional domain, but also the recognition helices show a behaviour that is distinct from the core region. Structural information could be transmitted between compact submolecular regions through crucial motifs that regulate the global flexibility of the protein. The plasticity of HPV-16 E2C is a salient feature of this fold and can be involved in the recognition process.

Magnetization transfer rates from water in HPV-16 bound E2C. Isotope exhange rates measurements on E2C complex were precluded because lyophilization prevents the solubilization of the sample in an irreversible way. However, the relatively fast exchange rates were measured using the mentioned method of Grzesiek and Bax (1993). In figure 29 a direct comparison of the bound and free form magnetization rate ζ_{NOE} is presented. This quantity is directly related to the transfer rate of magnetization from water.

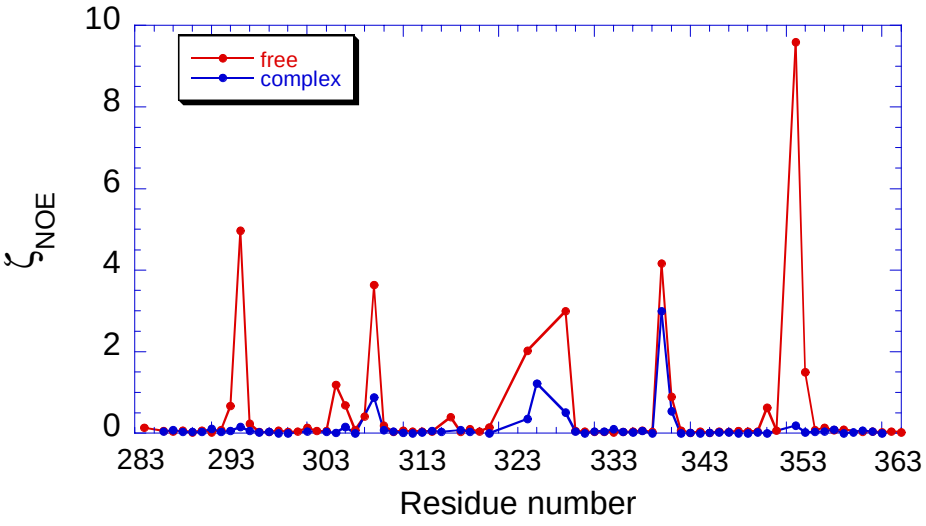

Figure 29. Magnetization transfer rate ζ_{NOE} for free (red) and bound (blue) HPV-16 E2C

It emerges that exchange is dramatically quenched in the bound form. Exception to this trend is the elongated β2/β3 indicating, as seen for relaxation data, that it exists in a flexible conformation in the complex. The turn that connects β2 and α2 appears to show low protection from solvent exchange as in the protein. Other turns connecting secondary structure element result quite protected. We must add that this kind of experiments are suitable to monitor exchange rates in the range 0.1-100 s^{-1} and it cannot be excluded that slightly slower exchange processes occur which escape this analysis. Nevertheless, when compared to the free protein the exchange profile is more uniform and the fast exchange

rates found for some regions of E2C are dramatically reduced, although these experiments were conducted at higher temperature (45°C) which is supposed to enhance exchange with the solvent.

In conjunction with the relaxation rates recorded for the complex, experimental NMR

δατα συγγεστ τηατ τηε οϖεραλλ φλεξιβιλιτψ δισπλαψεδ βψ τηε προτειν ιν τηε φρεε στατε ισ χονσιδεραβλψ θυενχηεδ. Τηε διφφερεντ συβμολεχυλαρ ρεγιονσ τηατ αρε πρεσεντ ιν τηε υνβουνδ στατε δισπλαψ α υνιφορμ βεηαϖιουρ ιν τηε βουνδ φορμ. Τηε ονλψ παρτ οφ τηε μολεχυλαρ χομπλεξ τηατ σηοωσ high mobility is the β2/β3 loop and this could have important consequences on the binding thermodynamics, limiting the entropy loss.

Assignemnt of ACCGACGTCGGT and ACCGAATTCGGT binding sites

Structural and biochemical data demonstrate that E2-BSs adopt B-DNA conformation. Therefore, nearly complete proton resonance assignment for ACCGACGTCGGT and ACCGAATTAACGGT duplexes was obtained following the standard approach used for double-stranded DNA in the B form (Wuthrich, 1986; Roberts, 1993). The cytosine H5-H6 and thymine Me-H6 resonances were readily identified by their correlation peaks in TOCSY spectra. Sugar spin systems appear dispersed in TOCSY spectra and were correlated to base protons of the same nucleotide examining intranucleotide sugar-base NOE crosspeaks. Therefore, the series of sequential intrastrand NOE connectivies between base H6-H8 to sugar H1', H2' and H2'' were individuated in NOESY spectra, allowing to follow the "sequential walk" along individual strands. Figure 30 shows the part of the H6/H8 region of the NOESY spectrum for BS(ACGT) oligomer. Assignment of exchangeable imino, amino and non-exchangeable adenine H2 protons was obtained from NOESY spectra in H_2O. Imino resonances were assigned by sequential imino-imino connectivities between adjacent base pairs, while amino proton and Adenine H2 resonances were assigned via cross-peaks with imino resonances. A1 and T12 labile protons exchange too fast with the solvent to be observable.

Heteronuclear correlation experiments (2D ^{1}H-^{13}C HSQC, ^{1}H-^{13}C HMQC-TOCSY and 2D ^{1}H-^{15}N HSQC) were performed to assign ^{13}C and ^{15}N nuclei of the two dodecamers and to assist with the proton assignment. Figure 31 shows the ^{1}H-^{13}C HMQC-TOCSY spectrum for BS(AATT) in which is possible to identify carbon and proton spin systems for the sugar moieties. Proton chemical shifts are reported at the end of the Results session

of this thesis work.

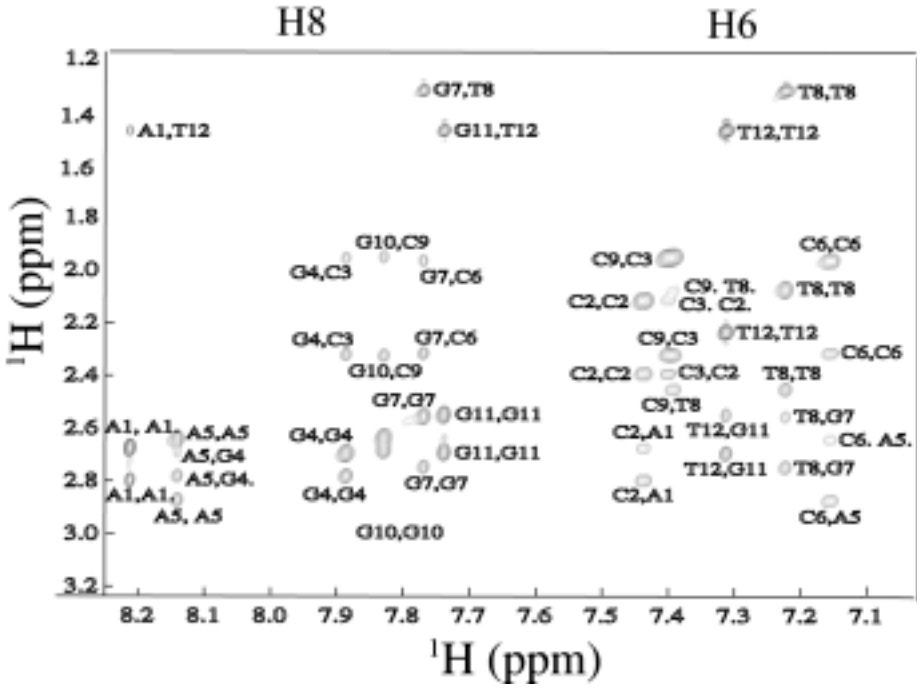

Figure 30. Region of NOESY spectrum acquired with 80 ms mixing time for BS(ACGT) duplex. The region corresponds to crosspeaks of H6/H8 protons with H2', H2'', Me protons.

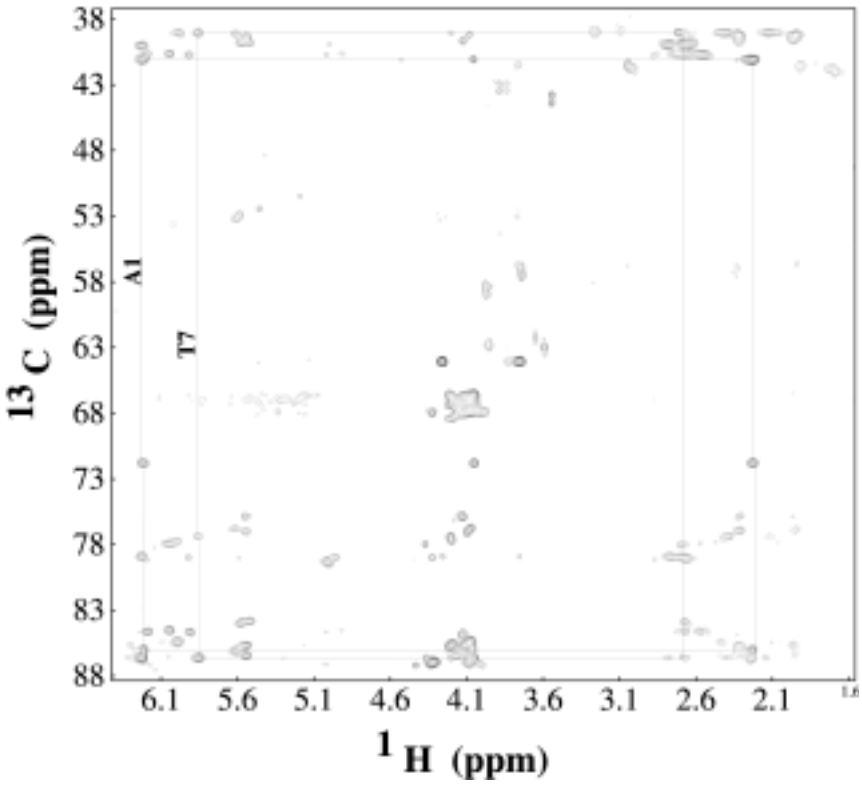

Figure 31. HMQC-TOCSY of BS(AATT) duplex. The sugar spin system

connectivity for two nucleotides is evidenced drawing a joining line.

Chemical shifts are sensitive and precise indicators of the surrounding magnetic environment. While clear correlations between chemical shifts of some types of nuclei and secondary structure were found for peptides and proteins (Wishart et al., 1994), these results are far to be achieved in the case of nucleic acids. The strong effects of local magnetic anisotropies and base ring-currents perturb profoundly DNA chemical shift in a way that is base-specific. Nevertheless, many nucleotide resonances are very sensitive to variations of defined torsional angles, but the conformation-dependent contributions to the shifts are difficult to extract. In fact, the correlated changes in torsional angles that occur in DNA conformation usually complicate and mask direct dependence on specific angles. This could be a discouraging observation if one deals with a specific structure, but could be an advantage when conducting a structural comparison between two or more related structures. If only few base pairs are different in the sequence of two duplexes, the difference of chemical shifts may be used to investigate the effect of these base substitutions on the rest of the molecule, that is the sequence-specific modifications in the structure. This kind of study is particularly feasible in E2 system because the two flanking tetranucleotide consensus half-sites of E2 binding sites present identical sequence (ACCG/CGGT), while variability is all localized in the central tetranucleotide spacer. If we compare the two dodecamers we examined, it results closer sequence similarity because they differ only in two positions: A6 and T7 in BS(AATT) *versus* C6 and G7 in BS(ACGT). Figure 32 shows the C.S. difference for H1', H4' and H6/H8 protons.

It appears clear that significant changes are substantially limited to the central tetranucleotide sequence. This result suggests that if differences are present between the two molecules, the source of structural diversity is situated in the conformational and mechanical properties of the central nucleotides. The comparison of the X-Ray high-

resolution structures solved for the two dodecamers (Table I) reveals that the consensus half-sites appear very similar between the two molecules, but important differences are found in the central zone and in the global shape. BS(ACGT) appears straight while BS(AATT) appears intrinsically curved. The bend in the latter molecule is obtained by negative rolling and minor groove compression in the spacer sequence. Chemical shifts are strictly local parameters and information about the overall conformation can be obtained from methods that are sensitive to the influence of the global shape (see below).

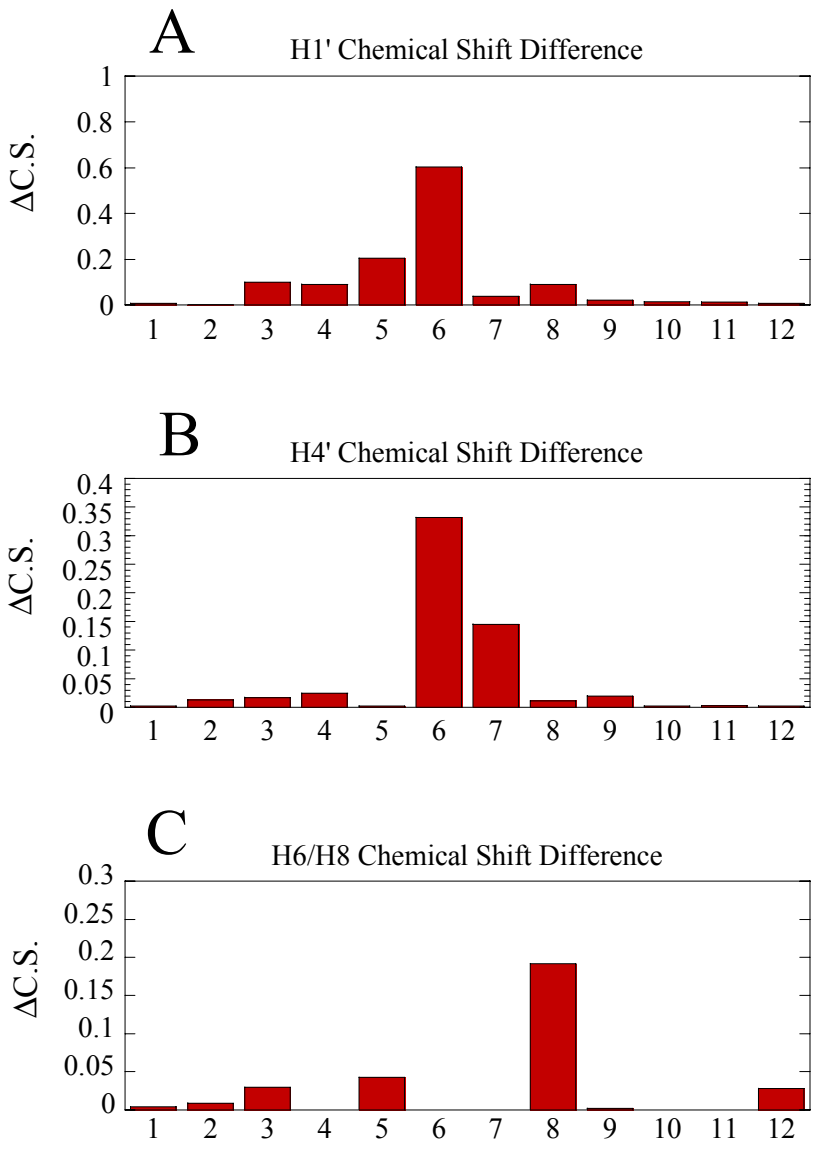

Figure 32. Root square chemical shift difference for H1' (a), H4' (b) and H6/H8 (c) protons of BS(ACGT) and BS(AATT). Note that H6/H8 protons for the two central nucleotides are missing and cannot

be compared because purine-pirimidine substitution occurs between the two duplexes.

In their crystallographic study, Hizver and coworkers observed that the conformational features of BS(AATT) are specific of the central CGAATTCG octameric sequence (Hizver et al., 2001). This sequence is found also in the Drew-Dickerson dodecamer (CGCGAATTCGCG), a DNA site recognized by the EcoRI endonuclease that was extensively studied by X-Ray crystallography and NMR. The rmsd in the octameric common sequence between BS(AATT) and the Dickerson dodecamer is 0.4-0.8 Å when only base atoms are superimposed and 0.9-1.2 Å when all atom coordinates are compared. In order to investigate if this similarity exists in solution too, we measured the chemical shifts difference between the two dodecameric sequences, using the assignment provided by Lane for the Dickerson dodecamer (Lane et al., 1991). Figure 33 shows the C.S. plots for the entire sequence. As it results evident from visual inspection of the plots, minimimal C.S. difference exists for the central octanucleotide sequence (rmsd of 0.07 ppm for H1', 0.01 ppm for H4', 0.04 ppm for H6/H8 protons). Deviation is seen for the sugar moieties of the first two and the last two residues that are the nucleotides not shared by the two sequences. The impressive closeness of chemical shifts indicates that BS(AATT) adopts a very similar conformation in the central zone with respect to the Dickerson dodecamer. The result corroborates the hypothesis that structural properties are strictly sequence-specific. The BS(AATT) crystallographic structure reveals a static curvature distributed along the entire sequence and achieved by negative rolling at the central tetranucleotide sequence together with positive rolling at the flanking tracts. A NMR study of local parameters, such as sugar pucker and base roll, twist, tilt in the solution conformation of the two dodecameric E2-BSs can provide detailed information on the sequence-dependent conformational properties of these two oligomers.

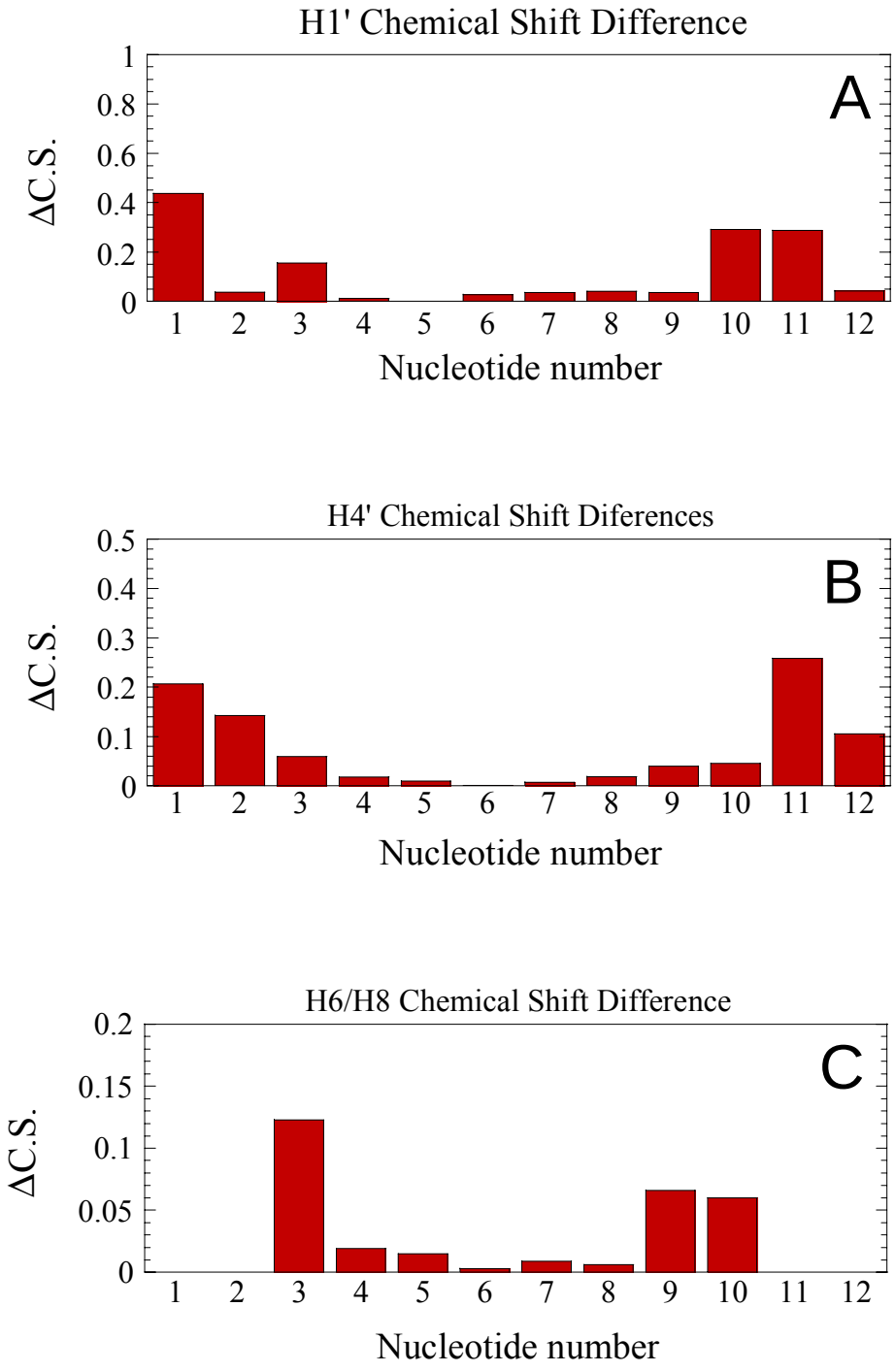

Figure 33. Root square chemical shift difference for H1' (a), H4' (b) and H6/H8 (c) protons of the Dickerson-Drew dodecamer and BS(AATT). Note that H6/H8 protons for the first two and for the last two nucleotides are missing and cannot be compared because base substitution occurs between the two duplexes.

E2Binding Sites diffusion studies

The global conformation of DNA molecules impacts the molecular motional properties and determines the apparent hydrodynamic parameters. The translational self-diffusion is related to the translational frictional coefficient by the Einstein equation:

$$D_t = kT / f_t$$

The frictional interaction of a molecule with the water medium is dependent on temperature and viscosity and on the overall and local shape of the hydrated molecule. Short DNA fragments can be modelled as prolate ellipsoids or, even better, as symmetric cylinders characterized by specific axial ratio (Fernandes et al., 2002). If curvature occurs the axial ratio decreases and this property affects self-diffusion rates. Also flexibility affects the diffusion behaviour. Therefore, sequence-specific features that affect the global shape or conformational characteristics of a duplex can be explored by diffusion experiments.

The original experiment conducted by Tanner to obtain diffusion constants of biomolecules by NMR (Tanner, 1970) was subjected to various variants. We employed the PFG-LED-STE (Gibbs and Johnson, 1991). The spin-echo scheme includes an additional delay before acquisition to remove the effects of the Eddy currents which would introduce serious artefacts in the spectrum. The magnetization is stored along the z-axis during the delay between the defocusing and refocusing pulsed field gradients. This strategy minimizes T_2 relaxation loss which is very effective in biomolecules of large molecular size.

It was shown that diffusion rates of DNA molecules can be measured with high precision and reproducibility by NMR methods under a wide range of experimental conditions (Lapham et al., 1997). The diffusion coefficients of double-stranded DNA

oligomers are demonstrated to be dependent on sequence length, on duplex concentration, on temperature and viscosity (Lapham et al., 1997), and on specific structural properties such as curvature and flexibility (Jerkovich and Bolton, 2000). Our objective was the structural comparison of the two dodecameric E2 binding sites BS(ACGT) and BS(AATT) in the same experimental context to assess if sequence-dependent properties can differentially affect the diffusion rates. To this end, we don't need to normalize our measurements for the influence of external factors (concentration, temperature, viscosity) if we perform experiments using identical conditions.

The diffusion constants measured for both duplexes at three different temperature points and in absence and presence of magnesium ion are shown in figure 34. We chose these conditions because temperature and magnesium are reported to profoundly affect the curvature of short DNA molecules (Jerkovich and Bolton, 2000).

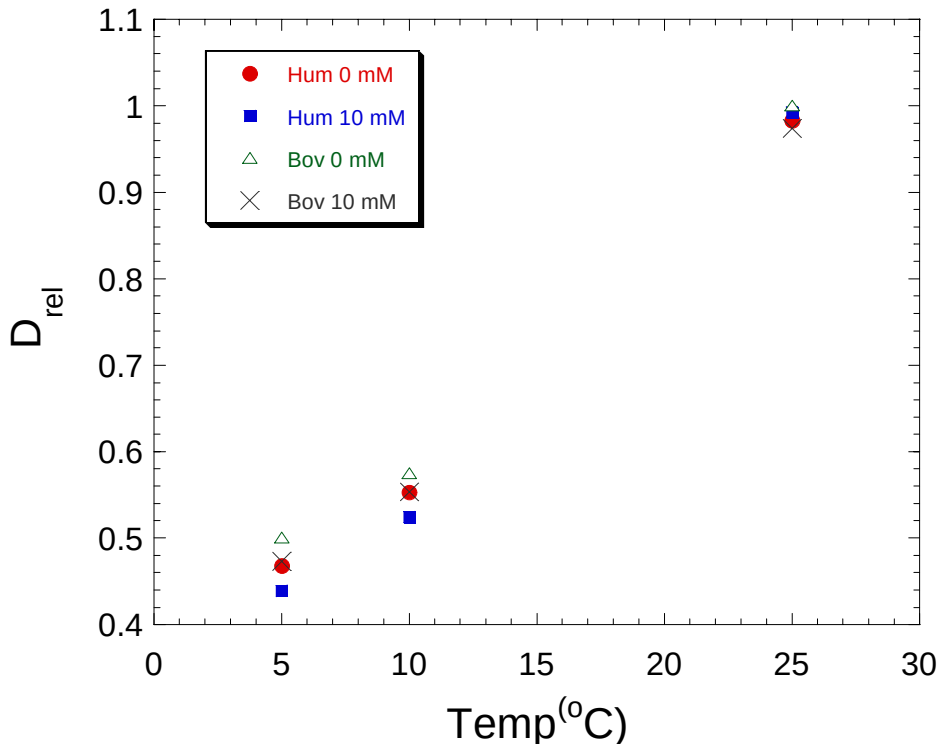

Figure 34. Diffusion constants of BS(ACGT) and BS(AATT) as a function of temperature. The diffusion coefficient is expressed relative to diffusion constant of BS(ACGT) at 25 °C when measured in absence of magnesium

The first observation regards the coincidence, within the experimental error, of diffusion rates at 25 °C, also when magnesium is added to the solution. Because self-diffusion are substantially dependent on the hydrodynamic parameters, that is on the global shape of the hydrated molecule, this result could surprise in light of the structural and biochemical data reported for these two oligomers. Experimental measures of E2 binding sites curvature were carried out using different methodologies such as cyclization kinetics methods (Zhang et al., 2004), circular permutation assays (Thain et al., 1997), electrophoretic phasing assays (Zimmerman et al., 2003). All these studies converged to the conclusion that sequences incorporating E2(AATT) possess a somewhat greater degree of bending relative to sequences incorporating BS(ACGT). This result was corroborated by the analysis of the crystal structures of these two dodecameric E2 binding sites (Table I). In many studies, these conformational differences were also correlated to the differential affinity of HPV E2 proteins toward their binding sites. HPV-16 E2 protein displays an affinity for BS(AATT) that is 300-fold higher than affinity for BS(ACGT). It was proposed that the structural basis of this preferential binding is the predisposition of BS(AATT) to adopt a conformation much closer to the conformation of the protein-bound state, because this requires smaller deformations to achieve optimal interaction with the recognition helices. A significant difference in the curvature should affect the global molecular shape and hence should be reflected in appreciable diversity in the diffusion properties. The discrepancy between our measures and the prediction based on previous studies may be explained taking into account the different experimental conditions used in the other experiments. With this regard, we think that diffusion coefficient measurements by NMR methods constitute a confident approach because don't involve interactions between DNA and other macromolecules or chemical agents or polymeric matrices.

Secondly, as we expect, diffusion constants of both oligomers are very sensitive to

temperature changes. According to the Einstein equation, the frictional properties depend directly on temperature and indirectly by means of temperature-dependent variations in the solution viscosity. We again underlie the fact that these physical changes of the system are the same for all molecules that are investigated using the same conditions. Therefore, the temperature dependence should be the same unless structural alterations occur that are inherent of the specific molecule. Jerkovich and Bolton (2001) show that curvature-dependent diffusion of short A-tracts is highly dependent on temperature. In their article they suggest that temperature represents a physical determinant that amplifies intrinsic conformational propensities of the molecule. The authors conclude that A-tracts don't have intrinsic temperature-independent curvature but display a tendency to curve in function of temperature as well as sequence.

The differential influence of temperature on E2-BSs diffusion rates is evident in figure 34. While at 25 °C the two 12-mers don't present appreciable difference in diffusion coefficients, lowering the temperature induces a diversity in their behaviour. BS(ACGT) shows higher diffusion constants at 5 and 10 °C with respect to BS(AATT). At 25 °C the difference in the diffusion coefficient between BS(ACGT) and BS(AATT) is about 2%, within the experimental error. At 10 °C the difference becomes circa 4% and at 5° it becomes 7%. The temperature could induce small local or global conformational modifications revealed by changes in the diffusion properties between the two duplexes.

In different crystallographic studies of curved DNAs magnesium ions are often found to be next to the regions responsible for conformational deviations. Magnesium is suggested to have an important role in facilitating the global axis to bend, by interacting with specific sites of the duplex or acting to reduce the electrostatic repulsion between the negatively charged backbone phosphate groups. When we added magnesium to the solution the same duplex exhibits slightly different diffusion rate relative to magnesium-

free sample at the temperatures od 5 and 10°C (figure 34). DNA oligonucleotide diffusion rates are only minimally sensitive to the ionic strength and the consequent changes of viscosity (Lapham et al., 1997). Therefore, magnesium can induce conformational modifications in the two E2 binding sites. It's notable that crystallographic studies of DNA fragments are often conducted with high magnesium concentration to facilitate the crystallization process. In the case of BS(ACGT) crystal were obtained in the presence of 10 mM Mg and in the case of BS(AATT) 76 mM Mg was used.

As found for A-tracts diffusion rates, the temperature seems to influence the intrinsic diffusion properties of E2-BSs. Temperature can enhance the flexibility of the sugar-phosphate backbone, alterating the global average conformation of the molecule. Curvature and flexibility effects are difficult to distinguish by this kind of approach. In the previous studies, BS(ACGT) is shown to possess intrinsic curvature while BS(ACGT) doesn't display overall bending. In addition, th CG doublet located at the center of the BS(ACGT) sequence is known to exhibit larger torsional and bending flexibility with respect to the AT step of BS(AATT). Further studies aiming at a characterization of the fine local details need to be conducted for these two double strands.

Table II.[a] Experimental restraints and structural statistics for the 20 lowest-energy structures for free E2C

number of experimental restraints	2904
distance restraints from NOEs	2398
hydrogen bond distance restraints	76
backbone dihedral angle restraints	244
HN-Hα scalar J coupling constants	64
H-N Residual Dipolar Coupling constants	56
intermonomer distance restraints	66
average number of restraints per residue	17.9
XPLOR energies (kcal/mol)[b]	
E_{total}	275 ±9
E_{bond}	4.2 ±0.5
E_{angle}	156 ±4
$E_{improper}$	28 ±2
E_{vdw}	8 ±3
E_{NOE}	4 ±1
E_{cdih}	0.9 ±0.4
E_{coup}	6 ±3
E_{sani}	7 ±13
E_{rama}	60 ±1
E_{NCS}	0.7 ±0.1
rms deviations from experimental restraints	
average distance restraints violation (Å)	0.006 ±0.001
average dihedral angle restraints violation (deg)	0.25 ±0.06
average scalar J coupling constant violation (Hz)	0.11 ±0.03
average RDC constant violation (Hz)	0.18 ±0.03
rms deviatios from idealized covalent geometry	
bond	0.0013 ±0.0001
angle	0.46 ±0.01
improper	0.36 ±0.01
Ramachandran analysis	
residues in favored regions (%)	83±3
residues in additional allowed regions (%)	16±4
residues in generously allowed regions (%)	1.0±0.1
residues in disallowed regions (%)	0.0±0.1

Coordinates precision [c]

	backbone (Å)	all heavy atoms (Å)
monomer (residues 284-319,329-362)	0.54 ±0.06	0.97 ±0.08
dimer (residues 284-319,329-362)	0.67 ±0.09	1.07 ±0.09

[a] numbers refer to data per dimer

[b] The force constants for energy potential were set to 50 Kcal mole^{-1} Å^{-2} for E_{NOE}, 200 Kcal mole^{-1} rad^{-2} for E_{CDIH}, 5 Kcal mole^{-1}s^{2} for E_{coup}, 4 Kcal mole^{-1}s^{2} for E_{sani}, , 0.02 Kcal mole^{-1}rad^{-2} for E_{rama}, 2 Kcal mole^{-1} Å^{-2} for E_{NCS}

[c] relative to the average of the 20 structures

BS(AATT) and BS(ACGT) chemical shift assignments

Site ACCGAATTCGGT: proton chemical shifts												
BASE	***H2***	***H5***	***H6***	***H8***	***Me***	***H1'***	***H2'***	***H2''***	***H3'***	***H4'***	***H5'***	***H5''***
A1	8,013			8,215		6,208	2,664	2,789	4,829	4,256	3,742	3,742
C2		5.403	7.432			5,843	2,091	2,374	4,790	4,187	4,068	4,139
C3		5.549	7.373			5,445	1,895	2,255	4,786	4,061	4,056	-
G4				7.849		5,429	2,653	2,74	4,980	4,308	3,969	4,065
A5	7,219			8,105		5,980	2,687	2,910	5,050	4,440	4,150	4,195
A6	7,595			8,093		6,134	2,549	2,909	4,989	4,449	4,238	4,238
T7		-	7.091		1.523	5,875	1,962	2,549	4,797	4,183	4,14	4,325
T8		-	7.354		1.511	6,07	2,131	2,513	4,833	4,211	4,086	4,086
C9		5.586	7.394			5,624	1,948	2,332	4,838	4,098	4,156	4,156
G10				7.830		5,559	2,649	2,649	4,966	4,315	3,999	4,089
G11				7.757		6,037	2,557		4,929	4,378	4,164	4,165
T12		-	7.334		1.500	6,202	2,228	2,228	4,532	4,045	4,050	4,217

Site ACCGAATTCGGT: exchangeable protons chemical shift			
BASE	***NH2***		***NH***
A1			
C2	8,11	6,62	
C3	8,51	6,73	
G4			12,648
A5			
A6			
T7			13,603
T8			13,748
C9	8,42	6,77	
G10			12,988
G11			12,840
T12			

Site ACCGACGTCGGT: proton chemical shifts												
BASE	***H2***	***H5***	***H6***	***H8***	***Me***	***H1'***	***H2'***	***H2''***	***H3'***	***H4'***	***H5'***	**H5''**
A1	8,002			8,219		6,2	2,675	2,797	4,825	4,258	3,747	3,748
C2		5.382	7.441			5,845	2,116	2,391	4,792	4,2	3,26	
C3		5.558	7.403			5,545	1,954	2,322	4,812	4,078	-	
G4				7.893		5,519	2,698	2,782	5,008	4,333	3,991	4,086
A5	7,778			8,148		6,184	2,643	2,875	5,012	4,442	4,174	4,208
C6		5.183	7.165			5,531	1,965	2,316	4,775	4,117	-	
G7				7.776		5,913	2,555	2,747	4,892	4,328	4,057	4,117
T8		-	7.23		1.319	5,98	2,075	2,453	4,833	4,2	4,102	4,102
C9		5.579	7.396			5,603	1,951	2,321	4,81	4,078	-	
G10				7.838		5,574	2,637	2,684	4,957	4,313	3,992	4,075
G11				7.745		6,025	2,55	2,695	4,92	4,375	4,163	4,163
T12		-	7.32		1.472	6,194	2,235	2,235	4,53	4,043	4,219	4,219

Site ACCGACGTCGGT: exchangeable protons chemical shift			
BASE	NH2		NH
A1			
C2	8,11	6,63	
C3	8,52	6,75	
G4			12,810
A5			
C6	8,10	6,44	
G7			12,71
T8			13,69
C9	8,51	6,81	
G10			13,07
G11			12,92
T12			

Site ACCGAATTCGGT: heteronuclei chamical shift										
BASE	***C2***	***C5***	***C6***	***C8***	***C1'***	***C2'***	***C3'***	***C4'***	***C5'***	***N1/N3***
A1	153,680			141,201	86,070	40,010	78,926	89,135	64,068	-
C2		98.45	141.67		86,585	38,960	-	84,887	-	97,56
C3		98.69	142.06		86,388	39,160	-	85,532	-	97,35
G4				-	83,730	-	78,445	86,959	-	146,8
A5	152,827			139,960	84,187	-	79,300	87,140	-	-
A6	153,370			-	84,320	41,269	79,477	86,830	-	-
T7		-	137.11		84,977	-	-	85,658	-	158,712
T8		-	138.34		85,155	38,880	-		-	159,043
C9		98.57	142.32		85,981	-	76,747	85,202	-	96,89
G10				137.16	83,869	39,830	78,910		-	146,477
G11				136.81	84,510	40,673	78,030	86,608	-	147,072
T12		-	138.81		85,959	41,079	71,745	86,613	67,700	-

Site ACCGACGTCGGT: heteronuclei chemical shifts										
BASE	***C2***	***C5***	***C6***	***C8***	***C1'***	***C2'***	***C3'***	***C4'***	***C5'***	***N1/N3***
A1	153,86			141,282	86,607	39,970	78,87	89,117	64,059	-
C2		98.4	141.75		86,590	39,030	77,40	85,66	-	97,56
C3		-	142.56		86,404	39,220	76,93		-	97,49
G4				137.29	83,811	-	-	-	-	-
A5	154,11			139,974	84,603	40,723	79,33	87,165	-	-
C6		98.05	140.87		85,681	39,220	75,81	84,755	-	97,15
G7				137.14	84,666	40,730	-	86,941	-	-
T8		-	137.53		85,362	39,060	77,83		-	-
C9		98.72	142.29		86,095	-	76,93	85,468	---	97,35
G10				137.38	83,922	39,950	78,96		-	-
G11				137.25	84,510	39,060	78,08	86,941	-	-
T12		-	138.46		86,050	41,100	71,72	86,62	-	-

CONCLUSIONS

We have studied the principles behind the structural plasticity of the high-risk HPV16 E2 DNA binding domain in solution. Structure determination, backbone dynamics through ^{15}N relaxation studies, and quantitative amide exchange rates were combined in order to obtain information about the behavior of the protein in solution. It was already clear from the known structures of E2C domains that the overall conformation of the monomeric unit is highly conserved in all viral strain proteins studied so far. In fact, the monomers do not differ significantly from one another in secondary and tertiary structure. The quaternary structure, on the contrary, is the main source of differentiation between the dimeric E2C domains (Liang et al., 1996; Hegde, 2002). When one of the monomers is superimposed, the orientations of the second monomers are such that the four structures can be divided into two families depending on the relative orientation of the two recognition helices (figure 5). This diversity was proposed to be significant in defining the bending imposed to the DNA to form the complex.

Protein-DNA complexes of BPV-1 and HPV-18 E2C show very similar bending angles of bound DNA. Differences could be appreciated at the central minor groove, which is narrower in the HPV-18 complex. On the other hand, there are no structures available for the complexes formed by HPV-16 and HPV-31 E2C with their respective DNA targets. The different orientation of the two recognition helices could have a significant impact on the deformation of DNA necessary to make the symmetrical contact with the recognition helices . Structural data on the complex with DNA will reveal if HPV-16 E2C requires to adapt itself to bind the DNA and induces a bending similar to that observed for HPV-18 E2C DNA complex, or it is the DNA that adopts a different conformation, as spectroscopic studies suggested (Ferreiro et al., 2000).

In order to obtain this information in solution, one has to rely on the use of NMR derived constraints to accurately measure the quaternary structure of the dimer, both in the free and in the complexed form. The use of only NOEs to define the relative orientation of the two monomers has a severe drawback when we look at regions, such as the recognition helices, that are distant from the dimeric interface. Little deviations at the interface cause larger deviations at the position of the two helices. Residual dipolar couplings are better constraints as they carry the same informational content for all regions of the protein, both regarding the conformation at the monomeric and dimeric levels. The structure described in the present paper is the first of this family in which RDC were used in the calculation. ^{15}N relaxation measurements have provided a correlation time of 10.8 ns, with a global isotropic tumbling. Both observations are in agreement with the size and shape of the molecule.

The most relevant result of our structure determination is that the loop connecting β_2 and β_3 is found to be disordered and highly solvent accessible in our solution DNA-free structure as well as in the complex. The first observation of this dynamic property was made for HPV-31 E2C, and it was speculated, by comparison with the DNA complex of BPV-1 E2C, that this region could adopt a more defined conformation upon binding to the DNA. A low electron density for this loop was observed in the complex of HPV-18 E2C and its cognate DNA (Kim et al., 2000). However, if the conformation of bound DNA is different in the HPV-16 E2C complex, as anticipated by the different orientation of the recognition helices and other studies, one cannot predict what would be the dynamic behavior of the β_2-β_3 loop in the complex. The amino acid composition of this loop is highly variable among the different viral strains. If we compare the length and charges of the residues that form this loop in the four strains for which we have structural information, we observed that HPV-16 E2C presents the highest number of positively charged residues

(4 positively charged and none negatively charged out of 8). A similar charge pattern can be found for HPV-31 E2C (3 positively charged, 1 negatively charged out of 8). Significantly less charged residues are found for HPV-18 E2C (1 positively charged and 1 negatively charged out of 7) and BPV-1 E2C (1 positively charged and 2 negatively charged out of 10). When comparing the electrostatic surfaces that HPV-16 and BPV-1 E2C present for their interaction with DNA, it can be concluded that there are more positively charged residues close to the recognition helices in BPV-1 E2C. The different distribution of charges between HPV-16 and BPV-1 E2C (more uniformly distributed along the entire surface in the first case and more concentrated near the recognition helices in the second) could be reflected in the dissimilar ability of the two proteins in assisting the necessary bending of the DNA during complex formation (Kim et al., 2000; Hegde, 2002). This fact, in turn, can be the basis of the difference in affinities for DNA binding sites with different nucleotide composition in the spacer region that HPV-16 and BPV-1 E2C show .

The quaternary structure differences are directly related to a different hydrogen bonding network connecting the two β_4 strands (Liang et al., 1996; Hegde, 2002). It is interesting to observe that the β_4 strands of HPV-16 E2C show fluctuations in the μs to ms time scale, as detected with ^{15}N relaxation measurements. These movements can be part of a breathing-like movement of the whole β-barrel structure, causing the high solvent exposition observed.

The recognition helix is very well conserved throughout the different Papillomavirus strains not only on its DNA binding side but also on the side that faces the barrel. The recognition helix exchanges its NH backbones very fast but it appears to be in a well defined conformation in the DNA-free conformation. In contrast, dramatic quenching of flexibility characterizes the entire protein conformation in the complex. Binding between the recognition helices and the DNA half-sites may require a certain flexibility of the

binding motifs to facilitate the induced-fit recognition process. In addition, the generalized mobility of this protein could have important functions in the transmission of structural information from the transactivation domain (the N-terminal module of the full-length protein) which in turn is related to the promotion and modulation of E2 acitivity.

The structural plasticity we now describe may not be reflected in a substantial conformational change upon DNA binding but could be essential for the kinetic binding mechanism, where there are several intermediate conformations through which the complex must evolve before the final complex is consolidated.

The E2 binding sites examined by us seem to adopt distinct conformational features depending on the experimental conditions. Specifically, temperature and magnesium ion have important effects on the global average conformation sampled by diffusion rates measurements, affecting the local bending propensity and flexibility. Site-specific structural parameters have to be inspected by high-resolution NMR methods to investigate the sequence-dependent contributions to conformational and dynamics of these two DNA molecules.

Adaptability and flexibility are recognized to be essential determinants in the DNA-protein recognition mechanism and have profound effects on the energetics pathways of macromolecular interactions. A fine balance of the enthalpic gain due to favourable direct contacts between the two molecules and the entropic costs because of the constraints imposed upon complex formation is exserted in these efficient biomolecular machines.

BIBLIOGRAPHY

Al-Hashimi, H.M., Majumdar, A., Gorin, A., Kettani, A., Skripkin, E., Patel, D.J.: "Field- and phage-induced dipolar couplings in a homodimeric DNA quadruplex: relative orientation of G.(C-A) triad and G-tetrad motifs and direct determination of C2 symmetry axis orientation". *J Am Chem Soc* (2001), **123**:633-640.

Annila, A., Aitio, H., Thulin, E. and Drakenberg,T.: "Recognition of protein folds via dipolar couplings". *J. Biomol. NMR* (1999), **14**:223-230.

Antson, A.A., Burns, J.E., Moroz, O.V., Scott, D.J., Sanders, C.M., Bronstein, I.B., Dodson, G.G., Wilson, K.S., Maitland, N.J.: "Structure of the intact transactivation domain of the human papillomavirus E2 protein". *Nature* (2000), **403**:805-809.

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), **25**:7751-7759.
Barbato, G., Ikura, M., Kay, L.E., Pastor, R.W., Bax, A.: "Backbone dynamics of calmodulin studied by 15N relaxation using inverse detected two-dimensional NMR spectroscopy: the central helix is flexible". *Biochemistry* (1992), **31**:5269-5278.

Barbic, A., Zimmer, D.P., Crothers, D.M.: "Structural origins of adenine-tract bending". *Proc Natl Acad Sci U S A* (2003), **100**:2369-2373.

Bax, A., Kontaxis, G. and Tjandra, N.: "Dipolar Couplings in macromolecular structure determination". *Methods Enzymol* (2001), **339**:127-174.

Bax, A.a.G., S.: "Methodological advances in protein NMR". *Acc.Chem. Res.* (1993), **26**:131-138.
Bax, A.a.L., L.: "Measurement of 1H-1H coupling constants in DNA fragments by 2D NMR". *J. Magn. Reson.* (1988), **79**:429-438.

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.

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.

Bewley, C.A.a.C., G.M.: "Determination of the relative orientation of the two halves of the domain-swapped dimer of Cyanovirin-N in solution using residual dipolar couplings and rigid body minimization". *J. Am. chem. Soc.* (2000), **122**:6009-6016.

Bochkarev, A., Barwell, J.A., Pfuetzner, R.A., Furey, W., Jr., Edwards, A.M., Frappier, L.: "Crystal structure of the DNA-binding domain of the Epstein-Barr virus origin-binding protein EBNA 1". *Cell* (1995), **83**:39-46.

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.

Bussiere, D.E., Kong, X., Egan, D.A., Walter, K., Holzman, T.F., Lindh, F., Robins, T., Giranda, V.L.: "Structure of the E2 DNA-binding domain from human papillomavirus serotype 31 at 2.4 A". *Acta Crystallogr D Biol Crystallogr* (1998), **54**:1367-1376.

Byun, K.S., Beveridge, D.L.: "Molecular dynamics simulations of papilloma virus E2 DNA sequences: dynamical models for oligonucleotide structures in solution". *Biopolymers* (2004), **73**:369-379.

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 15N-1H NMR spectroscopy". *Biochemistry* (1990), **29**:7387-7401.

Clore, G.M., 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.

Clore, G.M.S., M.R., Bewley, C.A., Cai, M. and Kuszewski, J: "Impact of Residual Dipolar Couplings on the accuracy of NMR structures determined from a minimal number of NOE restraints". *J. Am. Chem. Soc.* (1999), **122**:6513-6514.

Clore, M.G.a.G., D.S.: "R-Factor, free R and complete cross-validation for Dipolar Coupling refinement in NMR structures". *J Am Chem Soc* (1999), **121**:9008-9012.

Crothers, D.M.: "DNA curvature and deformation in protein-DNA complexes: a step in the right direction". *Proc Natl Acad Sci U S A* (1998), **95**:15163-15165.

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.

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., Gaston, K.: "Human papillomaviruses and their role in cervical cancer". *Cell Mol Life Sci* (2001), **58**:1923-1942.

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.

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.

Dotsch, V., Wagner, G.: "New approaches to structure determination by NMR spectroscopy". *Curr Opin Struct Biol* (1998), **8**:619-623.

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.
Fagan, P.A., Spielmann, H.P., Sigurdsson, S., Rink, S.M., Hopkins, P.B., Wemmer, D.E.: "An NMR study of [d(CGCGAATTCGCG)]2 containing an interstrand cross-link derived from a distamycin-pyrrole conjugate". *Nucleic Acids Res* (1996), **24**:1566-1573.

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 15N NMR relaxation". *Biochemistry* (1994), **33**:5984-6003.

Feigon, J., Sklenar, V., Wang, E., Gilbert, D.E., Macaya, R.F., Schultze, P.: "1H NMR spectroscopy of DNA". *Methods Enzymol* (1992), **211**:235-253.

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

Fernandes, M.X., Ortega, A., Lopez Martinez, M.C., Garcia de la Torre, J.: "Calculation of hydrodynamic properties of small nucleic acids from their atomic structure". *Nucleic Acids Res* (2002), **30**:1782-1788.

Ferreiro, D.U., de Prat-Gay, G.: "A protein-DNA binding mechanism proceeds through multi-state or two-state parallel pathways". *J Mol Biol* (2003), **331**:89-99.

Ferreiro, D.U., Lima, L.M., Nadra, A.D., Alonso, L.G., Goldbaum, F.A., de Prat-Gay, G.: "Distinctive cognate sequence discrimination, bound DNA conformation, and binding modes in the E2 C-terminal domains from prototype human and bovine papillomaviruses". *Biochemistry* (2000), **39**:14692-14701.

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.

Fowler, C.A., Tian, F., Al-Hashimi, H.M. and Prstegard, J.H.: "Rapid determination of protein folds using residual dipolar coupling". *J. Mol. biol.* (2000), **304**:447-460.

Gibbs, S.J.a.J., C.S.: "A PFG-NMR expriment for accurate diffusion and flow studies in the presence of Eddy currents". *J. Magn. Reson.* (1991), **93**:395-402.

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 13C/15N-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 H2O 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.

Hegde, R.S.: "The papillomavirus E2 proteins: structure, function, and biology". *Annu Rev Biophys Biomol Struct* (2002), **31**:343-360.

Hegde, R.S., Androphy, E.J.: "Crystal structure of the E2 DNA-binding domain from human papillomavirus type 16: implications for its DNA binding-site selection mechanism". *J Mol Biol* (1998), **284**:1479-1489.

Hegde, R.S., Grossman, S.R., Laimins, L.A., Sigler, P.B.: "Crystal structure at 1.7 A of the bovine papillomavirus-1 E2 DNA-binding domain bound to its DNA target". *Nature* (1992), **359**:505-512.

Hegde, R.S., Wang, A.F., Kim, S.S., Schapira, M.: "Subunit rearrangement accompanies sequence-specific DNA binding by the bovine papillomavirus-1 E2 protein". *J Mol Biol* (1998), **276**:797-808.

Hines, C.S., Meghoo, C., Shetty, S., Biburger, M., Brenowitz, M., Hegde, R.S.: "DNA structure and flexibility in the sequence-specific binding of papillomavirus E2 proteins". *J Mol Biol* (1998), **276**:809-818.

Hizver, J., Rozenberg, H., Frolow, F., Rabinovich, D., Shakked, Z.: "DNA bending by an adenine--thymine tract and its role in gene regulation". *Proc Natl Acad Sci U S A* (2001), **98**:8490-8495.

Hosur, R.V.a.G., G.: "Application of two-dimansional NMR spectroscopy in the determination of solution conformation of nucleic acids". *Magn. reson. in Chem.* (1988), **26**:927-944.

Ikura, M., Kay, L.E., Bax, A.: "A novel approach for sequential assignment of 1H, 13C, and 15N spectra of proteins: heteronuclear triple-resonance three-dimensional NMR spectroscopy. Application to calmodulin". *Biochemistry* (1990), **29**:4659-4667.

Ikura, M., Kay, L.E., Bax, A.: "Improved three-dimensional 1H-13C-1H correlation spectroscopy of a 13C-labeled protein using constant-time evolution". *J Biomol NMR* (1991), **1**:299-304.

Ishii, Y., Markus, M.A., Tycko, R.: "Controlling residual dipolar couplings in high-resolution NMR of proteins by strain induced alignment in a gel". *J Biomol NMR* (2001), **21**:141-151.

Ishima, R., Torchia, D.A.: "Protein dynamics from NMR". *Nat Struct Biol* (2000), **7**:740-743.

Jerkovic, B., Bolton, P.H.: "Magnesium increases the curvature of duplex DNA that contains dA tracts". *Biochemistry* (2001), **40**:9406-9411.

Johnson, B.A.a.B., R.A.: "NMRView: a computer program for the visualization and analysis of NMR data". *J. Biomol. NMR* (1994), **4**:603-614.

Kay, E.L., Xu, G.Y., Singer, A.U., Muhandiram, D.R. and Formankay, J.D.: "A gradient-enhanced HCCH-TOSCY experiment for recording side chain 1H and 13C correlation in H2O samples of proteins". *J. Magn. Reson. B* (1993), **101**:333-337.

Kay, L., Keifer, P. and Saarinen, T.: "Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity". *J. Am. Chem. Soc.* (1992), **114**:10663-10665.

Kay, L.E.: "Field gradient techniques in NMR spectroscopy". *Curr Opin Struct Biol* (1995), **5**:674-681.

Kay, L.E., Gardner, K.H.: "Solution NMR spectroscopy beyond 25 kDa". *Curr Opin Struct Biol* (1997), **7**:722-731.

Kay, L.E., Torchia, D.A., Bax, A.: "Backbone dynamics of proteins as studied by 15N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease". *Biochemistry* (1989), **28**:8972-8979.

Kim, K.S., Fuchs, J.A., Woodward, C.K.: "Hydrogen exchange identifies native-state motional domains important in protein folding". *Biochemistry* (1993), **32**:9600-9608.

Kim, S.S., Tam, J.K., Wang, A.F., Hegde, R.S.: "The structural basis of DNA target discrimination by papillomavirus E2 proteins". *J Biol Chem* (2000), **275**:31245-31254.

Kordel, J., Skelton, N.J., Akke, M., Palmer, A.G., 3rd, Chazin, W.J.: "Backbone dynamics of calcium-loaded calbindin D9k studied by two-dimensional proton-detected 15N NMR spectroscopy". *Biochemistry* (1992), **31**:4856-4866.

Lane, A.N., Jenkins, T.C., Brown, T., Neidle, S.: "Interaction of berenil with the EcoRI dodecamer d(CGCGAATTCGCG)$_2$ in solution studied by NMR". *Biochemistry* (1991), **30**:1372-1385.

Lapham, J., Rife, J.P., Moore, P.B. and Crothers, D.M: "Measurements of diffusion constants for nucleic acids by NMR". *J. Biomol. NMR* (1997), **10**:255-262.

Laskowski, R.A., Rullmannn, J.A., MacArthur, M.W., Kaptein, R., Thornton, J.M.: "AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR". *J Biomol NMR* (1996), **8**:477-486.

Lefebvre, A., Fermandjian, S., Hartmann, B.: "Sensitivity of NMR internucleotide distances to B-DNA conformation: underlying mechanics". *Nucleic Acids Res* (1997),

25:3855-3862.

Liang, H., Petros, A.M., Meadows, R.P., Yoon, H.S., Egan, D.A., Walter, K., Holzman, T.F., Robins, T., Fesik, S.W.: "Solution structure of the DNA-binding domain of a human papillomavirus E2 protein: evidence for flexible DNA-binding regions". *Biochemistry* (1996), **35**:2095-2103.

Lima, L.M., de Prat-Gay, G.: "Conformational changes and stabilization induced by ligand binding in the DNA-binding domain of the E2 protein from human papillomavirus". *J Biol Chem* (1997), **272**:19295-19303.

Lipari, G.a.S., A.: "Model-free approach to th interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity". *J. Am. Chem. Soc.* (1982), **104**:4546-4559.

MacDonald, D., Herbert, K., Zhang, X., Pologruto, T., Lu, P., Polgruto, T.: "Solution structure of an A-tract DNA bend". *J Mol Biol* (2001), **306**:1081-1098.

Mandel, A.M., Akke, M., Palmer, A.G., 3rd: "Backbone dynamics of Escherichia coli ribonuclease HI: correlations with structure and function in an active enzyme". *J Mol Biol* (1995), **246**:144-163.

Marathias, V.M., Jerkovic, B., Arthanari, H., Bolton, P.H.: "Flexibility and curvature of duplex DNA containing mismatched sites as a function of temperature". *Biochemistry* (2000), **39**:153-160.

Marion, D., Wuthrich, K.: "Application of phase sensitive two-dimensional correlated spectroscopy (COSY) for measurements of 1H-1H spin-spin coupling constants in proteins". *Biochem Biophys Res Commun* (1983), **113**:967-974.

Marion, D.a.B., A.: "P.COSY, a sensitive alternative for double-quantum-filtered COSY". *J. Magn. Reson.* (1988), **80**:528-533.

Mauffret, O., Tevanian, G., Fermandjian, S.: "Residual dipolar coupling constants and structure determination of large DNA duplexes". *J Biomol NMR* (2002), **24**:317-328.

Mok, Y.K., Alonso, L.G., Lima, L.M., Bycroft, M., de Prat-Gay, G.: "Folding of a dimeric beta-barrel: residual structure in the urea denatured state of the human papillomavirus E2 DNA binding domain". *Protein Sci* (2000), **9**:799-811.

Mok, Y.K., Bycroft, M., de Prat-Gay, G.: "The dimeric DNA binding domain of the human papillomavirus E2 protein folds through a monomeric intermediate which cannot be native-like". *Nat Struct Biol* (1996), **3**:711-717.

Mok, Y.K., de Prat Gay, G., Butler, P.J., Bycroft, M.: "Equilibrium dissociation and unfolding of the dimeric human papillomavirus strain-16 E2 DNA-binding domain". *Protein Sci* (1996), **5**:310-319.

Muhandiram, D.R.a.K., L.E.: "Gradient-enhanced triple-resonance three-dimensional NMR experiments with improved sensitivity". *J. Magn. Reson. B* (1994), **103**:203-216.

Neri, D., Szyperski, T., Otting, G., Senn, H., Wuthrich, K.: "Stereospecific nuclear magnetic resonance assignments of the methyl groups of valine and leucine in the DNA-binding domain of the 434 repressor by biosynthetically directed fractional 13C labeling". *Biochemistry* (1989), **28**:7510-7516.

Neuhaus, D., Wagner, G., Vasak, M., Kagi, J.H., Wuthrich, K.: "Systematic application of high-resolution, phase-sensitive two-dimensional 1H-NMR techniques for the identification of the amino-acid-proton spin systems in proteins. Rabbit metallothionein-2". *Eur J Biochem* (1985), **151**:257-273.
Nilges, M.: "A calculation strategy for the structure determination of symmetric dimers by 1H NMR". *Proteins* (1993), **17**:297-309.

O'Donoghue, S.I., King, G.F. and Nilges, M.: "Calculation of symmetric multimer structures from NMR data using a priori knowledge of the monomer structure, co-monomer restraints, and interface mapping: the case of leucine zippers". *J. Biomol. NMR* (1996), **8**:193-206.

Ojha, R.P., Dhingra, M.M., Sarma, M.H., Shibata, M., Farrar, M., Turner, C.J., Sarma, R.H.: "DNA bending and sequence-dependent backbone conformation NMR and computer experiments". *Eur J Biochem* (1999), **265**:35-53.

Orbons, L.P., van der Marel, G.A., van Boom, J.H., Altona, C.: "An NMR study of polymorphous behaviour of the mismatched DNA octamer d(m5C-G-m5C-G-A-G-m5C-G) in solution. The B-duplex and hairpin forms". *Eur J Biochem* (1987), **170**:225-239.

Orekhov, V.Y., Nolde, D.E., Golovanov, A.P., Korzhnev, D.M and Arseniev, A.S.: "Processing of heteronuclear NMR relaxation data with the new software DASHA". *Appl. Magn. Reson.* (1995), **9**:581.

Pabo, C.O., Nekludova, L.: "Geometric analysis and comparison of protein-DNA interfaces: why is there no simple code for recognition?" *J Mol Biol* (2000), **301**:597-624.

Palmer, A.G., 3rd: "Probing molecular motion by NMR". *Curr Opin Struct Biol* (1997), **7**:732-737.
Parella, T.: "Pulsed field gradient: a new tool for routine NMR". *Magn. Reson. Chem.* (1998), **36**:467-495.

Pawley, N.H., Wang, C., Koide, S., Nicholson, L.K.: "An improved method for distinguishing between anisotropic tumbling and chemical exchange in analysis of 15N relaxation parameters". *J Biomol NMR* (2001), **20**:149-165.

Pelton, J.G., Torchia, D.A., Meadow, N.D., Roseman, S.: "Tautomeric states of the active-site histidines of phosphorylated and unphosphorylated IIIGlc, a signal-transducing protein from Escherichia coli, using two-dimensional heteronuclear NMR techniques". *Protein Sci* (1993), **2**:543-558.

Piotto, M., Saudek, V., Sklenar, V.: "Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions". *J Biomol NMR* (1992), **2**:661-665.

Plateau, P.a.G., M: "Exchangeable proton NMR without base-line distortion, using new strong-pulse sequences". *J. Am. Chem. Soc.* (1982), **104**:7310-7311.

Prestegard, J.H.: "New techniques in structural NMR--anisotropic interactions". *Nat Struct Biol* (1998), **5 Suppl**:517-522.

Raschke, T.M.a.M., S.: "Mechanisms and uses of hydrogen exchange". *Curr Opin Struct Biol* (1998), **9**:80-86.

Roberts, C.G.K.: *NMR of macromolecules*. Oxford, England: Oxford University Press; 1993.

Rozenberg, H., Rabinovich, D., Frolow, F., Hegde, R.S., Shakked, Z.: "Structural code for DNA recognition revealed in crystal structures of papillomavirus E2-DNA targets". *Proc Natl Acad Sci U S A* (1998), **95**:15194-15199.
Sanders, C.M., Maitland, N.J.: "Kinetic and equilibrium binding studies of the human papillomavirus type-16 transcription regulatory protein E2 interacting with core enhancer elements". *Nucleic Acids Res* (1994), **22**:4890-4897.

Sass, H.J., Musco, G., Stahl, S.J., Wingfield, P.T., Grzesiek, S.: "Solution NMR of proteins within polyacrylamide gels: diffusional properties and residual alignment by mechanical stress or embedding of oriented purple membranes". *J Biomol NMR* (2000), **18**:303-309.

Saunders, M., Wishnia, A., and Kirkwood, J.G.: "The nuclear magnetic resonance spectrum of ribonuclease". *J Am Chem Soc* (1957), **79**:3289.

Schurr, J.M., Babcock, H.P., Fujimoto, B.S.: "A test of the model-free formulas. Effects of anisotropic rotational diffusion and dimerization". *J Magn Reson B* (1994), **105**:211-224.

Schwieters, C.D., Kuszewski, J.J., Tjandra, N., Marius Clore, G.: "The Xplor-NIH NMR molecular structure determination package". *J Magn Reson* (2003), **160**:65-73.

Sitkoff, D.a.C., D.A.: "Theories of chemical shif anisotropies in proteins and nucleic acids". *Prog. NMR Spectr* (1994), **32**:165-190.

Stone, M.J., Fairbrother, W.J., Palmer, A.G., 3rd, Reizer, J., Saier, M.H., Jr., Wright, P.E.: "Backbone dynamics of the Bacillus subtilis glucose permease IIA domain determined from 15N NMR relaxation measurements". *Biochemistry* (1992), **31**:4394-4406.

Suzuki, M., Yagi, N.: "Stereochemical basis of DNA bending by transcription factors". *Nucleic Acids Res* (1995), **23**:2083-2091.

Tanner, J.E.: "Use of the stimulated echo in NMR diffusion studies". *J. Chem. Phys.* (1970), **52**:2523-2526.

Thain, A., Webster, K., Emery, D., Clarke, A.R., Gaston, K.: "DNA binding and bending by the human papillomavirus type 16 E2 protein. Recognition of an extended binding site". *J Biol Chem* (1997), **272**:8236-8242.

Tian, F., Al-Hashimi, H.M., Craighead, J.L., Prestegard, J.H.: "Conformational analysis of

a flexible oligosaccharide using residual dipolar couplings". *J Am Chem Soc* (2001), **123**:485-492.

Tjandra, N., Feller, S.E., Pastor, R.W. and Bax, A.: "Rotational diffusion anisotropy of human ubiquitin from 15N NMR relaxation". *J. Am. Chem. Soc.* (1995), **117**:12562-12566.

Tjandra, N., Tata, S-i, Ono, A., Kainosho, M. and Bax, A.: "The NMR structure of a DNA dodecamer in an acqueous dilute liquid crystalline phase". *J.Am. Chem. Soc.* (2000), **122**:6190-6200.

Tolman, J.R., Flanagan, J.M., Kennedy, M.A., Prestegard, J.H.: "Nuclear magnetic dipole interactions in field-oriented proteins: information for structure determination in solution". *Proc Natl Acad Sci U S A* (1995), **92**:9279-9283.

Tonelli, M., James, T.L.: "Insights into the dynamic nature of DNA duplex structure via analysis of nuclear Overhauser effect intensities". *Biochemistry* (1998), **37**:11478-11487.

Ulyanov, N.B., Gorin, A.A., Zhurkin, V.B., Chen, B.C., Sarma, M.H., Sarma, R.H.: "Systematic study of nuclear Overhauser effects vis-a-vis local helical parameters, sugar puckers, and glycosidic torsions in B DNA: insensitivity of NOE to local transitions in B DNA oligonucleotides due to internal structural compensations". *Biochemistry* (1992), **31**:3918-3930.

Urbani, A., Bazzo, R., Nardi, M.C., Cicero, D.O., De Francesco, R., Steinkuhler, C., Barbato, G.: "The metal binding site of the hepatitis C virus NS3 protease. A spectroscopic investigation". *J Biol Chem* (1998), **273**:18760-18769.

Van Dijk, A.A., Scheek, R.M., Dijkstra, K., Wolters, G.K., Robillard, G.T.: "Characterization of the protonation and hydrogen bonding state of the histidine residues in IIAmtl, a domain of the phosphoenolpyruvate-dependent mannitol-specific transport protein". *Biochemistry* (1992), **31**:9063-9072.

Veeraraghavan, S., Mello, C.C., Androphy, E.J., Baleja, J.D.: "Structural correlates for enhanced stability in the E2 DNA-binding domain from bovine papillomavirus". *Biochemistry* (1999), **38**:16115-16124.
Veeraraghavan, S., Mello, C.C., Lee, K.M., Androphy, E.J., Baleja, J.D.: "1H, 15N, and 13C NMR resonance assignments for the DNA-binding domain of the BPV-1 E2 protein". *J Biomol NMR* (1998), **11**:457-458.

Vuister, G.W.a.B., A.: "Resolution enhancement and spectral editing of uniformly 13C-enriched proteins by homonuclear broadband 13C decoupling". *J. Magn. Reson.* (1992), **98**:428-435.

Vuister, G.W.a.B., A.: "Quantitative J correlation: a new approach for measuring homonuclear three-bond J (HNH.alpha) coupling constant in 15N-enriched proteins". *J.Am. chem. soc.* (1993), **115**:7772-7777.

Wagner, G.: "NMR relaxation and protein mobility". *Curr. Op. Struct. Biol.* (1993), **3**:748-754.

Wishart, D.S., Bigam, C.G., Yao, J., Abildgaard, F., Dyson, H.J., Oldfield, E., Markley, J.L., Sykes, B.D.: "1H, 13C and 15N chemical shift referencing in biomolecular NMR". *J Biomol NMR* (1995), **6**:135-140.

Wishart, D.S., Sykes, B.D.: "The 13C chemical-shift index: a simple method for the identification of protein secondary structure using 13C chemical-shift data". *J Biomol NMR* (1994), **4**:171-180.

Wishart, D.S., Sykes, B.D.: "Chemical shifts as a tool for structure determination". *Methods Enzymol* (1994), **239**:363-392.

Wu, Z., Delaglio, F., Tjandra, N., Zhurkin, V.B., Bax, A.: "Overall structure and sugar dynamics of a DNA dodecamer from homo- and heteronuclear dipolar couplings and 31P chemical shift anisotropy". *J Biomol NMR* (2003), **26**:297-315.

Wuthrich, K.: *NMR of protins and nucleic acids*: John Wiley & Sons; 1986.

Yee, A., Chang, X., Pineda-Lucena, A., Wu, B., Semesi, A., Le, B., Ramelot, T., Lee, G.M., Bhattacharyya, S., Gutierrez, P., et al.: "An NMR approach to structural proteomics". *Proc Natl Acad Sci U S A* (2002), **99**:1825-1830.

Zhang, Y., Xi, Z., Hegde, R.S., Shakked, Z., Crothers, D.M.: "Predicting indirect readout effects in protein-DNA interactions". *Proc Natl Acad Sci U S A* (2004), **101**:8337-8341.

Zhu, L., Reid, B.R. and Drobny, G.P.: "Errors in Measuring and interpreting values of coupling constants J from PE.COSY experiments". *J. Magn. Reson. Series A* (1995), **115**:206-212.

Zidek, L., Stefl, R., Sklenar, V.: "NMR methodology for the study of nucleic acids". *Curr Opin Struct Biol* (2001), **11**:275-281.

Zidek, L., Wu, H., Feigon, J., Sklenar, V.: "Measurement of small scalar and dipolar couplings in purine and pyrimidine bases". *J Biomol NMR* (2001), **21**:153-160.

Zimmerman, J.M., Maher, L.J., 3rd: "Solution measurement of DNA curvature in papillomavirus E2 binding sites". *Nucleic Acids Res* (2003), **31**:5134-5139.

Zuiderweg, E.R.: "Mapping protein-protein interactions in solution by NMR spectroscopy". *Biochemistry* (2002), **41**:1-7.

Zweckstetter, M.a.B., A: "Prediction of sterically induced alignment in a dilute liquid crystalline phase: aid to protein structure determination by NMR". *J. Am. Chem. Soc.* (2000), **122**:3791-3792.

AKNOWLEDGEMENTS

The production of the isotope-enriched protein was carried out in Buenos Aires, Instituto-Fundaciòn Leloir, Patricias Argentinas 434, Argentina

NMR experiments were collected on Avance400 and Avance 700 spectrometers located in Rome, Università di Tor Vergata, Italy.

I'm very grateful to Prof. Gonzalo de Prat-Gay of UBA (Universidad de Buenos Aires) for the period I spend in Buenos Aires working on this project and for his precious suggestions on the E2C system.

I'm very grateful to Prof. Maurizio Paci and his NMR group at "Università di Tor Vergata", for the collaboration in this project and his practical and scientific suggestions. In particular, I want to thank Prof. Daniel O. Cicero who introduced me to the NMR world and followed my training in this discipline, exciting my scientific interest.

I share the project on E2C system with Dr. Alejandro D. Nadra who is following a Ph.D. course at UBA. Without his help, I couldn't conduct this interesting work.