BIOCHEMISTRY

© Copyright 2006 by the American Chemical Society Volume 45, Number 21 May 30, 2006

Accelerated Publications

Structural and Thermodynamic Basis for the Enhanced Transcriptional Control by the Human Papillomavirus Strain-16 E2 Protein[†]

Daniel O. Cicero,*[,§] Alejandro D. Nadra,[‡] Tommaso Eliseo,[§] Mariano Dellarole,[‡] Maurizio Paci,[§] and Gonzalo de Prat-Gay*[,‡]

Instituto Leloir and CONICET, and Department of Chemical Sciences and Technologies, University of Rome 'Tor Vergata', via della Ricerca Scientifica, 00133 Rome, Italy Tor Vergata

Received January 19, 2006; Revised Manuscript Received March 17, 2006

ABSTRACT: Strain 16 of the human papillomavirus is responsible for the largest number of cases of cervical cancers linked to this virus, and the E2 protein is the transcriptional regulator of all viral genes. We present the first structure for the DNA binding domain of HPV16 E2 bound to DNA, and in particular, to a natural cognate sequence. The NMR structure of the protein backbone reveals that the overall conformation remains virtually unchanged, and chemical shift analysis of the protein bound to a shorter DNA duplex uncovered a contact out of the minimal E2 DNA binding site, made by lysine 349. This contact was confirmed by titration calorimetry and mutagenesis, with a contribution of 1.0 kcal mol^{-1} to binding energy. HPV16 E2 has the highest DNA binding affinity and exerts a strict transcriptional control, translated into the repression of the E6 and E7 oncogenes. These novel features provide the structural and thermodynamic basis for this tight transcriptional control, the loss of which correlates with carcinogenesis.

Papillomaviruses are a family of double-stranded DNA viruses that infect epithelia from mammals, and the outcome ranges from benign warts to cervical cancer, with particular impact on the female population of third world countries (*1*, *2*). Among the more than 100 human papillomaviruses (HPVs) identified and fully sequenced, some are associated with cervical cancer and are known as high risk strains, where the most frequent are HPV-16 and -18, followed by HPV-31 and -45 (*3*). In contrast, HPV types that are rarely found in cancers but are associated with genital and other type of warts are considered low risk HPVs, and the ones most frequently found are HPV-6 and -11 (*4*−*6*). From the eight viral products encoded, the HPV E2 protein can either enhance or repress viral transcription depending on the stage of infection or cell differentiation (*7*) and participates in viral DNA replication and episomal copy number maintenance of the viral genome (*8*, *9*). In particular, E2 represses the transcription of E6 and E7, and this ability is lost upon viral genome integration into the host cell. This phenomenon takes place through the disruption of the E2 open reading frame, leading to the deregulation of the expression of HPV E6 and E7 oncoproteins, essential for the transformation process (*5*).

The E2 proteins are ca. 400 amino acid polypeptides consisting of an *N*-terminal transactivation domain and a *C*-terminal DNA binding and dimerization domains (E2C), separated by a nonconserved hinge domain (*10*). They specifically bind the palindromic DNA sequence AC-CgNNNNcGGT known as the E2 binding site (E2BS,

[†] This work was supported by Wellcome Trust Grant GR077355AIA.

* To whom correspondence should be addressed. Tel: +39 06 7259 4835. Fax: +39 06 7259 4328 (D.O.C.) (Tel: 54 11 5238-7500 (ext 3209). Fax: 54 11 5238-7501. E-mail: gpratgay@leloir.org.ar (G.P.-G.).

[‡] Instituto Leloir and CONICET.

[§] Department of Chemical Sciences and Technologies.

10.1021/bi060123h CCC: $33.50 © 2006 American Chemical Society
Published on Web 05/05/2006

lowercase letters indicate preferred nucleotides and NNNN the variable non contacting spacer region) with 4−17 sites in the PV genome depending on the viral strain and host. The discrimination between binding sites is specific for the different virus strains; spacers rich in AT are preferred by all of the human strains. However, bovine papillomavirus 1 (BPV-1) E2 protein displays no apparent spacer sequence preference (*10*). This difference is reflected in the corresponding viral genomes: the mucosal HPV genomes have E2 binding sites with AT-rich spacers, whereas the genomes of nonprimate animal viruses (including BPV-1) have no such predominance of AT-rich spacers (*10, 11*). The binding site of HPV-16 E2C was found to have a preference for A/T or G/C base pairs flanking the minimal binding site (*12*).

The first structure of the *C*-terminal DNA binding domain (BPV1) bound to its DNA target revealed a novel fold, the dimeric β-barrel domain (*13*). Since then, several other related structures were determined, and the overall topology is very conserved (*10, 14*). Each monomer contributes with four β-strands and two helices, one of which, the recognition helix, is used to contact two successive major grooves of the palindromic DNA and operate specific DNA recognition. Although all of the structures are very similar, there is a significant difference at the level of quaternary structure: the DNA binding domain of the E2 protein belonging to HPV-16 (E2C-16) (*15*) and HPV-31 (E2C-31) (*16*) differs from the others in the relative orientation of the recognition helices. This is the consequence of a different hydrogen-bond register in the $\beta4-\beta4$ interaction at the monomer interface that shifts one monomer with respect to the other, significantly modifying the quaternary structure (*17*). This feature allows the classification of the E2C proteins into two families, depending on the orientation of the DNA binding helices. The available information regarding the DNA-bound conformation of the E2C proteins is limited; only two structures, those of E2C-BPV1 (*13*) and E2C-18 (*17*) complexed with DNA were reported. They both showed very similar quaternary structures and were obtained with consensus DNA sequences but not with natural E2 binding sites present in the corresponding viral genomes. As a consequence of the similar quaternary structures of BPV-1 and HPV-18 proteins, they both induce the same global deformation to the bound DNA. In contrast, no structural information is available on the bound conformation of E2C of HPV-16, the high risk strain that causes over 60% of the cases of HPV-linked cervical cancer and is a strong candidate for drug design. When one monomer of free E2C-16 is superimposed to the DNA-bound protein of E2C-18, the recognition helix of the nonsuperimposed subunit of E2C-16 is displaced by approximately 4.0 Å and tilted by 25° relative to the corresponding recognition helix of E2C-18 (*10*). As a consequence, if DNA binding occurs without a quaternary rearrangement of E2C-16, then a larger bending will be necessary to allow contacts between two successive DNA major grooves and the two recognition helices. However, it is still possible that upon DNA binding the quaternary structure of E2C-16 and consequently the $\beta4-\beta4$ hydrogen-bond register could change, making the relative positions of the two recognition helices closer to that observed in E2C-18 and E2C-B1 (*16*).

We have been investigating the folding (*18*−*21*) and DNA-binding mechanisms (*22*−*24*) of the HPV16 E2C DNA binding domain and addressed a number of questions regarding fundamental protein folding and DNA recognition issues in addition to the medical relevance of this virus. All evidence points to a stable but flexible unusual fold, which prompted us to determine the structure of HPV16 E2C in solution to be able to correlate structure, function, and dynamics (ref *25*, and unpublished results). Only two NMR structures of an HPV E2C domain were determined (16 and 31(*16, 25*)), and no structure for an E2C−DNA complex was determined in solution. In an attempt to fill this gap and to gain understanding of the molecular mechanism behind the enhanced transcriptional activity of this key viral protein from this particular strain (*26*), we have tackled the NMR characterization of E2C-16 (residues 283−363) bound to site 35 (E2BS2), one of the four natural E2 binding sites in the upstream regulatory region (URR) of HPV-16. In this work, we present the first NMR analysis for HPV16 E2C bound to a natural cognate site, which shows an unexpected almost identical backbone conformation to the unbound domain. We present novel structural features, a key contact away from what has been considered the entire DNA contact region, and support them with mutagenesis, binding, and thermodynamic analysis. These features provide a structural and thermodynamic basis for the high DNA discrimination and transcriptional activity of this particular strain, and we discuss the possible implications on its high cancer risk properties.

MATERIALS AND METHODS

DNA Duplexes. Double-stranded oligonucleotides containing the E2 recognition sequence (site 35 in the HPV-16 genome) were purchased Integrated DNA Technologies (Coralville, IA) and HPLC purified from. The site 35 18mer 5' GT*A ACCG* AAAT *CGGT* TGA 3' and the 14mer (underlined) were synthesized for NMR and ITC experiments and a variant with fluorescein modification for fluorescence spectroscopy (*22*). Annealing was performed by mixing equal amounts of the oligos in 10 mM sodium phosphate buffer at pH 7.0 and 200 mM NaCl and incubating for 5 min at 95 °C and slowly cooling to 25 °C for 16 h. This yielded a double-stranded oligonucleotide, and no detectable single-stranded oligonucleotide was present as judged by PAGE (not shown). Single-stranded oligonucleotide concentration was calculated using the molar extinction coefficient obtained from the nucleotide composition.

Protein Production. Inverse PCR mutagenesis (*27*) was used to produce the HPV-16 E2-C K349A mutant, which was expressed and purified as the wild-type unlabeled protein. (*20*). Uniformly ^{15}N-labeled and ^{15}N/^{13}C-labeled proteins were produced by transforming *Escherichia coli* JM109 strain cells with ptz18U-E2 plasmid. Overnight cultures of cells in M9 minimal medium were inoculated (1%) into an 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 0.6 O.D. 600 units. Protein expression was induced by adding the phage M13/T7 (Invitrogen, San Diego, CA) at a multiplicity of infection of 5 and 0.3 mM IPTG (isopropyl-*b*-thio-D-galactoside). The cells were grown overnight at 37 °C before harvesting. Protein purification was conducted as described (*20*). Once chemically pure proteins were obtained, they were unfolded in 6 M guanidine chloride, 10 mM DTT for 2 h at room temperature for a sample that could remain unmodified in

the NMR tube for longer periods. Refolding was accomplished by dialysis against 100 volumes of 50 mM sodium phosphate at pH 6.5 and 1 mM DTT. The purified protein was stored at −70 °C after snap freezing in liquid nitrogen. Protein concentration was determined using an extinction coefficient of 4194 M^{-1} cm^{-1} (*20*). The best conditions for NMR studies of the complex were found to include 50 mM sodium acetate at pH 5.6, 250 mM NaCl, 5 mM DTT, and a small excess (10−30%) of target DNA. Appropriate aliquots of protein were added to concentrated DNA provided with the calculated quantity of buffer components. Turbidity was observed during mixing, which disappeared when the solution was gently shaken.

NMR Spectroscopy. NMR experiments were all performed at 45 °C on a Bruker Avance700 spectrometer equipped with a triple resonance probe incorporating self-shielded gradient coils. Pulsed-field gradients were appropriately employed to achieve suppression of the solvent signal and spectral artifacts. Quadrature detection in the indirectly detected dimensions was obtained using the States-TPPI method. The NMR data were processed on Silicon Graphics workstations using NMRPipe and analyzed using NMRView. Linear prediction and apodization 90°-shifted squared sine-bell functions were typically applied before Fourier transformation. The final concentration of the complex for NMR studies was 0.6 mM.

Resonance Assignment. Nearly all 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, HNCA, HN(CO)CA, HNCO, and CBCA(CO)NH. A 3D ^{15}N-NOESY spectrum with 80 ms mixing time was acquired for the ^{15}N-labeled sample, a 3D ^{13}C-edited NOESY spectrum with 80 ms mixing time was acquired on the $^{15}N/^{13}C$-labeled sample, and a 3D ^{15}N-edited NOESY spectrum with 300 ms long mixing time was acquired on a perdeuterated $^{2}H/^{15}N$-labeled sample in water. The resonance assignment of most side chain nuclei was achieved from an analysis of NOESY spectra and by comparison with Chemical Shifts and NOESY cross-peak patterns of the DNA-free HPV16 E2C protein, whose NMR structure was previously determined by our group (*25*).

Approximate interproton distances were derived from NOESY spectra, and the corresponding 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, obtaining 532 intramonomer and 24 intermonomer distance restraints. The hydrogen bonds were recognized by measuring the exchange rates of amide protons and by evaluating the spatial relationships of slow exchangeable protons with potential acceptors in the DNA-free structure. Two distance restraints were defined for each hydrogen bond: 1.8−2.3 Å for the H−O distance and 2.3−3.3 Å for the N−O distance, obtaining 64 intramonomer and 6 intermonomer hydrogen bond restraints. A set of 51 one-bond ^{1}H-^{15}N residual dipolar couplings per monomer were measured by recording (F1) ^{1}H-coupled ^{1}H-^{15}N HSQC spectra with the IPAP version on the ^{15}N-labeled DNA−protein complex samples in isotropic medium and in liquid crystalline medium consisting of 17 mg/mL filamentous phage Pf1 (Asla Labs, Sweden) to induce the molecular alignment (*28*).

Structure calculation of the E2C-16 bound conformation was performed with XPLOR (version 3.85) using the two-step refinement strategy developed by Chou and co-workers (*29*). This protocol was slightly modified to incorporate NOE restraint information and appropriate energy terms to maintain the symmetry of the dimer (*30*). With this approach, the X-ray structure of HPV-16 E2C (pdb 1BY9) was used from the starting model, and all its backbone ϕ and ψ angles were used as tight dihedral restraints during step 1 of the structural refinement. The molecule was allowed to differentiate from the starting model by subjecting it to a cycle of low-temperature simulated annealing stage (by cooling the system from 200 to 20 K) and using the experimental information (RDCs, NOEs, and hydrogen bonds) and symmetry restraints. Subsequently, the structure with the lowest dipolar coupling energy obtained by the first-stage refinement was subjected to a second step of low-temperature (20 K) refinement in which dihedral angle restraints were gradually weakened during the molecular dynamics. Details can be found in the original reference (*29*).

The lowest global energy dimer among the generated structures displaying no violations of the restraints less than a defined threshold (0.3 Å for distances, and 1 Hz for RDCs) was chosen to represent the solution structure of the E2C bound-DNA conformation. To avoid bias toward the structure of HPV16 E2C, two other rounds of calculations were performed, only differing in the starting structure. In the first case, the X-ray structure of DNA-free HPV16 E2C was employed as the starting structure, whereas in the second case, a model consisting of an E2C-16 amino acid sequence but possessing the backbone dihedral angles of the X-ray structure of DNA-bound HPV18 E2C (pdb 1JJ4, representing a distinct structural E2C subfamily) was used as the starting structure. In both cases, intermonomer hydrogen bonds belonging to the β4 strands were not included as restraints. The above-described two-step refinement protocol was used, and the resulting lowest energy structures were selected with the same criteria and analyzed. No doubling in protein resonances was observed, considering the palindromic nature of the DNA target.

DNA Binding. Fluorescence measurements were recorded using an Aminco Bowman series 2 luminescence spectrometer assembled in L geometry. For fluorescein anisotropy measurements, excitation was set to 495 nm with a 4 nm slit, and emission was recorded at 520 nm. The temperature was kept constant at 25 °C. All titrations were performed adding small amounts of a concentrated solution of the variable ligand to a fixed amount of the other and allowed to equilibrate for 3 min. In all cases, maximal dilution was 20%, and the data were corrected accordingly. Dissociation constants for the E2C−DNA interaction were performed in 50mM sodium acetate at pH 5.6, 250 mM NaCl, 5mM $MgCl_2$, and 1 mM DTT by measuring the steady-state fluorescence anisotropy and the fluorescence intensity of the fluorescein-labeled DNA as a function of added E2C. Data were fitted to a quadratic equation as described (*22*).

Isothermal Titration Calorimetry. Experiments were performed using VP-ITC (MicroCal, USA), and Microcal ORIGIN 5.0 software was used for both data acquisition and analysis. In each titration, 12 μM E2C protein, in the cell, was titrated with several injections of 120 μM ligand in the buffer used for NMR experiments at 25 °C. The volume of

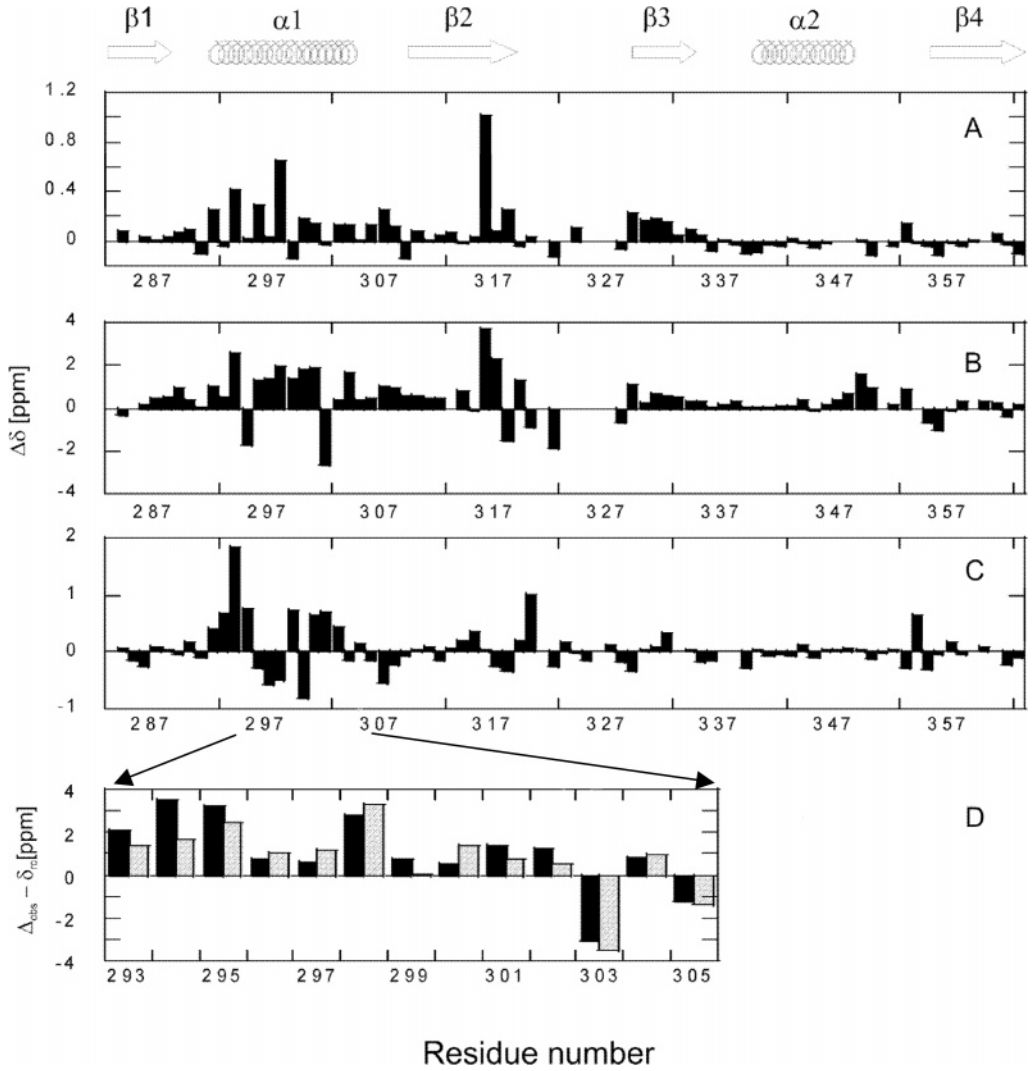


FIGURE 1: Chemical shift differences between DNA-bound and free E2C-16. (A) HN, (B) N, (C) C^{α} nuclei, and (D) chemical-shift differences for C^{α} between the observed shift and the random coil value for the recognition helix in the DNA-bound (black) and free (gray) E2C-16 forms.

each injection was of 7 μL, except for the first injection, which was 2 μL. Injections were continued beyond saturation levels to allow for the determination of the heat of ligand dilution. The resulting data were fit to a single-site binding isotherm after subtracting the heat of dilution, using the ORIGIN software supplied with the calorimeter. Estimated errors in thermodynamic values were automatically calculated by the software.

RESULTS

Chemical Shift Analysis of the DNA-Bound Protein. The chemical-shift assignment of backbone atoms and most of the side chains of the DNA-bound E2C-16 domain was obtained by multidimensional NMR experiments conducted on the $^{15}N/^{13}C$-labeled protein (the resulting chemical shifts were deposited at the BMRB data bank, accession number: 6877). An analysis of the chemical-shift differences between the free and bound states of E2C-16 allowed the first insight into conformational changes induced by DNA binding. Figure 1A−C shows the HN, N, and Cα chemical-shift differences. Not surprisingly, the region showing most of the changes induced by DNA binding is the recognition helix (residues 294−304). This helix makes most of the direct contacts with the DNA major groove and shows rather solvent exposed backbone NH groups, which was interpreted as an evidence of conformational flexibility (*16*), and a more structured recognition helix was expected to be formed upon DNA binding. In this respect, the Cα chemical shift is of great help: the secondary chemical shift induced by helical conformation yields a Cα chemical shift higher than the random coil value (*31*). Analysis of the differences between the observed Cα chemical shifts and the random coil values for the free and bound E2C-16 are shown in Figure 1D. This result shows that the recognition helix presents two distinct regions: the *N*-terminus (residues 294−298) behaves as a normal α-helix, and shows a stabilization of the first three residues at the N cap upon DNA binding. The second half of the helix (residues 299−305) shows deviations from α-helical character, and particularly, Phe303, presenting a negative chemical-shift difference, both in the free and complexed forms. On the basis of the available crystal structure of the DNA complex of E2C-18, it is expected that only residues 294−301 will be in direct contact with the DNA. We show that within this region the first three residues show a global enhancement of helical character upon DNA binding, probably due to the conformational stability pro-

vided by the major groove contacts.

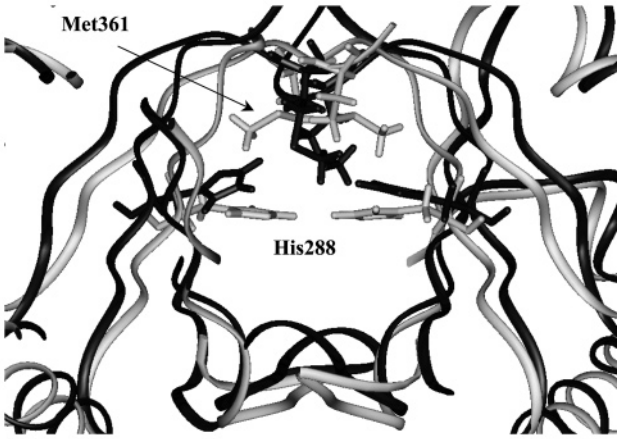
In addition, a large downfield shift (~1 ppm) is observed for HN of Thr316. In general, shifts of this magnitude are observed when a strong hydrogen bond is formed by the HN atom in the final structure. In the free protein, Thr316 is located in a bulge of the β2 strand, pointing outward the plane of the β-sheet without being involved in a hydrogen bond (*25*). The corresponding residue in E2C-18, Thr319, makes a hydrogen bond with the phosphate that links the last T of the spacer with the subsequent C in the DNA complex (*17*). Our data strongly suggest that this interaction is also present in the E2C-16 DNA complex here reported.

As already discussed, the main structural difference between the HPV-18 and the HPV-16 strains resides in the intermonomer β4−β4 sheet. A shift in hydrogen bond would cause a significant change in the relative orientation of the two recognition helices (*11*). The chemical shift difference data between bound and free E2C-16 for backbone nuclei of the β4 region reveal only minor adjustments and are, therefore, not compatible with a substantial conformational rearrangement implying changes in the hydrogen-bond register and the overall quaternary arrangement (*17*) (Figure 1A−C).

To further investigate possible changes that occur in the dimer interface upon DNA binding, we moved our attention to key side chains that may change their chemical environment and, hence, the chemical shift of their nuclei when comparing the dimeric interface of E2C-16 to that of E2C-18. An analysis of the available structures suggest that most of the side chains forming the interface remain in the same conformation but that mainly two side chains may potentially change their chemical shift following a hypothetical rearrangement: His288 and Met361. The two His288 side chains in the dimeric interface of E2C-16 are very close to one another, and in the crystal structure, a water molecule was modeled as a bridge between them (*15*). The free NMR structure showed a somewhat larger distance but still, with a relatively close disposition of the two N$^{\epsilon}$2 atoms (*25*). In contrast, the same two histidines of E2C-18, both in the free and bound forms, are more separated, and the resulting space between them is filled by the side chains of the two Met363 (the corresponding residue to Met361 of E2C-16) (*17*) (Figure 2A). The observed chemical shifts of the ϵ-CH_3 of Met361 are almost coincident between the free and bound forms of E2C-16 (Figure 2B), confirming that no major rearrangement of the dimer interface occurred in solution. However, His288 displays a slight shift for both N atoms (3 ppm, which is not considerable with respect to the 70 ppm scale for nitrogen nuclei in histidines). It can reflect a slight change in pK_a for His288. In fact, in the complexed state, a 5% increase of the protonated form can account for this change. Alternatively, it can be speculated that a more stable dimeric interface is formed by DNA binding, and the presence of a more stable bridge through a water molecule or a cation can account for the upfield shift observed for the N$^{\epsilon}$2 nuclei.

The loop connecting β2 to β3 also displays novel features in the complexed state. This region, which on the basis of the structures of the E2C-18 and E2C-BPV1 complexes is expected to lie just in front of the spacer sequence, did not show density in the crystal structures of unbound E2C, suggesting a substantial degree of flexibility (*15*, *17*).

A



B

Chemical Shifts (ppm)			
Residue	Nucleus	Free	Complex
His288	$N^{\delta 1}$	165.1	168.3
His288	$N^{\epsilon 2}$	246.7	242.5
Met361	C^{ϵ}	14.60	14.54
Met361	H^{ϵ}	2.354	2.401

FIGURE 2: Differences at the dimer interface of E2C-16 and E2C-18. (A) Side chain positioning of H288 and M361 of E2C-16 (black) compared with the legend correspondent positions in E2C-18 (grey). (B) Chemical shifts of the same nucleus in free and DNA-bound forms of E2C-16.

Moreover, it was also shown to be disordered in the E2C-18−DNA complex, whereas in the E2C-BPV1−DNA complex, it makes contact with the phosphate groups of the spacer region (*13*, *17*). An inspection of H-N HSQC spectra of bound E2C-16 reveals that only a few HNs of the residues belonging to the β2−β3 loop (residues 326−335) are detectable in the complexed state, as already observed in the free form of the protein (*16*, *25*). This fact along with low ^{1}H-^{15}N NOE measured values for the surviving cross-peaks (below 0.5, data not shown), allows us to infer that the β2−β3 loop remains both solvent exposed and flexible in the E2C-16−DNA complex. In this respect, E2C-16 and E2C-18 seem to behave in a similar way, and both retain a flexible β2−β3 loop in the complexed state.

Structure of the DNA-Bound Protein Using Residual Dipolar Coupling Analysis. A more quantitative picture about structural changes that occur in the E2C-16 protein upon DNA binding was obtained by using residual dipolar couplings and NOEs. Orientation of the HN-bond vectors in the bound state of E2C-16 were obtained through the measurement of 102 HN−N residual dipolar couplings using the Pf1 phage as the alignment medium. We have tested the conformational change that the protein undergoes upon DNA binding by back-calculation of the expected RDCs using the crystal structure of free E2C-16 (pdb 1BY9). Figure 3A shows the correlation between the observed and calculated RDCs. A Q factor (*32*) of 25% is obtained, which is a clear indication that no extensive change in the conformation of the protein has occurred upon DNA binding.

A large number of NOEs (including 670 long rage NOEs) were assigned from the ^{15}N- and ^{13}C-edited NOESYs. This information was used as experimental constraints in the refinement protocol described by Chou (*29*). Using this

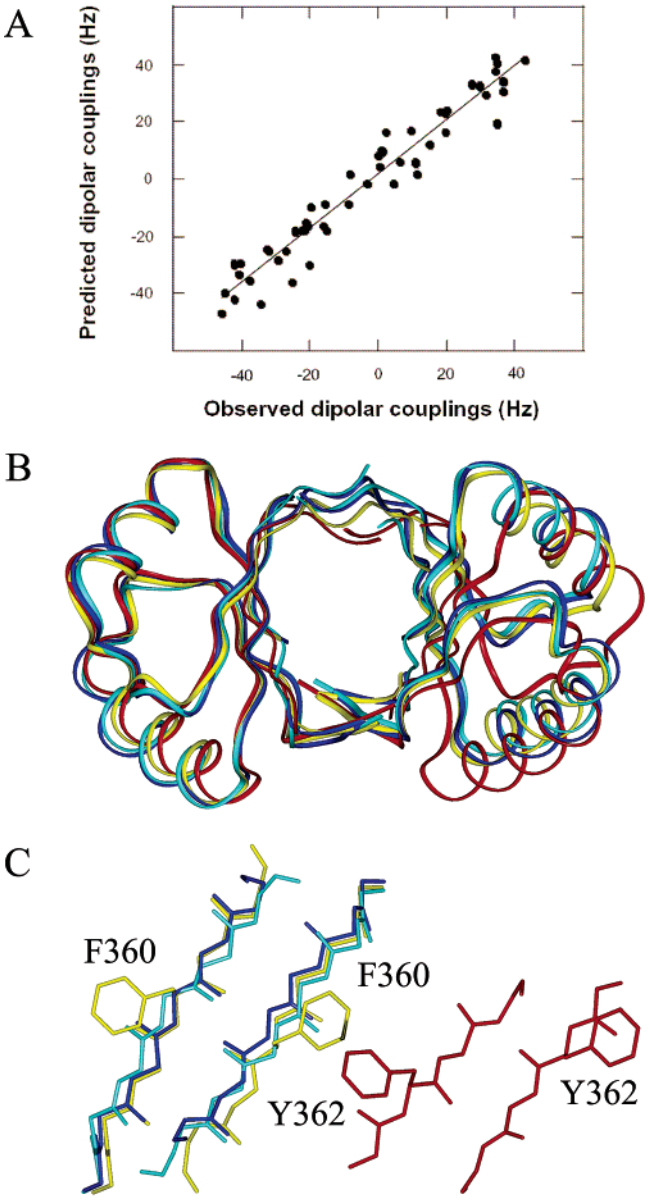


FIGURE 3: Comparison of free and DNA-bound forms of E2C-16 with the crystal structure of E2C-18. (A) Correlation between observed HN−N dipolar couplings and couplings predicted for the X-ray structure of free E2C-16. The correlation coefficient R_{SVD} is 0.97, and Q is 25%. The allignment tensor relative to this structure D_a^{NH} is 24.2 Hz, and rhombicity R is 0.52. (B) Superimposition of free (yellow) E2C-16 with (i) DNA-bound E2C-16, using as a starting point for refinement, the E2C-18 (blue) and E2C-16 (light blue) crystal structures and (ii) DNA-bound E2C-18 (red). Only the left monomer was superimposed (residues 284−319 and 329−364 of E2C-16). (C) Alignment of β4 strands for the four structures. The coordinates of the DNA-bound E2C-16 were deposited in the Protein Data Bank (id: 1ZZF).

approach, it is possible to derive the structural information of a given state of a protein (in our case, the bound form) starting from another (the free form of the protein). We have performed two different calculations, starting either from the free E2C-16 or the DNA-complexed E2C-18 crystal structures. Along with the NOE and RDC information, the hydrogen-bond network is also used in this protocol. However, in the first step of refinement, we have omitted those hydrogen bonds involving β4−β4 intermonomer interaction because this is the main difference between the two backbone conformations. The two structures showing the lowest RDC and NOE energies are presented in Figure 3B, and the backbone RMSDs are shown in Table 1.

Table 1: Backbone RMSD values of E2C Structures

	E2C-16[a]		E2C-18[b]		E2C-16C[c]		E2C−R18[d]	
	mono[e]	dim[f]	mono	dim	mono	dim	mono	dim
E2C-16			1.06	1.68	0.87	0.91	0.85	0.88
E2C-18					1.45	2.11	1.38	1.95
E2C-16C							0.61	0.66

[a] The X-ray structure of DNA-free E2C-16 (pdb 1BY9). [b] The X-ray structure of DNA-bound E2C-18 (pdb 1JJ4). [c] The NMR structure of DNA-bound E2C-16 starting from DNA-free E2C-16 (pdb 1ZZF).[d] The NMR structure of DNA-bound E2C-16 starting from DNA-bound E2C-18. [e] Residues 284−319 and 329−364 of the E2C-16 single monomer. [f] Same residues of both dimers.

Although these structures cannot be regarded as high-resolution structures of the complex because only HN−N RDC were used and no intermolecular NOEs between the protein and the DNA were included in the calculation, Figure 3 and Table 1 clearly show that the two calculated structures are almost identical, and both converge to the quaternary structure observed for free E2C-16. The relative orientation of the two β4 strands and the two recognition helices are almost identical to that observed in the NMR structure of the unbound E2C-16 (Figure 3B and C) but differ substantially from that of DNA-bound E2C-18. This result confirms our conclusion drawn from the chemical-shift analysis, that is, DNA binding does not lead to a significant quaternary structure rearrangement of E2C-16.

K349 Is Not in the Known DNA Binding Site and Makes an Unexpected Contact with DNA. To investigate the effect of DNA length, we carried out an HSQC experiment of HPV16 E2C bound to a 14 base pair duplex, containing the same E2 site, and looked for changes in chemical shifts. We observed an unexpected chemical-shift perturbation of the K349 HN with respect to the free protein signal in the complex with an 18 bp DNA duplex, not observed in the complex between E2C-16 and a 14 bp DNA duplex (Figure 4). K349 is located in the C cap of the minor α-helix but faces the side where the DNA binds, strongly suggesting additional, not previously observed contacts with bases flanking the E2 site sequence ACCGN4CGGT (Figure 4B). As a further test, we performed isothermal titration calorimetry (ITC) of the 14mer duplex and compared with the 18mer reference duplex under the same conditions in which the NMR experiments were carried out. Figure 5A (inset) shows the raw calorimetric data for the site 35 18mer duplex showing the steep change in specific heat, expected for a tight binding. Whereas the 18mer duplex has a ΔH value of -16.2 ± 0.1 kcal mol^{-1}, the 14mer duplex yields a ΔH value of -5.8 ± 0.1 kcal mol^{-1} (Figure 5A). The 10 kcal mol^{-1} drop in the measured binding enthalpy confirms that there are additional contacts beyond the known minimal E2 binding sequence. Moreover, such a large difference strongly suggests that it is highly cooperative with the rest of the specific DNA-contacting residues and cannot be assigned to a single interaction being broken.

Because tight binding affinities cannot be accurately determined from ITC experiments, we determined the dissociation constants for the 18mer and 14mer duplexes under the NMR buffer conditions by fluorescence spectroscopy. The dissociation constant (K_D) of the 18mer duplex

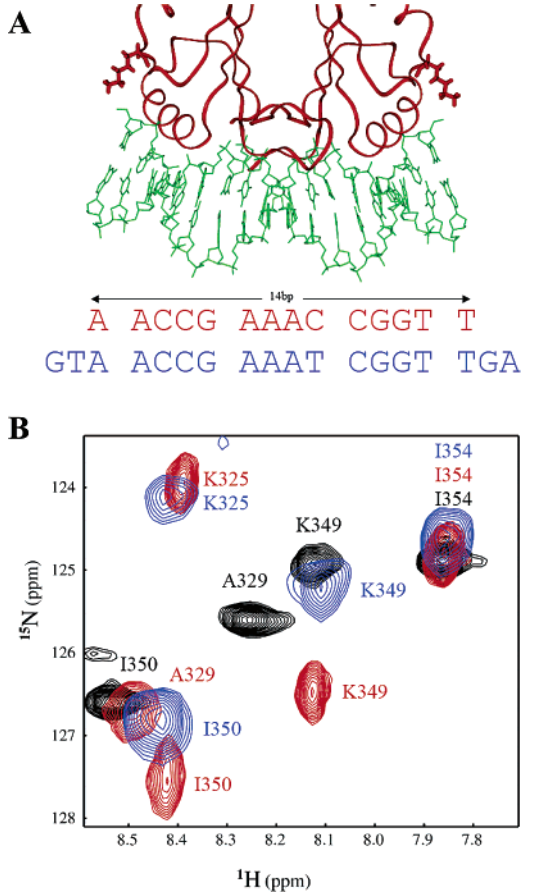


FIGURE 4: Structure of E2C-16 bound to the 14 bp E2 binding site 35 (E2BS2). (A) Structural model of E2C-16 bound to the 14 bp site showing the position of the K349 side chain. The bottom part shows the base sequence of the 18 bp and 14 bp E2 binding sites. (B) Section of the E2C-16 ^{15}N HSQC spectrum showing the cross-peaks for K349 in the free protein (black) and those for the protein bound to the 18 bp site (red) or complexed to the 14 bp site (blue).

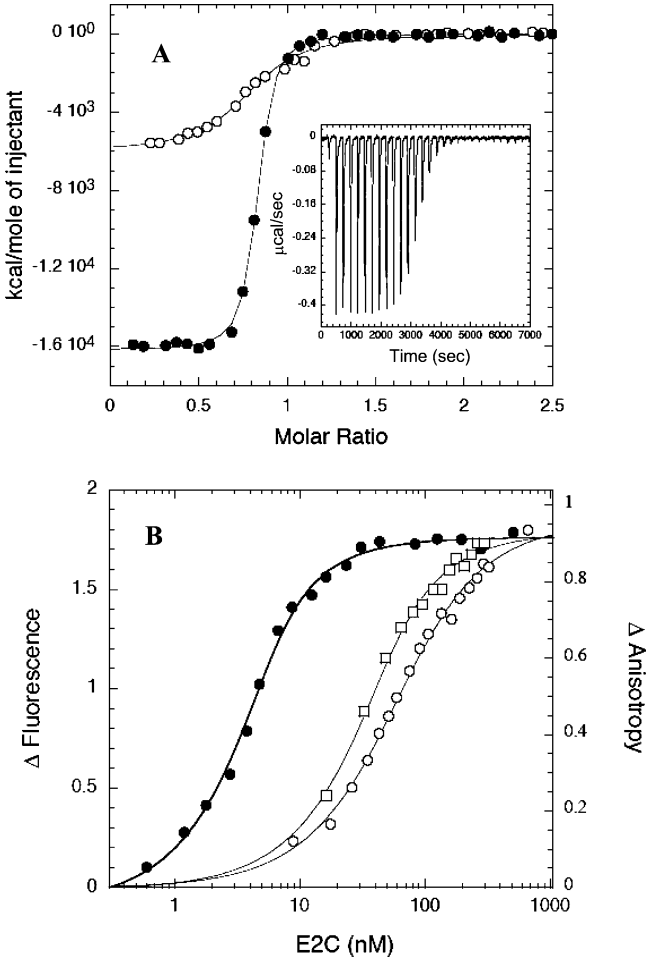


FIGURE 5: Binding of wild-type E2C to 14mer and 18mer sites. (A) ITC binding isotherms resulting from the integration of the specific heats with respect to time, with the appropriate molar correction for site 35-18 (●) or site 35-14 (○). Inset: raw ITC data for the titration with site 35-18. (B) Titration of site 35-18 with E2C monitored by fluorescence intensity (●) or site 35-14 monitored by fluorescence anisotropy (○) and the binding of site 35-18 with the K349A mutant followed by anisotropy (□). The Wiseman *c* value for the weaker interaction (14 bp, K_D 0.34 ± 0.03 μM) is 35, well within the accepted range (10−500) (*41*).

in these conditions (pH 5.6) is 1.3 ± 0.3 nM, slightly weaker than that measured at pH 7.0 (Figure 5B, (*22*)). The affinity of the 14mer duplex in identical conditions drops to 38.4 ± 4 nM, corresponding to a 2.0 kcal mol^{-1} change in the binding free energy of the dimer and to ∼1.0 kcal mol^{-1} per symmetric interaction, in excellent agreement with the average binding free energy values for each of the residues of the major helix contacting the DNA, measured in our recent site-directed mutagenesis analysis (*24*). Finally, we mutated the K349 to alanine and evaluated the binding affinity to the 18mer duplex by fluorescence spectroscopy (Figure 5B). The affinity is decreased by 1.4 kcal mol^{-1} (K_D 14 ± 2 nM), 0.7 kcal mol^{-1} per symmetric interaction, a magnitude similar to the effect of shortening the duplex to 14 bases, thus confirming the additional contact. The K_D value for the interaction with the 14mer duplex is 20 ± 2 nM. Thus, the binding free energy change between the 18mer and the 14mer in the case of the K349A mutant is 0.2 kcal mol^{-1}, compared to 2.0 kcal mol^{-1} for wild-type E2C-16, confirming the absence of the interaction between the mutated side chain and the additional DNA sequence in the 18mer E2 site.

DISCUSSION

To date, only a handful of protein−DNA complexes have been analyzed in detail by NMR techniques, despite it being a method of choice for investigating the flexibility and dynamics of these complexes, essential for the regulatory roles in transcription control (*33*). The recent progress in NMR applied to larger protein−DNA complexes allowed the description of ternary complexes such as in the case of Oct-1 and Sox-2 bound to adjacent DNA sites (*34*) and allowed the comparison between specific and nonspecific binding modes in the Lac repressor (*35*).

In this article, we present the first NMR analysis of the HPV16 E2C domain bound to a natural, specific DNA site and determine novel features behind the tight transcription control in this high cancer risk strain. The NMR model presented here indicates that E2C-16 does not significantly change its conformation upon DNA binding. This behavior is common to E2C-18 (*17*) but not to E2C-BPV1 (*36*). For the latter, a slight difference in quaternary structure between the free and DNA bound forms was observed, although some distortions in the free protein structure can be anticipated by the oxidized state of the crystallized protein (*36*). What differentiates E2C-16 from the other two complexes already characterized is the larger induced DNA bending. To

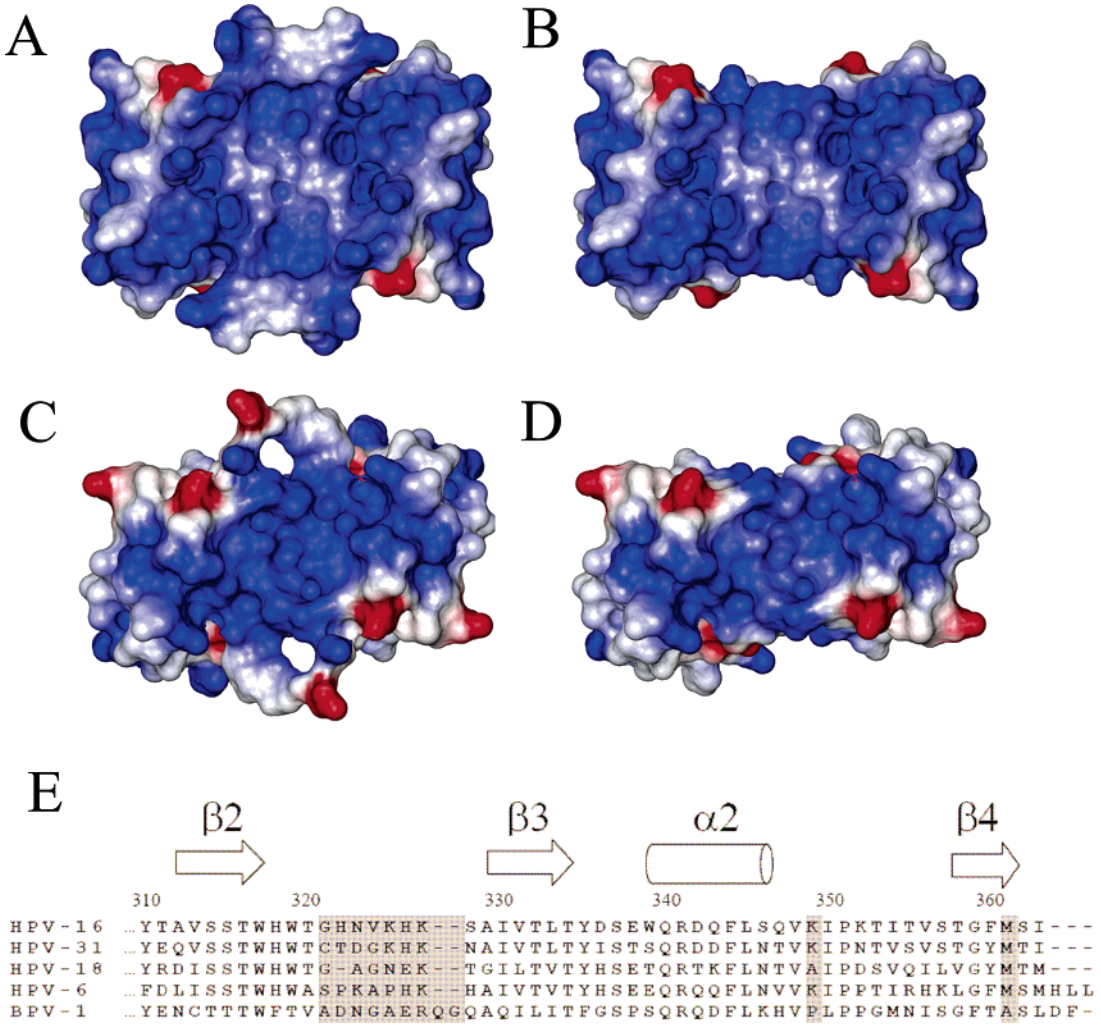


FIGURE 6: Comparison of electrostatic potential energy surfaces of E2C-16 and E2C-18. (A) E2C-16 NMR structure, (B) E2C-16 X-ray structure (does not include the central $\beta2-\beta3$ loop, residues 321−328), (C) E2C-18 crystal structure with a modeled $\beta2-\beta3$ loop (residues 324−330), and (D) E2C-18 crystal structure. Note the largely different contribution to the electrostatic energy potential of the $\beta2-\beta3$ loop in E2C-16 and E2C-18. (E) Amino acid sequence alignment of the *C*-terminal portion of HPV16 E2C and related types discussed in this work. The $\beta2-\beta3$ loop, K349, and M361 are shaded in gray. The alignment was carried out using the program Tcoffee, which uses structural information (*42*).

accomplish such a large structural deviation, the spacer sequence must be particularly deformable, explaining the preferences shown by E2C-16 for DNA sequences containing A tracts. The presence of an AT rich spacer in E2C-6 was shown to facilitate contacts with the additional flanking base (±7) (*10, 14*). In addition to the deformability of the DNA sequence that accommodates the two recognition helices, the protein may also contribute to DNA bending upon formation of the complex by using positive charges that complement the negatively charged phosphate backbone of the nucleic acid. Clearly, there are some peculiarities of the HPV-16 E2C domain that can contribute to a more pronounced DNA bending. As already pointed out (*10*), this domain shows a less positively charged surface in the recognition helices compared with that of E2C-BPV1. However, E2C-16 shows two additional regions that can help in DNA bending, which are the $\beta2-\beta3$ loop and the $\alpha2-\beta4$ loop. In the first case, it can be observed that E2C-16 presents the highest number of positively charged residues among all the proteins for which the structure is known (two lysines and the potential positive charges of two histidines; Figure 6A). This loop is flexible in the free form of the protein and was not observed in the crystal structure of E2C-16, thus underestimating the electrostatic potential (*10*). However, the loop was found to be ordered in the E2C domain from HPV6, likely due to the presence of two proline residues, but is almost as positively charged as E2C-16 (Figure 6E) (*10, 14*).

In the case of the second loop, the presence of K349, which is positioned very close in space to the DNA interacting R304 of the recognition helix, appears now as an additional factor contributing to DNA bending. A lysine residue is also present at position 349 in E2C-31 and E2C-6 (Figure 6E). The kinetic DNA binding mechanism of HPV16 E2C to E2 site 35 reveals two late rearrangement steps, one of which must involve DNA bending to yield a consolidated complex (*23*).

The enthalpy difference ($\Delta\Delta H$) of 10.0 kcal mol^{-1} between the 18mer and the 14mer site is too large to be assigned to the single interaction between K349 and the DNA, especially so when compared to the binding free energy change ($\Delta\Delta G$) obtained from fluorescence titrations at equilibrium, which is 2.0 kcal mol^{-1}. In addition, the average binding free energy change measured for all individual DNA-contacting side chains of E2-16 is ~1.0 kcal

mol^{-1}. We believe that most of this enthalpy difference is due to conformational changes in the DNA, which agrees very well with the fact that it is the DNA that must adapt and that the change is more pronounced in HPV16 E2C (*17*, *37*−*39*). It is not possible at this stage to quantify precisely because the interaction with K349 is coupled to the DNA deformation, and both phenomena cannot be analyzed separately in the absence of structural studies on the E2 DNA binding site bound to E2C. We shall await NMR studies on the E2 DNA binding site bound to E2C.

E2C-18 and E2C-BPV1 do not present such charge distributions (Figure 6B); the $\beta2-\beta3$ loop of E2C-18 is shorter by one amino acid and does not contain a positively charged residue. K349 is replaced by an alanine in E2C-18, the mutation that weakens the DNA binding of the E2C-16 domain, and by a proline in E2C−BPV1 that may possibly restrict the conformation of the loop but involves no positive charges. In this respect, both E2C-18 and E2C-BPV1 represent exceptions because there is always a lysine or arginine in this position among the more frequently found high- and low-risk viral strains.

A possible link between the different structural features of the E2C−DNA complex for HPV-16 compared to those of HPV-18 and BPV-1 and their biological roles has emerged in a study about the transcriptional activity of different E2 proteins (*26*). HPV16 E2 was shown to bind with a higher affinity to DNA fragments containing E2BS1 and E2BS2 compared to that of HPV18 E2, HPV11 E2, and BPV1 E2 (*26*). These two E2BSs are directly related to the ability of E2 proteins to act as transcriptional repressors of E6 and E7 oncogenes. It is interesting to note that in HPV11 and HPV18 the spacer sequence of E2BS1 and E2BS2 that mediate the transcriptional regulation by E2 contain only AAAA sequences. However, HPV16 E2BS1 and E2BS2 show AAAC and AAAT spacers, respectively. It was already shown that HPV16 E2 exhibits a relatively high affinity for sequences containing AAAA spacers in the same order as that shown for AATT sequences (*10*). HPV18 E2C, however, shows high affinity for AATT sequences but only moderate affinity for AAAA sequences. This could explain the differences in the observed transcriptional activities of the E6 and E7 promoters. The consensus recognition sequence initially described for the E2 sites was ACCGN4CGGT (Figure 4, (*40*)) and the first structure determined for a complex used a perfect 16mer palindromic duplex, where contacts were observed up to base pairs ±6 from the center of the palindrome (*13*) (Figure 4B). Base replacement and mutagenesis studies, demonstrated that the site of HPV16 E2 extended to preferentially recognize A/T or G/C base pairs at position ±7 ((*12*), Figure 4B). In this article, we show that at least an 18 base pair site is required for high-affinity binding recognition by HPV16 E2C and that a key residue out of the major DNA-contacting helix is involved in this interaction.

It was speculated that HPV-16 developed a more transcriptionaly active E2 protein to tightly regulate its highly active E6 and E7 oncoproteins and govern the viral life cycle more stringently than other HPVs, which express less potent E6 and E7 proteins (*26*). Once regulation of the E6/E7 promoter by high-risk E2 proteins is lost because of viral integration, cancer may then develop. The larger DNA bending necessary for binding and the high deformability of sequences containing the AAAN spacer can add up to yield an increased affinity of E2C-16 for E2BS1 and E2BS2, contributing to its enhanced transcriptional activity. Thus, the new structural and thermodynamic data we present here indicate (i) that the protein backbone remains virtually unchanged in the E2C−DNA complex suggesting that it is the DNA that must deform significantly; (ii) the presence of a new key contact out of the DNA-binding helix contributing to binding energy, which interacts with a yet unidentified base flanking the minimal E2 binding sequence; and (iii) a particular positive charge distribution only observable in the NMR structure, which, together with the newly defined unexpected electrostatic contact, strongly suggests a main role in DNA deformation. All of these features contribute to the high transcriptional activity of HPV16 E2, required in particular to repress the expression of the E6 and E7 oncogenes. Further NMR studies will reveal details of the DNA conformation and the side chain−base contacts.

ACKNOWLEDGMENT

We thank Alison McBride and Fernando Goldbaum for helpful criticisms to the manuscript. A.D.N. holds a fellowship from CONICET.

REFERENCES

1. Howley, P. M. (1996) Papillomaviridae: The viruses and their replication, in *Fields Virology* (Fields, B. N., Knipe, D. M., and Howley, P. M., Eds.) 3rd ed., Vol. 65, pp 2045−2076, Lippincott-Raven Publishers, Philadelphia, PA.
2. Cohen, J. (2005) Public health. High hopes and dilemmas for a cervical cancer vaccine, *Science 308*, 618−621.
3. Munoz, N., Bosch, F. X., de Sanjose, S., Herrero, R., Castellsague, X., Shah, K. V., Snijders, P. J., and Meijer, C. J. (2003) Epidemiologic classification of human papillomavirus types associated with cervical cancer, *N. Engl. J. Med. 348*, 518−527.
4. zur Hausen, H. (2000) Papillomaviruses causing cancer: evasion from host-cell control in early events in carcinogenesis, *J. Natl. Cancer Inst. 92*, 690−698.
5. Dell, G., and Gaston, K. (2001) Human papillomaviruses and their role in cervical cancer, *Cell Mol. Life Sci. 58*, 1923−1942.
6. Munoz, N. (2000) Human papillomavirus and cancer: the epidemiological evidence, *J. Clin. Virol. 19*, 1−5.
7. Steger, G., and Corbach, S. (1997) Dose-dependent regulation of the early promoter of human papillomavirus type 18 by the viral E2 protein, *J. Virol. 71*, 50−58.
8. Piirsoo, M., Ustav, E., Mandel, T., A., S., and Ustav, M. (1996) Cis and trans requirements for stable episomal maintenance of the BPV-1 replicator, *EMBO J. 15*, 1−11.
9. Bastien, N., and McBride, A. A. (2000) Interaction of the papillomavirus E2 protein with mitotic chromosomes, *Virology 270*, 124−34.
10. Hegde, R. S. (2002) The papillomavirus E2 proteins: structure, function, and biology, *Annu. Rev. Biophys. Biomol. Struct. 31*, 343−60.
11. Hines, C. S., Meghoo, C., Shetty, S., Biburger, M., Brenowitz, M., and Hegde, R. S. (1998) DNA structure and flexibility in the sequence-specific binding of papillomavirus E2 proteins, *J. Mol. Biol. 276*, 809−818.
12. Thain, A., Webster, K., Emery, D., Clarke, A. R., and Gaston, K. (1997) DNA binding and bending by the human papillomavirus type 16 E2 protein. Recognition of an extended binding site, *J. Biol. Chem. 272*, 8236−8242.
13. Hegde, R. S., Grossman, S. R., Laimins, L. A., and Sigler, P. B. (1992) Crystal structure at 1.7 A of the bovine papillomavirus-1 E2 DNA-binding domain bound to its DNA target, *Nature 359*, 505−12.
14. Dell, G., Wilkinson, K. W., Tranter, R., Parish, J., Leo Brady, R., and Gaston, K. (2003) Comparison of the structure and DNA-

binding properties of the E2 proteins from an oncogenic and a non-oncogenic human papillomavirus, *J. Mol. Biol. 334*, 979−991.
15. Hegde, R. S., and Androphy, E. J. (1998) Crystal structure of the E2 DNA-binding domain from human papillomavirus type 16: implications for its DNA binding-site selection mechanism, *J. Mol. Biol. 284*, 1479−1489.
16. Liang, H., Petros, A. M., Meadows, R. P., Yoon, H. S., Egan, D. A., Walter, K., Holzman, T. F., Robins, T., and Fesik, S. W. (1996) Solution structure of the DNA-binding domain of a human papillomavirus E2 protein: evidence for flexible DNA-binding regions, *Biochemistry 35*, 2095−2103.
17. Kim, S., Tam, J., Wang, A., and Hegde, R. (2000) The structural basis of DNA target discrimination by papillomavirus E2 proteins, *J. Biol. Chem. 275*, 31245−31254.
18. de Prat-Gay, G., Nadra, A. D., Corrales-Izquierdo, F. J., Alonso, L. G., Ferreiro, D. U., and Mok, Y. K. (2005) The folding mechanism of a dimeric beta-barrel domain, *J. Mol. Biol. 351*, 672−682.
19. Mok, Y. K., Bycroft, M., and de Prat-Gay, G. (1996) The dimeric DNA binding domain of the human papillomavirus E2 protein folds through a monomeric intermediate which cannot be native-like, *Nat. Struct. Biol. 3*, 711−717.
20. Mok, Y. K., de Prat Gay, G., Butler, P. J., and Bycroft, M. (1996) Equilibrium dissociation and unfolding of the dimeric human papillomavirus strain-16 E2 DNA-binding domain, *Protein Sci. 5*, 310−319.
21. Mok, Y. K., Alonso, L. G., Lima, L. M., Bycroft, M., and de Prat-Gay, G. (2000) Folding of a dimeric beta-barrel: residual structure in the urea denatured state of the human papillomavirus E2 DNA binding domain, *Protein Sci. 9*, 799−811.
22. Ferreiro, D. U., Lima, L. M., Nadra, A. D., Alonso, L. G., Goldbaum, F. A., and de Prat-Gay, G. (2000) Distinctive cognate sequence discrimination, bound DNA conformation, and binding modes in the E2 C-terminal domains from prototype human and bovine papillomaviruses, *Biochemistry 39*, 14692−14701.
23. Ferreiro, D. U., and de Prat-Gay, G. (2003) A protein-DNA binding mechanism proceeds through multi-state or two-state parallel pathways, *J. Mol. Biol. 331*, 89−99.
24. Ferreiro, D. U., Dellarole, M., Nadra, A. D., and de Prat-Gay, G. (2005) Free energy contributions to direct readout of a DNA sequence, *J. Biol. Chem. 280*, 32480−32484.
25. Nadra, A. D., Eliseo, T., Mok, Y. K., Almeida, C. L., Bycroft, M., Paci, M., de Prat-Gay, G., and Cicero, D. O. (2004) Solution structure of the HPV-16 E2 DNA binding domain, a transcriptional regulator with a dimeric beta-barrel fold, *J. Biomol. NMR 30*, 211−214.
26. Hou, S. Y., Wu, S. Y., and Chiang, C. M. (2002) Transcriptional activity among high and low risk human papillomavirus E2 proteins correlates with E2 DNA binding, *J. Biol. Chem. 277*, 45619−45629.
27. Clackson, T., Detlef, G., and Jones, P. (1991) in *PCR, a Practical Approach* (McPherson, M., Quirke, P., and Taylor, G., Eds.) pp 202, IRL Press, Oxford, U.K.
28. Bax, A., Kontaxis, G., and Tjandra, N. (2001) Dipolar couplings in macromolecular structure determination, *Methods Enzymol. 339*, 127−174.
29. Chou, J. J., Li, S., and Bax, A. (2000) Study of conformational rearrangement and refinement of structural homology models by the use of heteronuclear dipolar couplings, *J. Biomol. NMR 18*, 217−227.
30. O'Donoghue, S. I., King, G. F., and Nilges, M. (1996) 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 8*, 193−206.
31. Wishart, D. S., and Sykes, B. D. (1994) Chemical shifts as a tool for structure determination, *Methods Enzymol. 239*, 363−392.
32. Cornilescu, G., Marquardt, J. L., Ottiger, M., and Bax, A. (1998) Validation of Protein Structure from Anisotropic Carbonyl Chemical Shifts in a Dilute Liquid Crystalline Phase, *J. Am. Chem. Soc. 120*, 6836−6837.
33. Bonvin, A. M., Boelens, R., and Kaptein, R. (2005) NMR analysis of protein interactions, *Curr. Opin. Chem. Biol. 9*, 501−508.
34. Williams, D. C., Jr., Cai, M., and Clore, G. M. (2004) Molecular basis for synergistic transcriptional activation by Oct1 and Sox2 revealed from the solution structure of the 42-kDa Oct1.Sox2.Hoxb1-DNA ternary transcription factor complex, *J. Biol. Chem. 279*, 1449−1457.
35. Kalodimos, C. G., Biris, N., Bonvin, A. M., Levandoski, M. M., Guennuegues, M., Boelens, R., and Kaptein, R. (2004) Structure and flexibility adaptation in nonspecific and specific protein-DNA complexes, *Science 305*, 386−389.
36. Hegde, R. S., Wang, A. F., Kim, S. S., and Schapira, M. (1998) Subunit rearrangement accompanies sequence-specific DNA binding by the bovine papillomavirus-1 E2 protein, *J. Mol. Biol. 276*, 797−808.
37. Zimmerman, J. M., and Maher, L. J., III. (2003) Solution measurement of DNA curvature in papillomavirus E2 binding sites, *Nucleic Acids Res. 31*, 5134−5139.
38. Zhang, Y., Xi, Z., Hegde, R. S., Shakked, Z., and Crothers, D. M. (2004) Predicting indirect readout effects in protein-DNA interactions, *Proc. Natl. Acad. Sci. U.S.A. 101*, 8337−8341.
39. Rohs, R., Sklenar, H., and Shakked, Z. (2005) Structural and energetic origins of sequence-specific DNA bending: Monte Carlo simulations of papillomavirus E2-DNA binding sites, *Structure 13*, 1499−1509.
40. Hawley-Nelson, P., Androphy, E. J., Lowy, D. R., and Schiller, J. T. (1988) The specific DNA recognition sequence of the bovine papillomavirus E2 protein is an E2-dependent enhancer, *EMBO J. 7*, 525−531.
41. Turnbull, W. B., and Daranas, A. H. (2003) On the value of c: can low affinity systems be studied by isothermal titration calorimetry? *J. Am. Chem. Soc. 125*, 14859−14866.
42. Poirot, O., Suhre, K., Abergel, C., O'Toole, E., and Notredame, C. (2004) 3DCoffee@igs: a web server for combining sequences and structures into a multiple sequence alignment, *Nucleic Acids Res. 32*, W37−40.

BI060123H