Journal of Biomolecular NMR **30:** 211–214, 2004.

Letter to Editor: Solution structure of the HPV-16 E2 DNA binding domain, a transcriptional regulator with a dimeric β-barrel fold

Alejandro D. Nadra[a,*], Tommaso Eliseo[b,*], Yu-Keung Mok[c], Fabio C.L. Almeida[d], Mark Bycroft[e], Maurizio Paci[f], Gonzalo de Prat-Gay[a,**] & Daniel O. Cicero[f,**]

[a]*Instituto de Investigaciones Bioquímicas-Fundación Leloir, Facultad de Ciencias Exactas y Naturales-UBA and CONICET, Patricias Argentinas 435 (1405) Buenos Aires, Argentina;* [b]*Dipartimento di Chimica, Universitá di Roma 'La Sapienza', Rome, Italy;* [c]*Department of Biochemistry, The Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong, Kowloon, P.R. China;* [d]*Centro Nacional de Ressonancia Magnetica Nuclear, Departamento de Bioquimica Medica, ICB, Universidade Federal do Rio de Janeiro, Cidade Universitaria, Rio de Janeiro 21914-590, Brazil;* [e]*MRC Centre of Protein Engineering and MRC Laboratory of Molecular Biology, Hills Road, Cambridge, CB2 2QH, U.K.;* [f]*Dipartimento di Scienze e Tecnologie Chimiche, Universita' di Roma 'Tor Vergata', via della Ricerca Scientifica 1 (00133), Rome, Italy*

Received 5 April 2004; Accepted 17 June 2004

Key words: dimer protein structure, E2 DNA binding domain, papillomavirus, RDC

Biological context

Papillomaviruses are DNA viruses that infect a wide range of vertebrate species. Of the over one hundred strains known so far, a few of them are strongly linked to cervical cancer in woman (Dell and Gaston, 2001). Among these, human papillomavirus type 16 (HPV-16) is the most frequently found high risk strain. Gene transcription in papillomavirus is regulated by the E2 protein (Hedge, 2002) which can operate either as an activator or a repressor of transcription. In addition, the E2 protein plays an accessory role in viral genome DNA replication. The protein consists of fairly well conserved N-terminal and C-terminal domains, linked by a poorly conserved flexible region termed the 'hinge'. The DNA binding activity is located in the C-terminal dimerization domain which bears a unique fold, the dimeric β-barrel, only shared by the Epstein–Barr replication origin binding protein EBNA1 (Hedge et al., 1992). The dimeric β-barrel fold consists of a central eight-stranded β-barrel, where each symmetric monomer contributes to one half of the barrel, with a major DNA recognition helix, a small helix and two loops connecting the β-strands. Crystal structures were reported for E2 DNA binding domains (E2C), including that of HPV-16 (Hedge and Androphy, 1998), HPV-18 (Kim et al., 2000; 55% sequence identity), HPV-31 (Bussiere et al., 1998; 65% sequence identity), BPV-1 (Hedge and Androphy, 1998; 38% sequence identity) and recently that of the non-oncogenic strain HPV-6 (Dell et al., 2003; 57% sequence identity). NMR structures of HPV-31 E2C (Liang et al., 1996) and BPV-1 E2C domains (Veeraraghavan et al., 1999) were also reported. They all share the same fold with minor modifications, where the most salient feature is the relative orientation of one monomer with respect to the other that was proposed to be related to DNA bending (Hegde, 2002). In the present work we present the solution structure of the dimeric 81 residues HPV-16 DNA binding domain (HPV-16 E2C) by multidimenstional NMR, emphasizing the use of residual dipolar couplings (RDCs) on structure determination.

Methods and results

Uniformly ^{15}N-labeled and uniformly ^{15}N-^{13}C-labeled recombinant proteins were expressed and purified as previously described (Mok et al., 1996).

NMR experiments were performed at 30 °C on Broker Avance700 and Avance400 spectrometers equipped with triple resonance probes incorporating self-shielded gradient coils. The concentration of the ^{15}N and ^{15}N-^{13}C labelled protein in buffer solution (50 mM sodium phosphate, 5 mM DTT, pH 6.5) was 1.8 mM and 0.9 mM, respectively. The NMR data were processed on Silicon Graphics workstations using NMRPipe and analyzed using NMRView. Structure calculation and refinement were performed with a modified version of XPLOR 3.85 (kindly supplied by M.G. Clore) that includes RDC refinement (Clore et al., 1998). Nearly all backbone ^{15}N, ^{13}C and ^{1}H

*Both authors contributed equally to this work.
**To whom correspondence should be addressed. E-mails: cicero@scienze.uniroma2.it; gpratgay@leloir.org.ar

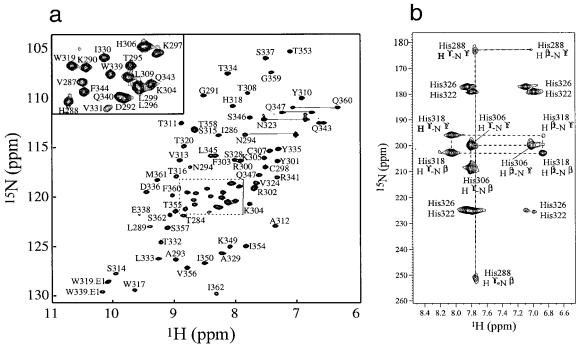

Figure 1. (a) 700 MHz ^{1}H-^{15}N HSQC spectrum of HPV-16 E2C in 50 mM sodium phosphate, 5 mM DTT, 0.05% sodium azide and 95% H_2O, 5% D_2O at pH 6.5, collected at 303 K, reporting backbone resonance assignments (labelled peaks). The offset shows an enlarged view of the central crowded region of the spectrum. (b) 400 MHz histidine ^{1}H-^{15}N long-range HMQC spectrum of HPV-16 E2C, reporting resonance assignments for $N^{\delta 1}$ and $N^{\varepsilon 2}$ crosspeaks of the five histidine residues.

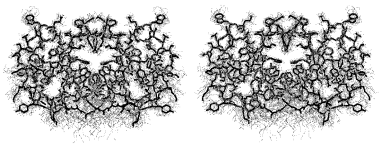

Figure 2. Stereoview of the superimposition of the 20 lowest-energy structures calculated for HPV 16 E2C dimer. The average structure is represented as black cylinder model.

resonances were sequentially assigned using standard double- and triple-resonance heteronuclear spectra. Amide signals 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 comprising residues 320–327. Figure 1a shows the ^{1}H-^{15}N correlation spectrum with the observed peaks assigned. $N^{\delta 1}$ and $N^{\varepsilon 2}$ chemical shifts of the five histidine residues were also assigned by performing a long-range ^{1}H-^{15}N HMQC experiment (Van Dijk, 1992), as shown in Figure 1b. ^{1}H and ^{13}C resonances were assigned for nearly all side chains from ^{15}N-separated TOCSY (25 ms spin locking time), HCCH-COSY and HCCH-TOCSY (16 and 24 ms spin locking time) spectra analysis. Stereospecific assignments of valine and leucine residues were obtained by means of a biosynthetic approach (Neri et al., 1989). Stereospecific assignment of other protons was achieved, where possible, from the inspection of the NOE patterns and the analysis of structures generated by the iterative structure calculation process. Chemical shifts for HPV-16 E2C were deposited in the BioMagResBank database, accession number 5952.

Approximate interproton distances were derived from 3D ^{15}N-edited NOESY and ^{13}C-edited NOESY using a mixing time of 100 ms. Only NOEs classified as unambiguous intramonomer and unambiguous intermonomer distances on the basis of a close inspection of monomeric and dimeric structures generated in the course of the iterative calculation process were used in the last stages of structure generation. A set of $^{3}J_{HNH\alpha}$ was determined by measuring cross peak intensities in a HNHA spectrum. In addition, backbone ϕ and ψ torsional angles were restrained in the later stages of the structure calculations, when it was possible to identify defined secondary structure elements. ^{1}H-^{15}N RDCs 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 molecular alignment, respectively (Sass et al., 2000). In these conditions, the values of D_A^{NH} and R were found to be 4.8 Hz and 0.13, respectively. 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. ^{15}N relaxation data are consistent with a dimeric state in solution for HPV-16 E2C (data not shown). The simulated annealing protocol for dimer calculation proposed by O'Donoghue et al. (1996) was slightly modified in order to incorporate RDCs (Clore et al., 1998), a direct potential for J scalar couplings and refinement against a database of Ramachandran plot dihedral angles. The ensemble of the twenty lowest-energy target function structures from a total of 200, was chosen to represent the solution structure of the dimer (Figure 2) and the corresponding statistics are summarized in Table 1. A cross-validation for the dipolar coupling refinement (Clore and Garrett, 1999) was performed by excluding different sets of 8 RDC from the calculation (4 per monomer). The resulting structures were used to back-calculate the missing values, with an average R_{free} of 34.5%. The twenty lowest-energy structures and the constraint lists were deposited in the Protein Data Bank, ID code 1R8P.

The homodimer displays the monomeric secondary structure topology (β_1–α_1–β_2–β_3–α_2–β_4) and the dimeric β-barrel fold shared by all the E2 protein structures solved so far. The 'recognition helices' ($\alpha 1$) are well defined and symmetrically disposed while the central (β_2-β_3 loop (residues 320–327) appears flexible and disordered in solution. When superimposed to its crystal structure (Figure 3a, PDB code: IBY9, obtained at 2.2 Å resolution (Hedge and An-

Table 1. Experimental restraints and structural statistics for the 20 lowest-energy structures[a].

Experimental restraints[b]	
Number of experimental restraints	2904
Distance restraints from NOEs	2398 (0.006 ± 0.001)[c]
Intraresidue	706
Interresidue short distance (<i+3)	625
Interresidue long range (>i+3)	1067
Hydrogen bond distance restraints[d]	76
Backbone dihedral angle restraints	244
Phi constraints	146
Psi constraints	98
HN-Hα scalar J coupling constants	64 (0.11 ± 0.03)[c]
H-N Residual Dipolar Coupling constants	56 (0.18 ± 0.03)[c]
Intermonomer distance restraints	66
Average number of restraints per residue	17.9
R_{free} (%)	34.5
Ramachandran analysis	
Residues in favoured 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 (relative to the average of the 20 structures)

	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]None of the structures exhibited distance violations >0.5 Å, dihedral angle violations >5° or RDC violations >2 Hz.
[b]Numbers refer to data per dimer.
[c]RMS deviations from experimental restraints.
[d]HN-O and N-O distances were constrained to 2.8 ± 0.5 Å and 3.4 ± 0.5 Å, respectively.

drophy, 1998)), the average NMR structure ofHPV-16 E2C shows a rmsd of 0.81 Å between the monomers and 1.15 Å for the entire dimers for backbone atoms excluding the β_2-β_3 loop. Some differences were detected with respect to the distances separating the two halves of the dimeric β-barrel. The average distance of the two $N^{\varepsilon 2}$ of His288 located in 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 water molecule was modeled as 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 $S04^{2-}$, or another metal ion fortuitously carried through the purification process (Hedge and Androphy, 1998). 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.

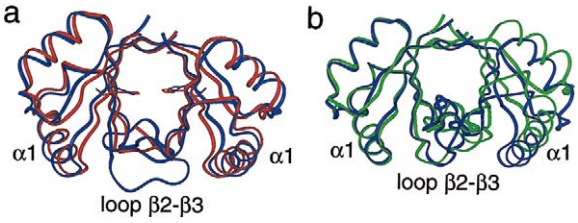

Figure 3. (a) Ribbon representation of the backbone superimposition of HPV-16 E2C structure obtained by NMR (blue) and by X-Ray (red). Residues 284–319, 329–362 of both subunits were aligned (rmsd 0.81 Å for the backbone atoms between monomers and 1.15 Å fot the entire dimer). The 'recognition helix' and β2-β3 loop are labelled. The His288 sidechain is depicted. (b) Superimposition of the left monomer of averaged structures calculated with (blue) and without (green) RDCs.

The chemical shift for the $N^{\delta 1}$ is around 251.3 ppm (Figure 1b) suggesting that there is no strong interaction involving the nitrogen lone pair, which would heavily shift the non-protonated nitrogen resonance upfield, from the reported mean value reported of 249.5 ppm (Bachovchin, 1986). 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).

In order to assess the impact of RDCs in the final structure, we have generated 200 structures excluding the 56 RDC constraints. The twenty lowest-energy structures show a good convergence (rmsd of 0.85 Å and 1.25 Å were obtained for backbone and heavy atoms with respect to the averaged structure excluding the β2-β3 loop). At the single'monomer level, the averaged structures calculated with and without RDCs show a reasonable agreement (rmsd 0.77 Å for backbone atoms, excluding the β2-β3 loop), but they deviate at the level of relative orientation of the two monomers in the dimeric structure (rmsd 1.50 Å) (Figure 3b). The averaged structure calculated without RDCs also presents higher rmsd values when compared with the crystal structure: 1.08 Å (monomer) and 1.53 Å (dimer) for backbone atoms superposition.

Discussion and conclusion

The overall conformation of the monomeric unit has a high structural homology in all viral strain proteins studied so far. In fact, 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 (Hedge, 2002). When one of the monomers is superimposed, the orientation of the second monomer is such that the five known structures can be divided into two families depending on the relative orientation of the two recognition helices, one family consisting of BPV-1 and HPV-18 and the other consisting of HPV-16, HPV-31 and HPV-6 E2C. This diversity has been proposed to be significant in defining the bending imposed to the DNA to form the complex (Hedge, 2002), pointing out the importance of an accurate determination of the quaternary structure. 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. RDCs 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. When dealing with dimers presenting a C2 symmetry like HPV-16 E2C, one of the axis of the alignment tensor must be parallel and the other two orthogonal to the molecular 2-fold symmetry axis (Bewley et al., 2000). This fact constitutes a stringent restraint for monomer orientation. The structure described in the present paper is the first of this family in which RDCs were used in the calculation.

Figure 3b shows the difference in the relative monomer orientation that is observed between structures calculated with and without RDCs. When comparing the distances between the two recognition helices, it can be noted that the structure calculated without RDCs place them circa 3 Å more distant. This helix shift makes a large difference in the predicted degree of bending for the bound DNA. The RDCs effect on structure calculation in the present case shows some analogies with DNA bending characterization in solution (Vermeulen et al., 2000). In both cases, NOEs alone are not able to accurately determine structural features that propagate significantly along the structure while the introduction of RDCs can improve the accuracy of the NMR structure.

The loop connecting β_2 and β_3 is found to be disordered and highly solvent accessible in our solution structure. 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 (Liang et al., 1996). 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 (Ferreiro et al., 2000), one cannot predict what would be the dynamic behavior of the β_2-β_3 loop in the complex.

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 (Hedge and Androphy, 1998). This fact, in turn, can be correlated with the deformability that different DNA binding sites show and can be one of the reasons for the different behavior that the two proteins show in terms of DNA discrimination (Ferreiro et al., 2000).

Acknowledgements

Alejandro D. Nadra holds a doctoral fellowship from CONICET, Argentina. This work was supported by a Wellcome Trust Grant (CRIG OIA U41 RG27994). Funding from Italian Ministry of University Research MIUR (PRIN and FIRB) and the technical assistance of Mr. Fabio Bertocchi are gratefully acknowledged.

References

Bachovchin, W.W. (1986) *Biochemistry*, **25**, 7751–7759.
Bewley, C.A. and Clore, G.M. (2000). *J. Am. Chem. Soc.*, **122**, 6009–6016.
Bussiere, D.E. et al. (1998) *Acta Crystallogr. Biol. Crystallogr.*, **54**, 1367–1376.
Clore, G.M. et al. (1998) *J. Magn. Reson.*, **131**, 159–162.
Clore, G.M. and Garrett, D.S. (1999) *J. Am. Chem. Soc.*, **121**, 9008–9012.
Dell, G. and Gaston, K. (2001) *Cell. Mol. Life Sci.*, **58**, 1923–1942.
Dell, G. et al. (2003) *J. Mol. Biol.*, **334**, 979–991.
Ferreiro, D.U. et al. (2000) *Biochemistry*, **39**, 14692–14701.
Hedge, R.S. (2002) *Annu. Rev. Biophys. Biomol. Struct.*, **31**, 343–360.
Hedge, R.S. and Androphy, E.J. (1998) *J. Mol. Biol.*, **284**, 1479–1489.
Hedge, R.S. et al. (1992) *Nature*, **359**, 505–512.
Kim, S.-S. et al. (2000) *J. Biol. Chem.*, **275**, 31245–31254.
Liang, H. et al. (1996) *Biochemistry*, **35**, 2095–2103.
Mok, Y.-K. et al. (1996) *Nat. Struct. Biol.*, **3**, 711–717.
Neri, D. et al. (1989) *Biochemistry*, **28**, 7510–7516.
O'Donoghue, S.I. et al. (1996) *J. Biomol. NMR*, **8**, 193–206.
Sass, H.-J. et al. (2000) *J. Biomol. NMR*, **18**, 303–309.
Van Dijk, A.A. et al. (1992) *Biochemistry*, **31**, 9063–9072.
Veeraraghavan, S. et al. (1999) *Biochemistry*, **38**, 16115–16124.
Vermeulen, A. et al. (2000) *J. Am. Chem. Soc.*, **122**, 9638–9647.