Journal of Biomolecular NMR (2005) 31: 129–141

^{1}H and ^{15}N resonance assignment, secondary structure and dynamic behaviour of the C-terminal domain of human papillomavirus oncoprotein E6

Yves Nominé[a], Sebastian Charbonnier[a], Laurent Miguet[b], Noelle Potier[b], Alain Van Dorsselaer[b], R. Andrew Atkinson[c], Gilles Travé[a,*] & Bruno Kieffer[c,*]
[a]*Laboratoire d'Immunotechnologie, CNRS UMR 7100, École Supérieure de Biotechnologie de Strasbourg, 67400 Illkirch, France;* [b]*Laboratoire de Spectrométrie de Masse Bio-Organique, CNRS UMR 7509, Faculté de Chimie, 67087 Strasbourg, France;* [c] *Laboratoire de RMN, CNRS UMR 7104, École Supérieure de Biotechnologie de Strasbourg, 67400 Illkirch, France*

Received 23 August 2004; Accepted 10 December 2004

Key words: C-terminal domain, dynamics, HPV oncoprotein E6, 3D NMR

Abstract

E6 is a viral oncoprotein implicated in cervical cancers, produced by human papillomaviruses (HPVs). E6 contains two putative zinc-binding domains of about 75 residues each. The difficulty in producing recombinant E6 has long hindered the obtention of structural data. Recently, we described the expression and purification of E6-C 4C/4S, a stable, folded mutant of the C-terminal domain of HPV16 E6. Here, we have produced ^{15}N-labelled samples of E6-C 4C/4S for structural studies by NMR. We have assigned most ^{1}H and ^{15}N resonances and identified the elements of secondary structure of the domain. The domain displays an original α/β topology with roughly equal proportions of α-helix and β-sheet. The PDZ-binding region of E6, located at the extreme C-terminus of the domain, is in a random conformation. Mass spectrometry demonstrated the presence of one zinc ion per protein molecule. Kinetics of replacement of zinc by cadmium followed by ^{1}H,^{15}N-HSQC experiments revealed specific frequency changes for the zinc-binding cysteines and their immediate neighbours. NMR spectra were affected by severe line-broadening effects which seriously hindered the assignment work. Investigation of these effects by ^{15}N relaxation experiments showed that they are due to heterogeneous dynamic behaviour with μs–ms time scale motions occurring in localised regions of the monomeric domain.

Introduction

E6 is one of two oncoproteins produced by 'high-risk' genital human papillomaviruses (HPVs) responsible for cervical cancers (zur Hausen, 1991). E6 interacts with more than 20 cellular proteins (Dell and Gaston, 2001; Du et al., 2002; Filippova et al., 2002; Iftner et al., 2002; Kumar et al., 2002), including tumour suppressor p53 (Werness et al., 1990), cellular ubiquitin ligase E6-AP (Huibregtse et al., 1993), and a family of proteins containing PDZ domains (Thomas et al., 2002). The interaction of E6 with its target proteins often leads to their cellular degradation. For example, E6-mediated degradation of p53 occurs via formation of a trimeric complex involving E6, p53 and the cellular ubiquitin ligase E6-AP (Scheffner et al., 1993). E6 is also a transcription modulator. In particular, it activates the

*To whom correspondence should be addressed. E-mail: trave, kieffer@esbs.u-strasbg.fr

transcription of the gene encoding the retrotranscriptase activity of human telomerase (Gewin and Galloway, 2001; Oh et al., 2001; Veldman et al., 2001). Finally, E6 is a DNA-binding protein which specifically recognises four-way DNA junctions (Ristriani et al., 2000).

Although the first DNA sequences encoding E6 were described in 1985 (Schwarz et al., 1985), no structural data are yet available for the wild-type E6 protein. Two forms of HPV16 E6 are produced *in vivo*, with lengths of 151 and 158 residues, depending on which methionine is used as a start codon (Androphy et al., 1987). Sequence alignments of E6 proteins from numerous HPV subtypes suggested the presence of two zinc-binding motifs (Cole and Danos, 1987) (Figure 1). However, biochemical studies that have been reported so far do not agree on the number of zinc ions bound to the protein. A first study reported that full-length E6 lacking nine C-terminal residues binds two zinc ions whereas the N-terminal domain of E6 binds only one zinc ion (Lipari et al., 2001). However, it has since been suggested that the biologically active form of full-length E6 may contain only one zinc ion per molecule (Degenkolbe et al., 2003).

Recently, we designed and produced a stable, folded and monomeric form of the C-terminal domain of HPV16 E6, referred to henceforth as E6-C 4C/4S, in which four non-conserved cysteines were mutated into serines (Nominé et al., 2003). This domain encompasses residues 80–151 of the shorter transcript of full-length HPV16 E6. It interacts specifically with four-way DNA junctions (Ristriani et al., 2001; Nominé et al., 2003) and contains many residues important for p53 degradation (Crook et al., 1991; Pim et al., 1994; Nakagawa et al., 1995; Liu et al., 1999; Ristriani et al., 2002). Furthermore, E6-C 4C/4S contains the C-terminal sequence responsible for PDZ-domain recognition (Pim et al., 2002). Here, we describe the production of ^{15}N-labelled E6-C 4C/4S, the near complete assignment of ^{1}H and ^{15}N resonances and the subsequent identification of the elements of secondary structure. We have established the stoichiometry of zinc binding and identified the four chelating residues. ^{15}N relaxation rates have been measured to characterise the heterogeneous nature of the backbone dynamics of E6-C 4C/4S.

Material and methods

M9 medium

1 l of M9 ^{15}N minimal medium contained 7.7 g $Na_2HPO_4{\cdot}2H_2O$, 3 g KH_2PO_4, 0.5 g NaCl, 0.5 g $^{15}NH_4Cl$, 8.3 mg $FeCl_3{\cdot}6H_2O$, 0.8 mg $ZnCl_2$, 0.1 mg $CuCl_2$, 0.1 mg $CoCl_2{\cdot}6\ H_2O$, 0.1 mg H_3BO_3, 11.1 mg $MnCl_2{\cdot}2H_2O$, 0.04% glucose, 120 mg $MgSO_4$, 36.6 mg $CaCl_2$, 1 mg biotin, 1 mg thiamine. Under these standard conditions, the zinc concentration was 6 μM. This concentration was varied from 0 to 200 μM, depending on the specific experiment.

Screening of conditions for optimal folding

The E6-C 4C/4S construct was subcloned within the pETM-10 vector allowing expression with a N-terminal HisTag. The DNA construct was electroporated into BL21 DE3 *E. coli* cells. About

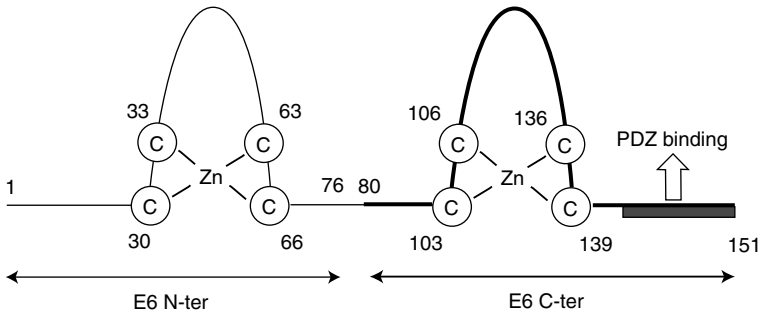

Figure 1. Schematic representation of the HPV E6 sequence with putative zinc binding sites deduced from the conservation of cysteine residues. The C-terminal domain encompasses residues 80–151 which are highlighted using a bold line.

1/200th of electroporated bacteria were transferred into M9 ^{15}N minimal medium containing 15 μg/ml kanamycin and were incubated for 3 days at 25 °C and 250 rpm. The pre-culture was diluted 40-fold in fresh medium and grown at 37 °C until OD (600 nm) reached 0.6–0.7. Cells were harvested by centrifugation at 2300 × *g* and 25 °C during 10 min and transferred to fresh medium containing five possible zinc concentrations (0, 6, 25, 100 and 200 μM). IPTG (0.5 mM) was added for induction and the cultures were grown further at three possible temperatures (37 °C, 27 °C and 22 °C) for respectively 3, 6 or 18 h. Samples of 1 ml of expression culture corresponding to each combination of zinc concentration and expression temperature were pelleted at 1800 × *g* in a benchtop centrifuge at 4 °C, re-suspended in 500 μl of buffer A (Tris-HCl 50 mM, NaCl 150 mM, DTT 2 mM, pH 6.8 containing 2.5 μg/ml DNase I, 2.5 μg/ml RNase I and eq. 1/100th tablet of complete EDTA-free protease inhibitor cocktail (Roche) in a 1.5 ml Eppendorf tube and sonicated with the 3 mM probe of a Vibracell 72412 sonicator (Bioblock Scientific) performing 2 runs of 15 s with a pulse frequency of 1 s pulse for 1 s pause and an amplitude of 6%. Fifty μl of these sonicated extracts were centrifuged at 16000 × *g* for 15 m. The supernatant was taken up in 50 μl sample buffer and the pellet in 50 μl sample buffer and 50 μl H_2O. 20 μl samples of pellet and supernatant were tested by Tris–Tricine PAGE and the ratio of soluble to insoluble protein was determined with Quantity One™ quantification software (BIORAD) and validated by eye.

Production of ^{15}N-labelled E6-C 4C/4S

BL21 DE3 *E. coli* cells freshly electroporated with the MBP-E6-C 4C/4S expression construct (Nominé et al., 2003) were transferred directly into M9 ^{15}N minimal medium containing 15 μg/ml kanamycin and 6 μM $ZnSO_4$. The pre-culture and culture protocols were identical to those described for the screening. Cells from 2 l of expression culture were harvested by centrifugation at 2300 × *g* and 4 °C during 20 min. To minimise oxidation problems, all buffers were degassed using a water vacuum pump and bubbled extensively with argon. The pellet was re-suspended in 150 ml of buffer A containing 5% glycerol, 1 μg/ml DNase I, 1 μg/ml RNase I and three tablets of complete EDTA-free protease inhibitor cocktail. Cells were broken by sonication on ice, then centrifuged at 18000 × *g* at 6 °C for 30 min. The supernatant was filtered (Millipore 0.22 μm) and loaded onto a 80 ml column of amylose resin (New England Biolabs) pre-equilibrated with buffer A. The column was washed stepwise with 1 volume, 3 volumes and 12 volumes of buffer A containing respectively 100%, 50%, and 12% of initial anti-protease concentration. Remarkably, MBP-E6-C 4C/4S eluted as a relatively pure form by leaking from the column in the late washing steps. The protein was incubated for 12–24 h at 6 °C with an appropriate concentration of recombinant TEV protease until full separation of E6-C 4C/4S from the MBP tag was achieved. The TEV cleavage site results in two additional residues (Gly–Ala) on the N-terminus of the construct, prior to the methionine residue. The digestion product was concentrated to a 1 ml volume using a 5 kDa-centriprep concentration device (Amicon) and applied on a Hiload 16/60 superdex 75 gel-filtration column (Amersham Biosciences) pre-equilibrated with buffer A. Pure monomeric E6-C 4C/4S peptide eluted as a single peak at the volume expected for a monomer according to the calibration of the column. The buffer was adjusted to 20 mM Tris-HCl, 50 mM NaCl, 1 mM DTT, pH 6.8 by performing dilution/concentration steps using a 15 ml Ultrafree Biomax 5K NMWL Membrane (Millipore). The final concentration was raised to 1.0–1.2 mM.

Mass spectrometry

Analyses were performed on a ESI-TOF mass spectrometer (LCT, Micromass U.K.). The sample buffer was exchanged by dilution/concentration steps against a 50 mM ammonium acetate buffer at pH 6.8, then diluted to 10 pmol/μl and continuously infused into the ESI ion source at a flow rate of 8 μl/min through a Harvard syringe pump. Parameters were optimised such that non-covalent interactions survive the ionisation desorption process. The accelerating voltage was set to 80 V and the pressure to 3 mbar. ESI-MS data were acquired in the positive ion mode, in the range 500–4000 m/z. The instrument was calibrated using multiply charged ions produced by a separate injection of horse heart myoglobin

diluted to 2 pmol/μl in a 1:1 water/acetonitrile mixture (v/v), acidified with 1% of formic acid.

NMR spectroscopy

Homonuclear ^{1}H spectra were acquired at 5 °C, 15 °C and 25 °C on a Bruker DRX600 spectrometer whereas heteronuclear experiments were recorded at 15 °C on a Bruker DRX500 spectrometer equipped with a z-gradient triple resonance cryoprobe. In all spectra, the water signal was suppressed using the WATERGATE sequence (Piotto et al., 1992). Proton assignments were based on homonuclear and ^{15}N-edited NOESY and TOCSY spectra. Heteronuclear ^{15}N-edited spectra were recorded with mixing times of 200 and 60 ms for the NOESY and TOCSY spectra, respectively. Data were processed using either NMRPipe (Delaglio et al., 1995) or XWINNMR (Bruker) and analysed with XEASY (Bartels et al., 1995). Dihedral ϕ angles were obtained from $^{3}J_{NH-H\alpha}$ scalar coupling constants determined from the ratio of diagonal to cross-peak intensities in a HNHA spectrum (Vuister and Bax, 1993). The ^{15}N resonance assignments of His118 and His126 rings and subsequently their protonation states were determined using a long-range ^{1}H,^{15}N-HSQC spectrum (Pelton et al., 1993) in which the ^{15}N spectral width was set to 200 ppm. Two ^{15}N relaxation datasets were recorded for sample concentrations of 1.2 and 0.3 mM. Longitudinal (R_1) and transverse (R_2) ^{15}N relaxation rates together with heteronuclear ^{1}H-^{15}N NOE values were measured using proton-detected heteronuclear experiments (Kay et al., 1989). ^{15}N R_1 rates were obtained from 9 experiments with relaxation delays of 8, 40, 101, 243, 405, 568, 810, 1216 and 1824 ms and ^{15}N R_2 rates from 12 experiments with delays of 16, 32, 48, 72, 96, 119, 143, 183, 223, 263, 303 and 350 ms. During the R_2 relaxation delay, ^{15}N 180° pulses with a field strength of 2.0 kHz were applied every 1.2 ms ($\nu_{CPMG} = 0.83$ kHz) and ^{1}H 180° pulses were applied every 7.2 ms. For error estimation, the spectra with relaxation delays of 101 ms (^{15}N R_1) and 72 ms (^{15}N R_2) were recorded twice. In these sets of experiments, the recycle delay was 2 s. In ^{15}N, R_1 and heteronuclear ^{1}H-^{15}N NOE experiments, the protons were saturated using a train of 180° pulses separated by 4 ms. In the heteronuclear ^{1}H-^{15}N NOE experiment, the proton magnetisation was saturated during 4 s to achieve the steady state. Data sets were typically recorded as 100 × 1024 complex points with spectral widths of 6.6 and 1.75 kHz in F1 and F2, respectively. The total acquisition time for one complete set of relaxation data measurements (R_1, R_2, ^{1}H-^{15}N NOE) was 2 days. Peak intensities were measured using Felix 2.10 (Accelrys Inc.) and exponential decay rates were obtained from a non-linear two-parameter least-squares fit using the Levenberg–Marquardt algorithm implemented in Matlab (The MathWorks Inc.). A set of 200 Monte Carlo simulations were used to estimate the statistical error on the parameters. Residues with ^{15}N resonances affected by exchange contributions were identified by recording relaxation-compensated CPMG dispersion experiments with a constant transverse relaxation time of 40 ms (Loria et al., 1999) but different values of the spacing delay (τ_{CP}) in the spin–echo pulse train. The extreme τ_{CP} delays were set to 20 and 0.2 ms corresponding to CPMG field strengths of 50 and 5000 Hz, respectively. The ^{15}N refocussing pulse length was set to 80 μs. Each 2D spectrum was recorded as 128 × 1024 complex points matrix with 24 scans per FID and a relaxation delay of 1.5 s resulting in a total acquisition time of 3 h per dispersion point. The presence of chemical exchange on the amide proton was derived from the ratios of peak intensities measured on a ^{1}H,^{15}N-HSQC experiment and a ^{1}H,^{15}N-CPMG-HSQC experiment in which both $^{1}J_{HN}$ transfer delays are replaced by a train of 16 refocussing pulses on both ^{15}N and ^{1}H nuclei (Mulder et al., 1996). Translational diffusion experiments were recorded using the LED pulse scheme as described previously (Nominé et al., 2003). To identify the zinc-binding residues, we substituted the initially bound zinc by cadmium by adding to the sample a solution of cadmium–EDTA, up to a final ratio of 3:1 (Cd:Zn), and following the displacement of amide resonances on ^{1}H,^{15}N-HSQC spectra.

Results

Production of ^{15}N-labelled E6-C 4C/4S

Preliminary studies allowed us to express the C-terminal domain of E6 as a folded, monomeric

peptide (Nominé et al., 2003). The expression strategy was based on a fusion of the sequence of the C-terminal domain of E6 with maltose-binding protein (MBP) together with a systematic conversion of non-conserved cysteine residues into serines. Sequential assignment using homonuclear methods proved nonetheless to be extremely difficult due to spectral overlap and heterogeneous line-broadening. Initial attempts to produce a ^{15}N-labelled protein using M9 ^{15}N minimal media produced insoluble E6-C 4C/4S after proteolytic removal of the MBP carrier. We have shown previously that expression under improper conditions can lead to misfolded proteins, solubilised by the MBP carrier, forming high molecular weight soluble aggregates (Nominé et al., 2001). To avoid this solubilisation artefact, we performed a screen of expression conditions using a N-terminal HisTag fusion of E6-C 4C/4S. Three temperatures (22 °C, 27 °C and 37 °C) and five zinc concentrations (0, 6, 25, 100 and 200 μM) were probed and the solubility of the expressed material was analysed by pellet/supernatant assays. At all temperatures, we observed minimal solubility (30% of expressed material) in the absence of zinc, maximal solubility at low to intermediate zinc concentrations (6–25 μM), and lower solubility (40% of expressed material) at higher zinc concentrations (100–200 μM). The highest solubility (up to 70% of expressed material) was obtained for expression at 22 °C in the presence of 6 μM zinc. A minimum concentration of zinc in the expression medium is therefore required for proper folding of the domain, but an excess is detrimental and promotes aggregation, and this trend is reinforced upon raising the expression temperature. These conditions were subsequently applied for large scale production of the E6-C 4C/4S using the MBP version of the construct which provides higher purity and folding homogeneity as compared to the His-tagged version. Production of soluble ^{15}N-labelled E6C 4C/4S was successfully performed by expressing the His-MBP-E6C 4C/4S fusion at 22 °C using 2 l of M9 ^{15}N minimal medium with a zinc content of 6 μM. Purification and concentration of the product led to a ca. 1 mM protein sample that was suitable for NMR analysis.

^{1}H and ^{15}N resonance assignments, secondary structure and topology of E6-C 4C/4S

The ^{1}H,^{15}N-HSQC spectrum of ^{15}N-labelled E6-C 4C/4S exhibits good dispersion of the ^{1}H and ^{15}N resonances, as expected for a folded protein (Figure 2). However, both the line-shapes and the intensities of the cross-peaks display great variability, indicative of differential line-broadening along the peptide chain. This feature is striking when one compares, for example, the line-shapes of Gly130 and Gly134, the former displaying a much larger line-width in the ^{15}N dimension. Since this behaviour may originate from dynamic oligomerisation of the protein (Pfuhl et al., 1999), the oligomeric state was probed by measuring the translational diffusion coefficient (D_t) at two concentrations (0.3 and 1.2 mM) using a modified LED experiment (Dingley et al., 1995). The measured values were compatible with that expected for a monomer and were independent of the concentration, indicating that the observed line-broadening is due to μs–ms time-scale motions within the peptide chain (data not shown). Despite this line-broadening phenomenon, the quality of the ^{15}N-edited NOESY and TOCSY spectra (Figure 2) allowed near-complete assignment of the backbone and side-chain ^{1}H and ^{15}N nuclei. No spin system could be found for Thr86, probably due to severe line-broadening. The resonances of Asn127 show significant deviation from expected 'random coil' values (Hα: 4.28; Hβ: 1.83/-0.30; Hδ: 5.75/3.86 and Nδ: 107.0 ppm) for a reason which remains to be elucidated. To assign the histidine side-chains unambiguously we used long-range ^{1}H,^{15}N-HSQC spectra (Pelton et al., 1993). Analysis of ^{15}Nδ and ^{15}Nε chemical shifts indicated that the Nδ atom of His118 is protonated while His126 is in equilibrium between a fully and Nδ-protonated form, giving an estimated pKa value of 6.8 for this residue. ^{1}H and ^{15}N chemical shifts have been deposited in the BioMagResBank (Accession code 6407).

NOEs observed in homonuclear and ^{15}N-edited NOESY experiments allow the definition of three helical segments (Figure 3). For these segments, characteristic $H\alpha_i$-HN_{i+3} NOEs are supported by both low values of $^{3}J_{HN\text{-}H\alpha}$ coupling constants, deduced from the HNHA spectrum,

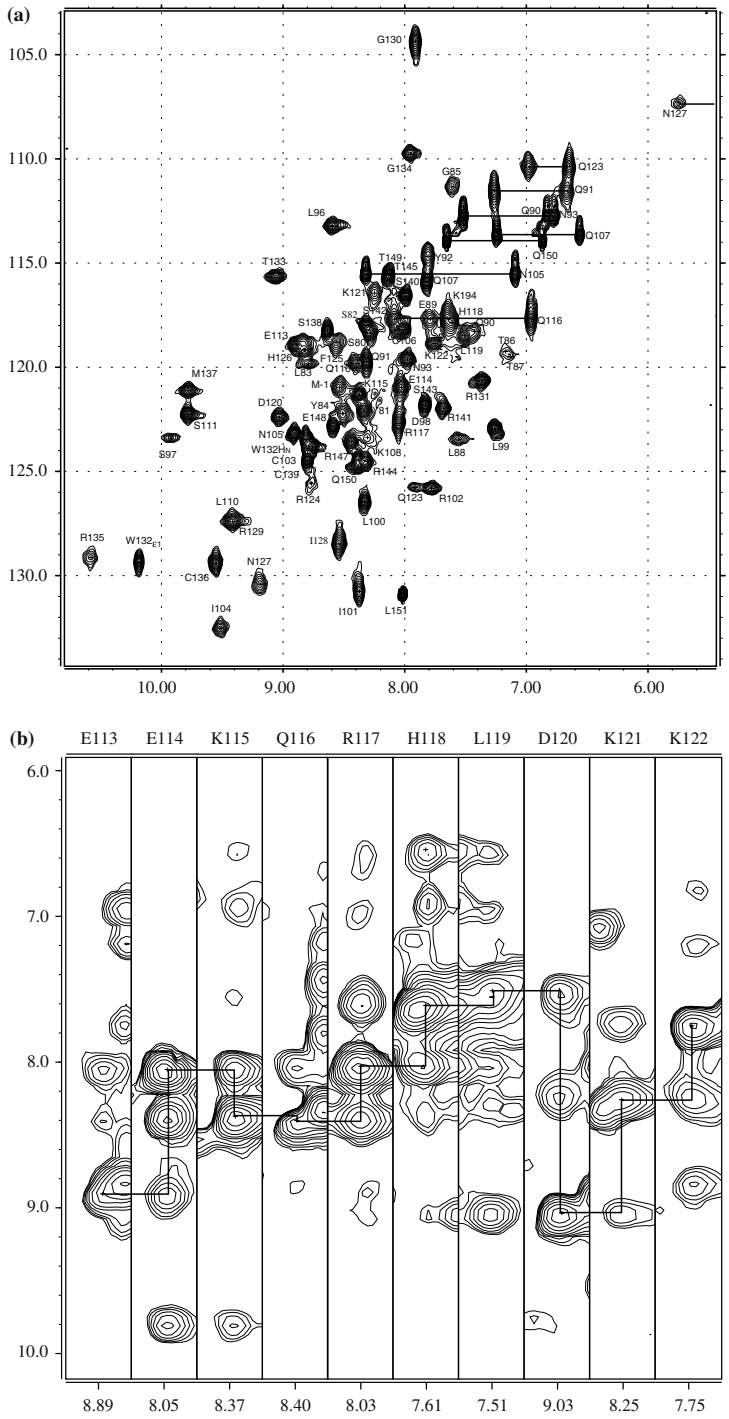

Figure 2. Representative spectra of the ^{15}N-labelled E6-C 4C/4S domain recorded at pH 6.8 and 15 °C. (a) 2D ^{1}H,^{15}N-HSQC spectrum. Cross-peaks are numbered according to the full-length HPV16 E6 (151-residue form). Side-chain NH_2 resonances of asparagine and glutamine residues are connected by horizontal bars, excepted for Asn127, for which only one proton appears in the region of the spectrum shown. (b) Regions of 3D ^{15}N-edited NOESY-HSQC spectrum (mixing time 200 ms) showing sequential connectivities for residues between Glu113 and Lys122.

and the chemical shift index (Wishart et al, 1992). In addition, three short β-strands were identified, encompassing residues 81–83, 125–127 and 131–133. Several inter-strand NOEs between amide protons belonging to residues Leu83–Phe125, His126–Thr133 and Ile128–Arg131 were observed together with NOEs between Hα protons of residues Asn127–Trp132 and Ser82–His126, indicating the presence of an anti-parallel β-sheet structure. This definition of secondary structure elements is in good agreement with circular dichroism results which estimated the α-helical and β-sheet content to be 22 ± 1% and 22 ± 3%, respectively (Nominé et al., 2003). 'Random coil' values for Hα chemical shifts and a lack of medium-range NOEs suggest that the last 10 residues are unfolded.

Zinc binding properties of E6-C 4C/4S

The zinc content of the E6-C 4C/4S domain was probed by ESI mass spectrometry. The difference in mass of 63.7 Da between experiments carried out under native and denaturing conditions (Table 1) is consistent with the binding of a single zinc ion. To further characterise the zinc-binding properties of E6-C 4C/4S and to identify those residues directly involved in zinc binding, a zinc/cadmium substitution experiment was monitored by ^{1}H,^{15}N-HSQC measurements. Chemical shift perturbations upon replacement of zinc by cadmium were localised in two regions encompassing residues Cys103–Gln107 and Cys136–Arg141, identifying cysteines 103, 106, 136 and 139 as the residues involved in zinc coordination. The perturbation was found to be systematically higher for the amide group of the second cysteine of each pair, which could be explained by with the involvement of this amide proton in hydrogen bond with the sulphur atom of the preceding cysteine residue (Summers et al., 1990; Allen et al., 1997). The possibility of zinc coordination by the two histidine residues, His118 and His126, can be discounted by the lack of chemical shift perturbation of both backbone and side-chain ^{1}H and ^{15}N nuclei upon cadmium substitution.

The identification of the zinc-binding residues together with the assignment of inter β-strand NOEs provided enough constraints to establish the topology of E6-C 4C/4S (Figure 4). The three short β-strands form an anti-parallel β-sheet which, together with two α-helices between strands β1 and β2, define the limits of two 5-residue loops (L1 and L3) and a longer, 18-residue loop (L2) encompassing the first two zinc-coordinating cysteine residues. The C-terminal α-helix contains the third and fourth cysteine residues involved in zinc coordination and precedes an unfolded C-terminal tail of 10 residues.

Dynamic behaviour of E6-C 4C/4S

Structural analysis of E6-C 4C/4S was severely hindered by the presence of broadened resonances in NMR spectra. We therefore investigated the dynamic behaviour of E6-C 4C/4S via a set of ^{15}N relaxation experiments (Figure 5), with the aim of identifying possible sources of increased transverse relaxation rates. The profile of ^{1}H-^{15}N NOE values along the peptide sequence indicates that the peptide backbone is fairly rigid on the ps–ns time-scale from residue Ser82 to Ser140, with the exception of Lys108, localised within loop L2. Reduced values of the heteronuclear NOE for the 10 C-terminal residues reflect dynamic disorder for this portion of the peptide chain. The profile of ^{15}N transverse relaxation rates exhibits much larger variations along the sequence, indicative of conformational exchange contributions to these rates. This behaviour made the extraction of a global correlation time from the relaxation data rather difficult and hinders further interpretation of these rates in terms of molecular motions. In order to identify those residues not affected by exchange contributions and to further characterise these intermediate time-scale motions, dispersion experiments were conducted. A fast and convenient way of mapping exchange contributions to the transverse relaxation rates of amide protons is to compare peak intensities measured in ^{1}H,^{15}N-HSQC and ^{1}H, ^{15}N-CPMG-HSQC spectra (van Tilborg et al., 2000; Palmer et al., 2001). The profile of the ratio of peak intensities shows a steady increase followed by a decrease for residues located in loops L1 and L3 (Figure 6). Other residues also display increased values such as Ser97 and Ile104, the latter located between the first two zinc-coordinating cysteine residues, and Leu110, sandwiched between two proline residues. From this profile, it is readily apparent that the line-broadening phenomena affecting the

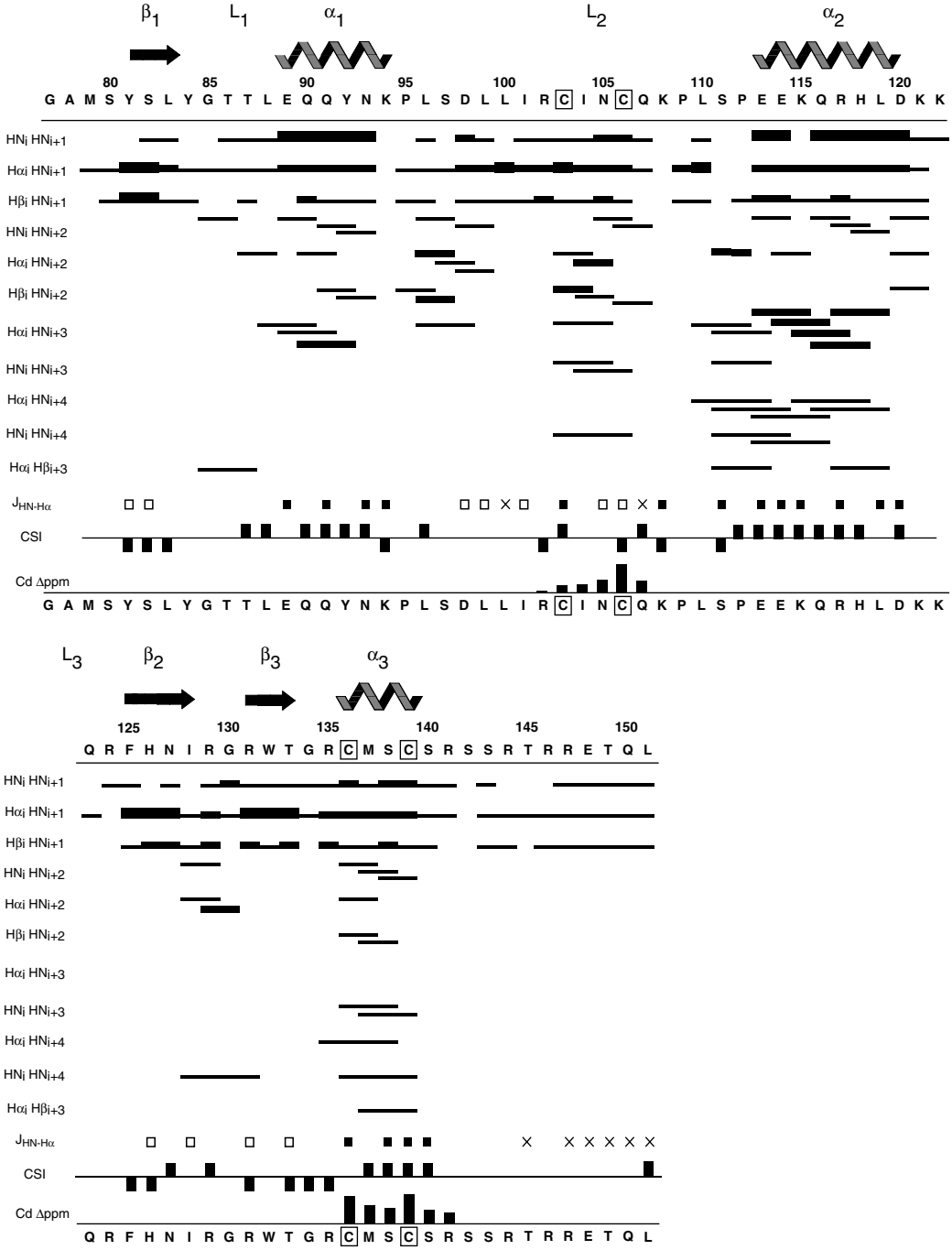

Figure 3. Secondary structure of E6-C 4C/4S deduced from short- and medium-range NOEs, $^{3}J_{HN-H\alpha}$ coupling constants and Hα chemical shift index (Wishart et al., 1992). The first two residues (77–78) originate from the TEV protease cleavage site. The presence of a NOE cross peak is indicated by a line with a thickness proportional to its intensity. $^{3}J_{HN-H\alpha}$ coupling constants greater than 6.5 Hz and lower than 7.5 Hz are indicated by closed and open squares, respectively. The lack of coupling information (due to overlapped or unobserved peaks) is indicated by a cross. CSI represents the H_α chemical shift index where 'random coil' values are taken from (Merutka et al., 1995). A vertical bar indicates an absolute difference between experimental and 'random coil' values greater than 0.2 ppm. The line Cd ?ppm reports the composite chemical shift differences for ^{1}H-^{15}N cross-peaks recorded for the zinc- and the cadmium- loaded proteins calculated using: $\Delta ppm = [\ 5(\delta^{H}(Cd^{2+}) - \delta^{H}(Zn^{2+}))^2 + (\delta^{N}(Cd^{2+}) - \delta^{N}(Zn^{2+}))^2\ ]^{1/2}$.

Table 1. Mass deduced from ESI mass spectrometry experiments The expected mass deduced from the sequence of E6-C 4C/4S is 9010.1 Da (assuming uniform ^{15}N-labelling)

	E6-C 4C/4S (Zn^{2+})	E6-C 4C/4S (Cd^{2+})
native	9073.4 ± 0.6 Da	9120.0 ± 1.0 Da
denaturated	9009.7 ± 0.1 Da	9010.1 ± 0.2 Da

NMR spectra of E6-C 4C/4S probably originate from several different motions. The profile of peak intensity ratios obtained from the extreme values of the delay between ^{15}N refocussing pulses ($\tau_{cp} = 20$ ms and $\tau_{cp} = 0.2$ ms) in a series of ^{15}N relaxation dispersion experiments (Figure 6) is very similar to that obtained from peak intensities measured in ^{1}H,^{15}N-HSQC and ^{1}H, ^{15}N-CPMG-HSQC spectra. Most residues affected by exchange contributions are localised in loops L1 and L3, β-strands β1 and β2 and α-helices α1 and α2 (see Figure 4). An analysis of these relaxation profiles allowed the identification of a small number of residues (97–103 excluding 101) whose ^{15}N transverse relaxation rates are not affected by exchange contributions. The R2/R1 ratio yielded estimates of the global rotational correlation time of 8.7 ± 1.5 ns and 7.8 ± 1.7 ns at high (1.2 mM) and low (0.3 mM) concentrations, respectively, compatible with the domain behaving as a monomer (expected correlation time: 7.6 ns), and in agreement with translational diffusion data obtained for unlabelled E6-C 4C/4S.

Discussion

Over the past 20 years, a large number of publications has established the key role played by oncoprotein E6 in HPV-induced carcinogenesis (Rapp and Chen, 1998; Finzer et al., 2002). E6 activity is based on its ability to target numerous cellular targets (Mantovani and Banks, 2001). Thus E6 has been identified as an important subject for structural studies aimed at designing potential inhibitors of its various interactions. Such studies have long been hindered however, by the difficulty of producing well-behaved samples of either full-length E6 or its subdomains. In a recent publication, we reported the production of the C-terminal domain of E6 (E6-C) in a pure, monomeric form amenable to NMR studies (Nominé et al., 2003). Analysis of preliminary NMR spectra showed that several resonances of this domain were affected by significant line-broadening so that the production of an ^{15}N-labelled sample was required. However, the use of M9 ^{15}N minimal medium in place of LB

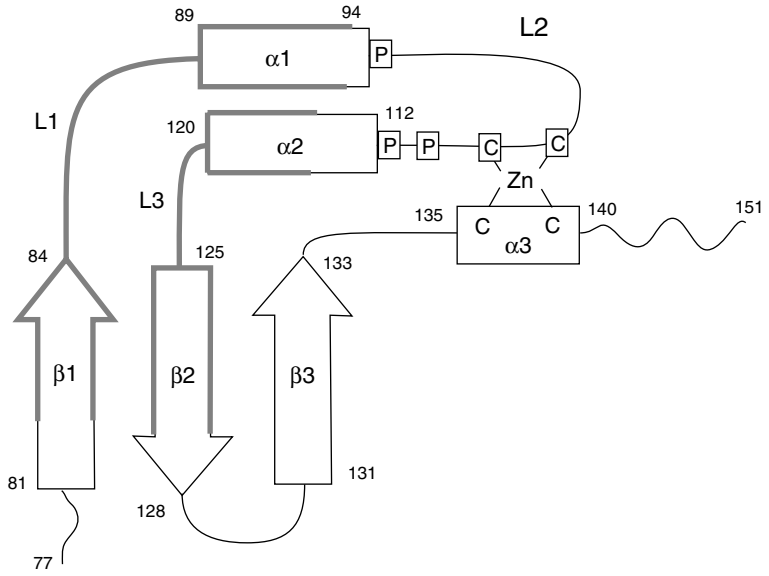

Figure 4. Topology of E6-C 4C/4S domain. Elements of secondary structures are shown as arrows and rectangles for β-strands and α-helices respectively. Regions of the peptide affected by chemical exchange are highlighted using a bold grey line.

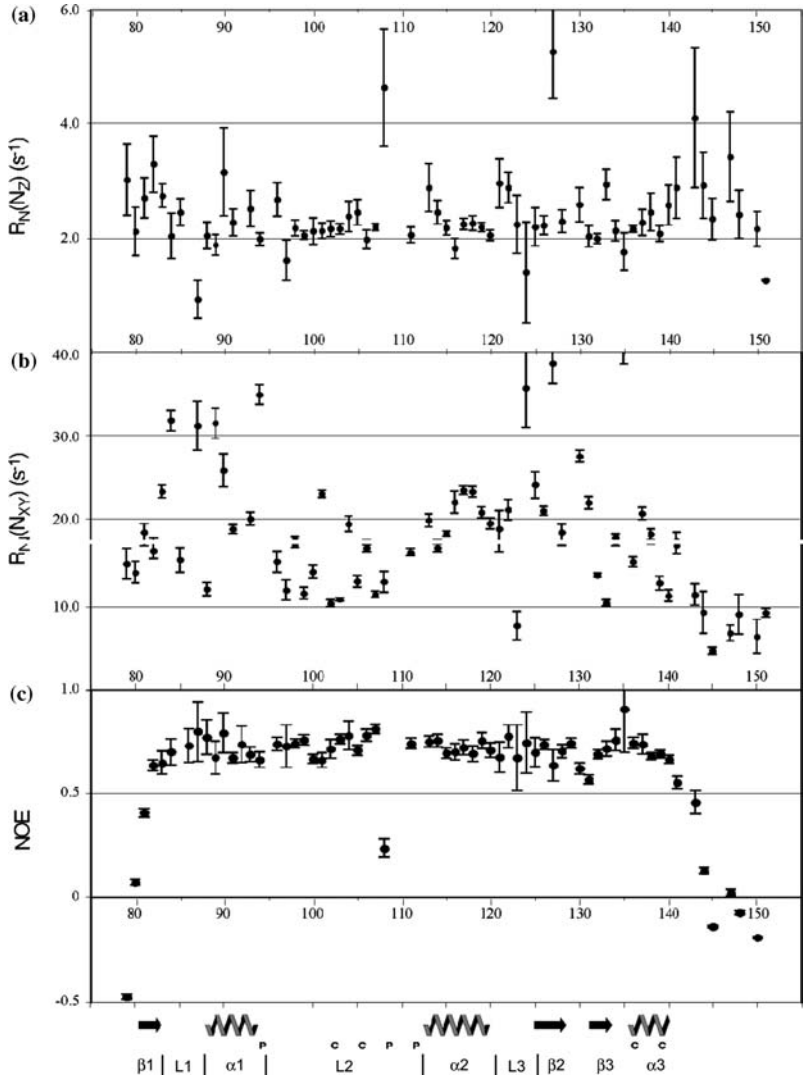

Figure 5. Longitudinal (a) and transverse (b) ^{15}N relaxation rates and steady state ^{1}H-^{15}N NOE values (c) for E6-C 4C/4S measured at 15 °C.

led unexpectedly to the production of misfolded E6-C 4C/4S samples. Only by careful screening of the expression conditions could the origin of this behaviour be identified: optimal solubility required fine tuning of the zinc content within the M9 ^{15}N expression medium, and a lower temperature of expression. This optimisation of zinc content and expression temperature by screening proved critical for the obtention of folded ^{15}N-labelled E6-C 4C/4S and subsequent assignment of resonances of the domain.

The strict requirement for zinc in the expression medium provided the first evidence of the zinc binding to E6-C 4C/4S. Mass spectrometry showed that E6-C 4C/4S indeed contains a single zinc atom. Finally, zinc–cadmium substitution experiments identified the four zinc-binding cysteines of the domain. These data provide sound experimental confirmation of the earlier prediction of a zinc-binding site in this region of E6 based on sequence analysis of all HPV strains (Cole and Danos, 1987). The analysis of the secondary structure and the topology of the E6-C domain (Figure 4) suggests a structural role for the zinc ion which contributes to the stabilisation of a short C-terminal helix (α3) by anchoring it

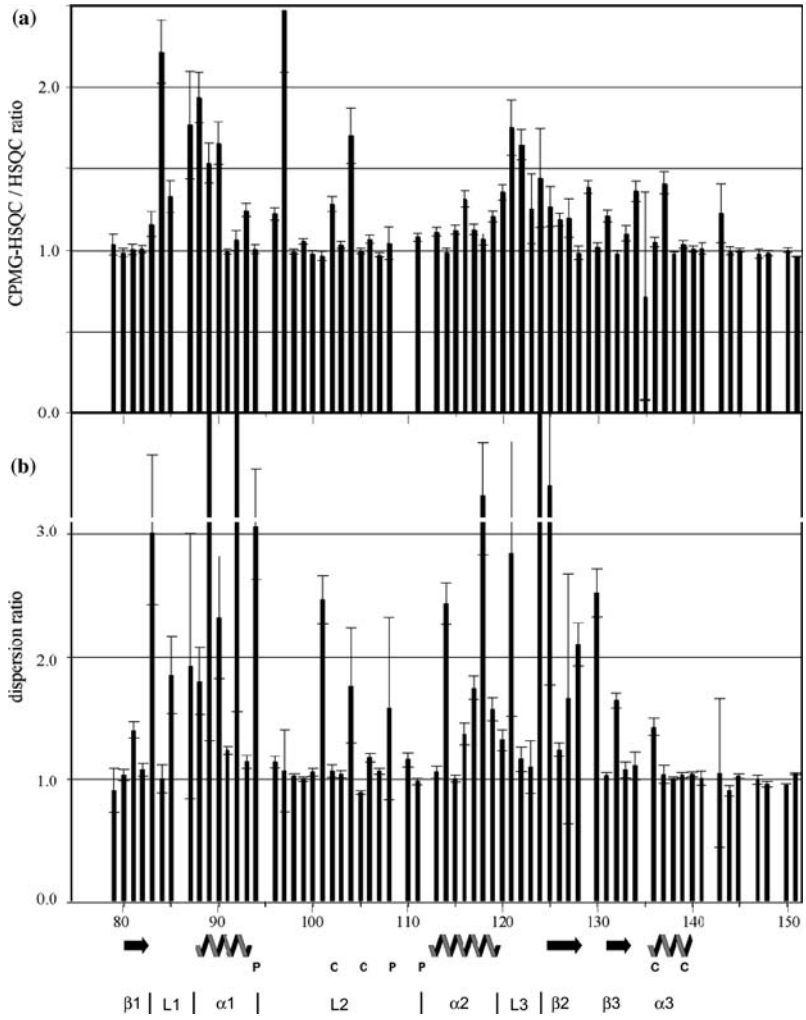

Figure 6. Identification of intermediate exchange contributions to transverse relaxation. (a) HSQC-CPMG/HSQC ratio as a function of sequence. (b) Ratio between relaxation-compensated CPMG dispersion experiments recorded with inter-pulse delays of 20 and 0.2 ms. In both cases, statistical errors were estimated from the standard deviation of the noise.

to the large L2 loop connecting helices α1 and α2.

Analysis of medium range NOEs (Figure 3) shows that the minimal folding unit begins with Tyr81 initiating the first β strand and ends at Arg141. These boundaries for the C-terminal domain of E6 are consistent with the domain definition based on results from mild proteolysis experiments (Lipari et al., 2001; Nominé et al., 2003). The strong sequence similarity between the N- and C-terminal domains suggests that the N-terminal domain should adopt the same topology. Both the lack of medium-range NOE and ^{15}N relaxation studies show that the last 10 C-terminal residues (Ser142–Leu151) are not structured and rather flexible on the ps–ns timescale. This region encompasses the PDZ domain-binding site consisting of the last four residues (sequence ETQL) and is likely to fold upon binding to its partner (Cowburn, 1997; Hung and Sheng, 2002).

A specific feature of E6-C 4C/4S is the presence of motions on the μs–ms time scale in several regions of the peptide (Figure 6). The L2 loop contains a number of isolated residues affected by exchange broadening which may arise from cis–trans isomerisation of the peptide bonds of proline residues in this loop. However, most residues affected by exchange broadening are clustered within two continuous sequence

elements (β1–L1–α1) and (α2–L3–β2) (see grey elements in Figure 4). This suggests that these two regions sense a common conformational exchange phenomenon which may indicate a spatial proximity of these two regions within the folded domain. A precise characterisation of time-scales (τ_{ex}), frequency differences and the populations of the exchanging states would require an extensive set of relaxation dispersion experiments (Palmer et al., 2001). Preliminary attempts to fit the relaxation dispersion dataset recorded at 500 MHz with the general equation of a two-state exchange system (Carver and Richards, 1972) led to an estimated time range of 0.2–3 ms and unequal populations of the states (data not shown). This indicated that the equilibrium is on the slow-intermediate time-scale ($\tau_{ex}.\Delta\nu > 1$) (where $\Delta\nu$ is the chemical shift difference between the two states) and is consistent with our repeated observations that increased temperature induced the concerted disappearance of a subset of peaks in NMR spectra of E6-C 4C/4S. Indeed, an increase in temperature is expected to reduce the value of τ_{ex} according to Arrhenius' law and to shift $\tau_{ex}.\Delta\nu$ towards 1, leading to enhanced line-broadening (Mandel et al., 1996). Moreover, it was observed that the exchange contribution was independent of sample concentration, which excludes the possibility that this effect is due to fractional self-association of E6-C 4C/4S. Therefore, the occurrence of slow-intermediate exchange in specific regions of the primary sequence of E6-C 4C/4S may rather be related to a transition from a folded state toward a less populated locally unfolded state resulting from the lack of stabilising interactions, normally provided by the N-terminal domain. This would suggest that the two loops are involved in the domain interface, an hypothesis which is reinforced by the fact that two hydrophobic residues within L1 loop (Tyr84 and Leu88) are conserved among all HPV strains. These residues may thus prove to be potentially valuable mutation targets for stabilising the domain. Following this detailed characterisation of the dynamic behaviour of the C-terminal domain of E6, we are now in a position to interpret structural information in NMR spectra of E6-C 4C/4S and expect soon to be able to report on its three-dimensional structure.

Acknowledgements

This program was supported by a grant from Association pour la Recherche contre le Cancer (ARC). We are grateful to Claude Ling for his excellent management of the computing and NMR facilities. We thank Rieko Ishima for kindly providing pulse sequences for dispersion experiments and advice.

References

Allen, F.H., Bird, C.M., Rowland, R.S. and Raithby, P.R. (1997) *Acta Crystallogr. Sect. B-Struct. Commun.*, **53**, 696–701.

Androphy, E.J., Hubbert, N.L., Schiller, J.T. and Lowy, D.R. (1987) *EMBO J.*, **6**, 989–992.

Bartels, C., Xia, T.H., Billeter, M., Güntert, P. and Wüthrich, K. (1995) *J.Biomol. NMR*, **5**, 1–10.

Carver, J.P. and Richards, R.E. (1972) *J. Magn. Reson.*, **6**, 89–105.

Cole, S.T. and Danos, O. (1987) *J. Mol. Biol.*, **193**, 599–608.

Cowburn, D. (1997) *Curr. Opin. Struct. Biol.*, **7**, 835–838.

Crook, T., Tidy, J.A. and Vousden, K.H. (1991) *Cell*, **67**, 547–556.

Degenkolbe, R., Gilligan, P., Gupta, S. and Bernard, H.U. (2003) *Biochemistry*, **42**, 3868–3873.

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J. and Bax, A. (1995) *J. Biomol. NMR*, **6**, 277–293.

Dell, G. and Gaston, K. (2001) *Cell. Mol. Life Sci.*, **58**, 1923–1942.

Dingley, A.J., Mackay, J.P., Chapman, B.E., Morris, M.B., Kuchel, P.W., Hambly, B.D. and King, G.F. (1995) *J. Biomol. NMR*, **6**, 321–328.

Du, M., Fan, X., Hong, E. and Chen, J. (2002) *Biochem. Biophys. Res. Commun.*, **296**, 962–969.

Filippova, M., Song, H., Connolly, J., Dermody, T. and Duerksen-Hughes, P. (2002) *J. Biol. Chem.*, **277**, 21730–21739.

Finzer, P., Aguilar-Lemarroy, A., Rösl, F., Rapp, L., Chen, J.J., Mantovani, F. and Banks, L. (2002) *Cancer Lett.*, **188**, 15–24.

Gewin, L. and Galloway, D.A. (2001) *J. Virol.*, **75**, 7198–7201.

Huibregtse, J.M., Scheffner, M. and Howley, P.M. (1993) *Mol. Cell. Biol.*, **13**, 775–784.

Hung, A.Y. and Sheng, M. (2002) *J. Biol. Chem.*, **277**, 5699–5702.

Iftner, T., Elbel, M., Schopp, B., Hiller, T., Loizou, J., Caldecott, K. and Stubenrauch, F. (2002) *EMBO J.*, **21**, 4741–4748.

Kay, L.E., Torchia, D.A. and Bax, A. (1989) *Biochemistry*, **28**, 8972–8979.

Kumar, A., Zhao, Y., Meng, G., Zeng, M. Srinivasan, S., Delmolino, L., Gao Q., Dimri, G., Weber, G., Wazer, D., Band, H. and Band, V. (2002) *Mol. Cell. Biol.*, **22**, 5801–5812.

Lipari, F., McGibbon, G.A., Wardrop, E. and Cordingley, M.G. (2001) *Biochemistry*, **40**, 1196–1204.

Liu, Y., Chen J.J., Gao Q., Dalal S., Hong Y., Mansur C.P., Band V. and Androphy, E.J. (1999) *J. Virol.*, **73**, 7297–7307.

Loria, J.P., Rance M. and Palmer, A.G. (1999) *J. Am. Chem. Soc.*, **121**, 2331–2332.

Mandel, A.M., Akke M. and Palmer, A.G. (1996) *Biochemistry*, **35**, 16009–16023.

Mantovani, F. and Banks, L. (2001) *Oncogene*, **20**, 7874–7887.

Merutka, G., Dyson, H.J. and Wright, P.E. (1995) *J. Biomol. NMR*, **5**, 14–24.

Mulder, F.A., Spronk, C.A. E.M., Slijper, M., Kaptein, R. and Boelens, R. (1996) *J. Biomol. NMR*, **8**, 223–228.

Nakagawa, S., Watanabe, S., Yoshikawa, H., Taketani, Y., Yoshiike, K. and Kanda, T. (1995) *Virology*, **212**, 535–542.

Nominé, Y., Charbonnier, S., Ristriani, T., Stier, G., Masson, M., Cavusoglu, N., Van Dorsselaer, A., Weiss, E., Kieffer, B. and Travé, G. (2003) *Biochemistry*, **42**, 4909–4917.

Nominé, Y., Ristriani, T., Laurent, C., Lefèvre, J.F., Weiss, E. and Travé, G. (2001) *Protein. Expr. Purif.*, **23**, 22–32.

Oh, S.T., Kyo, S. and Laimins, L.A. (2001) *J. Virol.*, **75**, 5559–5566.

Palmer, 3rd, A.G., Kroenke, C.D. and Loria, J.P. (2001) *Meth. Enzymol.*, **339**, 204–238.

Pelton, J.G., Torchia, D.A., Meadow, N.D. and Roseman, S. (1993) *Protein Sci.*, **2**, 543–558.

Pfuhl, M., Chen, H.A., Kristensen, S.M. and Driscoll, P.C. (1999) *J Biomol NMR*, **14**, 307–320.

Pim, D., Storey, A., Thomas, M., Massimi, P. and Banks, L. (1994) *Oncogene*, **9**, 1869–1876.

Pim, D., Thomas, M. and Banks, L. (2002) *Oncogene*, **21**, 8140–8148.

Piotto, M., Saudek, V. and Sklenár, V. (1992) *J. Biomol. NMR*, **2**, 661–665.

Rapp, L. and Chen, J.J. (1998) *Biochim. Biophys. Acta*, **1378**, 1–19.

Ristriani, T., Masson, M., Nominé, Y., Laurent, C., Lefèvre, J.F., Weiss, E. and Travé, G. (2000) *J. Mol. Biol.*, **296**, 1189–1203.

Ristriani, T., Nominé, Y., Laurent, C., Weiss, E. and Travé, G. (2002) *Protein Expr. Purif.*, **26**, 357–367.

Ristriani, T., Nominé, Y., Masson, M., Weiss, E. and Travé, G. (2001) *J. Mol. Biol.*, **305**, 729–739.

Scheffner, M., Huibregtse, J.M., Vierstra, R.D. and Howley, P.M. (1993) *Cell*, **75**, 495–505.

Schwarz, E., Freese, U.K., Gissmann, L., Mayer, W., Roggenbuck, B., Sremlau, A. and zur Hausen, H. (1985) *Nature*, **314**, 111–114.

Summers, M.F., South, T.L., Kim, B. and Hare, D.R. (1990) *Biochemistry*, **29**, 329–340.

Thomas, M., Laura, R., Hepner, K., Guccione, E., Sawyers, C., Lasky, L. and Banks, L. (2002) *Oncogene*, **21**, 5088–5096.

van Tilborg, P.J., Czisch, M., Mulder, F.A., Folkers, G.E., Bonvin, A.M., Nair, M., Boelens, R. and Kaptein, R. (2000) *Biochemistry*, **39**, 8747–8757.

Veldman, T., Horikawa, I., Barrett, J.C. and Schlegel, R. (2001) *J. Virol.*, **75**, 4467–4472.

Vuister, G.W. and Bax, A. (1993) *J. Am. Chem. Soc.*, **115**, 7778–7782.

Werness, B.A., Levine, A.J. and Howley, P.M. (1990) *Science*, **248**, 76–79.

Wishart, D.S., Sykes, B.D. and Richards, F.M. (1992) *Biochemistry*, **31**, 1647–1651.

zur Hausen, H. (1991) *Virology*, **184**, 9–13.