Domain Substructure of HPV E6 Oncoprotein: Biophysical Characterization of the E6 C-Terminal DNA-Binding Domain†

Yves Nominé,‡ Sebastian Charbonnier,‡ Tutik Ristriani,‡ Gunter Stier,§ Murielle Masson,‡ Nukhet Cavusoglu,|| Alain Van Dorsselaer,|| Étienne Weiss,‡ Bruno Kieffer,⊥ and Gilles Travé*,‡

Laboratoire d'Immunotechnologie, UMR CNRS 7100, École Supérieure de Biotechnologie de Strasbourg, 67400 Illkirch, France, EMBL, Meyerhofstrasse 1, D-69117 Heidelberg, Germany, Laboratoire de RMN, UMR CNRS 7104, École Supérieure de Biotechnologie de Strasbourg, 67400 Illkirch, France, and Laboratoire de Spectrométrie de Masse Bio-Organique, UMR CNRS 7509, Faculté de Chimie, 67087 Strasbourg, France

Received October 9, 2002; Revised Manuscript Received February 19, 2003

ABSTRACT: E6 is a viral oncoprotein implicated in cervical cancers, produced by high-risk human papillomaviruses (HPVs). Structural data concerning this protein are scarce due to the difficulty of producing recombinant E6. Recently, we described the expression and purification of a stable, folded, and biologically active HPV16 E6 mutant called E6 6C/6S. Here, we analyzed the domain substructure of this mutated E6. Nonspecific proteolysis of full-length E6 6C/6S (158 residues) yielded N-terminal and C-terminal fragments encompassing residues 7−83 and 87−158, respectively. The C-terminal fragment of residues 87−158 was cloned, overexpressed, and purified at concentrations as high as 1 mM. The purified domain retains the selective four-way DNA junction recognition activity of the full-length E6 protein. Using UV absorption, UV fluorescence, circular dichroism, and nuclear magnetic resonance, we show that the peptide is primarily monomeric and folded with equal proportions of α-helix and *β*-sheet secondary structure.

E6 is one of the two oncoproteins produced by "high-risk" human papillomaviruses (HPVs)[1] responsible for cervical cancers (*1*). E6 is thought to promote tumorigenesis by stimulating cellular degradation of the tumor suppressor p53 via formation of a trimeric complex comprising E6, p53, and the cellular ubiquitination enzyme E6AP (*2*, *3*). However, recent findings suggest that E6 displays other activities unrelated to p53. These include recognition of a variety of other cellular proteins: transcription coactivators p300/CBP (*4*, *5*) and ADA3 (*6*), transcription factors c-Myc (*7*) and IRF3 (*8*), replication protein hMCM7 (*9*), DNA repair proteins MGMT (*10*) and XRCC1 (*11*), protein kinases PKN (*12*) and Tyk2 (*13*), Rap-GTPase activating protein E6TP1 (*14*), tumor necrosis factor receptor TNF-R1 (*15*), apoptotic protein Bak (*16*), clathrin-adaptor complex AP-1 (*17*), focal adhesion component paxillin (*18*, *19*), calcium-binding proteins E6BP (*20*) and fibulin-1 (*21*), and several members of the PDZ protein family such as hDLG (*22*), hScrib (*23*), MAGI-1 (*24*), and MUPP1 (*25*). In addition, E6 activates or represses several cellular or viral transcription promoters (*8*, *26*−*28*). In particular, E6 induces transcriptional activation of the gene encoding the retrotranscriptase of human telomerase (*29*−*31*). Finally, we have recently demonstrated that E6 is a DNA-binding protein which recognizes four-way DNA junctions (*32*, *33*). In contrast to the large amount of functional data, there are very few biophysical and structural studies concerning E6 due to the difficulty of expressing and purifying this protein, or its domains, in a stable folded form. Most efforts have focused on E6 from HPV16 since this is the highest-risk HPV strain for cervical cancers. Two HPV16 E6 polypeptides are thought to be produced *in vivo* depending on which methionine is used as a start codon: a 158-residue form and a 151-residue form (*34*) (here, we will use the 158-residue numbering). Sequence alignments of E6s from numerous HPV subtypes suggested the presence of two zinc-binding motifs (*35*). Both motifs are 37 residues long and contain four cysteines distributed in a CxxC-(29x)-CxxC sequence. Zinc binding to these motifs was experimentally demonstrated by zinc blotting assays (*35*−*38*), site-directed mutagenesis (*39*), and atomic and electronic absorption methods (*40*). On this basis, several authors (*41*−*43*) divide the E6 sequence into five "regions": the N-terminal loop (residues 1−36), the N-terminal zinc-binding motif (residues 37−73), the linker region (residues 74−110), the C-terminal zinc-binding motif (residues 110−146), and the C-terminal loop (residues 147−158). However, the structural boundaries of the zinc-binding domains may include additional residues provided by the N-terminal and/or C-terminal regions flanking the 37-residue motifs. Recently, we expressed and purified two fragments containing the N-terminal and the C-terminal zinc-binding motifs of wild-type HPV16 E6 (*33*). On the basis of multiple

† This work was supported by a grant from the Association pour la Recherche sur le Cancer and by a "PCV" grant from the CNRS. T.R. was a recipient of a grant from the Ligue Nationale Contre le Cancer.

* To whom correspondence should be addressed. Telephone: 0033 3 90 24 47 20. Fax: 0033 3 90 24 47 70. E-mail: trave@esbs.u-strasbg.fr.

‡ UMR CNRS 7100.

§ EMBL.

|| UMR CNRS 7509.

⊥ UMR CNRS 7104.

[1] Abbreviations: HPV, human papillomavirus; CD, circular dichroism; UV, ultraviolet; NMR, nuclear magnetic resonance; DTT, dithiothreitol; LB, Luria broth; SDS−PAGE, sodium dodecyl sulfate−polyacrylamide gel electrophoresis; IPTG, isopropyl 1-thio-*β*-D-galactopyranoside.

10.1021/bi026980c CCC: $25.00
Published on Web 04/08/2003

sequence alignments, we deliberately included nine residues upstream of the first cysteine and five residues downstream of the fourth cysteine of each motif. The N-terminal and the C-terminal fragments each contained 51 residues and spanned residues 28−78 and 101−151, respectively. Both constructs were produced as soluble fusions to the C-terminus of maltose-binding protein (MBP). The MBP−ZD2 fusion containing the C-terminal zinc-binding motif was monomeric and was shown to retain the DNA recognition activity of full-length HPV16 E6 (*33*). However, in that work, we did not attempt to separate the ZD2 peptide from the MBP carrier. In another work, Lipari et al. (*40*) produced an N-terminal fragment of wild-type HPV16 E6 spanning residues 9−84. This fragment was therefore 75 residues long, and it included 28 residues upstream of the first cysteine and nine residues downstream of the fourth cysteine. The fragment was expressed and purified in a soluble form which could be raised at high concentrations. It was folded with a high content of α-helical and β-sheet secondary structures and contained a single zinc ion (*40*). The same authors also expressed and purified a longer construct which spanned almost the entire sequence of wild-type HPV16 E6 (residues 9−149). This construct was also folded with α-helical and β-sheet structure, but it displayed low solubility (*40*). The authors concluded that the C-terminal region (residues 85−149), containing the second zinc-binding motif, was probably unstable or insoluble and contributed to the low solubility of full-length HPV16 E6.

Recently, we described a stabilized mutant of HPV16 E6 in which six nonconserved cysteines were changed into serines (*44*, *45*). This mutant, called E6 6C/6S, can be purified as a stable, folded, and biologically active form (*45*). Here, we have investigated the domain substructure of the E6 6C/6S mutant. Nonspecific proteolysis of the folded protein yielded two soluble fragments of ∼75 residues each which were separated by chromatography and analyzed by N-terminal sequencing and mass spectrometry. One of these fragments corresponds to the N-terminal domain already described for wild-type HPV16 E6 (*40*). More interestingly, the second fragment corresponds to the complementary C-terminal half. This C-terminal domain, derived from our E6 6C/6S mutant, is soluble and mainly monomeric at concentrations as high as 1 mM. Circular dichroism and NMR data show that the protein is properly folded and contains a high content of α-helical and β-sheet secondary structures.

EXPERIMENTAL PROCEDURES

Expression and Purification of E6 6C/6S and E6-C 4C/4S. Full-length E6 6C/6S (158 residues) was expressed as a fusion to the C-terminus of MBP and purified as described previously (*45*). The MBP−E6-C 4C/4S construct corresponds to residues 87−158 of E6 6C/6S fused to the C-terminus of His6-MBP via a TEV protease-sensitive linker. The sequence encoding E6-C 4C/4S was PCR-amplified with appropriate oligomers and inserted into the *Nco*I and *Kpn*I sites of PETM-41, a modified pET24d expression vector (Novagen) containing an N-terminal His6-MBP tag followed by a TEV protease cleavage site (www.embl-heidelberg.de/ExternalInfo/geerlof/draft_frames/index.html). The sequence was checked using a Thermo Sequenase Cy5.5 dye terminator cycle sequencing kit (Amersham Pharmacia Biotech). Sequences were run on a Seq4×4 Basecaller and processed using an ALFwin sequence analyzer. BL21 DE3 *Escherichia coli* cells freshly electroporated with the MBP−E6-C 4C/4S expression construct were directly transferred in LB, and 15 μg/mL kanamycin, and grown overnight for 12−14 h at 37 °C. The preculture was diluted 40-fold in fresh LB and kanamycin and grown at 37 °C until an OD_{600} of 0.6−0.7. Cells were harvested by centrifugation at 2300*g* and 25 °C for 10 min, transferred to fresh medium pre-equilibrated at 37 °C, adjusted to 0.5 mM IPTG, and then grown for 6−8 h at 22 °C. Cells from 2 L of either expression culture were harvested by centrifugation at 2300*g* and 4 °C for 20 min. To minimize oxidation problems, all buffers used for purification were degassed using a water vacuum pump and then bubbled extensively with argon. The pellet was resuspended in 150 mL of buffer A [50 mM Tris, 150 mM NaCl, and 2 mM DTT (pH 6.8)] containing 5% glycerol, 1 μg/mL DNase I, 1 μg/mL RNase I, and 3 tablets of anti-protease cocktail (EDTA-free) (Boehringer Mannheim). Cells were broken by sonication on ice and then centrifuged at 18000*g* and 6 °C for 30 min. The supernatant was filtered (Millipore, 0.22 μm) and loaded on an 80 mL column of amylose resin (New England Biolabs) pre-equilibrated with buffer A. The column was washed stepwise with 1, 3, and 12 volumes of buffer A adjusted at 100, 50, and 12%, respectively, of the anti-protease concentration initially present in sonication buffer. Remarkably, pure MBP−E6-C 4C/4S construct was mainly recovered as a fraction leaking from the column in the late washing steps. The protein was mixed with recombinant TEV protease (G. Stier, unpublished results) with a ratio of 10^{-2} mol of TEV protease/mol of MBP fusion. Incubation was performed at 6 °C for 12−24 h 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. The digestion product was concentrated to a volume of 1 mL using a 10 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 column's calibration. The sample concentration was increased to 1−1.4 mM using a 15 mL Ultrafree Biomax 5K NMWL Membrane (Millipore).

Mass and Sequence Analysis. Mass measurements were carried out on a Bruker (Bremen, Germany) BIFLEX III matrix-assisted laser desorption time-of-flight mass spectrometer equipped with the SCOUT high-resolution optics with the *X*−*Y* multisample probe and griddless reflector. This instrument was used at a maximum accelerating potential of 19 kV and was operated in the linear mode. Ionization was accomplished with a 337 nm beam from a nitrogen laser with a repetition rate of 3 Hz. The output signal from the detector was digitized at a sampling rate of 2 GHz. The mass accuracy was approximately 0.1%. A saturated solution of α-cyano-4-hydroxycinnamic acid in acetone was used as a matrix. A first layer of fine matrix crystals was obtained by spreading and fast evaporation of 0.5 μL of the matrix solution. On this fine layer of crystals, a droplet of 0.5 μL of an aqueous HCOOH (5%) solution was deposited. Afterward, 0.5 μL of a sample solution was added and a second droplet (0.2 μL) of a matrix-saturated solution (in

50% H_2O/50% ACN) was added. The preparation was dried under vacuum. The instrument was calibrated in the mass range of 5000−20000 Da using a mixture of chicken cytochrome *c* and lysozyme (Sigma). To perform a linear calibration in this mass range, mono- and bi-charged species at *m*/*z* 6181 and 12 361 for cytochrome *c* and *m*/*z* 7153 and 14 306 for lysozyme were used successively.

For N-terminal sequence analysis, the peptides were separated with an SDS gel and transferred onto nitrocellulose membranes. The bands were isolated and subjected to N-terminal Edman degradation using an Applied Biosystems model 492 N-terminal protein sequencer. PTH amino acids are identified with an Applied Biosytems model 140 PTH analyzer and a model 785A programmable detector set at a wavelength of 269 nm.

Spectroscopic Measurements. Fluorescence measurements were performed with a SPEX Fluorolog-2 spectrofluorimeter (SPEX Industries, Inc., Edison, NJ) equipped with a 450 W Xe lamp, a double-grating excitation monochromator, and a single-grating emission monochromator. Data were acquired with a photon counting photomultiplier (linear up to 10^7 counts/s) with high voltages fixed at 800 V. Slit widths were adjusted to avoid saturation of detectors. The solution (2 mL) was placed in a cuvette maintained at 15 °C in a thermostated cuvette handler. Emission spectra recordings were typically sampled at every half-nanometer. In all experiments, slit widths were set to 1.8 nm both for excitation and emission. Ultraviolet absorption measurements were performed with a Perkin-Elmer λ-2 spectrophotometer. Circular dichroism was performed using a Jobin-Yvon spectropolarimeter. Secondary structure content was estimated by the VARiable SELECtion method (VARSELEC) (*46*).

Nuclear Magnetic Resonance Experiments. ^{1}H NMR measurements were obtained from a 1 mM sample of E6-C 4C/4S in 20 mM deuterated Tris (pH 6.8), 50 mM NaCl, 1 mM DTT, and 10% D_2O. The two-dimensional (2D) ^{1}H−^{1}H NOESY spectrum was recorded at 15 °C on a Bruker DRX-600 spectrometer with a mixing time of 200 ms and a spectral width of 8 kHz in both dimensions. The water signal was suppressed using the WATERGATE sequence (*47*). Forty-eight transients of 2K data points were acquired per t_1 value, with 512 t_1 increments and a relaxation delay of 2.25 s. Data were processed using NMR-pipe (*48*) and analyzed with XEASY (*49*).

The translational diffusion constants were measured on a Bruker AMX-500 spectrometer equipped with a *Z*-gradient 5 mm probe using a modified version of the longitudinal encoded diffusion (LED) experiment (*50*, *51*). The protein signal decay was followed by recording 32 LED spectra with a gradient length δ varying from 1 to 16 ms and other constant delays set to the following values: encoding delay $\tau = 20$ ms and longitudinal diffusion delay $T = 50$ ms. Square gradient pulses were applied with an intensity of 0.28 T/m. This intensity was calibrated by imaging a hole of known length in a Teflon phantom and by measuring the rate of diffusion of a water sample. An additional check was performed by measuring the self-diffusion coefficient of lysozyme and comparing it to previously reported values (*52*). The measured value was similar to the reported value with a deviation of <16%. Moreover, the value which could be extrapolated from our curve for hemoglobin was also consistent with the experimental value (*53*). FELIX 2.1 (Accelrys Inc., Burlington, MA) software was used to process the spectra and to measure the peak height of 10 well-resolved protein resonances. The diffusion coefficients were obtained by fitting the signal intensities with a three-parameter monoexponential decay using a simplex algorithm implemented within the Matlab software (The Mathworks Inc., Natick, MA). The expression of the exponential was

$$R(x) = R_\infty + R_0 \exp(-D_t x) \text{ where}$$

$$x = \gamma H^2 G_z^2 \delta^2 (\Delta - \delta/3)$$

We chose a three-parameter fit to avoid any bias during the nonlinear fitting procedure. Examination of the resulting R_∞ parameter was used as an independent check of the monoexponential nature of the decay curve. Errors on the diffusion coefficients were estimated by calculating the standard devation of the 10 diffusion values obtained at each temperature.

Construction of the DNA Probe for the Gel-Retardation Assay. A cruciform based upon the sequence of junction 1 originally described by Duckett et al. (*54*) was constructed by hybridizing four oligonucleotides of 24 nucleotides each termed b, h, r, and x, with the b-strand 5′-^{32}P-labeled: b-strand, 5′-GTCCTAGCAAGCCGCTGCTACCGG-3′; h-strand, 5′-CCGGTAGCAGCGAGAGCGGTGGTT-3′; r-strand, 5′-AACCACCGCTCTTCTCAACTGCAG-3′; and x-strand, 5′-CTGCAGTTGAGAGCTTGCTAGGAC-3′.

Gel-Retardation Analysis. Purified proteins were mixed with labeled DNA junction in 25 mM Tris-HCl (pH 7.5), 100 mM NaCl, 10 mM KCl, 1.5 mM $MgCl_2$, 100 μg/mL genomic DNA, and 5% glycerol in a total volume of 10 μL. Samples were then incubated for 15 min on ice and electrophoresed in 6.5% polyacrylamide gels (29:1 acrylamide: bisacrylamide ratio) with 45 mM Tris-borate in the presence of 1 mM EDTA, at 10 V/cm and 8 °C. Gels were dried on Whatman 3M paper and analyzed on a phosphorimager (Amersham Pharmacia Biotech).

RESULTS

Nonspecific Proteolysis of E6 6C/6S Yields Soluble Fragments Corresponding to the N-Terminal and C-Terminal Halves of E6 6C/6S. We routinely purified samples of the E6 6C/6S mutant at a final concentration of 5−10 μM (*45*). When stored at 4 °C in high-salt phosphate buffer [20 mM sodium phosphate (pH 7.4) and 500 mM NaCl], samples did not undergo detectable proteolysis for several months. However, when the purified protein was stored in high-salt Tris buffer [20 mM Tris (pH 7.4) and 500 mM NaCl] instead of phosphate, the samples underwent nonspecific proteolysis, generating several soluble fragments of ∼8−10 kDa (Figure 1). Over five months of storage, the full-size protein gradually disappeared to be completely replaced by the fragments (Figure 1, lanes 1−4). Under low-salt conditions, the slowest-migrating fragment (peptide 3) and a faster minor band (peptide 2) bound to anion-exchange resin and could be eluted at higher ionic strengths (lane 6). The fastest-migrating band (peptide 1) did not bind to resin and was thus recovered in the flow-through of the low-salt incubation step (lane 7). The purified peptides were subjected to N-terminal sequencing and MALDI-TOF mass spectrometry. The results of this analysis are summarized in Table 1. Peptide 1, which did

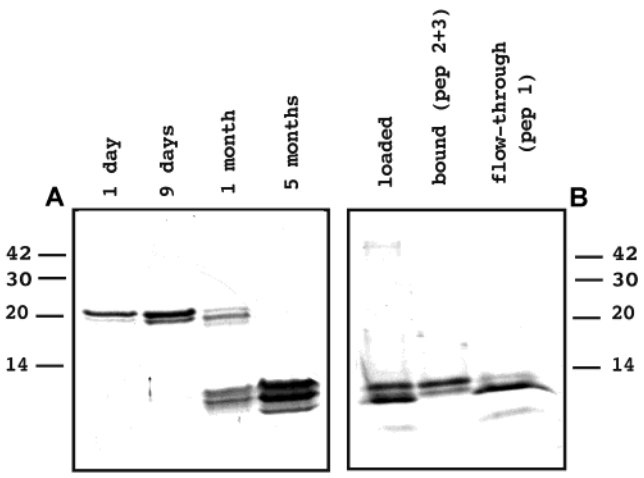

FIGURE 1: Nonspecific proteolysis of E6 6C/6S yields soluble fragments that can be separated by ion-exchange chromatography. (A) Time-dependent proteolysis of E6 6C/6S in Tris buffer. E6 6C/6S was purified as previously described (*45*) but was kept in Tris buffer (pH 7.4) instead of phosphate buffer. SDS gel analysis of samples after storage for 1 day, 9 days, 1 month, and 5 months at 4 °C (lanes 1−4, respectively). (B) Fragment separation. The soluble fraction from E6 6C/6S stored for 5 months in Tris buffer was diluted into low-salt Tris buffer (20 mM NaCl) and then allowed to incubate with S-Sepharose anion-exchange resin (Amersham Biosciences) overnight. The resin was washed three times with low-salt buffer and then incubated with high-salt Tris buffer (750 mM NaCl) to allow elution of the bound protein: lane 5, total protein before incubation with resin; lane 6, high-salt eluate containing peptides 2 and 3 which did bind to resin under low-salt conditions; and lane 7, flow-through containing peptide 1 which did not bind to resin under low-salt conditions.

not bind to anion-exchange resin under low-salt conditions, starts at alanine 7 and displayed an experimental mass of 9251 ± 9 Da. Peptides 2 and 3, which bound resin under low-salt conditions and were eluted under high-salt conditions, both start at serine 87 and display experimental masses of 8011 ± 8 and 8637 ± 8 Da, respectively. Therefore, as compared to the 158 residues of full-length E6 6C/6S, peptide 1 corresponds to residues 7−83, peptide 2 to residues 87−153, and peptide 3 to residues 87−158, respectively. For each peptide, the theoretical mass computed from sequence is in good agreement with the experimental mass (Table 1). The calculated isoelectric points of the peptides also fit with their behavior on anion-exchange resin. Peptide 1 is neutral at pH 7.4, and therefore, it does not bind S-Sepharose resin. In contrast, peptides 2 and 3 are positively charged at pH 7.4 which explains their strong binding to the resin.

Cloning, Expression, and Purification of E6-C 4C/4S (Residues 87−158 of E6 6C/6S). The results presented above showed that residues 87−158 constituted a soluble fragment which resisted mild proteolysis and could be efficiently purified by chromatography. This pointed to this peptide as an interesting candidate for biophysical studies. Residues 87−158 were fused to the C-terminus of maltose-binding protein (MBP) via a linker sensitive to TEV protease. The resulting MBP fusion (called MBP−E6-C 4C/4S) was overexpressed (Figure 2, lane 1), purified by affinity on MBP-specific amylose resin (lanes 2 and 3), and then subjected to TEV digestion, yielding a 40 kDa protein corresponding to MBP and an approximatively 10 kDa protein corresponding to the E6-C 4C/4S domain (lane 4). E6-C 4C/4S was then separated from MBP and TEV by gel filtration (lane 5). E6-C 4C/4S was eluted as a single peak which corresponded to a monomer according to the column's calibration. This purified E6-C 4C/4S sample was completely soluble and could be concentrated up to 1 mM without undergoing any detectable aggregation.

In parallel, we tried a similar approach with a shorter construct, called MBP−ZD2, which consists of residues 101−151 of wild-type HPV16 E6 fused to the C-terminus of MBP via a thrombin-sensitive liker (*33*). We have previously shown that MBP−ZD2 retains the DNA binding properties of full-length E6 (*33*). The ZD2 construct includes the C-terminal zinc-binding motif, but it lacks 14 N-terminal residues (S_{87}YSLYGTTLEQQYN$_{100}$) when compared to E6-C 4C/4S. In addition, ZD2 corresponds strictly to a fragment of the sequence of wild-type HPV16 E6, whereas E6-C 4C/4S contains four mutations of nonconserved cysteines into serines. However, when ZD2 was separated from the MBP carrier by thrombin digestion, it became prone to nonspecific proteolysis and fast precipitation and could not be purified further (data not shown).

E6-C 4C/4S Retains Monomeric Binding to Four-Way DNA Junctions. We have previously demonstrated that E6 is a DNA-binding protein which recognizes four-way junctions (*32*), and that this activity is localized in the C-terminus of the protein (*33*). The purified E6-C 4C/4S sample retains this activity as shown by a gel-retardation assay (Figure 3, lanes 5−8). In our previous work, we have shown that only one monomer of E6 6C/6S can bind to the junction (*32*). Interestingly, E6-C 4C/4S generated only one retardation band even at the highest concentration that was tested (1 μM) (lane 8). In addition, this band migrated faster than the band generated by E6 6C/6S (lanes 2−4). This strongly suggests that the E6-C 4C/4S domain binds to the junction only as a monomer, as observed for full-length E6 6C/6S.

Biophysical Characterization of the E6-C 4C/4S Domain. To check the purity and quality of the E6-C 4C/4S preparation, the sample was first subjected to UV absorption measurements (Figure 4A). This spectrum has a particular shape due to the aromatic residues of E6-C 4C/4S: three tyrosines (absorption maximum of 274 nm) and one tryptophan (absorption maximum of 280 nm). The CD (circular dichroism) spectrum spanning the far-UV region is characteristic of a folded protein with defined secondary structure elements (Figure 4B). On the basis of that spectrum, back-calculations of the secondary structure content of E6-C 4C/4S give proportions of 22 ± 1% α-helix, 22 ± 3% antiparallel β-sheet, 3 ± 1% parallel β-sheet, and 19 ± 2% turn. The fluorescence spectrum (Figure 4C) was recorded at an excitation wavelength of 280 nm. The maximum of intrinsic fluorescence was found at 331 nm, which is shifted to lower wavelengths as compared to the normal maximum emission of tryptophan in water (∼350 nm). This indicates that the single tryptophan of E6-C 4C/4S is buried inside the protein in a hydrophobic environment, suggesting that the protein is folded. Finally, the protein was subjected to NOESY NMR measurements (Figure 5). Spectral dispersion in both the amide and aliphatic region confirms that the protein is folded. The occurrence of numerous HN−HN NOESY cross-peaks indicates the presence of α-helical regions, whereas several Hα resonances with chemical shifts greater than 4.8 ppm are indicative of β-sheet structure. However, a detailed analysis of the spectra shows a large

Table 1: Characteristics of the Three Proteolytic Peptides

	peptide 1	peptide 2	peptide 3
S-Sepharose binding[a]	not bound	bound	bound
N-terminal sequence[b]	AMFQDPQE	SYSLYGTT	SYSLYGTT
expected mass[c] (Da)	9251 ± 9	8011 ± 8	8637 ± 8
sequence[d]	^{7}AMFQDP...SKISEY^{83}	^{87}SYSLYG...RSSRTR^{153}	^{87}SYSLYG...RRETQL^{158}
calcd mass[e] (Da)	9242.7	8002.2	8629.9
calcd pI[f]	7.04	10.8	10.8

[a] Binding to S-Sepharose resin at pH 7.4 and under low-salt conditions. [b] N-Terminal peptidic sequence. [c] Molecular mass obtained by mass spectrometry. [d] Sequence deduced from experimental data. [e] Theoretical mass recalculated from the sequence. [f] Isoelectric point recalculated from the sequence.

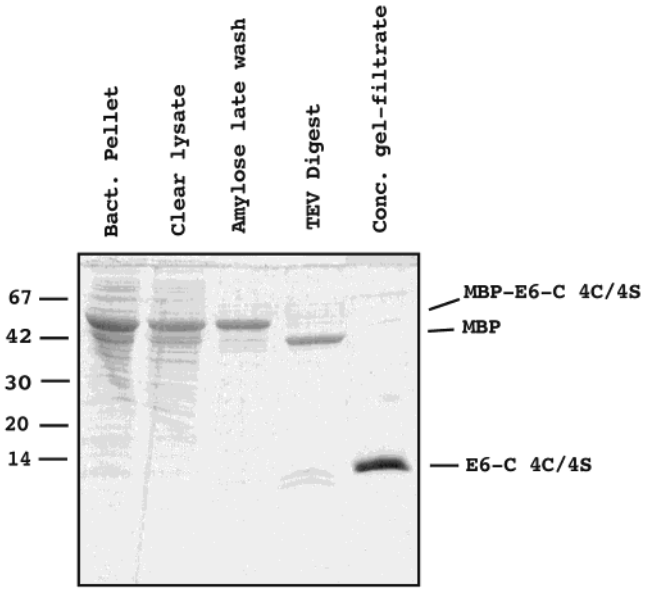

FIGURE 2: Expression and purification of the E6-C 4C/4S domain. Pellets of XL1 Blue bacteria expressing the MBP−E6-C 4C/4S (lane 1) construct were sonicated and then ultracentrifuged. The cleared supernatant (lane 2) was loaded on an amylose column. Pure MBP−E6-C 4C/4S fusion construct leaked from the column in the late washing steps (lane 3). Thrombin incubation led to the proteolytic separation of the MBP moiety and the E6-C 4C/4S moiety (lane 4). Digestion products were concentrated and separated by gel filtration. Fractions containing E6-C 4C/4S were pooled and concentrated to 50 μM (lane 5).

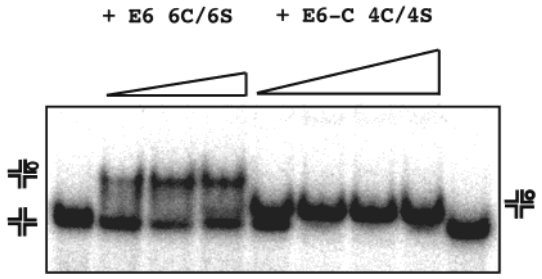

FIGURE 3: Binding of the pure E6-C 4C/4S domain to four-way DNA junctions. Radioactively ^{32}P-labeled junction 1 (1 nM) (*54*), with four arms of 12 bp each, was incubated with increasing concentrations of either pure E6 6C/6S or pure E6-C 4C/4S in the presence of 100 μg/mL genomic DNA and 1 mM magnesium ions. The mixtures were electrophoresed in the presence of 1 mM EDTA and then analyzed by phosphorimaging. In lanes 1−8, 1 nM cruciform DNA was incubated with 0, 80, 125, and 250 nM E6 6C/6S (lanes 1−4, respectively) and with 125, 250, 500, 1000, and 0 nM E6-C 4C/4S (lanes 5−9, respectively). The labeled junction is schematized as a square cross, the E6-C 4C/4S domain as a single sphere, and full-length E6 6C/6S as an ellipsoid.

heterogeneity in peak line width along the sequence, from 16 to 47 Hz. This may be due either to the occurrence of intermediate microsecond to millisecond time scale motions within the domain or to the formation of dimers or higher-order oligomers.

The E6-C 4C/4S Domain Diffuses as a Monomer at 1 mM. To check the possible origins of heterogeneous line broadening, we first recorded spectra on the sample at various concentrations, and observed that dilution had no effect on ^{1}H line shapes (data not shown). This gave us a first indication that heterogeneous line broadening was not due to oligomerization. To further investigate this point, we used pulsed field gradient LED experiments to measure the translational diffusion coefficient (D_t) of a 1 mM sample of E6-C 4C/4S for temperatures ranging from 15 to 50 °C. Figure 6A shows the D_t values plotted as a function of the temperature:viscosity ratio (T/η). The observed linear relationship between D_t and T/η reflects the accuracy of the diffusion measurement and indicates that the shape and the oligomeric state of the protein were conserved over this range of temperatures. For temperatures above 50 °C, the accuracy of the diffusion measurements was diminished by convection. To determine the oligomeric state of E6-C 4C/4S, we compared normalized D_t values (20 °C in water) measured for proteins with sizes ranging from 1100 Da (a peptidic siderophore) to 14 500 Da (lysozyme) (Figure 6B). For a given molecular mass, the normalized D_t values depend on the mass and shape of the proteins. The comparison of normalized D_t values suggests that E6-C 4C/4S exists primarily as a monomeric protein whose hydrodynamic shape does not differ significantly from the spherical model. Therefore, heterogeneous line broadening is probably due to the occurrence of particular motions in the monomeric domain.

DISCUSSION

In this work, we have analyzed the domain substructure of E6 6C/6S, a stabilized mutant of HPV16 E6. Nonspecific proteolysis of the protein generated three soluble peptidic fragments which span almost the entire sequence of E6 (Table 1). One fragment spans the N-terminal half (residues 7−83) and includes the first cluster of conserved cysteines known to bind zinc (*36*, *37*, *39*, *40*) (positions 37, 40, 70, and 73). The two other fragments span the C-terminal half (residues 87−153 and 87−158) and include the second zinc-binding motif (positions 110, 113, 143, and 146). Proteolysis was extremely selective and attacked only a few positions on the protein sequence, i.e., peptide bonds between residues 6 and 7, 83 and 84, 86 and 87, and 153 and 154. Under limiting conditions, proteases are known to operate preferentially on interdomain linkers as well as N-terminal and

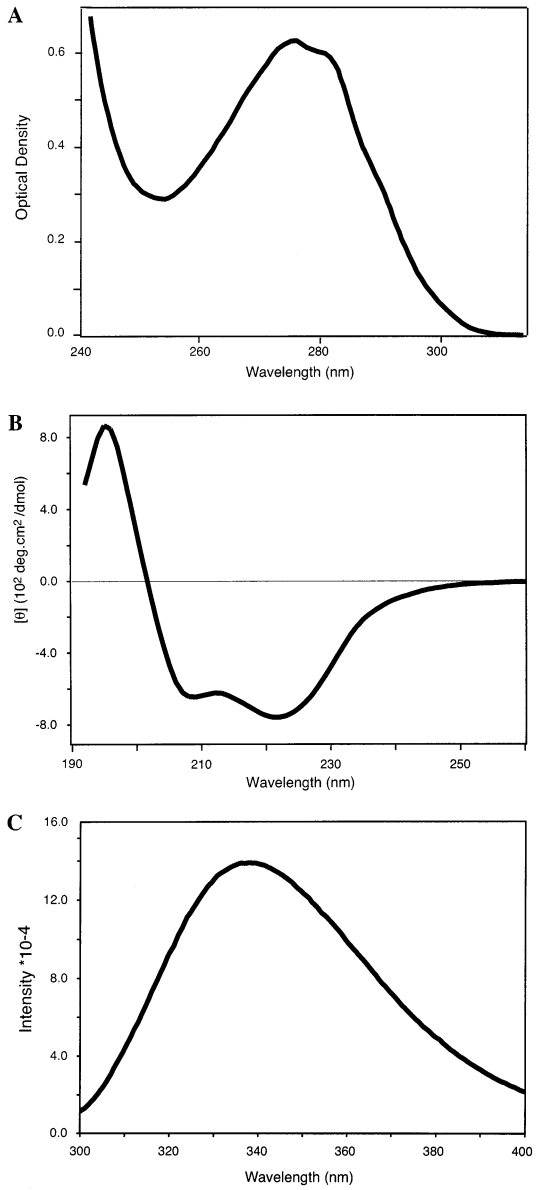

FIGURE 4: Optical characteristics of the E6-C 4C/4S domain: (A) UV absorption spectrum (10 μM), (B) circular dichroism spectrum (10 μM), and (C) UV fluorescence emission spectrum (excitation wavelength of 280 nm) (0.5 μM).

C-terminal unstructured regions (*55*). Our results therefore suggest that E6 6C/6S consists mainly of two folded zinc-binding domains corresponding to residues 7−83 and 87−153, connected by a four-residue linker (residues 83−86) and flanked by an N-terminal loop (residues 1−6) and a C-terminal loop (residues 154−158). This result confirms the recent work by Lipari et al. (*40*) which also suggested a two-domain substructure for E6 proteins. Therefore, the familiar division of the E6 sequence into five regions (*41*−*43*) is probably inaccurate in terms of three-dimensional structure. Remarkably, the circular dichroism spectrum of the E6-C 4C/4S domain is very similar to that previously obtained for the E6-N domain (*40*), and back-calculations from both spectra give comparable estimates of secondary structure content for both domains. This supports previous suggestions, based on sequence alignments, that the two domains have a similar fold and originate from duplication of a common structural ancestor (*35*).

E6 proteins and E6 domains have long resisted expression and purification attempts. One of the reasons for this might be their anomalously high cysteine content (*45*, *56*). We have observed that whenever E6 and E6 fragments are handled without extreme control of redox conditions, aggregation and precipitation systematically occur. It is this particularity of E6 proteins which led us to design the E6 6C/6S mutant, in which six nonconserved cysteine residues were changed into serines (*45*). The present E6-C 4C/4S construct is derived from E6 6C/6S mutant and therefore contains only four cysteines instead of the eight cysteines present in the corresponding region of wild-type E6. The four remaining cysteines of E6-C 4C/4S are all implicated in zinc binding, which should inhibit their ability to form disulfide bridges. Despite this fact, the purification of E6-C 4C/4S again required particular care against oxidation, as indicated in Experimental Procedures. In this regard, it is also interesting to compare our results with the work of Lipari et al. (*40*). By analyzing the proteolysis products of a "minimal core construct" of wild-type HPV16 E6, the authors obtained only one stable peptide corresponding to residues 9−84. This peptide (called E6-N) is nearly identical to our N-terminal peptide 1 (residues 7−83 of E6 6C/6S). In their report, Lipari et al. mentioned their failure to detect a complementary C-terminal peptide, which should have been situated between residues 85 and 150. The authors deduced that the C-terminal half of E6 was "unstable or insoluble" and contributed to the instability and insolubility of full-length E6. In contrast, our C-terminal E6-C 4C/4S construct (residues 87−158) is folded and highly soluble. It therefore appears that the four cysteine-to-serine mutations present in our E6-C 4C/4S construct have dramatically improved the biochemical stability, probably by decreasing the Cys-bonding reactivity of the domain. In fact, the better results obtained by Lipari et al. for the E6-N domain compared to the E6-C domain may also be explained by differences in Cys-bonding reactivity. Six cysteine residues are present in wild-type HPV16 E6-N compared to eight in E6-C. If we take into account the four zinc-binding cysteines of each domain, we are left with two free cysteines in E6-N against four free cysteines in E6-C. Therefore, two molecules of wild-type E6-N can in principle establish four distinct intermolecular Cys−Cys bonds, whereas two molecules of wild-type E6-C can form 16 distinct intermolecular Cys−Cys bonds, thereby increasing the likelihood of oligomerization leading to precipitation. It is also possible that one or several cysteines in E6-C display anomalously high Cys-bonding reactivity promoted by a particular chemical environment in the folded domain.

Our NMR diffusion data suggest that E6-C 4C/4S is primarily monomeric at NMR concentrations (1 mM). Lipari et al. (*40*) also have found that the minimal core of wild-type HPV16 (residues 9−149) and the E6-N domain (residues 9−84) were monomeric at physiological concentrations (i.e., 1−10 μM). In contrast, older works often suggested that E6 proteins form dimers or higher-order oligomers (*57*, *58*). Considering the sensitivity of E6 proteins to oxidation, we think that reported E6 dimers or oligomers were probably associated via cysteine bonds, and that this was a purification artifact. We have recently studied by immunogold electron microscopy the intracellular localization of wild-type HPV16 E6 protein in HPV-positive cervical

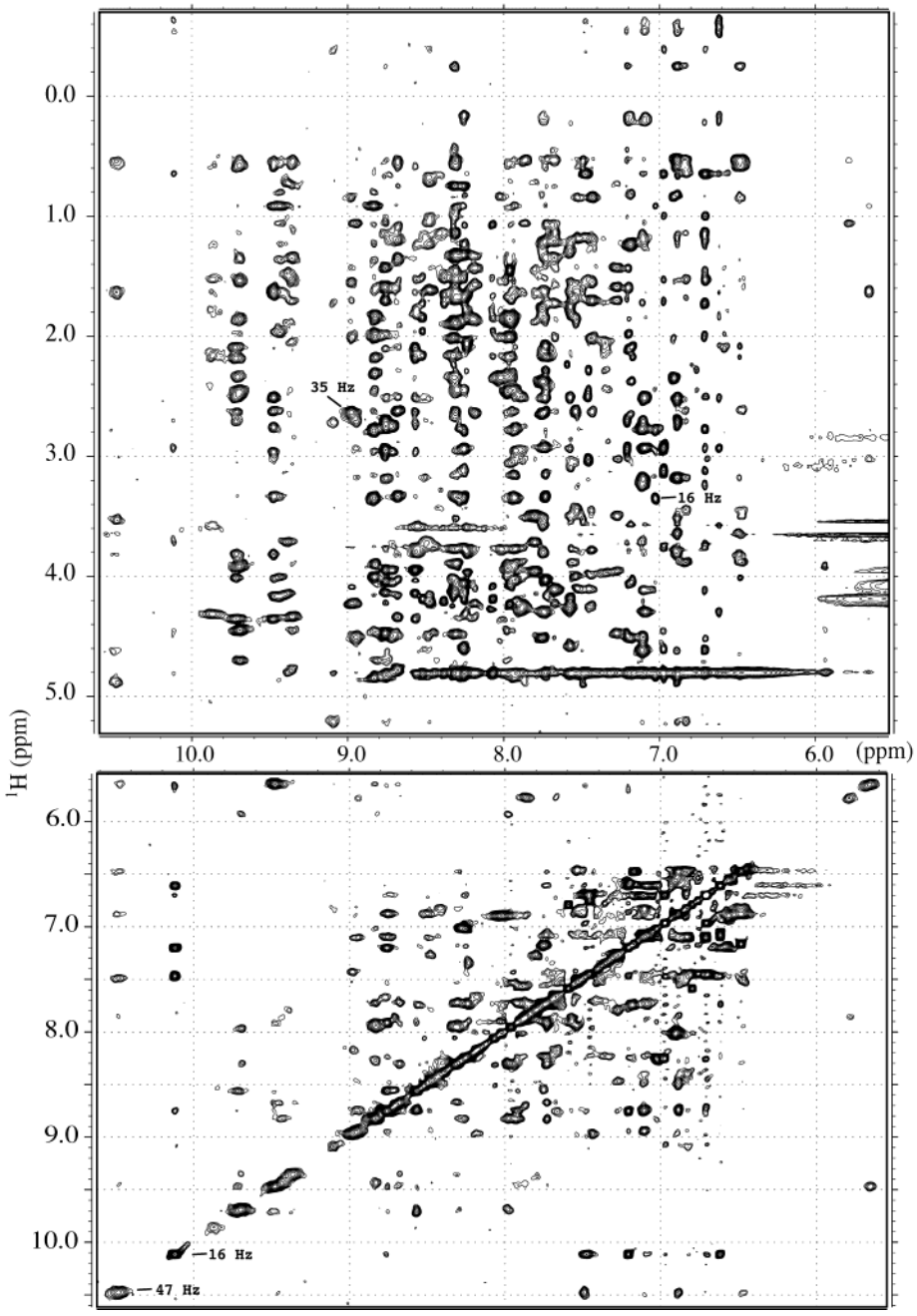

FIGURE 5: 2D NOESY spectrum of E6-C 4C/4S. Amide−aliphatic region of a 600 MHz NOESY spectrum recorded on a 1 mM sample of E6-C 4C/4S at pH 6.8 and 15 °C. The line widths of four NOE cross-peaks are indicated to illustrate the heterogeneous relaxation behavior along the sequence.

carcinoma cells (M. Masson et al., manuscript submitted to publication). E6 is predominantly found in the nucleus, but a fraction of the protein is also found in cytoplasm. These two compartments are reducing environments which cannot promote E6 oligomerization via cysteine bond formation.

The E6-C 4C/4S domain appears as the minimal autonomous folding unit retaining the DNA junction recognition activity of full-length E6. In a previous paper, we demonstrated that this activity was also retained by the MBP−ZD2 construct (*33*). The ZD2 fragment also includes the C-terminal zinc-binding motif, but it lacks 14 N-terminal residues (S_{87}YSLYGTTLEQQYN$_{100}$) compared to E6-C 4C/4S. In contrast to E6-C 4C/4S, the ZD2 construct binds DNA only in the presence of the MBP carrier, and ZD2 is not stable after proteolytic separation from MBP. Therefore, the $S_{87}-N_{100}$ region is not necessary for DNA binding, but it must contain secondary structure elements essential for the folding and/or stability of the domain. E6-C 4C/4S and ZD2 also differ in their DNA binding stoichiometry. The natural stoichiometry of full-length E6−DNA complexes is 1:1 (*32*). We have shown here that E6-C 4C/4S also binds junctions as a single monomer, consistent with the behavior of full-length E6. In contrast, two monomers of the MBP−ZD2 construct bind to one junction on independent and equivalent sites (*33*). Thus, when one monomer of E6 4C/4S is bound to DNA, a second interaction site should be available on the junction for addition of a second monomer. However, access to this site is probably masked because of steric hin-

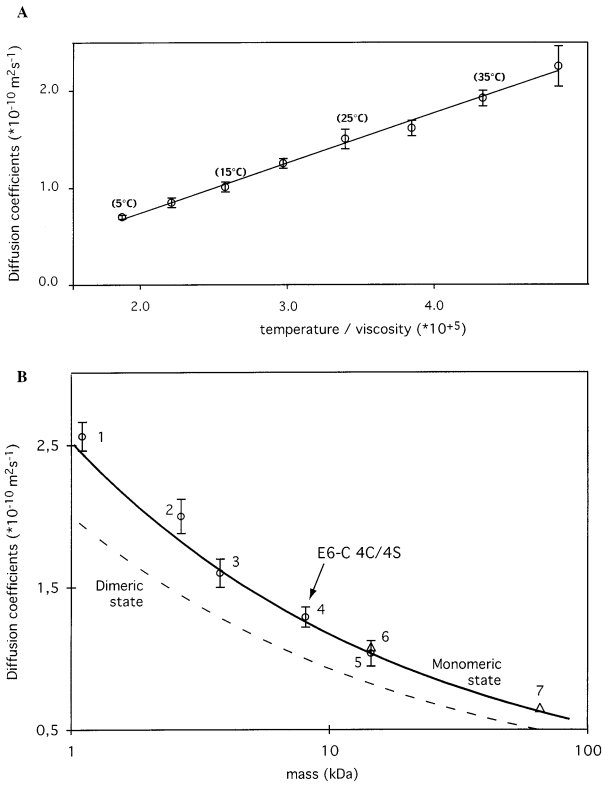

FIGURE 6: Diffusion of E6-C 4C/4S as measured using NMR pulsed field gradients. (A) Temperature dependence of the diffusion coefficients of E6-C 4C/4S from 5 to 40 °C. The linear fit of the data points is shown by a solid line, and the correlation coefficient of the fit is 0.996. (B) Comparison of normalized D_t values (20 °C in water) measured for proteins of various sizes. The theoretical mass-dependent diffusion value of a sphere is shown by a solid line using the measured lysosyme diffusion coefficient as a reference for a monomeric protein. The dotted line corresponds to diffusion coefficients that would be obtained for dimers. Circles and triangles indicate values deduced from our experiments (unpublished results) and values published by other groups, respectively. The peptides shown on the figure are (1) G4R4 (1100 Da), (2) conotoxin MVIIA (2646 Da), (3) PMPD2 (3750 Da), (4) E6-C 4C/4S (8630 Da), (5) lysosyme (14 500 Da), (6) lysosyme [measured by Haner et al. (*52*)], and (7) human hemoglobin (65 500 Da) [measured by Everhart et al. (*53*)].

drance by the $S_{87}-N_{100}$ region. This situation of indirect steric hindrance has already been raised to interpret monomeric binding of HMG box proteins to four-way junctions (*59*).

Our NMR data have indicated heterogeneous line broadening probably due to the occurrence of slow motions (nano- to millisecond time scale) within localized regions of the monomeric domain. The occurrence of such motions has been reported for several DNA-binding proteins (*60*−*62*). New NMR approaches have recently been developed for detailed investigation of protein motions (*62*). It will be interesting to apply such methods in analyzing the dynamic behavior of E6 and E6 domains, and in studying the possible biological relevance of these motions. More generally, these data constitute a definitive step toward obtaining structural information about papillomavirus E6 protein. We have delimited the C-terminal domain of E6 and have produced samples allowing biophysical analysis by a variety of methods, including NMR. Completion of the solution structure of the E6-C 4C/4S domain will provide precious information about the molecular basis of E6 activities.

ACKNOWLEDGMENT

We are grateful to Claude Ling for his excellent management of the computing and NMR facilities, to Dr. Adrien Staub for chemical N-terminal sequencing of the peptides, and to Dr. Andrew Atkinson for critical reading of the manuscript.

REFERENCES

1. zur Hausen, H. (1991) *Virology 184*, 9−13.
2. Scheffner, M., Werness, B. A., Huibregtse, J. M., Levine, A. J., and Howley, P. M. (1990) *Cell 63*, 1129−1136.
3. Scheffner, M., Huibregtse, J. M., Vierstra, R. D., and Howley, P. M. (1993) *Cell 75*, 495−505.
4. Patel, D., Huang, S. M., Baglia, L. A., and McCance, D. J. (1999) *EMBO J. 18*, 5061−5072.
5. Zimmermann, H., Degenkolbe, R., Bernard, H. U., and O'Connor, M. J. (1999) *J. Virol. 73*, 6209−6219.
6. 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.
7. Gross-Mesilaty, S., Reinstein, E., Bercovich, B., Tobias, K., Schwartz, A., Kahana, C., and Ciechanover, A. (1998) *Proc. Natl. Acad. Sci. U.S.A. 95*, 8058−8063.
8. Ronco, L. V., Karpova, A. Y., Vidal, M., and Howley, P. M. (1998) *Genes Dev. 12*, 2061−2072.
9. Kukimoto, I., Aihara, S., Yoshiike, K., and Kanda, T. (1998) *Biochem. Biophys. Res. Commun. 249*, 258−262.
10. Srivenugopal, K., and Ali-Osman, F. (2002) *Oncogene 21*, 5940−5945.
11. Iftner, T., Elbel, M., Schopp, B., Hiller, T., Loizou, J., Caldecott, K., and Stubenrauch, F. (2002) *EMBO J. 21*, 4741−4748.
12. Gao, Q., Kumar, A., Srinivasan, S., Singh, L., Mukai, H., Ono, Y., Wazer, D. E., and Band, V. (2000) *J. Biol. Chem. 275*, 14824−14830.
13. Li, S., Labrecque, S., Gauzzi, M. C., Cuddihy, A. R., Wong, A. H., Pellegrini, S., Matlashewski, G. J., and Koromilas, A. E. (1999) *Oncogene 18*, 5727−5737.
14. Gao, Q., Srinivasan, S., Boyer, S. N., Wazer, D. E., and Band, V. (1999) *Mol. Cell. Biol. 19*, 733−744.
15. Filippova, M., Song, H., Connolly, J., Dermody, T., and Duerksen-Hughes, P. (2002) *J. Biol. Chem. 277*, 21730−21739.
16. Thomas, M., and Banks, L. (1998) *Oncogene 17*, 2943−2954.
17. Tong, X., Boll, W., Kirchhausen, T., and Howley, P. M. (1998) *J. Virol. 72*, 476−482.
18. Tong, X., and Howley, P. M. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 4412−4417.
19. Vande Pol, S. B., Brown, M. C., and Turner, C. E. (1998) *Oncogene 16*, 43−52.
20. Chen, J., Reid, C. E., Band, V., and Androphy, E. J. (1995) *Science 269*, 529−532.
21. Du, M., Fan, X., Hong, E., and Chen, J. (2002) *Biochem. Biophys. Res. Commun. 296*, 962−969.
22. Kiyono, T., Hiraiwa, A., Fujita, M., Hayashi, Y., Akiyama, T., and Ishibashi, M. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 11612−11616.
23. Nakagawa, S., and Huibregtse, J. M. (2000) *Mol. Cell. Biol. 20*, 8244−8253.
24. Glaunsinger, B., Lee, S., Thomas, M., Banks, L., and Javier, R. (2000) *Oncogene 19*, 5270−5280.
25. Lee, S., Glaunsinger, B., Mantovani, F., Banks, L., and Javier, R. (2000) *J. Virol. 74*, 9680−9693.
26. Sedman, S. A., Barbosa, M. S., Vaa, W. C., Hubbert, N. L., Haas, J. A., Lowy, D. R., and Schiller, J. T. (1991) *J. Virol. 65*, 4860−4866.
27. Morosow, A., Phelps, W. C., and Raychaudhuri, P. (1994) *J. Biol. Chem. 269*, 18434−18440.
28. Dey, A., Atcha, I. A., and Bagchi, S. (1997) *Virology 228* (2), 190−199.
29. Gewin, L., and Galloway, D. A. (2001) *J. Virol. 75*, 7198−7201.
30. Veldman, T., Horikawa, I., Barrett, J. C., and Schlegel, R. (2001) *J. Virol. 75*, 4467−4472.
31. Oh, S. T., Kyo, S., and Laimins, L. A. (2001) *J. Virol. 75*, 5559−5566.

32. Ristriani, T., Masson, M., Nominé, Y., Laurent, C., Lefèvre, J. F., Weiss, E., and Travé, G. (2000) *J. Mol. Biol. 296*, 1189−1203.
33. Ristriani, T., Nomine, Y., Masson, M., Weiss, E., and Trave, G. (2001) *J. Mol. Biol. 305*, 729−739.
34. Androphy, E. J., Hubbert, N. L., Schiller, J. T., and Lowy, D. R. (1987) *EMBO J. 6*, 989−992.
35. Cole, S. T., and Danos, O. (1987) *J. Mol. Biol. 193*, 599−608.
36. Grossman, S. R., and Laimins, L. A. (1989) *Oncogene 4*, 1089−1093.
37. Barbosa, M. S., Lowy, D. R., and Schiller, J. T. (1989) *J. Virol. 63*, 1404−1407.
38. Kanda, T., Watanabe, S., Zanma, S., Sato, H., Furuno, A., and Yoshiike, K. (1991) *Virology 185*, 536−543.
39. Lipari, F., McGibbon, G. A., Wardrop, E., and Cordingley, M. G. (2001) *Biochemistry 40*, 1196−1204.
40. Foster, S. A., Demers, G. W., Etscheid, B. G., and Galloway, D. A. (1994) *J. Virol. 68*, 5698−5705.
41. Pim, D., Storey, A., Thomas, M., Massimi, P., and Banks, L. (1994) *Oncogene 9*, 1869−1876.
42. Thomas, M., Pim, D., and Banks, L. (1999) *Oncogene 18*, 7690−7700.
43. Nominé, Y., Ristriani, T., Laurent, C., Lefèvre, J. F., Weiss, E., and Travé, G. (2001) *Protein Expression Purif. 23*, 22−32.
44. Nominé, Y., Ristriani, T., Laurent, C., Lefèvre, J. F., Weiss, E., and Travé, G. (2001) *Protein Eng. 14*, 297−305.
45. Johnson, W. C. (1999) *Proteins 35*, 307−312.
46. Piotto, M., and Saudek, V. (1992) *J. Biomol. NMR 2*, 661−665.
47. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) *J. Biol. NMR 6*, 277−293.
48. Bartels, C., Xia, T. H., Billeter, M., Güntert, P., and Wüthrich, K. (1995) *J. Biomol. NMR 5*, 1−10.
49. Gibbs, S. J., Morris, K. F., and Johnson, C. S., Jr. (1991) *J. Magn. Reson. 94*, 165−169.
50. 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.
51. Haner, R., and Schleich, T. (1989) *Methods Enzymol. 176*, 418−446.
52. Everhart, C. H., and Johnson, C. S. J. (1982) *J. Magn. Reson. 50*, 466−474.
53. Duckett, D. R., Murchie, A. I. H., Diekmann, S., Von Kitzing, E., Kemper, B., and Lilley, D. M. J. (1988) *Cell 55*, 79−89.
54. Pisani, F. M., Manco, G., Carratore, V., and Rossi, M. (1996) *Biochemistry 35*, 9158−9166.
55. Dell, G., and Gaston, K. (2001) *Cell. Mol. Life Sci. 58*, 1923−1942.
56. Pim, D., Massimi, P., and Banks, L. (1997) *Oncogene 15*, 157−264.
57. Daniels, P. R., Sanders, C. M., Coulson, P., and Maitland, N. J. (1997) *FEBS Lett. 416*, 6−10.
58. Pohler, J. R. G., Norman, D. G., Brahmam, J., Bianchi, M. E., and Lilley, D. M. J. (1998) *EMBO J. 17*, 817−826.
59. Zhu, L., Hu, J., Lin, D., Whitson, R., Itakura, K., and Chen, Y. (2001) *Biochemistry 40*, 9142−9150.
60. Ramboarina, S., Srividya, N., Atkinson, R. A., Morellet, N., Roques, B. P., Lefevre, J. F., Mely, Y., and Kieffer, B. (2002) *J. Mol. Biol. 316*, 611−627.
61. McIntosh, P., Taylor, I., Frenkiel, T., Smerdon, S., and An, L. (2000) *J. Biomol. NMR 16*, 183−196.
62. Cavanagh, J., and Venters, R. A. (2001) *Nat. Struct. Biol. 8*, 912.

BI026980C