Protein Science (2000), *9*:799–811. Cambridge University Press. Printed in the USA.

Folding of a dimeric β-barrel: Residual structure in the urea denatured state of the human papillomavirus E2 DNA binding domain

YU-KEUNG MOK,[2,4] LEONARDO G. ALONSO,[1] L. MAURICIO T.R. LIMA,[3] MARK BYCROFT,[2] AND GONZALO DE PRAT-GAY[1]

[1]Instituto de Investigaciones Bioquímicas, Fundación Campomar, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Patricias Argentinas 435, (1405) Buenos Aires, Argentina
[2]MRC Unit for Protein Function and Design, Cambridge University Chemical Laboratory, Lensfield Road, Cambridge CB2 1EW, United Kingdom
[3]Departamento de Bioquimica Medica, Universidade Federal do Rio de Janeiro, Cidade Universitaria, Rio de Janeiro 21914-590, Rio de Janeiro, Brazil

(RECEIVED August 10, 1999; FINAL REVISION December 9, 1999; ACCEPTED February 18, 2000)

Abstract

The dimeric β-barrel is a characteristic topology initially found in the transcriptional regulatory domain of the E2 DNA binding domain from papillomaviruses. We have previously described the kinetic folding mechanism of the human HPV-16 domain, and, as part of these studies, we present a structural characterization of the urea-denatured state of the protein. We have obtained a set of chemical shift assignments for the C-terminal domain in urea using heteronuclear NMR methods and found regions with persistent residual structure. Based on chemical shift deviations from random coil values, $^3J_{NHN\alpha}$ coupling constants, heteronuclear single quantum coherence peak intensities, and nuclear Overhauser effect data, we have determined clusters of residual structure in regions corresponding to the DNA binding helix and the second β-strand in the folded conformation. Most of the structures found are of nonnative nature, including turn-like conformations. Urea denaturation at equilibrium displayed a loss in protein concentration dependence, in absolute parallel to a similar deviation observed in the folding rate constant from kinetic experiments. These results strongly suggest an alternative folding pathway in which a dimeric intermediate is formed and the rate-limiting step becomes first order at high protein concentrations. The structural elements found in the denatured state would collide to yield productive interactions, establishing an intermolecular folding nucleus at high protein concentrations. We discuss our results in terms of the folding mechanism of this particular topology in an attempt to contribute to a better understanding of the folding of dimers in general and intertwined dimeric proteins such as transcription factors in particular.

Keywords: dimer; DNA binding; E2; folding; urea-denatured state

The full experimental description of protein folding pathways aims at the characterization of the different species and events involved, i.e., unfolded states, intermediates and transition states, and the structure of the native-folded protein (Arcus et al., 1994). In aqueous and native buffer conditions, fully folded proteins are in equilibrium with their denatured states, only the latter represent a minimal population. Nevertheless, this equilibrium dictates the conformational pathway and stability of proteins, and the study of denatured states is imperative for understanding the general mechanistic and thermodynamic aspects of protein folding, in particular those related to early events.

We are interested in the folding mechanism of the DNA binding and dimerization domain of the transcriptional regulatory E2 protein from human papillomavirus (HPV) strain-16 (Hawley-Nelson et al., 1988; McBride et al., 1989). The dimeric E2 protein binds the 12 bp palindromic sequence, $ACCN_6GGT$, via its C-terminal DNA binding domain (Hawley-Nelson et al., 1988) and shows high sequence homology to its bovine papillomavirus-1 (BPV-1) counterpart and that of HPV-31, a closely related human strain. The crystal structure of the BPV-1 E2 DNA binding domain bound

Reprint requests to: Gonzalo de Prat-Gay, Instituto de Investigaciones Bioquímicas, Fundación Campomar, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Patricias Argentinas 435, (1405) Buenos Aires, Argentina; e-mail pratgay@iib.uba.ar.

[4]Present address: Department of Biochemistry, The Hong Kong University of Science and Technology, Clear Water Bay, Sai Hung, Hong Kong.

Abbreviations: CD, circular dichroism; DTT, dithiothreitol; HMQC, heteronuclear multiple quantum coherence; HPV, human papillomavirus; HSQC, heteronuclear single quantum coherence; IPTG, isopropyl-β-D-thiogalactopyranoside; NOE, nuclear Overhauser effect; NOESY, NOE spectroscopy; OBP, origin binding protein; TOCSY, total correlation spectroscopy.

to a DNA oligonucleotide revealed a novel folding topology termed the dimeric β-barrel (Hegde et al., 1992), and crystal structures of the free bovine and human domains were recently solved (Hegde et al., 1998; Hegde & Androphy, 1998). This peculiar DNA binding fold consists of an eight-stranded (four per subunit) β-barrel, with a major DNA binding and minor α-helices, packed against opposite sides of the barrel (Hegde et al., 1992). The majority of small DNA binding and transcriptional regulatory proteins that bind palindromic DNA sequences are dimers with a highly intertwined dimeric interface (Travers, 1993).

We have described the equilibrium dissociation and unfolding of the HPV-16 E2 DNA binding domain (E2-C) (Mok et al., 1996a), which was shown to follow a two-state mechanism with only the native dimer or unfolded monomer populated across the unfolding transition. The E2-C dimeric domain was largely stabilized by its cognate DNA (Lima & Prat Gay, 1997). In kinetic refolding experiments, we found that E2-C folds through the formation of a monomeric intermediate, which appears to be nonnative like (Mok et al., 1996b). This raised a number of questions regarding the structural nature of the transient intermediate and any related residual structures that precede its formation, i.e., in the denatured state.

NMR proved to be an excellent tool for studying the residual structure in unfolded states of proteins (Evans et al., 1991; Shortle, 1993; Wüthrich, 1994). Most of the best-described protein folding models correspond to small monomeric proteins; similarly, the unfolded states of proteins that have been studied in detail by NMR techniques are mostly monomeric. These include the 434 repressor (Neri et al., 1992a, 1992b, 1992c), lysozyme (Buck et al., 1995), barnase (Arcus et al., 1995), FK506 (Logan et al., 1994), and drk SH3 domain (Zhang et al., 1997). Less information is available on denatured states and folding mechanisms of dimers. We have undertaken a structural characterization of the unfolded state of E2-C from HPV-16, mainly by heteronuclear NMR techniques, complemented by fluorescence and circular dichroism (CD) spectroscopy. We discuss our findings in the light of the overall kinetic folding mechanism (Mok et al., 1996b).

Materials and methods

Sample preparation

Recombinant HPV-16 E2-C used in fluorescence and CD experiments was purified from culture of BL21 (DE3) pLysS cells containing the plasmid ptz18U-E2 as described previously (Mok et al., 1996a). Uniformly ^{15}N-labeled and uniformly ^{15}N–^{13}C-labeled proteins were expressed by transforming the plasmid ptz18U-E2 into *Escherichia coli* JM109 strain. Overnight cultures of the cells in M9 minimal medium containing ^{13}C glucose were inoculated (1%) into M9 medium containing $^{15}NH_4Cl$ and ^{13}C glycerol as the sole nitrogen and carbon sources. The cells were grown at 37 °C to an $O.D._{600}$ of 0.35. Protein expression was induced by adding 1.5% (v/v) of M13 phage carrying the T7 polymerase gene (1×10^{11} pfu/mL, Invitrogen, San Diego, California) and 0.2 mM IPTG. The cells were grown overnight at 37 °C before harvesting. The purification procedure for the labeled protein is the same as for that grown in rich medium (Mok et al., 1996a).

The unfolded E2-C sample for NMR analysis was prepared by concentrating the protein in water to a final concentration of around 2 mM and volume of 0.5 mL using an Amicon Centriprep 10 concentrator. The concentrated protein was lyophilized and then dissolved in 0.5 mLof buffer containing 3.0 M deuterated urea (Sigma, St. Louis, Missouri) and 50 mM deuterated sodium acetate, pD 5.6 in 10% D_2O. The concentrations for both the uniformly ^{15}N labeled and ^{15}N/^{13}C labeled sample were 2 mM.

Unfolding experiments

Fluorescence spectra were acquired on a Aminco Bowman Series 2 Spectrofluorimeter (Spectronic, Rochester, New York). Samples containing 1 μM HPV-16 E2-C were measured in folding and unfolding conditions to probe the environment of the Trp residues and overall degree of unfolding by different denaturants. To assess the spectral shifts arising from unfolding, the center of mass for each spectra were calculated (Foguel et al., 1998). Urea unfolding experiments were performed in 50 mM sodium acetate buffer, pH 5.6, using protein concentrations indicated in the caption for Figure 1. Due to the large difference in fluorescence intensity at different urea concentrations, the degree of unfolding by urea was followed by calculating the center of mass of the fluorescence spectra at each urea concentration. CD spectra were acquired using a Jasco 720 spectropolarimeter with a 0.1 m path quartz cuvette, at 25 °C. Ten spectra were averaged for each represented spectrum.

NMR spectroscopy

All NMR spectra were recorded on a Bruker AMX 500 spectrometer equipped with a pulsed-field gradient device and a triple-resonance ^{1}H/^{15}N/^{13}C probe optimized for proton detection. The data were converted and processed using the FELIX software (Version 2.3, Biosym Technologies, San Diego, California) on a Silicon Graphics Indy workstation. Proton chemical shifts are referenced to internal sodium 3-(trimethylsilyl)-d^4-propionate, ^{15}N chemical shifts are referenced to external ^{15}N NH_4Cl and ^{13}C chemical shifts are indirectly referenced to the solvent ^{1}H frequency (Bax & Subramanian, 1986). Unless indicated, experiments were carried out at 25 °C.

1D ^{1}H-NMR spectra

One-dimensional (1D) NMR spectra of E2-C at different concentrations of urea were acquired. The spectra were recorded with 4,096 complex points over a spectral width of 8,064.52 Hz. A total of 128 transients were collected. Water suppression was achieved by on-resonance presaturation of the water signal.

Heteronuclear ^{15}N–^{1}H NMR spectroscopy

Two-dimensional (2D) ^{1}H–^{15}N HSQC spectrum (Bodenhausen & Ruben, 1980; Bax et al., 1990) was acquired with 2,048 complex data points in t_2 and 256 t_1 increments. Spectral width were 6,578.95 Hz in t_2 and 2,000 Hz in t_1. Sixty-four transients were recorded per increment. The experimental data were zero filled to 8,192 real data points in a final matrix of size $1{,}024 \times 512$. The data were processed using an exponential window function in t_2 and a shifted ($\pi/3$) sine-bell function in t_1 prior to Fourier transformation. Water suppression was achieved using a spin-lock sequence (Messerle et al., 1989). The water signal was additionally suppressed during data processing by convolution of the t_2 time domain data (Marion et al., 1989b).

Three-dimensional (3D) ^{1}H–^{15}N TOCSY-HMQC and 3D ^{1}H–^{15}N NOESY-HMQC spectra (Marion et al., 1989a) were acquired with the ^{15}N–^{13}C labeled sample. The NOESY spectrum was mod-

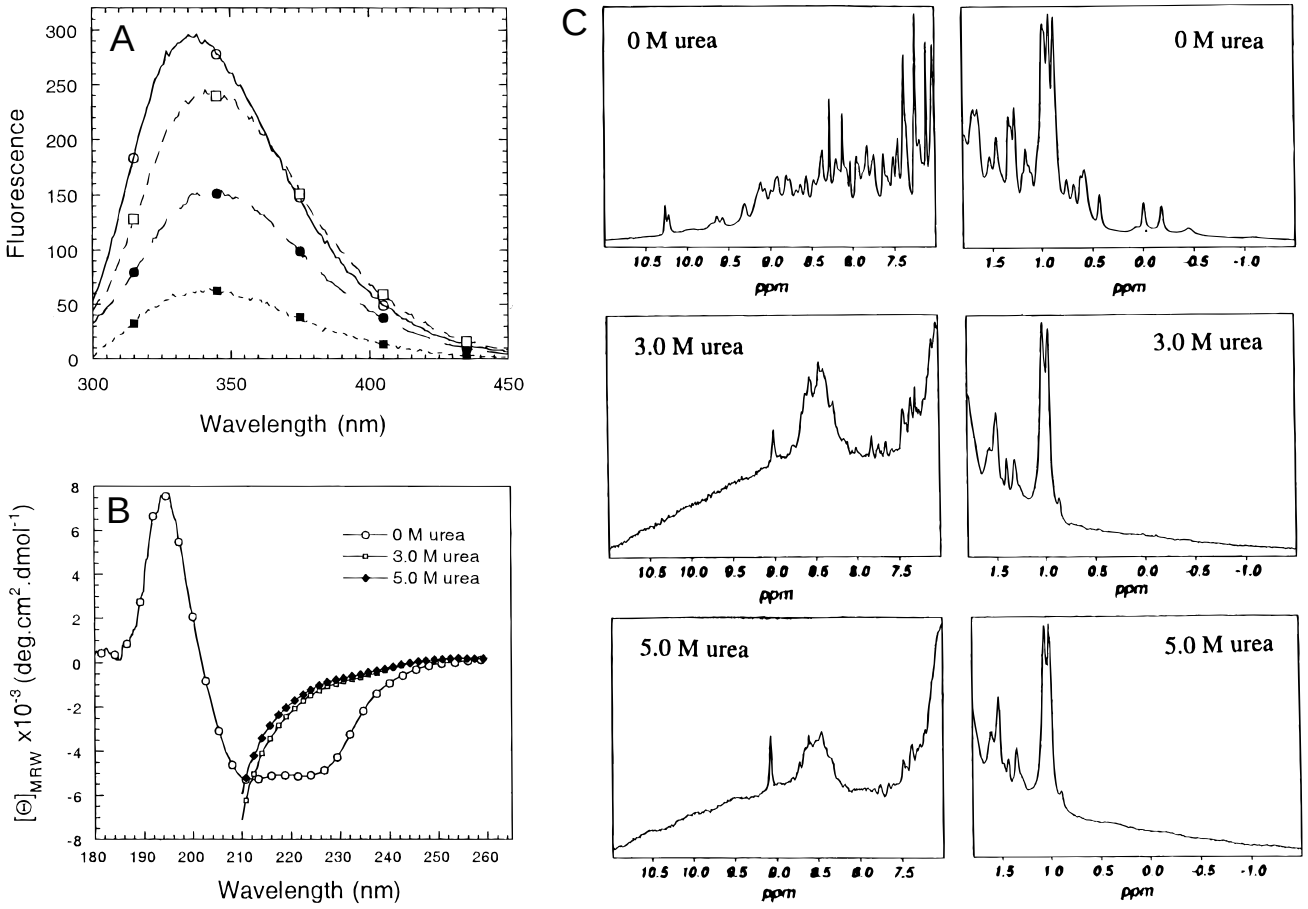

Fig. 1. A: Fluorescence spectra of different denatured states of E2-C. E2-C (1 μM) in: (○) 50 mM sodium acetate pH 5.5 at 25 °C, center of spectral mass (CM) 28,746 cm^{-1}; (□) 50 mM sodium acetate pH 5.5 at 25 °C plus 3M urea, CM 28,365 cm^{-1}; (●) 50 mM HCl, pH 1.5 at 25 °C, CM 28,375 cm^{-1}; (■) 50 mM sodium acetate pH 5.5 at 70 °C, CM 28,619 cm^{-1}. **B:** Far-UV CD spectra of the native and denatured HPV-16 E2 DNA binding domain in 3.0 and 5.0 M urea. Protein concentration was 20 μM in 50 mM sodium acetate buffer, pH 5.6 with 1 mM DTT. **C:** Effect of urea on the amide and methyl regions of the 1D proton NMR spectrum of HPV-16 E2 DNA binding domain investigated at 0.0, 3.0, and 5.0 M urea. Samples of 2 mM HPV-16 E2 DNA binding domain in 50 mM sodium acetate buffer, pH 5.6 were used for NMR spectroscopy at 25 °C and the specified urea concentrations.

ified with gradient selection for enhanced sensitivity and water suppression (Kay et al., 1992). Both spectra were acquired with 1,024 and 196 complex data points in t_3 and t_1 dimensions over a spectral width of 7,042.25 and 5,998.54 Hz, respectively. The t_2 dimensions were acquired with 64 and 62 increments covering 1,750 and 1,773 Hz in the NOESY and TOCSY experiment, respectively. Quadrature detection was achieved in t_2 and t_1 dimension of the TOCSY experiment by the TPPI method (Marion & Wüthrich, 1983). In the NOESY experiment, TPPI-States (Marion et al., 1989c) and TPPI was used in the t_2 and t_1 dimension, respectively. The TOCSY and NOESY mixing times were 73.35 and 300 ms, respectively. The number of transients acquired per t_1/t_2 increment were 16 in both the TOCSY and NOESY experiments. Solvent suppression in the TOCSY spectrum was achieved by the gradient sequence WATERGATE (Sklenar et al., 1993). The 3D TOCSY and NOESY spectra were zero filled to give a final matrix size of 512 × 256 × 128 real data points. Except in the acquired proton dimension of the NOESY spectrum, where a Lorentzian–Gaussian window function was applied, $\pi/3$ shifted sine-bell were applied to all the other dimensions of both spectra.

Triple resonance NMR spectroscopy

Two pairs of complementary triple resonance experiments were used for the sequential assignment of the unfolded state of the HPV-16 E2 DNA binding domain. The CBCA(CO)NH experiment, which correlates the amide proton to the $^{13}C\beta$ and $^{13}C\alpha$ resonances of the preceding residue, was modified with gradient selection for enhanced sensitivity and water suppression (Kay et al., 1992). The HNCA experiment (Kay et al., 1990) is a complementary experiment to CBCA(CO)NH and correlated the amide proton to the $^{13}C\alpha$ of its own spin system, and in some cases, if the coupling constant $J_{NHiC\alpha(i-1)}$ is large enough, with the $^{13}C\alpha$ of the preceding residue. In the acquired proton dimension (t_3), 2,048 and 1,024 complex points were acquired in the CBCA(CO)NH and HNCA experiment, respectively, both over a spectral width frequency of 6,024.1 Hz. In the ^{15}N dimension (t_1), 64 complex points were acquired over a spectral width frequency of 1,572.32 Hz in both experiments. In the CBCA(CO)NH experiment, 116 complex points were acquired in the ^{13}C dimension (t_2) over a spectral width of 8,799.04 Hz covering 70 ppm. In the HNCA experiment,

128 complex points were acquired in the ^{13}C dimension (t_2) over a spectral width of 3,771.02 Hz covering 30 ppm. Except in the t_2 dimension of the HNCA experiment, where TPPI was used, TPPI-States quadrature detection was used in all the other t_1 and t_2 dimensions of both spectra. Solvent suppression was achieved by the WATERGATE sequence in the HNCA experiment. Thirty-two transients were acquired per increment for both experiments. The spectra were zero filled to $512 \times 256 \times 128$ real points. The ^{15}N dimension in both spectra was linear predicted to 40 real data points before transformation. A weak Lorentzian–Gaussian window function was applied to the acquired proton dimensions, while $\pi/3$ shifted sine-bell functions were applied to the other dimensions of both spectra.

The HNCO spectrum, which correlates the amide proton to the carbonyl ^{13}C resonance of the preceding residue, was modified with gradient selection for enhanced sensitivity and water suppression (Muhandiram & Kay, 1994). The HN(CA)CO experiments (Clubb et al., 1992) correlates the amide proton with the carbonyl ^{13}C resonance of the same residue. In the acquired proton dimension (t_3), 2,048 and 1,024 complex points were acquired in the HNCO and HN(CA)CO experiment, respectively, both over a spectral width frequency of 6,024.1 Hz. In the ^{15}N dimension (t_1), 64 and 56 complex points were acquired in the HNCO and HN(CA)CO experiment, respectively, both over a spectral width frequency of 1,572.32 Hz. In the HNCO experiment, 128 complex points were acquired in the ^{13}C carbonyl dimension (t_2) over a spectral width of 1,256.90 Hz covering 10 ppm. In the HN(CA)CO experiment, 64 complex points were acquired in the ^{13}C carbonyl dimension (t_2) over a spectral width of 1,282.05 Hz covering 10.2 ppm. Except in the t_2 dimension of the HNCO experiment, where TPPI was used, TPPI-States quadrature detection was used in all the other t_1 and t_2 dimensions of both spectra. Sixteen and 64 transients per increment were acquired in the HNCO and HN(CA)CO experiment, respectively. Constant time evolution (Grzesiek & Bax, 1992) was used in the ^{15}N dimension of the HN(CA)CO experiment, which allowed mirror image linear prediction (Zhu & Bax, 1992). The spectra were zero filled to $512 \times 256 \times 128$ real data points. The ^{15}N dimension in the HNCO spectrum was linear predicted to 40 real data points before transformation. A Lorentzian–Gaussian window function was applied to the acquired proton dimensions, while $\pi/3$ shifted sine-bells were applied to the other dimensions of both spectra.

$^3J_{HNH\alpha}$ coupling constants

The $^3J_{HNH\alpha}$ coupling constant was measured by a method described in Stonehouse and Keeler (1995), based on the splitting of the amide proton cross peak in the HSQC spectrum caused by coupling between the amide proton and Hα. The digital resolution of the amide region was increased by zero filling to 6,144 points, such that 2,048 real data points of the matrix covered a spectral width of just 2,347.41 Hz (4.66 ppm) in the acquired amide proton dimension. The data points were shifted by 400 points to the left so that the amide region is situated at the middle of the spectrum.

Deviation from random coil chemical shift values

Deviation from random coil chemical shift value was calculated by subtracting random coil chemical shift values obtained for model peptides (Wishart et al., 1995) from the experimentally assigned chemical shift values. Additional corrections were made for residues that were followed by proline residues, as suggested (Wishart et al., 1995).

Results

Unfolding of the HPV-16 e2 DNA binding domain

Figure 1a shows the fluorescence spectra of the folded and various unfolded forms of HPV-16 E2 DNA binding domain (E2-C) at low pH, high temperature, and in urea. The red shifts in the spectra resulting from unfolding were analyzed by calculating the center of spectral mass. Results indicated that the maximum shift of the Trp probes is obtained by using urea, followed by low pH. Two of its three tryptophan residues lie at the center of the barrel, buried from the solvent (Liang et al., 1996). We therefore assume that a greater red shift represents a more solvent exposed chromophore, indicating a higher degree of unfolding. Only a small change was observed for the temperature denaturation. Previous studies suggested that the denatured state at high temperature has a tendency to aggregate, making the denaturation process irreversible (Mok et al., 1996a).

The far-ultraviolet (UV) CD spectrum of the protein at 3.0 M urea is characteristic of a highly disordered polypeptide. Further increase in urea concentration to 5.0 M did not produce a significant change in the far-UV CD spectrum (Fig. 1B), which implies that all native regular secondary structure is lost at 3.0 M urea. 1D ^1H-NMR spectra of E2-C were acquired at different concentrations of urea to confirm the degree of unfolding found in the spectroscopic studies described above. As shown in Figure 1C, the amide proton region of the native protein is substantially spread over a spectral width of 2.5 ppm from 7 to 9.5 ppm, as expected for a folded conformation. Addition of 3.0 M urea reduced the dispersion of the spectrum to 0.6 ppm (from 8 to 8.6 ppm), which is characteristic of a denatured polypeptide or loose conformations without persistent tertiary interactions. Further addition of urea to 5.0 M has no effect on the spectrum (Fig. 2A). The absence of highly shifted upfield methyl protons ($<$0.5 ppm) upon addition of 3.0 M of urea constitute further proof of loss of persistent tertiary structure of E2-C at this urea concentration. Similarly to the amide region, subsequent addition of urea to 5.0 M causes no further change on this region of the spectrum.

The denatured state, which can provide valuable information of folding events, is that devoided of all native secondary and tertiary structure, but in the mildest possible denaturation conditions to be able to observe residual structures. Our fluorescence, CD, and 1D-NMR data indicate that at 3.0 M of urea in sodium acetate buffer pH 5.6, this criterion was met. This denatured state would clearly differ from what can be obtained in extreme denaturing conditions, i.e., a fully unfolded polypeptide.

A two-state unfolding/dissociation of the HPV-16 E2 DNA binding domain was described previously from urea denaturation experiments (Mok et al., 1996a). At low protein concentrations, the urea midpoint for the denaturation transition ([U]50%) is dependent on protein concentration, as expected for a dissociation process. If this dependence was linear, the [U]50% at a protein concentration of 2 mM used in NMR experiments would exceed 50 M, in theory, for denaturation. Clearly, there must be a limit for the concentration dependence. We have analyzed the urea denaturation of E2-C at different protein concentrations under buffer conditions used for our NMR experiments, which

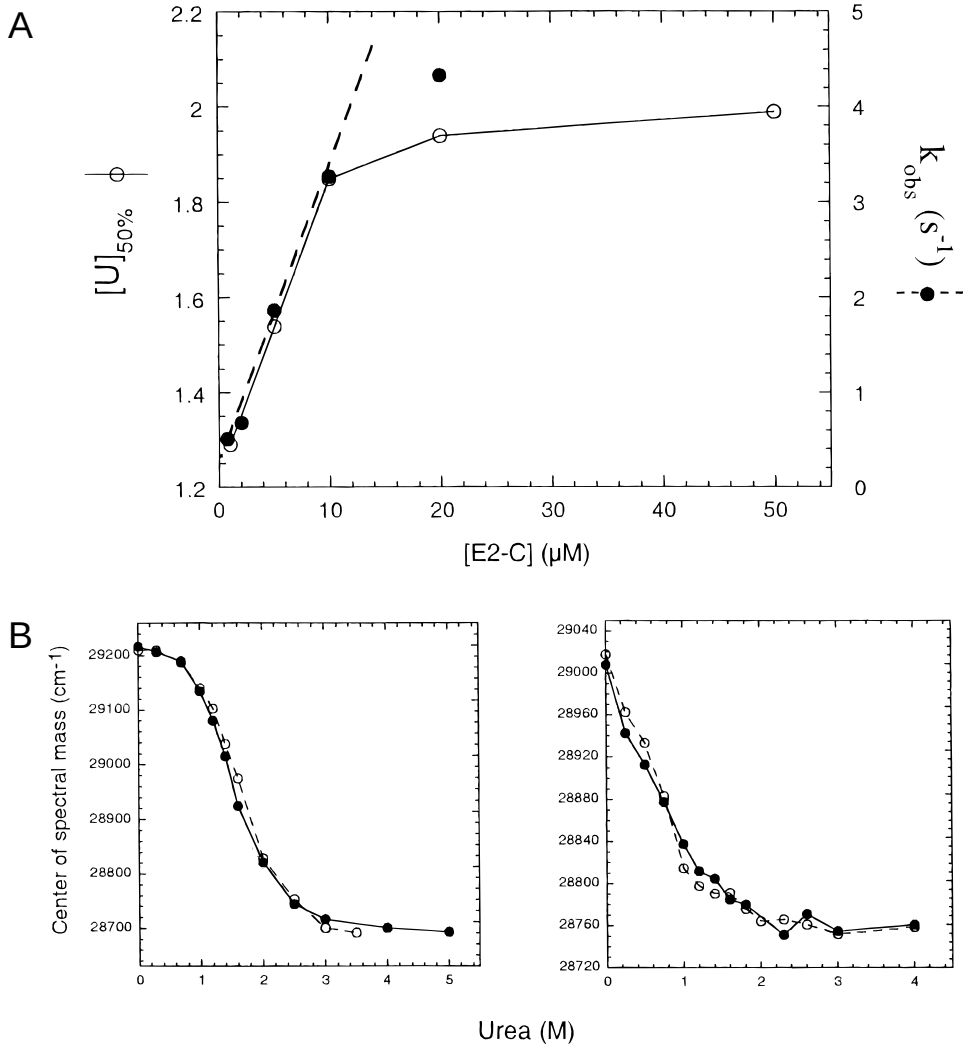

Fig. 2. A: Dependence of E2-C equilibrium urea denaturation midpoint on protein concentration. Each point represents the urea midpoint of an independent equilibrium denaturation experiment at the indicated protein concentration in 50 mM sodium acetate, pH 5.5 (○). Data for the protein concentration dependence of the observed rate constant (●) was taken from Mok et al. (1996b) and replotted here for direct comparison. **B:** Urea denaturation experiments at (○) 20 μM or (●) 50 μM E2-C concentration, monitored by changes in the center of fluorescence spectral mass. **C:** Urea denaturation experiments in the presence (●) or absence (○) of 10% D_2O. E2-C concentration was low (0.5 μM) to detect any marginal stabilization effect of the D_2O.

followed typical sigmoideal transitions. A concentration dependence was observed, but there was no change in [U]50% when the protein concentration was further increased from 10 to 50 μM. A plot of the [U]50% against concentration of E2-C shows this brake in the concentration dependence (Fig. 2A). Therefore, at high protein concentrations, the unfolding at equilibrium appears not to be driven by the dissociation of the dimer. The deviation in the protein concentration dependence parallels that of the rate constant for the dimerization/folding process previously described, and those data are plotted here to show the parallelism (Mok et al., 1996b; Fig. 2A).

The lack of protein concentration dependence can be clearly visualized from the superposition of the urea unfolding transitions at 20 and 50 μM, shown in Figure 2B. Finally, because NMR experiments will be conducted in the presence of D_2O, we wanted to test whether the solvent conditions could have a stabilizing effect on the protein. Figure 2C shows that there is no difference in the presence or absence of 10% D_2O.

Weak cross-peak intensity of HSQC indicative of restricted backbone movement

An HSQC experiment was performed as a first step toward the sequential assignment and detailed characterization of the urea denatured state of E2-C. Figure 3 shows the 2D ^{15}N–^{1}H HSQC spectrum for HPV-16 E2 DNA binding domain unfolded in 3.0 M urea at pH 5.6. Sequence-specific assignments were obtained for all 78 backbone amide protons, except for the two proline residues, which do not bear an amide proton. All the backbone amide cross peaks appear to be splitted. This is due to the high digital resolution in the amide proton dimension and the narrow line width of the unfolded protein, and to the $^{3}J_{HNH\alpha}$ coupling. The resulting

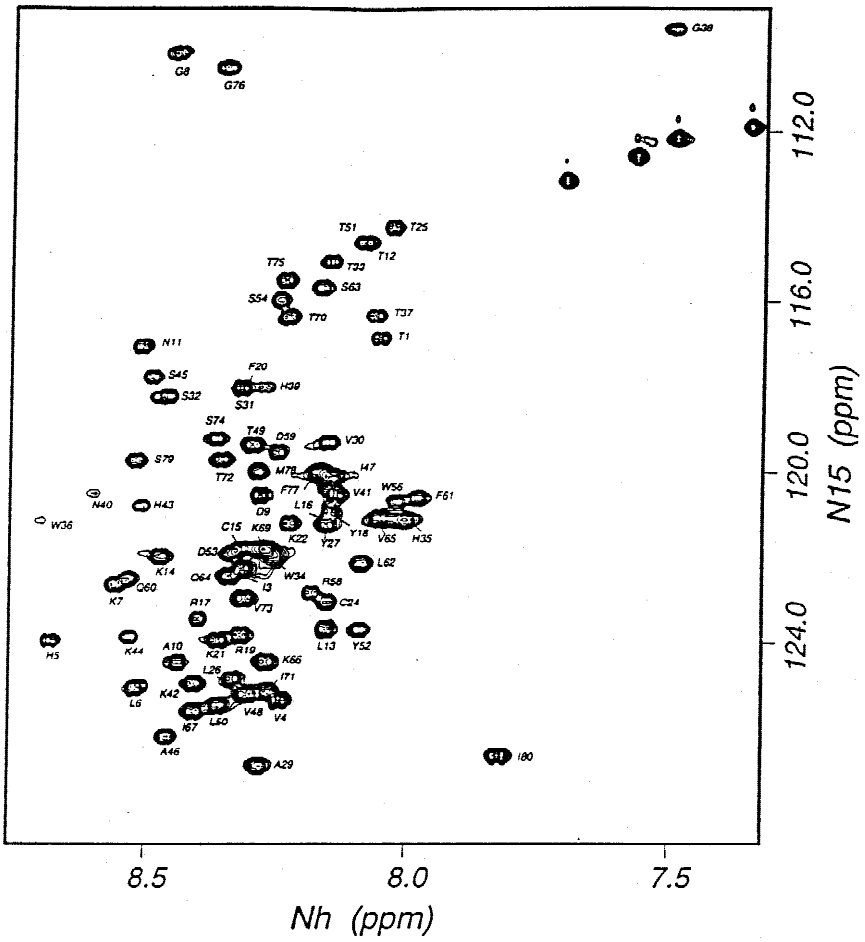

Fig. 3. The 2D ^{15}N–^{1}H HSQC spectrum of uniformly ^{15}N-labeled HPV-16 E2 DNA binding domain unfolded in 3.0 M urea. A protein sample of 2 mM in 50 mM sodium acetate (pH 5.6) was used in the NMR experiment at 25 °C. Cross peaks with particularly low intensity like W36, H39, and N40 are probably involved in nonrandom coil conformations.

split in the cross peaks can be used for $^{3}J_{HNH\alpha}$ coupling constant measurements. As observed in 1D experiments, there is very little dispersion in the amide proton dimension (8.0–8.6 ppm). Considerable dispersion in the ^{15}N dimension, however, is retained even in the denatured state.

When we analyzed the intensities of HSQC cross peaks, significant variations were observed (Fig. 3). Residues W36, H39, and N40 have much weaker intensities when compared with the rest of HSQC backbone amide cross peaks. Among the three Gly residues, the amide proton resonance of G38 is shifted upfield significantly to 7.5 ppm. According to the secondary structure assignment of the homologous structure of HPV-31 E2 DNA binding domain (Liang et al., 1996), residues W36, G38, H39, and N40 are situated at the end of β-strand 2 (β2), involved in the dimeric interface of the native conformation.

NMR sequential assignment of the urea denatured state of HPV-16 E2-C

The ^{1}H–^{15}N TOCSY was used to assign the side-chain proton resonances to each particular amide proton. Amino acid-type information can be obtained for some residues having a unique pattern, for example, Ala, Gly, Thr, Ser, and Val, by comparison with random coil chemical shift values (Wüthrich, 1989). A much longer TOCSY mixing time of 73 ms could be used due to the slower relaxation property of the unfolded protein, which allowed efficient TOCSY transfer up to the γ proton in the side chain before the signal decayed.

The identified spin systems were sequentially linked together by means of two pairs of complementary triple resonance experiments. The CBCA(CO)NH experiment (Fig. 4A) defined the amino acid type of the preceding residue based on comparison with $^{13}C\alpha$ and $^{13}C\beta$ random coil chemical shifts (Grzesiek & Bax, 1993). When combined with the HNCA experiment (Fig. 4B), the spin systems could be linked together through the $^{13}C\alpha$ chemical shifts. The HNCA spectrum provides increased resolution in the carbon dimension, which is essential in distinguishing the highly degenerate $^{13}C\alpha$ chemical shifts. When ambiguity arose in the CBCA(CO)NH and HNCA experiments, either due to the presence of a redundant amino acid pair or the overlapping $^{13}C\alpha$ chemical shifts, sequential information was obtained from the HNCO and HN(CA)CO experiments (Fig. 5). The ^{13}C carbonyl

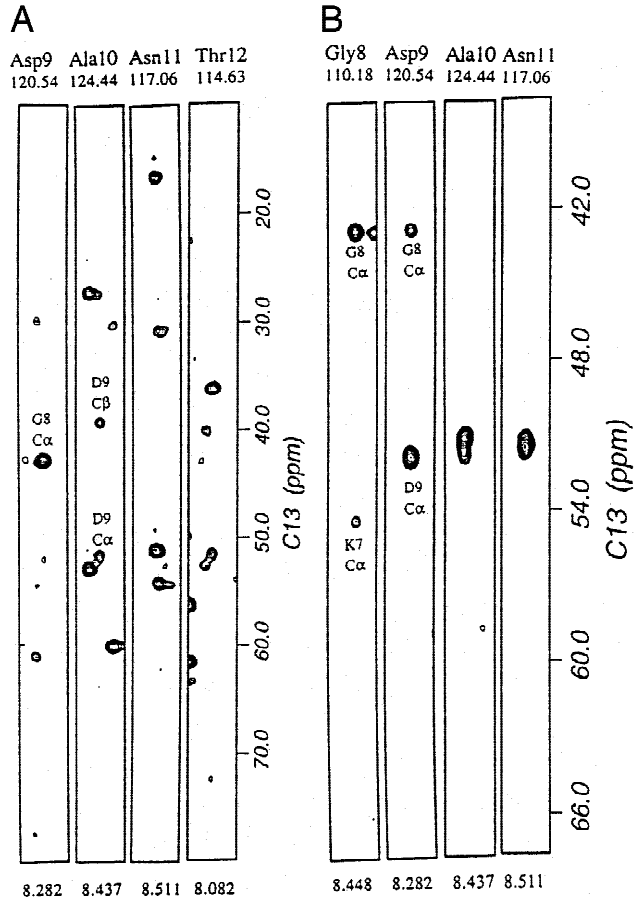

Fig. 4. The CBCA(CO)NH and HNCA spectra of unfolded HPV-16 E2 DNA binding domain in 3.0 M urea. **A:** Series of amide strips taken out from the CBCA(CO)NH spectrum. Comparison of $^{13}C\alpha$ and $^{13}C\beta$ chemical shifts with that of random coil values defined residue type information of the preceding residue. **B:** Series of amide strips taken out from the HNCA spectrum. Besides giving the $^{13}C\alpha$ chemical shift of its own spin system, the $^{13}C\alpha$ chemical shift of the preceding residue can also be observed.

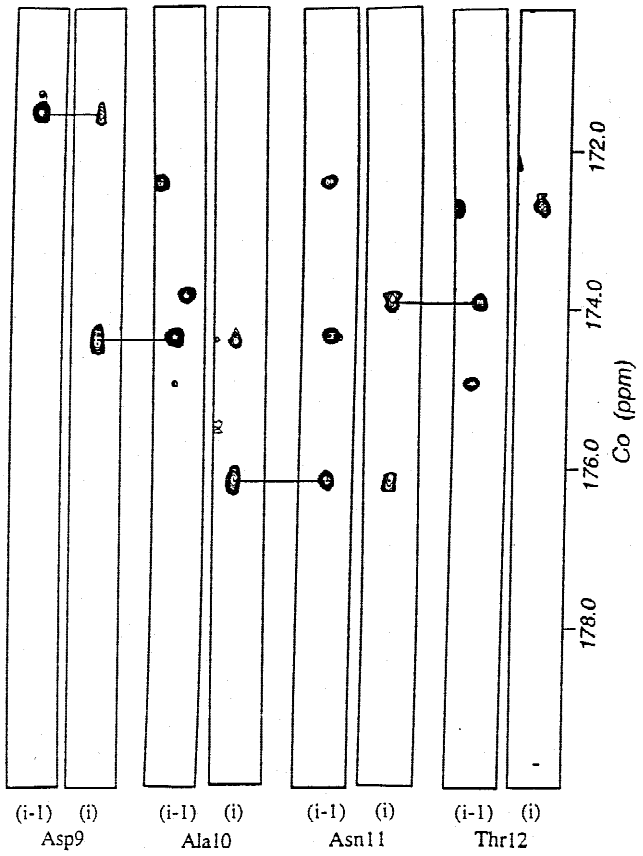

Fig. 5. The HNCO and HN(CA)CO spectra of the unfolded HPV-16 E2 DNA binding domain in 3.0 M urea. Regions of the 3D HNCO and HN(CA)CO spectra of HPV-16 E2 DNA binding domain in 3.0 M urea showing sequential connection of ^{13}C carbonyl chemical shifts were selected and drawn in pairs. The left strip of each pair is taken from the HNCO spectrum, which correlates the amide ^{1}H and ^{15}N resonances of residue i with the ^{13}C carbonyl resonance of residue $i-1$. The right strip of each pair is taken from the HN(CA)CO spectrum, which correlates the amide ^{1}H and ^{15}N resonances of residue i with the ^{13}C carbonyl resonances of both residue $i-1$ and i. In HN(CA)CO, the ^{13}C carbonyl resonance of residue $i-1$ usually has a weaker intensity. ^{13}C carbonyl resonances for residue $i-1$ are connected between spectra with horizontal lines. Sequential residue assignments are given below each paired strips.

dimension is usually less degenerate and has a higher digital resolution than the $^{13}C\alpha$ dimension, thus providing an additional link for sequential assignment of the urea denatured E2-C.

Deviation of chemical shifts from random coil values

Deviation of chemical shift values from random coil values indicates the presence of residual structure in denatured states of proteins. By subtracting the random coil values (Wishart et al., 1995, Tables 1, 2) from the ^{1}HN, ^{1}Ha, $^{13}C\alpha$, and $^{13}C\beta$ chemical shift values, their deviation from random coil values, or secondary shifts, can be obtained. Figures 6A and 6B show deviations from random coil values for ^{1}HN and $^{1}H\alpha$, respectively. Figures 6C and 6D show deviations from random coil values for $^{13}C\alpha$ and $^{13}C\beta$, respectively. The deviations from $^{13}C\alpha$ and $^{13}C\beta$ showed a systematic referencing error of about 2 ppm, which make the observation of deviation from normal values more difficult. It has been shown that the difference between both the $^{13}C\alpha$ and $^{13}C\beta$ chemical shifts of a given residue from their random coil values can indicate the type of secondary structure in which the residue is involved in the folded protein (Spera & Bax, 1991). Subtracting random coil Cα shift from the Cα shifts of a protein tends to give a positive secondary shift for Cαs in helical backbone conformation and a negative secondary shift for those in an extended strand conformation. This trend is opposite in sign for Cβ secondary shifts. By subtracting the Cβ secondary shift of a residue from its Cα secondary shift, this correlation is enhanced, as shown in Figure 6E for the unfolded state of E2. Further, by subtracting Cβ secondary shift from Cα secondary shift, any internal systematic referencing error is canceled out. Only those residues that have both Cα and Cβ chemical shifts assigned are shown in Figure 6E.

Figure 6 shows that the majority of ^{1}H and ^{13}C resonances are in close agreement with random coil values, indicating that the protein is largely unfolded. Those resonances that deviate to a significant extent ($\delta > 0.3$ ppm for labile amide proton, $\delta >$ 0.1 ppm for nonlabile Hα proton, and $\delta > 1$ ppm for 13Cα–

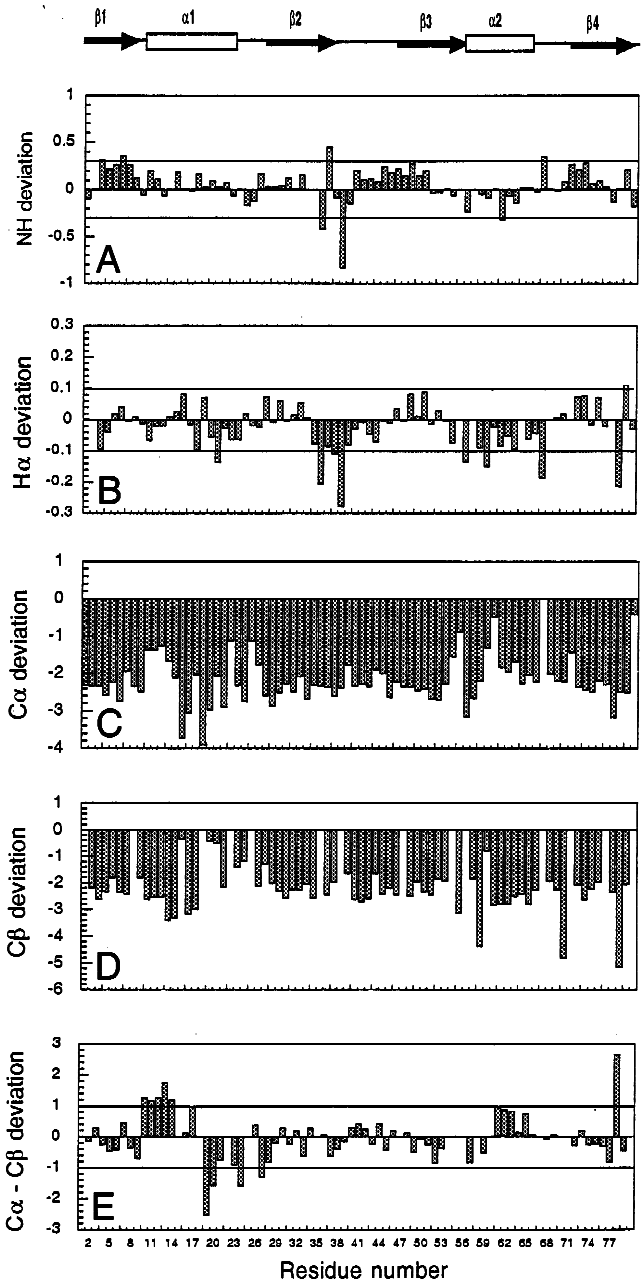

Fig. 6. Deviation from random coil chemical shifts values for urea denatured HPV-16 E2 DNA binding domain. (**A**) Deviation for amide proton, (**B**) deviation for Ha proton, (**C**) deviation for $^{13}C\alpha$, (**D**) deviation for $^{13}C\beta$, and (**E**) deviation for $^{13}C\alpha$ minus deviation for 13 Cβ. The horizontal lines show the accepted range outside which deviations are considered to be significant. The Hα proton deviation were centered about zero by the addition of 0.06 ppm to normalize the referencing difference. The secondary structure alignment was based on the the NMR structure of HPV-31 E2 DNA binding domain (Liang et al., 1996).

13Cβ) and cluster together were found in three distinct regions: (1) Ala10–Lys14:13Cα–13Cβ chemical shifts of A10, N11, T12, L13, K14; (2) Arg19–Phe20:13Cα–13Cβ chemical shifts of R19 and F20 and Hα of F20; (3) His35–Gly38:H35:NH, Hα; W36:NH; T37:Hα; G38:NH, Hα. According to the secondary structure assignment of the homologous structure of HPV-31 E2 DNA binding domain (Liang et al., 1996), residues Ala10–Lys14 correspond to the first helical turn (C-cap) of the DNA binding helix (α1), Arg19 and Phe20 are located at the end of α1, while His35–Gly38 spanned the end of β2, which is involved in the dimeric interface. All structures mentioned refer to the folded conformation.

Deviation from "random coil" $^3J_{HNH\alpha}$ coupling constants

In denatured states of proteins, $^3J_{HNH\alpha}$ coupling constants for residues of the same type are closely coincident and display "random coil" $^3J_{HNH\alpha}$ coupling constant values (Arcus et al., 1995), depending on the backbone dihedral angle preference of each type of amino acid residues. For each residue type in which more than two $^3J_{HNH\alpha}$ coupling constants are available, then an internal "random coil" value can be obtained. Figure 7 shows the backbone $^3J_{HNH\alpha}$ coupling constants plotted according to residue type. For instance, threonine appears to have an internal "random coil" coupling constant of 7.71 Hz and tryptophan has an internal "random coil" value of 6.31 Hz. Besides internal "random coil" values, a set of external "random coil" values of $^3J_{HNH\alpha}$ is also available based on distribution of main-chain torsion angles in a database of 85 protein crystal structures (Smith et al., 1996). Deviation from these "random coil" values would indicate nonrandom backbone conformation probably due to the presence of residual structures. As the result of spectral overlap in certain regions of the HSQC spectrum, $^3J_{HNH\alpha}$ coupling constants could be measured for only 65 of the 80 residues in the urea denatured HPV-16 E2 DNA binding domain (see Supplementary material in the Electronic Appendix).

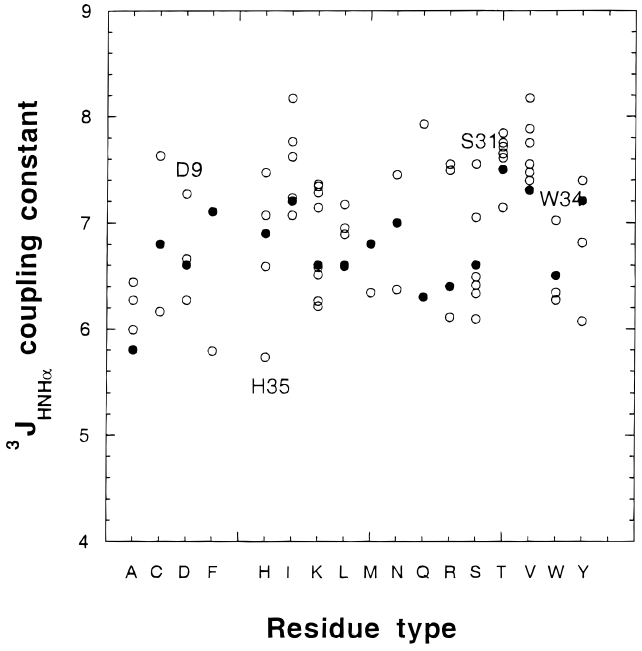

Fig. 7. $^3J_{HNH\alpha}$ coupling constants (open circles) for urea-denatured HPV-16 E2 DNA binding domain plotted according to residue type. Coupling constants fall within a narrow range of internal "random coil" value for residues of the same type. External "random coil" values taken from Lorna Smith et al. (1996). "ALL" dataset were also indicated (closed circles). Residues that have their coupling constant values deviated significantly from both the external "random coil" value and the rest of their own amino acid type are labeled.

Only residues having their coupling constants significantly deviated (>0.5 Hz) from both their external and internal random coil value are indicated in the figure. Residues D9 and R17 are both located in the region corresponding to the DNA binding helix $\alpha 1$ in the native structure. Residues S31, W34, and H35 are clustered on a region at the end of $\beta 2$, involved in the dimeric interface of the folded dimer.

NOESY data

The ^{1}H–^{15}N NOESY spectrum of urea-denatured HPV-16 E2 DNA binding domain is dominated by intraresidue, $\alpha N(i, i + 1)$ and $\beta N(i, i + 1)$ NOE contacts typical of a random coil conformation as shown in Figure 8. Nonrandom coil and nonnative NOE contacts, however, have also been observed, and their positions on the protein sequence are shown in Figure 9. $\alpha N(i, i + 1)$, $\beta N(i, i + 1)$, and $\gamma N(i + 1, i)$ were observed at the start of the major helix, $\alpha 1$, between residues A10 and D9, at the end of $\alpha 1$ between residues F20 and K21, and at the start of β-strand 2 between residues V30 and A29. This kind of nonrandom coil and nonnative NOE can only be observed in loop-like structures, but not in standard secondary structures. Longer range NOE contacts, for example, $\delta N(i, i + 3)$ and $NN(i, i + 3)$ were also observed at the start of $\alpha 1$ and $\beta 2$.

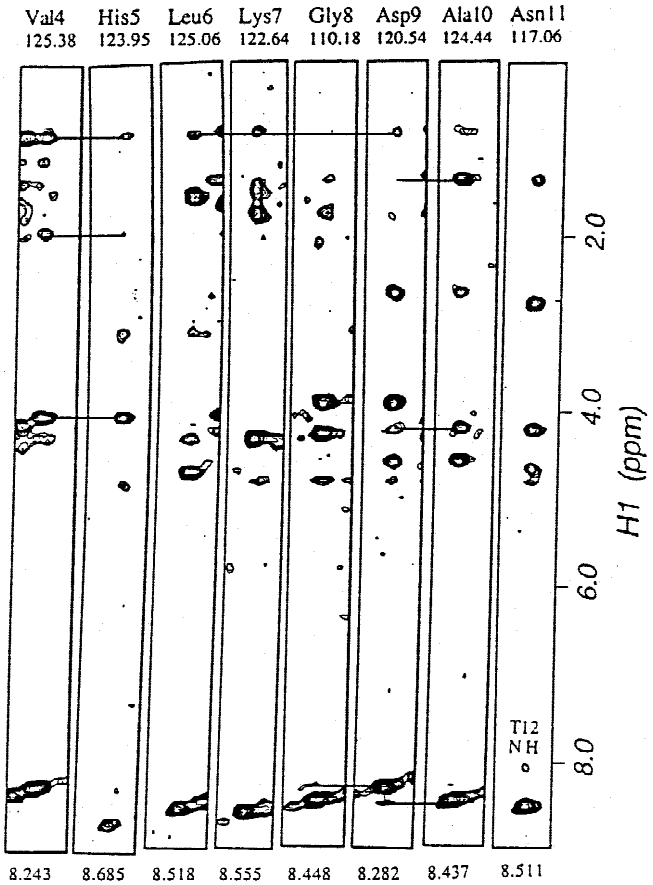

Fig. 8. The 3D NOESY-HMQC spectrum of HPV-16 E2 DNA binding domain in 3.0 M urea. The ^{1}H–^{15}N NOESY spectrum of the unfolded protein is dominated by intraresidue, $d_{\alpha N}(i, i + 1)$, $d_{\beta N}(i, i + 1)$, and even $d_{\gamma N}(i, i + 1)$ NOE such as those observed between Val4 and His5, which is typical of random coil conformation. $d_{NN}(i, i + 1)$ NOE is observed in some but not all of the residues, an example is the $d_{NN}(i, i + 1)$ NOE between Gly8, Asp9, and Ala10. Nonrandom coil and nonnative NOEs are also observed, including $d_{\alpha N}(i, i + 1)$ and $d_{\beta N}(i, i + 1)$ NOE from Ala10 to Asp9, and some longer range $d_{\delta N}(i, i + 3)$ NOE from Leu6 to Asp9.

Discussion

The unfolded state of E2-C: Folding mechanism and biological function

The dimeric β-barrel is a newly described fold found in the DNA binding domain of the E2 transcriptional regulatory proteins of papillomaviruses (Hegde et al., 1992). This fold is so far shared only by the EBNA1 nuclear antigen from the Epstein–Barr virus (Bochkarev et al., 1996). EBNA1 shows no sequence homology with the E2-C domains from HPVs, yet their folding topologies are strikingly similar. Both viral proteins bind DNA and are grouped within the origin binding proteins (OBPs) (Edwards et al., 1998). However, the mechanism of binding of DNA appears to be different (Bochkarev et al., 1996). Thus, the uniqueness of the dimeric β-barrel topology together with different DNA binding mechanisms or base sequence recognition strongly suggest possible role of the folding/dimerization mechanism in the biological function. This system not only serves as a model for studying the folding mechanism of dimeric proteins, but investigating residual structures in the denatured state of E2-C might result of relevance for understanding the molecular basis for the action of OBPs such as E2.

Sequential assignment process

A complete set of ^{1}H, ^{15}N, and ^{13}C backbone assignments was obtained using various heteronuclear triple resonance NMR experiments. From the analysis of different NMR parameters such as deviation from random coil chemical shifts, $^{3}J_{HNH\alpha}$ coupling constants, and NOE, we have characterized the residual structures in the denatured state of HPV-16 E2 DNA binding domain in 3.0 M of urea. Evidence of residual structure in the region that corresponds to the DNA binding helix ($\alpha 1$, A10–K22) in the native structure was obtained mainly from the deviation of random coil $^{13}C\alpha$ and $^{13}C\beta$ chemical shifts. The observed chemical shift in denatured states of proteins is a population weighted average of the shifts in the individual conformations, interconverting rapidly in the NMR timescale (Wüthrich, 1994). Only clustered regions with significant deviation from random coil chemical shift values are considered to have residual structure involved in "nonrandom coil" preferential conformations. $^{13}C\alpha$ and $^{13}C\beta$ chemical shifts are less susceptible to solution conditions if compared to amide proton chemical shifts. They are also less prone to sequence effects as in the case of ^{15}N chemical shifts, although these have a much greater dispersion (Yao et al., 1997). The direction in which ^{13}C chemical shifts are deviated can be indicative of the type of secondary structure (Wishart & Sykes, 1994). For $^{13}C\alpha$ chemical shifts, a downfield shift would indicate α-helical structure, whereas an upfield shift is indicative of β-strand conformations. $^{13}C\beta$ chemical shifts display similar behavior to Hα. The combined deviation of both $^{13}C\alpha$ and $^{13}C\beta$ shifts provides stronger support to the type of secondary structure found in a particular region. In the case of

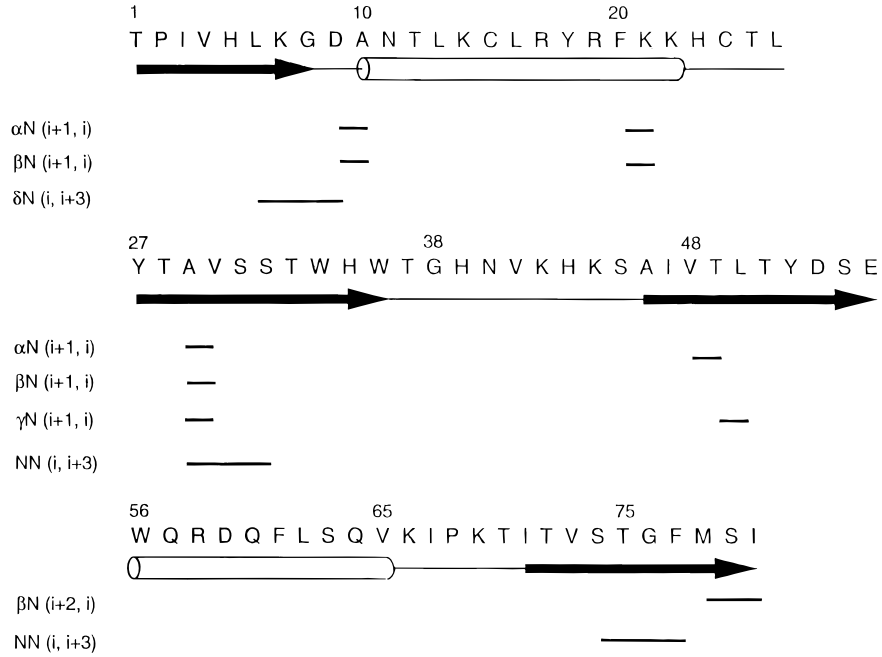

Fig. 9. Nonrandom coil NOEs observed in urea denatured HPV-16 E2 DNA binding domain. The sequence of HPV-16 E2 DNA binding domain together with a summary of nonrandom coil NOEs observed in the unfolded protein. Clusters of nonnative NOEs like $d_{\alpha N}(i, i + 1)$, $d_{\beta N}(i, i + 1)$, and $d_{\gamma N}(i, i + 1)$ were observed at the start of the DNA binding helix (D9–A10) and at the start of the β-strand 2 (A29–V30). The secondary structures assignment is based on the solution structure of the HPV-31 E2 DNA binding domain (Liang et al., 1996).

HPV-16 E2 DNA binding domain, an upfield shift is observed from residue A10–K14, while a downfield shift is observed for residues R19 and F20 of the DNA binding helix. This would lead us to hypothesize that both α-helical and nonnative β-strand residual secondary structures are observed in the DNA binding helix α1. However, such interpretations are based on secondary structure found in fully folded conformations of proteins. Because structures found here appear to be mainly of local nature, we cannot rule out that the upfield or downfield deviations correspond to nonclassical or nonnative secondary structure. We believe this consideration must be taken into account in the analysis of structure in denatured states or disordered peptides in general.

Evidence of hydrophobic clustering in β-strand 2

There is a group of hydrophobic residues, namely, W34, H35, and W36 situated on β-strand 2, the region corresponding to the dimer interface at the center of the β-barrel in the fully folded native conformation. The two Trp residues are highly conserved among HPV E2-C domains (Hegde et al., 1992). We show here that these residues are involved in hydrophobic clustering in the denatured state of the protein. The first evidence of residual structure in the region from H35 to G38 came from the deviation of random coil NH and ^{1}Hα chemical shifts. A closer inspection revealed that the highly shifted HN of G38 could be a result of ring current effect due to its proximity to the aromatic ring of W36. A similar ring current effect has also been observed by Merutka and coworkers in random coil chemical shifts from reference peptides (Merutka et al., 1995). Therefore, the chemical shift value alone may not be enough to prove residual structure in residue G38.

Further evidence of hydrophobic clustering in this region came from the fact that the HSQC peaks of W36, H39, and N40 are extremely broadened and weak when compared with HSQC peaks of other residues. HSQC peaks of narrow linewidth and high intensity are expected for a denatured protein, which adopt a highly flexible conformation and interconvert between different conformations at a very fast rate. Broadened peaks strongly suggest that the protein backbone in this region is less flexible and undergoing certain types of intermediate conformational exchange (10^3 s^{-1}), which are indicative of "nonrandom" coil behaviors. Restriction in protein backbone movement is likely to be the result of local preferences in backbone dihedral angles.

The last piece of evidence of residual structure in this region came from the $^3J_{HNH\alpha}$ coupling constant measurements of residues S31, W34, and H35. $^3J_{HNH\alpha}$ coupling constant is related to the backbone dihedral ϕ angle by the Karplus equation (Karplus, 1959). The $^3J_{HNH\alpha}$ for native α-helix and β-strand are 4.8 and 8.5 Hz, respectively. Using a database of 85 X-ray structures, the $^3J_{HNH\alpha}$ coupling constants for different amino acid types in peptide and denatured proteins has been found to range from 5.9 and 6.1 Hz for Gly and Ala to 7.7 Hz for Val (Smith et al., 1996). Intrinsic ϕ,ψ propensities of different residue types are driven mainly by electrostatic and, to a lesser extent, steric interactions between side-chain and local peptide units. This type of structural preference helps to reduce the conformational space explored by a polypeptide chain and presumably facilitates the folding process (Swindells et al., 1995). Significant deviations of the $^3J_{HNH\alpha}$ coupling constants from the "random coil" values of other residues of the same type is indicative of persistent residual structure. The $^3J_{HNH\alpha}$ coupling constants of S31 and W34 were found to be shifted

toward a higher value of 7.55 and 7.02 Hz, respectively. This is in agreement with a β-conformation of the hydrophobic cluster found on β-strand 2. The $^3J_{HNH\alpha}$ of H35, however, is shifted toward a lower value of 5.73 Hz, indicative of a more α-helical conformation. This strongly suggests that the β-conformation found near this hydrophobic cluster is not likely to be a typical β-strand but rather some type of nonnative conformation.

Nonnative turn-like structures

For a denatured protein in which multiple conformations are interconverting rapidly, the observed NOE intensities are a complex average over nonvanishing contributions from all conformations (Wüthrich, 1994). NOEs, which can be observed in all conformations, are therefore classified as random-coil NOEs. The distance $d\alpha N(i, i + 1)$ varies between 2.2 and 3.6 Å for all the possible conformations (Wüthrich, 1994) and should be observed throughout the entire polypeptide sequence. Besides $\alpha N(i, i + 1)$ NOEs, $\beta N(i, i + 1)$, and $\gamma N(i, i + 1)$, NOEs are also commonly observed because of their proximity to the amide proton. Finally, $NN(i, i + 1)$ are random-coil NOEs that are predicted for all the residues in the unfolded protein by a model based on main-chain dihedral angles (Fiebig et al., 1996). Not all the $NN(i, i + 1)$ NOEs, however, are observed in the denatured state of HPV-16 E2-C.

The presence of a nonrandom coil NOE cross peak between two distinct protons provides evidence for preferential occurrence of a particular conformation in denatured states of proteins. Based on our NOESY experiment, nonrandom coil and nonnative NOEs such as $\alpha N(i, i + 1)$, $\beta N(i, i + 1)$, and $\gamma N(i, i + 1)$ NOEs, together with longer range $\delta N(i, i + 3)$ and $NN(i, i + 3)$ NOEs, were found at the turns between $\beta 1$ and $\alpha 1$ as well as 1 and $\beta 2$. Residues involved in these turn-like structures included D9 and A10 at the start of $\alpha 1$, F20 and K21 at the end of $\alpha 1$, and A29 and V30 at the start of $\beta 2$. The $^3J_{HNH\alpha}$ coupling constant of D9 is also shifted to a higher value of 7.27 Hz, indicating restricted backbone flexibility at the start of $\alpha 1$. This type of nonnative NOE, $(i, i + 1)$, is not likely to be found in native folded proteins and can be explained by nonnative loop or turn-like conformations in the denatured state.

Implications for the folding pathway

Residual structures were mainly found in the region corresponding to the DNA binding helix from D9 to K14 and from R19 to K21, and in the second β-strand at the dimeric interface from A29 to V30 and W34 to N40. The presence of persistent residual structures indicate the occurrence of conformational preferences, which significantly reduce the conformational space to be searched by the polypeptide chain. Residual structures present in urea-denatured states are indicative of structure nucleation and highly relevant for folding mechanisms.

A proposed scheme to explain our results in terms of the folding mechanism of this dimer is shown in Figure 10. The unfolded monomer is largely disordered; however, significant amount of residual structure is found in the regions corresponding to the DNA binding helix $\alpha 1$ and the second β-strand, most likely in the form of hydrophobic clustering. From previous kinetic (Mok et al., 1996b) and equilibrium denaturation experiments in this work, a coincident deviation from linearity in the protein concentration dependence of the refolding rate and $[U]_{50\%}$ was observed above 10 μM. This strongly suggests that at low protein concentration an early

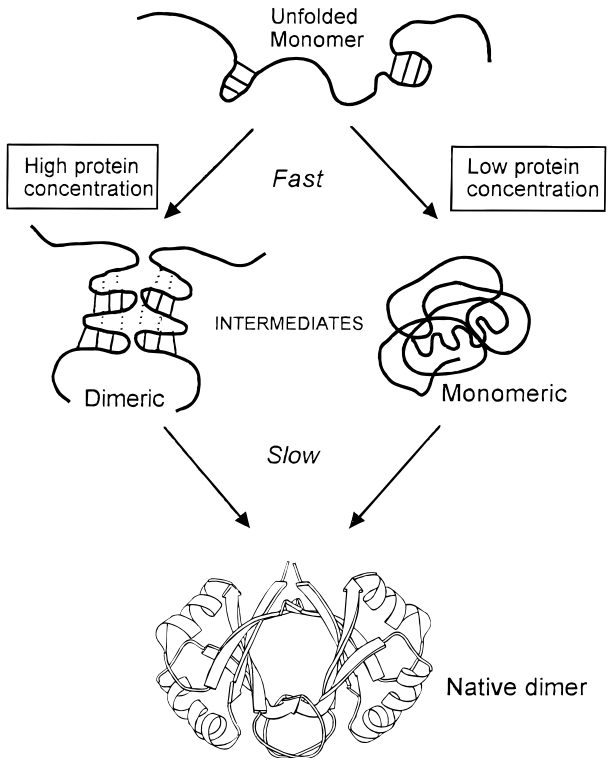

Fig. 10. Proposed model of alternative folding pathways at high and low protein concentration.

monomeric intermediate is formed, but the rate-limiting step is the subsequent dimerization step that includes substantial folding, as expected for the highly intertwined structure of the domain (Mok et al., 1996b). As previously suggested (Mok et al., 1996b) in this alternative pathway, the early intermediate would dimerize at high protein concentration, and the rate-limiting step becomes a first-order folding event. Denatured E2-C or early monomeric intermediate molecules could collide faster and recognize each other as the protein concentration is increased, with a higher dissociation constant than that for the folded dimer. Upon collision/refolding, classic secondary structures build up from these sites in the dimeric intermediate, and native tertiary interactions start to appear from this stage onward. In any event, a more detailed mechanistic characterization for the pathway at high concentration will be required.

Figure 11 summarizes the evidence for residual structure in the urea denatured state of E2-C. Most of the structures that can form, according to the NMR data, are likely to be of nonnative nature or, at least, cannot be fully native. This is in agreement with the previously described monomeric kinetic intermediate (Mok et al., 1996b). However, this intermediate may or may not be structurally and/or thermodynamically equivalent to the denatured state. Structures found in intermediates in general should not be thought exclusively in terms of native or nonnative structures; it should be considered whether they are productive for folding. The final native structure is the outcome of the balance between thermodynamics and kinetics, but at early stages when large unstructured regions of the polypeptide chain are exposed to the solvent, native-like structures may not be the only route down the folding funnel.

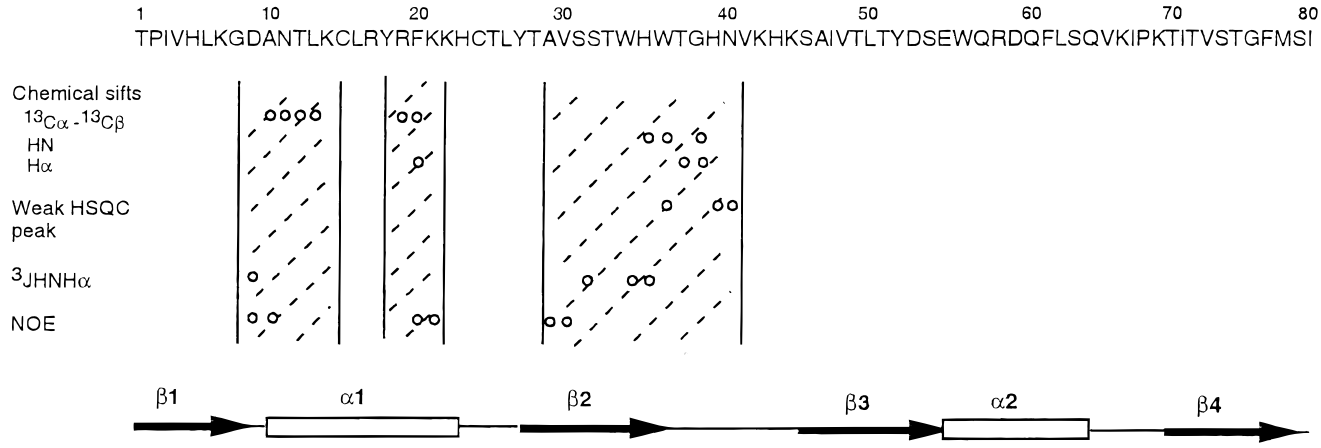

Fig. 11. NMR evidence for residual structure in the urea denatured state of HPV-16 E2-C.

Supplementary material in the Electronic Appendix

Supplementary material including the complete chemical shift assignments is available.

Acknowledgments

G.P.G. wishes to acknowledge the support of the Wellcome Trust (Grant OIA U41 RG27994), Fundación Bunge y Born and Fundación Antorchas. We thank Prof. Alan Fersht for his support, and Dr. Jose Luis Neira for helpful comments and criticisms of the manuscript. G.PG. is a John Simon Guggenheim Fellow.

References

Arcus VL, Vuilleumier S, Freund SMV, Bycroft M, Fersht AR. 1994. Toward solving the folding pathway of barnase: The complete backbone ^{13}C, ^{15}N and ^{1}H NMR assignments of its pH-denatured state. *Proc Natl Acad Sci USA 91*:9412–9416.

Arcus VL, Vuilleumier S, Freund SMV, Bycroft M, Fersht AR. 1995. A comparison of the pH, urea, and temperature-denatured states of barnase by heteronuclear NMR: Implications for the initiation of protein folding. *J Mol Biol 254*:305–321.

Bax A, Ikura M, Kay LE, Torchia DA, Tschudin R. 1990. Comparison of different modes of two-dimensional reverse-correlation NMR for the study of proteins. *J Magn Reson 86*:304–318.

Bax A, Subramanian S. 1986. Sensitivity-enhanced two-dimensional heteronuclear shift correlation NMR spectroscopy. *J Magn Reson 67*:565–569.

Bochkarev A, Barwell JA, Pfuetzner RA, Bochkareva E, Frappier L, Edwards AM. 1996. Crystal structure of the DNA-binding domain of the Epstein-Barr virus origin-binding protein, EBNA1, bound to DNA. *Cell 84*:791–800.

Bodenhausen G, Ruben DJ. 1980. Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chem Phys Lett 69*:185–189.

Buck M, Schwalbe H, Dobson CM. 1995. Characterization of conformational preferences in a partly folded protein by heteronuclear NMR spectroscopy: Assignment and secondary structure analysis of hen egg-white lysozyme in trifluoroethanol. *Biochemistry 34*:13219–13232.

Clubb RT, Thanabal V, Wagner G. 1992. A constant-time three-dimensional triple-resonance pulse scheme to correlate intraresidue $^{1}H^{N}$, ^{15}N, and $^{13}C'$ chemical shifts in ^{15}N-^{13}C-labeled proteins. *J Magn Reson 97*:213–217.

Edwards AM, Bochkarev A, Frappier L. 1998. Origin DNA-binding proteins. *Curr Opin Struct Biol 8*:49–53.

Evans PA, Topping KD, Woolfson DN, Dobson CM. 1991. Hydrophobic clustering in nonnative states of a protein: Interpretation of chemical shifts in NMR spectra of denatured states of lysozyme. *Proteins 9*:248–266.

Fiebig KM, Schwalbe H, Buck M, Smith LJ, Dobson CM. 1996. Toward a description of the conformations of denatured states of protein. Comparison of a random coil model with NMR measurements. *J Phys Chem 100*:2661–2666.

Foguel D, Silva JL, Prat Gay G de. 1998. Characterization of a partially folded monomer of the DNA-binding domain of human papillomavirus E2 protein obtained at high pressure. *J Biol Chem 273*:9050–9057.

Grzesiek S, Bax A. 1992. Improved 3D triple-resonance NMR techniques applied to a 31 kDa protein. *J Magn Reson 96*:432–440.

Grzesiek S, Bax A. 1993. Amino acid type determination in the sequential assignment procedure of uniformly $^{13}C/^{15}N$-enriched protein. *J Biomol NMR 3*:185–204.

Hawley-Nelson P, Androphy EJ, Lowy D, Schiller J. 1988. The specific DNA recognition sequence of bovine papillomavirus E2 protein is an E2-dependent enhancer. *EMBO J 7*:525–531.

Hegde RS, Androphy EJ. 1998. Crystal structure of the E2 DNA-binding domain from human papillomavirus type 16: Implications for its DNA binding-site selection mechanism. *J Mol Biol 284*:1479–1489.

Hegde RS, Grossman SR, Laimins LA, Sigler PB. 1992. Crystal structure at 1.7 A of the bovine papillomavirus-1 E2 DNA-binding domain bound to its DNA target. *Nature 359*:505–512.

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

Karplus M. 1959. Contact electron-spin coupling of nuclear magnetic moments. *J Chem Phys 30*:11–15.

Kay LE, Ikura M, Tschudin R, Bax A. 1990. Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. *J Magn Reson 89*:496–514.

Kay LE, Keifer P, Saarinen T. 1992. Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *J Am Chem Soc 114*:10663–10665.

Liang H, Petros AM, Meadows RP, Yoon HS, Egan DA, Walter K, Holzman TF, Robins T, Fesik SW. 1996. Solution structure of the DNA-binding domain of a human papillomavirus E2 protein: Evidence for flexible DNA-binding regions. *Biochemistry 35*:2095–2103.

Lima LMTR, Prat Gay G de. 1997. Conformational changes and stabilization by ligand binding in the C-terminal DNA-binding domain of the E2 protein from human papillomavirus strain-16. *J Biol Chem 272*:19295–19303.

Logan TM, Theriault Y, Fesik SW. 1994. Structural characterization of the FK506 binding protein unfolded in urea and guanidine hydrochloride. *J Mol Biol 236*:637–648.

Marion D, Driscoll PC, Kay LE, Wingfield PT, Bax A, Gronenborn AM, Clore GM. 1989a. Overcoming the overlap problem in the assignment of 1H NMR spectra of larger proteins by use of three-dimensional heteronuclear 1H-15N Hartmann-Hahn-multiple quantum coherence spectroscopy: Application to interleukin 1β. *Biochemistry 28*:6150–6156.

Marion D, Ikura M, Bax A. 1989b. Improved solvent suppression in one and two-dimensional NMR spectra by convolution of time domain data. *J Magn Reson 84*:425–430.

Marion D, Ikura M, Tschudin R, Bax A. 1989c. Rapid recording of 2D NMR spectra without phase cycling. Application to the study of hydrogen exchange in proteins. *J Magn Reson 85*:393–399.

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

McBride AA, Byrne JC, Howley PM. 1989. E2 polypeptides encoded by bovine

papillomavirus type 1 form dimers through the common carboxyl-terminal domain: Transactivation is mediated by the conserved amino-terminal domain. *Proc Natl Acad Sci USA 86*:510–514.

Merutka G, Dyson HJ, Wright PE. 1995. "Random coil" ^{1}H chemical shifts obtained as a function of temperature and TFE concentration for the peptide series GGXGG. *J Biomol NMR 5*:14–24.

Messerle BA, Wider G, Otting G, Weber C, Wüthrich K. 1989. Solvent suppression using a spin lock in 2D and 3D NMR spectroscopy with H_2O solutions. *J Magn Reson 85*:608–613.

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

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

Muhandiram DR, Kay LE. 1994. Gradient-enhanced triple-resonance three-dimensional NMR experiments with improved sensitivity. *J Magn Reson B 103*:203–216.

Neri D, Billeter M, Wider G, Wüthrich K. 1992a. NMR determination of residual structure in a urea-denatured protein, the 434-repressor. *Science 257*:1559–1563.

Neri D, Wider G, Wüthrich K. 1992b. Complete ^{15}N and ^{1}H NMR assignments for the amino-terminal domain of the phage 434 repressor in the urea-unfolded form. *Proc Natl Acad Sci USA 89*:4397–4401.

Neri D, Wider G, Wüthrich K. 1992c. ^{1}H, ^{15}N and ^{13}C NMR assignments of the 434 repressor fragments 1–63 and 44–63 unfolded in 7 M urea. *FEBS Lett 2*:129–135.

Shortle D. 1993. Denatured states of proteins and their roles in folding and stability. *Curr Opin Struct Biol 3*:66–74.

Sklenar V, Piotto M, Leppik R, Saudek V. 1993. Gradient-tailored water suppression for ^{1}H–^{15}N HSQC experiments optimized to retain full sensitivity. *J Magn Reson A 102*:241–245.

Smith LJ, Bolin KA, Schwalbe H, MacArthur MW, Thornton JM, Dobson CM. 1996. Analysis of main chain torsion angles in proteins: Prediction of NMR coupling constants for native and random coil conformations. *J Mol Biol 255*:494–506.

Spera S, Bax A. 1991. Empirical correlation between protein backbone conformation and Ca and Cb 13C nuclear magnetic resonance chemical shifts. *J Am Chem Soc 113*:5490–5497.

Stonehouse J, Keeler J. 1995. A convenient and accurate method for the measurement of the values of spin-spin coupling constants. *J Magn Reson A 112*:43–57.

Swindells MB, MacArthur MW, Thornton JM. 1995. Intrinsic ϕ,ψ propensities of amino acids, derived from the coil regions of known structures. *Nat Struct Biol 2*:596–603.

Travers A. 1993. *DNA–protein interactions*. London: Chapman & Hall.

Wishart DS, Bigam CG, Holm A, Hodges RS, Sykes BD. 1995. ^{1}H, ^{13}C and ^{15}N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects. *J Biomol NMR 5*:67–81.

Wishart DS, Sykes BD. 1994. The 13C chemical-shift index: A simple method for the identification of protein secondary structure using 13C chemical-shift data. *J Biomol NMR 4*:171–180.

Wüthrich K. 1989. *NMR of proteins and nucleic acids*. New York: John Wiley & Sons, Inc.

Wüthrich K. 1994. NMR assignments as a basis for structural characterization of denatured states of globular proteins. *Curr Opin Struct Biol 4*:93–99.

Yao J, Dyson HJ, Wright PE. 1997. Chemical shift dispersion and secondary structure prediction in unfolded and partly folded proteins. *FEBS Lett 419*:285–289.

Zhang O, Forman-Kay JD, Shortle D, Kay LE. 1997. Triple-resonance NOESY-based experiments with improved spectral resolution: Applications to structural characterization of unfolded, partially folded and folded proteins. *J Biomol NMR 9*:181–200.

Zhu G, Bax A. 1992. Improved linear prediction of damped NMR signals using modified forward backward linear prediction. *J Magn Reson 100*: 202–207.