Solution Structure of the DNA-Binding Domain of a Human Papillomavirus E2 Protein: Evidence for Flexible DNA-Binding Regions‡

Heng Liang, Andrew M. Petros, Robert P. Meadows, Ho Sup Yoon, David A. Egan, Karl Walter, Thomas F. Holzman, Terry Robins, and Stephen W. Fesik*

Pharmaceutical Discovery Division, Abbott Laboratories, Abbott Park, Illinois 60064

Received August 16, 1995; Revised Manuscript Received November 15, 1995⊗

ABSTRACT: The three-dimensional structure of the DNA-binding domain of the E2 protein from human papillomavirus-31 was determined by using multidimensional heteronuclear nuclear magnetic resonance (NMR) spectroscopy. A total of 1429 NMR-derived distance and dihedral angle restraints were obtained for each of the 83-residue subunits of this symmetric dimer. The average root mean square deviations of 20 structures calculated using a distance geometry-simulated annealing protocol are 0.59 and 0.90 Å for the backbone and all heavy atoms, respectively, for residues 2−83. The structure of the human virus protein free in solution consists of an eight-stranded β-barrel and two pairs of α-helices. Although the overall fold of the protein is similar to the crystal structure of the bovine papillomavirus-1 E2 protein when complexed to DNA, several small but interesting differences were observed between these two structures at the subunit interface. In addition, a β-hairpin that contacts DNA in the crystal structure of the protein−DNA complex is disordered in the NMR structures, and steady-state $^{1}H-^{15}N$ heteronuclear NOE measurements indicate that this region is highly mobile in the absence of DNA. The recognition helix also appears to be flexible, as evidenced by fast amide exchange rates. This phenomenon has also been observed for a number of other DNA-binding proteins and may constitute a common theme in protein/DNA recognition.

Human papillomaviruses are the causative agents of epithelial warts of the anogenital tract (Orth & zur Hausen, 1994), and several human papillomavirus serotypes have been linked to malignancies such as cervical neoplasia (Orth & zur Hausen, 1994; Richart & Wright, 1992). The E2 protein of papillomaviruses offers an attractive therapeutic target (Pepinsky *et al.*, 1994), since it is a transcription factor involved in the transactivation of several viral genes (Spalholz *et al.*, 1985; Haugen *et al.*, 1987) and, along with the E1 protein, is required for the DNA replication of papillomaviruses (Mohr *et al.*, 1990; Ustav & Stenlund, 1991).

The E2 protein consists of three functional domains: an N-terminal transactivation domain, a hinge domain, and a C-terminal domain of approximately 82 amino acids that is responsible for both DNA-binding and dimerization (Prakash *et al.*, 1992). A crystal structure of the DNA-binding domain (DBD)[1] of the E2 protein from bovine papillomavirus-1 (BPV-1) complexed with its cognate DNA site has been determined to a resolution of 1.7 Å (Hegde *et al.*, 1992). Here we describe the high-resolution solution structure of the DNA-binding domain of the E2 protein from human papillomavirus-31 (HPV-31) that was determined using heteronuclear multidimensional NMR spectroscopy. The solution structure of the HPV-31 E2-DBD determined in the absence of DNA is compared to the previously determined X-ray structure of BPV-1 E2 protein when bound to DNA. In addition, we characterize the backbone dynamics of the protein in solution by steady-state $^{1}H-^{15}N$ nuclear Overhauser effect (NOE) and amide exchange measurements.

MATERIALS AND METHODS

Protein Expression and Purification and NMR Sample Preparation. A gene fragment encoding the E2-DBD of HPV-31 (Goldsborough *et al.*, 1989) consisting of the C-terminal 82 amino acids was cloned into the pET-3b vector (Novagen, Madison, WI) for overexpression in *Escherichia coli* using the T7 system (Studier *et al.*, 1990). To obtain sufficient expression on minimal media for isotopic labeling, it was necessary to change the codon usage of two consecutive isoleucine residues (Ile6 and Ile7) near the N-terminus of the truncated protein from ATA to ATC. Cells of *E. coli* strain BL21(DE3)pLysS (Novagen, Madison, WI) transformed with the E2 expression plasmid were grown at 37 °C on M9 minimal media containing ^{15}N-labeled NH_4Cl and, for ^{13}C-labeling, uniformly ^{13}C-labeled glucose.

For protein purification, cells were resuspended in 50 mM Tris-HCl, 100 mM NaCl, 1 mM EDTA, and 5 mM dithiothreitol pH 8.3 (buffer A) plus 1 mM phenylmethanesulfonyl fluoride and lysed by repeated freezing/thawing followed by sonication. After centrifugation at 22700*g* for 45 min, the soluble cell lysate was loaded onto an S-Sepharose column. A linear gradient of buffer A containing from 100 to 500 mM NaCl was applied. Fractions containing the E2-DBD as identified by SDS−PAGE were pooled and diluted by at least 2-fold in 10 mM sodium phosphate, pH

‡ Atomic coordinates have been deposited with Protein Data Bank, Brookhaven National Laboratories, Long Island, NY 11973, under the accession code 1DHM.

* Address correspondence to this author at Abbott Laboratories, Department 47G, Building AP10, 100 Abbott Park Road, Abbott Park, IL 60064. Telephone: (847) 937-1201. Fax: (847) 938-2478. E-mail: fesik@steves.abbott.com.

⊗ Abstract published in *Advance ACS Abstracts,* February 1, 1996.

[1] Abbreviations: NMR, nuclear magnetic resonance; HPV, human papillomavirus; BPV, bovine papillomavirus; DBD, DNA-binding domain; 3D, three dimensional; NOE, nuclear Overhauser effect; rms, root mean square.

0006-2960/96/0435-2095$12.00/0

7.0. This solution was applied to a Mono-S column, which was washed with buffer A and eluted by a linear gradient of buffer A containing from 100 to 300 mM NaCl. Fractions containing the E2-DBD which were apparently homogeneous as judged by SDS−PAGE were pooled. N-terminal amino acid sequencing of the purified protein confirmed its identity and revealed that the additional N-terminal methionine encoded by the translation start codon in the expression vector was present in only roughly half of the recombinant protein, presumably due to the excision of the N-terminal methionine by the *E. coli* methionyl aminopeptidase (Hirel *et al.*, 1989).

The NMR samples were prepared by exchanging the purified protein into a H_2O/D_2O (9:1) or D_2O buffer containing 20 mM sodium phosphate (pH 6.5) and 5 mM dithiothreitol (d_{11}). The final concentration of the NMR samples was approximately 1.7 mM E2 monomer. An extinction coefficient of $\epsilon_{280} = 36\,040\ M^{-1}\ cm^{-1}$ was used for measuring the protein concentration.

A sample of the E2-DBD containing asymmetrically labeled heterodimers was prepared by mixing equal molar amounts of ^{13}C-labeled and unlabeled proteins (Weiss, 1990; Arrowsmith *et al.*, 1990; Folkers *et al.*, 1993). Urea was added to this mixture to the final concentration of 9 M to denature the protein in order to allow the exchange of labeled and unlabeled subunits. The protein was refolded by slowly reducing the concentration of urea by dialysis.

NMR Spectroscopy. NMR spectra were recorded at 30 °C on a Bruker AMX500, DMX500, or AMX600 NMR spectrometer. The NMR data were processed and analyzed on Silicon Graphics computers using in-house written software. Pulsed field gradients were applied where appropriate as described (Bax & Pochapsky, 1992) to afford the suppression of the solvent signal and spectral artifacts. Quadrature detection in indirectly detected dimensions was accomplished by using the States−TPPI method (Marion *et al.*, 1989b). Linear prediction was employed as described (Olejniczak & Eaton, 1990).

To obtain backbone ^1H, ^{13}C, and ^{15}N assignments, three-dimensional (3D) HNCA (Kay *et al.*, 1990), HN(CO)CA (Bax & Ikura, 1991), and CBCA(CO)NH (Grzesiek & Bax, 1992) spectra were acquired. In addition, a 3D HNHA spectrum (Vuister & Bax, 1993) was obtained. For side chain assignments, an ^{15}N-separated 3D TOCSY-HSQC spectrum (Marion *et al.*, 1989a) was acquired using the clean TOCSY mixing scheme (Griesinger *et al.*, 1988) with a mixing time of 38 ms, and a 3D constant-time HNHB spectrum (Archer *et al.*, 1991) was recorded. A 3D HCCH-TOCSY spectrum (Kay *et al.*, 1993) was acquired with a mixing time of 13 ms using the DIPSI-2 sequence (Rucker & Shaka, 1989) for ^{13}C isotropic mixing. In order to obtain assignments for side chains of the aromatic residues, ^1H, ^1H homonuclear 2D TOCSY and NOESY spectra were acquired using an unlabeled protein sample dissolved in the D_2O buffer. The mixing times were 25 and 80 ms for the TOCSY and NOESY spectrum, respectively. Stereospecific assignments of methyl groups of the valine and leucine residues were obtained by using a biosynthetic approach (Neri *et al.*, 1989) on the basis of the $^{13}C-^{13}C$ one-bond coupling pattern observed in a high-resolution $^1H,^{13}C$-HSQC spectrum (Bodenhausen & Ruben, 1980) of a fractionally ^{13}C-labeled protein sample.

Distance restraints were derived from an ^{15}N-separated 3D NOESY-HSQC spectrum (Fesik & Zuiderweg, 1988; Marion *et al.*, 1989c) and a ^{13}C-separated 3D NOESY-HMQC spectrum (Fesik & Zuiderweg, 1988; Marion *et al.*, 1989c), both with a mixing time of 80 ms. To detect NOEs between amide groups with degenerate proton chemical shifts, a 3D (^{15}N, ^{15}N, and ^1H) HMQC-NOESY-HSQC spectrum was acquired with a mixing time of 80 ms. The pulse sequence was essentially the same as the HMQC-NOESY-HMQC (Ikura *et al.*, 1990; Frenkiel *et al.*, 1990) except that the second HMQC unit was replaced with an HSQC sequence employing gradient coherence order selection and sensitivity enhancement schemes (Kay *et al.*, 1992). To detect NOEs between two monomeric subunits of the protein, a 3D ^{13}C-filtered, ^{13}C-separated NOESY spectrum was collected using an asymmetrically labeled sample described above. The pulse scheme consisted of a double ^{13}C-filter sequence (Gemmecker *et al.*, 1992) concatenated with a NOESY-HMQC sequence (Fesik & Zuiderweg, 1988; Marion *et al.*, 1989c). The spectrum was recorded with a mixing time of 80 ms.

To identify amide groups that exchange slowly with the solvent, a series of ^1H, ^{15}N-HSQC spectra (Bodenhausen & Ruben, 1980) were recorded at 20 °C at 2 h intervals after the lyophilized protein was dissolved in D_2O. The acquisition of the first HSQC spectrum was started 8 min after the addition of D_2O.

Steady-state $^1H-^{15}N$ heteronuclear NOEs were measured by using a pulse sequence (Farrow *et al.*, 1994) that incorporates gradient coherence selection and sensitivity enhancement schemes. Three data sets each consisting of a spectrum with proton saturation (NOE spectrum) and one without proton saturation (NONOE spectrum) were collected. To ensure the accurate measurement of NOEs of amide groups that are in fast exchange with water, care was taken to avoid saturation of the water resonance in the NONOE spectrum (Farrow *et al.*, 1994). A cycle time of 7 s was used for each scan.

NMR-Derived Restraints and Structure Calculations. Distance restraints derived from the NOE data were classified into six categories based on the NOE cross peak intensity and given a lower bound of 1.8 Å and upper bounds of 2.5, 3.0, 3.5, 4.0, 4.5, and 5.0 Å, respectively. Restraints for ϕ torsional angles were derived from $^3J_{HN\alpha}$ coupling constants measured from the 3D HNHA spectrum (Vuister & Bax, 1993). The ϕ angle was restrained to $120° \pm 40°$ for $^3J_{HN\alpha} > 8.5$ Hz and $60° \pm 40°$ for $^3J_{HN\alpha} < 5$ Hz. Hydrogen bonds, identified for slowly exchanging amides based on initial structures, were defined by two restraints: 1.8−2.5 Å for the H−O distance and 1.8−3.3 Å for the N−O distance.

Structures were calculated with the X-PLOR 3.1 program (Brünger, 1992) on Silicon Graphics computers using a hybrid distance geometry−simulated annealing approach (Nilges *et al.*, 1988). As the E2 protein is a symmetric dimer, NOE-derived distance restraints were treated differently based on their origin with respect to whether they are due to intra- or intermonomer correlations. NOEs observed in the ^{13}C-filtered, ^{13}C-edited NOESY spectrum were treated as unambiguous intermonomer distance restraints and given lower and upper distance bounds of 1.8 and 5.0 Å, respectively. NOEs observed between protons located near the monomer−monomer interface that may contain contribu-

tions from intersubunit correlations were considered ambiguous and treated with a target function that computes the $\langle r^{-6} \rangle$ average of the intra- and intermonomer distances as described by Nilges (1993). Other NOEs were treated as intramonomer distance restraints. Structure calculations were carried out in two steps. First, the structure of a monomer was calculated with only intramonomer restraints from a linear representation of the polypeptide chain as the starting structure using the distance geometry and simulated annealing protocol described in the X-PLOR manual (Brünger, 1992). Converged low-energy structures of the folded monomer were then used as starting structures for the structure calculations of the dimer. To generate starting structures of the dimer, coordinates of a monomer were first duplicated. Then one monomer was translated in space in a random fashion with respect to the other. This was followed by a rigid body energy minimization (Brünger, 1992) using intermonomer distance restraints which allows the two monomers to come together and to be correctly oriented with respect to one another. These starting structures of the dimer were then subjected to a standard dynamical simulated annealing protocol (Brünger, 1992) to yield structures derived from NMR restraints. Only one monomer in the dimer was subjected to experimental distance and angular restraints during the calculation. Both monomeric subunits were subjected to symmetry restraints, which were applied in the form of the X-PLOR noncrystallographic symmetry restraint and distance symmetry restraints as described (Nilges, 1993).

RESULTS AND DISCUSSION

NMR Assignments. The ^{1}H, ^{13}C, and ^{15}N resonances of the DNA-binding domain of the HPV-31 E2 protein were assigned from an analysis of several 3D double- and triple-resonance NMR spectra (Bax & Grzesiek, 1993). The only amides that could not be assigned (Ala2, Cys41, Lys47, and Thr58) are located either at the N-terminus of the protein or in solvent-accessible loops or turns. The amide signals corresponding to these residues could not be detected, as essentially all correlations observed in the HNCA/HN(CO)-CA data sets have been accounted for. In addition, ^{1}H and ^{13}C chemical shifts were assigned for nearly all side chain resonances. Stereospecific assignments of all the methyl groups of Leu and Val residues were obtained by using a biosynthetic approach following the method of Wüthrich and co-workers (Neri *et al.*, 1989). The ^{1}H, ^{13}C, and ^{15}N chemical shift assignments of the HPV-31 E2-DBD are given as Supporting Information.

Structure Determination. On the basis of the ^{1}H, ^{13}C, and ^{15}N chemical shift assignments, the intramonomer NOEs obtained from 3D ^{15}N- and ^{13}C-edited NOESY spectra were assigned. A total of 1263 approximate interproton distance restraints were derived from the NOE data for each subunit. In addition, 19 unambiguous intermonomer distance restraints were derived from a 3D ^{13}C-filtered, ^{13}C-edited NOESY spectrum acquired on a sample that contains heterodimers composed of a ^{13}C-labeled monomeric subunit and an unlabeled monomeric subunit prepared by asymmetric isotopic labeling (Weiss, 1990; Arrowsmith *et al.*, 1990; Folkers *et al.*, 1993). An additional 50 restraints that involve protons located near the monomer−monomer interface were included in the structure calculations. Since these NOEs could have arisen from intra- and/or intermonomer interactions, they were treated differently using a target function that takes these possibilities into account (Nilges, 1993). Of the 1263 intra-monomer NOE restraints, 354 were intraresidue, 532 were sequential or short-range between residues separated in the primary sequence by less than five amino acids, and 377 were long-range involving residues separated by at least five residues. The distribution of NOE restraints along the primary sequence of the protein is shown in Figure 1A.

In addition to the NOE distance restraints, 29 ϕ dihedral angle restraints were included in the structure calculations that were derived from three-bond coupling constants ($^3J_{\mathrm{HN}\mathrm{H}\alpha}$) determined from an HNHA spectrum (Vuister & Bax, 1993). The experimental restraints also included 68 distance restraints corresponding to 34 hydrogen bonds. The amides involved in hydrogen bonds were identified on the basis of their characteristically slow exchange rate and the hydrogen bond partners from initial NMR structures calculated without the hydrogen bond restraints. The total number of nonredundant, experimentally derived restraints was 1429 for each monomer, corresponding to 17.2 restraints per residue.

Quality of the NMR Structures. The structural statistics for 20 converged low-energy structures are summarized in Table 1. The structures are in good agreement with the NMR experimental restraints; there are no distance violations greater than 0.12 Å and no dihedral angle violations by greater than 1°. In addition, the simulated energy for the van der Waals repulsion term is small, indicating that the structures are devoid of bad interatomic contacts. As shown in Table 1, the NMR structures also exhibit good covalent bond geometry, as indicated by small bond-length and bond-angle deviations from the corresponding idealized parameters.

A stereoview of the superposition of the 20 NMR structures of the HPV-31 E2-DBD is shown in Figure 2. The average atomic root mean square (rms) deviation of the 20 structures of dimer from the mean coordinates is 0.59 Å for backbone heavy atoms (N, C_α, and C′) and 0.90 Å for all non-hydrogen atoms. The residue-based rms deviations are shown in Figure 1 (panels B and C). Except for the two residues at the N-terminus and a disordered loop (residues 40−48), the three-dimensional structure of the E2-DBD of HPV-31 in solution is very well defined by experimental NMR restraints. When these residues (1, 2, 40−48) are excluded, the average atomic rms deviation is 0.21 and 0.56 Å for backbone and all non-hydrogen atoms, respectively.

Further evidence for the convergence of the NMR structures is provided by examining the torsion angle variability of these structures using the angular order parameter (S^{angle}) (Hyberts *et al.*, 1992) shown in Figure 1 (panels D and E), where values of S^{angle} approaching zero correspond to an essentially random distribution of dihedral angles, whereas those near unity indicate torsional angle convergence. For the 20 NMR structures of the HPV-31 E2-DBD, 67 out of the 81 residues (excluding N- and C-termini) have S^{angle} values for both Φ and Ψ angles greater than 0.95, indicating that the NMR-derived structures are highly converged. A Ramachandran plot of the Φ and Ψ angles is shown in Figure 3. The Φ and Ψ dihedral angles of all non-glycine residues lie in the stereochemically allowed region, with 89% of the residues with converged dihedral angles in the most favored region of the Ramachandran plot (Figure 3).

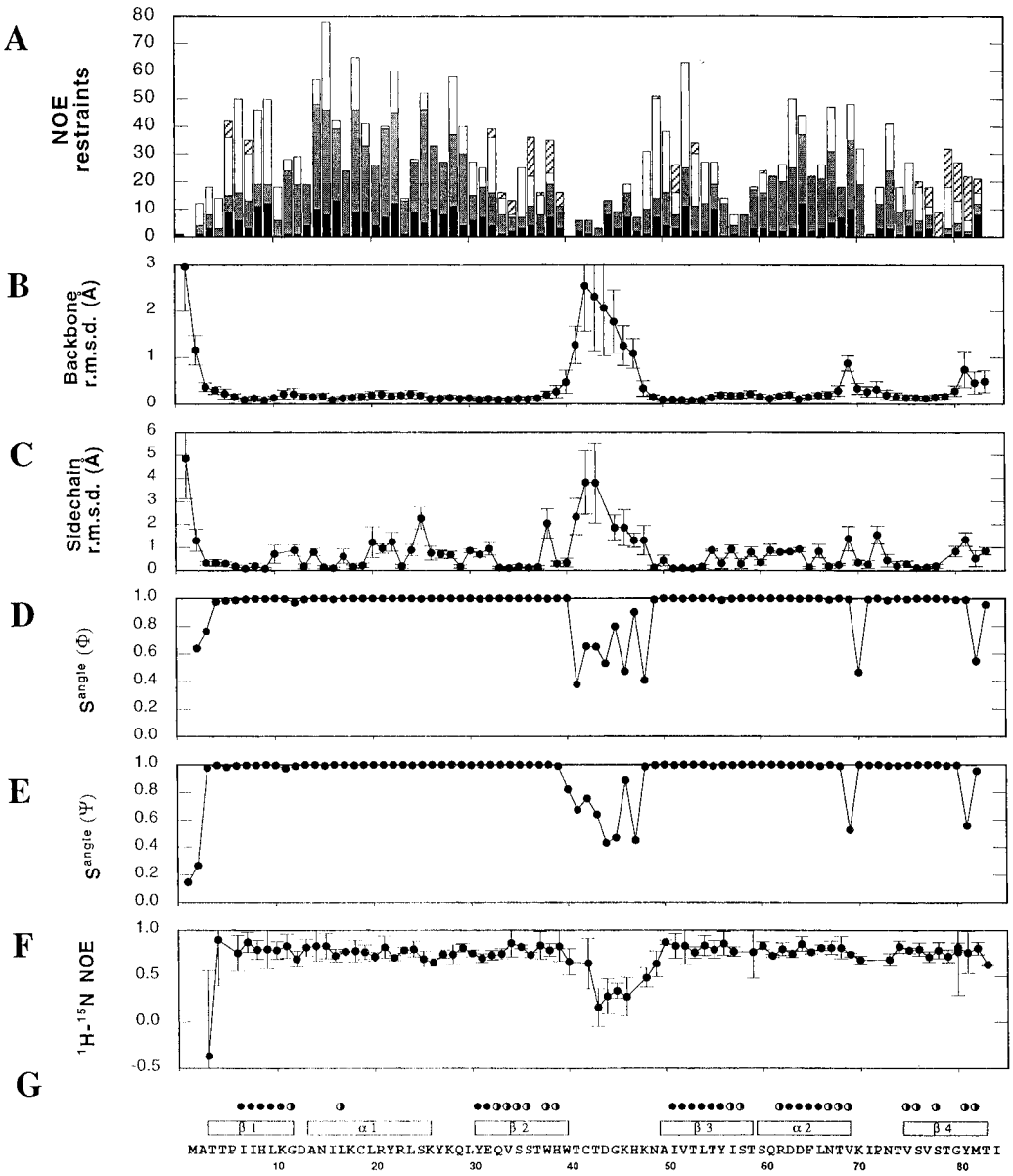

FIGURE 1: (A) Distribution of NOE restraints. The height reflects the total number of NOE-derived restraints for each residue while the black, gray, white, and hatched bars correspond to intraresidue restraints, sequential or short-range restraints, long-range, and intersubunit (both ambiguous and unambiguous) restraints, respectively. The average root-mean-square deviations of coordinates of 20 NMR-derived structures from the averaged coordinates are shown in panel B for backbone heavy atoms (C_α, C′, N, and O) and in panel C for all non-hydrogen atoms. Error bars represent the variation (one σ) among the 20 structures. Angular order parameters, S^{angle}, of each residue are shown for the ϕ (panel D) and ψ (panel E) dihedral angles, respectively. (F) Heteronuclear $^1H-^{15}N$ NOE. Error bars represent the standard deviation of NOE values measured in three data sets. (G) Slowly exchanging amide groups. Filled circles represent amides that remained protected from the solvent after 24 h, and half-filled circles represent amides that were protected after 2 h but were nearly completely exchanged after 24 h. Boxes represent secondary structural elements of the E2-DBD as defined by NMR structures. The amino acid sequence of the E2-DBD from HPV-31 used in the NMR studies is shown at the bottom. Residues 2−83 correspond to residues 291−372 of the full-length E2 protein from HPV-31.

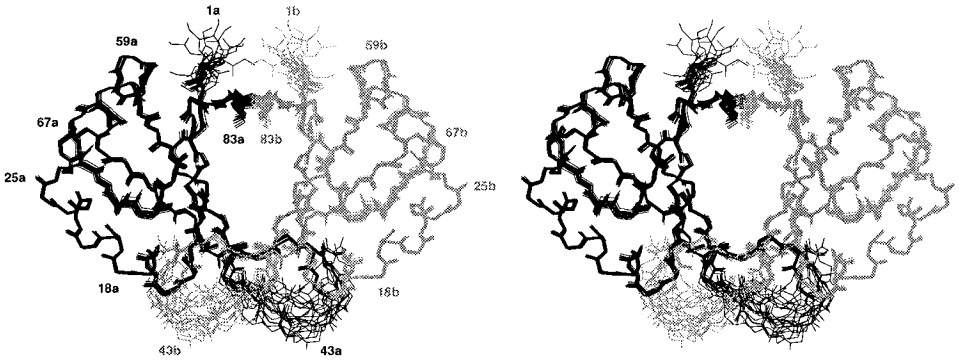

FIGURE 2: Stereoview of the superposition of 20 simulated-annealing structures calculated from NMR-derived restraints. Residue numbers are indicated and the two monomeric subunits (shaded differently) are designated “a” and “b”, respectively.

Table 1: Structural Statistics of NMR Structures[a]

structural statistics	⟨SA⟩	$\overline{\langle SA \rangle}_r$
rms deviation from experimental distance restraints (Å)[b,c]		
intraresidue (354)	0.005 ± 0.003	0.006
sequential (304)	0.010 ± 0.001	0.009
short range (228)	0.010 ± 0.001	0.012
long range (377)	0.012 ± 0.001	0.013
unambiguous intermonomer (19)	0.023 ± 0.004	0.023
ambiguous inter- or intramonomer (50)	0.014 ± 0.001	0.015
hydrogen bonds (68)	0.022 ± 0.002	0.020
rms deviation from experimental torsional angle restraints (deg)[c]		
Φ angles (29)	0.07 ± 0.05	0.00
X-PLOR potential energies (kcal mol^{-1})[d]		
E_{total}	130 ± 2	134
E_{bond}	10.0 ± 0.2	10.4
E_{angle}	77 ± 1	80
$E_{improper}$	14.4 ± 0.4	14.5
E_{repel}	15.9 ± 0.8	16.3
E_{NOE}	9 ± 1	9
E_{CDIH}	0.01 ± 0.01	0
E_{NCS}	3.6 ± 0.4	3.4
Lennard-Jones potential energy (kcal mol^{-1})[e]	−942 ± 18	−934
rms deviation from idealized covalent geometry		
bonds (Å)	0.002 ± 0.000	0.002
angles (deg)	0.3 ± 0.0	0.3
impropers (deg)	0.07 ± 0.05	0.38
Cartesian coordinate rms differences (Å) ⟨SA⟩ vs $\overline{\langle SA \rangle}$ of the dimer	N, C_α, and C′	all heavy atoms
residues 2−83	0.59 ± 0.16	0.90 ± 0.13
residues 3−39 and 49−83	0.24 ± 0.05	0.63 ± 0.05

[a] ⟨SA⟩ is the ensemble of 20 simulated-annealing structures; $\overline{\langle SA \rangle}$ is the averaged Cartesian coordinates obtained by averaging individual structures following a least-squares superposition of the backbone heavy atoms (N, C_a, C′) for residues 3−39 and 49−83; $\overline{\langle SA \rangle}_r$ is the energy minimized averaged Cartesian coordinates. [b] NOE-derived distance restraints consist of intraresidue, sequential (between protons in adjacent residues), short-range (involving residues separated in sequence by more than one but less than five residues), and long-range (separated by at least five residues) restraints for each monomer. Unambiguous intermonomer restraints were derived from ^{13}C-edited and ^{13}C-filtered NOE spectrum obtained on the asymmetrically ^{13}C-labeled heterodimer. Ambiguous inter- or intramonomer restraints were satisfied by the $\langle r^{-6} \rangle$ average of the corresponding inter- and intramonomer distances. [c] No distance restraint was violated by more than 0.12 Å, and no torsional angle restraints by more than 1° in any of the 20 structures. [d] The square-well NOE (E_{NOE}) and torsional angle (E_{CDIH}) potentials were calculated with a force constant of 50 kcal mol^{-1} Å^{-2} and 200 kcal mol^{-1} rad^{-1}, respectively. The quadratic van der Waals repulsion term E_{repel} was calculated with a force constant of 4 kcal mol^{-1} Å^{-4} with atomic radii set to 0.8 times their CHARMM (Brooks *et al.*, 1983) values. The noncrystallographic symmetry potential (E_{NCS}) was calculated with a force constant of 100 kcal mol^{-1} Å^{-2}. The parameters used were essentially the same as in the X-PLOR parallhdg.pro parameter set. [e] The Lennard-Jones potential was not used during any stage of the structure calculations.

Description of the Structure. A ribbon representation of the solution structure of the E2-DBD from HPV-31 is shown in Figure 4. The structure consists of a homodimer with an eight-stranded β-barrel at the center of the molecule that is formed from a four-stranded antiparallel β-sheet from each monomer. The secondary structure of each monomer contains two $\beta-\alpha-\beta$ repeats, and, as a result, the dimer has an approximate 4-fold symmetry (Figure 4). The first helix (α_1) of each monomer corresponds to the "recognition helix"

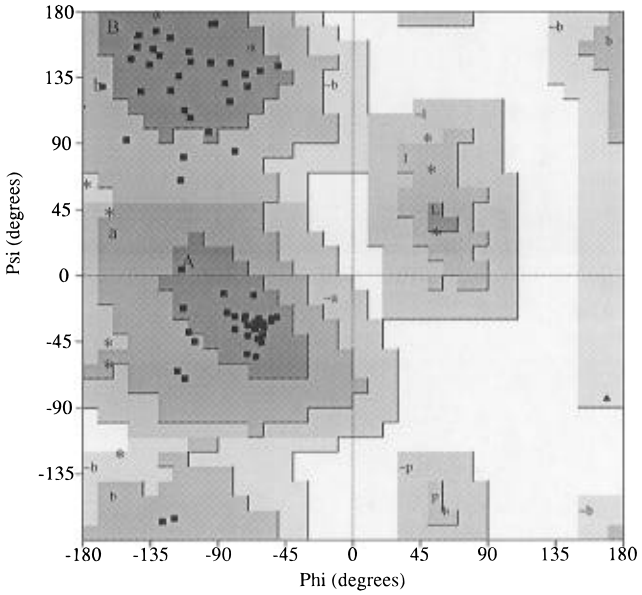

FIGURE 3: Ramachandran plot of the Φ and Ψ angles observed in the averaged and energy-minimized coordinates. Glycines are plotted as triangles, and asterisks are residues for which the dihedral angles residues are not well defined (with S^{angle} for Φ or Ψ less than 0.90). The shaded regions correspond to the most favorable and allowed regions as defined by ProCheck (Morris *et al.*, 1992).

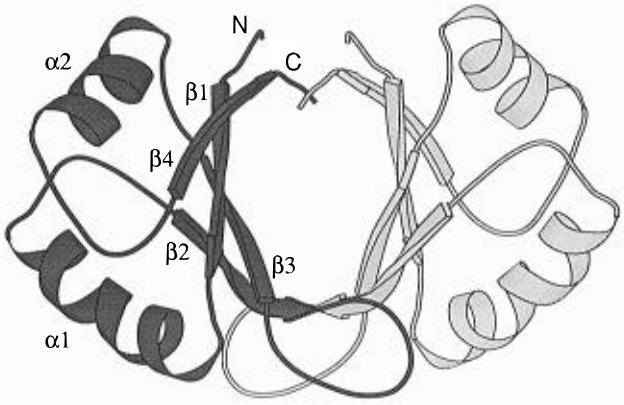

FIGURE 4: Ribbon representation (Kraulis, 1991) of the solution structure of the E2-DBD of HPV-31. The same view as in Figure 2 is shown, and the two monomeric subunits are shaded differently. Secondary structural elements as well as amino and carboxy termini are indicated.

that forms specific base interactions in the crystal structure of the E2/DNA complex (Hegde *et al.*, 1992).

The hydrophobic core of each monomer consists of residues on the outer face of the dome-shaped β-barrel and the buried hydrophobic residues of the amphipathic α-helices. Essentially all of these residues are conserved among E2 proteins from different papillomavirus strains. Most notably, Leu 16 is invariant among all known human papillomavirus E2 proteins (Figure 5). In addition to the conserved hydrophobic residues and residues that make specific protein−DNA contacts, Gly11 is also invariant among E2 proteins from different papillomavirus strains. In our NMR structure of the E2-DBD from HPV-31, this residue is part of the tight turn leading to the recognition helix. It exhibits dihedral angles that cannot be accommodated by other amino acids (Figure 3), which may account for the requirement of a glycine at this position.

Comparison to the Structure of the BPV-1 E2 Protein. A crystal structure of the DNA-binding domain from BPV-1

```
                10        20        30        40        50        60        70        80
                 |         |         |         |         |         |         |         |
HPV-31   ATTPIIHLKG DANILKCLRY RLSK-YKQLY EQVSSTWHWT CTD-GKHKN- AIVTLTYIST SQRDDFLNTV KIPNTVSVST -GYM-TI

BPV-1    --SCFALISG TANQVKCYRF RVKKNHRHRY ENCTTTWFTV ADNGAERQGQ AQILITFGSP SQRQDFLKHV PLPPGMNIS- -GFTASLDF

HPV-5    RDPPVIIVKG AANTLKNVRN RAKIKYMGLF RSFSTTWSWV AGDGTERLGR PRMLISFSSY TQRRDFDEAV RYPKGVD-KA YGNLDSL
HPV-8    IDPPVILVRG EANTLKCFRN RARVRYRGLF KYFSTTWSWV AGDSTERLGR SRMLILFTSA GQRKDFDETV KYPKGVD-TS YGNLDSL
HPV-14   LDPPVILVRG DPNTLRCFRN RAKQKFTGLY RAFSTAWSWV AGDGTERLGR SRMLISFFSF NQRRDFDQTV KYPKGVD-RS FGSFDSL
HPV-16   NTTPIVHLKG DANTLKCLRY RFKK-HCTLY TAVSSTWHWT GHN-VKHKS- AIVTLTYDSE WQRDQFLSQV KIPKTITVST -GFM-SI
HPV-17   RDPPVILLRG EANKLKCFRY RAKKRYGSLV KYYSTTWSWV GANTNDRIGR SRMLLAFNTY DERELFIQKM KLPPGVD-WS LGHLDDL
HPV-18   NTTPIIHLKG DRNSLKCLRY RLRKH-SDHY RDISSTWHWT GAGNEK---T GILTVTYHSE TQRTKFLNTV AIPDSVQILV -GYM-TM
HPV-30   KTTPVVHLKG EPNRLKCLRY RCQK-HKHLF VNISSTYHWT NT----HTEY SYITVVYKDE TQRANFLNVV KIPPSIKIVM -GHMTGVDM
HPV-33   NVAPIVHLKG ESNSLKCLRY RLKP-YKELY SSMSSTWHWT SDN-KNSKN- GIVTVTFVTE QQQQMFLGTV KIPPTVQIST -GFM-TL
HPV-34   NVAPIVHLKG DKNSLKCLRY RMHKGYSHLF NNVTTTWHWT NNTNSK---C GVITFMFSST SQQKQFLQCA KIPPTISVSS -GYM-SI
HPV-35   TTTPIVHLKG DANTLKCLRY RLGK-YKALY QDASSTWRWT CTN-DKKQI- AIVTLTYTTE YQRDKFLTTV KIPNTVTVSK -GYM-SI
HPV-39   NTTPIIHLKG DKNGLKCLRY RLQK-YDTLF ENISCTWHWI RG--KGTKNA GILTVTYATE SQRQKFLDTV KIPSSVHVS- LGYM-TL
HPV-40   SSTPIIQLEG EANCLKCFRY RLGKV-SHLF CNSSTTWRWT TESRTEKN-- AIITLTYSSV QQRSDFLAIV KIPKTIKHS- LGML-TLM
HPV-41   EGHYLVGAKG PVNSLRCLRY KWKNKYSGDI MYLGTTFTWT ESDGTERCGS GRFFCAFSNE TKREKFLKSV KIPKNIGLFR AHAEKL
HPV-42   QATPVIHLQG DPNCLKCLRF RLKRNCSHLF TQVSSTWHLT ENDCTRDTKT GIITIHYYDE AQRNLFLNTV KIPSGIKSC- IGYMSMLQFI
HPV-45   NTTPIIHLKG DKNSLKCLRY RLR-KYADHY SEISSTWHWT -GC---NKNT GILTVTYNSE VQRNTFLDVV TIPNSVQIS- VGYM-TI
HPV-51   QTAFIVHLKG DTNCLKCFRY RFTK-HKGLY KNVSSTWHWT SNTKT----- GIVTIVFDSA HQRETFIKTI KVPPSVTLS- LGIM-TL
HPV-52   CTAPIIHLKG DPNSLKCLRY RVKT-HKSLY VQISSTWHWT SNECTNNKL- GIVTITYSDE TQRQQFLKTV KIPNTVQVIQ -GVM-SL
HPV-56   KTTPVVHLKG EPNRLKCCRY RFQK-YKTLF VDVTSTYHWT STD---NKNY SIITIIYKDE TQRNSFLSHV KIPVVYRLVW DK
HPV-57   DAVPVIHLQG EANCLKCFRY RVQKHKDVLF VKASSTWHWA ---CGNGDKT AFVTLWYKSQ EQRAEFLTRV HLPKGVKAL- PGYM-SAFV
HPV-58   KVSPIVHLKG DPNSLKCLRY RLKP-FKDLY CNMSSTWHWT SDD-KGDKV- GIVTVTYTTE TQRQLFLNTV KIPPTVQIST -GVM-SL
```

FIGURE 5: Sequence alignment of the DNA-binding domain of the E2 proteins from different human papillomavirus strains that are associated with malignancy (Doeberitz, 1992), along with that of the BPV-1 protein for which the crystal structure was determined. Numbering indicated on the top corresponds to that of the E2 DNA-binding domain from HPV-31 shown in Figure 1.

when complexed to its target DNA was previously reported (Hegde *et al.*, 1992). The E2 proteins from the bovine virus BPV-1 and the human virus HPV-31 share an amino acid sequence identity of 31% over approximately 80 residuesof the DNA-binding domain (Figure 5). A comparison of the NMR structure of the uncomplexed E2-DBD from HPV-31 and the crystal structure of the E2 from BPV-1 when complexed to DNA is shown in Figure 6. The secondary structures of the two proteins are essentially identical, including the β-bulge observed in the second β-strand. Thus, it appears that this β-bulge which enables Thr359 of BPV-1 E2 to form a hydrogen bond with the phosphate of DNA in the crystal structure of the protein−DNA complex (Hegde *et al.*, 1992) is preformed when the protein is free in solution.

The tertiary structure of the individual monomers of the two proteins is very similar to each other, with an rms difference of 1.3 Å for the 68 aligned C_α atoms corresponding to residues 3−39 and 49−80 of the HPV protein (Figure 6A). However, if we compare the protein dimers upon superposition of only one monomer, there is a systematic deviation of the second monomer as shown in Figure 6A. This difference is apparently due to a change in the alignment of the C-terminal β-strands of the two monomers at the subunit interface as illustrated in Figure 6B. When one subunit (subunit "a" on the left in Figure 6A) is superimposed, residue Y80a of the HPV protein aligns with T405a of the BPV protein (Figure 6B). However, in subunit b, there is a shift between Y80b of the HPV-31 E2 protein free in solution and T405b of the BPV-1 E2 protein when complexed to DNA.

Due to the coincidence of this region with the dyad axis of the dimer, the chemical shift degeneracy of resonances in this region and the small difference between the two structures, relative few NOEs could be relied on to distinguish the two structures. However, in the ^{13}C-edited and ^{13}C-filtered NOE spectrum, we observed intersubunit NOEs between the aromatic ring protons of Y80 in one monomer and the methyl groups of both T78 and T82 in the other monomer. This indicates that, in the HPV E2 protein in solution, Y80 is at the dyad axis of the dimer and is opposite to the other Y80 from the second subunit as shown in Figure 6B. The intersubunit NOEs observed between the methyl group of Thr78 and the H^δ and H^ϵ ring protons of Y80 would clearly be inconsistent with alignment of the C-terminal strands based on the crystal structure of BPV-1 E2 bound to DNA. This suggests that there is a real structural difference in this region between the structures of HPV-31 protein in solution and the BPV-1 protein when complexed to DNA. As a result of the difference in the alignment of C-terminal β-strands at the subunit interface, the recognition helix of subunit "b" in the DNA-bound form of the BPV-1 E2 tilts toward DNA relative to its counterpart in the free form of the HPV protein (Figure 6A). It is tempting to speculate that this may represent a conformational change induced by DNA binding. However, it is also possible that the differences between the two structures are due to difference in their amino acid sequences. It would be interesting to determine the structure of either the DNA-bound form of the HPV-31 E2 or the uncomplexed form of the BPV-1 protein to address this question.

Additional differences are observed at the subunit interface between the structures of the HPV-31 and BPV-1 E2 proteins. The HPV protein contains a second pair of tryptophans (W39), in addition to the first pair (W37) which are also present in the bovine virus protein and were found to stabilize of the dimer of the BPV-1 E2 protein (Corina *et al.*, 1993). As shown in Figure 7, two indole rings of W37 of the HPV-31 E2-DBD, one from each monomer, stack in parallel to one another, similar to those in the bovine virus protein. The indole rings of a second pair of tryptophans (W39), which are not found in the BPV protein but are conserved among most E2 proteins from different HPV serotypes (Figure 5), stack roughly perpendicular to the first pair (W37) giving rise to favorable aromatic−aromatic interactions (Burley & Petsko, 1985; Singh & Thornton, 1985). Due to this stacking, the aromatic ring proton Hζ3 of W37 is placed on top of the indole ring of W39 (Figure 7), and, as a result, its proton resonance is shifted to high field by more than 3 ppm to 3.91 ppm.

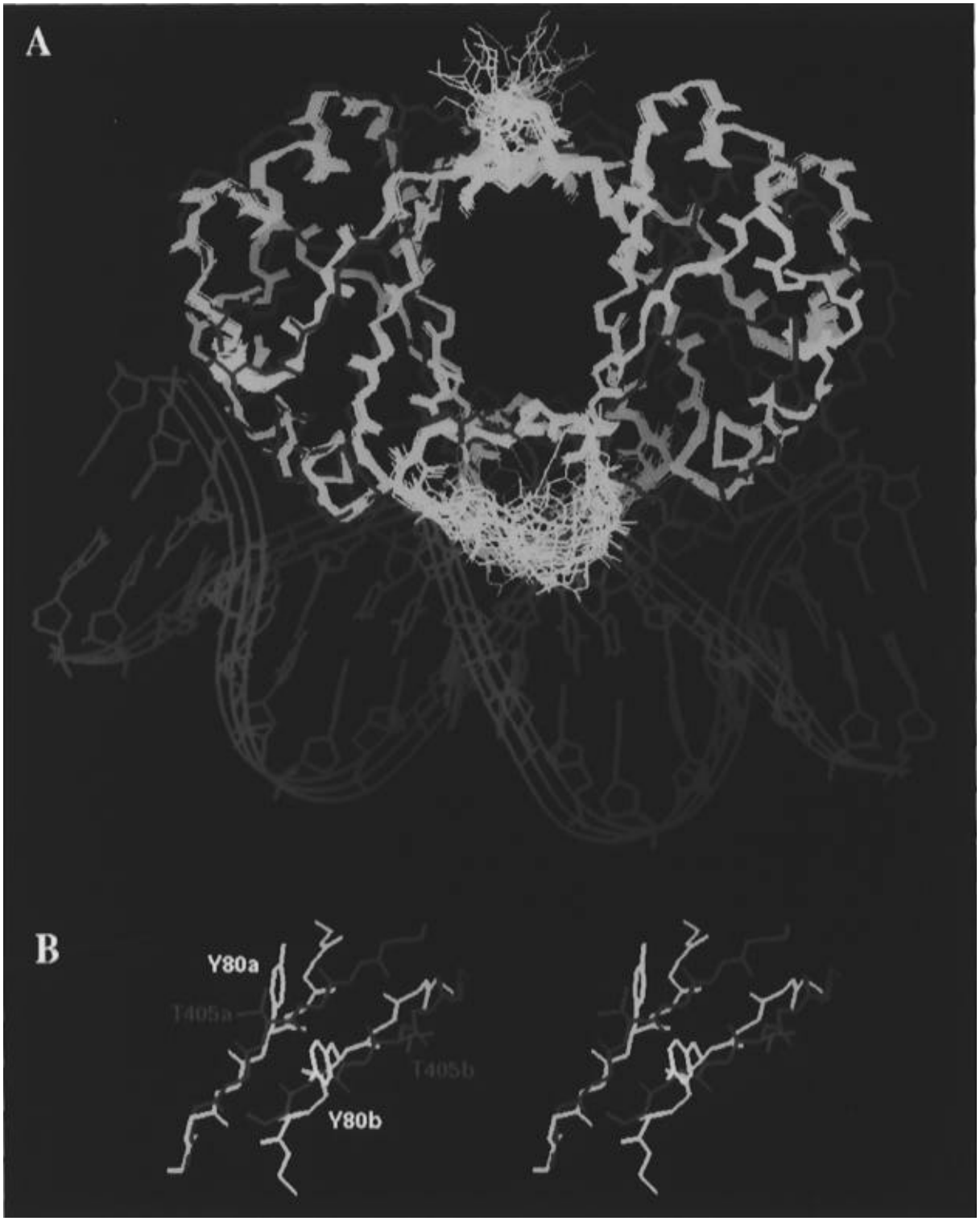

FIGURE 6: (A) Overlay of the ensemble of 20 NMR structures of the E2-DBD of HPV-31 (white) and the crystal structure of E2-DBD from BPV-1 (red) complexed with DNA (blue). Shown is the same view as in Figures 2 and 4. Structures of only one subunit (subunit "a" on the left) are best superposed. (B) View from top down of the overlay shown in panel A showing the alignments of the C-terminal β-strands in the NMR structure of HPV-31 E2 free in solution (white) and the crystal structure of the BPV-1 E2 complexed to DNA (red). The best superposition of the two structures is obtained for subunit "a" as shown in panel A. The NMR structure is represented by the averaged and energy-minimized coordinates. Side chains of Y80a and Y80b of the HPV-31 E2-DBD, as well as those of T405a and T405b of BPV-1 E2, are drawn in addition to the backbone atoms.

Flexibility of Regions That Interact with DNA. Protein−DNA contacts observed in the co-complex crystal structure of the E2-DBD of BPV-1 are provided by two regions of the protein: α-helix-1 that contacts base pairs of the DNA in the major groove ("recognition helix") and a β-hairpin that interacts with the minor groove. In the solution structures of E2-DBD of HPV-31, residues 40−48, which correspond to the DNA-binding β-hairpin, are not well-defined by NMR data (Figures 1B−E and 6A). In order to determine whether this represents structural flexibility of this region or is simply due to the lack of experimental restraints, we measured the steady-state $^{1}H-^{15}N$ heteronuclear NOE of the E2 DBD. The data are shown in Figure 1F. The loop region (residues 40−48) exhibits small steady-state

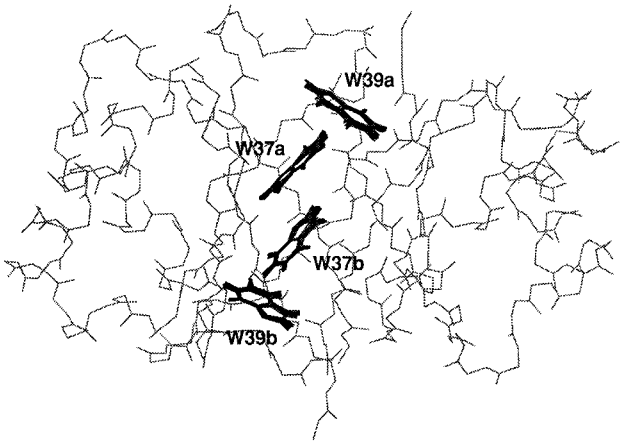

FIGURE 7: Stacking of tryptophans at the subunit interface in HPV-31 E2-DBD. The indole rings of Trp37 and Trp39 in the ensemble of 20 NMR structures are shown. Light gray lines represent the backbone of the protein in the averaged and energy-minimized structure.

heteronuclear NOEs, indicating that this loop is highly mobile in solution in the absence of DNA.

Evidence for the flexibility of the recognition helix in solution is obtained from amide exchange studies. Patterns of sequential and short-range NOE connectivities, $^3J_{HNH\alpha}$ coupling constants, and the systematic deviations of C_α and H_α chemical shifts from random coil values (data not shown) all indicate that residues 14−26 exist in a helical conformation. This is confirmed by the structure calculations which yield a well-defined α-helix for these residues (Figure 2). However, most amide groups of the recognition helix exchange relatively fast with the solvent, with time constants faster than hours as measured by H−D isotope exchange (Figure 1G). This is unlike the amide groups of a typical α-helix such as those of the second helix (residues 59−68) which exchange slowly with life times of days. This suggests that even though the time-averaged conformation of this region as defined by the NOE data is α-helical, there exists at least one alternative conformation in which the amide groups of these residues are not protected from the solvent by the protein hydrogen-bonding network. A plausible explanation would be that this segment is in a dynamic conformational equilibrium between the helical conformation and a locally unfolded conformation (Englander & Kallenbach, 1984). As the amides in this helix are irreversibly exchanged after unfolding, the local "breathing" of the recognition helix must occur on a relative slow time scale, slower than the intrinsic exchange rates of amide groups of unstructured peptides which are roughly on the order of minutes under our conditions at pH 6.5. As shown in Figure 1F, the $^1H-^{15}N$ heteronuclear NOE data did not indicate any fast motions on a time scale of nanoseconds in this helix.

These results suggest that both regions of the E2 protein that interact with DNA in the protein−DNA complex are mobile when the protein is free in solution. Flexibility of DNA-contacting regions and the modulation of this flexibility upon DNA complex formation have been also reported for several other DNA-binding proteins. In GCN4, which is a prototype of the leucine-zipper motif, the DNA-binding basic domain is largely unstructured in the absence of DNA but becomes α-helical upon binding DNA (Weiss *et al.*, 1990). A recent study (Gryk *et al.*, 1995) indicated that the helix−turn−helix region of the uncomplexed *trp* repressor is mobile on a time scale of seconds. Our previous data on the amide exchange of the *ets* domain of Fli-1 (Liang *et al.*, 1994) suggested that its recognition helix is also flexible in solution in the absence of DNA.

It is still not clear what roles the flexibility of the DNA-contacting regions plays in specific DNA recognition. It is possible that the flexibility may facilitate any adjustment of the binding surface of the protein to achieve optimal interaction with DNA. Energetically, mobile recognition helices and DNA-interacting regions in the uncomplexed form would be expected to be unfavorable for entropic considerations. However, this mobility and the reduction of the mobility upon DNA complex formation may contribute to the large negative heat capacity change (Ladbury *et al.*, 1994), which is a hallmark of specific protein-DNA recognition (Ha *et al.*, 1989).

ACKNOWLEDGMENT

We thank V. Giranda for sharing data and helpful discussions, M. Sattler and E. Olejniczak for help with NMR experiments, and M. Nilges for making available computer files for the structure calculation of a dimer using ambiguous restraints.

SUPPORTING INFORMATION AVAILABLE

One table containing the ^1H, ^{13}C, and ^{15}N chemical shifts of the E2-DBD from HPV-31 (8 pages). Ordering information is given on any current masthead page.

REFERENCES

Archer, S. J., Ikura, M., Torchia, D. A., & Bax, A. (1991) *J. Magn. Reson, 95*, 636−641.
Arrowsmith, C. H., Pachter, R., Altman, R. B., Iyer, S. B., & Jardetzky, O. (1990) *Biochemistry 29*, 6332−6341.
Bax, A., & Ikura, M. (1991) *J. Biomolec. NMR 1*, 99−104.
Bax, A., & Pochapsky, S. (1992) *J. Magn. Reson. 99*, 638−640.
Bax, A., & Grzesiek, S. (1993) *Acc. Chem. Res. 26*, 131−137.
Bodenhausen, G., & Ruben, D. J. (1980) *J. Chem. Phys. Lett. 69*, 185−189.
Brooks, B., Bruccoleri, R., Olafson, B., States, D., Swaminathan, S., & Karplus, M. (1983) *J. Comput. Chem. 4*, 187−217.
Brünger, A. T. (1992) *X-PLOR 3.1 Manual*, Yale University Press, New Haven, CT.
Burley, S. K., & Petsko, G. A. (1985) *Science 229*, 23−28.
Corina, K., Grossman, S. R., Barsoum, J., Prakash, S. S., Androphy, E. J., & Pepinsky, R. B. (1993) *Virology 197*, 391−396.
Doeberitz, M. v. K. (1992) *Eur. J. Med. 1*, 415−423.
Englander, S. W., & Kallenbach, N. (1984) *Q. Rev. Biophys. 16*, 521−655.
Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Forman-Kay, J. D., & Kay, L. E. (1994) *Biochemistry 33*, 5984−6003.
Fesik, S. W., & Zuiderweg, E. R. P. (1988) *J. Magn. Reson. 87*, 588−593.
Folkers, P. J. M., Folmer, R. H. A., Konings, R. N. H., & Hilbers, C. W. (1993) *J. Am. Chem. Soc. 115*, 3798−3799.
Frenkiel, T., Bauer, C., Carr, M. D., Birdsall, B., & Feeney, J. (1990) *J. Magn. Reson. 90*, 420−425.
Gemmecker, G., Olejniczak, E. T., & Fesik, S. W. (1992) *J. Magn. Reson. 96*, 199−.
Goldsborough, M. D., Disilvestre, D., Temple, G. F., & Loriincz, A. T. (1989) *Virology 171*, 306−311.
Griesinger, C., Otting, G., Wüthrich, K., & Ernst, R. R. (1988) *J. Am. Chem. Soc. 110*, 7870−7872.
Gryk, M. R., Finucane, M. D., Zheng, Z., & Jardetzky, O. (1995) *J. Mol. Biol. 246*, 618−627.
Grzesiek, S., & Bax, A. (1992) *J. Am. Chem. Soc. 114*, 6291−6293.

Ha, J.-H., Spolar, R. S., & Record, M. T., Jr. (1989) *J. Mol. Biol. 209*, 801−816.
Haugen, T. H., Cripe, T. P., Ginder, G. D., Karin, M., & Turek, L. P. (1987) *EMBO J. 6*, 145−152.
Hegde, R. S., Grossman, S. R., Laimonis, L. A., & Sigler, P. B. (1992) *Nature 359*, 505−512.
Hirel, P.-H., Schmitter, J.-M., Dessen, P., Fayat, G., & Blanquet, S. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 8247−8251.
Hyberts, S. G., Goldberg, M. S., Havel, T. F., & Wagner, G. (1992) *Protein Sci. 1*, 736−751.
Ikura, M., Bax, A., Clore, G. M., & Gronenborn, A. M. (1990) *J. Am. Chem. Soc. 112*, 9020−9022.
Kay, L. E., Ikura, M., Tschudin, R., & Bax, A. (1990) *J. Magn. Reson. 89*, 496−514.
Kay, L. E., Keifer, P., & Saarinen, T. (1992) *J. Am. Chem. Soc. 114*, 10663−10665.
Kay, L. E., Xu, G.-Y., Singer, A. U., Muhandiram, D. R., & Forman-Kay, J. D. (1993) *J. Magn. Reson. 86*, 110−126.
Kraulis, P. J. (1991) *J. Appl. Crystallogr. 24*, 946−950.
Ladbury, J. E., Wright, J. G., Sturtevant, J. M., & Sigler, P. B. (1994) *J. Mol. Biol. 238*, 669−681.
Liang, H., Olejniczak, E. T., Mao, X., Nettesheim, D. G., Yu, L., Thompson, C. B., & Fesik, S. W. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 11655−11659.
Marion, D., Driscoll, P. C., Kay, L. E., Wingfield, P. T., Bax, A., Gronenborn, A. M., & Clore, G. M. (1989a) *Biochemistry 28*, 6150−6156.
Marion, D., Ikura, M., Tschudin, R., & Bax, A. (1989b) *J. Magn. Reson. 85*, 393−399.
Marion, D., Kay, L. E., Sparks, S. W., Torchia, D. A., & Bax, A. (1989c) *J. Am. Chem. Soc. 111*, 1515−1517.
Mohr, I. J., Clark, R., Sun, S., Androphy, E. J., MacPherson, P., & Botchan, M. R. (1990) *Science 250*, 1694−1699.
Morris, A. L., MacArthur, M. W., Hutchinson, E. G., & Thornton, J. M. (1992) *Proteins 12*, 345−364.
Neri, D., Szyperski, T., Otting, G., Senn, H., & Wüthrich, K. (1989) *Biochemistry 28*, 7510−7516.
Nilges, M. (1993) *Proteins 17*, 297−309.
Nilges, M., Clore, G. M., & Gronenborn, A. M. (1988) *FEBS Lett. 239*, 317−324.
Olejniczak, E. T., & Eaton, H. L. (1990) *J. Magn. Reson. 87*, 628−632.
Orth, G., & zur Hausen, H. (1994) in *Encyclopedia of Virology* (Webster, R. G., & Granoff, A., Eds.) pp 1013−1026, Academic Press, London.
Pepinsky, R. B., Androphy, E. J., Corina, K., Brown, R., & Barsoum, J. (1994) *DNA Cell Biol. 13*, 1011−1019.
Prakash, S. S., Grossman, S. R., Pepinsky, R. B., Laimins, L. A., & Androphy, E. J. (1992) *Genes Dev. 6*, 105−116.
Richart, R. M., & Wright, T. C., Jr. (1992) *Curr. Opin. Obstet. Gyn. 4*, 663−669.
Rucker, S. P., & Shaka, A. J. (1989) *Mol. Phys. 68*, 509−517.
Singh, J., & Thornton, J. M. (1985) *FEBS Lett. 191*, 1−6.
Spalholz, B. A., Yang, Y.-C., & Howley, P. M. (1985) *Cell 42*, 183−191.
Studier, F. W., Rosenberg, A. H., Dunn, J. J., & Dubendorff, J. W. (1990) *Methods Enzymol. 185*, 60−89.
Ustav, M., & Stenlund, A. (1991) *EMBO J. 10*, 449−457.
Vuister, G. W., & Bax, A. (1993) *J. Am. Chem. Soc. 115*, 7772−7777.
Weiss, M. A. (1990) *J. Magn. Reson. 86*, 626−632.
Weiss, M. A., Ellenberger, T., Wobbe, C. R., Lee, J. P., Harrison, S. C., & Struhl, K. (1990) *Nature 347*, 575−578.

BI951932W