Increased Stability and DNA Site Discrimination of "Single Chain" Variants of the Dimeric β-Barrel DNA Binding Domain of the Human Papillomavirus E2 Transcriptional Regulator†

Mariano Dellarole,‡ Ignacio E. Sánchez,‡ Eleonora Freire, and Gonzalo de Prat-Gay*

Instituto Leloir and IIBBA-Conicet, Patricias Argentinas 435 (1405), Buenos Aires, Argentina

Received June 5, 2007; Revised Manuscript Received August 14, 2007

ABSTRACT: Human papillomavirus infects millions of people worldwide and is a causal agent of cervical cancer in women. The HPV E2 protein controls the expression of all viral genes through binding of its dimeric C-terminal domain (E2C) to its target DNA site. We engineered monomeric versions of the HPV16 E2C, in order to probe the link of the dimeric β-barrel fold to stability, dimerization, and DNA binding. Two single-chain variants, with 6 and 12 residue linkers (scE2C-6 and scE2C-12), were purified and characterized. Spectroscopy and crystallography show that the native structure is unperturbed in scE2C-12. The single chain variants are stabilized with respect to E2C, with effective concentrations of 0.6 to 6 mM. The early folding events of the E2C dimer and scE2C-12 are very similar and include formation of a compact species in the submillisecond time scale and a non-native monomeric intermediate with a half-life of 25 ms. However, monomerization changes the unfolding mechanism of the linked species from two-state to three-state, with a high-energy intermediate. Binding to the specific target site is up to 5-fold tighter in the single chain variants. Nonspecific DNA binding is up to 7-fold weaker in the single chain variants, leading to an overall 10-fold increased site discrimination capacity, the largest described so far for linked DNA binding domains. Titration calorimetric binding analysis, however, shows almost identical behavior for dimer and single-chain species, suggesting very subtle changes behind the increased specificity. Global analysis of the mechanisms probed suggests that the dynamics of the E2C domain, rather than the structure, are responsible for the differential properties. Thus, the plastic and dimeric nature of the domain did not evolve for a maximum affinity, specificity, and stability of the quaternary structure, likely because of regulatory reasons and for roles other than DNA binding played by partly folded dimeric or monomeric conformers.

Papillomaviruses are small double stranded DNA viruses that are capable of infecting epithelia and cause a variety of lesions that go from innocuous to malignant, where the most widespread threat is cervical cancer (*1*). The virus contains only 8 open reading frames (*2*), and thus it is extremely effective in leading a productive life cycle in infected epithelia. The E2 protein was first identified as a transcriptional activator, but at present it is known to act as repressor depending of the yet unknown features in the virus life cycle, linked to the epithelial cell differentiation (*3*). In addition, several other essential functions in DNA replication, episome migration in mitosis, apoptosis, and replication origin binding, establish it as a multifunctional protein (*4*−*6*). Its architecture is composed of an N-terminal transactivation domain and a C-terminal dimerization and DNA binding domain (E2C[1]), separated by a nonconserved and flexible domain termed "hinge" (*7*).

One of the essential reactions in the biological activity of E2 is the primary DNA site recognition by the E2C domain. The binding of E2C to its target DNA takes place thanks to its unusual dimeric β-barrel topology, only shared with the Epstein−Barr EBNA1 origin binding protein (*8*), that positions two successive DNA binding helices along the major grooves of a pseudo-palindromic site (*9*). We have been investigating the folding and DNA binding mechanisms of E2C from HPV16 as a model for DNA recognition by proteins, besides its key role in the virus life cycle. The HPV16 strain accounts for around 60% of the cervical cancers linked to HPV. The structural topology of HPV16 E2C is overall similar to other E2C domains (*5*, *10*, *11*). Studies on the HPV31 E2C suggested a rather dynamic domain (*12*). The dimer unfolds and dissociates in a concerted manner (*13*), and a well-defined compact intermediate of non-native nature was found on its kinetic pathway (*14*, *15*). There is local persistent residual structure in the urea denatured state, and there are two parallel routes for folding/association, one at high and one at low protein concentration, respectively (*16*). Equilibrium DNA binding

† This work was supported by Wellcome Trust Grant GR077355AYA, and PICT 2000 01-08959 from ANPCyT. M.D. holds a doctoral fellowship from Consejo Nacional de Investigaciones Científicas y Técnicas of Argentina (CONICET), I.E.S. is the recipient of a MUTIS postdoctoral fellowship from the A.E.C.I. (Spanish Ministry for Foreign Affairs and Cooperation), and G.d.P.G. is a Career Investigator from CONICET.

* Corresponding author.

‡ Contributed equally to this work.

[1] Abbreviations: scE2C-6, scE2C-9, scE2C-12, monomeric variants of the HPV16 E2C domain with 6-, 9-, and 12-residue linkers; GdmCl, guanidinium chloride.

10.1021/bi701104q CCC: $37.00
Published on Web 10/04/2007

of HPV16 to a specific site showed higher affinity and specificity than the domain form BPV1, a prototypic bovine strain, even for the BPV site (*17*). Kinetics of DNA binding proceed through two parallel routes, one, a single-step encounter, and the other, a multistep pathway (*18*). Systematic mutagenesis of the DNA binding helix showed that the contributions of protein side chains are additive (*19*). No "hot-spots" were found, uncovering an essential difference with model protein−protein interfaces (*19*). More recently, NMR analysis of the HPV16 E2C−DNA complex and mutagenesis studies uncovered a previously undescribed contribution of Lys 349, out of the consensus DNA binding site, to the affinity of binding (*20*).

E2C binding to its target DNA takes place in its native, dimeric state. Thus, stability and function of this domain are coupled, as in many transcription factors (*21*). Several groups have engineered monomerized variants of dimeric transcription factors in order to decouple stability and DNA binding (*22*−*27*). In most cases, the stability at physiological protein concentrations was improved, while retaining wild type binding affinity (*22*−*27*). We set out to engineer monomeric or "single chain" versions of the HPV16 E2C in order to address the folding and DNA binding mechanisms of this model domain. We determined the crystal structure of one of the variants and analyzed its stability and folding mechanism (*11*). Binding to specific DNA is up to 5-fold stronger in the single chain proteins, while the binding to nonspecific DNA is weaker. This translates to an increased DNA sequence discrimination capacity. Overall, these results show the E2C as a robust and unique domain with improved properties, and suggest its possible use as an engineering scaffold.

MATERIALS AND METHODS

Construction, Expression, and Purification of Single Chain E2C. The starting point for the construction was the previously described E2C ptzU18-based expression vector (*13*). In this plasmid, the open reading frame of E2C is flanked at the 5′ end by a restriction site for *Eco*RI, a ribosome binding site and a restriction site for *Cla*I at the ATG codon, and at the 3′ end by a restriction site for *Bam*HI downstream of the Stop codon. We constructed a monomeric version of E2C by inserting a sequence coding for E2C, followed in tandem by a sequence coding for a linker at the restriction site for *Cla*I. The inserted sequence thus precedes the original ORF for E2C. We synthesized the insert by a PCR reaction from the *Eco*RI to the *Bam*HI restriction sites of ptzU18-E2C with the following primers: N-term AGGGAATTCA AAGAGGAGAA ATTACATATG ACACCCATAG TA, C-term scE2C-6 CATCGATCCA CCGGTACCGC CTATAGACAT AAATCCAGT, C-term E2Csc-9 CATCGATCCG CCTGAGCCAC CGGTACCGCC TATAGACATA AATCCAGT, C-term scE2C-12 CATCGATCCG CCACCTGAAC CGCCACCGGT ACCGCCTCCT ATAGACATAA ATCCAGT. The PCR product yields an E2C ORF without the stop codon and with a linker sequence and a restriction site for *Cla*I at the 3′ end. At the 5′ end we mutated the original *Cla*I restriction site to an *Nde*I restriction site. The PCR product was then cloned in tandem with the original gene in the same T7 expression vector ptzU18, between the *Eco*RI and *Cla*I sites, generating the E2C-linker-E2C gene. Finally, the ligation product was sequenced, stored, and transformed into the Bl21(DE3) plys *E. coli* expression strain.

Wild-type HPV16 E2C and its single chain versions scE2C-6 and scE2C-12 were recombinantly expressed, purified, and stored as previously described for E2C (*17*). Briefly, the proteins were overexpressed in *Escherichia coli* as soluble folded dimers. All constructions had a very high affinity for the Heparin Hyper D column, even in the presence of 0.6 M NaCl at pH 8.0. Protein eluted from this column is >90% pure. Purity after size exclusion chromatography is >95%. Protein concentration was determined spectrophotometrically using a molar extinction coefficient $\epsilon_{280} = 41920\ M^{-1}\ cm^{-1}$.

Circular Dichroism. Samples containing 10 μM of E2C proteins were extensively dialyzed against 50 mM Sodium phosphate buffer, pH 7.0, 1 mM DTT at 298 K and far UV CD spectra recorded in a Jasco (Japan) J810 equipment in a 0.1 cm path length cuvette. Ten scans were accumulated and averaged for each measurement.

Crystallization, X-ray Data Collection, and Structure Determination. The E2Csc-12 protein was crystallized by the vapor diffusion method using hanging drops against a reservoir solution at room temperature. A mixture of 1 μL of protein stock solution (6 mg/mL) and 1 μL of precipitating agent (80 mM sodium citrate (pH 5.21), 1.53 M $(NH_4)_2SO_4$, and 150 mM sodium and potassium tartrate) was equilibrated against 150 μL of the same precipitating agent. Hexagonal crystals grew in a few weeks. X-ray diffraction data were collected at the X9A beamline at the National Synchrotron Light Source (Brookhaven National Laboratory, New York) during the RapiData 2005 course using a MarCCD detector. Crystals were cooled to 100 K, and the resolution was 1.85 Å. Data indexing, integration, scaling, and reduction were performed with the programs MOSFLIM, Scala, and Truncate from de CCP4 package (*28*). A total of 5% of the measured reflections were flagged for cross-validation. The structure was solved by molecular replacement method with the program AMoRe (*29*) using the dimeric model of E2C HPV16 (PDB entry 1BY9) (*11*). Due to the minimal differences between the search model and the unknown structure, phasing was straightforward. Model building was carried out with O (*30*), and the structure was refined with CNS (*31*). The $2F_o - F_c$ and $F_o - F_c$ maps built showed a strong and well-defined electronic density. Model building included two additional residues in the β3 sheet, A327, and A328 and few side chain shifts. Refinement was carried out using the maximum likelihood algorithm and cycles of simulated annealing. During the final stages of refinement, solvent molecules and a sulfate anion were introduced. Details of data collection parameters and processing statistics are shown in Table 1.

Equilibrium Denaturation. Equilibrium denaturation experiments were performed in MES (2-(*N*-morpholino)-ethanesulfonic acid) 100 mM, DTT 1 mM, pH 6.1 or sodium phosphate buffer 50 mM pH 7.0, 1 mM DTT, at 298 K as described (*13*). The variables monitored were the molar ellipticity at 222 nm and the center of mass of the fluorescence emission spectrum after excitation at 280 nm. The data for E2C were fitted to a two-state coupled unfolding/dissociation model $N_2 \leftrightarrow 2D$ (*13*), and the data for a monomeric variant to a two-state model $N \leftrightarrow U$ (*32*) using the ProFit software (Quantumsoft, Zurich). In all cases

Table 1: X-ray Data Collection Parameters and Processing Statistics for scE2C-12

Data Collection	
wavelength (Å)	0.9791
number of frames	90
oscillation step (deg)	1
crystal-to-detector distance (mm)	95.1
Indexing And Scaling	
cell parameters (Å)	$a = b = 43.17$, $c = 74.92$
cell angles (deg)	$\alpha = 90$, $\beta = 90$, $\gamma = 120$
space group	$P3_121$
resolution limit (Å)	1.80
total reflections (unique)	40056 (7833)
multiplicity[a]	5.1 (5.2)
I/σ	25.5 (6.8)
R_{merge} (%)[b]	4.5 (18.5)
completeness (%)[b]	99.3 (100)
B-factor (Wilson plot, Å^2)	23.9
Refinement	
cutoff	$F > 2\sigma$
dimers/asymmetric unit	1/2
number of protein atoms	1208
number of water molecules	80
number of ion atoms	10
R-factor[c]	0.2266
free *R*-factor [d]	0.2407
rmsd bond lengths (Å)	0.005226
rmsd bond angles (deg)	1.07807
B-factor protein (average, Å^2)	26.27
B-factor waters (average, Å^2)	38.05
B-factor anion (average, Å^2)	63.106

[a] Values in parentheses correspond to the highest resolution shell: 1.90−1.80 Å. [b] $R_{merge} = \sum_i |I_i - \langle I \rangle| / \sum_i |I_i|$. [c] $R = \sum_{hkl} |F_o - F_c| / \sum_i |F_o|$. [d] R_{free} is the same as *R*, but for 5% of the data excluded from the refinement.

the free energy for unfolding ΔG_{eq} was assumed to vary linearly with denaturant concentration, with a proportionality constant *m* (*13*, *32*). The denaturant concentration at which half of the protein molecules are unfolded, $[D]_{50\%}$, was calculated as described (*13*, *32*). All fits were carried out using the ProFit software (Quantumsoft, Zurich).

Folding Kinetics. Refolding and unfolding kinetics were recorded in an Applied Photophysics SX-18MV spectrophotometer. The sample was excited at 280 nm (4 nm slit), and fluorescence was recorded using a 320 nm cutoff filter. In the refolding experiments one volume of 10 μM unfolded protein in HCl 10 mM, pH 2.0 was mixed with one volume of refolding buffer, to final conditions of MES 100 mM, DTT 1 mM, pH 6.1. In the unfolding experiments, one volume of 55 μM native protein in MES 100 mM, DTT 1 mM, pH 6.1 was mixed with ten volumes of the same buffer with varying urea concentrations.

The unfolding data for E2C were fitted to a simple two-state model for unfolding, N → U (*14*). The unfolding data for scE2C-12 were fitted to a three-state model for unfolding with an obligatory, high energy intermediate N ↔ I → U (*33*). All activation free energies were assumed to vary linearly with denaturant concentration, with a proportionality constant *m* (*33*). Since the intermediate is not populated, only the ratio of the rate constants for the steps leading from the intermediate, k_{IN}/k_{IU}, can be obtained from the data. For practical reasons we did not fit the ratio directly but kept k_{IN} and its denaturant dependence fixed at a value that will not make this step rate-limiting under any conditions ($k_{IN} = 10^5$ s^{-1}, $m_{IN} = 0$ (kcal/mol·M)) (*33*).

DNA Binding Followed by Fluorescence. Equilibrium binding of E2C protein and the monomeric variants to a specific E2 HPV16 DNA site (GTAACCGAAATCGGTTGA) and a nonspecific randomized binding site with the same nucleotide composition (ACATGGACCTGTCAAGTA) was carried out by fluorescence spectroscopy in a Aminco Bowman series 2 luminescence spectrometer assembled in "L" geometry (*17*). The buffer conditions were 150 to 300 mM sodium phosphate, pH 7.0, 1 mM DTT, at 298 K. 5 to 20 nM solutions of the fluorescein labeled DNA oligonucleotide were titrated with increasing amounts of the corresponding protein. Samples were equilibrated for 2 minutes before each measurement. In all cases, maximal dilution was 20%, and the data were corrected accordingly. The excitation wavelength was 495 nm (4 nm slit), and emission was recorded at 520 nm. The data were fitted to a quadratic binding equation (*17*):

$$[P:L] = 0.5\Delta F([P] + [L] + K_D) - \sqrt{([P] + [L] + K_D)^2 - 4[P][L]} \quad (1)$$

where P is the E2C variant, L is the DNA, and P:L is the protein−DNA complex; ΔF is the difference in the signal between the E2C−DNA complex and free DNA; and K_D is the dissociation constant for the interaction.

Isothermal Titration Calorimetry. Experiments were performed using the VP-ITC (MicroCal). In each titration 10 μM E2C protein, in the cell, was titrated with several injections of 100 μM ligand in sodium phosphate 200 mM pH 7.0, 1 mM DTT, at 298 K. The volume of each injection was of 5 μL, except the first injection, which was 2 μL. Injections were continued beyond saturation levels to allow for determination of heats of ligand dilution. The resulting data were fitted to a single-site binding isotherm, after subtracting the heat of dilution, using the ORIGIN 5.0 software supplied with the calorimeter.

RESULTS

Construction and Expression of "Single-Chain" Variants of E2C. Three single-chain variants of HPV16 scE2C were constructed, expressed, and purified as described in Materials and Methods. We engineered linkers of 6, 9, and 12 residues, respectively. The length of the linker was constrained by the need to join approximately 17 Å, the distance between the N- and C-termini of two different subunits in the E2C dimer (*10*, *20*) (Figure 1C). The glycine-rich linker sequences (GGGTGGGSGGGS, SGGSGGTGG, and GGTGGS) were chosen in an attempt to maintain maximum flexibility and reasonable solubility, on the basis of previous studies on monomerized transcription factors (*22*−*27*). While the scE2C-9 variant expressed only at very low levels in *E. coli*, the scE2C-6 and scE2C-12 variants showed good expression levels and could be obtained in mg amounts to >95% purity (see Materials and Methods).

The Structures and Conformations of E2C and Its Monomeric Variants Are Indistinguishable. We first checked the structural integrity of the single chain variants by spectroscopic methods. The fluorescence spectra of both variants are indistinguishable from the dimeric HPV16 E2C domain, not only in the native form but also in the unfolded state in 6.0 M GdmCl, anticipating that the environment of the tryptophan residues in both states are similar (Figure 1A).

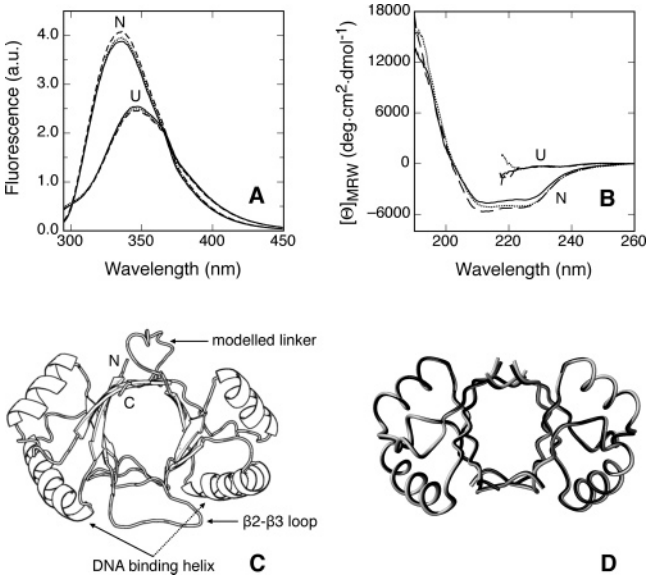

FIGURE 1: Structural characterization of monomeric variants of E2C. A and B: Spectra of native and unfolded (6 M GdmCl) E2C (solid line), scE2C-6 (dashed line), and scE2C-12 (dotted line). All proteins were at 10 μM in sodium phosphate buffer 50 mM pH 7, 1 mM DTT. (A) Fluorescence emission spectra. (B) Far-UV CD spectra. (C) Putative structure of scE2c-12 modeled from the solution HPV16 E2C structure (pdb entry 1R8P) with MODELLER 9v1 software (*57*). The figure was prepared using PyMol (Delano Scientific LLC). (D) Superposition of the polypeptide backbones of E2C (pdb entry 1BY9) (white) and scE2C-12 (pdb entry 2Q79) (black). The figure was prepared using MOLMOL 2.0 (*58*).

Far UV CD analysis shows that the overall shapes of the spectra, including the position of the minima at 212 and 226 nm, are similar (Figure 1B). The small differences observed could correspond to the contribution of the additional linker residues or result from the intrinsic flexibility of the domain (*34*).

The structures of E2Csc-6, E2Csc-12, and E2C are very similar, according to spectroscopic evidence. However, small changes in the structure of a protein may not be evidenced by circular dichroism or fluorescence spectroscopy. In order to check for changes in atomic detail, we crystallized E2Csc-12 and determined its structure by X-ray diffraction (Table 1, Protein Data Bank ID 2Q79). The structure revealed two additional residues in the β3 sheet, A327 and A328. The electron density of the linker is visible at low σ (0.4). Despite this, the density between A283 and D362 is not clear enough for the linker to be completely traced. Both dimeric E2Csc-12 and E2C superimpose very well with a C^{α} root-mean-square deviation (rmsd) of 0.27 Å (Figure 1D), indicating that monomerization of E2C by a 12-residue linker does not change the structure of the protein.

Increased Stability of Monomeric E2C Variants. The stability of the engineered variants was evaluated by chemical denaturation using GdmCl and urea, followed by fluorescence and CD. We carried out the experiments in 50 mM sodium phosphate, pH 7.0 (Figure 2A) and 100 mM MES, pH 6.1 (Figure 2B), in order to match the conditions used in the DNA binding and folding studies in this and previous work (*14*, *15*). All E2C variants showed cooperative transitions, in which secondary and tertiary structure are lost concomitantly (Figure 2). This is in agreement with previous studies on E2C, including the observed stabilization effect by phosphate (*13*). All the data could be fitted confidently

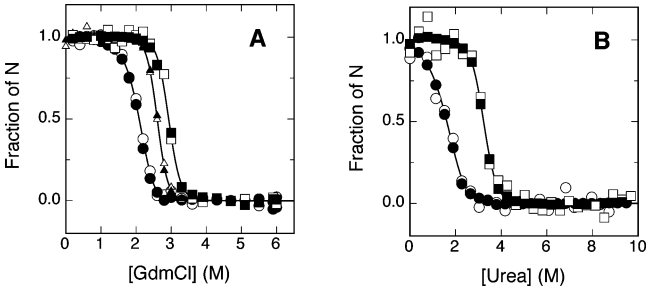

FIGURE 2: Equilibrium unfolding of E2C (circles), scE2C-12 (squares), and scE2C-6 (triangles) followed by fluorescence (filled symbols) and CD (empty symbols). (A) Experiments in sodium phosphate pH 7.0 50 mM, 1 mM DTT at 298 K. (B) Experiments in MES pH 6.1, 100 mM, 1 mM DTT at 298 K. Protein concentrations were at 10 μM in all cases.

to two-state folding (monomeric variants) or folding/dissociation (dimer) models, without populated intermediates (see Materials and Methods). The parameters obtained are shown in Table 2. The results for the single chain variants do not depend on protein concentration, as expected (data not shown). The *m*-values for the momomerized variants are very similar to those for E2C, indicating that the change in solvent accessible surface area upon unfolding remains roughly the same upon linking the dimer subunits covalently (*35*).

Both scE2C-6 and scE2C-12 unfold at higher denaturant concentrations than E2C, indicating that they are more stable than the dimeric parent protein at the micromolar monomer concentrations used in these experiments (Figure 2 and Table 2). scE2C-12 is more stable than scE2C-6. The free-energy change (ΔG) values obtained from the fit cannot be compared directly because the equilibrium for the dimer depends on protein concentration while those for the monomers do not (*36*, *37*). The appropriate way to compare the stability of monomeric and dimeric variants ruled by equilibria of different molecularity is the use of the effective concentration C_{eff} (*36*, *37*). C_{eff} is the protein concentration at which dimer

$$C_{\text{eff}}(\text{M}) = \frac{K_{\text{eq}}^{\text{monomer}}}{K_{\text{eq}}^{\text{dimer}}} \qquad (2)$$

and monomer have the same free energy for unfolding (*36*, *37*). The theory predicts that a dimer will be less stable than a monomer at concentrations lower than C_{eff}, and vice versa (*36*, *37*). The effective concentrations of scE2C-6 and scE2C-12 are shown in Table 2, and indicate that the monomeric variants are more stable than the parent dimer at concentrations lower than 0.5−5 mM.

The Early Folding Events of scE2C-12 and E2C Are Very Similar. Folding of dimeric E2C from its monomeric, acid-denatured state is a multistep process. Upon dilution into refolding buffer, the protein folds into a monomeric intermediate (I_M) in about 25 ms (*14*, *15*). Formation of the intermediate appears as an overshoot in the fluorescence kinetic trace (*14*, *15*). I_M folds into the native state through two parallel reactions, one of which takes place through a dimeric intermediate (*14*). In this work, we investigated whether the monomeric intermediate observed in the refolding of E2C is also populated in the folding mechanism of scE2C-12.

Table 2: Equilibrium Denaturation of E2C and Its Monomeric Variants

buffer	protein	ΔG (kcal•mol^{-1})	m (kcal•mol^{-1}•M^{-1})	$[D]_{50\%}$ (M)	C_{eff}[a] (mM)
sodium phosphate 50 mM, 1 mM DTT, pH 7.0	E2C	-14.80 ± 0.04	4.00 ± 0.01	1.99 ± 0.01	
	scE2C-12	-10.40 ± 0.77	3.55 ± 0.26	2.93 ± 0.30	5.9
	scE2C-6	-10.30 ± 0.81	3.96 ± 0.30	2.60 ± 0.28	0.63
MES 100 mM, 1 mM DTT, pH 6.1	E2C	-10.43 ± 0.01	2.03 ± 0.01	1.41 ± 0.01	
	scE2C-12	-6.79 ± 0.83	2.10 ± 0.25	3.23 ± 0.55	1.5

[a] At 3 M denaturant. $C_{eff} = K_{eq}^{monomer}/K_{eq}^{dimer}$.

We carried out stopped-flow refolding kinetics of acid-unfolded scE2C-12 followed by fluorescence (see Materials and Methods). The observed rate constants for this variant do not depend on protein concentration, as expected for a unimolecular process (data not shown). Figure 3A shows that there is a large change in the fluorescence of scE2C-12 in the dead time of the instrument, approximately 2 ms. Such fast changes in fluorescence are often observed in the refolding of small proteins or domains (*38*). We have compared the dead-time changes in the fluorescence of scE2C-12 and dimeric E2C as a function of urea concentration (Figure 3B). The curves for the two proteins superimpose well, suggesting that formation of an equivalent collapsed species takes place within the first 2 ms of refolding of E2C and scE2C-12.

The inset in Figure 3A shows the observable changes in scE2C-12 fluorescence during refolding. The data are well described by the sum of three exponential functions, with rate constants of 37 ± 1, 3.6 ± 0.2, and 0.22 ± 0.01 s^{-1}, respectively. The rate constant for the fastest process is very similar to the rate constant for formation of the monomeric intermediate during refolding of E2C, 31 s^{-1} (*15*). The overshoot in fluorescence intensity in the first 100 ms is also similar to the one observed during refolding of E2C (*15*). Moreover, the urea dependence of the natural logarithm of the rate constant is -0.78 ± 0.02 M^{-1}, in excellent agreement with that for the formation of I_M in E2C (-0.86 ± 0.01 M^{-1}) (Figure 3C). From these data, we propose that a species equivalent to the monomeric intermediate observed during refolding of E2C is also populated during refolding of scE2C-12, which is monomeric but a duplicate of the polypeptide chain. We conclude from the data shown in Figure 3 that monomerization of E2C by a flexible linker does not alter substantially the early folding steps.

Comparative Analysis of the Unfolding Mechanism of E2C and Single-Chain Variants. Kinetic unfolding of E2C by addition of urea is a simple two-state process that follows single-exponential kinetics (*14*). There are no changes in the fluorescence signal in the dead time of the instrument (data not shown), and the natural logarithm of the observed rate constant depends linearly on urea concentration (Figure 4A). Unfolding of scE2C-12 under the same conditions also follows single-exponential kinetics (data not shown). The complete amplitude of the fluorescence change between the native and unfolded states is resolved in the stopped-flow measurements (Figure 4B), indicating the absence of rapid processes taking place in the dead time of the instrument. The urea dependence of the natural logarithm of the rate constant for unfolding of scE2C-12 shows a sharp change between two limiting slopes at high and low urea concentration (Figure 4A). This is indicative of a denaturant-induced change of rate-limiting step on a linear pathway with an obligatory high-energy intermediate (*33*). Consequently, the data were well described by the following three-state model for unfolding (Figure 4A).

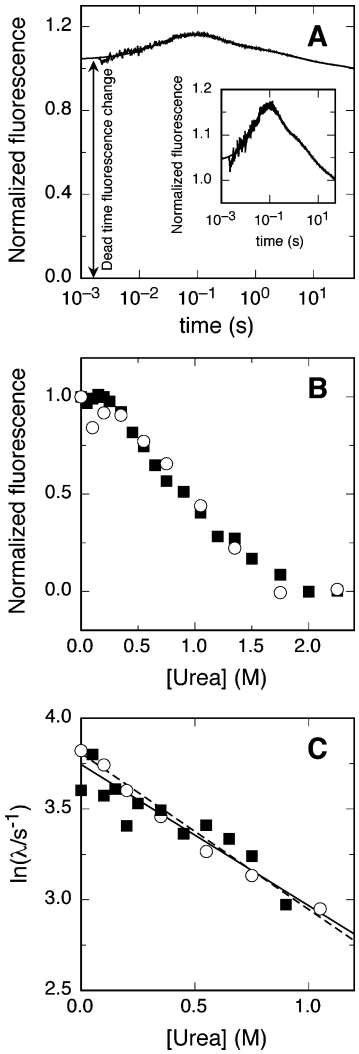

FIGURE 3: Refolding kinetics of scE2C-12 in 100 mM MES, pH 6.1 at 298 K, monitored by fluorescence. (A) Kinetic trace at 0 M urea. The data are normalized so that the fluorescence signal for the native state is one and the fluorescence signal for the unfolded state at pH 2 is zero. The line is a fit of the data to a sum of three exponentials. The inset highlights the measurable changes in fluorescence. (B) Initial fluorescence signal of the refolding kinetics of E2C (○) and scE2C-12 (■). The data are normalized by setting the fluorescence signal for the unfolded state (extrapolated from high urea concentrations) to zero and the value at 0 M urea to one. (C). Urea dependence of the natural logarithm of the rate constant for formation of the non-native intermediate of E2C (○, dashed line) and scE2C-12 (■, continuous line). The lines are linear fits to the data.

$$N \leftrightarrow I \rightarrow U \quad (3)$$

The observed change in the mechanism for unfolding may

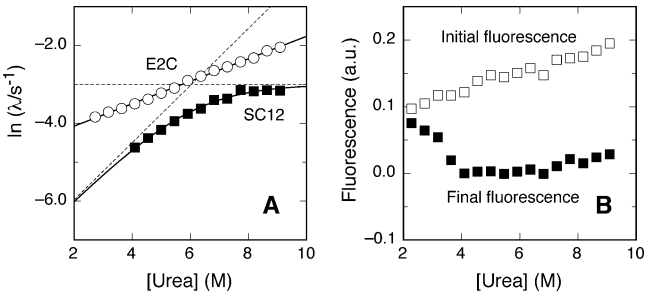

FIGURE 4: Unfolding kinetics of scE2C-12 in 100 mM MES, pH 6.1 at 298 K, monitored by fluorescence. (A) Dependence of the observed rate constant for unfolding of E2C (○) and scE2C-12 (■) on urea concentration. The continuous lines are fits to a two-state model for E2C and to a linear three-state model with a high-energy intermediate for scE2C-12. The dashed lines correspond to the rate constants for the individual transition states in the three-state model. (B) Dependence of the initial (□) and final (■) fluorescence signals for unfolding of scE2C-12 on urea concentration.

be a switch to an alternative unfolding pathway or the observation of a previously hidden transition state along the same pathway. We have investigated whether one of the two transition states in the unfolding mechanism of scE2C-12 corresponds to the transition state for unfolding of E2C by examining the corresponding microscopic rate constants and their urea dependencies (Figure 4A, dashed lines). Clearly, none of the two transition states in the unfolding mechanism of scE2C-12 correspond to the transition state ensemble for unfolding rate of E2C. Thus, E2C and scE2C-12 unfold by different mechanisms.

Single-Chain Variants Display Increased DNA Binding Affinity and Target Discrimination. We have measured the binding of E2C and its two monomeric variants for a specific E2C target DNA site using fluorescence spectroscopy (*17*) in sodium phosphate buffer, pH 7.0, 1 mM DTT at 298 K (see Materials and Methods). Titration of 100 nM fluorescein-labeled DNA with scE2C-12 and E2C showed identical 1:1 binding stoichiometries (Figure 5A, inset). The same result was obtained for scE2C-6 (data not shown). We determined the dissociation constants, K_D, for the protein−DNA complexes with experiments at lower concentrations of DNA and protein (Figure 5A and Table 3). At 200 mM sodium phosphate, the affinity of scE2C-12 for DNA is 0.6 kcal/mol greater than for both E2C and scE2C-6 (Table 3). We confirmed the greater affinity of scE2C-12 relative to E2C using an electrophoresis mobility shift assay (see Supporting Information).

We have measured the enthalpic contribution to the free energy of binding ΔG by isothermal titration calorimetry (Figure 5B, inset). This allows us to dissect ΔG into enthalpic and entropic contributions (Table 3). Binding is enthalpy-driven in all three cases, and the enthalpy term ΔH is very similar for all three variants, as the superposition of the titrations shows in Figure 5B. The increased affinity of scE2C-12 comes from a larger ΔH, opposed but not counterbalanced by a smaller increase in the entropic term $-T\Delta S$ (Table 3). Interestingly, although the affinities of E2C and scE2C-6 for DNA are the same, both ΔH and $-T\Delta S$ have smaller values in scE2C-6 than in E2C (Table 3). This change in the underlying thermodynamics of the interaction suggests that scE2C-6 and E2C bind DNA in a different manner.

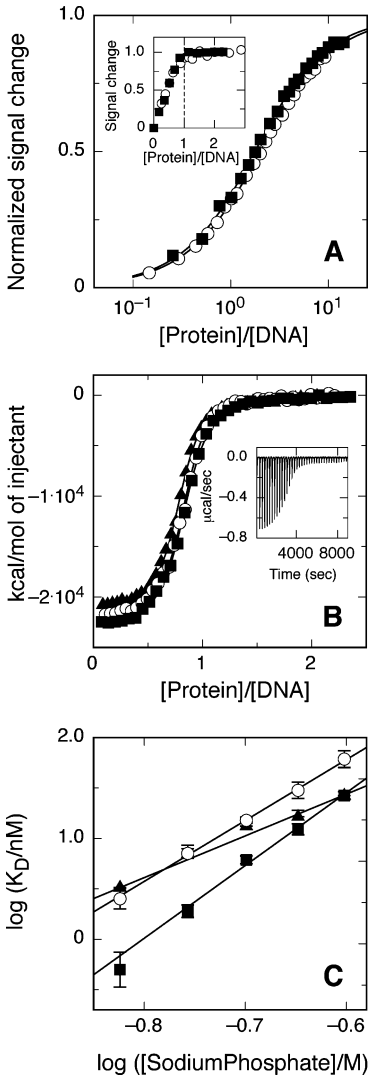

FIGURE 5: DNA binding of E2C (○), scE2C-12 (■), or scE2C-6 (▲). (A) Determination of dissociation constants (K_D). Binding experiments were performed by adding protein to a fixed amount of oligonucleotide at dissociating conditions and followed by the emission intensity of the fluorescein probe. Inset: Stoichiometry of protein−DNA interaction measured at 100 nM of DNA. (B) ITC binding isotherms resulting from integration of the specific heats with respect to time, with the appropriate molar correction. Inset: Raw ITC data for the titration with scE2C-12. (C) Dependence of K_D on sodium phosphate concentration from 150 to 300 mM. The lines are linear fits to the data for each variant.

We have further characterized the interaction of the E2C variants with DNA by measuring their dependence on the concentration of sodium phosphate (*39*). The results confirm that the affinity of scE2C-12 for DNA is higher than that of E2C (Figure 5C). For all variants, the logarithm of the dissociation constant increases linearly with the logarithm of phosphate concentration. This is indicative of net cation release from the DNA backbone upon binding (*39*). The slopes for E2C and scE2C-12 are identical within the experimental error (Table 3), indicating that the energetic contribution of cation release to the free energy of binding is the same for the two proteins (*39*). The slope for scE2C-6 is clearly smaller than for the other two variants (Table 3). This abrupt change in the energetic contribution of cation release to binding suggests an altered mode of binding for scE2C-6 relative to E2C and scE2C-12, in agreement with

Table 3: Thermodynamics of DNA Binding and Sequence Discrimination by E2C and Its Monomeric Variants

protein	ΔG^a (kcal·mol^{-1})	ΔH^b (kcal·mol^{-1})	$-T\Delta S^c$ (kcal·mol^{-1})	$(\partial \log K_D)/(\partial \log [\text{phosphate}])$	$K_D^{\text{specific}\ d}$ (nM)	$K_D^{\text{nonspecific}\ d}$ (μM)	specificity^e
E2C	-10.6 ± 0.1	-21.9 ± 0.1	11.3 ± 0.1	-7.1 ± 0.4	2.5 ± 0.3	8 ± 1	3200 ± 600
scE2C-12	-11.2 ± 0.1	-23.2 ± 0.1	12.0 ± 0.1	-6.0 ± 0.5	0.5 ± 0.2	15 ± 2	30000 ± 13000
scE2C-6	-10.7 ± 0.1	-20.9 ± 0.2	10.1 ± 0.2	-4.1 ± 0.1	3.3 ± 0.1	59 ± 11	18000 ± 3400

[a] From fluorescence measurements at 200 mM sodium phosphate, 298 K. [b] From isothermal titration calorimetry at 200 mM sodium phosphate, 298 K. [c] Calculated from columns 2 and 3: $\Delta G = \Delta H - T\Delta S$. [d] From fluorescence measurements at 150 mM sodium phosphate, 298 K. [e] Specificity ratio calculated as $K_D^{\text{nonspecific}}/K_D^{\text{specific}}$ (*17*).

the thermodynamic measurements. Our experiments show that monomerization increased the affinity of E2C for DNA modestly: 2- to 5-fold in the case of the 12-residue linker and up to 2-fold in the case of the 6-residue linker. However, in the case of scE2C-6 the thermodynamics of the interaction changed significantly.

Increased Sequence Discrimination Capacity in Linked E2C Species. In vivo binding of transcription factors to DNA requires not only a high affinity of the protein for its target site but also a high affinity relative to nonspecific sequences, in order to discriminate among the large number of sites in the genome (*40*), that is, a high specificity of binding. The specificity of binding or discrimination capacity can be defined as the ratio of the dissociation constants for nonspecific and specific DNA (*17*):

$$\text{specificity} = \frac{K_D^{\text{nonspecific}}}{K_D^{\text{specific}}} \quad (4)$$

We have measured the affinity of binding of the monomerized variants of E2C to a randomized target site as a model for their ability to bind nonspecific DNA (*17*) (see Materials and Methods). Remarkably, the binding affinity of the monomeric variants for nonspecific DNA is lower than that of E2C, with a 2-fold increase in K_D in the case of scE2C-12 and 7-fold in the case of scE2C-6 (Table 3). This, together with the changes in the affinity for the target DNA, leads to an overall increase in binding specificity of 10- and 6-fold for scE2C-12 and scE2C-6, respectively (Table 3).

DISCUSSION

Spectroscopic evidence suggested that monomerization of dimeric transcription factors using flexible linkers does not lead to substantial changes in the structure (*22*−*27*). Our results confirm that flexible linkers of appropriate length and composition have little, if any, influence on the average native state structure of the protein or domain (Figure 1). On the other hand, E2C is known to be highly dynamic in solution as seen in molecular dynamic simulations (*34*), NMR spectroscopy (*12*) (Cicero et al., in preparation), and binding to DNA (*5*, *20*). Thus, highly sensitive methods such as circular dichroism and fluorescence spectroscopy, as well as the X-ray structure, indicate absolute conservation of the fold even though neither the linker nor the β2-β3 loop shows electron density, the latter being expected from other dimer crystal structures, suggesting flexibility in both regions which are curiously positioned at opposed faces of the barrel (Figure 1D).

Both single-chain variants are stabilized relative to dimeric E2C at physiological nano- to micromolar concentrations (*36*, *37*, *41*), with effective concentrations ranging from 0.6 to 6 mM (Figure 2, Table 2). These numbers fall within the lower limit for previous work on linked dimeric transcription factors, in which effective concentrations ranged from 2.4 to 240 mM (Table 4) (*22*−*27*). scE2C-6 has a smaller effective concentration than scE2C-12, suggesting that the shorter linker of scE2C-6 induces a strained conformation in the protein. Complementary fragments of the monomeric proteins CI-2 and barnase have been generated by chemical cleavage (reviewed in ref (*42*)). We will consider the construction of a linked variant of E2C as the reverse experiment. The effective concentrations reached for the noncovalent complexes of CI2 and barnase are 8 mM and 8 M, respectively. These large differences can be ascribed to the fact that CI2 is a slow-release protease inhibitor that evolved to remain bound to the protease after cleavage of the target bond at the inhibitory loop. In addition, barnase is modular and presents an intermediate, and the isolated fragments have stable native-like or native-compatible residual structure, while in the case of the CI2 fragments, one is unstructured and the other presents non-native collapsed hydrophobic local patches, and must overcome a larger energetic barrier for attaining the native structure. This is reflected in the nearly 4-fold slower association rates of CI2 compared to barnase noncovalent complexes (*42*).

In a similar way, the association rate of the E2C dimer is slower than that of the model dimeric transcription factor arc (*15*, *26*). We ascribe this to the fact that E2C shows a detectable, stable, cooperative, and non-native monomeric intermediate (*14*) while arc has an inferred dimeric intermediate (*43*), and no structure has been described in its unfolded state ensemble, while persistent residual structure was determined in E2C unfolded state ensemble (*16*). Remarkably, linking two polypeptides of E2C does not affect the formation of the "monomeric intermediate", strongly supporting a non-native structure in the E2C monomer, and

Table 4: Effect of Monomerization on the Stability and DNA Binding Properties of Dimeric Transcription Factors

protein	linker length	$C_{\text{effective}}$ (mM)^a	affinity (monomer)/affinity (dimer)^b	specificity (monomer)/specificity (dimer)^c	ref
E2C	6	0.6	0.8−2.3^d	5.6	this work
	12	1.5−5.9^e	2.3−5.0^d	9.1	this work
Arc	15	2.4	8.4, 14^f	1.9, 3.2^f	(*25*, *26*)
Cro	8, 16	5.3			(*22*, *56*)
Gene V	5, 6	150	1.0, 0.8		(*24*)
GCN4	6	240			(*23*)
MASH-1	8, 17		10, 14	2.5, 2.3	(*27*)

[a] $C_{\text{eff}} = K_{\text{eq}}^{\text{monomer}}/K_{\text{eq}}^{\text{dimer}}$. [b] Affinities for specific DNA. [c] Specificity ratio calculated as $K_D^{\text{nonspecific}}/K_D^{\text{specific}}$ (*17*). [d] From 150 to 300 mM sodium phosphate. [e] In MES and sodium phosphate buffer. [f] For two different DNA half-sites.

indicating that the formation of the intermediate species does not involve intermonomer interactions even in the case of a single-chain species, which would be entropically favored for an early dimerization. The isolated E2C monomer has not been so far observable by dilution methods, because of the limitations on the detection under such low protein concentrations. Combined fragmentation and mutagenesis approaches should be required to attempt this goal.

Theory predicts that, below the C_{eff}, the dimeric form is less stable than the covalently linked species (*36*, *37*). However, the E2C dimer displays a loss in concentration dependence of the stability toward chemical denaturation above 10 μM, and this holds for the association rate as well, where the rate-limiting step switches to a first-order reaction (*14*, *16*). In the folding pathway this means a weak dimeric intermediate preceding the main folding event, through a different channel than the monomeric intermediate. At the equilibrium, we believe it is due to global inter- and intramonomer interactions predominating over those stabilizing the interface at lower protein concentrations. A similar effect is observed for the noncovalent CI2 complex (*44*), and for the thermal denaturation of the arc dimer (*45*). Thus, the linear increase of stability with the logarithm of concentration in a second-order process appears to hold experimentally for small compounds or for dimers built up from two natively structured monomers. In the case of the other known dimeric β-barrel domain, the DNA binding domain of the Epstein−Barr origin binding protein EBNA1, we have observed that there is no concentration dependence for chemical denaturation at protein concentrations as low as 0.2 μM (*46*).

On the other hand, monomerization of E2C induces a drastic change in the mechanism of unfolding (Figure 4). This is at odds with the results for monomeric versions of the Arc and GCN4 transcription factors, which retain the two-state mechanism of their dimeric counterparts (*23*, *26*) and with studies on complementary fragments of chymotrypsin inhibitor 2 and barnase, which show identical folding/unfolding mechanisms and very similar transition state structures to the uncleaved species (*42*, *47*). ScE2C-12 unfolds via a linear pathway with a high-energy on-pathway intermediate. Similar dimeric intermediates have been identified in the folding pathways of the dimeric transcription factors Trp repressor and factor for inversion stimulation (*48*, *49*), suggesting that dimeric intermediates are common in the free energy landscapes of protein homodimers. The denaturant dependence of the unfolding reaction from the native to the intermediate state is close to zero (Figure 4A), indicating that there is no net change in the interactions of the protein with the solvent in the first step of unfolding (*33*). This is reminiscent of the subtle quaternary structure rearrangements that take place in several E2C domains upon DNA binding (*5*, *20*). We speculate that the unfolding intermediate observed for scE2C-12 is also present in the energy landscape of dimeric E2C and that it might be related to regulation of DNA binding (*50*), and in the route for amyloid formation that we recently described for this domain (*51*).

The affinity of scE2C-12 for binding to its target DNA increases upon monomerization evenly along the tested phosphate concentration range, while the affinity for nonspecific DNA decreases to a larger extent, resulting in a 10-fold increase in specificity (Figure 5, Table 3). Specific binding of E2C to DNA induces changes in the stability and dynamics of the protein (*52*), and nonspecific binding is known to induce larger changes, at least in one protein (*53*). We interpret that monomerization of E2C restricts the structural and dynamic changes that take place upon nonspecific DNA binding but favors the changes that take place upon specific DNA binding. The effects on nonspecific binding are larger for the scE2C-6 variant (Table 3), suggesting that its shorter linker leads to a more drastic reduction in the adaptability of the domain. On the other hand, this reduced adaptability is detrimental for specific binding. The end result is a 6-fold increase in specificity against the 10-fold increase of scE2C-12 (Table 3 and 4). In fact, the scE2C-6 variant displays a significantly smaller salt dependence than the dimer and the unstrained scE2C-12 (Figure 5C), which suggests a decrease in the electrostatic component of the interaction, as counterions are released from both macromolecules upon binding (*39*, *54*). This would correspond to a local structural or dynamic rearrangement of the protein DNA binding by influence of the strain introduced by the 6-residue as opposed to 12-residue linker. We plan to test this hypothesis using nuclear magnetic resonance.

The improved target discrimination of the DNA by scE2C agrees with other monomerized transcription factors (*24*, *25*, *27*), but is the largest increase in specificity for engineered single-chain DNA binding domains described to date (Table 4). Discrimination between specific and nonspecific sequences is crucial for *in vivo* binding of proteins to DNA (*40*), but transcriptional regulators, in particular the E2 example we now describe, appear not to evolve to affinity or specificity limits for functional and regulatory reasons. In addition, E2 was shown to have other functions where an equilibrium between dimers and monomer would be required. By linking the domain, the dissociation into monomers is eliminated, and in the case of the BPV1 E2 linked heterodimer recently described, it appears not to affect its role in transcription activation or DNA replication but does affect episomal migration (*55*). The biological activity of the HPV16 E2 could be tested in the context of an engineered viral genome.

On speculative grounds, one can argue that the conserved dimeric β-barrel topology of E2C has not evolved for maximum stability, binding affinity, or sequence discrimination for regulatory reasons. In addition, a monomer must exist, depending on concentration, but most likely formed by yet unknown cell environment conditions. This monomer will be non-native and unlikely to bind DNA with affinity comparable to the symmetric "double" site in the dimer. We propose that it participates in other functions of the protein that involve oligomerization, such as DNA linking, recruiting of cellular transcription machinery or in replication, together with E1. Future experiments will aim at the characterization of the solution structure and dynamics of the scE2C-12 variant in its free form and in the complex with its target site.

ACKNOWLEDGMENT

We would like to thank Daniel Cicero for helpful discussions and for preparing Figure 1D, Gastón París for crystal-

lography data collection, and Thierry Rose for help with the Modeller software.

SUPPORTING INFORMATION AVAILABLE

Figure depicting DNA binding of E2C and scE2C-12 followed by electrophoresis mobility shift assay. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Bosch, F. X., Sanjosé, S., Castellsagué, X., Moreno, V., and Muñoz, N. (2006) Epimediology of Human Papillomavirus Infections and Associations with Cervical Cancer: New Opportunities for Prevention, in *Papillomavirus Research: From Natural History to Vaccines and Beyond* (Saveria Campo, M., Ed.) pp 19−40, Caisreir Academic Press, Wymondham.
2. Howley, P. M. (1996) *Papillomaviridae: The viruses and their replication* (Fields, B. N., Knipe, D. M., Howley, P. M., Eds.) pp 2045−2076, Lippincott-Raven Press, Philadelphia.
3. Kalantari, M., and Bernard, H.-U. (2006) Gene Expression of Papillomaviruses, in *Papillomavirus Research: From Natural History to Vaccines and Beyond* (Saveria Campo, M., Ed.) pp 41−52, Caisreir Academic Press, Wymondham.
4. Chow, L. T., and Broker, T. R. (2006) Mechanisms and Regulation of Papillomavirus DNA Replication, in *Papillomavirus Research: From Natural History to Vaccines and Beyond* (Saveria Campo, M., Ed.) pp 53−71, Caisreir Academic Press, Wymondham.
5. Hegde, R. S. (2002) The papillomavirus E2 proteins: structure, function, and biology, *Annu. Rev. Biophys. Biomol. Struct. 31*, 343−360.
6. Dell, G., and Gaston, K. (2001) Human papillomaviruses and their role in cervical cancer, *Cell. Mol. Life Sci. 58*, 1923−1942.
7. Giri, I., and Yaniv, M. (1988) Structural and mutational analysis of E2 trans-activating proteins of papillomaviruses reveals three distinct functional domains, *EMBO J. 7*, 2823−2829.
8. Cruickshank, J., Shire, K., Davidson, A. R., Edwards, A. M., and Frappier, L. (2000) Two domains of the Epstein-Barr virus origin DNA-binding protein, EBNA1, orchestrate sequence-specific DNA binding, *J. Biol. Chem. 275*, 22273−22277.
9. Hegde, R. S., Grossman, S. R., Laimins, L. A., and Sigler, P. B. (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.
10. Nadra, A. D., Eliseo, T., Mok, Y. K., Almeida, C. L., Bycroft, M., Paci, M., Prat-Gay, G. d., and Cicero, D. O. (2004) Solution structure of the HPV-16 E2 DNA binding domain, a transcriptional regulator with a dimeric beta-barrel fold, *J. Biomol. NMR 30*, 211−214.
11. Hegde, R. S., and Androphy, E. J. (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.
12. Liang, H., Petros, A. M., Meadows, R. P., Yoon, H. S., Egan, D. A., Walter, K., Holzman, T. F., Robins, T., and Fesik, S. W. (1996) Solution structure of the DNA-binding domain of a human papillomavirus E2 protein: evidence for flexible DNA-binding regions, *Biochemistry 35*, 2095−2103.
13. Mok, Y. K., Prat-Gay, G. d., Butler, P. J., and Bycroft, M. (1996) Equilibrium dissociation and unfolding of the dimeric human papillomavirus strain-16 E2 DNA-binding domain, *Protein Sci. 5*, 310−319.
14. de Prat-Gay, G., Nadra, A. D., Corrales-Izquierdo, F. J., Alonso, L. G., Ferreiro, D. U., and Mok, Y. K. (2005) The folding mechanism of a dimeric beta-barrel domain, *J. Mol. Biol. 351*, 672−682.
15. Mok, Y. K., Bycroft, M., and Prat-Gay, G. d. (1996) 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.
16. Mok, Y. K., Alonso, L. G., Lima, L. M., Bycroft, M., and de Prat-Gay, G. (2000) Folding of a dimeric beta-barrel: residual structure in the urea denatured state of the human papillomavirus E2 DNA binding domain, *Protein Sci. 9*, 799−811.
17. Ferreiro, D. U., Lima, L. M., Nadra, A. D., Alonso, L. G., Goldbaum, F. A., and Prat-Gay, G. d. (2000) Distinctive cognate sequence discrimination, bound DNA conformation, and binding modes in the E2 C-terminal domains from prototype human and bovine papillomaviruses, *Biochemistry 39*, 14692−14701.
18. Ferreiro, D. U., and Prat-Gay, G. d. (2003) A protein-DNA binding mechanism proceeds through multi-state or two-state parallel pathways, *J. Mol. Biol. 331*, 89−99.
19. Ferreiro, D. U., Dellarole, M., Nadra, A. D., and de Prat-Gay, G. (2005) Free energy contributions to direct readout of a DNA sequence, *J. Biol. Chem. 280*, 32480−32484.
20. Cicero, D. O., Nadra, A. D., Eliseo, T., Dellarole, M., Paci, M., and Prat-Gay, G. d. (2006) Structural and thermodynamic basis for the enhanced transcriptional control by the human papillomavirus strain-16 E2 protein, *Biochemistry 45*, 6551−6560.
21. Beckett, D. (2001) Regulated assembly of transcription factors and control of transcription initiation, *J. Mol. Biol. 314*, 335−352.
22. Jana, R., Hazbun, T. R., Fields, J. D., and Mossing, M. C. (1998) Single-chain lambda Cro repressors confirm high intrinsic dimer-DNA affinity, *Biochemistry 37*, 6446−6455.
23. Moran, L. B., Schneider, J. P., Kentsis, A., Reddy, G. A., and Sosnick, T. R. (1999) Transition state heterogeneity in GCN4 coiled coil folding studied by using multisite mutations and crosslinking, *Proc. Natl. Acad. Sci. U.S.A. 96*, 10699−10704.
24. Liang, H., Sandberg, W. S., and Terwilliger, T. C. (1993) Genetic fusion of subunits of a dimeric protein substantially enhances its stability and rate of folding, *Proc. Natl. Acad. Sci. U.S.A. 90*, 7010−7014.
25. Robinson, C. R., and Sauer, R. T. (1996) Covalent attachment of Arc repressor subunits by a peptide linker enhances affinity for operator DNA, *Biochemistry 35*, 109−116.
26. Robinson, C. R., and Sauer, R. T. (1996) Equilibrium stability and sub-millisecond refolding of a designed single-chain Arc repressor, *Biochemistry 35*, 13878−13884.
27. Sieber, M., and Allemann, R. K. (1998) Single chain dimers of MASH-1 bind DNA with enhanced affinity, *Nucleic Acids Res. 26*, 1408−1413.
28. Walton, R. I., Millange, F., Loiseau, T., O'Hare, D., and Ferey, G. (2000) Crystallization of a Large-Pore Three-Dimensional Gallium Fluorophosphate under Mild Conditions, *Angew. Chem., Int. Ed. 39*, 4552−4555.
29. Navaza, J. (2001) Implementation of molecular replacement in AMoRe, *Acta Crystallogr., Sect. D: Biol. Crystallogr. 57*, 1367−1372.
30. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models, *Acta Crystallogr. A 47* (Part 2), 110−119.
31. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination, *Acta Crystallogr., Sect. D: Biol. Crystallogr. 54*, 905−921.
32. Santoro, M. M., and Bolen, D. W. (1988) Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl alpha-chymotrypsin using different denaturants, *Biochemistry 27*, 8063−8068.
33. Sanchez, I. E., and Kiefhaber, T. (2003) Evidence for sequential barriers and obligatory intermediates in apparent two-state protein folding, *J. Mol. Biol. 325*, 367−376.
34. Falconi, M., Santolamazza, A., Eliseo, T., de Prat-Gay, G., Cicero, D. O., and Desideri, A. (2007) Molecular dynamics of the DNA-binding domain of the papillomavirus E2 transcriptional regulator uncover differential properties for DNA target accommodation, *FEBS J. 274*, 2385−2395.
35. Myers, J. K., Pace, C. N., and Scholtz, J. M. (1995) Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding, *Protein Sci. 4*, 2138−2148.
36. Zhou, H. X. (2001) Single-chain versus dimeric protein folding: thermodynamic and kinetic consequences of covalent linkage, *J. Am. Chem. Soc. 123*, 6730−6731.
37. Zhou, H. X. (2004) Loops, linkages, rings, catenanes, cages, and crowders: entropy-based strategies for stabilizing proteins, *Acc. Chem. Res. 37*, 123−130.

38. Bachmann, A., Segel, D., and Kiefhaber, T. (2002) Test for cooperativity in the early kinetic intermediate in lysozyme folding, *Biophys. Chem. 96*, 141−151.
39. Record, M. T., Jr., Ha, J. H., and Fisher, M. A. (1991) Analysis of equilibrium and kinetic measurements to determine thermodynamic origins of stability and specificity and mechanism of formation of site-specific complexes between proteins and helical DNA, *Methods Enzymol. 208*, 291−343.
40. von Hippel, P., and Berg, O. (1986) On the specificity of DNA-protein interactions, *Proc. Natl. Acad. Sci. U.S.A. 83*, 1608−1612.
41. Ghaemmaghami, S., Huh, W. K., Bower, K., Howson, R. W., Belle, A., Dephoure, N., O'Shea, E. K., and Weissman, J. S. (2003) Global analysis of protein expression in yeast, *Nature 425*, 737−741.
42. Prat-Gay, G. d. (1996) Association of complementary fragments and the elucidation of protein folding pathways, *Protein Eng. 9*, 843−847.
43. Jonsson, T., Waldburger, C. D., and Sauer, R. T. (1996) Nonlinear free energy relationships in Arc repressor unfolding imply the existence of unstable, native-like folding intermediates, *Biochemistry 35*, 4795−4802.
44. Mohana-Borges, R., Silva, J. L., Ruiz-Sanz, J., and Prat-Gay, G. d. (1999) Folding of a pressure-denatured model protein, *Proc. Natl. Acad. Sci. U.S.A. 96*, 7888−7893.
45. Robinson, C. R., Rentzeperis, D., Silva, J. L., and Sauer, R. T. (1997) Formation of a denatured dimer limits the thermal stability of Arc repressor, *J. Mol. Biol. 273*, 692−700.
46. Freire, E., Oddo, C., Frappier, L., and de Prat-Gay, G. (2007) Kinetically driven refolding of the hyperstable EBNA1 origin DNA-binding dimeric beta-barrel domain into amyloid-like spherical oligomers, *Proteins*, submitted.
47. Ruiz-Sanz, J., Prat-Gay, G. d., Otzen, D. E., and Fersht, A. R. (1995) Protein fragments as models for events in protein folding pathways: protein engineering analysis of the association of two complementary fragments of the barley chymotrypsin inhibitor 2 (CI-2), *Biochemistry 34*, 1695−1701.
48. Topping, T. B., Hoch, D. A., and Gloss, L. M. (2004) Folding mechanism of FIS, the intertwined, dimeric factor for inversion stimulation, *J. Mol. Biol. 335*, 1065−1081.
49. Gloss, L. M., Simler, B. R., and Matthews, C. R. (2001) Rough energy landscapes in protein folding: dimeric E. coli Trp repressor folds through three parallel channels, *J. Mol. Biol. 312*, 1121−1134.
50. Lee, B. C., Pandit, A., Croonquist, P. A., and Hoff, W. D. (2001) Folding and signaling share the same pathway in a photoreceptor, *Proc. Natl. Acad. Sci. U.S.A. 98*, 9062−9067.
51. Wetzler, D. E., Castano, E. M., and de Prat-Gay, G. (2007) A quasi-spontaneous amyloid route in a DNA binding gene regulatory domain: The papillomavirus HPV16 E2 protein, *Protein Sci. 16*, 744−754.
52. Lima, L. M., and Prat-Gay, G. d. (1997) Conformational changes and stabilization induced by ligand binding in the DNA-binding domain of the E2 protein from human papillomavirus, *J. Biol. Chem. 272*, 19295−19303.
53. Kalodimos, C. G., Biris, N., Bonvin, A. M., Levandoski, M. M., Guennuegues, M., Boelens, R., and Kaptein, R. (2004) Structure and flexibility adaptation in nonspecific and specific protein-DNA complexes, *Science 305*, 386−389.
54. Ladbury, J. E., and Williams, M. A. (2004) The extended interface: measuring non-local effects in biomolecular interactions, *Curr. Opin. Struct. Biol. 14*, 562−569.
55. Kurg, R., Tekkel, H., Abroi, A., and Ustav, M. (2006) Characterization of the functional activities of the bovine papillomavirus type 1 E2 protein single-chain heterodimers, *J. Virol. 80*, 11218−11225.
56. Zhou, H. X. (2001) The affinity-enhancing roles of flexible linkers in two-domain DNA-binding proteins, *Biochemistry 40*, 15069−15073.
57. Fiser, A., Do, R. K., and Sali, A. (2000) Modeling of loops in protein structures, *Protein Sci. 9*, 1753−1773.
58. Koradi, R., Billeter, M., and Wuthrich, K. (1996) MOLMOL: a program for display and analysis of macromolecular structures, *J. Mol. Graphics 14*, 51-55, 29−32.

BI701104Q