DNA Binding and Bending by the Human Papillomavirus Type 16 E2 Protein

RECOGNITION OF AN EXTENDED BINDING SITE*

(Received for publication, August 6, 1996, and in revised form, December 10, 1996)

Alison Thain‡, Kenneth Webster, Dave Emery§, Anthony R. Clarke, and Kevin Gaston¶
From the Department of Biochemistry, Centre for Molecular Recognition, School of Medical Sciences, University of Bristol, Bristol BS8 1TD, United Kingdom

The human papillomavirus (HPV) 16 E2 protein (hE2) binds to four sites present upstream of the P97 promoter and regulates transcription of the viral E6 and E7 oncogenes. We have determined the relative binding constants for the interaction of the full-length hE2 protein with these sites. Our results show that hE2 binds tightly to site 4, less tightly to sites 1 and 2, and weakly to site 3. Similar results have previously been obtained using a C-terminal fragment of the hE2 protein suggesting that the C-terminal domain is the sole determinant of DNA binding affinity and specificity. Using circular permutation assays we show that binding of the hE2 protein induces the formation of a significant DNA bend and that the hE2-induced DNA bend angle is the same at both tight and weak hE2-binding sites. An alignment of the four hE2-binding sites from the HPV 16 genome suggests that this protein recognizes an extended binding site when compared with the bovine papillomavirus E2 protein. Here we show that the hE2 protein binds tightly to sites containing an A:T or a G:C base pair at position 7 of its binding site but weakly to sites containing either C:G or T:A at this position. Using site-directed mutagenesis we demonstrate that an arginine at position 304 of the hE2 protein is responsible for the recognition of specific base pairs at this position.

The recognition of specific DNA sequences by transcription factors is often the first step in the regulation of gene expression. An understanding of how these proteins recognize their target sequences, and the effect that this has on DNA conformation, is central to the question of how genes are controlled. We are studying the human papillomavirus E2 protein, a sequence-specific DNA-binding protein involved in the regulation of viral gene expression and DNA replication. Papillomaviruses infect epithelial cells and induce the formation of benign hyperproliferative lesions or warts. Over 70 distinct types of human papillomavirus (HPV)³ have been described. Some of these viral types produce lesions that have the potential to undergo malignant transformation. HPV 16 and HPV 18, for example, are thought to play a primary role in the development of cervical cancer (for a review, see Ref. 1). The products of the viral E6 and E7 genes form complexes with the cellular tumor suppressor proteins p53 and Rb, respectively. These interactions bring about a change in cell growth rate and promote cell immortalization (reviewed in Ref. 2). In HPV 16, transcription of the E6 and E7 genes is under the control of a single promoter (P97) that lies immediately upstream of the E6 gene (3) (9). The activity of the P97 promoter is regulated by a variety of cellular transcription factors and by the viral E2 protein (4–6).

Much early work concentrated on the bovine papillomavirus (BPV) E2 protein. The BPV E2 protein (bE2) binds as a dimer to 12 inverted repeats (consensus sequence 5′-ACCGN₄CGGT-3′) present upstream of the BPV early genes and activates transcription (7, 8). Binding of bE2 protein to DNA is cooperative (9, 10), and once bound, bE2 dimers can associate further to form DNA loops (11). The bE2 protein contains an N-terminal transcription activation domain and a C-terminal DNA-binding/dimerization domain separated by a flexible hinge (12). The structure of the bE2 C-terminal domain bound to a specific DNA sequence has been determined by x-ray crystallography (13). The two subunits of the bE2 dimer each form half of a β-barrel over which the DNA is bent, allowing the interaction of two α-helices with the exposed edges of the base pairs in two successive major grooves of the DNA.

The HPV 16 E2 protein (hE2) and the BPV E2 protein share a high degree of sequence similarity at the protein level. The amino acids involved in the recognition of specific DNA sequences, for example, are conserved between bE2 and hE2; the single exception being a bE2 phenylalanine (Phe-343) that is replaced by a tyrosine in the hE2 protein (Tyr-303). The hE2 protein binds to four sites present upstream of the HPV 16 P97 promoter (14). The precise function of the hE2 protein in the regulation of this promoter is the subject of some controversy. Over-expression of the hE2 protein has been shown to repress the activity of some P97 promoter fragments linked to reporter genes (15, 16). However, in similar experiments over-expression of hE2 was found to activate transcription from other promoter derivatives (4, 5, 17, 18). This confusion over the role of the hE2 protein is probably brought about by differences in the contribution of each of the four hE2-binding sites toward P97 promoter activity. The organization of the P97 promoter is shown diagrammatically in Fig. 1. Two hE2-binding sites (separated by only 3 base pairs) are located immediately upstream of the P97 TATA box and are flanked on the 5′ side by a binding site for the cellular transcription factor Sp1. The binding of hE2 (or bE2) to these sites has been shown to block the binding of Sp1 and TBP (the TATA box binding factor) and to repress the promoter activity of DNA fragments carrying these sequences

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.
‡ Supported by SmithKline Beecham and a Biotechnology and Biological Sciences Research Council Ph.D. studentship. Present address: Program in Molecular Medicine, University of Massachusetts Medical Center, Worcester, MA 01605.
§ Present address: National Institute for Medical Research, Mill Hill, London, NW7 1AA United Kingdom.
¶ To whom correspondence should be addressed. Tel.: 0117-9287439; Fax: 0117-9288274.
³ The abbreviations used are: HPV, human papillomavirus; BPV, bovine papillomavirus; PCR, polymerase chain reaction; GST, glutathione S-transferase.
P1

\[ \text{DNA Binding and Bending by the HPV 16 E2 Protein} \]

**Figure 1.** The HPV 16 P97 promoter region. The top line shows the HPV 16 P97 promoter and positions of the HPV 16 early genes: E6, E7, E1, E2, and E4. The bottom line shows an expanded view of the P97 promoter. The transcription start point is indicated by the bent arrow. The E2-binding sites are represented by the open boxes and labeled 1–4. The Sp1-binding site and TATA box sequence are represented by S and T, respectively. Not to scale.

(16, 19, 20). The two remaining hE2-binding sites are located around 150 and 550 base pairs upstream of the P97 transcription start point. The binding of hE2 to sites in this “enhancer configuration” has been shown to activate transcription from artificial promoter constructs (21, 17).

To understand the role of the hE2 protein in the regulation of the P97 promoter it is important to know the relative affinities of the hE2-binding sites present within this promoter and to characterize fully the DNA binding activity of this protein. To this end we have purified the full-length hE2 as a GST fusion protein and determined the relative binding constants of the four hE2-binding sites present in the HPV 16 genome. Using gel retardation assays we show that the DNA bend angle induced by the binding of hE2 is the same at both tight- and weak-binding sites. We also demonstrate that unlike hE2, the hE2 protein binds preferentially to an extended recognition sequence that contains a purine at the −7 position and we use site-directed mutagenesis to show that this specific interaction is mediated by arginine 304 in the hE2 DNA-binding domain.

**EXPERIMENTAL PROCEDURES**

**Plasmids Used in This Study**—The HPV 16 E2 gene was amplified by PCR from HPV 16 DNA using the oligonucleotide primers: 5′-GACTGGAGACTCATGGAGAGCTTCTTGGCAACGG-3′ and 5′-GACTGAATTCATCATATAGACATAAATCCAG-3′. These primers placed BamHI and EcoRI restriction sites at the 5′ and 3′ ends of the hE2 coding sequence, respectively. The PCR product was cloned between the BamHI and EcoRI restriction sites of the prokaryotic expression vector pGEX-2T (Pharmacia), creating an in-frame GST-hE2 fusion (pGEX-hE2). The R304A mutation of hE2 was made by PCR. The primers 5′-GCTTCTTTTATTGATATCTTTTAAAACGGCAGCCACGG-3′ and 5′-GGCTGGCAAGCCACGTTTGG-3′ were used to amplify the 5′ end of the hE2 gene. In a separate reaction, the primers 5′-GCTTCTTTTATTGATATCTTTTAAAACGGCAGCCACGG-3′ and 5′-GACTGAATTCATCATATAGACATAAATCCAG-3′ were used to amplify the 3′ end of the hE2 gene. The underlined bases mismatch the hE2 sequence and introduce the R304A mutation. The 5′ and 3′ end PCR products were purified from agarose gels, mixed, and re-amplified using the pGEX-specific primers. This generated a full-length hE2 gene containing the R304A mutation. This PCR product was digested with BamHI and EcoRI, and cloned into pGEX-2T as described above. The resulting pGEX-hE2 and pGEX-R304A plasmids were sequenced using a panel of E2-specific sequencing primers to check for the occurrence of any point mutations generated by the Taq polymerase.

**Protein Purification**—Escherichia coli XL-1 blue cells containing either pGEX-hE2 or pGEX-hE2 R304A were grown at 37 °C in Terrific broth supplemented with 0.01% ampicillin to an OD600 of 0.5. Fusion protein expression was then induced by the addition of 0.15 mM isopropyl-1-thio-p-D-galactopyranoside and the cells were grown at 28 °C overnight. Bacteria were harvested by centrifugation, resuspended in phosphate-buffered saline, and sonicated at 4 °C. The cell lysate was cleared by centrifugation (15,000 × g for 30 min at 4 °C) then incubated with 0.1% DNase I for 30 min at 4 °C. After re-centrifugation, the supernatant was loaded onto a glutathione-Sepharose 4B column pre-equilibrated in 50 mM Tris, pH 8.0, 500 mM NaCl, and then eluted in buffer 1 (50 mM trithionolamine buffer, 50 mM KCl, 20 mM MgCl₂, pH 7.5). The column was then washed with 20 column volumes of 5 mM ATP in buffer 1 to remove co-purifying GroEL (Cpn 60). After re-equilibration in 50 mM Tris, pH 8.0, the fusion proteins were eluted using 50 mM reduced glutathione in 50 mM Tris, pH 8.0. The eluted proteins were dialyzed against 50 mM Tris, pH 8.0, overnight at 4 °C, then stored in 10% glycerol, 200 mM NaCl at −70 °C. Circular dichroism spectra of the wild type GST-E2 and R304A GST-E2 proteins (0.2 mg/ml in 25 mM phosphate buffer, pH 7.9) were obtained using a JY CD 6 spectrometer with 2-mm path length cells. Spectra were collected at 1-nm increments, using a 20-s integration time.

**Gel Retardation Assays**—Single stranded oligonucleotides (100 ng) were 5′-end labeled with [γ-32P]ATP using T4 polynucleotide kinase. After annealing to the complementary oligonucleotide, free label was removed using Sephadex G-50 columns (Stratagene). Labeled oligonucleotides (10,000 cpm) were incubated with purified proteins (in the quantities indicated in the Figures) in binding buffer (20 mM HEPES pH 7.9, 25 mM KCl, 1 mM dithiothreitol, 0.1% Nonidet P-40, 10% glycerol, 0.5 μg μl⁻¹ bovine serum albumin, 80 ng μl⁻¹ poly(dI-C)]. After 20 min at 20 °C the complexes were resolved on 6% nondenaturing polyacrylamide gels run in 0.5 × TBE, visualized by autoradiography, and quantified using a PhosphorImager. Non-specific competitor DNA was omitted from gel retardation assays involving the GST-hE2 R304A protein.

**Circular Permutation Assays**—The oligonucleotides shown below carry hE2-binding sites 3 and 4, respectively, and were cloned into pBluescript KS (Stratagene) as NcoI or EcoRI fragments. The oligonucleotides were 5′-end labeled with [γ-32P]ATP using T4 polynucleotide kinase. Unlabeled oligonucleotides (10,000 cpm) were incubated with purified GST-hE2 in the binding buffer described above. Free and bound DNA was resolved and visualized as described in the previous section. Each experiment was repeated at least three times and the DNA bend angle (α) was calculated using the empirical relationship,

\[
\mu_m/\mu_e = \cos(\alpha/2)
\]

where \(\mu_m\) equals the mobility of the complex with hE2 bound in the middle of the fragment, and \(\mu_e\) equals the mobility of the complex with hE2 bound near the end of the fragment (22).

**RESULTS**

**Purification of a GST-hE2 Fusion Protein**—The full-length hE2 protein is poorly expressed in bacteria, shows low solubility, and is unstable (data not shown). To avoid these problems we cloned the HPV 16 E2 gene into the GST expression vector pGEX-2T and used a modification of the standard GST fusion purification procedure (23) to obtain pure GST-hE2. The use of GST-hE2 fusion proteins to improve solubility and stability of hE2 and also allows rapid purification. However, GST-hE2 was found to co-purify with a bacterial protein of around 60 kDa (Fig. 2, lane 2). N-terminal peptide sequencing showed that the co-purifying protein was GroEL, also known as Cpn 60 (data not shown). The GroEL chaperonin is overexpressed by bacteria on cell stress and frequently co-purifies with recombi-
The binding of ATP to GroEL has been shown to induce a conformational change from a tight-binding form of the protein to a weak-binding form (24). To remove the co-purifying GroEL we included an ATP washing step in the purification procedure (25). GST-hE2/GroEL, bound to a glutathione-Sepharose column, was washed with 5 mM ATP in column buffer. GST-hE2 free of GroEL could then be eluted from column using reduced glutathione (Fig. 2a, lane 3).

**Binding of hE2 to the HPV 16 hE2-binding sites.**—We used gel retardation assays to study the binding of hE2 to the four hE2-binding sites found in the HPV 16 genome. Increasing amounts of purified GST-hE2 were added to labeled oligonucleotides carrying hE2 sites 1–4 (shown in Fig. 3a). Free and bound oligonucleotides were then separated by polyacrylamide gel electrophoresis and visualized by autoradiography. Binding to hE2 site 1 was repeated on each gel as an internal control. Fig. 3b shows the results of a typical experiment; in this case, the binding of GST-hE2 to hE2 site 1 (Fig. 3b, lanes 1–7) was compared with the binding GST-hE2 to hE2 site 2 (Fig. 3b, lanes 8–14). The binding of GST-hE2 results in the formation of two retarded bands (indicated by arrowheads in Fig. 3) probably as a consequence of the dimerization of GST. The quantity of free and bound DNA (both complexes) was determined using a PhosphorImager and the percentage of oligonucleotide bound ([bound oligonucleotide] / [free oligonucleotide] x 100) is plotted against the amount of hE2 added. Although site 3 shows poor binding and does not reach saturation, each of the binding curves extrapolates to the same end point (within experimental error). The experiments were repeated a minimum of four times and weak binding curves were fitted to the data to determine the dissociation constants shown in Table I. Binding to site 1 was repeated on each gel as an internal control. ●, site 1; ○, site 2; ▲, site 3; △, site 4.

**DNA Bending by the hE2 Protein.**—In the x-ray crystallography-derived structure of the hE2 DNA-binding domain-DNA complex, the DNA is bent around the protein through an angle of around 50° (13). As the HPV 16 P97 promoter contains four hE2-binding sites, any DNA bending caused by the binding of hE2 would be expected to have consequences for promoter architecture. We used gel retardation assays to investigate the DNA bending properties of the full-length hE2 protein at both tight and weak hE2-binding sites. hE2 sites 3 and 4, the weakest and tightest of the HPV 16 hE2 sites, respectively, were cloned into the DNA-bending vector pBend3 (26). Restriction digestion of the resulting pBend3-hE2(3) and pBend3-hE2(4) plasmids generated a set of circularly permuted fragments in which the position of the E2-binding site is varied (see Fig. 4a).

**Fig. 2.** Removal of co-purifying GroEL from GST-hE2. An ATP washing step was used to remove co-purifying GroEL from preparations of GST-hE2. The figure shows protein samples taken before (lane 2) and after (lane 3) the ATP wash and analyzed by SDS-PAGE. The GST-hE2 fusion protein (indicated by the arrowhead) migrates slightly slower than the 60-kDa GroEL band. The low molecular weight band present after (lane 3) is a breakdown product of GST-hE2 and corresponds in size to GST alone.

**Fig. 3.** Binding of hE2 to the HPV 16 hE2-binding sites. a, synthetic oligonucleotides carrying the four hE2-binding sites from the HPV 16 P97 promoter. b, labeled oligonucleotides carrying hE2 site 1 (lanes 1–7) or hE2 site 2 (lanes 8–14) were incubated with increasing amounts of purified hE2 (40, 80, 160, 320, 640, and 1050 nM, respectively). Free and bound DNA (hE2) were separated on a 6% polyacrylamide gel and visualized by autoradiography. c, the graph shows the binding of hE2 to sites 1–4 assayed as described above. At each protein concentration the amount of free and bound DNA was determined using a PhosphorImager and the percentage of oligonucleotide bound ([bound oligonucleotide] / [bound oligonucleotide] x 100) is plotted against the amount of hE2 added. Although site 3 shows poor binding and does not reach saturation, each of the binding curves extrapolates to the same end point (within experimental error). The experiments were repeated a minimum of four times and weak binding curves were fitted to the data to determine the dissociation constants shown in Table I. Binding to site 1 was repeated on each gel as an internal control. ●, site 1; ○, site 2; ▲, site 3; △, site 4.
Between βE2 and hE2, we reasoned that Arg-304 of hE2 (the Arg equivalent of Arg-344 in bE2) might be responsible for the recognition of a purine at the −7/+7 position of the hE2-binding site.

Using site-directed mutagenesis we changed Arg-304 of hE2 to alanine. hE2 R304A was expressed in bacteria as a GST fusion protein and purified exactly as described for the wild-

Table I

| E2 site | Apparent dissociation constanta | Relative dissociation constantb | ΔG |
|---------|--------------------------------|-------------------------------|-----|
| 1       | 203                            | 1                             | 0.0 |
| 2       | 110                            | 0.54 ± 0.26                   | −1.6 ± 0.20 |
| 3       | 568                            | 2.79 ± 1.04                   | +2.6 ± 0.15 |
| 4       | 75                             | 0.37 ± 0.16                   | −2.6 ± 0.25 |

a The apparent dissociation constant (Kd(apparent)) was obtained using the equation: [bound DNA] = [maximum bound DNA] [protein]/[protein] + Kd(apparent). For weak binding curves ([DNA] ≪ Kd), the apparent dissociation constant is equal to the protein concentration at half-maximum oligonucleotide binding. The values shown are the average of at least four independent experiments.

b The relative dissociation constant, Krel = Kd/Kx, and ΔG = RTln(Kx/Kd), where Kx = apparent dissociation constant for E2 binding site 1, and Kd = apparent dissociation constant for sites 2, 3, or 4.

hE2 to site 3 (the weak-binding site) and site 4 (the tight-binding site) induced DNA bends of 61.4° ± 3.0° and 61.2° ± 1.8°, respectively. Thus, hE2 induces very similar, if not identical, DNA bends at both weak- and tight-binding sites. It is interesting to note that the experimentally determined hE2-induced DNA bend angle is in close agreement with the bE2-induced DNA bend angle (50°) measured from the crystal structure of the bE2-DNA complex (13).

hE2 Recognizes an Extended Binding Site—An alignment of the four hE2-binding sites present within the HPV 16 P97 promoter produces a consensus sequence that differs from that obtained when the bE2-binding sites are aligned (Fig. 5a). While the consensus bE2 site shows no base pair preference at the −7 position, or the symmetrically related +7 position, the HPV 16 hE2 sites all contain an adenine at −7 and a thymine at +7. To investigate the significance of this difference we assayed the binding of hE2 to a series of otherwise identical oligonucleotides containing A:T, T:A, G:C, C:G, or A:U at the −7/+7 position (shown in Fig. 5b). The binding of hE2 to each of these sites was assayed exactly as described in Fig. 3 and is shown graphically in Fig. 5c. The effect of different bases at the −7/+7 position on the relative dissociation constant is shown in Table II. These data show that changing the −7/+7 position from A:T to either T:A or C:G significantly reduces the binding of hE2 (8-fold and 4-fold reductions in binding, respectively). In contrast, changing the −7/+7 position from A:T to G:C has little or no effect. Given that the structure of the bE2-DNA complex derived by x-ray crystallography indicates that the hE2 protein makes specific contacts with base pairs in the major groove, these data suggest that hE2 contacts the N-7 atoms of adenine and guanine when either A:T or G:C are at the −7/+7 position of the hE2-binding site. Changing the −7/+7 position from A:T to A:U resulted in only a slight reduction in the binding of hE2. This would also seem to indicate that the hE2 protein makes specific contact with the purine at the −7/+7. Taken together these data suggest that the hE2 protein forms a specific contact with a purine at the −7/+7 position and that this contact stabilizes the binding of hE2 to the HPV 16 E2 sites.

Arg-304 of hE2 Contacts the −7/+7 Position—Although the structure of the bE2-DNA complex has been determined by x-ray crystallography, no structural information is available on the hE2-DNA interaction. The hE2 protein makes no direct contacts with the G:C base pair at the −7/+7 position of the hE2-binding site. However, Arg-344 in bE2 appears to form a water-mediated hydrogen bond with the guanine at this position as well as side chain hydrogen bonds to the phosphate groups of the −7 and −8 base pairs (13). Since the amino acids involved in sequence-recognition appear to be conserved between bE2 and hE2, we reasoned that Arg-304 of hE2 (the equivalent of Arg-344 in bE2) might be responsible for the recognition of a purine at the −7/+7 position of the hE2-binding site.

Using site-directed mutagenesis we changed Arg-304 of hE2 to alanine. hE2 R304A was expressed in bacteria as a GST fusion protein and purified exactly as described for the wild-type bE2 protein. While the consensus bE2 site shows no base pair preference at the −7 position, or the symmetrically related +7 position, the HPV 16 hE2 sites all contain an adenine at −7 and a thymine at +7. To investigate the significance of this difference we assayed the binding of hE2 to a series of otherwise identical oligonucleotides containing A:T, T:A, G:C, C:G, or A:U at the −7/+7 position (shown in Fig. 5b). The binding of hE2 to each of these sites was assayed exactly as described in Fig. 3 and is shown graphically in Fig. 5c. The effect of different bases at the −7/+7 position on the relative dissociation constant is shown in Table II. These data show that changing the −7/+7 position from A:T to either T:A or C:G significantly reduces the binding of hE2 (8-fold and 4-fold reductions in binding, respectively). In contrast, changing the −7/+7 position from A:T to G:C has little or no effect. Given that the structure of the bE2-DNA complex derived by x-ray crystallography indicates that the hE2 protein makes specific contacts with base pairs in the major groove, these data suggest that hE2 contacts the N-7 atoms of adenine and guanine when either A:T or G:C are at the −7/+7 position of the hE2-binding site. Changing the −7/+7 position from A:T to A:U resulted in only a slight reduction in the binding of hE2. This would also seem to indicate that the hE2 protein makes specific contact with the purine at the −7/+7. Taken together these data suggest that the hE2 protein forms a specific contact with a purine at the −7/+7 position and that this contact stabilizes the binding of hE2 to the HPV 16 E2 sites.

Using site-directed mutagenesis we changed Arg-304 of hE2 to alanine. hE2 R304A was expressed in bacteria as a GST fusion protein and purified exactly as described for the wild-type bE2 protein.
type protein. Circular dichroism (CD) was used to test whether the presence of the R304A mutation altered the folding or dimerization of hE2. The CD spectra for the wild-type and mutant proteins (shown in Fig. 6a) are very similar, suggesting that this mutation has little or no effect on these properties. The R304A protein did show reduced affinity for DNA compared with the wild-type protein (data not shown) and consequently, binding to the position 7 variants was assayed in the absence of competitor DNA. Fig. 6b shows the results of a gel retardation assay in which increasing amounts of hE2 R304A were added to labeled oligonucleotides carrying E2-binding sites with either A:T or T:A at position 27/17. As can be seen from the Figure, hE2 R304A binds equally well to both sites indicating that this mutant has lost the ability to discriminate between A:T and T:A at this position. These data imply that arginine 304 of the hE2 protein makes a specific contact with the N-7 of purine bases at the 27/17 position of the hE2-binding site.

**DISCUSSION**

The DNA-binding domain of the bE2 protein is almost identical in structure to the core domain of the Epstein-Barr virus origin binding protein, EBNA1 (27). However, unlike the bE2 C-terminal domain which forms multiple contacts with the base pairs in the E2-binding site, the core domain of the EBNA1 protein makes no direct interactions with the DNA (28). The sequence-specific contacts between EBNA1 and its
DNA binding site appear to be formed by two “flanking domains” that lie outside the E2 homology region. Each flanking domain consists of an α-helix and an extended polypeptide chain that form major and minor groove contacts with the DNA (28). Sanders and Maitland (29) have assayed the binding of the isolated hE2 C terminus to the four hE2-binding sites present within the HPV 16 P97 promoter (29). As amino acids outside this C-terminal region could form important contacts with the DNA we have used the full-length protein to determine relative dissociation constants for the interaction of hE2 with these sites. The values obtained for the full-length protein are similar to those which have previously been obtained using the isolated C-terminal domain (29). This suggests that, unlike the EBNA1 core domain, the hE2 C-terminal domain is the sole determinant of DNA binding affinity and specificity.

The data obtained using both the full-length hE2 protein, and the isolated C-terminal domain, indicate that hE2 binds most tightly to E2 site 4, then sites 1 and 2, and most weakly to site 3. The weak binding to site 3 is unsurprising given that unlike sites 1, 2, and 4, this site is not a perfect match to the “consensus” binding site (see Fig. 5a). The tight binding of hE2 to site 4 suggests that at low E2 concentrations this site would be occupied first. E2 site 4 is located around 550 base pairs upstream of the P97 promoter in an “enhancer configuration.”

As the binding of hE2 to sites in this configuration has been shown to activate transcription from downstream promoters (21, 17), the binding of hE2 to E2 site 4 would be expected to activate transcription from the P97 promoter. At higher E2 concentrations E2 sites 1 and 2 would become occupied, possibly displacing Sp1 and the TATA box-binding protein (20). The binding of hE2 to these promoter proximal sites would thus be expected to result in transcriptional repression. The P97 promoter directs transcription of the viral early genes, including E2. Our binding data suggest that the hE2 protein might positively and negatively regulate transcription of its own gene at low and high E2 concentrations, respectively. This hypothesis is consistent with the results of previous studies which have shown that although hE2 activates transcription from the P97 promoter, increased expression of the hE2 protein results in transcriptional repression (18). Similar concentration-dependent effects of HPV 8 E2 have been observed at the HPV 8 late gene promoter (30).

The binding of many transcription factors has been shown to induce DNA bending and, in some cases, these protein-induced DNA bends have important consequences for the regulation of gene expression. The DNA bend induced by E2F, for example, is important for the transcriptional activity of the E2F1 promoter (31). The DNA-binding domain of the hE2 protein has previously been shown to induce DNA bending in an enhancer fragment containing three hE2-binding sites (32). Here we have shown that binding of the full-length hE2 protein induces a significant bend in the hE2-binding site (around 60°). As the P97 promoter contains four binding sites for hE2, the binding of this protein probably has dramatic effects on the architecture of this DNA. This conformational change might be of importance in the regulation of P97 promoter activity and, as hE2 interacts with the HPV origin recognition protein E1 (33), the regulation of viral DNA replication. Interestingly, the degree of DNA bending induced by the binding of the hE2 protein is the same at both tight and weak hE2-binding sites. In this respect hE2 appears to be similar to the 434 repressor protein which has also been shown to induce identical DNA bends at both tight- and weak-binding sites (34). In contrast, the degree of DNA bending induced by the cAMP receptor protein (CRP or CAP) has been shown depend on the strength of the protein-DNA interaction (35).

An alignment of the four E2-binding sites present in the HPV 16 P97 promoter produces a consensus sequence which differs from that obtained when the E2-binding sites from the BPV genome are aligned. All four E2-binding sites from HPV 16 contain an A:T base pair at the −7 position and a T:A base pair at the symmetrically related +7 position (shown in Fig. 5a). We have shown that mutations at these positions can significantly reduce the binding of hE2. These results confirm and extend the previous observation that a change from A:T to C:G at this position weakens the binding of the hE2 C-terminal domain (29). In contrast, the consensus hE2-binding site shows no preference for A:T at position 7 and a change from A:T to C:G at this position has been shown to result in a slight increase in the binding of hE2 (8).

Having shown that the hE2 protein discriminates between base pairs at position 7 of its binding site, we set out to determine the amino acid responsible for this effect. Inspection of the hE2-DNA complex showed that arginine 344 (equivalent to arginine 304 in hE2) makes a water-mediated hydrogen bond to the N-7 and O-6 of the guanine base at the −7 position of the hE2-binding site used in the co-crystal (13). We reasoned that arginine 304 in hE2 might make a contact (either direct or water-mediated) with the N-7 of adenine or guanine present at the position 7 of the hE2-binding site. To test this hypothesis we mutated arginine 304 of hE2 to alanine (R304A) and assayed the binding of the mutant protein to sites containing either an A:T or a T:A at position 7. Unlike the wild-type protein, which showed an 8-fold preference for A:T over T:A at this position, the E2 R304A protein bound equally to both sites, albeit with reduced affinity. This loss of specificity suggests that arginine 304 of hE2 makes a specific contact with position 7 of the hE2-binding site. Given the flexibility of surface arginine residues it would seem possible that hE2 R344 could make direct contact with a purine at position 7 of the hE2-binding site. However, the lack of sequence-specificity at position 7 of the hE2 site suggests that this contact does not occur. The recognition of the base pair at position 7 of the hE2-binding site might not be solely attributable to the Arg-304 side chain but might also involve other contacts that are disrupted when this amino acid is changed to alanine. The lack of a preferred base pair at position 7 of the hE2 consensus site would also seem to indicate that the water-mediated contact between hE2 Arg-344 and the hE2 site makes little contribution to the specificity of the hE2-DNA interaction, however, the interactions with backbone phosphates made by this residue might make an important contribution toward the affinity of hE2 for DNA.

There is a growing body of evidence which suggests that the hE2 and hE2 proteins have significantly different biological properties. For example, the hE2 protein has been shown to activate transcription from the HPV 16 P97 promoter whereas, under exactly the same conditions, the hE2 protein has been shown to repress P97 promoter activity (18). Our results highlight a further difference between these proteins; in comparison to hE2, the hE2 protein recognizes an extended binding site. These functional differences between hE2 and hE2 suggest that caution must be exercised when extrapolating from one system to the other.

Acknowledgment—We thank Professor Steve Halford for comments on the manuscript.

REFERENCES

1. zur Hausen, H. (1991) Virology 184, 9–13
2. Mansur, C. P., and Androphy, E. J. (1993) Biochim. Biophys. Acta 1155, 323–345
3. Smotkin, D., and Wettstein, F. O. (1986) Proc. Natl. Acad. Sci. U. S. A. 83, 4680–4684
4. Cripe, T. P., Haugen, T. H., Turk, J. P., Tabatabai, F., Schimid, P. G., Dürst, M., Gissman, L., Roman, A., and Turek, L. P. (1987) EMBO J. 6, 3745–3753
5. Phelps, W. C., and Howley, P. M. (1987) J. Virol. 61, 1630–1638
DNA Binding and Bending by the HPV 16 E2 Protein

6. Ishiji, T., Lace, M. J., Parkkinen, S., Anderson, R. D., Haugen, T. H., Cripe, T. P., Xiao, J. H., Davidson, I., Chambon, P., and Turek, L. P. (1992) *EMBO J.* 11, 2271–2281
7. Spalholz, B. A., Yang, Y.-C., and Howley, P. M. (1985) *Cell* 42, 183–191
8. Li, R., Knight, J., Bream, G., Stenlund, A., and Botchan, M. (1989) *Genes Dev.* 3, 510–526
9. Spalholz, B. A., Byrne, J. C., and Howley, P. M. (1988) *J. Virol.* 62, 3143–3150
10. Monini, P., Grossman, S. R., Pepinsky, B., Androphy, E. J., and Laimins, L. A. (1991) *J. Virol.* 65, 2124–2130
11. Knight, J. D., Li, R., and Botchan, M. (1991) *Proc. Natl. Acad. Sci. U. S. A.* 88, 3204–3208
12. McBride, A. A., Byrne, J. C., and Howley, P. M. (1989) *Proc. Natl. Acad. Sci. U. S. A.* 86, 510–514
13. Hegde, R. S., Grossman, S. R., Laimins, L. A., and Sigler, P. B. (1992) *Nature* 359, 505–512
14. Chong, T., Chan, W. K., and Bernard, H. U. (1990) *Nucleic Acids Res.* 18, 465–470
15. Romanzuk, H., Thierry, F., and Howley, P. M. (1990) *J. Virol.* 64, 2849–2859
16. Tan, S.-H., Gloss, B., and Bernard, H. U. (1992) *Nucleic Acids Res.* 20, 251–256
17. Ushikai, M., Lace, M. J., Yamakawa, Y., Kono, M., Anson, J., Ishiji, T., Parkkinen, S., Wicker, N., Valentine, M.-E., Davidson, I., Turek, L. P., and Haugen, T. H. (1994) *J. Virol.* 68, 6655–6666
18. Bouvard, V., Storey, A., Pim, D., and Banks, L. (1994) *EMBO J.* 13, 5451–5459
19. Dostatni, N., Lambert, P. P., Souza, R., Ham, J., Howley, P. M., and Yaniv, M. (1991) *Genes Dev.* 5, 1657–1671
20. Tan, S.-H., Leong, L. E.-C., Walker, P. A., and Bernard, H.-U. (1994) *J. Virol.* 68, 6411–6420
21. Ham, J., Steger, G., and Yaniv, M. (1994) *EMBO J.* 13, 147–157
22. Thompson, J. F., and Landy, A. (1988) *Nucleic Acids Res.* 20, 9687–9705
23. Smith, D. B., and Johnson, K. S. (1988) *Gene (Amst.)* 67, 31–40
24. Badcoe, I. G., Smith, C. J., Wood, S., Halsall, D. J., Holbrook, J. J., Lund, P., and Clarke, A. R. (1991) *Biochemistry* 30, 9185–9200
25. Thain, A., Gaston, K., Jenkins, O., and Clarke, A. R. (1996) *Trends Genet.* 12, 209–210
26. Kim, J., Zwieb, C., Wu, C., and Adhya, S. (1989) *Gene (Amst.)* 85, 15–23
27. Bochkarev, A., Barwell, J., Pfuetzner, R., Purey, W. J., Edwards, A. M., and Frappier, L. D. (1995) *Cell* 83, 39–46
28. Bochkarev, A., Barwell, J. A., Pfuetzner, R. A., Bochkareva, E., Frappier, L., and Edwards, A. M. (1996) *Cell* 84, 791–800
29. Sanders, C. M., and Maitland, N. J. (1994) *Nucleic Acids Res.* 22, 4890–4897
30. Stubenrauch, F., Leigh, I. M., and Pfister, H. (1996) *J. Virol.* 70, 119–126
31. Cress, W. D., and Nevins, J. R. (1996) *Mol. Cell Biol.* 16, 2119–2127
32. Moskaluk, C., and Bastia, D. (1988) *Proc. Natl. Acad. Sci. U. S. A.* 85, 1826–1830
33. Frattini, M. G., and Laimins, L. A. (1994) *Proc. Natl. Acad. Sci. U. S. A.* 91, 12398–12402
34. Koudelka, G. B. (1991) *Nucleic Acids Res.* 19, 4115–4119
35. DalmaWeiszhausz, D. D., Gartenberg, M. R., and Crothers, D. M. (1991) *Nucleic Acids Res.* 19, 611–616