4890–4897 Nucleic Acids Research, 1994, Vol. 22, No. 22

Kinetic and equilibrium binding studies of the human papillomavirus type-16 transcription regulatory protein E2 interacting with core enhancer elements

Cyril M.Sanders and Norman J.Maitland*
Cancer Research Unit, Department of Biology, University of York, Heslington, York YO1 5DD, UK

Received September 20, 1994; Accepted October 12, 1994

ABSTRACT

The human papillomaviruses (HPVs) are a family of DNA viruses which cause benign tumours of the skin and mucosa that infrequently progress to malignant carcinoma. The E2 open reading frame of HPV is thought to encode a papillomavirus-specific transcription factor which also has a role in viral replication. The E2 proteins of all papillomaviruses studied to date have been shown to bind specifically to the common conserved sequence ACC$(N)_6$GGT found at multiple locations in their genomes. In the case of HPV-16, a 'high risk' genital papillomavirus, the E2 protein is thought to negatively regulate expression of the major viral transforming genes E6 and E7, which have been directly implicated in the oncogenic process. However, little information exists concerning the relative or absolute affinities of the native HPV-16 protein for its palindromic recognition sequences; moreover, interpretation of any transcription or replication phenomena attributed to this protein is more complicated in the absence of such data. Here we describe the overexpression, purification and characterisation of the C-terminal 89 amino acids of the protein encompassing the DNA binding/dimerisation domain. We show that the recombinant protein purified from *E.coli* by a combination of non-group-specific chromatography steps retains high biological activity and is able to bind to all sites in the HPV-16 genome with high affinity ($\sim 8 \times 10^{-11}$ M). In addition, kinetic studies show that the E2 – DNA complexes are very stable, with half-lives ranging from 2.15 to greater than 240 min, and that nucleotides internal and external to the conserved palindrome appear to influence stability.

INTRODUCTION

The papillomaviruses are a diverse group of small DNA tumour viruses that infect a wide variety of vertebrate hosts, including reptiles, avian species and humans, where over 70 types are known (1,2). Viral infection is restricted to mucosal or cutaneous sites where viral gene expression is tightly linked to the differentiation status of the cell: complete viral particles are shed from the terminally differentiated cells of the outer dermis, while the undifferentiated stem cells abutting the basement membrane are presumed to maintain the viral genome. The nature of the virus life cycle has severely restricted, or rendered the propagation of these viruses *in vitro* impossible, thus hampering the definition of viral genes and regulatory circuits governing the viral life cycle. Viral infection may be latent, sub-clinical or apparent as benign epithelial tumours (papillomas). However a small sub-group of the human papillomaviruses, the 'high risk' genital papillomaviruses, most notably HPV-16 but also including HPV-18, -31 and -33, have been implicated in the aetiology of squamous cell carcinoma of the cervix and other anogenital tumours (3). The viral early genes E6 and E7, the products of which bind the tumour suppressor genes *p53* and *Rb* respectively (4–6), are selectively retained in these malignancies, suggesting strongly that papillomaviruses may participate directly in the oncogenic process. Elevated expression of E6 and E7 has been associated with the loss of the viral E2 gene (7–9), which is presumed to encode a papillomavirus-specific transcription factor.

The E2 open reading frame initially aroused interest and speculation that it was a modulator of papillomavirus transcription on account of the pleiotropic effect on viral functions, such as replication and oncogenic transformation *in vitro*, which resulted after mutation or disruption of the open reading frame (10,11). The BPV-1 protein was subsequently shown to be a site-specific DNA binding protein which recognised the sequence A-CC$(N)_6$GGT (12,13). This motif has been found in the genomes of all papillomaviruses sequenced to date (14), and is repeated 17 times in the BPV-1 genome, 12 sites being clustered in the viral regulatory region or enhancer (URR), and at least one site being associated with each of the major viral promoters (15). In the bovine system, the E2 protein is a necessary activator of viral transcription capable of stimulating transcription from the BPV E2 conditional enhancer by 300-fold (16–21). The protein is also required in association with the viral E1 protein for replication (22,23). In addition to the full-length E2 protein, BPV-1 expresses two functional N-terminally truncated polypeptides which may antagonise the activity of the full-length

*To whom correspondence should be addressed

transcriptional activator and have a role in autoregulation of viral transcription and replication (24–27).

The full-length BPV-1 E2 protein is a 48 kDa phosphoprotein (28) which dimerises spontaneously in the absence of DNA to form stable molecules (29,30). The E2 proteins of all papillomaviruses are organised into three domains, as shown in Figure 1: a well-conserved N-terminal domain involved in transactivation, a central hinge of variable sequence and length, and a C-terminal domain responsible for DNA binding and dimerisation (31). Recent crystallographic data for the core BPV-1 E2 DNA binding domain (C-terminal 85 amino acids) has revealed a novel dimeric DNA binding/dimerisation motif previously undescribed or unpredicted in a eucaryotic transcription factor (32). Two monomers interact to form an eight-stranded antiparallel β-barrel which is stabilised by an extensive network of hydrogen bonds encompassing a central hydrophobic core where two conserved tryptophans critical for dimerisation (30) are in van der Waals contact. Target DNA is bent over the barrel to align a pair of symmetrically disposed recognition helices with the major groove of the conserved nucleotides in each palindromic half site. Four discriminating amino acids in each helix are involved in multiple protein–DNA interactions. Of these, a conserved cysteine (Cys340) has been implicated in redox regulation of DNA binding (33).

BPV-1 E2 has been shown to bind to a single palindromic recognition sequence with a K_d of 2×10^{-11} M, the core DNA binding domain with a K_d of 4×10^{-10} M and a co-operative parameter of 8.5 has been measured for two full-length dimers interacting with two adjacent sites (34). Nuclease protection and chemical probing techniques had previously indicated symmetrical contacts with the G residues in each half site and possible interactions extending beyond the conserved core (15,35). A comprehensive and extensive mutational analysis of a typical E2 binding site and individual native sites from the BPV-1 enhancer has addressed the question of DNA context with regard to DNA binding affinities (15). The authors of this survey concluded that, overall, binding sites that contained the extended palindrome A-CCG(N)$_4$CGGT have a higher affinity; for a particular palindrome, a single nucleotide difference in the internal core or flanking residues could result in a change in relative binding affinities over a 24-fold range, and that in some instances, two native E2 enhancer binding sites in the BPV-1 genome could vary in relative affinity $\geq$300-fold.

Genital papillomaviruses contain four binding sites for the E2 protein arranged in a characteristic fashion (14,36). In HPV-16 three sites are in a promoter proximal position 5′ to the major promoter P97, which is responsible for the transcription of the oncogenes E6 and E7 (7,9). The two most 5′ sites are juxtaposed in a tandem array and are thought to overlap and negatively interfere with the binding of the SP-1 transcription factor and the TATA binding protein (TBP), thus facilitating a mode of E2-mediated down-regulation of transcription (37,38). The three promoter proximal sites also encompass, or are within, a putative origin of replication (39). The fourth E2 binding site is at the distal end of the viral enhancer. All of the HPV-16 E2 binding sites contain an internal A/T-rich core, which is proposed to increase the affinity of the sites for the native viral protein (40), and are flanked by conserved residues, 5′-A and 3′-T, respectively. Here we describe the cloning, expression, purification and characterisation of the HPV-16 E2 DNA binding domain. We show that the purified protein retains high specific activity and is capable of interacting with all four native core E2 enhancer sequences with high affinity. The stability of the various complexes was measured in terms of the half-life, and was shown to vary over greater than a 100-fold range. In addition, the conserved nucleotides flanking the core sequence were also shown to influence stability. This information may be important in interpreting any transcription or replication phenomena that could be attributed to the activity of the HPV-16 E2 protein.

MATERIALS AND METHODS

Construction of recombinant plasmids

Paired oligonucleotide primers 5′-TAAGACTCCATGGC-ACATATGGGACGGATTAACTGTAATAGT-3′ and 5′-AGC-AGCAAGTGGATCCGCTAGCAGCACGCCAGTAATGTT-3′ were used to amplify a 352 bp fragment containing the sequence from nucleotides 3583 to 3891 in the 5′ region of the HPV-16 E2 open reading frame for direct insertion into the TA-cloning vector pT7blue (Novagen). The flanking sequences of the PCR primers contained restriction enzyme sites to facilitate further manipulations. An *Nco*I–*Bam*HI sub-fragment was released from the resulting clone and inserted into the vector pET-15b (Novagen) which had been cleaved with the same enzymes. The resulting plasmid pET-E2CT retained the promoter and *lac* operator control regions of the parent vector, but not the sequences encoding the His-Tag affinity leader. Construct pET-E2CT contained an open reading frame consisting of the C-terminal 90 codons of the E2 gene fused to 4 codons (methionine, alanine, histidine, methionine), under the control of the bacteriaphage T7 promoter. The DNA sequence was verified by Δ-Taq cycle sequencing (USB Corporation).

Protein expression and purification

The E2 C-Terminal DNA binding domain was expressed in the *E.coli* strain JM109(DE3) transformed with the plasmid pET-E2CT. Bacteria were grown at 25°C in Circlegrow medium (BIO101 Inc.) to an OD_{600} = 0.8–1.0 and induced in the presence of 1 mM IPTG for 5–6 h. Bacteria were harvested by centrifugation, washed in 0.1 culture volumes of 20 mM Tris–HCl (pH 7.9, 4°C), 200 mM NaCl and stored at −80°C. Frozen pellets were lysed at 4°C by sonication in 3 ml/g of lysis buffer (50 mM Tris–HCl, pH 7.9, 200 mM NaCl, 2 mM EDTA, 2 mM EGTA, 10 mM DTT, 0.5 mM PMSF) and the extract clarified by centrifugation at 20 000 *g* for 1 h at 4°C. The soluble fraction was applied to an ion exchange column (SP Sepharose; Pharmacia Biotech) equilibrated with low salt buffer [50 mM bicine, pH 8.9 (4°C), 350 mM NaCl, 1 mM EDTA, 5 mM DTT], and the column developed on a simple linear gradient over 10 column volumes from 350 to 850 mM NaCl in bicine, EDTA, and DTT. Peak fractions were pooled, concentrated (Centriprep 3; Amicon), and resolved on a pre-calibrated Sephacryl S-100 size exclusion column (SEC; 95×1.6 cm) equilibrated in 20 mM Tris–HCl, 400 mM NaCl, 5 mM EDTA, 5 mM DTT, pH 8.4 (4°C). Trace contaminants from the peak SEC fractions were removed by heparin chromatography (Hi-Trap heparin Sepharose; Pharmacia Biotech) on a linear gradient from 400 to 1000 mM NaCl in 20 mM Tris–HCl, pH 7.9 (25°C), 5 mM DTT, 2 mM EDTA. The purified protein was dialysed against 20 mM HEPES, pH 7.9 (4°C), 250 mM NaCl, 20% glycerol, 0.2 mM EDTA, 5 mM DTT, aliquoted, snap frozen and stored at −80°C. Protein concentration was determined by absorbance at 280 nm using a calculated molar extinction coefficient of 15 700 (1 A_{280} =

0.68 mg/ml). Protein identity was confirmed by N-terminal sequencing of eight amino acids carried out at the University of Leicester (Department of Biochemistry).

Binding site oligonucleotides

Single-stranded oligonucleotides were synthesised on an Applied Biosystems model 391 DNA synthesiser and purified on the basis of trityl affinity. Two sets of complementary oligonucleotides were synthesised. One set (designated sites 1–4) consisted of the four E2 binding sites from the HPV-16 URR flanked by their native genomic sequences. Tandem sites 3 and 4 contained minor nucleotide changes to reduce identity to the adjacent half sites. Oligonucleotide sequences and map locations of the HPV-16 E2 binding sites are given in Figure 1. A second series (designated oligonucleotides 7450, 7857, 35 and 50) represented each E2 core palindrome flanked by identical nucleotides and modelled on the standard sequence 5′-CTAAGGGCGTAACCGAAATCGGTTGA-3′ upper strand, and 5′-GGATTCAACCGATTTCGGTTACG-3′ for the lower strand, derived from site 3 in the viral genome. Binding site oligonucleotide site 3 and oligonucleotide 35 were identical. Paired, partially complementary oligonucleotides at equimolar concentrations were hybridised for 1 h at the calculated melting temperature after thermal denaturation in 10 mM Tris–HCl, 50 mM NaCl, 1 mM EDTA, resolved on 15% polyacrylamide gels and the specific double-stranded products recovered by soak elution and ethanol precipitation. Nucleic acid concentrations were determined by absorbance at 260 nm using the calculated molar extinction coefficients for each oligonucleotide (corrected for the hyperchromic effect).

A defined quantity of synthetic duplex molecules were labelled to high specific activity by the Klenow fill-in reaction [200 μm each dATP, dGTP, dTTP, 50 μCi [α-^{32}P]dCTP (3000 Ci/mmol, 10 mCi/ml), 3 U Exo$^-$ Klenow (USB Corporation)] and rendered completely double-stranded with a 200 μm dCTP cold chase. Unincorporated nucleotides were removed by gel filtration (Pharmacia NAP-5) in 10 mM Tris–HCl, 100 mM NaCl, 1 mM EDTA and the nucleic acid recovery quantified by Cerenkov counting of DNA bound to DEAE–cellulose after washing in 0.5 M Na_2HPO_4 (average of three counts).

Gel retardation assays

Protein–DNA binding reactions were carried out in 22 mM HEPES pH 7.9, 50 mM KCl, 10 mM NaCl, 5 mM $MgCl_2$, 20 μM EDTA, 5.5 mM DTT, 10% glycerol (v/v), 0.55% Triton X-100, 110 μg/ml BSA at 21°C for 25 min. Reaction products were resolved on native 8% polyacrylamide gels (29:1) containing 10 mM DTT, 2.5% glycerol (v/v) and 22.5 mM Tris–borate/500 μM EDTA (0.25× TBE electrophoresis buffer, pH 8.0), 10 V/cm, 4–8°C for 3–4 h. After electrophoresis gels were dried and subjected to autoradiography.

Quantification of E2 C-terminal DNA binding activity

A fixed quantity of purified E2 C-terminal protein (1×10^{-10} M final) was titrated against an increasing concentration of labelled binding site oligonucleotide to saturation (up to a 50-fold excess) under the conditions described above, and the products analysed by polyacrylamide gel electrophoresis and autoradiography. Autoradiograms were quantified by computer-assisted densitometry and the proportion of active DNA binding protein calculated from the signal of maximum shifted oligonucleotide compared to a known quantity of free probe.

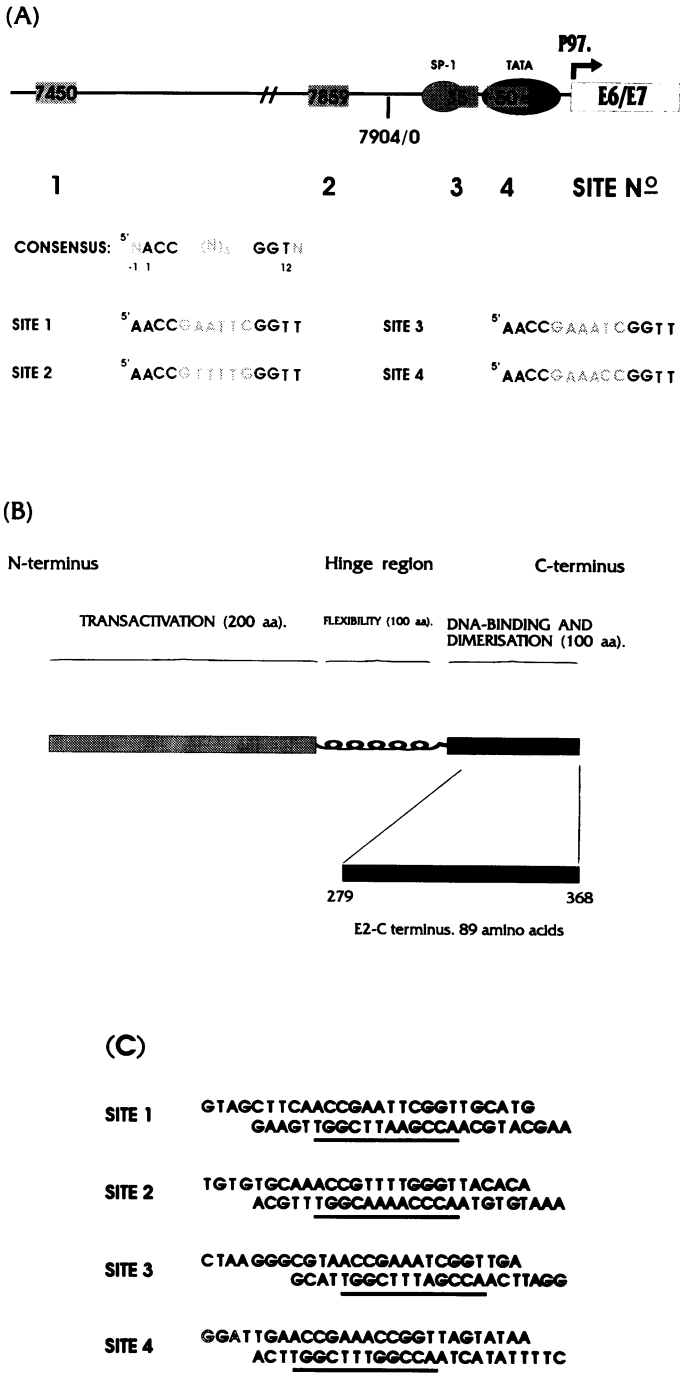

Figure 1. (A) Map location, sequence and relationship between the E2 binding sites in the HPV-16 upstream regulatory region (URR) or enhancer. The numbers are the map positions of the first nucleotides in the 12 bp consensus E2 binding sites, and the sites are labelled 1–4. Three binding sites are proximal to the major viral promoter P97, the cap site for transcripts encoding the viral oncogenes E6 and E7. Two (sites 3 and 4) form a tandem array separated by 3 bp. Occupancy of sites 3 and 4 by E2 protein may negatively interfere with the binding of the SP-1 and TATA binding proteins respectively (37,38). (B) Schematic representation of the HPV-16 E2 protein. The 368 amino acid protein is composed of three distinct domains: an N-terminal transactivation domain, a central 'hinge' and the C-terminal DNA binding/dimerisation domain. The C-terminal 89 amino acids as indicated were cloned, expressed in *E.coli* and purified to homogeneity as described. (C) HPV-16 binding site oligonucleotides. Sequences were derived from the four core motifs in the HPV-16 genome. Flanking sequences are native with the exception of minor changes for sites 3 and 4, without the immediate flanking nucleotides, as indicated.

Determination of apparent equilibrium constants for DNA–protein interactions

Apparent equilibrium constants for the interaction of purified E2 DNA binding domain with native and non-native HPV-16 E2 core enhancer elements were determined as follows. A fixed quantity of [α-^{32}P]-labelled binding site oligonucleotide, varied between 1×10^{-10} M and 5×10^{-12} M in 5 separate experiments, was titrated against a standard protein dilution series (common to the oligonucleotides within an experimental set) in a 10 μl reaction under the conditions described above. Reaction products were resolved and quantified as described above. Numerical data was used to construct a standard binding curve from which the free protein concentration when 50% of the probe is shifted could be determined.

The equilibrium reaction for the protein–DNA interaction can be represented as follows:

$$[\text{protein}_{\text{free}}] + [\text{DNA}_{\text{free}}] \underset{k_2}{\overset{k_1}{\longleftrightarrow}} [\text{protein–DNA complex}]$$

$K_{eq} = ([\text{protein}_{\text{free}}]\,[\text{DNA}_{\text{free}}])/[\text{protein–DNA complex}] = k_2/k_1$

when 50% of the DNA is bound to protein then $K_{eq} = [\text{protein}_{\text{free}}]$, and K_{eq} is the ratio of the forward and reverse rate constants (41).

Dissociation rate determination

Under the reaction conditions described above, each oligonucleotide (at 1×10^{-10} M) was mixed with a 5-fold excess of purified E2 C-terminal protein and the reaction allowed to reach equilibrium over 15 min. At time zero, 10 μl of the reaction was removed prior to the addition of a 500- or 1000-fold excess, depending on the experiment, of specific unlabelled competitor oligonucleotide in a small volume (<1% reaction volume). Equivalent volumes were removed at time intervals and loaded onto a continuously run 8% polyacrylamide gel.

Autoradiographs were processed and quantified as described above and complex half-lives determined from a semi-log plot of fraction bound at time T versus time.

RESULTS

Expression and purification of the HPV-16 DNA binding domain

Enzymatic manipulation of the PCR-generated E2 open reading frame resulted in the generation of plasmid pET-E2CT which expressed the C-terminal 89 amino acids of the HPV-16 E2 gene fused to 4 amino acids at the N-terminus. The ORF was fully characterised by nucleic acid sequencing and maintained 100% identity to the reference clone; N-terminal sequencing of the purified protein confirmed its identity. Protein expression was carried out in the *E.coli* strain JM109(DE3) where the desired product accumulated to ~10% of the soluble protein of the bacterial lysate following 6 h induction at 25°C in the presence of 1 mM IPTG. The induced product exhibited a molecular weight of ~10 kDa in 15% SDS–polyacrylamide gels, which is in agreement with the predicted molecular mass (10.6 kDa) for the E2 C-terminal ORF (Fig. 2). The E2 C-terminal polypeptide was purified to virtual homogeneity, as judged from SDS–polyacrylamide gels, stained with Coomassie Brilliant Blue, essentially in two non-group-specific chromatographic steps (Fig. 2). A first cation exchange step at alkaline pH (bicine, pH 8.9) resulted in a purification to ~80% homogeneity (Fig. 2 Lane 3). E2-CT eluted late in the ion exchange chromatogram under the defined conditions, indicating that the protein's isoelectric point is close to the predicted value of 9.65. A second gel filtration step resulted in purification to near homogeneity (Fig. 2, lane 4) and determined the molecular mass of the protein to be 13 kDa under non-denaturing conditions, which deviates from the predicted dimeric mass of 21 kDa. Trace contaminants were removed by heparin chromatography (Fig. 2, lane 5). Approximately 10 mg of recombinant protein were obtained per g wet wt *E.coli*.

Protein concentrations were determined spectrophotometrically using a calculated molar extinction coefficient of 15 700 and absorbance readings at 280 nm (1 A_{280} =0.68 mg/ml). Values obtained were approximately half those obtained using the Bio-Rad assay.

Quantification of DNA binding activity

The proportion of purified E2 C-terminal polypeptide which was active in DNA binding was determined by titrating a fixed quantity of purified protein against an increasing concentration of purified specific E2 binding site oligonucleotide to saturation as described in Materials and Methods. The quantity of shifted oligonucleotide at saturation, which was assumed to be equivalent to the molar quantity of active DNA binding protein, was determined by densitometry and comparison to a known fixed quantity of free probe run on the same gel. An average value of 79% activity was derived from three separate experiments employing oligonucleotides 7450, 35 and 50 (79.6%, 83.3% and 74.2% respectively). The results of the binding assay for oligonucleotide 7450 are shown in Figure 3.

Determination of apparent equilibrium constants

To assess the contribution of base composition and context in determining the relative binding affinity of the E2 DNA binding domain for consensus motifs in the HPV-16 URR, duplex

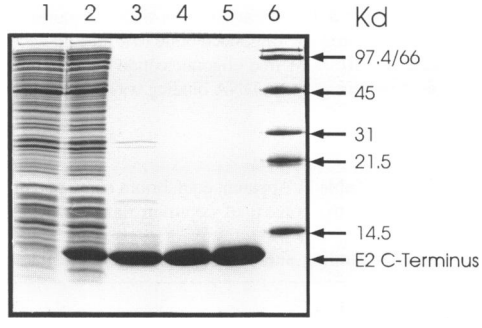

Figure 2. Expression and purification of the HPV-16 E2 C-terminal protein in *E.coli*. The E2 DNA binding domain was expressed from plasmid pET-E2CT in the *E.coli* strain JM109(DE3) and purified to virtual homogeneity as described in Materials and Methods. Proteins were analysed on 15% SDS–polyacrylamide gels stained with Coomassie Brilliant Blue. Lane 1, total protein from *E.coli* strain JM109(DE3) harbouring plasmid pET-15b induced by the presence of 1 mM IPTG. Lane 2, induced cells harbouring plasmid pET-E2CT. Lane 3, pooled peak fractions after ion exchange chromatography. Lane 4, pooled fractions after size exclusion chromatography. Lane 5, ~30 μg of purified protein after heparin chromatography. The faint band at 20 kDa is considered to be unresolved dimeric E2 CT. Lane 6, molecular weight marker proteins.

oligonucleotides were prepared which contained the sequences of each of the four core palindromes in the context of native (oligonucleotide sites 1–4) and non-native, but common, (oligonucleotides 7450, 7857, 35 and 50) flanking sequences. Each oligonucleotide was a 30 bp molecule when rendered completely double-stranded. Oligonucleotide site 3NW (3 non-wild-type) was constructed to directly test the contribution of the conserved A and T residues at positions −1 and +13 that flank all the HPV-16 E2 core motifs. The A and T residues were mutated to C and G respectively. A mutant control oligonucleotide was identical to site 3 except that two of three residues in each conserved E2 half site were altered to non-recognition nucleotides.

The results shown in Table 1 are the average of at least three experiments performed at fixed oligonucleotide concentrations of 1×10^{-10}, 2×10^{-11} and 5×10^{-12} M. Three independent probe labellings and quantifications are represented in the results, and a standard protein dilution series was common to each titration within an experimental set. All E2 binding sites examined showed high affinity and the equilibrium constant values determined (K_{eq}) ranged from 7.66×10^{-11} M (site 1) to 2.8×10^{-10} M (site 2) for the native binding sites, and 6.95×10^{-11} M to 8.278×10^{-11} M for core motifs with non-wild-type flanking nucleotides. For oligonucleotide 3NW, the flanking sequence mutant of wild-type site 3, a K_{eq} of 8.04×10^{-11} M was calculated, which compares to 7.99×10^{-11} M for the parent oligonucleotide. The purified protein had an appreciable non-specific DNA binding affinity ($K_{eq} = 2\times10^{-9}$ M), which was similar for a totally unrelated probe.

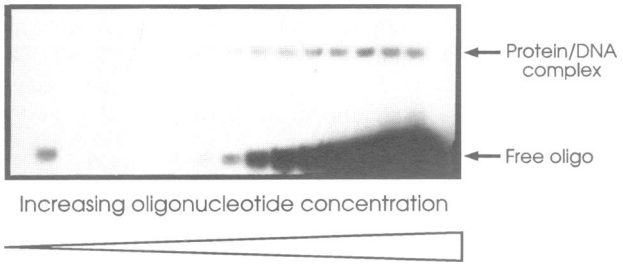

Figure 3. DNA binding activity of purified recombinant E2 DNA binding domain. 1×10^{-10} M final protein in a 10 μl reaction was titrated against an increasing quantity of ^{32}P-labelled binding site oligonucleotide to a 50-fold excess. The lane on the far left contains 1×10^{-10} M free oligonucleotide with a single E2 binding site. The proportion of protein active in DNA binding was determined as described in Materials and Methods.

Off-rate analysis

The dissociation rate for a protein–DNA complex can be measured in terms of the half-life following the addition of a vast excess of specific unlabelled competitor to a reaction that has already reached equilibrium. In these experiments, each ^{32}P-labelled oligonucleotide at a final concentration of 1×10^{-10} M was mixed with a 5-fold molar excess of protein, the reaction was allowed to reach equilibrium, and the dissociation rate determined following the addition of a 500- or 1000-fold excess of homologous competitor. The results are tabulated in Table 2. In contrast to the apparent equilibrium constants, the values show marked differences, ranging from 2.15 min (2 min 9 s) to greater than 240 min for the native binding sites. Binding site 1 was the most stable and site 2 the least, correlating with the measured equilibrium constants (K_{eq}). When the interactions of the core motifs with the E2 protein were assayed in the context of alternative flanking nucleotides, similarly marked differences in complex stability were observed. The half-life of the control mutant oligonucleotide could not be measured accurately using this technique, and is certainly less than 30 s.

DISCUSSION

The E2 open reading frame of the papillomaviruses is presumed to encode a transcription and replication factor that binds to the conserved palindromic motif ACC(N)$_6$GGT found in the enhancers of all papillomaviruses. The HPV-16 E2 protein has been shown to bind to four such sites in the viral enhancer by footprint analysis (40), however, no information exists concerning the relative or absolute affinity of the protein for these sites. Given that the 17 E2 binding sites in the BPV-1 genome interact with E2 with relative affinities ranging over 300-fold, it is important to obtain similar information for the HPV-16 enhancer if it is to be used as a model for HPV gene expression. Moreover, the interpretation of any biological phenomena attributed to the HPV-16 E2 protein is complicated in the absence of such data. In order to quantify the native E2 protein–DNA interactions we have expressed in *E.coli* and purified to homogeneity a 93 amino acid polypeptide containing the C-terminal 89 amino acids of the

Table 1. Apparent equilibrium constants for wild-type E2 binding sites (sites 1–4; A) and core E2 binding palindromes in the context of common flanking nucleotides (B)

Binding site/binding site oligonucleotide	Core sequence	$K_{(eq)}$ apparent
(A)		
Site 1	AACCGAATTCGGTT	7.66×10^{-11}
Site 2	AACCGTTTTGGGTT	2.8×10^{-10}
Site 3 (oligo 35)	AACCGAAATCGGTT	7.83×10^{-11}
Site 4	AACCGAAACCGGTT	9.47×10^{-11}
(B)		
7450	AACCGAATTCGGTT	7.06×10^{-11}
7857	AACCGTTTTGGGTT	8.278×10^{-11}
35 (site 3)	AACCGAAATCGGTT	7.99×10^{-11}
50	AACCGAAACCGGTT	6.95×10^{-11}
3 Non-wild-type	CACCGAAATCGGTG	8.04×10^{-11}
Mutant	ACACGAAATCGTGT	2×10^{-9}

Constants are the average values of at least three experiments (two for the mutant oligonucleotide).

HPV-16 E2 protein encompassing the DNA binding/dimerisation domain. The expressed product accumulated to ~10% of the total cell protein, was entirely soluble at 25°C, and was purified to homogeneity in essentially two non-group-specific chromatography steps. The identity of the protein was verified indirectly by nucleic acid sequencing and directly by amino acid sequencing of the purified product. Purified recombinant E2 C-terminus was shown to retain high biological activity; a value of ~80% activity at the level of DNA binding was measured. That this value represents an upper or lower limit for binding activity cannot be known, since precise quantitation requires an exact measurement of protein concentration, which is difficult to determine without direct calibration of a photometric or colourimetric assay.

As a preliminary characterisation of the HPV-16 E2 protein, the apparent equilibrium constants (K_{eq}) for the interaction of protein with native and non-native consensus binding sites were determined using a quantitative electrophoretic mobility shift assay. The values given in Table 1 ranged over less than four-fold from 7.66×10^{-11} M to 2.8×10^{-10} M for native sites. For core motifs assayed in the context of common flanking sequences, measured K_{eq}s were of the same order of magnitude, $6.95-8.278\times10^{-11}$ M, but fell within a tighter range. Since the range of values is small, it is probably not appropriate to rank sites in terms of binding affinity. In addition, absolute affinity may not be the critical determinant in the E2–DNA interaction. It is noteworthy that these values are intermediate between those reported for the full length BPV-1 protein (2×10^{-11} M) and the isolated DNA binding domain (C-terminal 86 amino acids; 4×10^{-10} M) (34). While it may be significant that the BPV-1 proteins were produced and purified from two different systems, baculovirus and *E.coli* respectively, alternatively it may also reflect genuine differences in binding affinity due to the presence or absence of the conserved N-terminal domain. Whether the full-length HPV-16 E2 protein may display a similar increase in affinity over the isolated DNA binding domain remains unresolved. Overall, the most salient point to emerge from this analysis is that HPV-16 E2 protein has a high affinity for its cognate binding sites, and saturation of E2 binding sites in an *in vivo* situation could occur when the steady-state levels of E2 in the cell are relatively low.

The apparent equilibrium constant for a protein–DNA interaction is the ratio of the rate constants for the forward and reverse reactions. Despite having similar apparent K_{eq} values, the binding sites may differ appreciably in rates of association with and dissociation from E2 protein. The dissociation rate of a protein–DNA complex is inversely proportional to its half-life, which in turn is a measure of the stability of the interaction. When the half-lives of the various E2 binding site complexes were determined, they varied over a greater than 100-fold range, the least stable interaction being for binding site 2, the only motif that does not contain the extended palindrome

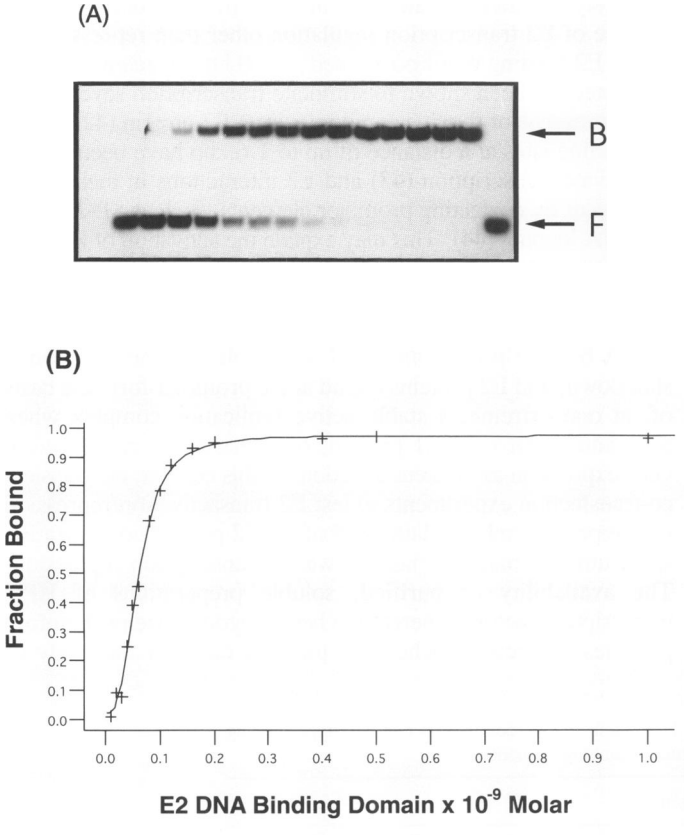

Figure 4. Determination of the apparent equilibrium constant for the interaction of the HPV-16 E2 DNA binding domain with a core binding site. (**A**) Titration of ^{32}P-labelled 30 bp oligonucleotide 50 containing consensus site 4. The oligonucleotide concentration was fixed at 2×10^{-11} M and titrated in a 10 μl reaction volume against a standard protein dilution series ranging from 1×10^{-11} to 1×10^{-9} M final. The products were resolved on an 8% polyacrylamide gel and quantified as described in Materials and Methods. F, free probe; B, protein–DNA complex. (**B**) Numerical data from quantified autoradiographs was used to construct a standard binding curve. The protein concentration at half saturation was determined from this graph to be 6×10^{-11} M, which corresponds to the apparent equilibrium constant K_{eq}.

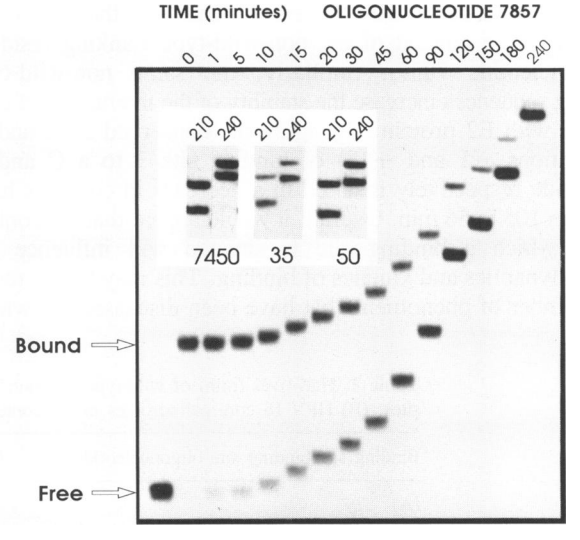

Figure 5. Off-rate analysis for E2 binding oligonucleotides. Labelled binding site oligonucleotide was incubated with protein as described in Materials and Methods and the decay of the labelled complexes was analysed as a function of time on a continuously run polyacrylamide gel, following the addition of a vast excess of unlabelled competitor. The autoradiograph shows the entire time course for a typical experiment where oligonucleotide 7857 was under analysis. The elapsed time in minutes is shown above each lane and the half-life was determined in this case to be 49 min. The inset panel shows the distribution of bound and free probe at T = 210 and 240 min for oligonucleotides 7455, 35 and 50.

ACCG(N)$_4$CGGT. There was a perfect correlation between the ranking order of sites based on affinity compared to stability, although the range over which the half-life values spanned was much greater. It should be noted that for binding sites 3 and 4 minor base alterations from the wild-type were made in flanking nucleotides (5′ to site 3 and 3′ to site 4), well outside the conserved region, in order to minimise the identity to a second half site within the same oligonucleotide, a natural constraint imposed by the existence of these motifs in an adjacent tandem array in the HPV-16 DNA (see Figure 1). We have observed the appearance of complexes of higher mobility when consensus binding sites are assayed in this system, which may be mediated by oligonucleotide sequences resembling imperfect half sites, and may indicate that the kinetics of binding to the core site could be altered. Indeed, *bona fide* E2 binding sites, that contain only one perfect half site, have been characterised in the BPV-1 enhancer. However, we consider these minor changes to be insignificant, since when oligonucleotide site 3 is modified exclusively by polynucleotide kinase to incorporate a ^{32}P label into the molecule we observe the same dissociation kinetics as when the oligonucleotide is labelled and rendered completely double-stranded by Klenow fill-in. The half-lives of 37.3, 105 and greater than 240 min for the HPV E2 C-terminus are considerably larger than those for the BPV-1 E2 protein (N-terminal truncated form, 287 amino acids) interacting with native enhancer elements, where values ranging from 17 min (ACCG(N)$_4$CGGT motif) to less than 30 s were recorded (15). The data presented here may reflect a similar tendency in the HPV-16 system where a stabilising influence due to the conserved residues of the extended palindromic motif may provide additional contacts and contribute to a complex of lower free energy. A similar effect may also result from the external flanking residues, which were shown here to contribute directly to stability. Most notably, the 12 bp E2 consensus from binding site 2 forms a complex with a half-life 20 times less than that of the same motif assayed in the context of the non-wild-type flanking residues (oligonucleotide 7857). Similarly, the same non-wild-type flanking sequences increase the stability of the interaction of core motif 4 with E2 protein. Changing the conserved 5′ A and 3′ T (positions −1 and +13) of binding site 3 to a C and G nucleotide respectively resulted in a decrease in complex half-life from 105 to 66 min. Overall, it would appear that the context within which a binding site is situated will influence the thermodynamics and kinetics of binding. This may be the result of a number of phenomena that have been discussed elsewhere (15): non-conserved residues may provide chance stabilising contacts, hinder or diminish specific contacts, effect the local hydration of the target binding site, or locally contribute to stretches of DNA with special structure or flexibility that favours binding.

In the bovine papillomavirus model the full-length E2 protein is absolutely required for transcription and has been implicated in the regulation of at least four viral promoters. Specific regulation is thought to occur by virtue of the range of affinities displayed by the 17 sites (300-fold range; 15), and the antagonising activities of the truncated E2 repressors. The role of the E2 protein in transcription and replication is less clearly defined in HPV-16 and other genital HPVs. Since HPV enhancers display constitutive cell type-dependent activity in the absence of E2, it has been suggested that E2 binding may repress transcription by interacting with promoter proximal sequences that are adjacent to, or overlap with, binding sites for SP-1 and the TATA binding protein (14,36). Given the high affinities and stabilities of the E2−DNA complexes we have characterised, it is possible that the E2 protein could interact with binding sites 1 and 2, while sites 3 and 4 are occupied by cellular components of the transcription machinery, even when viral E2 is at a relatively low steady-state level in the cell, thus contributing to a mode of E2 transcription regulation other than repression. A single E2 binding motif positioned ~100 bp upstream of a core promoter has been shown to stimulate transcription several fold in the presence of (bovine papillomavirus) E2 protein (42). Also, E2 binding sites at a distance of up to 1.6 kbp have been shown to activate transcription (43) and E2 interactions in these cases may occur by contacting promoter elements, such as SP-1 or E2, by DNA looping (44). This may explain the activation of HPV-16 promoter P97 in the presence of HPV-16 E2 observed by others in *in vitro* transient transfection assays (45). As the levels of E2 increase in the cell, a critical point may be reached where the TATA binding protein and SP-1 are displaced, transcription is shut down, and E2 proteins bound at the promoter form the basis of, at one extreme, a stable active replication complex when associated with viral E1 protein, or alternatively repress viral gene expression as in latent infection. In this context, the transient co-transfection experiments to test E2 transactivation/repression may represent only the latter end of the E2 protein concentration spectrum and may be biased towards transcription repression. The availability of purified, soluble preparations of HPV transcription factors generated in heterologous systems therefore provides a precise biochemical tool not only for the study of

Table 2. Half-lives (min) of wild-type and non-wild-type E2 protein−DNA complexes: (A) wild-type binding sites; (B) HPV-16 core palindromes in the context of common flanking sequences

Binding site/binding site oligonucleotide	Core sequence	Half-life (min)
(A)		
Site 1	AACCGAATTCGGTT	>>240
Site 2	AACCGTTTTGGGTT	2.15
Site 3 (oligo 35)	AACCGAAATCGGTT	105
Site 4	AACCGAAACCGGTT	37.3
(B)		
7450	AACCGAATTCGGTT	>240
7857	AACCGTTTTGGGTT	47
35 (site 3)	AACCGAAATCGGTT	105
50	AACCGAAACCGGTT	178
3 Non-wild-type	CACCGAAATCGGTG	46
Mutant	ACACGAAATCGTGT	<<0.5

DNA–protein interaction kinetics, but also structural and transcription studies.

ACKNOWLEDGEMENTS

We would like to thank David Roper and James Brannigan in the Department of Chemistry at the University of York for fast and efficient oligonucleotide synthesis, and Peter Crosby and Meg Stark for photography. Peter Moody and Paula Daniels are thanked for helpful discussions. C.M.S. was in receipt of a Medical Research Council Studentship, and the research was supported by a program grant to N.J.M. from the Yorkshire Cancer Research Campaign.

REFERENCES

1. Lancaster,W.D. and Olson,C. (1982) J. Microbial. Rev., **40**, 191–207.
2. de Villiers,E.M. (1989) J. Virol., **63**, 4898–4903.
3. zur Hausen,H. and Schneider,A. (1987) In Salzman,N. and Howley,P.M. (eds), The Papovaviridae, Vol 2. Plenum Press, New York, pp. 245–265.
4. Dyson,N., Howley,P.M., Munger,K. and Harlow,E. (1989) Science, **243**, 934–937.
5. Lechner,M.S., Mack,D.H., Finicle,A.B., Crook,T., Vousden,K.H. and Laimins,L.A. (1992) EMBO J., **11**, 3045–3052.
6. Munger,K., Werness,B.A., Dyson,N., Phelps,W.C., Harlow,E. and Howley,P.M. (1989) EMBO J., **8**, 4099–4105.
7. Schwarz,E., Freese,U.K., Gissmann,L., Mayer,W., Roggenbuck,B., Stremlau,A. and zur Hausen,H. (1985) Nature, **314**, 111–113.
8. Schneider-Gaedicke,A. and Schwarz,E. (1986) EMBO J., **5**, 2285–2292.
9. Baker,C.C., Phelps,W.C., Lindgren,V., Braun,M.J., Gonda,M.A. and Howley,P.M. (1987) J. Virol., **61**, 962–971.
10. Sarver,N., Rabson,M.S., Yang,Y., Byrne,J.C. and Howley,P.M. (1984) J. Virol., **52**, 377–388.
11. DiMaio,D. (1986) J. Virol., **57**, 475–480.
12. Androphy,E.J., Lowy,D.R. and Schiller,J.T. (1987) Nature, **325**, 70–73.
13. Moskaluk,C.A. and Bastia,D. (1987) Proc. Natl. Acad. Sci. USA, **84**, 1215–1218.
14. Dartmann,K., Schwarz,E., Gissmann,L. and zur Hausen,H. (1986) Virology, **151**, 124–130.
15. Li,R., Knight,J., Bream,G., Stenlund,A. and Botchan,M. (1989) Genes Dev., **3**, 510–526.
16. Harrison,N.M., Gearing,K.L., Kim,S., Kingsman,A.J. and Kingsman,S.M. (1987) Nucleic Acids Res., **15**, 10267–10283.
17. Haugen,T.H., Cripe,T.P., Ginder,G.D., Karin,M. and Turek,L.P. (1987) EMBO J., **6**, 145–152.
18. Haugen,T.H., Turek,L.P., Mercurio,F.M., Cripe,T.P., Olson,B.J., Anderson,R.D., Seidl,D., Karin,M. and Schiller,J. (1988) EMBO J., **7**, 4245–4253.
19. Hawley-Nelson,P., Androphy,E.J., Lowy,D.R. and Schiller,J.T. (1988) EMBO J., **7**, 525–531.
20. Morissey,L., Barsoum,J. and Androphy,E.J. (1989) J. Virol., **63**, 4422–4425.
21. Spalholz,B.A., Yang,Y. and Howley,P.M. (1985) Cell, **42**, 183–191.
22. Ustav,M. and Stenlund,A. (1991) EMBO J., **10**, 449–457.
23. Blitz,I.L. and Laimins,L.A. (1991) J. Virol., **65**, 649–656.
24. Choe,J., Vaillancourt,P., Stenlund,A. and Botchan,M. (1989) J. Virol., **63**, 1743–1755.
25. Hubbert,N.L., Schiller,J.T., Lowy,D.R. and Androphy,E.J. (1988) Proc. Natl. Acad. Sci. USA, **85**, 5864–5868.
26. Lambert,P.F., Spalholz,B.A. and Howley,P.M. (1987) Cell, **50**, 69–78.
27. McBride,A.A. and Schlegel,R. and Howley,P.M. (1988) EMBO J., **7**, 533–539.
28. McBride,A.A., Bolen,J.B. and Howley,P.M. (1989) J. Virol., **63**, 5076–5085.
29. McBride,A.A., Byrne,J.C. and Howley,P.M. (1989) Proc. Natl. Acad. Sci. USA, **86**, 510–514.
30. Prakash,S.S., Grossman,S.R., Pepinsky,R.B., Laimins,L.A. and Androphy,E.J. (1992) Genes Dev., **6**, 105–116.
31. Giri,I. and Yaniv,M. (1988) EMBO J., **7**, 2823–2829.
32. Hegde,R.S., Grosmann,S.R., Laimins,L.A. and Sigler,P.B. (1992) Nature, **359**, 505–512.
33. McBride,A.A., Klausner,R.D. and Howley,P.M. (1992) Proc. Natl. Acad. Sci. USA, **89**, 7531–7535.
34. Monini,P., Grossman,S.R., Pepinsky,B., Androphy,E.J. and Laimins,L.A. (1991) J. Virol., **65**, 2124–2130.
35. Dostatni,N., Thierry,F. and Yaniv,M. (1988) EMBO J., **7**, 3807–3816.
36. Thierry,F., Heard,J.M., Dartmann,K. and Yaniv,M. (1987) J. Virol., **61**, 134–142.
37. Gloss,B. and Bernard,H. (1990) J. Virol., **64**, 5577–5584
38. Dostatni,N., Lambert,P.F., Sousa,R., Ham,J., Howley,P.M. and Yaniv,M. (1991) Genes Dev., **5**, 1657–1671.
39. Remm,M., Brain,R. and Jenkins,J.R. (1992) Nucleic Acids Res., **20**, 6015–6021.
40. Bedrosian,C.L. and Bastia,D. (1990) Virology, **174**, 557–575.
41. Garabedian,M.J. (1993) In Hanes,B.D. and Higgins,S.J. (eds), Gene Transcription, A Practical Approach.
42. Ham,J., Dostatni,N., Arnos,F. and Yaniv,M. (1991) EMBO J., **10**, 2931–2940.
43. Thierry,F., Dostatni,N., Arnos,F. and Yaniv,M. (1990) Mol. Cell. Biol., **10**, 4431–4437.
44. Li,R., Knight,J.D., Jackson,S.P., Tjian,R. and Botchan,M.R. (1991) Cell, **65**, 493–505.
45. Cripe,T.P., Haugen,T.H., Turk,J.P., Tabatabai,R., Schmid,P.G., Durst,M., Gissmann,L., Roman,A. and Turek,L.P. (1987) EMBO J., **6**, 3745–3753.