

# Comparison of the Structure and DNA-binding Properties of the E2 Proteins from an Oncogenic and a Non-oncogenic Human Papillomavirus

Ghislaine Dell, Kay W. Wilkinson, Rebecca Tranter, Joanna Parish  
R. Leo Brady and Kevin Gaston\*

Department of Biochemistry  
School of Medical Sciences  
University of Bristol  
University Walk  
Bristol BS8 1TD, UK

Human papillomaviruses (HPVs) that infect the genital tract can be divided into two groups: high-risk HPV types, such as HPV 16 and HPV 18, are associated with cancer, low-risk HPV types, such as HPV 6, are associated with benign warts. In both high-risk and low-risk HPV types, the papillomavirus E2 protein binds to four sites within the viral long control region (LCR) and regulates viral gene expression. Here, we present the crystal structure of the minimal DNA-binding domain (DBD) from the HPV 6 E2 protein. We show that the HPV 6 E2 DBD is structurally more similar to the HPV 18 and bovine papillomavirus type 1 (BPV1) E2 proteins than it is to the HPV 16 E2 protein. Using gel retardation assays, we show that the hierarchy of E2 sites within the HPV 16 and HPV 6 LCRs are different. However, despite these differences in structure and site preference, both the HPV 16 and 6 E2 DBDs recognise an extended version of the consensus E2 binding site derived from studies of the BPV1 E2 protein. In both cases, the preferred binding site is 5' **AACCGN<sub>4</sub>CGGTT** 3', where the additional flanking base-pairs are in bold and N<sub>4</sub> represents a four base-pair central spacer. Both of these HPV proteins bind preferentially to E2 sites that contain an A:T-rich central spacer. We show that the preference for an A:T-rich central spacer is due, at least in part, to the need to adopt a DNA conformation that facilitates protein contacts with the flanking base-pairs.

© 2003 Elsevier Ltd. All rights reserved.

**Keywords:** papillomavirus; E2 protein; DNA binding; sequence specificity; DNA conformation

\*Corresponding author

## Introduction

Papillomaviruses are DNA tumour viruses that infect a variety of vertebrate species. There are over 100 types of human papillomavirus (HPV), many of which infect the genital tract. The genital HPVs can be subdivided into two groups. Low-risk HPV types, such as HPV 6, cause benign

warts. In contrast, high-risk HPV types, such as HPV 16 and HPV 18, are associated with cervical cancer.<sup>1</sup> The papillomavirus E2 open reading frame (ORF) encodes a DNA-binding protein that regulates transcription of the viral genome and is required for efficient viral replication.<sup>2</sup> In cervical carcinoma cells, HPV DNA is often found integrated into the host genome. Viral integration is not thought to play a role in the normal life-cycle of HPV. However, integration usually results in the loss of E2 expression, and this de-regulates expression of the viral E6 and E7 oncoproteins, leading to cell transformation, and ultimately, to cancer.<sup>1</sup> The present study aims to shed light on the biological functions of E2 by comparing the DNA-binding properties of the E2 proteins from a high-risk and a low-risk HPV.

The recognition of specific DNA sequences by

Present address: J. Parish, University of Massachusetts Medical School, 364 Plantation Street, Worcester, MA 01605, USA.

Abbreviations used: HPV, human papillomavirus; ORF, open reading frame; DBD, DNA-binding domain; BPV1, bovine papillomavirus type 1; LCR, long control region.

E-mail address of the corresponding author:  
kevin.gaston@bristol.ac.uk

DNA-binding proteins is often a key event in the regulation of gene expression and the control of DNA replication. The E2 proteins represent an excellent model system in which to study protein–DNA interactions and their consequences.<sup>2</sup> The E2 protein is a dimer with three distinct domains: an N-terminal transcription activation/DNA replication domain, a central region thought to act as a flexible “hinge”, and a C-terminal DNA-binding domain (DBD).<sup>3</sup> Although the structure of the full-length E2 protein has yet to be determined, X-ray crystallography has revealed that the N-terminal domain of the HPV 16 E2 protein is made up of three  $\alpha$ -helices that form a twisted plane and a region that is predominantly anti-parallel  $\beta$ -sheet.<sup>4,5</sup> Conserved amino acid residues in the twisted plane mediate dimerisation of the N-terminal domain. Structural studies have shown that the minimal DBDs from the HPV 16, HPV 18, and HPV 31 E2 proteins and from the related bovine papillomavirus type 1 (BPV1) E2 protein, are also dimeric.<sup>6–9</sup> The two interacting polypeptides that make up the DBD form an anti-parallel  $\beta$ -barrel structure, with two pairs of surface  $\alpha$ -helices, one pair of which are DNA recognition helices.<sup>9</sup> Amino acid sequences from immediately upstream of the DBD can stabilise the protein–DNA interaction, but do not appear to influence sequence-specificity.<sup>10</sup> Detailed structural information on the E2 proteins from low-risk E2 proteins is lacking.

The E2 proteins bind to inverted repeats that conform to the consensus sequence 5' ACCGN<sub>4</sub> CCGT 3', where N<sub>4</sub> represents a 4 bp central spacer that can be any sequence.<sup>11,12</sup> Biochemical, genetic, and X-ray crystallographic studies have shown that the DNA recognition helices make contacts with two successive major grooves of the DNA.<sup>9</sup> Although the base-pairs in the central spacer sequence are not contacted by the protein, their nature does affect protein binding. The HPV 16 and HPV 18 E2 DBDs appear to bind preferentially to sites with A:T-rich central spacer regions, presumably because these sequences adopt a conformation that facilitates the protein–DNA interaction.<sup>12</sup> Somewhat surprisingly, the BPV1 E2 DBD does not share this preference for an A:T-rich spacer.<sup>12</sup> Several other differences between the HPV and BPV1 DBDs have been noted. For instance, the HPV DBD has a much less electro-positive surface than the BPV1 protein and appears to make fewer contacts with the DNA-phosphate backbone.<sup>2</sup> Furthermore, the HPV 16 E2 protein binds preferentially to an extended version of the consensus E2 binding site: 5' **A**ACCGN<sub>4</sub>**CGGTT** 3', where the additional base-pairs are in bold.<sup>13</sup> These differences have been taken as evidence to suggest that the BPV1 DBD may represent a low-risk E2 protein and that differences in DNA-binding activity may have a bearing on carcinogenesis. However, direct comparisons of low-risk and high-risk E2 DBDs are lacking.

The papillomavirus genome is a circular mole-

cule of around 8 kb that contains several ORFs arranged on one strand of the DNA.<sup>1</sup> Transcription of these ORFs is regulated by the E2 protein and by cellular transcription factors. The binding sites for these proteins lie mostly between the start of the E6 ORF and the end of the L1 ORF, a region known as the long control region (LCR). This region encompasses the viral origin of replication. The LCRs of all genital HPV types contain four E2 binding sites in conserved positions relative to the start of the E6 ORF.<sup>14</sup> The precise role of E2 in the regulation of the LCR activity is difficult to determine. However, the E2 proteins from genital HPV types are generally thought to activate viral transcription when they are at low concentrations and repress viral transcription when they are at higher concentrations.<sup>15–17</sup> During the course of an HPV infection, the concentration of E2 in the cell is thought to vary dramatically and this could influence the occupancy of the E2 binding sites on the HPV genome and thereby regulate HPV gene expression. The order in which E2 binds to its sites within the LCR might thus be important in the HPV life-cycle, and several attempts have been made to determine the affinity of E2 for these sites in the genital HPVs.<sup>13,18,19</sup> However, different fragments of E2 and different methodologies have been used in these studies, and little consensus exists as to the hierarchy of the sites, or the similarities and differences between high-risk and low-risk HPVs. It is important to point out that, unlike the BPV1 E2 DBD, the full-length BPV1 E2 protein binds co-operatively to DNA fragments with multiple E2 sites and forms DNA loops.<sup>20,21</sup> This suggests that the N-terminal domains of DNA-bound BPV E2 dimers and, by extension, HPV E2 dimers, interact, and that the order in which N-terminally truncated E2 proteins bind to these sites *in vitro* may not reflect what happens *in vivo*.

Here, we present the crystal structure of the DBD from the low-risk HPV 6 E2 protein and we compare this protein to the DBDs from the high-risk HPV 16 and 18 E2 proteins and to the BPV1 E2 protein. In addition, using the minimal DBDs from the HPV 16 and HPV 6 E2 proteins, we show that both of these proteins recognise the extended version of the consensus E2 binding site, and that both of these proteins bind preferentially to sites that contain an A:T-rich spacer sequence. We show that the hierarchy of the E2 binding sites within the HPV 16 and HPV 6 LCRs are different, and we discuss our results in terms of models that seek to explain how changes in E2 concentration modulate LCR activity.

## Results

To enable a detailed comparison of the minimal DBDs from the HPV 16 and HPV 6 E2 proteins, we expressed these proteins in bacteria and purified them to homogeneity (data not shown). The

structure of the HPV 6 DBD was determined at 1.9 Å resolution by X-ray crystallography (Figure 1 and Table 1). As might be expected on the basis of their high degree of sequence homology, the HPV 6 DBD is similar to the previously studied HPV 16, HPV 18, and BPV1 DBDs in terms of its overall fold (rms 1.034 Å for the C $\alpha$  atoms for monomers of the HPV 6 E2 and HPV 16 E2 DBDs) (Figure 1(c)). The HPV 6 E2 DBD is a dimer in which each monomer supplies four anti-parallel  $\beta$  strands that together form an eight-stranded  $\beta$  barrel. Each monomer also carries two surface-exposed  $\alpha$  helices. Comparison to the BPV1 E2–DNA complex suggests that the  $\alpha$ 1 helices from each subunit are DNA recognition helices that interact with the exposed edges of the base-pairs in two successive major grooves within the E2 binding site.<sup>9</sup>

On the basis of alignments of the DBDs from BPV1, HPV 16, HPV 18, and HPV 31, it has been suggested that these proteins fall into two distinct classes.<sup>7</sup> When one subunit of each DBD is superimposed, the non-superimposed recognition helices of the HPV 18 and BPV1 proteins and the HPV 16 and HPV 31 proteins are displaced by around 7 Å.<sup>7</sup> Superimposition of one subunit of the HPV 6 DBD with one subunit from the BPV1 and HPV 16 DBDs shows that the non-superimposed subunit of the HPV 6 protein aligns with the BPV1 E2 protein and is displaced by around 7 Å from the HPV 16 E2 protein (Figure 1(c)). This suggests that the HPV 6 E2 DBD is most structurally related to the BPV1 and HPV 18 E2 proteins in terms of overall fold.

Interestingly, the loop between strands  $\beta$ 2 and  $\beta$ 3 is ordered in our HPV 6 E2 structure (Figure 1(c) and (d)), whereas the equivalent loop is disordered in every E2 structure without DNA (HPV 18, 16, and BPV1).<sup>2</sup> In the BPV1 E2–DNA complex, the  $\beta$ 2/ $\beta$ 3 loop assumes an ordered conformation and forms side-chain contacts with the phosphate groups at positions  $-3$  and  $-4$ , and positions  $+3$  and  $+4$  in the E2 binding site (the nomenclature refers to the distance from the axis of symmetry).<sup>2</sup> Ordering of the  $\beta$ 2/ $\beta$ 3 loop is not seen in the HPV 18 E2–DNA complex. In the HPV 6 DBD, this loop is smaller than that in the BPV1 protein and it is oriented differently (Figure 1(c) and (d)). The HPV 6  $\beta$ 2/ $\beta$ 3 loop conformation appears to be maintained primarily by a network of charged interactions between Lys323 in the loop, and Asp311 and His336 in the other molecule in the dimer (Figure 1(d)). All other contacts with the loop appear to be weak, water-mediated interactions likely to be present only within the crystal lattice. The small size of the HPV 6  $\beta$ 2/ $\beta$ 3 loop suggests that it is unlikely to make the same contacts with the phosphate backbone observed in the BPV1 complex unless major structural changes occur on binding to DNA. However, the HPV 6  $\beta$ 2/ $\beta$ 3 loop contains two proline residues (Pro322 and Pro325) that are likely to restrain its conformation, arguing against a structural rearrangement on binding DNA. In the BPV1 E2 DBD–DNA com-

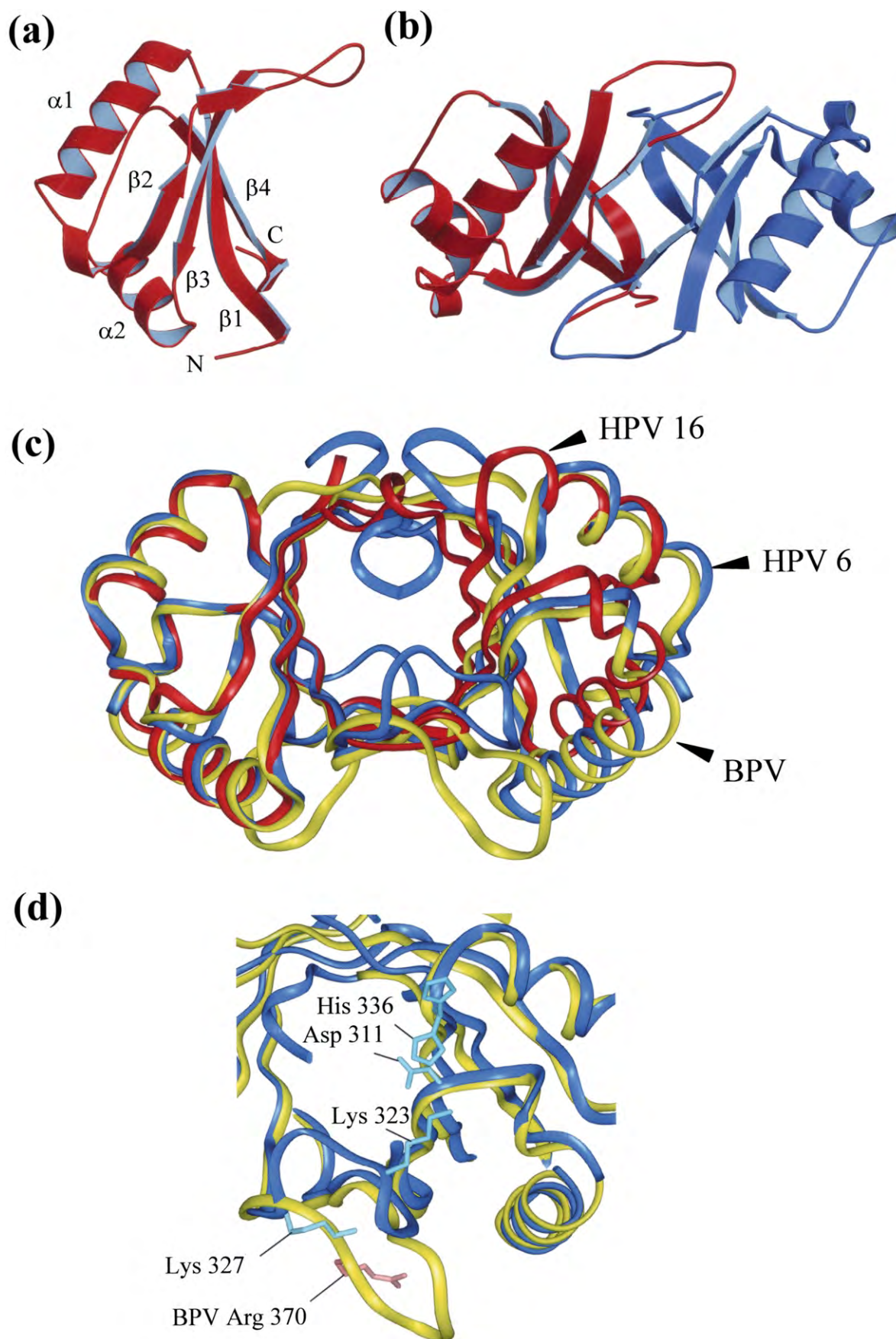
plex, arginine side-chains in the  $\beta$ 2/ $\beta$ 3 loop (Arg370) contact the DNA. There are no arginine residues in the HPV 6  $\beta$ 2/ $\beta$ 3 loop, but there are two lysine residues. One of these (Lys323) is bound in the salt-bridge described above. The other (Lys327) could conceivably contact DNA in the HPV 6 E2–DNA complex if the loop underwent a major rearrangement (Figure 1(d)). However, mutagenesis of this amino acid to alanine has little or no effect on the DNA-binding activity of this protein (data not shown).

Comparison of the electrostatic potential energy surfaces of the HPV and BPV1 DBDs has revealed interesting differences.<sup>2</sup> The DNA-binding surface of the BPV1 DBD is more positively charged than the equivalent surface of the HPV 16 DBD (Figure 2(a) and (b), respectively). The DNA-binding surface of the HPV 18 DBD is also more positively charged than that of the HPV 16 DBD, and in this case the positive charge is highest midway between the DNA recognition helices and opposite the minor groove of the E2 binding site when the protein is bound to DNA (Figure 2(c)). This comparison suggested that differences in the DNA-binding properties of these proteins might result from their different distribution of surface charge.<sup>2</sup> Interestingly, the charge distribution on the HPV 6 DBD DNA-binding surface most closely resembles that of the HPV 16 DBD and not those of the HPV 18 or BPV E2 DBDs (Figure 2(d)).

### DNA-binding activity

To compare the DNA-binding activity of the HPV 16 and HPV 6 E2 proteins, increasing amounts of each minimal DBD were added to labelled oligonucleotides carrying an idealised, symmetric E2 binding site. After 20 minutes at 20 °C, free and bound labelled DNA were separated by polyacrylamide gel electrophoresis and visualised using a PhosphorImager (Figure 3(a)). Previous experiments have shown that in the absence of competitor DNA this is sufficient time to allow equilibrium to be reached. The amount of free and bound DNA was determined using a PhosphorImager and binding curves fit to the data (Figure 3(b) and (c)). The resulting binding curves describe the data within experimental error and were used to calculate the apparent equilibrium constant for each DBD (Table 2). As can be seen from the data, the HPV 16 and HPV 6 DBDs bind to this E2 binding site with an apparent equilibrium constant ( $K_{eq(app)}$ ) of around 3 nM (Table 2, line 1). These results are in agreement with values reported by Hegde and co-workers for the binding of the HPV 16 DBD to a similar idealised E2 binding site.<sup>12</sup> However, they differ markedly from values reported in an earlier study in which much lower concentrations of salt were used in the binding buffer.<sup>18</sup>

To compare the gross architecture of the DBD–DNA complexes formed by the HPV 16 and HPV



**Figure 1.** The crystal structure of the HPV 6 E2 DNA-binding domain. (a) Topology of the HPV 6 E2 DBD monomer. (b) Topology of the HPV 6 E2 DBD dimer. The subunits are indicated in red and blue. (c) A comparison of the DBDs from the HPV 6 (blue), HPV 16 (red), and BPV1 (yellow) E2 proteins. The left subunit of each protein is superimposed.

**Table 1.** Data collection and refinement statistics

	Native	Xe derivative	Hg derivative	Pt derivative
Resolution range (Å)	20–1.9 (1.97–1.9)	30–2.0 (2.07–2.0)	50–2.80 (2.9–2.8)	30–3.1 (3.24–3.1)
Unique reflections	42,743	37,784	12,595	10,230
Redundancy	3.0 (1.4)	9.2 (3.5)	3.4 (2.1)	2.1 (1.9)
Completeness (%)	93.2 (58.8)	98.2 (86.5)	90.1 (77.0)	97.3 (91.6)
Average $I/\sigma(I)$	28.6 (2.5)	33.5 (2.2)	17.5 (3.8)	12.7 (1.5)
$R_{\text{merge}}$ (%)	4.4 (22.1)	7.0 (46.3)	6.0 (20.5)	8.1 (25.6)
Refined model				
$R_{\text{free}}$ (all data, 20–1.9 Å) (%)	26.28			
$R_{\text{cryst}}$ (all data, 20–1.9 Å) (%)	20.31			
rmsd values				
Bond lengths (Å)	0.019			
Bond angles (deg.)	1.691			

6 proteins, we incubated increasing amounts of each protein with a labelled DNA fragment carrying E2 site 4 from the HPV 16 genome and used DNase I to footprint the resulting complexes (Figure 4(a)). Both proteins protect a similar length of DNA from digestion by DNase I. Furthermore, the concentrations of protein required to obtain a clear footprint are roughly equivalent. Hydroxyl-radical footprinting revealed no significant differences in the DBD–DNA complexes formed by the HPV 16 and HPV 6 DBDs (data not shown). Circular permutation assays using a set of DNA fragments in which the position of the E2 binding site is varied were used to investigate the effects of each DBD on DNA bending. On binding to DNA, both proteins induce DNA bending and there is no significant difference in the DNA bend angles:  $30(\pm 1)^\circ$  for the HPV 16 DBD and  $27(\pm 3)^\circ$  for the HPV 6 DBD (Figure 4(b)). Although these data clearly indicate that the DNA bend angle in the HPV 6 and HPV 16 DBD–DNA complexes are very similar, the difficulties inherent in relating DNA bend angles determined using this approach to bend angles observed in protein–DNA co-crystals, make it difficult to say with any certainty whether these values differ significantly from the DNA bend angle of  $43\text{--}51^\circ$  seen in crystals of the HPV 18 and BPV1 E2 DBD–DNA complexes.<sup>7</sup> Since the uncomplexed HPV 6 DBD superimposes on the HPV 18 and BPV1 DBDs, these data raise the possibility that the 7 Å displacement observed in the structure of the HPV 16 DBD may not be replicated in the HPV 16 E2 DBD–DNA complex. In summary, the data from gel retardation assays and footprinting experiments suggest that the HPV 6 and HPV 16 E2 DBDs bind to a consensus E2 site with similar affinity, and that the protein–DNA complexes have similar architecture.

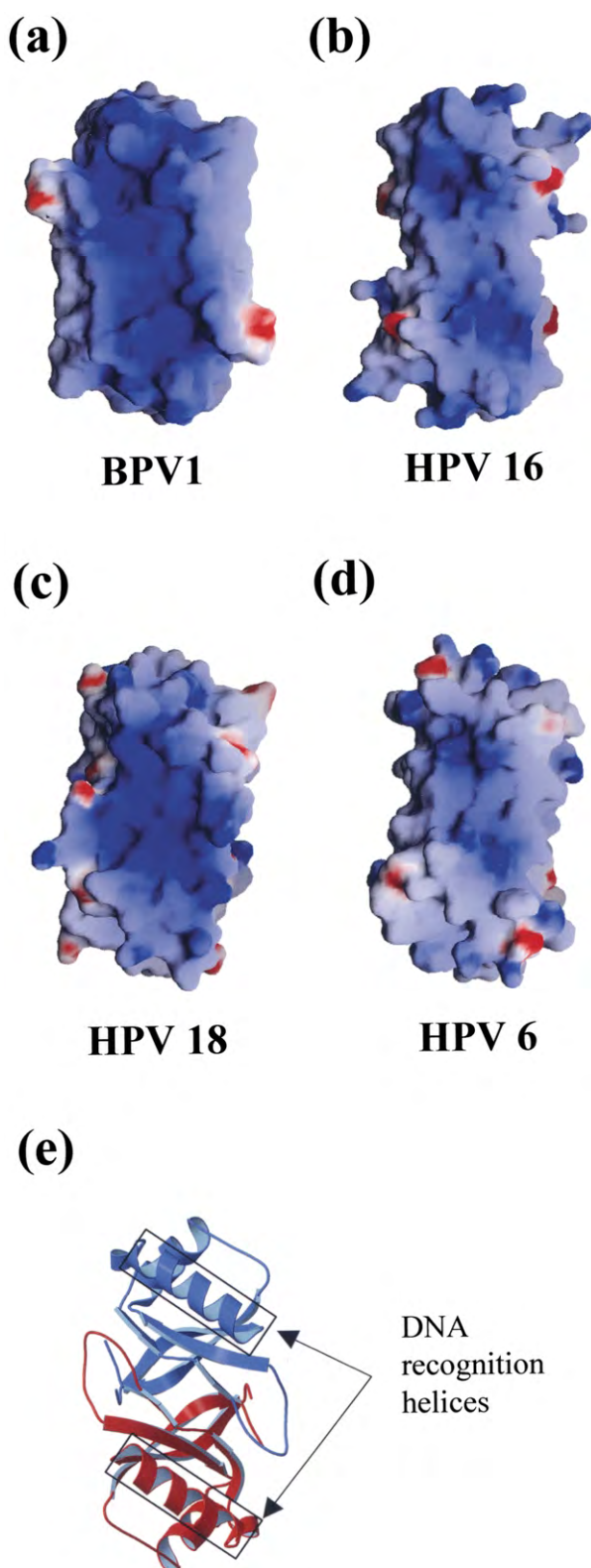
### HPV 6 and HPV 16 E2 bind preferentially to sites with an A:T-rich spacer

Unlike the BPV E2 DBD, the DBD from the HPV 16 and HPV 18 E2 proteins bind preferentially to sites with an A:T-rich central spacer.<sup>7,12</sup> Since the spacer region is not contacted directly by the DBD, this difference in binding specificity must be due to the overall conformation or flexibility afforded by these base-pairs to the surrounding DNA. To compare the binding of the HPV 16 and HPV 6 E2 DBDs to E2 sites with different central base-pairs, we made a series of E2 binding sites with either AATT, ACGT, or CCGG in the spacer region. Table 2 shows the affinity of the HPV 16 and HPV 6 E2 DBD for these sites ( $K_{\text{eq(app)}}$ ). Both proteins bind preferentially to an E2 site with an A:T-rich spacer (Table 2, line 1). Changing the central spacer to either ACGT or CCGG, decreases binding by a factor of around tenfold in the case of the HPV 16 DBD and by more than 1000-fold in the case of the HPV 6 DBD. The values for the HPV 16 E2 DBD differ somewhat from the 30-fold degree of preference reported previously for the binding of this protein to E2 sites with these spacer regions.<sup>12</sup> One possibility is that this could be a consequence of the different types of oligonucleotides used in each study. Whilst we have used pairs of short complementary oligonucleotides, the previous study made use of single long oligonucleotides that anneal intramolecularly to form a hairpin of double-stranded DNA.<sup>12</sup>

### HPV 16 and HPV 6 E2 recognise an extended binding site

An alignment of the four E2 binding sites present within the HPV 16 LCR produces a consensus sequence that differs from that derived from the

(d) A comparison of the  $\beta 2/\beta 3$  loop in the HPV 6 E2 DBD (blue) and BPV1 E2 DBD (yellow). BPV1 E2 Arg370 is indicated in pink. HPV 6 E2 Lys327, Lys323, Asp311 and His366 are indicated in light blue. His366 is present in two conformations in the crystal structure.



**Figure 2.** The electrostatic potential energy surface of the E2 DBD. The electrostatic potential energy surfaces of the DBDs from the (a) BPV1, (b) HPV 16, (c) HPV 18, and (d) HPV 6 E2 proteins were calculated using GRASP. Positive potentials are blue, negative potentials are red. The water-probe radius was 1.4 Å and the electrostatic potential is displayed on a scale of  $-13 k_B T$  to  $+13 k_B T$ . (e) The HPV 6 E2 DBD is shown to illustrate the position of the DNA recognition helices in this view.

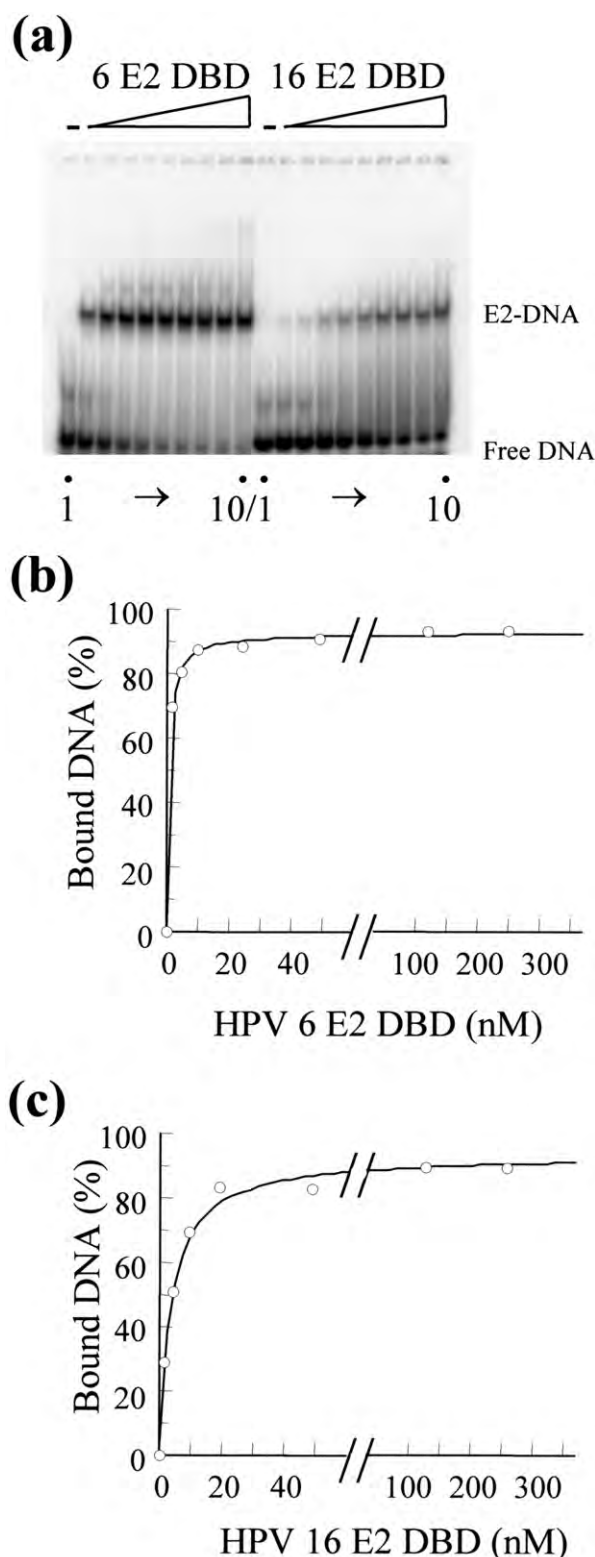
alignment of BPV E2 sites. In previous work we showed that the HPV 16 E2 protein binds preferentially to this extended version of the consensus E2 binding site.<sup>13</sup> We showed that Arg304 in the HPV 16 E2 dimer contacts the base-pairs at the  $-7$  and  $+7$  positions within the E2 binding site, a finding that was later confirmed by X-ray crystallography.<sup>7</sup> To determine whether the HPV 6 DBD shares this preference for an extended binding site, we assayed the binding of this protein to a series of otherwise identical E2 binding sites containing either A:T, T:A, G:C, or C:G, at the  $-7$  and  $+7$  positions ( $\pm 7$ ). Table 3 shows the values for the affinity of these sites relative to the consensus site ( $K_{rel} = K_{eq(app)} \pm 7 \text{ A:T} / K_{eq(app)} \text{ test}$ ). Changing the  $\pm 7$  position from A:T to either T:A, G:C, or C:G reduces binding of the HPV 6 E2 DBD. Table 3 shows the equivalent data for the HPV 16 E2 protein. Comparison of these data show that the HPV 6 E2 DBD has a preference for A:T over G:C at the  $\pm 7$  position, whereas the HPV 16 E2 protein binds equally well to sites that contain A:T or G:C at  $\pm 7$ . Thus, not only does the HPV 6 protein bind preferentially to sites containing an A:T at this position, it appears to be more base-specific at this position when compared to the HPV 16 E2 DBD.

#### The central spacer influences the recognition of flanking DNA

We have shown that the HPV 16 and HPV 6 E2 proteins recognise an extended version of the consensus E2 binding site and that they both bind preferentially to E2 sites with an A:T-rich spacer sequence. We wondered whether the preference for an A:T-rich spacer might be due to a need to correctly position the DNA in order to form the additional protein–DNA contacts at the outer ends of the E2 binding site, far from the bend centre induced during complex formation. To test this hypothesis we made a series of E2 binding sites containing CCGG in the spacer region and either A:T, T:A, G:C, or C:G, at the  $\pm 7$  position. We then tested the binding of the HPV 16 and HPV 6 DBD to each of these sites (Table 3). As might be expected in light of the weak binding of the HPV 6 E2 DBD to E2 sites with a CCGG spacer sequence, this protein binds very poorly to all of these sites. In contrast, the HPV 16 E2 DBD binds to each of these sites. However, although the affinity of HPV 16 E2 DBD for each site is reduced by a factor of around tenfold compared to the  $\pm 7$  A:T site with an AATT spacer, the base-pair at the  $\pm 7$  position has little if any influence on HPV 16 E2 binding in this context. These data suggest that an A:T-rich central spacer facilitates the interaction between E2 and the base-pair at the  $\pm 7$  position.

#### HPV 16 and HPV 6 have a different hierarchy of E2 binding sites

The four E2 binding sites within the HPV LCR



**Figure 3.** The DNA-binding properties of the HPV 16 and HPV 6 E2 proteins. (a) A representative gel retardation experiment. Increasing amounts of the HPV 6 E2 DBD and HPV 16 E2 DBD proteins were incubated with a labelled E2 binding site. Lanes 1–10 contain 0, 2, 5, 10, 25, 50, 125, 250, 500, and 1000 nM protein, respectively. After 20 minutes at 20 °C, free and bound labelled DNA were separated on a 6% polyacrylamide gel and visualised using a PhosphorImager. (b) A representative graph showing the binding of HPV 6 E2 DBD to a

are thought to contribute differently to transcriptional regulation. Figure 5 shows a representation of the LCRs from HPV 16 and HPV 6. The current model for the regulation of HPV 18 and HPV 16 LCR activity proposes that at low E2 concentrations, the high-affinity E2 binding site (E2 site 4) is occupied by E2 and that this results in transcription activation. As E2 levels increase during the course of a viral infection, E2 is proposed to bind to the low-affinity promoter proximal E2 sites. This leads to the displacement of cellular transcription factors and the repression of HPV transcription.<sup>22,23</sup> To compare directly the binding of the HPV 16 and HPV 6 E2 DBD proteins to their binding sites in their respective LCRs, we determined the apparent equilibrium constants for each protein at all of its binding sites. These results are summarised in Figure 5, where  $K_{rel} = K_{eq(app)} E2 \text{ site } 1 / K_{eq(app)} E2 \text{ site } 2, 3, \text{ or } 4$ . As can be seen from the data, the HPV 16 E2 DBD binds tightly to HPV 16 E2 sites 4 and 2, less tightly to E2 site 1, and least tightly to E2 site 3. In contrast, the HPV 6 DBD binds tightly to HPV 6 E2 site 2, and somewhat less tightly to sites 1, 3, and 4. These data suggest that any mechanism for the regulation of LCR activity that is mediated by E2 site affinity is not conserved between HPV 16 and HPV 6.

Finally, it is interesting to consider whether the data obtained using the mutated spacer and flanking sequences described above can help to explain the hierarchy of E2 binding sites seen in the HPV 16 and HPV 6 LCRs. All of the spacer sequences in these sites are predominantly composed of A:T or T:A base-pairs; of the 32 base-pairs within the spacer sequences of these eight E2 binding sites, 30 are A:T or T:A. Three of the HPV 6 E2 binding sites are AAAA or TTTT, whereas only one of the HPV 16 E2 binding sites is TTTT. These poly(A) tracts are known to be relatively inflexible and to adopt a bent DNA conformation.<sup>24–26</sup> This difference in spacer sequences between HPV 6 and HPV 16 may be related to the stronger preference for A:T-rich spacer sequences shown by the HPV 6 E2 DBD (Table 2). The flanking base-pairs at all eight of these E2 binding sites are either A:T (14 out of 16) or G:C (two out of 16). Interestingly, E2 sites 1 and 2 in the HPV 6 LCR are identical, except that site 1 has A:T at position –7, whereas site 2 has G:C. Since the HPV 6 E2 protein binds more

labelled E2 binding site. At each concentration of protein, the amount of free and bound DNA was determined using a PhosphorImager and the percentage of labelled DNA bound:

$$\frac{(\text{bound labelled DNA})}{(\text{bound labelled DNA} + \text{free labelled DNA})} \times 100$$

is plotted against the amount of protein added. (c) A representative graph showing the binding of HPV 16 E2 DBD to a labelled E2 binding site.

**Table 2.** Binding of HPV 16 and HPV 6 DBD to E2 sites with altered spacer sequences

Symmetric E2 site	HPV 16		HPV 6	
	$K_{eq(app)}$ (nM)	$\Delta G$ (kJ/mol)	$K_{eq(app)}$ (nM)	$\Delta G$ (kJ/mol)
CGCAACCG <b>aat</b> tCGGTTGCG	$3.5 \pm 1.1$	$-48 \pm 1$	$2.0 \pm 0.9$	$-49 \pm 2$
CGCAACCG <b>acgt</b> tCGGTTGCG	$24 \pm 14$	$-43 \pm 2$	$>2000$	n.d.
CGCAACCG <b>ccgg</b> tCGGTTGCG	$27 \pm 11$	$-43 \pm 1$	$>2000$	n.d.

The apparent equilibrium constant ( $K_{eq(apparent)}$ ) was obtained using the equation:

$$[\text{bound}_{DNA}] = [\text{maximum bound}_{DNA}][\text{protein}] / ([\text{protein}] + K_{eq(app)})$$

When  $[DNA] \ll K_{eq}$ , the apparent equilibrium constant is equal to the protein concentration at half maximum DNA binding.  $\Delta G = RT \ln K_{eq(app)}$ . The values shown are derived from at least three independent experiments. n.d., not determined.

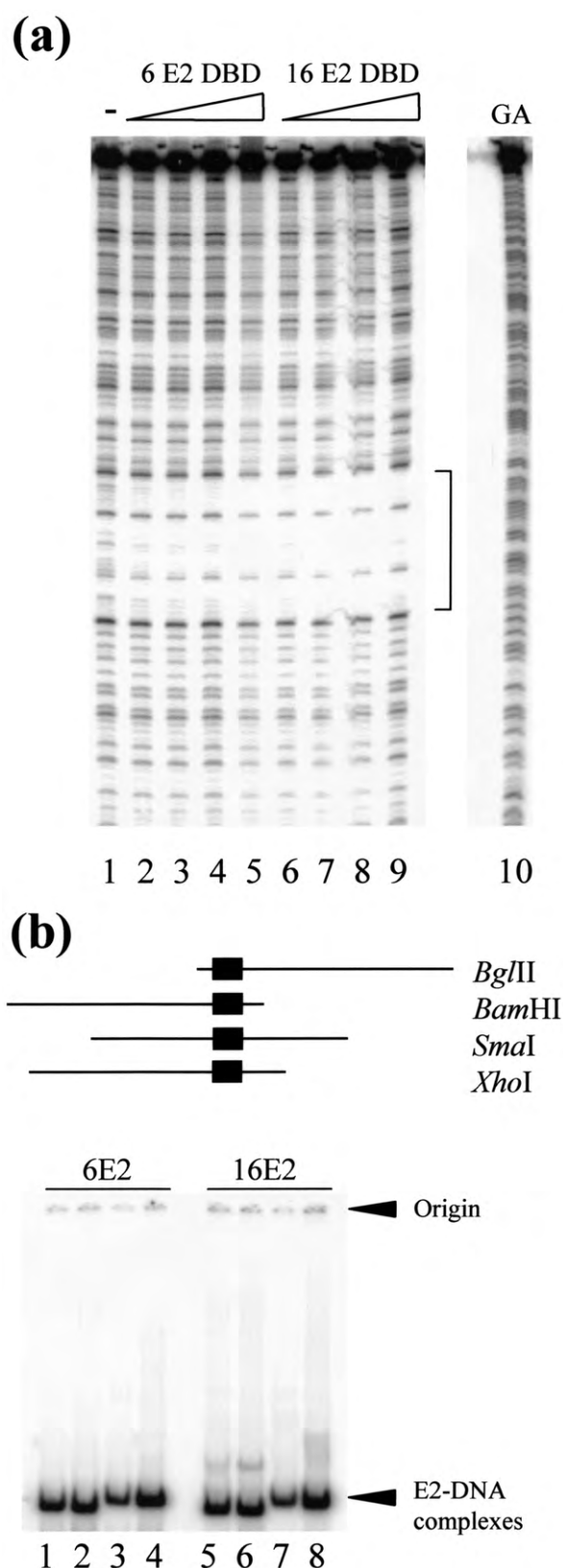
tightly to sites containing A:T at  $-7$  than it does to sites containing a G:C base-pair at this position (Table 3), one might expect HPV 6 E2 to bind better to site 1. However, the HPV 6 E2 protein binds equally well to E2 sites 1, 3, and 4 in the HPV 6 LCR but binds around tenfold better to E2 site 2. It is possible that sequences upstream of  $-7$  or downstream of  $+7$  might be responsible for these differences in binding, or that the full-length HPV 6 E2 protein has somewhat altered specificity compared to the isolated DBD. It is interesting to note that in the HPV 16 LCR all four E2 binding sites contain A:T at the  $\pm 7$  position. This is difficult to explain, since the HPV 16 E2 DBD binds equally well to E2 sites that contain A:T or G:C at this position (Table 3). One possibility is that G:C at  $\pm 7$  might create a binding site for a cellular transcription factor that could compete with E2. A variety of cellular proteins that can bind to E2 sites or to sites that overlap E2 binding sites have been described<sup>27</sup> and this may be an important constraint on the sequence of E2 binding sites *in vivo*. In summary, whilst all of the E2 binding sites within the HPV 16 and HPV 6 LCRs contain A:T-rich spacer sequences and generally contain A:T base-pairs at the  $\pm 7$  position, these features cannot in themselves fully explain the hierarchy of E2 binding sites seen in either case.

## Discussion

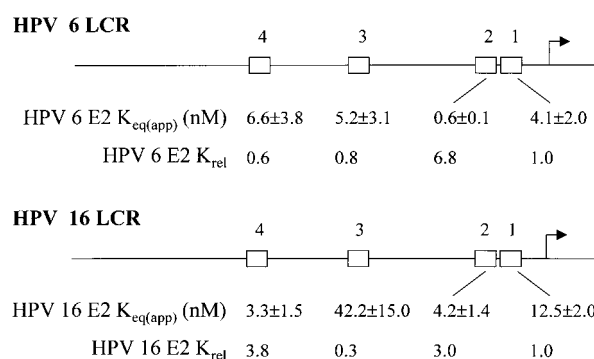
The recognition of specific DNA sequences by DNA-binding proteins is the first step in the initiation of both transcription and DNA replication. The papillomavirus E2 proteins are sequence-specific DNA-binding proteins that regulate transcription of the viral genome and are required for viral DNA replication.<sup>2</sup> The E2 family contains over 100 members and, as yet, only a few of these proteins have been studied in any depth. The structures of the DBDs from several high-risk HPV types and from the BPV1 E2 protein have been determined by X-ray crystallography but until now there has been no structural information available on the equivalent domain from a low-risk HPV type. We have determined the structure of the minimal DBD from the low-risk HPV 6 E2

protein. In terms of its overall fold, the HPV 6 E2 DBD is structurally most similar to the BPV1 and HPV 18 E2 proteins. However, the  $\beta 2/\beta 3$  loop that in BPV1 E2 makes a number of contacts with phosphate groups in the central spacer of the E2 binding site, is smaller in the HPV 6 E2 DBD, and may not be able to form the same contacts with the phosphate backbone. Interestingly, a comparison of the electrostatic potential energy surfaces of the BPV and HPV DBDs reveals that the DNA-binding surface of the HPV 6 is more similar to that of the HPV 16 E2 protein than it is to that of the BPV1 protein. Thus, the low-risk HPV 6 E2 shares some structural features with the BPV E2 protein but also shows significant differences in structure.

The HPV 16 and HPV 6 E2 DBDs are similar in terms of their DNA-binding activity. Both proteins bind preferentially to an extended version of the consensus E2 binding site derived from studies of the BPV1 E2 protein.<sup>11,13</sup> Both of these HPV proteins bind preferentially to DNA sequences that contain an A:T-rich central spacer. A:T-rich sequences can show intrinsic DNA bending and altered DNA flexibility and the ability of some E2 proteins to bind preferentially to these sites suggests that these are important determinants of E2 binding site selection.<sup>12,26</sup> In the E2–DNA complex the DNA is bent through around  $43$ – $51^\circ$  and pre-bent DNA sequences have been shown to facilitate binding of the HPV 16 and HPV 18 E2 DBDs.<sup>7,12</sup> However, the DNA bend angle appears to be similar in both low-affinity and high-affinity E2–DNA complexes.<sup>7,13</sup> We and others have shown that in the case of the HPV 16 and HPV 6 E2 DBD, an A:T-rich central spacer increases protein binding by around tenfold and at least 1000-fold, respectively.<sup>19</sup> Furthermore, we have shown that changes in the central spacer region can influence contacts made between E2 and the base-pairs at the  $+7$  and  $-7$  positions in the E2 site. Preferential binding to pre-bent DNA targets has been documented in several protein–DNA complexes and the increase in binding has often been attributed to a decrease in the entropic cost associated with bending the DNA into a favourable conformation.<sup>28,29</sup> Our data suggest that in the case of the E2 proteins, increased DNA binding to pre-bent DNA sequences may result from both a



**Figure 4.** Architecture of the HPV 16 and HPV 6 E2 DBD-DNA complexes. (a) DNase I footprinting. Increasing amounts of the HPV 6 E2 DBD and HPV 16 E2 DBD proteins were incubated with a labelled DNA fragment containing a single E2 binding site. Lane 1 contains no protein, lane 10 is a G + A marker, lanes 2-5 and lanes 6-9 contain 50, 100, 200, and 400 nM of HPV 6 E2 DBD



**Figure 5.** The hierarchy of E2 binding sites in the HPV 16 and HPV 6 LCR. A representation (not to scale) of the HPV 6 and HPV 16 LCRs showing the arrangement of the E2 binding sites (open boxes) relative to the start of the E6 open reading frame (bent arrow). For each E2 binding site, the apparent equilibrium constant ( $K_{eq(app)}$ ) for the relevant E2 protein was obtained as described in the text. For each LCR and its relevant E2 protein, the relative binding constant  $K_{rel} = K_{eq(app)}$  E2 site 1/ $K_{eq(app)}$  E2 site 2, 3, or 4. The values shown are derived from at least three independent experiments.

decrease in the entropic cost and to additional protein-DNA contacts formed at a distance from the bend centre.

Although the DNA-binding properties of the HPV 16 and HPV 6 E2 proteins are similar, these proteins differ significantly in their ability to activate transcription. Whilst the HPV 16 and 18 E2 proteins can be potent activators of transcription, the HPV 6 and HPV 11 E2 proteins activate transcription poorly.<sup>16,30,31</sup> These differences in transcriptional activity may be related to differences in the activation domains of these proteins and/or to differences in their DNA-binding activities. The full-length HPV 11 E2 protein appears to bind DNA with lower affinity than the full-length HPV 16 E2 protein.<sup>31</sup> However, this conclusion was drawn from the results of experiments using DNA fragments carrying multiple E2 binding sites

and HPV 16 E2 DBD, respectively. After 30 minutes at 30 °C the samples were incubated with DNase I under single-hit conditions. The reaction products were run on a sequencing 8% acrylamide gel and viewed using a PhosphorImager. The bracket indicates the protected region. (b) DNA bending. The upper panel shows the series of circularly permuted DNA fragments used to assay the DNA-bending activity of the HPV 6 and HPV 16 E2 DBDs. The fragments were produced by digestion of pBend3-hE2(4) with *Bgl*II, *Bam*HI, *Sma*I and *Xho*I, respectively. Each fragment contains the E2 binding site (filled rectangle) at a different position relative to the ends of the molecule. The lower panel shows the mobility of the protein-DNA complexes formed when the HPV 6 (lanes 1-4) and the HPV 16 (lanes 5-8) E2 DBD bind to the DNA fragments described above. Bend angles were estimated from three independent experiments. The free DNA was run off the bottom of the gel.

**Table 3.** HPV 16 and HPV 6 DBD bind preferentially to an extended E2 site

E2 site	HPV 6		HPV 16	
	$K_{rel}$	$\Delta\Delta G$ (kJ/mol)	$K_{rel}$	$\Delta\Delta G$ (kJ/mol)
-7                      +7				
CGCAACCGaattCGGTGCG	1	-	1	-
CGCTACCGaattCGGTAGCG	0.22	4.5 ± 2.5	0.11	5.3 ± 0.3
CGCCACCGaattCGGTGGCG	0.27	4.1 ± 1.6	0.28	3.1 ± 0.4
CGCGACCGaattCGGTCGCG	0.32	3.4 ± 1.2	1.07	-0.2 ± 0.4
-7                      +7				
CGCAACCGcgggCGGTGCG	<0.001	n.d.	0.10	5.9 ± 0.5
CGCTACCGcgggCGGTAGCG	<0.001	n.d.	0.09	6.2 ± 0.9
CGCCACCGcgggCGGTGGCG	<0.001	n.d.	0.08	6.6 ± 1.0
CGCGACCGcgggCGGTCGCG	<0.001	n.d.	0.16	4.8 ± 0.5

The relative binding constant  $K_{rel} = K_{eq(app)} \pm 7 A:T/K_{eq(app)}$ . The values shown are derived from at least three independent experiments. n.d., not determined.

and could therefore reflect differences in protein–protein interactions as well as protein–DNA interactions.<sup>31</sup> Here, we have shown the minimal DBDs from the HPV 6 and HPV 16 proteins bind to an E2 site conforming to the consensus E2 site with approximately equal affinity. This suggests that in this case any differences in DNA binding affinity must be due to amino acid sequences outside these domains or to other aspects of E2 function such as cooperativity. However, at E2 sites that differ from the consensus sequence, the HPV 6 E2 DBD is more sensitive to the base composition of the central spacer. Thus it is possible that differences in transcription activation could be due, in part, to this difference in DNA binding specificity.

The LCRs from both HPV 6 and HPV 16 contain four E2 binding sites in conserved positions relative to the transcription start point (Figure 5). Changes in the concentration of E2 that occur as infected cells move through the epithelium have been proposed to lead to the differential occupancy of these E2 sites and, as a consequence, to the modulation of HPV gene expression. The hierarchy of E2 binding sites within the HPV 16 LCR has been examined previously using gel retardation assays with either a GST-E2 fusion protein<sup>13</sup> or the HPV 16 E2 DBD.<sup>18</sup> The hierarchy of E2 binding sites within the HPV 11 LCR have been examined using fluorescence anisotropy with the HPV 11 E2 DBD.<sup>19</sup> Using equivalent protein fragments and identical experimental conditions, we have compared the binding of the HPV 6 and HPV 16 DBD to their respective binding sites. We have shown that the HPV 16 E2 DBD binds tightly to E2 sites 4 and 2, less tightly to E2 site 1, and least tightly to E2 site 3 from the HPV 16 LCR. This hierarchy is in agreement with that determined in previous studies.<sup>13,18</sup> In contrast, the HPV 6 E2 DBD binds tightly to E2 site 2, and only slightly less tightly to E2 sites 1, 3, and 4 from the HPV 6 LCR. These data suggest that in these different virus types the E2 proteins may use different mechanisms to regulate viral gene expression. However, since

the full-length E2 proteins bind co-operatively to DNA fragments with multiple E2 sites,<sup>20,21</sup> it is possible that even at low concentrations E2 occupies all four sites within each LCR simultaneously. Clearly, much further work will be needed to determine whether this is the case *in vivo*.

## Materials and Methods

### Plasmids

The HPV 16 E2 DBD (amino acid residues 280–365) and the HPV 6 E2 DBD (residues 281–368) were expressed in *Escherichia coli* XL1-blue cells using the expression vector pKK223-3 (Pharmacia Biotech). The pKK223-3 plasmid expressing the HPV 16 E2 DBD has been described. HPV 6 sequences encoding the DBD were obtained by PCR using pCB12-6aE2 as the template (a gift from Dr Veronica Sullivan, Roche Discovery, UK) and cloned into a unique *Eco*RI site downstream of the  $P_{tac}$  promoter in pKK223-3. The insert was sequenced using pKK223-3-specific primers. Circular permutation assays were performed using DNA fragments obtained from pBend3-hE2(4), which contains E2 site 4 from the HPV 16 genome cloned between two copies of a tandem array of restriction enzyme cleavage sites.<sup>13</sup>

### Proteins

*E. coli* XL1-blue cells expressing the HPV 16 E2 DBD and the HPV 6 E2 DBD were grown to an  $A_{600\text{ nm}}$  of 0.5. Protein expression was then induced by adding 1 mM IPTG. After incubation at 37 °C for four hours, the cells were harvested by centrifugation and resuspended in 50 mM Tris–acetate/EDTA buffer (pH 7.5) containing 1 mM MgCl<sub>2</sub>, 1% (v/v) 2-mercaptoethanol, 1 mM PMSF, and an EDTA-free protease inhibitor cocktail tablet (Roche). After sonication at 4 °C the lysate was cleared by centrifugation (15,000g for 30 minutes at 4 °C) and DNase I added (5 µg/ml) to reduce viscosity. Following dialysis overnight at 4 °C against 50 mM phosphate buffer (pH 5.7) containing 0.1% 2-mercaptoethanol, the cell extract was re-centrifuged and the supernatant loaded onto an SP-Sepharose fast-flow column equilibrated in 50 mM phosphate buffer (pH 5.7). After

washing with four column volumes of phosphate buffer, bound protein was eluted using a linear gradient of 0–1.5 M NaCl in the same buffer. Pooled fractions containing the DBD were dialysed against ten volumes of 50 mM phosphate buffer (pH 5.7) and then applied to a MonoS 5/5 column equilibrated in the same buffer. Bound protein was eluted using a 0–1.5 M NaCl gradient. Fractions containing the HPV 16 E2 DBD were pooled and then dialysed against PBS containing 1 mM DTT for three hours, before being snap-frozen and stored at  $-70^{\circ}\text{C}$ . Fractions containing the HPV 6 E2 DBD were pooled, dialysed against 50 mM phosphate buffer (pH 5.7), and then loaded onto a 5 ml Heparin-Sepharose column. Bound protein was eluted using a 0–1.5 M NaCl gradient. Fractions containing the HPV 6 E2 DBD were pooled, dialysed against PBS containing 1 mM DTT for three hours, and then snap-frozen and stored at  $-70^{\circ}\text{C}$ . Molecular mass values were confirmed on a VG Quattro triple quadrupole mass spectrophotometer with electrospray ionisation. Protein concentrations were determined from the  $A_{280\text{ nm}}$  value using their respective molar extinction coefficients and their activity determined using gel retardation assays.

### Crystallography

Diffraction-quality crystals were grown using the vapour-diffusion technique. Hanging drops containing 2 ml of protein at 2.5 mg/ml in PBS, 0.1%  $\beta$ -mercaptoethanol and 0.02% (w/v) sodium azide were equilibrated against 1 ml of well solution containing 1–2 M ammonium sulphate, 0.1 M NaCl and 0.1 M Hepes buffer (pH 7.5). For data collection, crystals were soaked progressively for 15 minutes in 5%, 10%, 15%, 20%, 25% and 30% (v/v) glycerol mixed with well solution, prior to freezing in a stream of liquid nitrogen. Attempts to solve the structure using molecular replacement proved unsuccessful. The structure was therefore solved using multiple isomorphous replacement with three heavy-atom derivatives: xenon, mercury and platinum. A native data set was collected at the SRS, Daresbury, on beamline 14.1 and diffracted to 1.9 Å. Crystals were found to belong to one of the two enantiomorphic space-groups  $P6_1$  or  $P6_5$  with cell dimensions  $a = b = 71.6$  Å and  $c = 195.0$  Å,  $\alpha = \beta = 90^{\circ}$  and  $\gamma = 120^{\circ}$ . The xenon derivative was prepared using the Oxford Cryosystems Xcell at the SRS, Daresbury. Data were collected on beamline 14.1, to 2.0 Å resolution. The mercury derivative was prepared by soaking crystals previously soaked in cryoprotectant in 4 mM mercury acetate dissolved in cryoprotectant overnight. Data were collected on a Bruker AXS PROTEUM R system including SMART 6000 CCD detector and rotating anode generator where diffraction was observed to 2.8 Å. The platinum derivative was prepared in a similar way to that used for the mercury derivative, using platinum nitrate and a soaking time of seven days. Data were collected on beamline 14.1 at the SRS, Daresbury, and diffracted to 3.1 Å. All data were processed with the programs DENZO and SCALEPACK.<sup>32</sup> An initial phasing solution was calculated using SOLVE, giving a figure of merit of 0.48, and RESOLVE was used to produce an initial model to 2.7 Å. The model was rebuilt using O and improved electron density maps were produced using DM (CCP4) at which point the phasing was extended to the resolution of the native data set. The structure was refined with REFMAC and CNS, with cycles of manual model building with O. The first round of refinement reduced  $R_{\text{cryst}}$  to 34.6% and  $R_{\text{free}}$  to 40.6%. Repeated cycles of refinement

**Table 4.** Oligonucleotides used in this study

E2 site	DNA sequence <sup>a</sup>
HPV 16 E2 site 1	TTGAACCGAAACCGGTTAGT
HPV 16 E2 site 2	CGTAACCGAAATCGGTTGAA
HPV 16 E2 site 3	GCAAACCGTTTTGGGTTACA
HPV 16 E2 site 4	TTCAACCGAATTCGGTTGCA
HPV 6 E2 site 1	TTCAACCGAAAACCGGTTGTA
HPV 6 E2 site 2	AGGGACCGAAAACCGTTCAA
HPV 6 E2 site 3	TGCGACCGGTTTCGGTTATC
HPV 6 E2 site 4	TGCAACCGTTTTTCGGTTTTT
E2 AATT	CGCAACCGAATTCGGTTGCG
E2 ACGT	CGCAACCGACGTCGGTTGCG
E2 CCGG	CGCAACCGCCGGCGGTTGCG
E2 7A	CGCAACCGAATTCGGTTGCG
E2 7T	CGCTACCGAATTCGGTAGCG
E2 7C	CGCCACCGAATTCGGTGC
E2 7G	CGCGACCGAATTCGGTCGCG
E2 7A CCGG	CGCAACCGCCGGCGGTTGCG
E2 7T CCGG	CGCTACCGCCGGCGGTTGCG
E2 7C CCGG	CGCCACCGCCGGCGGTTGCG
E2 7G CCGG	CGCGACCGCCGGCGGTTGCG

<sup>a</sup> The top strand sequence only is given in each case for clarity.

and rebuilding have lowered the  $R$  factors to the current values of  $R_{\text{cryst}}$  20.31 and  $R_{\text{free}}$  26.28%. A summary of the refinement statistics is presented in Table 1. The electrostatic potential energy surfaces of the E2 DBDs were calculated using GRASP.<sup>33</sup>

### Gel retardation assays

The oligonucleotides used in this work are listed in Table 4. Single-stranded oligonucleotides (200 ng) were 5'-end labelled with [ $\gamma$ -<sup>32</sup>P]ATP using phage T4 polynucleotide kinase in 50 mM Tris (pH 7.6), 10 mM MgCl<sub>2</sub>, 5 mM DTT, 0.1 mM spermidine-HCl, 0.1 mM EDTA. Complementary oligonucleotides were annealed in the same buffer by heating to 90 °C for one minute and then cooling to 20 °C over 20 minutes. Unincorporated label was removed using Micro Bio-Spin 6 columns (Bio-Rad). Labelled oligonucleotides (5000 cpm) were incubated with purified proteins in 20 mM Hepes (pH 7.9), 150 mM KCl, 5 mM MgCl<sub>2</sub>, 5 mM dithiothreitol, 0.1% (v/v) Nonidet P40, 10% glycerol, and 0.5  $\mu\text{g}/\mu\text{l}$  of bovine serum albumin. After 20 minutes at 20 °C, the complexes were resolved on non-denaturing 6% polyacrylamide gels run in TBE. Free and bound labelled DNA was visualised and quantified using a Phosphor-Imager with Molecular Dynamics ImageQuant software (version 3.3). The apparent equilibrium constant,  $K_{\text{eq( apparent)}}$ , was obtained using the equation below and Graft4 software:

$$\frac{[\text{bound}_{\text{DNA}}]}{([\text{protein}] + K_{\text{eq( apparent)}})} = \frac{[\text{maximum bound}_{\text{DNA}}][\text{protein}]}{([\text{protein}] + K_{\text{eq( apparent)}})}$$

When  $[\text{DNA}] \ll K_{\text{eq}}$ , the apparent equilibrium constant is equal to the protein concentration at half maximum DNA binding. All experiments were repeated at least three times.

Circular permutation assays were performed using DNA fragments obtained by digesting pBend3-hE2(4)<sup>13</sup> with the restriction enzymes *Bam*HI, *Xho*I, *Sma*I, and *Bgl*II. This generated a series of fragments in which the position of the E2 binding site varies evenly from one

end of the fragment to the other. The DNA fragments were dephosphorylated using calf alkaline phosphatase then 5'-end labelled with [ $\gamma$ - $^{32}$ P]ATP using T4 polynucleotide kinase and purified exactly as described above. The labelled fragments were incubated with purified proteins in the binding buffer described above. Free and bound DNA was then resolved and visualised as described above. Each experiment was repeated three times and the DNA bend angle ( $\alpha$ ) was calculated using the empirical relationship:

$$\mu\text{m}/\mu\text{e} = \cos(\alpha/2)$$

where  $\mu\text{m}$  equals the mobility of the complex with E2 bound in the middle of the fragment, and  $\mu\text{e}$  equals the mobility of the complex with E2 bound near the end of the fragment.<sup>34</sup>

### Footprinting

A *Pst*I/*Hind*III fragment carrying HPV 16 E2 site 4 was excised from pBend3.E2(4)<sup>13</sup> and labelled at the *Hind*III end using Klenow polymerase and [ $\alpha$ - $^{32}$ P]dATP. Unincorporated label was removed using a Micro Bio-Spin 6 column as described above. Purified DBD proteins were incubated with 1 nM labelled DNA for 30 minutes at 30 °C in 20 mM Hepes (pH 7.9), 150 mM KCl, 5 mM MgCl<sub>2</sub>, 5 mM DTT, 0.1% Nonidet P40, and 0.5  $\mu\text{g}/\mu\text{l}$  of bovine serum albumin; 5  $\mu\text{l}$  of DNase I (1:24,000 dilution of a 5 mg/ml stock) was then added and the reactions incubated at 30 °C for 60 seconds. The reaction products were processed as described,<sup>35</sup> separated on an 8% polyacrylamide sequencing gel, and viewed using a PhosphorImager. Hydroxyl radical footprinting was performed as described<sup>36</sup> and the reaction products visualised as described above.

### Atomic coordinates

The coordinates and structure factors have been deposited in the Protein Data Bank under accession number 1R8H.

### Acknowledgements

This work was made possible by project grants from the AICR and The Royal Society to K.G. We thank Dr Nigel Savery & Professor Steve Halford for comments on the manuscript. J.P. is grateful for a PhD studentship from the UK Biotechnology and Biological Sciences Research Council and Enact Pharma Plc. We thank the staff at the Daresbury SRS for assistance and access to the synchrotron facility.

### References

- Dell, G. & Gaston, K. (2001). Human papillomaviruses and their role in cervical cancer. *Cell Mol. Life Sci.* **58**, 1923–1942.
- Hegde, R. S. (2002). The papillomavirus E2 proteins: structure, function, and biology. *Annu. Rev. Biophys. Biomol. Struct.* **31**, 343–360.
- Giri, Y. & Yaniv, M. (1988). Structural and mutational analysis of E2 *trans*-activating proteins of papillomaviruses reveals three distinct functional domains. *EMBO J.* **7**, 2823–2829.
- Antson, A. A., Burns, J. E., Moroz, O. V., Scott, D. J., Sanders, C. M., Bronstein, I. B. *et al.* (2000). Structure of the intact transactivation domain of the human papillomavirus E2 protein. *Nature*, **403**, 805–809.
- Harris, S. F. & Botchan, M. R. (1999). Crystal structure of the human papillomavirus type 18 E2 activation domain. *Science*, **284**, 1673–1677.
- Hegde, R. S. & 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.
- Kim, S.-S., Tam, J. K., Wang, A.-F. & Hegde, R. S. (2000). The structural basis of DNA target discrimination by papillomavirus E2 proteins. *J. Biol. Chem.* **275**, 31245–31254.
- Bussiere, D. E., Kong, X., Egan, D. A., Walter, K., Holzman, T. F., Lindh, F. *et al.* (1998). Structure of the E2 DNA-binding domain from human papillomavirus serotype 31 at 2.4 Å. *Acta Crystallog. sect. D*, **54**, 1367–1376.
- Hegde, R. S., Grossman, S. R., Laimins, L. & Sigler, P. B. (1992). Crystal structure at 1.7 Å of the bovine papillomavirus-1 E2 DNA-binding domain bound to its DNA target. *Nature*, **359**, 505–512.
- Pepinsky, R. B., Prakash, S. S., Corina, K., Gressel, M. J., Barsoum, J. & Androphy, E. J. (1997). Sequences flanking the core DNA-binding domain of bovine papillomavirus type 1 E2 contribute to DNA-binding function. *J. Virol.* **71**, 828–831.
- Li, R., Knight, J., Bream, G., Stenlund, A. & Botchan, M. (1989). Specific recognition nucleotides and their DNA context determine the affinity of E2 protein for 17 binding sites in the BPV-1 genome. *Genes Dev.* **3**, 510–526.
- Hines, C. S., Meghoo, C., Shetty, S., Biburger, M., Brenowitz, M. & Hegde, R. (1998). DNA structure and flexibility in the sequence-specific binding of papillomavirus E2 proteins. *J. Mol. Biol.* **276**, 809–818.
- Thain, A., Webster, K., Emery, D., Clarke, A. R. & Gaston, K. (1997). DNA binding and bending by the human papillomavirus type 16 E2 protein. *J. Biol. Chem.* **272**, 8236–8242.
- Meyers, G., Bernard, H. U., Delius, H., Baker, C., Icenogle, J., Halpern, A., *et al.* (1995). *Human Papillomaviruses, 1995 Compendium*, Los Alamos National Laboratory, Los Alamos, NM.
- Bernard, B. A., Bailly, C., Lenoir, M. C., Darman, M., Thierry, F. & Yaniv, M. (1989). The human papillomavirus type 18 (HPV18) E2 gene product is a repressor of the HPV 18 regulatory region in human keratinocytes. *J. Virol.* **63**, 4317–4324.
- Bouvard, V., Storey, A., Pim, D. & Banks, L. (1994). Characterization of the human papillomavirus E2 protein: evidence of trans-activation and trans-repression in cervical keratinocytes. *EMBO J.* **13**, 5451–5459.
- Bouvard, V., Thierry, F. & Howley, P. M. (1994). Characterisation of the human papillomavirus E2 protein: evidence of trans-activation and trans-repression in cervical keratinocytes. *EMBO J.* **13**, 5451–5459.
- Sanders, C. M. & Maitland, N. J. (1994). Kinetic and equilibrium binding studies of the human papillomavirus type-16 transcription regulatory

- protein E2 interacting with core enhancer elements. *Nucl. Acids Res.* **22**, 4890–4897.
19. Alexander, K. A. & Phelps, W. C. (1996). A fluorescence anisotropy study of DNA binding by HPV-11 E2C protein: a hierarchy of E2-binding sites. *Biochemistry*, **35**, 9864–9872.
  20. Spalholz, B. A., Byrne, J. C. & Howley, P. M. (1988). Evidence for cooperativity between E2 binding sites in E2 trans-regulation of bovine papillomavirus type 1. *J. Virol.* **62**, 3143–3150.
  21. Knight, J. D., Li, R. & Botchan, M. (1991). The activation domain of the bovine papillomavirus E2 protein mediates association of DNA-bound dimers to form DNA loops. *Proc. Natl Acad. Sci. USA*, **88**, 3204–3208.
  22. Tan, S.-H., Gloss, B. & Bernard, H.-U. (1992). During negative regulation of the human papillomavirus-16 E6 promoter, the viral E2 protein can displace Sp1 from proximal promoter element. *Nucl. Acids Res.* **20**, 251–256.
  23. Tan, S.-H., Leong, L. E.-C., Walker, P. A. & Bernard, H.-U. (1994). The human papillomavirus type 16 E2 transcription factor binds with low cooperativity to two flanking sites and represses the E6 promoter through displacement of Sp1 and TFIID. *J. Virol.* **68**, 6411–6420.
  24. Zinkel, S. S. & Crothers, D. M. (1987). DNA bend direction by phase sensitive detection. *Nature*, **328**, 178–181.
  25. Packer, M. J., Dauncey, M. P. & Hunter, C. A. (2000). Sequence-dependent DNA structure: tetranucleotide conformational maps. *J. Mol. Biol.* **295**, 85–103.
  26. Zimmerman, J. M. & Maher, L. J., III (2003). Solution measurement of DNA curvature in papillomavirus E2 binding sites. *Nucl. Acids Res.* **31**, 5134–5139.
  27. Lewis, H., Webster, K., Sanchez-Perez, A. M. & Gaston, K. (1999). Cellular transcription factors regulate human papillomavirus type 16 gene expression by binding to a subset of the DNA sequences recognized by the viral E2 protein. *J. Gen. Virol.* **80**, 2087–2096.
  28. Koudelka, G. B. & Carlson, P. (1992). DNA twisting and the effects of non-contacted bases on affinity of 434 operator for 434 repressor. *Nature*, **355**, 89–91.
  29. Gaston, K., Kolb, A. & Busby, S. (1989). Binding of the *Escherichia coli* cyclic AMP receptor protein to DNA fragments containing consensus nucleotide sequences. *Biochem. J.* **261**, 649–653.
  30. Kovelman, R., Bilter, G. K., Glezer, E., Tsou, A. Y. & Barbosa, M. S. (1996). Enhanced transcriptional activation by E2 proteins from the oncogenic human papillomaviruses. *J. Virol.* **70**, 7549–7560.
  31. Hou, S. Y., Wu, S. Y. & Chiang, C. M. (2002). Transcriptional activity among high and low risk human papillomavirus E2 proteins correlates with E2 DNA binding. *J. Biol. Chem.* **277**, 45619–45629.
  32. Otwinowski, Z. (1993). DENZO and SCALEPACK—Data processing and scaling programmes. In *Proceedings of the Daresbury Study Weekend* (Sawyer, L., ed), Daresbury Laboratory, Warrington, UK.
  33. Nicholls, A., Sharp, K. A. & Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins: Struct. Funct. Genet.* **11**, 281–296.
  34. Thompson, J. F. & Landy, A. (1988). Empirical estimation of protein-induced DNA bending angles: applications to lambda site-specific recombination complexes. *Nucl. Acids Res.* **16**, 9687–9705.
  35. Spassky, A., Busby, S. & Buc, H. (1984). On the action of the cyclic AMP-cyclic AMP receptor protein complex at the *Escherichia coli* lactose and galactose promoter regions. *EMBO J.* **3**, 43–50.
  36. Tullius, T. D. & Dombroski, B. A. (1986). Hydroxyl radical “footprinting” high resolution information about DNA–protein contacts and application to lambda repressor and Cro protein. *Proc. Natl Acad. Sci. USA*, **83**, 5469–5473.

Edited by J. O. Thomas

(Received 19 May 2003; received in revised form 2 October 2003; accepted 3 October 2003)