Specific recognition nucleotides and their DNA context determine the affinity of E2 protein for 17 binding sites in the BPV-1 genome

Rong Li, Jonathan Knight, Gary Bream, Arne Stenlund,[1] and Michael Botchan[2]

University of California, Department of Molecular Biology, Berkeley, California 94720 USA

The DNA context of nucleotides that a protein recognizes can influence the strength of the protein–DNA interaction. Moreover, in prokaryotes, understanding the quantitative differences in binding affinities that result in part from the DNA context is often important in describing regulatory mechanisms. Nevertheless, these issues have not been a major focus yet for the investigation of protein–DNA interactions in eukaryotes. In this study, we explored the binding specificity and the range of affinities that the BPV-1 E2 transcriptional activator has for DNA. Because E2 binding sites are positioned near several different BPV-1 promoters, such quantitative information may be important to understand transcriptional regulatory mechanisms in BPV-1. Gel retardation assays and DNA footprinting were used to quantitate the affinities of the E2 binding sites in the viral genome. In the process, five sites were discovered, which, on the basis of sequence, had not been predicted previously to interact with the E2 protein. Equilibrium and kinetic studies show that the range of E2 affinities of the 17 sites varied over 300-fold. The sequence elements responsible for E2 recognition of DNA were determined by missing contact analysis of several sites and a point mutation analysis of one site. The results presented show that the affinity of an E2 binding site is to a large extent determined by the availability of specific contacts, but the data also strongly suggest that DNA structure plays an important role.

[*Key Words*: Papilloma virus; protein–DNA complex; DNA structure]

Received January 3, 1989; revised version accepted February 8, 1989.

Regulatory mechanisms that govern DNA transcription, replication, and recombination require specific interaction between proteins and DNA. In eukaryotes a multitude of site-specific DNA-binding proteins, primarily proteins involved in transcription, have been identified, and much progress has been made in the elucidation of their roles in the regulation of promoter activity (Ptashne 1988; Mitchell and Tjian 1989). In general, the nature of the DNA recognition sites for these proteins has been studied in two ways, both of which mainly are qualitative: by comparison of different binding sites for the same protein from which a so-called 'consensus' sequence can be derived, and by mutational analysis to establish which nucleotides within the consensus are required for binding or transcriptional activity. Although such qualitative studies are important, it has been found in prokaryotes that the quantitative aspect of variations in affinity of a protein for different binding sites often forms the crux of complex regulatory mechanisms. For example, in the case of phage λ, the different affinities of repressor and the related *cro* protein for six similar but not identical DNA binding sites control the crucial switch between lysogeny and lysis (Ptashne 1987). Another example is the interaction of *Escherichia coli* promoters with RNA polymerase, in which variations in recognition sequences and their contexts regulate promoter strength by determining the nature of the polymerase–DNA interactions (McClure 1985). Thus, a quantitative understanding of DNA–protein interactions is likely to be important to understand gene regulation.

Accurate determination of what is required for a binding site can be complicated for several reasons. For one, nucleotides may be conserved through evolutionary accidents or constraints, such as nucleotides that are important for an overlapping binding site but not for the one of interest. Also, a single protein may recognize remarkably different DNA sequences (Johnson et al. 1987; Pfeifer et al. 1987; Baumruker et al. 1988; Costa et al. 1988), thus limiting the usefulness of the concept of a consensus sequence. When such variation among binding sites occurs, the problem of explaining this degeneracy of recognition can be resolved in some instances (for example, see Baumruker et al. 1988) by the quantitative consideration of the sum of the effects of several single contact changes, where any single contact is not required absolutely.

[1]Present address: Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724 USA.
[2]Corresponding author.

A third complication is that there are examples of sequence elements that affect binding affinity by determining DNA structure rather than by providing protein contact points. In particular, two cases illustrate the importance of DNA structure in DNA–protein interactions. Crystallographic analysis of the bacteriophage 434 repressor–operator complex (Anderson et al. 1987) shows that each of the recognition α-helices of the repressor dimer can make contact with only 5 bp in each half-site, yet the sequence between half-sites can affect the binding affinity dramatically (Koudelka et al. 1987). These internal base pairs appear to affect the ability of DNA to twist along its long axis to accommodate the specific contacts between protein and DNA (Koudelka et al. 1988). A related situation exists for the catabolite activator protein (CAP). It has been suggested that nonspecific, electrostatic interactions external to the DNA sequence-specific contacts between DNA and CAP can contribute substantially to binding strength, provided that the DNA is flexible enough to bend sharply around the protein (Gartenberg and Crothers 1988). Thus, a CAP binding site whose flanking sequences allow more configurational freedom can form a complex of lower free energy and, therefore, greater stability. In the latter case, the context-dependent effect involves bending, whereas in the former it involves twisting of the DNA.

The early region of the bovine papillomavirus type 1 (BPV-1) provides an interesting system for the study of the role that quantitative differences between recognition sites may play in the biology of eukaryotic organisms because of the large number of different sites in BPV-1 that the E2 gene products recognize. Mutations within the BPV-1 E2 gene have a pleiotropic effect on viral gene functions, including oncogenic transformation, and it is believed that these effects are the result of the requirement for E2 expression to regulate viral transcription (for review, see DiMaio and Neary 1988). The BPV-1 E2 protein first was shown by Spalholz et al. (1985) to activate viral enhancers in *trans*. These authors defined two E2-responsive elements, E2-RE 1 and 2, of the upstream regulatory region (URR) of the virus that served as E2-dependent enhancers of heterologous promoters (Spalholz et al. 1987). That this *trans*-activation was the result of a direct effect of the E2 protein on these enhancers was indicated by the finding that the E2 open reading frame (ORF) encodes a site-specific DNA-binding protein that can bind to several sites within E2-RE 1 and 2 (Androphy et al. 1987; Moskaluk and Bastia 1987). All E2 recognition sites in BPV-1 that have been studied to date contain the sequence motif ACC N_6 GGT, where N is any nucleotide (Hawley-Nelson et al. 1988; Hirochika et al. 1988; McBride et al. 1988; Moskaluk and Bastia 1988a), and predictions that E2 binds to this palindrome as a dimer have been confirmed (Dostani et al. 1988; McBride et al. 1989).

Not all of these sites appear to bind E2 with the same strength. Experiments performed on a limited number of sites have suggested that sites with the motif ACCG N_4 CGGT bind better than sites that deviate in the fourth and ninth bases (Hawley-Nelson et al. 1988; Moskaluk and Bastia 1988b). Moreover, it seems as if some of the target sites for the protein have different capabilities for activation in vivo (Harrison et al. 1987; Haugen et al. 1987; Spalholz et al. 1987). Thus, in the E2 system, a number of naturally occurring sites with quantitatively different affinities for the E2 protein may play different roles in mediating E2 effects. This point is particularly intriguing as putative E2 binding sites are found proximal to all of the known viral promoters. However, little effort has been made to detect, let alone quantitate, the binding of E2 to many of its putative targets.

In this paper we demonstrate E2 binding to a total of 17 sites in the BPV-1 genome; until now only 7 have been shown to bind positively and only 12 have been predicted to exist on the basis of the consensus sequence. We discuss the use of both mutational analysis and a missing contact probing technique to determine which nucleotides in the binding site play a significant role in DNA–protein interaction. We provide a detailed analysis of the relative equilibrium binding constants of these 17 E2 binding sites and show that the affinities vary over a 300-fold range. We found that the presence of the perfectly conserved consensus ACCG N_4 CGGT did not necessarily guarantee that the binding site would be stronger than one with a deviant base. In one case, a site with two deviant nucleotides in one half-site is only twofold weaker than a perfect site. These and other anomalies show that the presence of the palindrome ACCG N_4 CGGT is not a sufficient parameter for predicting the strength of a binding site, even though our data do not indicate any consistent contacts other than these conserved nucleotides. We conclude that the context within which the recognition bases fall influences binding affinities significantly.

Results

E2 protein expression and purification of binding activity

To express E2 protein in large quantity, we made two expression constructs (whole E2 and truncated E2; see Materials and methods) by using vectors that carry a T7 promoter, which is responsive only to its own RNA polymerase (Studier and Moffatt 1986). The resultant plasmids, on induction by a phage λ derivative that carries the gene that encodes T7 RNA polymerase, produced fusion proteins of the expected sizes (47 kD for whole E2 and 36 kD for truncated E2, including an 11-amino acid, amino-terminal fusion of T7 gene *10*, see Materials and methods). When the crude extract, prepared as described, was incubated with labeled fragments that contained one known E2 binding site and assayed for DNA-binding activity by use of the gel retardation method, several shifted bands were observed. These bands could all be competed away simultaneously by addition of oligonucleotides that carried the binding sequence. No specific DNA-binding activity was detected by this method in a protein extract prepared from induced bacteria that carried the expression vector without the E2 ORF.

As described in Materials and methods, both proteins were purified by DNA affinity chromatography. When

the fractions from the affinity column were analyzed for DNA binding by the gel retardation assay (Fig. 1A), the profile showed that the bandshift pattern of the first fractions (lanes 8–11) was very similar to that of the crude extract, whereas the later fractions (lanes 15–17) were enriched to a large extent with the material that produced a simpler pattern that consisted mainly of one major shifted band. SDS–polyacrylamide gel electrophoresis shows that these later fractions corresponded to the peak of a single polypeptide of the predicted size and that this polypeptide was 70–80% pure (Fig. 1B). The purity of the preparations varied as did the relative proportion of the three bands in the gel retardation assay; however, no differences in the relative equilibrium binding properties of different preparations were detected (see Relative equilibrium binding below). It is clear that all bands seen in the gel retardation experiments were produced by the induced polypeptide, because the protein band when excised from an SDS gel and renatured, gave exactly the same gel retardation pattern as did the affinity-purified material.

Both the whole (47 kD) and truncated (36 kD) versions of the protein exhibited identical behavior in footprint and gel retardation assays that used several different binding sites. The experiments reported here were performed with the 36-kD protein, because this form was found to be more soluble and to have fewer copurifying degradation products.

E2 protein binds to 17 ACC N_6 $G^G/_TT$ sites in BPV-1

It has been proposed that ACC N_6 GGT is a minimal recognition sequence for E2 binding (Androphy et al. 1987; Hawley-Nelson et al. 1988). This sequence occurs 12 times within the BPV-1 genome, 10 times in the URR alone. In fact, only seven of these sites have been shown to bind E2 (Hawley-Nelson et al. 1988; McBride et al. 1988; Moskaluk and Bastia 1988a). We have used the DNase I footprint assay and the gel retardation assay to show that E2 bind to these seven sites, the five remaining palindromic putative E2 binding sites, and to five sites of sequence ACC N_6 GTT. Figure 2C shows a summary map of BPV-1 with the binding sites (BS) labeled 1–17.

The results of a DNase protection experiment are shown in Figure 2A. Four labeled DNA fragments (nucleotides 7142–7673, 7476–93, 767–1299, and 2775–3175) were used for the protection of BS 1–8, BS 6–12, BS 13, and BS 16–17, respectively. The first two fragments cover the whole URR. Figure 2A shows that all of the sites in BPV-1 that contained the ACC N_6 GGT sequence did indeed bind the E2 protein, with the exception of site 15. All attempts to obtain a footprint over this region were surprisingly unsuccessful.

Four binding sites with imperfect palindromes were detected also in our footprint assay (Fig. 2A; BS 4, nucleotide 7495; BS 12, nucleotide 16; BS 13, nucleotide 855; BS 16, nucleotide 2921). All of these bear the sequence ACC N_6 GTT. It is interesting to note, as we shall demonstrate in the next section, that although this sequence has gone undetected until now, the ACC N_6 GTT motif is not necessarily a weaker binder than the ACC N_6 GGT motif. All of these sites could be studied by footprint analysis with the exception of site 14.

The DNase I footprint analysis presents an approximate delineation of the range of affinities among the dif-

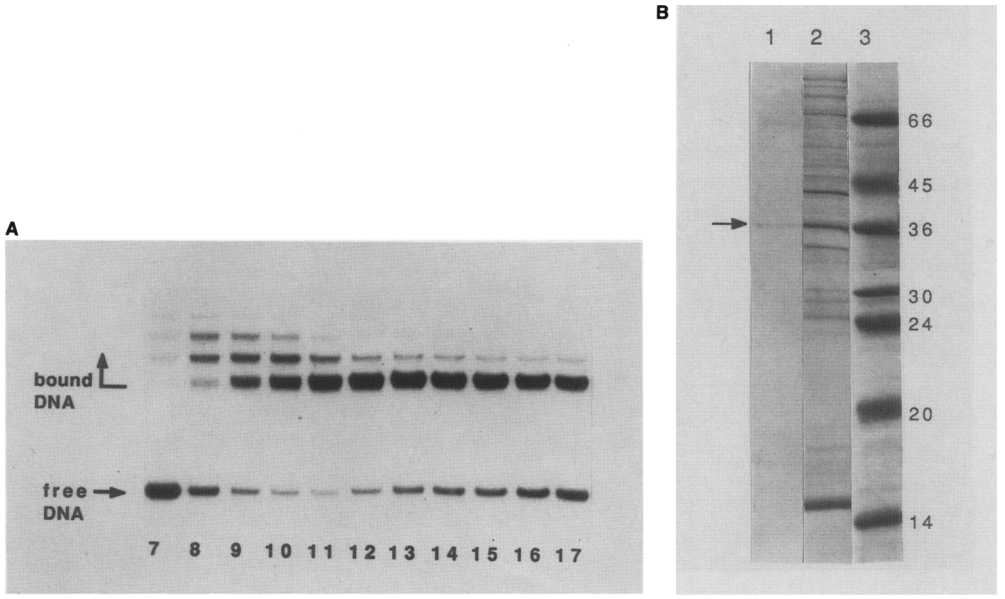

Figure 1. Affinity purification. (*A*) DNA affinity column fractions were tested in the gel retardation assay. The labeled DNA fragment is 173-bp long and contains BS 2. The separated free DNA bands and E2-bound DNA bands are indicated. The numbers at the bottom of the figure represent the fractions collected during elution with high-salt buffer (0.6 M KCl). (*B*) SDS–polyacrylamide gel showing crude extract of induced *E.coli* carrying the short E2 expression vector (lane *2*), and fraction 11 (see *A*) eluted from the affinity column (lane *1*). Molecular weights in kilodaltons are given for size markers in lane *3*. Arrow indicates the affinity-purified polypeptide at 36 kD.

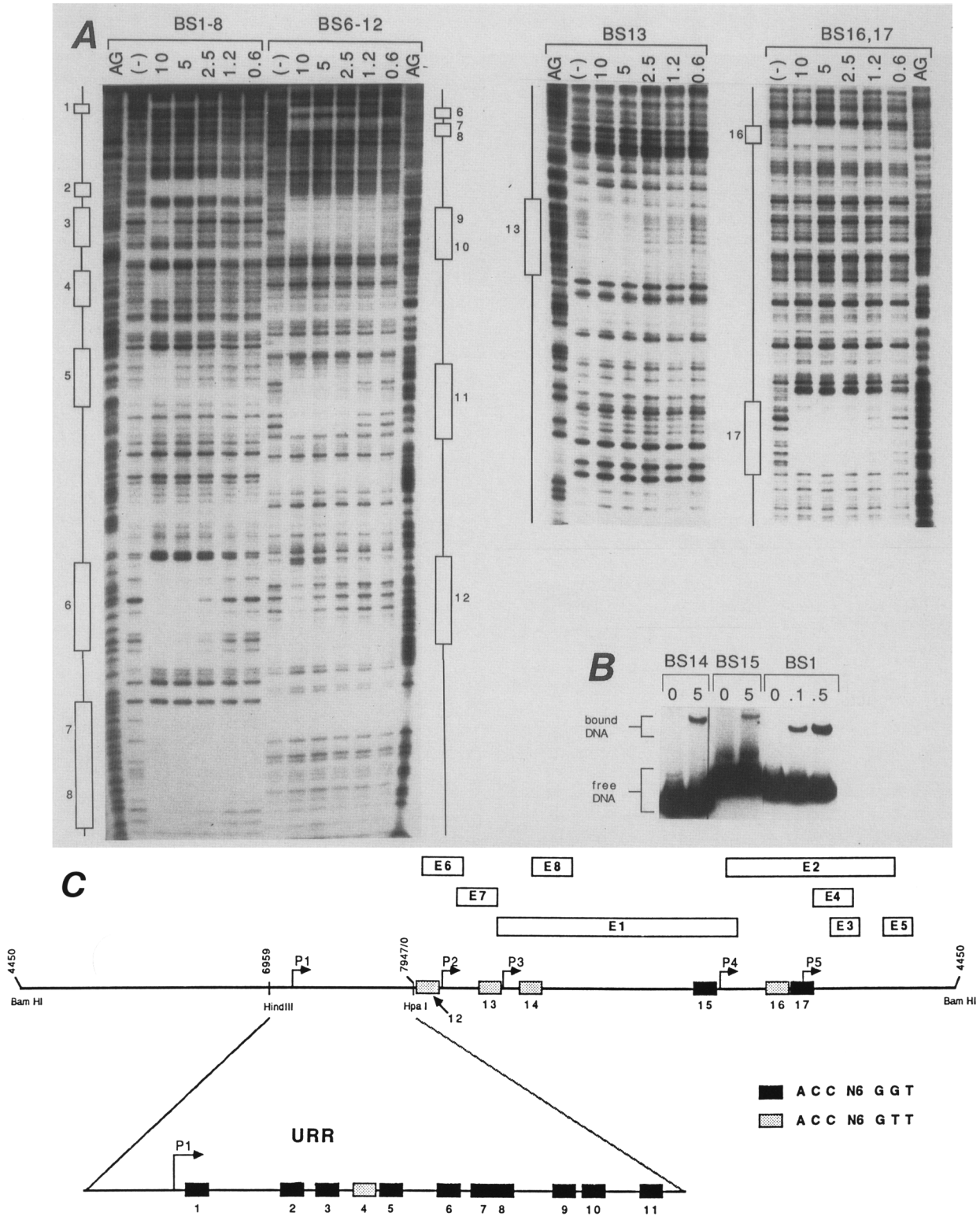

Figure 2. Summary of E2 binding sites in BPV-1 genome. (*A*) DNase I footprint assay showing the protection of 15 E2 binding sites. Numbers at the top of each lane indicate the volume (in microliters) of E2 affinity column eluate (peak fraction) incubated with labeled DNA fragments. AG lanes show the A + G sequence ladder, obtained by the method of Maxam and Gilbert, and the lane marked (−) contains the control samples without E2 protein. All protected regions are indicated by open bars on the side of the figure. BPV-1 sequences used to make labeled fragments are: BS 1–8, nucleotides 7142–7673; BS 6–12, nucleotides 7476–93; BS 13, nucleotides 767–1299; BS 16 and 17, nucleotides 2775–3175. (*B*) Gel retardation assay that compares the weak binding of BS 14 and BS 15 with that of BS 1. The amount of BS 14 and BS 15 DNA shifted by 5 μl of the peak affinity fraction is compared with the amount of BS 1 DNA shifted by 0.1 and 0.5 μl, as indicated above the lanes. (*C*) Map of BPV-1 early region showing the 17 binding sites found in our binding assays. Solid boxes represent the ACC N_6 GGT motif, while stippled boxes represent the ACC N_6 GTT motif. Open bars numbered from E1 to E8 represent the eight ORFs in the BPV-1 early region. Five promoters (P1–P5) found in this region are shown also.

ferent sites. From the amount of protein required to protect the various sites, it is apparent that, for example, BS 4, BS 12, and BS 13 (ACC N_6 GTT) are no weaker than BS 3 and BS 5 (ACC N_6 GGT), and therefore are stronger than BS 15 (ACC N_6 GGT), which could not be footprinted (see below). Also, BS 16 (ACC N_6 GTT) is comparable in strength to some of the intermediate binders with perfect palindromes, such as BS 6 and BS 17.

The two sites, with one of the two motifs discussed above that could not be footprinted, ACC N_6 GTT at nucleotide 1125 (BS 14) and ACC N_6 GGT at nucleotide 2396 (BS 15), were tested by gel retardation analysis. The results in Figure 2B show that fragments that contained these positions bind E2 weakly. For comparison, we show the binding of BS 1, but with 50- and 10-fold less protein (0.1 and 0.5 lanes, respectively). The gel shifts seen with both sites 14 and 15 are >50-fold higher than background, as demonstrated by our data with point mutants (see below). On this basis, we believe that these sites do interact with protein, albeit extremely weakly, and have labeled them accordingly, BS 14 and BS 15.

Relative equilibrium binding constants of the E2 binding sites

It is apparent from the footprint shown in Figure 2A that the affinities of the E2 binding sites for protein vary widely. To quantitate the variation, we have determined the equilibrium binding constant for the 17 sites relative to a binding site of intermediate affinity, BS 1.

The relative equilibrium constant (K_{rel}) of two independent binding sites can be determined with a binding reaction that contains three components: labeled DNA fragments that carry the first site, labeled fragments that carry the second site, and protein. The relative equilibrium constants are given by: $K_1/K_2 = ([C_1]/[D_1])/([C_2]/[D_2])$, where $[C]$ is the concentration of protein–DNA complex, and $[D]$ is the concentration of free DNA (Liu-Johnson et al. 1986). The advantage of this method is that the measurement is independent of the concentration of active protein and of the specific activities of the labeled fragments, both of which are often difficult to determine with precision.

The standard to which all binding sites were com-

BS	loc.	sequence	K_{rel}	$t_{1/2}$
1	7203	G A **A C C A** C A C C **C G G T** A C	0.97 +/- 0.03	2.2
2	7365	G C **A C C G** G C G G **C G G T** A G	6.6 +/- 1.5	ND
3	7408	C G **A C C T** A T C C **C G G T** A A	0.18 +/- 0.02	ND
4	7459	G G **A C C G** A A C A **C G T T** A T	0.12 +/- 0.01	ND
5	7510	G A **A C C A** G A A C **T G G T** A A	0.16 +/- 0.04	ND
6	7592	C C **A C C A** G T A A **T G G T** G C	1.2 +/- 0.1	ND
7	7621	G T **A C C G** C C A T **C G G T** G C	8.1 +/- 1.0	ND
8	7635	G C **A C C G** A T A T **A G G T** T T	1.9 +/- 0.2	8.5
9	7761	G T **A C C G** T T G C **C G G T** C G	4.6 +/- 0.8	17.0
10	7781	A A **A C C G** T C T T **C G G T** G C	11.3 +/- 1.8	ND
11	7896	T C **A C C G** A A A C **C G G T** A A	1.0 +/- 0.09	ND
12	16	A C **A C C A** T C A C **C G T T** T T	0.1 +/- 0.03	0.3
13	855	C A **A A C G** A T A A **A G G T** A G	0.09 +/- 0.01	ND
14	1125	A A **A A C A** G C A G **C G G T** T C	0.05 +/- 0.01	<0.3
15	2396	C C **A C C C** C T C C **T G G T** A A	0.03 +/- 0.01	<0.3
16	2921	A G **A A C C** T A A A **C G G T** G C	0.54 +/- 0.06	1.3
17	3088	G C **A C C A** T G G C **C G G T** G C	1.4 +/- 0.1	1.7
		-2 -1 1 2 3 4 5 6 7 8 9 10 11 12 13 14		

Figure 3. Sequences and relative affinities of the 17 E2 binding sites. The locations and sequences of all 17 E2 binding sites in the BPV-1 genome are shown alongside the equilibrium constant relative to BS 1 (see Materials and methods). Standard deviations are from two to six experiments. Bases are numbered below the figure as they are discussed in the text. The K_{rel} of BS 1 was not defined as 1, but was determined experimentally by testing a 210-bp fragment against the 29-base oligonucleotide to show that the oligonucleotide standard indeed did bind E2 with the same affinity as a fragment with significantly larger sequence flanking the site. Half-lives are also given where determined. (ND) Not determined.

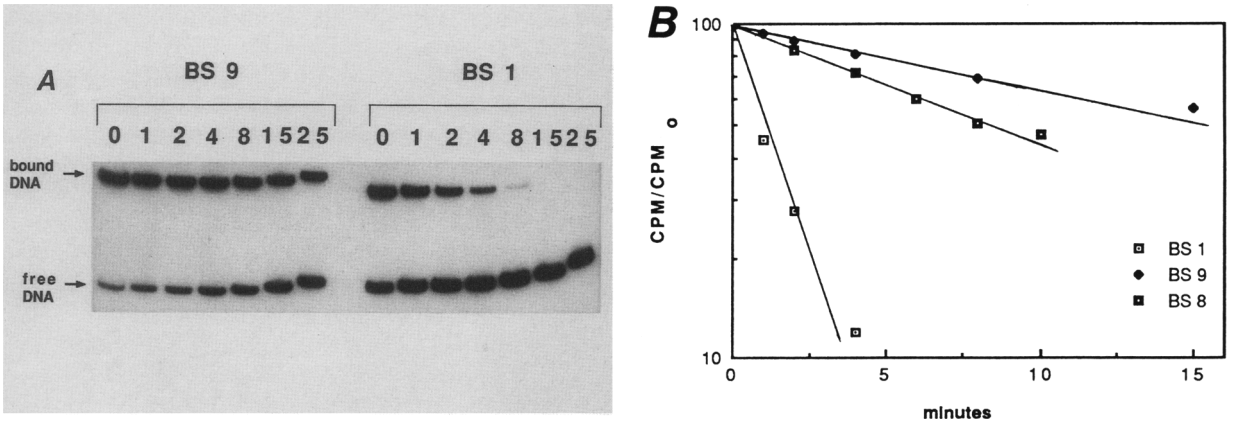

Figure 4. Off-rate determination for binding sites 1, 8 and 9. (*A*) Labeled DNA fragments bearing either BS 1 or BS 9 were incubated with protein as described in Materials and methods. After the reactions had reached equilibrium (a time course has shown this to be <5 min), a 1000-fold excess of plasmid DNA that carried the BPV-1 URR was added, and aliquots were loaded onto a running gel at selected time points. Because the gel ran continuously as time points were taken, apparent differences in mobility can be observed. Times are indicated in minutes above the lanes. Time point zero was taken before the addition of competitor DNA. (*B*) Data from *A* are plotted on a semi-log graph. Shifted cpm at time *x* over shifted cpm at time zero are plotted versus time in minutes. Data for BS 8 are also included. The half-lives determined from this plot are: BS 1, 2.2 min; BS 8, 8.5 min; BS 9, 17 min.

pared was a synthetic oligonucleotide containing BS 1. This molecule carried the BPV-1 sequence from nucleotide 7194 to nucleotide 7218 plus a 4-base overhang on each end. The sequence of this standard was:

GATCCTTTATTGGAACCACACCCGGTACA
GAAATAACCTTGGTGTGGGCCATGTCTAG

The end-labeled oligonucleotide was mixed with end-labeled fragments carrying the various binding sites and incubated with protein for 30 min at room temperature. The lengths of the added fragments were selected to insure that the resulting shifted bands could be distinguished from those of the oligonucleotide standard, and these lengths varied from 67 to 290 nucleotides (see Materials and methods for exact lengths). The result for BS 1 in which a 210-bp fragment was tested against the 29-base oligonucleotide demonstrates that the assay is insensitive to fragment length. The reactions were run on a nondenaturing polyacrylamide gel, and the relative intensities of the bands on an autoradiograph of the gel were measured by scanning densitometry.

The resulting relative equilibrium binding constants are given in Figure 3 along with the sequence and location of each of the sites. The values shown represent the average of several repetitions of the experiment with different protein concentrations and preparations. As expected, the binding constant of the restriction fragment that contains BS 1 is equivalent to that of the oligonucleotide.

The data show the variation in affinity of the protein for the 17 sites. The binding constant of BS 10 is 376 times greater than that of BS 15, and the other binding strengths are spread out between these two. The hierarchy of binding strengths determined by this technique corresponds well with that derived from the footprint analysis, in which protein concentration was increased in twofold steps (Fig. 2A). Because the footprint does not cover the entire range of E2–DNA affinity (protein concentration varies only 16-fold), the sites with the highest binding constants are protected fully at the lowest protein concentrations (BS 2, 7, 9, and 10) and those with the lowest binding constant are protected only partially at the higher protein concentration (BS 3, 4, and 12).

Dissociation rates

The dissociation rate of E2 from a binding site can be measured in terms of the half-life by addition of a vast excess of specific competitor DNA to a binding reaction that has already reached equilibrium. The ratio of bound to free DNA is then measured at subsequent time points (see Materials and methods for a complete description).

Figure 4 presents an example of such an experiment performed on BS 9 and 1, chosen because of the marked contrast in their half-lives. The data are plotted on the accompanying graph (Fig. 4B) as the log of cpm measured over cpm at time zero versus minutes after the addition of the competitor DNA. The curve for BS 8 also is included in the graph. BS 9 has the longest half-life, 17 min, whereas the half-lives of BS 8 and 1 are 8.5 min and 2.2 min, respectively. The half-lives of these and other sites are given in Figure 3.

Like the range of binding constants, the range of E2 complex half-lives, from 17 min to <20 sec, is also quite striking. E2 bound to sites 14 and 15 dissociates so rapidly that the half-lives are shorter than this technique can measure. In general, the stronger sites have longer half-lives.

Determination of nucleotides involved in E2–DNA interaction

The results presented above substantiate the notion that the strongest binding sites contain the perfect palin-

drome ACCG N_4 CGGT. However, as we will discuss later, anomalies exist in these data. We have used two different methods to study further the interaction of the E2 protein with DNA. Point mutations were generated at every base across one particular site, and the missing contact method was employed at several sites (Brunelle and Schleif 1987) to assess the roles of individual nucleotides. Because of its convenience, the latter method is suited particularly to examining binding sites in different contexts.

Point mutations in the E2 binding site BS 1 was chosen for initial study and point mutational analysis. BS 1 is one of the E2 binding sites that is known to be functionally important in vivo. It is located near the P1 promoter, which overlaps one of the regions in BPV-1 involved in viral replication control (Stenlund et al. 1987). Furthermore, it has been demonstrated recently that E2 can interfere with the action of cellular factors that bind to overlapping and proximal sites at BS 1 and by doing so repress transcription initiation from P1 (Stenlund et al., unpubl.).

The results in Figure 5 show the protection by E2 protein on both strands of BS 1. The protected region, as indicated by the sequence listed alongside the footprint, extends several bases (3–6 bp) beyond the sequence (ACC N_6 GGT). Twenty bases are protected on the top strand and 23 on the bottom.

The mutants generated in BS 1 are shown in Figure 6A. The *Hin*cII–*Mlu*I fragment that carried BS 1 wild type or one of the mutants was end-labeled and tested in the gel retardation assay. Reactions were incubated for 30 min at room temperature before being loaded onto the gel. Initial concentrations of DNA and protein were held constant so that the relative affinities for E2 of each of the mutants could be estimated from the relative amounts of DNA bound in parallel reactions. Figure 6B shows the binding of 16 of the mutants. It is clear that several types of mutations, changes in the conserved palindrome in particular, eliminate E2 binding, whereas others only slightly reduce or even increase affinity. We have quantified the binding affinities of each of the mutants more accurately by determining the equilibrium binding constant relative to wild type as described above. The resulting values are given in Figure 6A. In addition, we have determined the dissociation rate for three of the mutants and the wild type as described above. The results are shown in Figure 7. We point out that, although some of the mutational effects on the relative equilibrium constants are small, they parallel the effects on the dissociation rates measured. For example, as shown in Figure 7, the dissociation rate of mutant 7207-8 is about 40% of wild type and its relative equilibrium constant is 36% of wild type (Fig. 6A).

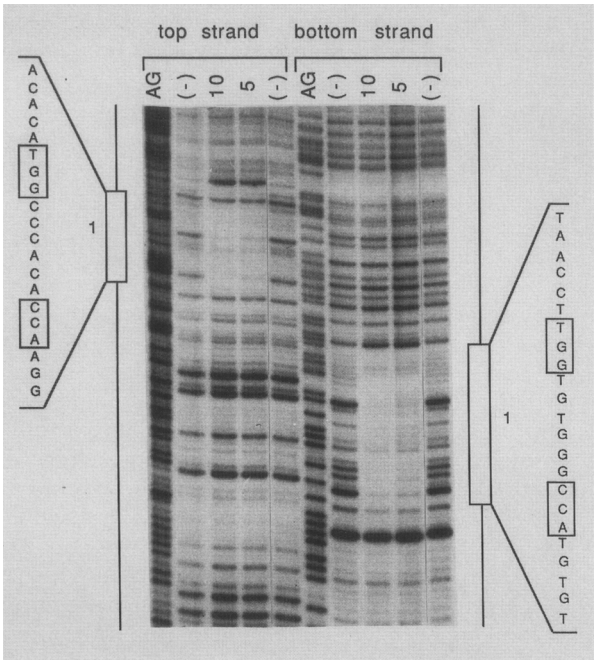

Figure 5. DNase I footprint assay showing the protection of both strands of BS 1. The BPV-1 fragment used extended from nucleotides 7100 to 7274. The sequence in the region of the binding site is listed on both sides of the figure.

Several important observations can be made from these data. First, the most devastating mutational effects occur within the conserved ACC or GGT portions of the binding site. The mutants 7204 (ACC to AAC) and 7212 (GGT to GTT) have less severe effects than others in the conserved region, which is consistent with our detection of the sequence AAC N_6 GGT as a bona fide E2 binding site.

Second, these data show the importance of the fourth and ninth nucleotide positions (our system of numbering the nucleotides across the site is given in Fig. 6A). Changing the C9 to an A destroys binding, and, changing the A4 to a G improves binding. This supports the proposal that ACCG N_4 CGGT is a better recognition sequence than is ACC$^{Py}/_{A}$ N_4 $^{Pu}/_{T}$GGT. Also, three of four possible bases have been tested at position 4 with the result that both purine options produce a better binder than cytosine. Possibly, ACCPu is better than ACCPy.

Third, there are effects of changing the nucleotide immediately preceding and immediately following the conserved nucleotides. Changing the A at −1 in mutant 7207 to a C strengthens binding 2.7 times. In contrast, changing the A at +13 in mutant 7215 to a C reduces the binding constant sixfold. Although there is no striking conservation of C at the −1 position when all sites are considered, if we take into consideration both strands of the 10 strongest binding sites (BS 1, 2, 6, 7, 8, 9, 10, 11, 16, 17), 15 out of 20 ACCs are preceded by a pyrimidine. It is possible that a pyrimidine is preferred at the −1 position.

A final observation is that several nucleotides in the center of the site appear to be important in the determination of the binding strength of BS 1. This is particularly interesting, because there is no apparent conservation of these nucleotides among the 17 sites. The number of nucleotides in the central region appears to be critical, because we have found that the insertion of a single base, a cytosine before position 7, reduces binding

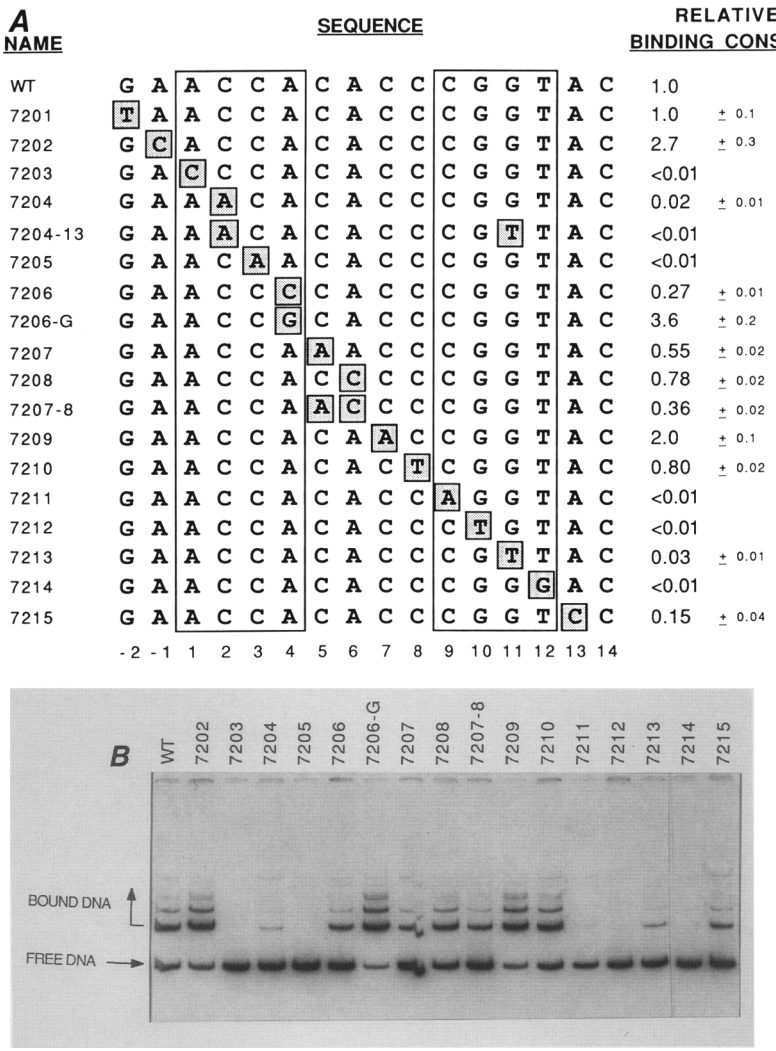

Figure 6. Point mutations in BS 1. (*A*) The effects of point mutations on E2 binding affinity in the region of BS 1 are given as equilibrium binding relative to wild type (WT). Altered nucleotides are in stippled boxes. The point mutant names reflect the position at which a base was changed. In all cases a *Hin*cII–*Mlu*I fragment (7143–7352) was excised from the mutant BPV DNA and tested against a 29-base oligonucleotide that carried wild-type BS 1 as described in Materials and methods. The values given are the mean of at least two experiments followed by the standard deviation. The value <0.01 is arbitrary and is based on the length of exposure of the gel. The actual value may be lower by an order of magnitude or more. Nucleotides are numbered at the bottom of the figure for reference in the text. (*B*) Gel retardation assay illustrating the effects of the BS 1 mutations. Wild-type or mutant DNA was incubated with the affinity-purified E2 protein. An early affinity column fraction was used, which resulted in multiple bands that are likely to be degradation products of E2 (see Fig. 1 and Results). The protein was in excess so that the amount of DNA shifted would reflect the relative affinities of the mutants. Numbers above the lanes correspond to mutants depicted in *A*.

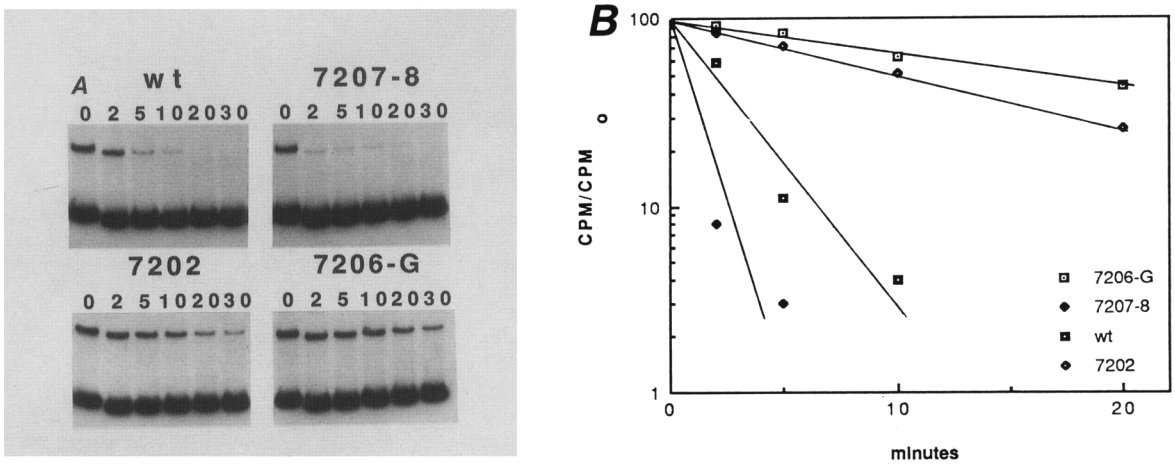

Figure 7. Off rates of BS 1 wild type and three point mutants. (See Fig. 4 legend.) (*A*) Off-rate gel of BS 1 wild type (wt) and three mutants. Times are indicated in minutes above the lanes. (*B*) Half-lives are: wt, 2.0 min; 7207-8, 0.6 min; 7207, 9.5 min; 7206-G, 17.0 min.

by greater than 100-fold (data not shown). But, more significantly, point mutations that do not change the size of the spacer region can also affect binding, and, in some cases, have an effect that is as dramatic as the effect of changing a conserved base. For example, mutant 7207-8 (in the nonconserved region) reduces E2 binding by the same factor that 7206-G increases it [the latter converts the site into a member of the highest-affinity group (ACCG N_4 CGGT)]. It seems that although no specific nucleotide is required at a specific position among the central 4 bp, these bases may provide a context that favors or opposes binding.

That DNA sequence context is a parameter in E2 protein binding is indicated also by the mutation data at positions 4 and 9. BS 1 is one nucleotide (position 4) away from being a perfect palindrome. Changing the deviant A to the correct G (mutant 7606-G) improves binding 3.6-fold. This mutant site now has roughly the affinity of BS 9 (which has a perfect motif) but still significantly less affinity than other sites with perfect motifs. Similarly, a mutation at position 9 (mutant 7211) lowers affinity of E2 greater than 100-fold, making this mutant site much weaker than all of the naturally occurring sites with deviations at both 4 and 9 (e.g., sites 5 and 6 in Fig. 3). This question of context is considered further in the following sections and in the discussion.

Missing contact probing A technique for testing the contributions of individual bases, rather than of base pairs, is the recently developed missing contact probing (Brunelle and Schleif 1987). The principle of this assay is to remove a certain base in or around the binding site either by depurination or by depyrimidation, and to determine how this influences the overall binding affinity. Compared with other chemical modification techniques such as DMS interference, which only tests guanines, missing contact probing has the advantage of being able to test the role of each base in binding to the protein.

Labeled DNA fragments that bore only one binding site were depurinated or depyrimidated sparingly and in-

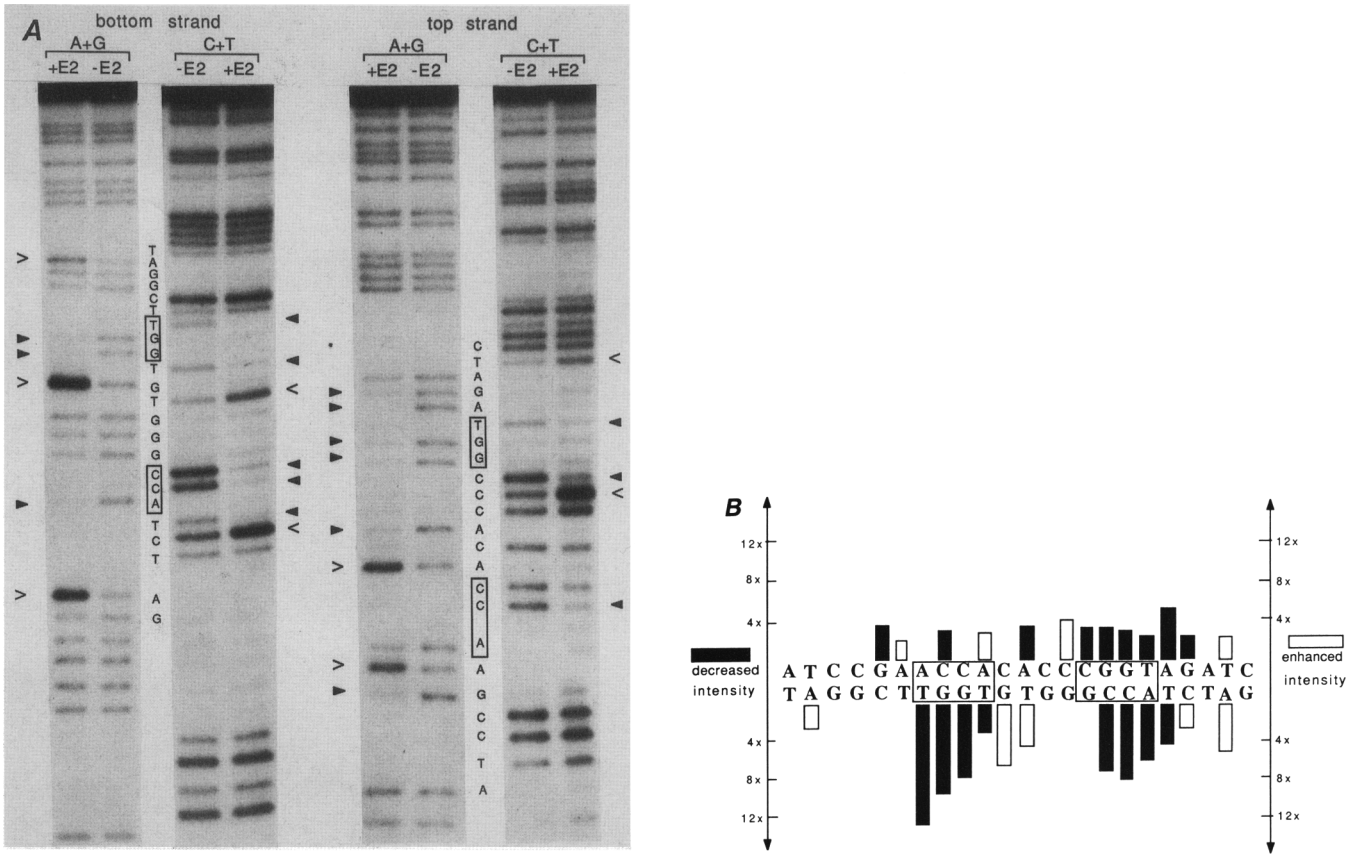

Figure 8. Missing contact probing on both strands of BS 1. (*A*) Autoradiograph showing the effects on binding of depurination (A + G) and depyrimidation (C + T) in and around BS 1. (Lane *−E2*) DNA isolated from free DNA bands; (lane *+E2*) DNA isolated from mobility-shifted DNA bands in gel retardation assay. Labeled fragments were taken from a pUC18 plasmid into which an oligonucleotide with the sequence 5′-GATCCGAACCACACCCGGTA-3′ (BS 1) had been cloned. As a result, the sequence in the figure deviates from wild-type BPV sequence. Solid triangles represent positions where base removal results in decreased binding affinity; carets represent positions where base removal enhances the binding. (*B*) Densitometric quantitation of the data shown above. Solid bars indicate decreased band intensity, open bars indicate enhanced band intensity. The numbers on both axes represent the fold increase or decrease. The sequence in the figure deviates from wild-type BPV-1 sequence in the first four and the last four nucleotides shown (see *A*). In BPV-1 these are ATTG and CACA, respectively.

cubated with E2 protein. E2-bound fragments and free DNA fragments subsequently were separated as in the gel retardation assay. Fragments with bases removed that are important for binding remained in the free-DNA pool. Both bound and unbound bands were excised and DNA fragments were eluted from the gel. After cleavage by piperidine at the positions of the removed bases, DNA fragments were resolved by denaturing polyacrylamide gel electrophoresis.

Figure 8A shows the missing base analysis of the interaction of E2 with BS 1. Bases, which, on their removal, reduced the binding affinity, resulted in bands of decreased intensity in the +E2 lane compared with the −E2 lane (indicated by solid triangles). In contrast, if the removal of a base facilitated binding, an enhanced band would result (indicated by carets). Bases irrelevant to the interaction with E2 had the same intensity in both lanes. Thus, this comparison allowed us to determine which bases played a role in E2 binding.

The results, quantitated by densitometry, are diagramed in Figure 8B. Only differences that are reproducibly greater than twofold are considered as strong effects and are included in the figure. Solid bars indicate decreased intensities and open bars indicate enhanced intensities. Consistent with the results of the point mutations, the most significant effects are located in the conserved region of the site, that is, C2, G10, G11, and T12 on the top strand; and A1, C2, C3, G10, G11, and T12 on the bottom strand. Interestingly, the missing contact experiments on this site and others (see below) indicate that the protein–DNA interactions are not perfectly symmetrical, as would be implied by the sequence. For example, the removal of A1 and C3 from the bottom strand (downstream half-site) had a much stronger effect than the removal of these bases from the top strand (upstream half-site). Furthermore, bases outside of the conserved regions, such as G-2, A-1, A6, C8, A13, and G14 on the top strand and T9, G8 T7, T-1, and C-2 on the bottom strand, also influenced binding affinity when removed.

We applied the missing contact probing to five other binding sites (BS 2, 7, 8, 9, and 11). An example, BS 9, is given in Figure 9A, and the results for all the sites examined are summarized in Figure 9B. Again, we see that the strongest effects are in the ACC and GGT regions, indicating contact between E2 and the conserved sequence motif.

Corroborating the point mutation result that a G is preferred at position 4 (and thus a C at position 9), removal of C9 always substantially decreased binding affinity. However, removal of the G on the other strand had little, if any, effect on E2 binding. Thus, it appears that the C of this base pair is contacted by protein whereas the G is not.

BS 8 and BS 11 exhibit an asymmetry with respect to E2 binding reminiscent of the results obtained with BS 1: one half-site was affected less by base removal than was the other half-site, and removal of certain bases outside of the conserved regions enhanced binding. These sites where shown above to have 5- to 10-fold lower

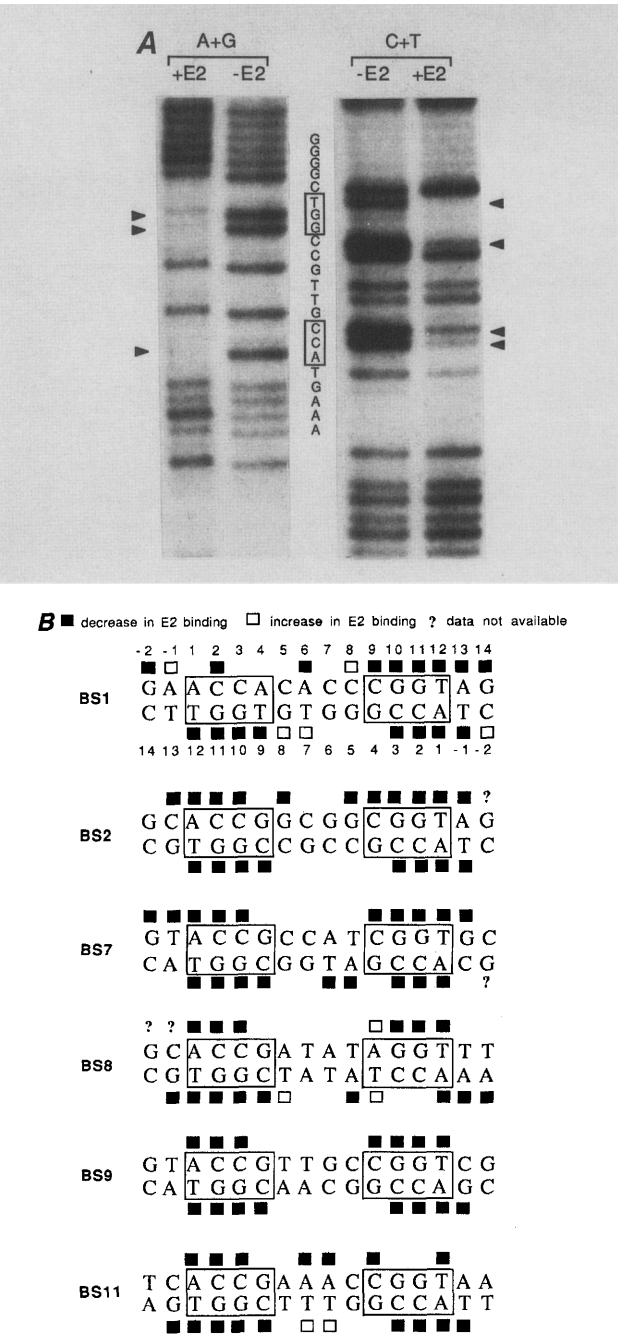

Figure 9. Missing contact probing on several E2 binding sites. (*A*) Autoradiograph showing acid depurination (A + G) and depyrimidation (C + T) results for the top strand of BS 9. (Lane *−E2*) DNA isolated from free DNA bands; (lane *+E2*) DNA isolated from E2-bound DNA bands in gel retardation assay. (Arrowheads) Positions where base removal causes more than twofold decrease in E2 binding affinity. DNA sequence is shown in the middle, with consensus regions highlighted. (*B*) Summary of data on all tested binding sites. For all tested bases, only effects larger than twofold are considered significant and included here. (■) Positions where base removal decreases E2 binding affinity; (□) positions where base removal enhances binding affinity; (?) data not available. For BS 1, the sequence tested and shown here has a G at position 14, whereas wild-type BPV-1 has a C at that position (see legend to Fig. 8).

equilibrium binding constants than the other three sites tested here. This asymmetry suggests that a decrease in E2 interaction with only one half-site may account for the lower binding affinity of BS 1, 8, and 11. It is interesting that binding site 11 behaves this way, despite its two perfect half-sites.

Sites with the most contact points, as suggested by the missing contact analysis and the mutational analysis of BS 1, can bind protein with the highest affinity. For example, BS 2, 7, 9, and 10 all have eight of 8 bp that are shown to be significant in binding (ACCG N_4 CGGT), and all are ~5- to 10-fold stronger than the BS 1 standard. The surprising exception is BS 11, which also has the perfect 4-bp inverted repeat, but which has a 5- to 10-fold lower affinity for E2 protein than any of the other perfect sites and only twice the affinity of BS 16, a site with two deviations. Thus, the presence of these conserved nucleotides is not the sole determinant in the strength of interaction between E2 and its binding sites.

Discussion

The results presented above emphasize two important points concerning the E2–DNA complexes: the wide range of affinities among the viral E2 binding sites and the large role played by context in determining these affinities. This conclusion is based not only on a quantitative comparison between the naturally occurring sites in BPV-1, but also on mutational and missing contact analysis. These points are particularly salient in that several new binding sites were discovered within the viral genome whose existence could not have been predicted from the results of previous studies.

At this time it is evident that E2 binds as a dimer (Dostani et al. 1988; McBride et al. 1989), each of whose subunits interacts in the major groove with the conserved nucleotides of one half-site. The results tabulated in Figure 3 demonstrate that, of the 17 E2 binding sites studied, those with the highest affinity all contain the conserved sequence motif ACCG N_4 CGGT. That this palindromic sequence provides specificity of binding is substantiated by a number of our observations. First, a mutation in BS 1, which converts this site into the perfect palindrome (7206-G), increases binding, whereas any of the mutations that bring the site further away from this perfect palindrome decrease binding (see Fig. 6). Second, the missing contact data show that these bases in general (with some interesting exceptions), when removed, decrease the affinities of the sites studied substantially. That the interaction is in the major groove has been supported by the methylation interference data of Moskaluk and Bastia (1988b) and Dostani et al (1988). Figure 10 illustrates this simple model for binding and shows one more possible reason why the C base of the G/C base pair at positions 9 and 4 of the palindromic sites is required for binding while the G base on the other strain is not (Fig. 8B): The functional groups of G base that potentially could mediate protein contact are presented on the face of the helix opposite the presumed binding face of the E2 protein.

We consider one binding site to be the consequence of two separate yet interacting half-sites, because of (1) the dimeric nature of the protein; (2) the twofold rotational symmetry of the sequence itself; and (3) the tendency for certain mutations or base removals within the perfectly conserved half-site to have a greater effect on binding than mutations in a divergent half-site. The sites can be grouped according to the number of deviations from the perfect palindrome that existed in each of the two half-sites (5′ and 3′ domains), as shown in Figure 11. Four groups result: perfect palindromes (five members), one deviation in one half-site (five members), one deviation in each half-site (five members), and two deviations in one half-site (two members). On the basis of the above model for binding, if the nucleotides in these domains are the primary determinants of binding affinity, then it follows that the binding strengths within each group should be similar and the groups should become progressively weaker as the number of deviations from the perfect palindrome increases.

It is obvious immediately from Figure 11 that this is not always the case. Yet the discrepancies cannot be explained in terms of the type (transition versus transversion) or the location of base differences within the conserved sequence. We propose that the large number of quantitative differences either within families or between families can best be understood in terms of what we will call a context effect. A few examples will illustrate this point. BS 11 is at least 10-fold lower in affinity for E2 protein than is BS 10, yet both contain the perfect ACCG N_4 CGGT motif (see Fig. 3 for sequences and binding strengths of each site). Sites 17 and 8 both are mismatched from the ideal sequence by one base pair (site 8 contains the sequence ACCG N_4 AGGT and site 17 ACCA N_4 CGGT, where the mismatch is underlined), yet both of these sites bind E2 better than does BS 11. Thus, BS 11 is anomalously weak.

In another example, BS 16, which contains two deviations from the conserved palindrome, AACC N_4 CGGT, is >10-fold better in its ability to bind E2 than BS 14, which also has two deviations, AACA N_4 CGGT. The only difference between these sites, if these nucleotides alone are considered, is at position 4. On the basis of the point mutant data, however, this change would be predicted to make BS 14 fourfold better than BS 16 and not 10-fold worse [compare wild type (WT) and 7206 in Fig. 6A]. The discrepancy between the prediction and the result implies that there are more complex differences between sites 14 and 16 than the single change in the conserved sequence. Perhaps the most dramatic anomaly is found in the 50-fold difference in E2 binding affinities between BS 15 and BS 6. These two sites are both members of the same group as defined above and in Figure 11, and differ in their conserved sequence by only one nucleotide: BS 15, ACCC N_4 TGGT; BS 6, ACCA N_4 TGGT. As discussed in the last example, this difference would account for no more than a fourfold difference in affinities, according to the point mutant data, if the conserved nucleotides were the only factors in binding affinity. In addition, BS 6 is anomalously

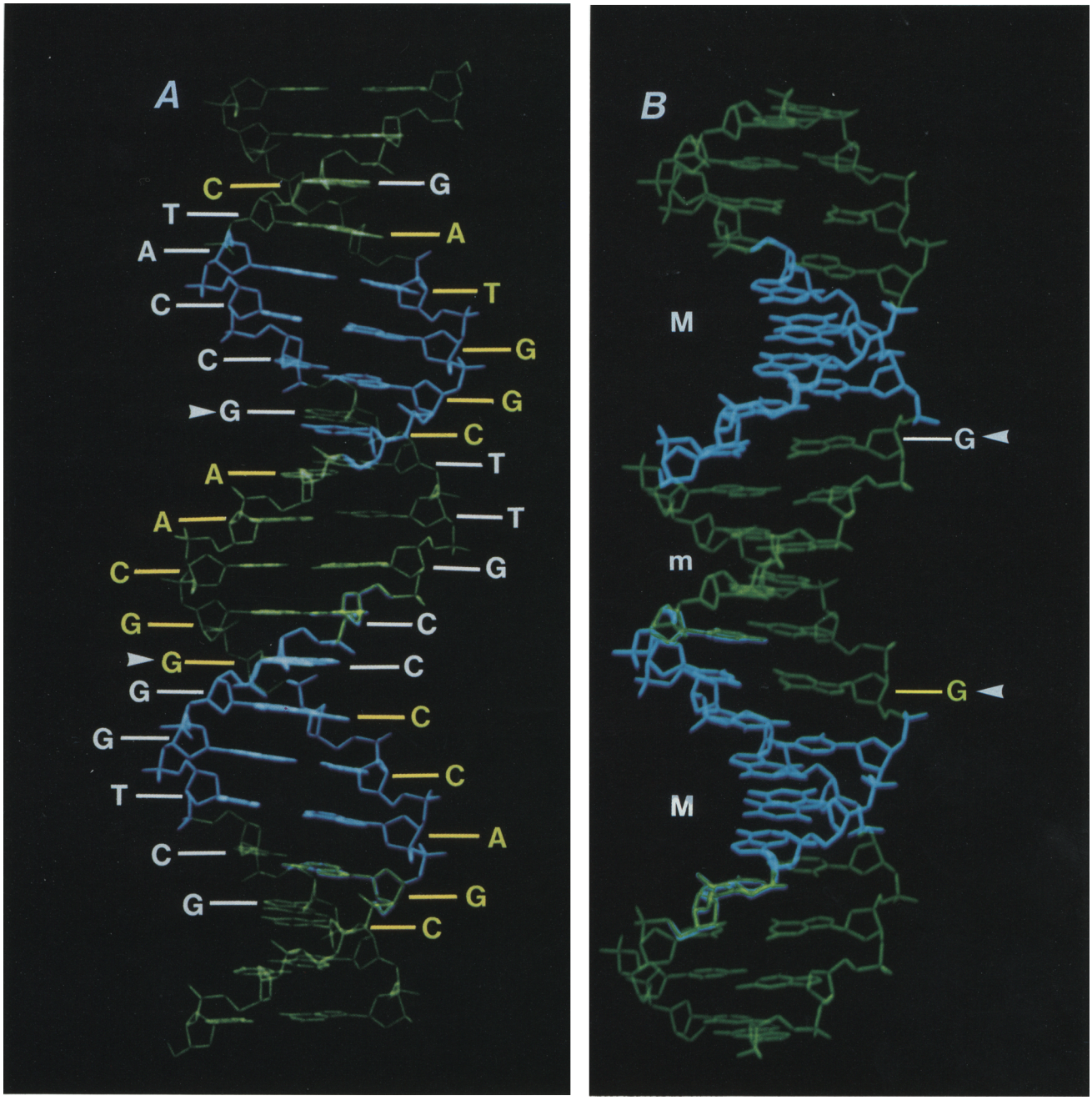

Figure 10. Three-dimensional computer graphic depiction of an E2 binding site in the B-DNA form to show how the E2 protein might make the contacts described in the text. The sequence corresponds to binding site 9. White-lettered bases are on the top strand running 5′ to 3′ from the top to the bottom of the figure; yellow-lettered bases are on the bottom strand. Nucleotides in blue are those that are both conserved and implicated as contact points for the protein by missing contact probing. The two views, *A* and *B*, are oriented 90° from each other. It is clear from this representation that all the contacts in blue are accessible to a protein binding in the major groove (front in *A*, left side in *B*). In contrast, the guanine residue indicated by the arrowhead might not be accessible to a protein that bound in the major groove but that did not wrap itself around the site. Such a model explains the indifference of E2 protein to the removal of the arrowhead G while the removal of its complementary C reduces binding severely (see Fig. 9).

stronger than some sites with fewer deviations, including the perfectly palindromic site 11 (see Fig. 11).

The notion of context is necessarily vague in the absence of crystallographic data. However, several different sorts of phenomena can be envisioned. A base within or external to a particular binding site may by chance interfere with binding as a result of its possession of a bulky group that blocks access to the specific contacts. Alternatively, a base may improve binding directly by providing a chance contact for the protein or

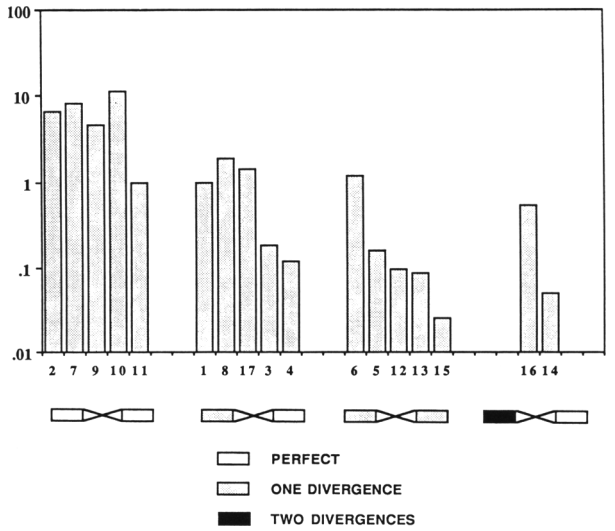

Figure 11. Relative affinities of the BPV-1 E2 binding sites grouped according to conservation of half-sites. The 17 sites are divided into four groups on the basis of the number of deviations from the perfect palindrome ACCG N_4 CGGT. Binding site numbers (BS; see Fig. 2) are given on the ordinate, and the binding affinities of each site relative to BS 1 are plotted on a logarithmic scale. Each group contains binding affinities that vary over at least an order of magnitude, indicating that factors besides the conservation of these nucleotides contribute to the strength of a site. (See text for discussion.)

indirectly by affecting the local hydration of the DNA such that the entropy gain from the release of bound water molecules favors protein binding.

Another view for which some support is found in our data is that the protein requires or induces some conformational change in the DNA on binding. That B-DNA polymorphisms exist that are dependent on base steps is an established fact (for example, see Nelson et al. 1987). Thus, one site that more closely approximates this shape before binding would have a higher affinity for E2 than a site with a very different shape. In such a case the notion that flexibility improves binding strength (Hogan and Austin 1987) may not apply, as the site may have a high affinity because it approximates the required shape and may be relatively stiff.

That DNA structure affects binding is consistent with our data. The base changes in the internal positions of BS 1 sometimes helped and sometimes hindered E2 binding, yet no evidence could be found in other sites for specific contacts in this internal region (see missing contact data in Fig. 9). For example, mutant 7209, which changes a C to an A at position 7, improves binding; yet in three cases tested by missing contact probing in which an A was present at this position on either strand, no effect on binding was observed after removal of the A or complementary T. Although it is conceivable that the sequence context of a binding site might by chance create such a contact, in the absence of crystallographic data, this is a difficult point to resolve.

However, certain results presented here can best be understood in terms of the effect of DNA conformation on binding. It is particularly striking in the missing contact data that certain nucleotides that interact with the E2 protein in high-affinity sites do not seem to contribute to the binding affinities for the weaker sites. For example in BS 11 the 3′ half-site shows little involvement of the GG in the CGGT motif and in BS 8 the CC on bottom strand of the 3′ half-site also shows little involvement. Both of these cases might be explained by the idea that the DNA is restrained conformationally and cannot be distorted to allow for such contacts and that this condition accounts for the low affinities of these sites. This idea is supported by the fact that only in the three sites in which all conserved nucleotides are not contacted (e.g., 1, 8, and 11) did the missing contact probing reveal bases whose removal will actually improve binding. Clearly the removal of a base within a site that causes improved binding affinities implies a structural constraint mediated by the deleted nucleotide, irrespective of the nature of the constraint. Moreover, this implies that the protein has a stringent requirement for a precise geometry with little ability to accommodate changes. This is supported by our plus-one mutation, which inserts a single base between the half-sites and results in greater than 100-fold reduction in binding. It would, of course, be interesting to test this concept by making internal mutations in BS 11 to increase its affinity and to see if these changes lead to contacts throughout the 3′ half-site.

Is there any data that may help to explain the presumed context effects described above? Although we are reluctant to ascribe all of the anomalous differences of affinities to one particular phenomenon, the ability of the DNA to bend and bring the major grooves of the half-sites closer together (see Fig. 10) could accommodate some of the data presented. Indeed, Moskaluk and Bastia (1988a) have shown that the strong E2–DNA complexes that they have studied migrate at an anomalously slow rate in gel retardation assays when the sites are positioned toward the center of the fragment, and they interpret this as a bending of the DNA in the complex. By use of the data presented by Gartenberg and Crothers (1988) one can predict which of two given sequences is more likely to bend in a particular direction. The application of these rules to the mutants in the internal region of BS 1 results in the prediction that all of the mutants that decrease binding should hinder bending toward the minor groove relative to wild type. Mutant 7209, which increases binding, helps bending toward the minor groove. With reference to the model system in Figure 10, bending toward the minor groove in the middle of the site clearly brings the major grooves where the recognition nucleotides are located (in blue) closer together. The possibility that such bending aids the E2–DNA interaction might also explain why removal of certain bases in the middle of the sites 1 and 11 increases binding if such a modification improves the flexibility of the helix toward the minor groove. We would like to emphasize here, however, that this simple idea cannot account for all of our data. For example, BS 2

has a significantly higher affinity for E2 than does BS 11, although one would predict, using the Gartenberg and Crothers data, that BS 11 would bend more favorably than BS 2 for E2 binding. Still, other structural features might be involved in making BS 2 a strong binder.

In this study, we have characterized what we believe to be all of the E2 binding sites in the BPV-1 genome. All 17 locations with the sequence ACC N_6 GG/TT were tested and found to bind E2, but we were unable to detect binding to sites that deviated from this sequence. Five unexpected binding sites (ACC N_6 GTT) have been identified, three of which are interesting in that they are located near downstream promoters and could conceivably be implicated in putative regulatory loops in the viral life cycle. Evidence exists that supports the notion that two tandem binding sites might be necessary for E2-induced activation (Hawley-Nelson et al. 1988), although single sites are known to have repressive effects in some contexts (A. Stenlund et al., in prep.). Therefore, the two binding sites (one of which, BS 16, was not previously recorded) proximal to the P5 promoter may play a role in the E2 activation of the E2C repressor (Lambert et al. 1987). BS 16 and 17, which are within 150 nucleotides of each other, both bind E2 well, albeit 10-fold less effectively than the highest affinity sites. One may posit that as E2 levels increase within the cell, BS 16 and 17 become occupied, and then this would activate a transcriptional repressor. Similarly, the P3 promoter has been shown recently to serve as the start site for a related repressor (Choe et al. 1989) and it also has proximal binding sites (see Fig. 2 for the positions of these binding sites in the whole genome). Finally, we are impressed particularly with how weak a binding site is BS 15 and wonder if such sites have any biological role. This site is positioned just upstream of the P4 promoter, which probably starts transcription of at least one form of the full-length E2 product (Yang et al. 1985) and the E5 product (Prakash et al. 1988). It will be interesting to mutate these sites within the context of the whole viral genome to determine their phenotypes.

The E2 proteins of the various papillomaviruses have evolved considerably (Giri and Yaniv 1988), and, considering the extent of their divergence, it is remarkable that the proteins from heterologous sources can *trans*-activate the enhancers of one another in recombinant constructs (Phelps and Howley 1987; Hirochika et al. 1987; Thierry and Yaniv 1987). Although all papillomaviruses to date contain E2 binding motifs (Dartmann et al. 1986), their distribution and numbers appear to be not at all conserved. It will be of interest to learn how the relative affinities and locations of these sites in the various viruses evolve and how these parameters contribute to the life-styles of the respective virus families.

Materials and methods

Induction and purification of E2 protein

Two expression constructs were generated by fusing two different BPV-1 fragments into the expression vector pAR3040, which contains the gene *10* promoter of bacteriophage T7 (Studier and Moffatt 1986). The first fragment (whole E2) is the *Sph*I–*Bam*HI fragment (2617–4450), which lacks three amino acids of the predicted E2 sequence beginning with the first ATG of the open reading frame (ORF). The stop codon of the ORF is at nucleotide 3838. Thus, the predicted product of this expression construct is a fusion of 11 amino acids of gene *10* and 407 amino acids of E2. The second fragment cloned (short E2) is the *Nar*I–*Bam*HI fragment (2944–4450), which includes only the 3′-terminal two-thirds of the E2 ORF. This fusion is 11 amino acids of gene *10* and the carboxy-terminal 287 amino acids of BPV.

Bacterial growth and protein induction were performed as described by Studier and Moffatt (1986). The bacterial pellet from a 1-liter culture was resuspended in 30 ml of LS buffer [50 mM Tris (pH 7.5), 10 mM EDTA, 0.1 M NaCl, 20% sucrose, 1 mM PMSF]. After the addition of 100 mg of lysozyme, the suspension was incubated on ice for 20 min. Then cells were sonicated discontinuously for ~3 min total, and the sample was centrifuged at 3000*g* for 5 min to remove the cell debris. The supernatant was spun again at 13,000*g* for 20 min. The insoluble pellet was extracted several times with U2 dilution buffer [2 M urea, 50 mM Tris (pH 7.5), 1 mM $MgCl_2$, 1 mM EDTA, 0.9 M NaCl, 10% glycerol, 1 mM DTT]. The extract was diluted with Z(0.0)U2 buffer [25 mM HEPES (pH 7.8), 12.5 mM $MgCl_2$, 1 mM DTT, 20% glycerol, 0.1% NP-40, 2 M urea] to 0.1 M final salt concentration. The diluted extract then was mixed with 600 mg of salmon sperm DNA, incubated on ice for 10 min, and then centrifuged at 17,000*g* for 20 min. The supernatant was applied to a DNA affinity column (Kadonaga and Tjian 1986) that consisted of a ligated oligonucleotide of sequence (only one strand is given) GGTCAAACCGTCTTCGGTGCTCGAAAA, coupled to a Sepharose CL2B resin. The column was washed with 0.1 M KCl-Z(0.1)U2 and then eluted with 4 ml of 0.6 M KCl-Z(0.6)U2. Fractions of 120 μl were collected and the DNA-binding activity of each fraction was determined by the gel retardation assay.

DNase I footprint assay

Footprint reactions were carried out essentially as described previously (Jones et al. 1985). Competitor DNA was not included in the reactions.

Gel retardation assay

End-labeled DNA fragments (1–5 ng; 5000 cpm/reaction) were incubated for 30 min at room temperature with 1–5 μl of the affinity-purified protein [in Z(0.6)U2 buffer, see Induction and purification] in a 20-μl reaction with 10 mM Tris-HCl (pH 7.0), 1 mM $MgCl_2$, 1 mM DTT, 1 mM EDTA, and 5% glycerol; KCl was adjusted to a final concentration of 100 mM. The reactions were loaded onto a 10% polyacrylamide gel (acrylamide/*bis*, 29 : 1), 1 mm thick, in 300 mM glycine, 50 mM Tris, and 1 mM EDTA (final pH 8.7), and run at 9 V/cm for 5–10 hr at room temperature. The dried gel was exposed to X-ray film for 4–24 hr (exposures were made without enhancement screens when films were used for densitometry). When relative binding affinities of different binding sites were being compared, the reactions were run in parallel with identical amounts of DNA and protein.

Rate constants

For each of the sites to be tested, labeled DNA fragments carrying the site were incubated with protein as described for the gel retardation assay, except that the reaction was scaled up to

100 μl. For time zero, an aliquot was removed and loaded directly onto a prerun, nondenaturing gel (see Gel retardation assay). Afterwards, 5 μg of plasmid DNA (a 1000-fold molar excess) bearing the entire BPV-1 URR was added to the reaction (producing a volume increase of no more than 2%), and aliquots were removed and loaded on the gel at subsequent time points. Because it is the electric field in the gel that separates the competitor DNA from the bound DNA and thus marks the time point, the gel was left running as the samples for each time point were loaded.

Autoradiographs of the dried gels were scanned densitometrically. The half-life was determined from a semi-log plot of the data (shifted cpm over shifted cpm at time = zero versus minutes), and the dissociation rate constant, K_d, determined from the equation, $K_d = 0.693 / t_{1/2}$.

Relative equilibrium constants

To determine the equilibrium binding constants of a binding site relative to BS 1, end-labeled fragments of BPV that carried the site of interest were incubated with an end-labeled oligonucleotide standard carrying BS 1 and protein for 30 min (equilibrium is reached in <5 min as determined by analysis for time points). The sequence of the doubled-stranded oligonucleotide standard is given in the text. Lengths of the fragments tested for each binding site are as follows: BS 1, 210 bp; BS 2, 173 bp; BS 3, 68 bp; BS 4, 68 bp; BS 5, 121 bp; BS 6*, 120 bp; BS 7*, 90 bp; BS 8*, 95 bp; BS 9*, 157 bp; BS 10, 67 bp; BS 11, 127 bp; BS 12, 124 bp; BS 13, 240 bp; BS 14, 290 bp; BS 15, 142 bp; BS 16, 215 bp; BS 17, 131 bp (* indicates that the binding site was cloned into a pUC plasmid, and that the fragment contains some pUC polylinker sequences). The reactions were run on a nondenaturing gel as described for the gel retardation assay. The ratio of bound to free DNA was determined by scanning densitometry of the autoradiographs. Generally, three shifted bands appeared and all were included in the calculations. The relative equilibrium constants of the two binding sites were calculated from: $K_1 / K_2 = [C_1]/[D_1] / [C_2]/[D_2]$, where $[C]$ is the concentration of protein–DNA complex, and $[D]$ is the concentration of free DNA (Liu-Johnson et al. 1986).

Missing contact probing

These assays were performed as described by Brunelle and Schleif (1987) with minor modifications. DNA fragments with the approximate sizes of 70–120 bp that contained a single E2 binding site were end-labeled. For the G + A reaction, DNA fragments dissolved in 15 μl TE [0.01 M Tris (pH 8), 0.01 M EDTA] with 5 μg tRNA were mixed with 3 μl of 1 M formate (pH 2.0) and incubated at 37°C for 2 hr. The reaction was then ethanol-precipitated twice. The C + T reaction was performed by adding 30 μl of hydrazine to DNA fragments in 20 μl of ddH_2O). After incubation for 10 min at 37°C, the reaction was terminated by the addition of 50 μl of 1.5 M sodium acetate, 100 μg of yeast tRNA per ml, and by two successive ethanol precipitations (Maxam and Gilbert 1980). The incubation time for each reaction was varied according to different lengths of labeled DNA fragments. Premodified DNA fragments were reannealed in 0.2 M NaCl at 65°C for 2 hr. The gel retardation assay was carried out as described above, only with five times as much protein and 30 to 50 times more labeled fragment. Free and bound fragments were eluted from the gel using isotachtophoresis, and then treated with 100 μl of 1 M of piperidine for 30 min at 90°C. Then piperidine was removed. Equal amounts of the free and bound fragments were resolved on a denaturing polyacrylamide sequencing gel.

Site-directed mutagenesis

Point mutants were generated by standard methods, essentially as described by Kramer and Fritz 1987. Details will be published elsewhere.

Densitometry and computer graphics

The data for missing contact probing and kinetic studies of E2 binding were quantitated by densitometric scanning of the autoradiographs with Hoefer GS300 densitometer.

Three-dimensional DNA helical models that depict the missing contact results on BS 9 were produced by Stephen Holbrook and Fan Jiang (Lawrence Berkeley Laboratory) with an Evans & Sutherland PS340. Using the PSFRODO program (Jones 1978; Pflugrath et al. 1984), the DNA sequence of BS 9 was visualized according to the coordinates of idealized B-DNA. Nucleotides which are indicated to be the strong contact points were displayed with blue lines.

Acknowledgments

We wish to thank Dr. D. Bohmann for his generous advice on preparing the DNA-affinity column, Josh LaBaer for technical advice on relative equilibrium constant determination, and Peter Vaillancourt and Richard Harland for comments on the manuscript. We are grateful to S. Holbrook and F. Jiang for their valuable time and expertise in performing the three-dimensional computer graphics. Expression vectors pAR3040, pAR3039, and pAR3038 were provided generously by Drs. W. Studier and B. Moffatt. G.B. was supported by a predoctoral fellowship from the Office of Naval Research. This work was funded by research grants to M.B. from the National Institutes of Health, CA-42414 and CA-30490.

References

Anderson, J.E., M. Ptashne, and S.C. Harrison. 1987. Structure of the repressor–operator complex of bacteriophage 434. *Nature* **326:** 845–852.

Androphy, E.J., D.R. Lowy, and J.T. Schiller. 1987. Bovine papillomavirus E2 trans-activating gene product binds to specific sites in papillomavirus DNA. *Nature* **325:** 70–73.

Baumruker, T., R. Strum, and W. Herr. 1988. OBP100 binds remarkably degenerate octamer motifs through specific interactions with flanking sequences. *Genes Dev.* **2:** 1400–1413.

Broker, T.R. and M. Botchan. 1986. Papillomaviruses: Retrospecitives and prospectives. In *Cancer Cells* **4:** 17–36.

Brunelle, A. and R.F. Schleif. 1987. Missing contact probing of DNA–protein interactions. *Proc. Natl. Acad. Sci.* **84:** 6673–6676.

Choe, J., P. Vaillancourt, A. Stenlund, and M. Botchan. 1989. Bovine papilloma virus type 1 encodes two forms of a transcriptional repressor: structural and functional analysis of new viral cDNAs. *J. Virol.* (in press).

Costa, R.H., D.R. Grayson, K.G. Xanthopoulous, and J.E. Darnell, Jr. 1988. A liver-specific DNA binding protein recognizes multiple nucleotide sites in regulatory regions of transthyretin, antitrypsin, albumin, and simian virus 40 genes. *Proc. Natl. Acad. Sci.* **85:** 3840–3844.

Dartmann, K., E. Schwarz, L. Gissmann, and H. Zur Hausen. 1986. The nucleotide sequence and genome organization of human papillomavirus type II. *Virology* **151:** 124–130.

Di Maio, D. and K. Neary. 1989. The genetics of bovine papillomavirus type 1 papillomaviruses and human cancer. (ed. N. Pfister), CRC Press, Boca Raton, Florida.

Dostani, N., F. Thierry, and M. Yaniv. 1988. A dimer of BPV-1 E2 containing a protease resistant core interacts with its target. *EMBO J.* **7:** 3807–3816.

Gartenberg, M.R. and D.M. Crothers. 1988. DNA sequence determinants of CAP-induced bending and protein binding affinity. *Nature* **333:** 824–829.

Giri, I. and M. Yaniv. 1988. Structural and mutational analysis of E2 trans-activaating proteins of papillomaviruses reveals three distinct functional domains. *EMBO J.* **7:** 2823–2829.

Harrison, S.M., K.L. Gearing, S.-Y. Kim, A.J. Kingsman, and S.M. Kingsman. 1987. Multiple *cis*-active elements in the long control region of bovine papillomavirus type 1 (BPV-1). *Nucleic Acids Res.* **15:** 10267–10283.

Haugen, T.H., T.P. Cripe, G.D. Ginder, M. Karin, and L.P. Turek. 1987. Trans-activation of an upstream early gene promoter of bovine papillomavirus-1 by a product of the viral E2 gene. *EMBO J.* **6:** 145–152.

Hawley-Nelson, P., E.J. Androphy, D.R. Lowy, and J.T. Schiller. 1988. The specific DNA recognition sequence of the bovine papillomavirus E2 protein is an E2-dependent enhancer. *EMBO J.* **7:** 525–531.

Hirochika, H., T.R. Broker, and L.T. Chow. 1987. Enhancers and *trans*-acting E2 transcriptional factors of papillomaviruses. *J. Virol.* **61:** 259–260.

Hirochika, H., R. Hirochika, T.R. Broker, and L.T. Chow. 1988. Functional mapping of the human papillomavirus type 11 transcriptional enhancer and its interaction with the *trans*-acting E2 proteins. *Genes Dev.* **2:** 54–67.

Hogan, M.E. and R.H. Austin. 1987. Importance of DNA stiffness in protein–DNA binding specificity. *Nature* **329:** 263–266.

Johnson, P.F., W.H. Landschultz, B.J. Graves, and S.L. McKnight. 1987. Identification of a rat liver nuclear protein that binds to the enhancer core element of three animal viruses. *Genes Dev.* **1:** 133–146.

Jones, K.A., K.R. Yamamoto, and R. Tjian. 1985. Two distinct transcription factors bind to the HSV thymidine kinase promoter *in vitro*. *Cell* **42:** 559–572.

Jones, T.A. 1978. A graphics model building and refinement system for macromolecules. *J. Appl. Cryst.* **11:** 268–272.

Kadonaga, J.T. and R. Tjian. 1986. Affinity purification of sequence-specific DNA binding proteins. *Proc. Natl. Acad. Sci.* **83:** 5889–5893.

Koudelka, G.B., S.C. Harrison, and M. Ptashne. 1987. Effect of non-contacted bases on the affinity of 434 operator for 434 repressor and Cro. *Nature* **326:** 886–888.

Koudelka, G.B., P. Harbury, S.C. Harrison, and M. Ptashne. 1988. DNA twisting and the affinity of bacteriophage 434 operator for bacteriophage 434 repressor. *Proc. Natl. Acad. Sci.* **85:** 4633–4637.

Kramer, W. and H. Fritz. 1987. Oligonucleotide-directed construction of mutations via gapped duplex DNA. *Methods Enzymol.* **154:** 350–367.

Lambert, P.F., B.A. Spalhoz, and P.M. Howley. 1987. A transcriptional repressor encoded by BPV-1 shares a common carboxy-terminal domain with the E2 transactivator. *Cell* **50:** 69–78.

Liu-Johnson, H., M.R. Gartenberg, and D.M. Crothers. 1986. The DNA binding domain and bending angle of *E. coli* CAP protein. *Cell* **47:** 995–1005.

Maxam, A.M. and W. Gilbert. 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods Enzymol.* (part 1) **65:** 499–560.

McBride, A.A., R. Schlegel, and P.M. Howley. 1988. The carboxy-terminal domain shared by the bovine papillomavirus E2 transactivator and repressor proteins contains a specific DNA binding activity. *EMBO J.* **7:** 533–539.

McBride, A.A., J.C. Byrne, and P.M. Howley. 1989. E2 polypeptides encoded by bovine papillomavirus type 1 form dimers through the common carboxyl-terminal domain: Transactivation is mediated by the conserved amino-terminal domain. *Proc. Natl. Acad. Sci.* **86:** 510–514.

McClure, W. 1985. Mechanism and control of transcription in prokaryotes. *Annu. Rev. Biochem.* **54:** 171–204.

Mitchell, P.J. and R. Tjian. 1989. Transcriptional regulation in mammalian cells by sequence-specific DNA binding protein. (in press).

Moskaluk, C. and D. Bastia. 1987. The E2 'gene' of bovine papillomavirus encoded an enhancer binding protein. *Proc. Natl. Acad. Sci.* **84:** 1215–1218.

———. 1988a. DNA bending is induced in an enhancer by the DNA-binding domain of the bovine papillomavirus E2 protein. *Proc. Natl. Acad. Sci.* **85:** 1826–1830.

———. 1988b. Interaction of the bovine papillomavirus type 1 E2 transcriptional control protein with the viral enhancer: Purification of the DNA-binding domain and analysis of its contact points with DNA. *J. Virol.* **62:** 1925–1931.

Nelson, H.C.M., J.T. Finch, B. Luisi, and A. Klug. 1987. The structure of an oligo(dA) oilgo(dT) tract and its biological implications. *Nature* **330:** 221–226.

Pabo, C.O. and R.T. Sauer. 1984. Protein–DNA recognition. *Annu. Rev. Biochem.* **53:** 293–3321.

Pfeifer, K., T. Prezant, and L. Guarente. 1987. Yeast HAP1 activator binds to two upstream activation sites of different sequences. *Cell* **49:** 19–27.

Pflugrath, J.W., M.A. Saper, and F.A. Quiocho. 1984. New generation graphics system for molecular modeling. In *Methods and applications in crystallographic computing* (ed. S. Hall and T. Ashira), pp. 404–407. Clarendon Press, Oxford.

Phelps, W.C. and P.M. Howley. 1987. Transcriptional transactivation by the human papillomavirus type 16 E2 gene product. *J. Virol.* **6:** 1630–1638.

Prakash, S.S., B.H. Horwitz, T. Zibello, J. Settleman, and D. Di Maio. 1988. Bovine papilloma virus E2 gene regulates expression of the viral E5 transforming gene. *J. Virol.* **63:** 3608–3613.

Ptashne M. 1987. *A genetic switch*. Cell press. California & Blackwell Publishing, Palo Alto, California.

———. 1988. How eukaryotic transcriptional activators work. *Nature* **335:** 683–689.

Spalholz, B.A., Y.C. Yang, and P.M. Howley. 1985. Transactivation of the bovine papilloma virus transcriptional regulatory element by the E2 gene product. *Cell* **42:** 183–191.

Spalholz, B.A., P.F. Lambert, C.L. Yee, and P.M. Howley. 1987. Bovine papillomavirus transcriptional regulation: Localization of the E2-responsive elements of the long control region. *J. Virol.* **61:** 2128–2137.

Stenlund, A., G.L. Bream, and M.R. Botchan. 1987. A promoter with an internal regulatory domain is part of the origin of replication in BPV-1. *Science* **263:** 1666–1671.

Studier, F.W. and B.A. Moffatt. 1986. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.* **189:** 113–130.

Thierry, F. and M. Yaniv. 1987. The BPV-1 E2 protein can be

either an activator or a repressor of the HPV 18 regulatory region. *EMBO J.* **6:** 3391–3397.

Yang, Y.C., A.B. Spalholz, M.S. Rabson, and P.M. Howley. 1985. Disassociation of transforming and transactivation functions for bovine papilloma virus type 1. *Nature* **318:** 575–577.

Specific recognition nucleotides and their DNA context determine the affinity of E2 protein for 17 binding sites in the BPV-1 genome.

R Li, J Knight, G Bream, et al.

Genes Dev. 1989, **3:**
Access the most recent version at doi:10.1101/gad.3.4.510