J. Biochem. 121, 138-144 (1997)

A Mutation Study of the DNA Binding Domain of Human Papillomavirus Type11 E2 Protein[1]

Takashi Matsumoto,*,[2] Nobuhiko Nakashima,† Kenji Takase,* Hirohiko Hirochika,* and Hiroshi Mizuno*

Department of Molecular Biology, National Institute of Agrobiological Resources, Tsukuba, Ibaraki 305; and †Department of Insect Physiology and Behavior, National Institute of Sericultural and Entomological Science, Tsukuba, Ibaraki 305

Received for publication, September 12, 1996

A site-specific mutation study was performed on the C-terminal domain, containing a cloned DNA binding region, of the human papillomavirus type11 (HPV11) E2 protein to determine the specific properties of residues directly involved in the DNA binding. The effect of a point mutations on the DNA binding was assessed by means of a gel mobility shift assay. The mutagenesis was concentrated on the residues in the third helix from the N-terminal, that is known as the "recognition helix," in the crystal structure of the bovine papillomavirus (BPV) E2 protein. Most point mutations caused a great decrease in the DNA binding activity. The leucine repeat in the DNA binding region was proved not to be a leucine prerequisite, as the leucines could be substituted by valine without significant loss of the DNA binding ability. Substitution of Leu for Glu caused a significant decrease in the DNA binding, indicating that the hydrophobicity of the residue at this position is important. The results suggest that the individual contribution of each amino acid residue in the DNA binding region is essential for the DNA binding.

Key words: DNA-binding protein, papillomavirus, site-directed mutagenesis.

Recently, many protein factors that regulate gene expression through binding to the promoter region have been identified (*1*). These *trans*-acting factors are structurally classified into several groups, such as the helix-turn-helix protein, bZIP protein, and zinc finger protein types. The papillomavirus E2 protein is a *trans*-acting protein that binds to the transcriptional enhancer sequences within the viral genome and stimulates the transcription of viral genes including viral oncogenes (*2*). The dimer forms of E2 proteins within the range of 45-48 kDa were shown to bind to DNA (*3, 4*). The minimal binding sequence of DNA is semi-palindromic $ACCN_6GGT$ (*5, 6*). Cooperativity among multiple E2-binding sites in the HPV genome for activating transcription has been reported (*7, 8*).

Recently, it was revealed that the E2 protein is required for papillomavirus DNA replication *in vitro*, indicating that this protein is an important factor for both viral DNA replication and gene transcription (*9, 10*).

Studies on the relationship between protein structure and function revealed that the E2 protein contains two functional domains (*11*). The N-terminal region of about 200 amino acids long has a function in *trans*-activation, and the C-terminal (*ca.* 90 amino acids) region is responsible for the specific DNA binding and dimerization (*5, 12, 13*). The two regions are separated by a hinge (non-conservative) region (*14, 15*). Chin *et al.* (*12*) cloned a part of the hinge region and the entire C-terminal DNA binding region of the HPV11 E2 protein, and called it the E2-C protein (186 amino acid residues).

The crystal structure of the DNA binding domain of the bovine papillomavirus (BPV) E2 protein complexed with its target DNA (*16*) revealed that the region contained two α-helices, and the N-terminus proximal helix, the "recognition helix," was shown to be predominantly involved in the DNA recognition/binding. Multiple interactions among the amino-acid side chains of the "recognition helix" and either the base or the backbone phosphate of the target DNA were assumed.

To assess the function of each amino acid residue by biochemical analysis, we have generated, by site-specific mutagenesis, a series of amino-acid substitutions in the DNA binding domain of the HPV11 E2 protein. The results of gel mobility shift assays indicated that many amino acid residues in the recognition helix are involved in the DNA binding, as in the case of the BPV E2 protein.

MATERIALS AND METHODS

Materials—The reagents and *Escherichia coli* host strain used for *in vitro* mutagenesis, and [γ-^{32}P]ATP were from Amersham. S-Sepharose and poly(dI-dC) were from Pharmacia. Triton X-100, NP-40, antipain and leupeptin were from Sigma. Restriction enzymes and T_4-polynucleotide kinase were from Takara Shuzo.

In Vitro Mutagenesis—The truncated HPV11 E2 gene used in this study was derived from plasmid pEVE2C,

[1] This work was supported by grants from the Ministry of Agriculture, Forestry and Fisheries of Japan.

[2] To whom correspondence should be addressed. Fax: +81 298 38 7408, e-mail: mat@abr.affrc.go.jp, Tel: +81 298 38 7014

Abbreviations: HPV11, human papillomavirus type11; BPV, bovine papillomavirus; CRPV, cottontail rabbit papillomavirus.

which was designed to overexpress the E2-C protein in *E. coli*. The construction of pEVE2C was described in the literature (*12*). A 1.2 kb *Bam*HI fragment containing the DNA binding domain of the HPV11 E2 gene was inserted into the M13 phage vector, mp18. *In vitro* mutagenesis was performed by the method of Eckstein using an Amersham *in vitro* mutagenesis kit (*17*). According to the manufacturer's instructions, oligodeoxyribonucleotide primers (listed in Table I) were hybridized with single-stranded phage DNA templates. Double-stranded circular DNA was synthesized with the Klenow fragment and T4 DNA ligase. *E. coli* TG1 was transformed with the ligated DNA and white plaques were picked up. Single and double stranded DNAs were purified from mutagenized M13 phage-infected *E. coli* cultures. Mutant E2 genes were confirmed by sequencing by the dideoxynucleotide chain-termination method. Double mutants (Table II; L107VL114V and L121VL128V) were created by performing second mutagenesis of the single-site mutants. The mutant E2 fragments were ligated into the pEVE2C expression vector and the ligated DNA was introduced into the *E. coli* RR1 strain containing pRK248cIts. The bacteria were first incubated at 30°C for 2 h, and then the E2-C protein was expressed on incubation at 42°C for 2 h (*12*).

Purification of Mutated E2-C Proteins from E. coli—The bacterial pellet was washed once with buffer A (50 mM Tris-HCl, pH 7.8, 5 mM EDTA, 2 mM DTT, 0.1 mM PMSF) and then resuspended in buffer A containing 5 mg/ml of lysozyme. The cell suspension was kept on ice for 20 min, and then Triton X-100 (final conc., 0.1%) was added, and the cells were disrupted by sonication. Insoluble materials were precipitated by centrifugation. The precipitate was resuspended in buffer C (50 mM Tris-HCl, pH 8.8, 2 mM DTT, 1 mM PMSF) containing 8 M urea as a denaturant. The suspension was briefly sonicated and then centrifuged. The E2-C protein was in the urea-soluble fraction. The supernatant was successively dialyzed against buffer C containing 6, 4, 2, and 0 M urea. This method yielded enough of the soluble E2-C protein for the DNA binding assay (about several hundred microgram protein from 1 g wet weight of *E. coli*), though much of the E2-C protein was re-precipitated upon refolding (especially during 0 M urea dialysis). Debris was removed by centrifugation, and the supernatant was loaded onto a S-Sepharose column equilibrated with buffer S (10 mM Tris-HCl, pH 7.8, 1 mM DTT, 0.5 mM EDTA, 0.5 mM PMSF, 0.05% NP-40, 10% glycerol, 0.1 μg/ml antipain, 0.1 μg/ml leupeptin) containing 10 mM NaCl. The column was then washed twice with buffer S containing 10 mM NaCl and 100 mM NaCl. The bound proteins were eluted with a linear gradient of NaCl, from 100 to 500 mM. The E2-C protein was eluted at about 150 mM NaCl. The fraction in which the E2-C protein was most abundant was used for the binding assay. The wild type E2-C protein was further purified by DNA-affinity chromatography, in which the E2 recognition sequence was immobilized on the matrix (*6*).

Gel Mobility Shift DNA Binding Assay—Since the E2-C protein contains both DNA binding and dimerization regions, it is believed to form a dimer and to bind to a semi-palindromic motif, ACCN$_6$GGT (*6*). Two probes for the gel-shift assay were prepared in this study. To construct a single E2 (dimer) binding site probe [(RS)$_1$], an oligodeoxyribonucleotide, 5′ GGTCGACA*ACCGGTTTC-GGT*TTCGACC 3′, and its complementary sequence were synthesized. ACC and GGT in italics are the minimal binding sites for the E2 protein. A multiple E2 binding site substrate was constructed from the pCAT-SN5 plasmid, which was originally constructed to evaluate the transcription stimulation by the E2 protein *in vivo* (*6*). *Hin*dIII digestion of this plasmid gave a 115 bp fragment [(RS)$_5$] which contains a five times tandemly repeated sequence, 5′ TCGACA*ACCGGTTTC*GGTT 3′ (the spacing between minimal recognition cores, ACCN$_6$GGT, is 7 bp).

The gel mobility shift assay was performed by the method of Garner and Revzin (*18*) with a modification. The 5′-terminus of each DNA probe was labeled with T$_4$-polynucleotide kinase and [γ-^{32}P]ATP, and the labeled probes were purified by polyacrylamide gel electrophoresis. The amount of E2-C protein used for the binding assay was estimated from the intensity of the protein band stained with Coomassie Brilliant Blue on the SDS-PAGE gel, compared with that of the protein marker standard. Proteins were incubated with the labeled probe for 30 min at 30°C, and then the reaction mixtures was applied to a 5% polyacrylamide gel equilibrated with 40 mM Tris-acetate–1 mM EDTA buffer (pH 7.5) for 2 h at 200 V at 4°C. The gel was then treated with 7% TCA, dried and autoradiographed. The regions of the dried gel corresponding to the retarded bands were cut out for Cherenkov counting.

RESULTS

Bacterial Expression and Partial Purification of the Wild Type and Mutant E2-Cs—Figure 1a shows expression of the E2-C protein in *E. coli*. In every bacterial culture harboring the E2-C plasmid, a *ca.* 23 kDa E2-C protein was expressed specifically under the induction conditions (at

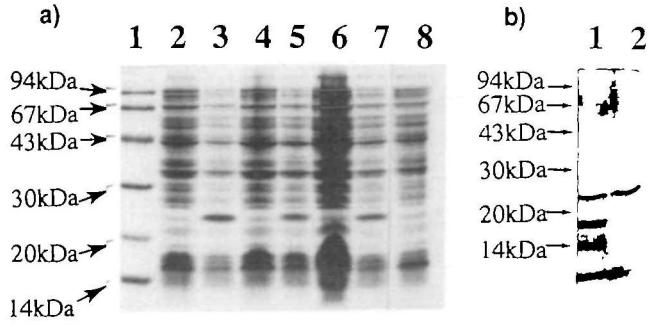


Fig. 1. **(a) Bacterial expression of the wild type and mutant E2-C proteins.** Coomassie Blue staining of the 12.5% SDS-PAGE patterns of the L114V mutant (lanes 2 and 3), L121V mutant (lanes 4 and 5), wild type E2-C protein (lanes 6 and 7), and *E. coli* cells which did not harbor pEVE2C (lane 8). The molecular mass markers are shown in lane 1: phosphorylase *b* (94 kDa), serum albumin (67 kDa), ovalbumin (43 kDa), carbonic anhydrase (30 kDa), trypsin inhibitor (20 kDa), and α-lactalbumin (14 kDa). Bacteria were grown at 30°C for 2 h and then at either 30°C (lanes 2, 4, and 6) or 42°C (lanes 3, 5, 7, and 8) for 2 h. One hundred microliter aliquots of the cultures were pelleted by centrifugation, and resuspended in 20 μl of 1 × SDS-PAGE sample buffer containing β-mercaptoethanol. **(b) Purification of the recombinant E2-C protein.** Silver-staining, after 12.5% SDS-PAGE, of the S-Sepharose peak fraction (lane 1, 10 ng) and the affinity-purified wild type E2-C protein (lane 2, 14 ng) is shown. The molecular mass markers are the same as in (a) (marker lane not shown).

42°C) (compare lanes 2 and 3, 4 and 5, 6 and 7), and was only expressed in *E. coli* containing pEVE2C (lane 8, as a negative control). Figure 1b shows the SDS-PAGE pattern of the purified wild type E2-C protein (lane 1, S-Sepharose purified; lane 2, affinity purified). The molecular weight of the E2-C protein estimated from the relative mobilities is 23 kDa. All of the mutant E2-C proteins showed the same elution profile on S-Sepharose column chromatography (data not shown), and the same protein pattern as the native E2-C protein (lane 1). We used the S-Sepharose fraction pool as the source for the gel mobility shift assay because some of the mutant E2-C proteins were expected to show weak or no binding activity towards the recognition sequence DNA that was used for the affinity chromatography. The lysozyme present in this fraction (the band corresponding to a molecular weight of 14 kDa in lane 1) showed no effect on the binding assay under the experimental conditions (data not shown).

Gel-Shift Analysis of Conserved Leu Mutants—The HPV11 E2 protein has a hepted leucine repeat (L107, L114, L121, and L128; the amino acid residues are numbered from the first methionine of the E2-C protein), which is highly conserved in HPVs (*19, 20*), and it resided in the recognition helix region in the C-terminal portion of the E2 protein.

We created four single Leu to Val mutants (see Table I; L107V, L114V, L121V, and L128V) and examined them for DNA binding activity. The results are shown in Fig. 2a. Using the single binding site substrate [$(RS)_1$], L107V, L114V, and L128V were found to show almost the same binding activity as that of the wild type E2-C protein (Fig. 2a, lanes 2, 3, 4, and 6). L121V gave a fainter band (lane 5),

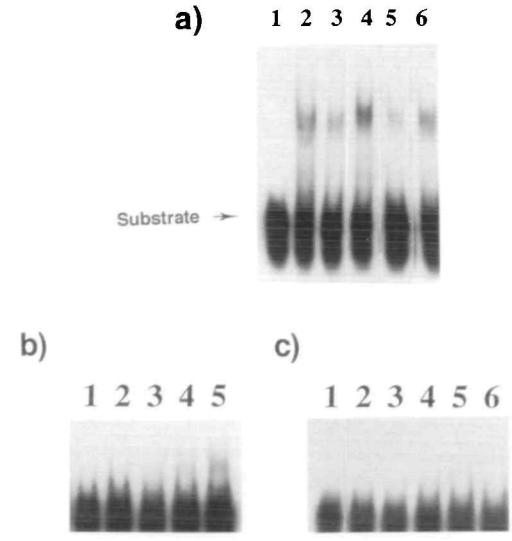


Fig. 2. **Gel mobility shift analysis of mutant E2-Cs.** (a) The leucine to valine mutant E2-Cs. The 20 μl reaction mixture contained 10 mM Tris-HCl (pH 7.5), 50 mM NaCl, 10% glycerol, 1 mM EDTA, 1 mM DTT, E2-C protein, and 1 μg of poly(dI-dC). 0.75 fmol of the $(RS)_1$ probe was used as a substrate. An autoradiogram of a 5% PAGE gel is shown. The arrow at the left indicates the unbound substrate. The substrates was incubated with only buffer (lane 1), wild type E2-C (10 ng; lane 2), L107V mutant (10 ng; lane 3), L114V mutant (10 ng; lane 4), L121V mutant (10 ng; lane 5), or L128V mutant (10 ng; lane 6). (b) The L114E mutant. 0.1 ng (lane 1), 0.5 ng (lane 2), 2 ng (lane 3), 5 ng (lane 4), or 10 ng (lane 5) of the mutant E2 protein was added to the reaction mixture. (c) The L121E mutant. 0 ng (lane 1), 0.1 ng (lane 2), 0.5 ng (lane 3), 2 ng (lane 4), 5 ng (lane 5), or 10 ng (lane 6) of the mutant E2 protein was added to the reaction mixture.

TABLE I. **Design of the primers used for *in vitro* mutagenesis.** The top line is the wild type HPV11 E2 gene sequence corresponding to the region from Ala101 to Glu130. The boxed amino acids are the conserved hepted leucines. The underlined nucleotides are the mutagenized bases. Changes of amino acids are simplified as L107V (Leu107 changed to Val), for example.

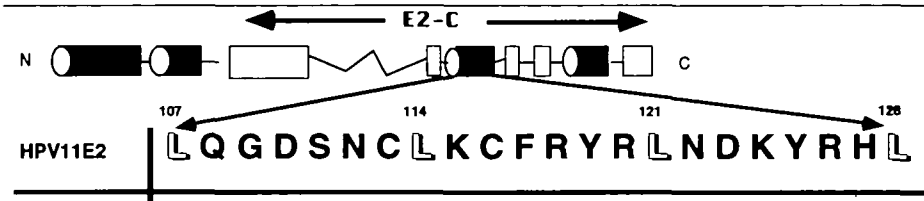
	101									110										120										130
E2-C	GCT	ACG	CCT	ATA	GTG	CAA	CTG	CAA	GGT	GAT	TCC	AAT	TGT	TTA	AAA	TGT	TTT	AGA	TAT	AGA	CTG	AAT	GAC	AAA	TAT	AGA	CAT	TTG	TTT	GAA
	Ala	Thr	Pro	Ile	Val	Gln	Leu	Gln	Gly	Asp	Ser	Asn	Cys	Leu	Lys	Cys	Phe	Arg	Tyr	Arg	Leu	Asn	Asp	Lys	Tyr	Arg	His	Leu	Phe	Glu
L107V					ATA	GTG	CAA	GTG	CAA	GGT	GAT																				
N112A									CAA	GGT	GAT	TCC	GCT	TGT	TTA	AAA	TGT														
C113A										GGT	GAT	TCC	AAT	GCT	TTA	AAA	TGT	TTT													
L114V												TCC	AAT	TGT	GTA	AAA	TGT	TTT													
L114E												TCC	AAT	TGT	GAA	AAA	TGT	TTT													
K115E													AAT	TGT	TTA	GAA	TGT	TTT	AGA												
K115T													AAT	TGT	TTA	ACA	TGT	TTT	AGA												
C116A													AAT	TGT	TTA	AAA	GCT	TTT	AGA	TAT	AGA										
R118A																	AAA	TGT	TTT	GCA	TAT	AGA	CTG								
Y119A																	AAA	TGT	TTT	AGA	GCT	AGA	CTG	AAT	GAC						
R120A																				TTT	AGA	TAT	GCA	CTG	AAT	GAC					
L121V																					AGT	TAT	AGA	GTG	AAT	GAC	AAA				
L121E																					AGT	TAT	AGA	GAG	AAT	GAC	AAA				
N122A																						TAT	AGA	CTG	GCTGAC	AAA	TAT				
K124A																							CTG	AAT	GAC	GCA	TAT	AGA	CAT		
R126A																								GAC	AAA	TAT	GCA	CAT	TTG	TTT	
L128V																									TAT	AGA	CAT	GTG	TTT	GAA	TTA

and the radioactivity incorporation into the shifted band amounted to 65% of that of wild type E2-C. In summary, it was apparent that all the Leu to Val mutants except L121V exhibit substantial DNA binding activity, and even L121V had some activity. Therefore, it was assumed that this hepted leucine repeat in the E2-C DNA binding domain was not a Leu prerequisite. We also constructed double mutants (L107VL114V and L121VL128V), but these mutant E2-Cs also showed good DNA binding activity. Next we tried to induce more drastic changes in the chemical nature of the conserved leucine residues. Substitution of leucine by glutamic acid leads to a *ca.* 20% decrease in the residue volume (*21*), and a change in the nature of the amino acid from hydrophobic to negatively charged, hydrophilic. The results obtained with these mutants on gel-shift assaying are shown in Fig. 2, b and c, as well as Table II. Both mutants, L114E and L121E, showed greatly reduced DNA binding activity (Fig. 2, b and c), with the $(RS)_1$ probe, indicating that the hydrophobic nature of the amino acids at these positions may be important for the E2-C protein to bind to its target DNA. The results described above are summarized in Table II [$(RS)_1$ column].

Gel-Shift Analysis of Other Mutants with the Monomer Probe [*$(RS)_1$*]—We carried out detailed scanning mutagenesis of the recognition helix of the HPV E2 protein, that contains the hepted leucine repeat, in order to assess the role of each amino acid residue in this region by biochemical analysis. Most residues from N112 to R126 were changed to alanine independently, except that K115 was substituted by glutamic acid or threonine (K115E and K115T; see Table II). No further mutants were constructed for the leucines.

Up to 10 ng of the partially purified mutant E2-C proteins was used for the gel-shift assay, with RS_1 as a probe. The results are summarized in Table II. Lys124 is located near the C-terminal end of the "recognition helix." Substitution of this lysine by alanine had no effect on the E2-C binding activity. Residue 126 was originally reported to be Lys, which was deduced from codon AAA (*22*), but our nucleotide sequencing analysis showed this codon constantly to be AGA, which corresponds to Arg instead of Lys, although the nature of the amino acid (basic) was unchanged by the replacement at this position. Furthermore, Arg is at this position in the case of the BPV E2

TABLE II. **Summary of the DNA binding activities of E2-C mutants on gel shift analysis.** The schematic secondary structure of the E2 protein (the E2-C region is indicated) is presented above left (from Ref. *14*). In the figure, the black cylinders are α-helices and the white boxes are β-strands. The amino acids are numbered from the N-terminal of the E2-C protein. The mutagenized residues and destined amino acids are shown in the table. The highlighted leucines are the leucine hepted repeat. $(RS)_1$ represents the binding site monomer, and $(RS)_5$ represents the pentamer substrate (described under "MATERIALS AND METHODS"). Relative binding activities are shown compared with that of the wild type E2-C protein.



		DNA binding Activity	
		$(RS)_1$	$(RS)_5$
HPV11E2	L Q G D S N C L K C F R Y R L N D K Y R H L	+++	+++
L107V	V	+++	+++
N112A	A	-	-
C113A	A	-	+
L114V	V	++++	+++
L114E	E	+	++
K115E	E	-	-
K115T	T	-	+
C116A	A	-	-
R118A	A	-	-
Y119A	A	-	-
R120A	A	-	-
L121V	V	++	+++
L121E	E	-	-
N122A	A	-	-
K124A	A	+++	+
R126A	A	++++	++
L128V	V	+++	+++
L107VL114V	V V	+++	+++
L121VL128V	V V	+++	+++

protein. Alanine substitution of this Arg (R126A) led to *ca.* 2.5-5 times higher activity (the positive effect). We also changed most of the residues from N112 to N122, exceptions being L114, K115, F117, and L121, to alanine (N112A, C113A, C116A, R118A, Y119A, R120A, and N122A). Surprisingly, they all showed no or only faintly detectable binding activity with the $(RS)_1$ probe, while the wild type E2C protein, used in the same range, showed substantial DNA binding.

Lys115 was assumed to be involved in the direct interaction with the bases of the target DNA (*23*). So we changed the positive charge of Lys to a negative charge (K115E) or non-charged but hydrophilic nature (K115T). Neither mutant protein showed detectable DNA binding, suggesting this Lys is also important for the DNA binding of the E2-C protein. These results indicate that multiple interactions among the amino acid residues and DNA are important for recognition of the E2-C protein. The results described above are summarized in Table II [$(RS)_1$ column].

Gel-Shift Assay with a Pentamer Probe [*$(RS)_5$*]—Multiple copies of E2 target sequences are found in the papillomavirus genome, and binding to two adjacent sites is required to activate a high level enhancer function (*6*), and cooperative binding of the E2 protein to adjacent sites was observed in *in vitro* studies (*7*). Therefore, in addition to the RS_1 probe, the DNA binding activities of the mutant E2-C proteins were evaluated by means of the gel mobility shift assay using a pentamer [$(RS)_5$]; hereafter, $(RS)_5$ is referred to as the S probe (115 bp, see "MATERIALS AND METHODS") to determine if site-specific mutagenesis had any effect on the cooperativity in DNA binding. Figure 3a shows the band shift pattern of the S probe with the purified wild type E2-C protein as a positive control, and Fig. 3b shows the result with the L114E mutant. In the figures, band O corresponds to the position of the probe in the unbound state (S), and four distinct shift bands can be seen in Fig. 3b (band A, [E2C dimer]$_1$S; band B, [E2C dimer]$_2$S; band C, [E2C dimer]$_3$S and band D, [E2C dimer]$_5$S, we assumed, respectively). We could not detect a band between bands C and D. Excess the E2-C protein led to aggregation of the protein-DNA complex, which stayed in the gel slot. In Fig. 3a, one can detect only a smear bandshift between bands A and D. We do not know the reason for this smear bandshift, but since binding of the active E2 protein to a multimer substrate causes DNA bending (*24*), we have assumed that a mixture of multiple bending conformations would give shifted bands exhibiting multiple mobilities.

All the Leu to Val mutants, either single or double, L107V, L114V, L121V, L128V, L107VL114V, and L121VL128V, showed the same bandshift patterns as those of the wild type E2-C protein (data not shown). Even the L121V mutant, which showed reduced activity with the $(RS)_1$ substrate, exhibited substantial DNA binding activity with the S substrate. Therefore, it was concluded that the Leu to Val substitution at these positions did not influence the binding to the multimer probe.

The results for the L114E and L121E mutants showed a good correlation to those as to $(RS)_1$ binding. Up to 10 ng of L114E gave only a [E2C dimer]$_3$S band (Fig. 3b), indicating that L114E exhibits only weak binding cooperativity in the binding to the enhancer probe. We observed no band shift with up to 10 ng of L121E (Fig. 3c), indicating that the hydrophobicity at this position is critical for the DNA binding.

The negative results with the pentamer substrate for N112A, K115E, C116A, R118A, Y119A, R120A, and N122A were the same as those with the $(RS)_1$ probe (data not shown). C113A (Fig. 3d) and K115T (Fig. 3e), which showed no detectable activity on RS_1 analysis, were found to be active in this assay although the activity was relatively low (we could detect only monomer [E2C dimer]$_1$S in K115T, and the bandshift pattern showed no cooperativity. But for C113A, we could detect a smear gelshift, which indicated cooperativity). The difference in the sensitivity of the multimer probe assay may be due to the cooperativity of the E2-C binding as to the tandem repeat of the binding core sequence. K124A was active as in the $(RS)_1$ binding assay, but less active with the S substrate (only to the [E2C dimer]$_2$S was observed, result not shown). R126A, which showed higher activity than the wild type E2-C protein with the $(RS)_1$ probe, showed substantially active DNA binding,

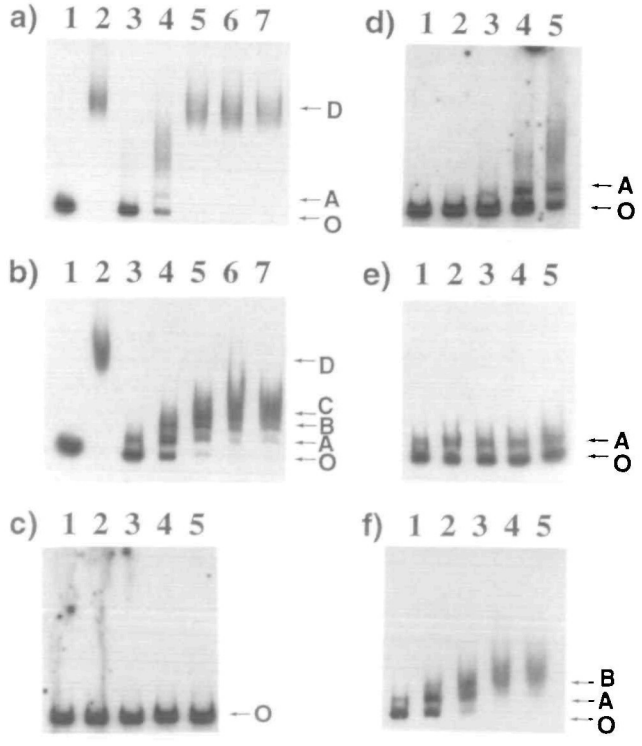


Fig. 3. **Gel mobility shift analysis of E2-C mutants using the $(RS)_5$ substrate.** 1.3 fmol of γ-^{32}P-labeled substrate was used. The contents of the assay mixtures were the same as in the experiments in Fig. 2. Autoradiograms of 5% PAGE gels are shown. (a) Titration of the pentamer substrate with the wild type purified E2-C protein. 0 ng (lane 1), 0.1 ng (lane 3), 0.5 ng (lane 4), 2 ng (lane 5), 5 ng (lane 6), or 10 ng (lanes 2 and 7) of the E2-C protein was added in the reaction mixture. (b) L114E mutant. 0.1 ng (lane 3), 0.5 ng (lane 4), 2 ng (lane 5), 5 ng (lane 6), or 10 ng (lane 7) of the L114E mutant E2-C was added. Reactions with the substrate only (lane 1) and the substrate with 10 ng of wild type E2-C (lane 2) were also performed as positive and negative controls, respectively. Each band in the autoradiogram corresponds to free substrate (O), one E2-C dimer bound (A), two E2-C dimers bound (B), three E2-C dimers bound (C), or five E2-C dimers bound (D). (c) L121E mutant. (d) C113A mutant. (e) K115T mutant. (f) R126A mutant. In each figure, 0.1 ng (lane 1), 0.5 ng (lane 2), 2 ng (lane 3), 5 ng (lane 4), or 10 ng (lane 5) of the mutant E2-C protein was added.

but the bandshift was only up to the [E2C dimer]$_3$S (Fig. 3f), indicating weak cooperativity. The results obtained with the pentamer substrate are summarized in Table II [$(RS)_5$ column]. In summary, there is a strong correlation between the results obtained with $(RS)_1$ and those with $(RS)_5$.

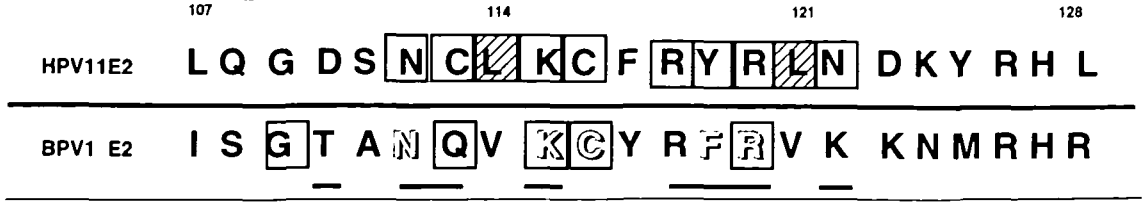
DISCUSSION

The C-terminal domain of the papillomavirus E2 protein has a hepted repeat of leucines or hydrophobic amino acid residues. Sequence alignment of the bovine, human, and cottontail rabbit papillomavirus (BPV, HPV, and CRPV) E2 proteins (*13*) revealed that the portion that includes two hypothetical α-helices (3rd and 4th ones in the E2 protein) is important for either DNA binding or transcriptional activation, or both (*14*). Several amino acids are highly conserved in these regions, suggesting that these residues are important for the biological function of the E2 protein. These regions are also responsible for dimerization of the E2 monomer (*4*), and we were tempted to assume that this region has a leucine zipper structure. On the other hand, the crystal structure of the DNA binding/dimerization domain (85 amino acids) of the bovine papillomavirus E2 protein (*16*) showed this region has a novel three-dimensional structure. A 18aa recognition helix (G109-R126 in the case of human papillomavirus E2) directly interacts with DNA, and a barrel structure consisting of many anti-parallel β-sheets outside this helix forms a strong dimer interface between E2-C monomers. Recently, furthermore, the solution structure of the DNA binding domain of the HPV31 E2 protein (not as a complex with DNA) was analyzed by NMR spectroscopy (*25*). Based on the results, the overall folding was found to be similar to the crystal structure of BPV E2. Therefore it is reasonable that we use the crystal structure of BPV E2 as a model of HPV E2. Based on the results of these structural analyses, in this study, we performed systematic mutagenesis for further understanding of the structure-function relationship of E2 proteins.

There have been several mutation studies on the BPV (*3, 11, 14*), CRPV (*14*), and HPV E2 proteins (*26*), but most of the results were based on insertion/deletion mutagenesis [except in one study by Prakash *et al.* (*23*), and one by Brokaw *et al.* (*27*)]. Our present study was a detailed mutagenesis one on the recognition helix in the HPV11 E2-C protein. The results of this study are schematically presented in Table III. Single or double leucine to valine mutants showed substantial binding activity toward the target DNA (Table II). This indicates that these leucines are not involved in leucine zipper structure, because in the case of rat liver C/EBP (*28*), which was shown to form a leucine zipper, a single substitution of critical Leu to either Ile, Val, or Met led to a dramatic loss of DNA binding activity. And in the case of C/EBP (*29*) and mammalian Jun (*30*), successive Leu to Val double mutants exhibited a great decrease in the binding activity toward the Fos protein and thus the binding to the AP-1 site was abolished. Nevertheless, the hydrophobic nature of this position may be important, since the introduction of an acidic charge (L114E and L121E) abolished some DNA binding activity, and the hydrophobic nature is conserved among papillomaviruses. We do not know the exact reason for the importance of the hydrophobicity, but from the crystal structure of BPV E2, we can assume that these two residues reside on opposite sides of the "DNA-contacting face." Therefore, they might form a "hydrophobic core" to stabilize the total structure of the E2 protein.

Then results of scanning mutagenesis of the recognition helix showed that most of the substitution mutants as to N112-N122 showed reduced or completely abolished DNA binding activity. There seems to be cooperative DNA binding through many residues in this helical region. Crystallographic data on BPV E2 indicated that N112, K115, C116, and Y119 (F119 in the case of BPV) could interact with the DNA base regions, and mutation of these residues (N112A, K115E, K115T, C116A, and Y119A) indeed abolished this activity (except in the case of K115T). C113 (Q113 in BPV) and R120 were considered to interact with the DNA backbone phosphate. The C113A and R120A mutants showed reduced activity, suggesting these residues are important for DNA binding. No direct interaction between N122 (K122 in BPV) and DNA was observed, although some interactions *via* water molecules were suggested (*16*). So the results for the N122A mutant, which lost its activity, are remarkable. The C-terminal proximal rim of the helical region of the E2 protein (K124-R126), which was not expected to interact with DNA, was rather insensitive to substitutions in terms of DNA binding. The structural changes induced by our mutagenesis should not affect the dimer interface, so the resultant negative DNA binding may indicate this region is not involved in the interactions between the E2 dimer and DNA. Prakash *et al.* (*23*) obtained DNA binding-defective mutants of BPV E2 on random mutagenesis. Their mutants had either a Gly109 to Val, Gln113 to Leu, Lys115 to Met,

TABLE III. **Comparison of the results of this study with those of previous studies on BPV E2.** The protein sequences of leucine107 to leucine128 of HPV E2-C and the corresponding region of BPV E2 are shown. The boxed residues are those involved in DNA binding identified in site-specific mutation studies [this study for HPV E2 and the study of Prakash *et al.* (*23*) for BPV E2]. The hatched boxes in HPV E2 are leucines114 and 121. A change in either of these leucines to glutamic acid caused a great loss of DNA binding activity. The shaded residues are those involved in base recognition, and the underlined residues are those involved in backbone phosphate interactions identified on crystal structure analysis [Hedge *et al.* (*16*)].



Cys116 to Arg, or Arg120 to Leu substitution (the numbers are adjusted as to the HPV11 E2-C sequence). Comparing our results with these data (see Table III), K115, C116, and R120 were proved to be important in both cases. We have not examined the Gly109 mutant, but restriction on a structural basis was assumed at this position (*16*). We can not comment on the Gln113 data because in HPV, position 113 is occupied by Cys. Thus, our biochemical data here almost completely confirm the results of three-dimensional structure analysis.

The interaction of the papillomavirus E2 protein and DNA was proved to be unique, involving multiple interactions among amino acids and DNA bases or backbone phosphates. Therefore, it would be very important to determine the function of each residue in the E2 protein. This is the first detailed mutagenesis study on the DNA binding domain of the E2 protein, and it revealed the residues important for DNA binding and those not important. Recently, the three-dimensional structure of the Epstein-Barr virus origin-binding protein, EBNA1, were revealed (*31*). It exhibits remarkably similarity to the structure of BPV E2. Therefore, it would be very intriguing to know if this structure is unique to E2 proteins or is conserved in some family of *trans*-acting factors.

REFERENCES

1. Ghosh, D. (1991) New developments of a transcription factors database. *TIBS* **16**, 445-447
2. Lambert, P.F., Baker, C.C., and Howley, P.M. (1988) The genetics of bovine papillomavirus type 1. *Annu. Rev. Genet.* **22**, 235-258
3. McBride, A.A., Byrne, J.C., and Howley, P.M. (1989) E2 polypeptides encoded by bovine papillomavirus type1 form dimers through the common carboxy-terminal domain: Transactivation is mediated by the conserved amino-terminal domain. *Proc. Natl. Acad. Sci. USA* **86**, 510-514
4. Moskaluk, C.A. and Bastia, D. (1989) The bovine papillomavirus type1 transcriptional activator E2 protein binds to its DNA recognition sequence as a dimer. *Virology* **169**, 236-238
5. Moskaluk, C.A. and Bastia, D. (1988) Interaction of the bovine papillomavirus type1 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
6. Hirochika, H., Hirochika, R., Broker, T.R., and Chow, L.T. (1988) Functional mapping of the human papillomavirus type11 transcriptional enhancer and its interaction with the trans-acting E2 proteins. *Genes Dev.* **2**, 54-67
7. Monini, P., Grossman, S.R., Pepinsky, B., Androphy, E.J., and Laimins, L.A. (1991) Cooperative binding of the E2 protein of bovine papillomavirus to adjacent E2-responsive sequences. *J. Virol.* **65**, 2124-2130
8. Spalholtz, B.A., Byrne, J.C., and Howley, P.M. (1988) Evidence for cooperativity between E2 binding sites in E2 *trans*-regulation of bovine papillomavirus type1. *J. Virol.* **62**, 3143-3150
9. Yang, L., Li, R., Mohr, I.J., Clark, R., and Botchan, M.R. (1991) Activation of BPV-1 replication *in vitro* by the transcription factor, E2. *Nature* **353**, 628-632
10. McBride, A.A. and Winokur, P.L. (1992) Separation of the transcriptional activation and replication functions of the bovine papillomavirus-1 E2 protein. *EMBO J.* **11**, 4111-4118
11. Haugen, T.H., Turek, L.P., Mercurio, F.M., Cripe, T.P., Olson, B.J., Anderson, R.D., Seidl, D., Karin, M., and Schiller, J. (1988) Sequence-specific and general transcriptional activation by the bovine papillomavirus-1 E2 trans-activator require an N-terminal amphipathic helix-containing E2 domain. *EMBO J.* **7**, 4245-4253
12. Chin, M.T., Hirochika, R., Hirochika, H., Broker, T.R., and Chow, L.T. (1988) Regulation of the human papillomavirus type-11 enhancer and E6 promoter by activating and repressing proteins from the E2 open reading frame: functional and biochemical studies. *J. Virol.* **62**, 2994-3002
13. Giri, I. and Yaniv, M. (1988) Study of the E2 gene product of the cottontail rabbit papillomavirus reveals a common mechanism of transactivation among papillomaviruses. *J. Virol.* **62**, 1573-1581
14. Giri, I. and Yaniv, M. (1988) Structural and mutational analysis of E2 trans-acting proteins of papillomaviruses reveals three distinct functional domains. *EMBO J.* **7**, 2823-2829
15. Hermonat, P.L., Spalholtz, B.A., and Howley, P.M. (1988) The bovine papillomavirus P_{2443} promotor is E2 *trans*-responsive: Evidence for autoregulation. *EMBO J.* **7**, 2815-2822
16. Hedge, R.S., Grossman, S.R., Laimins, L.A., and Siegler, P.B. (1992) Crystal structure at 1.7 Å of the bovine papillomavirus-1 E2 DNA-binding domain bound to its target. *Nature* **359**, 505-512
17. Taylor, J.W., Ott, J., and Eckstein, F. (1985) The rapid generation of oligonucleotide-directed mutations at high frequency using phosphothioate-modified DNA. *Nucleic Acids Res.* **13**, 8765-8785
18. Garner, M. and Revzin, A. (1981) A gel electrophoresis method for quantitating the binding of proteins to specific DNA regions: Application to components of the *Escherichia coli* lactose operon regulatory system. *Nucleic Acids Res.* **9**, 3047-3060
19. McBride, A.A., Romanczuk, H., and Howley, P.M. (1991) The papillomavirus E2 regulatory proteins. *J. Biol. Chem.* **266**, 18411-18414
20. Ham, J., Dostatni, N., Gauthier, J.-M., and Yaniv, M. (1991) The papillomavirus E2 protein: A factor with many talents. *TIBS* **16**, 440-444
21. Chothia, C. (1975) Structural invariants in protein folding. *Nature* **254**, 304-308
22. Dartman, K., Schwarz, E., Gissman, L., and zur Hausen, H. (1986) The nucleotide sequence and genome organization of human papilloma virus type11. *Virology* **151**, 124-130
23. Prakash, S.S., Grossman, S.R., Pepinsly, R.B., Laimins, L.A., and Androphy, E.J. (1992) Amino acids necessary for DNA contact and dimerization imply novel motifs in the papillomavirus E2 trans-activator. *Genes Dev.* **6**, 105-116
24. Moskaluk, C. and Bastia, D. (1988) DNA bending is induced in an enhancer by the DNA-binding domain of the bovine papillomavirus E2 protein. *Proc. Natl. Acad. Sci. USA* **85**, 1826-1830
25. Liang, H., Petros, A.M., Meadows, R.P., Yoon, H.P., Egan, D.A., Walter, K., Hozman, T.F., Robins, T., and Fesik, S.W. (1996) Solution structure of the DNA-binding domain of a human papillomavirus E2 protein: Evidence for flexible DNA-binding regions. *Biochemistry* **35**, 2095-2103
26. Hirochika, H., Broker, T.R., and Chow, L.T. (1987) Enhancers and trans-acting E2 transcriptional factors of papillomavirus. *J. Virol.* **61**, 2599-2606
27. Brokaw, J.L., Blanco, M., and McBride, A.A. (1996) Amino acids critical for the function of the bovine papillomavirus type1 E2 transactivator. *J. Virol.* **70**, 23-29
28. Landschultz, W.H., Johnson, P.F., and McKnight, S.L. (1988) The leucine zipper: A hypothetical structure common to a new class of DNA binding proteins. *Science* **240**, 1759-1764
29. Landschultz, W.H., Johnson, P.F., and McKnight, S.L. (1989) The DNA binding domain of the rat liver nuclear protein, C/EBP, is bipartite. *Science* **243**, 1681-1688
30. Gentz, R., Rauscher, III, F.J., Abete, C., and Curran, T. (1989) Parallel association of Fos and Jun leucine zippers juxtaposes DNA binding domains. *Science* **243**, 1695-1699
31. Bochkarev, A., Barwell, J.A., Pfuentzner, R.A., Furrey, W., Jr., Edwards, A.M., and Frappier, L. (1995) Crystal structure of the DNA-binding domain of the Epstein-Barr virus origin-binding protein EBNA1. *Cell* **83**, 39-46