Structural and mutational analysis of E2 *trans*-activating proteins of papillomaviruses reveals three distinct functional domains

Isabelle Giri and Moshe Yaniv

Institut Pasteur, Unité des Virus Oncogènes, UA CNRS 041149, Département de Biologie Moléculaire, 25 Rue du Dr Roux, 75724 Paris, Cedex 15, France

Communicated by M.Yaniv

The E2 proteins of papillomaviruses are able to trans-activate the viral enhancers by interacting with the sequence $ACCGN_4CGGT$ found in all papillomavirus long control regions. Analysis of the alignment of the amino acid sequences of 10 E2 proteins reveals three distinct regions: two partially conserved domains at the N and C termini of the proteins and a region variable in size and sequence in the middle. A computer prediction of the secondary structure of the 10 sequences outlines interesting conserved features, including two long amphiphilic α helices at the N terminus. To analyse the respective roles of the different segments of these proteins, we constructed a set of in-frame deletion and insertion mutations in the E2 coding sequences of the bovine papillomavirus type 1 (BPV1) and cottontail rabbit papillomavirus (CRPV). The test of their capacity to *trans*-activate or repress different viral constructs shows that the C-terminal domain of the E2 proteins is involved exclusively in DNA binding whereas the N-terminal domain is probably required for interaction with other components of the transcriptional machinery. The inner variable domain may confer flexibility to the protein such that it will facilitate contacts of the two others with their respective targets.
Key words: E2 proteins/papillomavirus/sequence analysis/transactivation

Introduction

Papillomaviruses are the causative agents of benign proliferative lesions (papillomas or warts) strictly localized on the epithelia of higher vertebrates. Some of these viruses, like the cottontail rabbit papillomavirus (CRPV), the bovine papillomavirus type 4 (BPV4) and the human papillomaviruses (HPV) types 16, 18 or 33 are strongly implicated in the occurrence of malignant lesions (Zur Hausen and Schneider, 1987). The genome of the papillomaviruses is composed of a circular DNA molecule of ~8000 bp. The overall organization of all the genomes already sequenced is quite similar (Giri and Danos, 1986), allowing the definition of a long control region (LCR) and of eight putative open reading frames (ORFs) identically located. The product of the E2 ORF has been characterized, first with the BPV1 model: this protein is able to *trans*-activate papillomavirus enhancers and promoters (Spalholz *et al.*, 1985), even in an heterologous manner (Phelps and Howley, 1987; Giri and Yaniv, 1988). The E2 gene product is a

DNA-binding protein (Androphy *et al.*, 1987; Moskaluk and Bastia, 1987), recognizing the sequence $ACCGN_4CGGT$ found in all the known papillomaviruses LCRs (Dartmann *et al.*, 1986). E2 proteins of several papillomaviruses were shown to *trans*-activate viral transcription (Hirochika *et al.*, 1987). They certainly play a critical role in the viral cycle as the activity level of most papillomavirus promoters is rather low in the absence of the E2 protein (Giri *et al.*, 1985b; Thierry *et al.*, 1987). The E2 binding target usually occurs several times along the LCRs but their exact location is different from one virus type to another. In the control region of BPV1 and CRPV, this motif is situated far upstream of the E6 promoter. In the case of human genital viruses like HPV16, HPV18 and HPV33, this palindrome is tandemly repeated very close to the initiator AUG of the E6 ORF between the putative CAAT and TATA boxes of the E6 proximal promoter. In contrast to the observations made with BPV1 or CRPV, transcription of HPV18 promoter is strongly inhibited by the E2 protein of BPV1 (Thierry *et al.*, 1987).

A negatively acting transcriptional regulatory factor is also encoded by the BPV1 E2 ORF. It corresponds to the 3' domain of the E2 protein (Lambert *et al.*, 1987). It has already been demonstrated that the C-terminal part of the E2 proteins from BPV1 (McBride *et al.*, 1988) and CRPV (Giri and Yaniv, 1988) is necessary for the binding of the protein to its target sequence, thus suggesting that the C terminus of the protein present in its repressor version is responsible for the DNA-binding function. We were interested in a more precise characterization of the functional domains of the E2 proteins. First, we have carefully analysed the sequences of 10 E2 proteins and defined three structural domains. We have then constructed a set of in-frame insertion and deletion mutations of the E2 proteins of CRPV, BPV1 and HPV18 (respectively named E2C, E2B and E2H18) and tested their *trans*-activation and *trans*-repression potential on papillomavirus promoters and enhancers with the help of CAT assays. Our results show that E2 proteins include three functional domains that overlap the structural domains: DNA binding in the C terminus, flexible region in the middle and transcription activation in the N terminus.

Results

Comparison of the amino acid sequences of 10 E2 proteins reveals three distinct regions
At first glance, the genomic organization of all the papillomaviruses sequenced is so similar that it is tempting to deduce a common function for each viral ORF (Giri and Danos, 1986). But at the amino acid level, the homology between the putative proteins of papillomaviruses is too low to allow such a conclusion: ~30% within the E6 ORFs and ~55% for the most conserved ORF, E1. The homology between the E2 ORFs of unrelated papillomaviruses such as HPV11

Fig. 1. Alignment of the amino acid sequences of 10 E2 proteins. Identical amino acids in the 10 sequences are boxed in grey. Conservation of eight residues among the 10 is outlined by open boxes. Amino acids identical in the five genital virus proteins are indicated by an open triangle below the sequence. A duplication in the HPV8 sequence is shown with dashed lines. The values to the right indicate the amino acid numbers of BPV1 E2.

and BPV1 is only ~35%. Even between viruses as closely related as the human genital viruses HPV16 and HPV18, the homology level is only 45%. It was thus important to study more precisely the amino acid sequences of the E2 proteins to examine if the conserved DNA binding and *trans*-activation functions are reflected by some conserved domains of the sequences, such domains probably being functionally important.

Our amino acid sequence analysis of the E2 proteins is based on the first 10 papillomavirus genomes sequences determined: two fibropapillomaviruses, BPV1 (Chen *et al.*, 1982) and the deer papillomavirus DPV (Groff *et al.*, 1985) which are closely related, three cutaneous viruses, CRPV (Giri *et al.*, 1985a), HPV1, the agent of deep plantar warts (Danos *et al.*, 1982) and HPV8, a virus associated with a rare cutaneous disease (Epidermodysplasia verruciformis) (Fuchs *et al.*, 1986), and five human genital viruses, three

of them (HPV16, HPV18 and HPV33) being strongly implicated in the aetiology of genital cancers (Seedorf *et al.*, 1985; Cole and Streeck, 1986; Cole and Danos, 1987); whereas HPV11 and HPV6 are rarely found in invasive tumours (Dartmann *et al.*, 1986; Schwarz *et al.*, 1983). As shown in Figure 1, only 9% of the amino acids are conserved within the 10 proteins. Analysis of the amino acids conserved in at least eight out of the 10 sequences studied, representing ~25% of the residues, reveals conserved motifs and allows the division of the amino acid sequence into three regions: the N-terminal part of the protein (domain A, about half of the length of the polypeptide) is much better conserved than the rest of the sequence. The C-terminus domain (domain C, ~20% of the sequence length) contains only three short conserved blocks and is more basic than the rest of the protein. It has been shown to be necessary for the DNA-binding function of the protein (Giri and Yaniv, 1988;
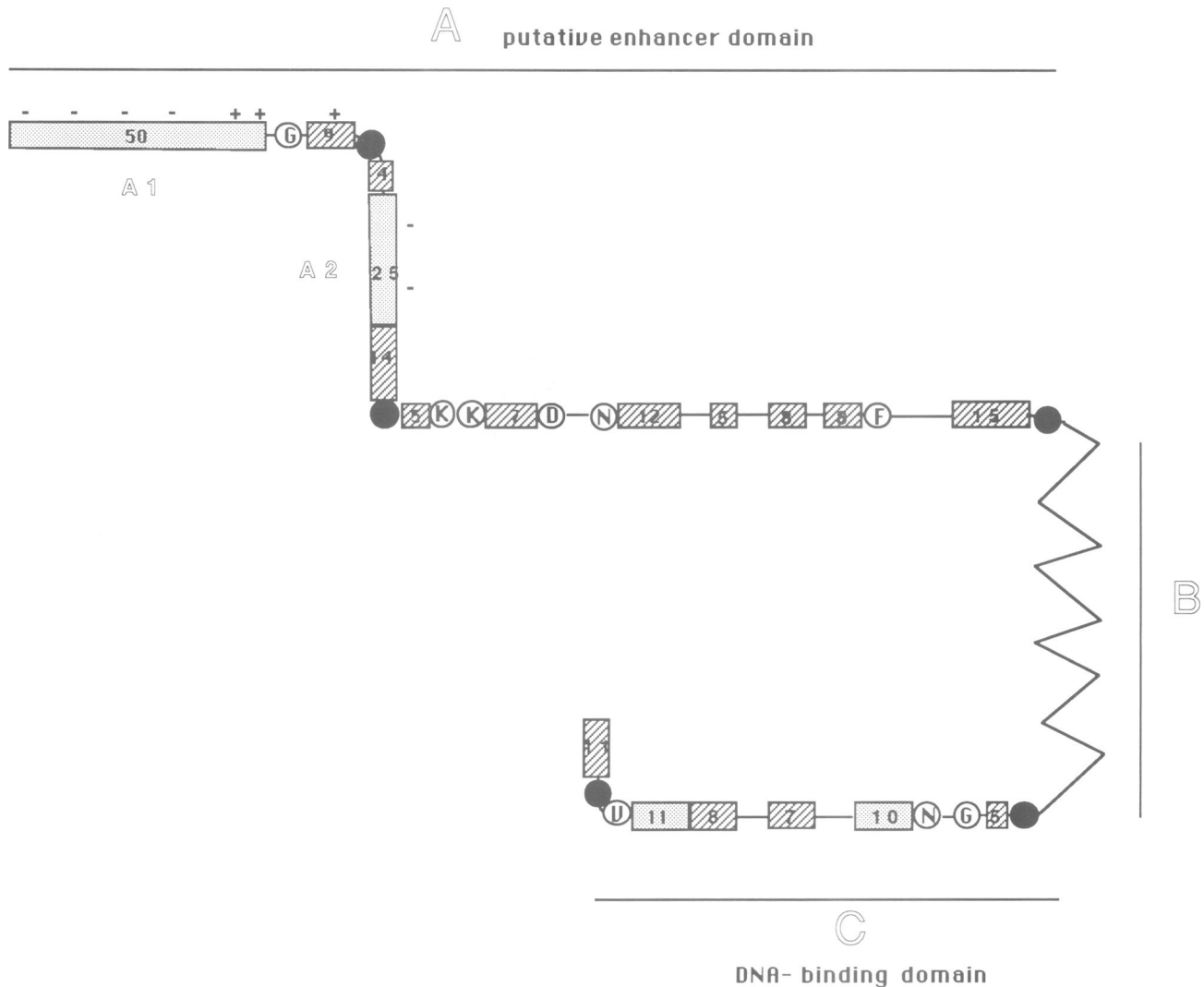
Fig. 2. Diagram of secondary structure computer predictions of E2 proteins. Conserved residues at the border of conserved structures are indicated by white circles and prolines by black ones. α helices are schematized by grey rectangles, β sheets with dashed ones, the length of which is proportional to the amino acid numbers indicated inside. Sequences of undetermined structure are symbolized by black lines. Broken lines indicate the absence of reliable structures for B domain. Overall charges of A1 helix are shown.

McBride *et al.*, 1988). The internal part of the protein (domain B, ~30% of the sequence length) is quite variable in size: from 70 amino acids in genital virus sequences to 160 amino acids for HPV8 E2 protein which harbours a duplication shown in Figure 1. No conserved amino acids could be found in this domain, but some features common to the 10 sequences can be outlined: a very high number of prolines (15% P instead of 3% in the rest of the E2C protein for example) and arginine (14% R in domain B versus 5% in domain A and 7% in domain C), a particularly basic motif $(R-X-R)_n$ found interspersed in the domain B of all the E2 proteins but only once in E2B and E2D sequences. Domain B of CRPV, HPV1 and HPV8 contains more basic amino acids than acidic ones and is thus positively charged (+7 to +24). In contrast, this domain of E2B is acidic with a resulting net charge of −6, due in particular to an acidic motif around amino acid 270 (EEEE---D--EEE). Such an acidic stretch of amino acids is not present in the other viral sequences. Domain B of the six other sequences is neutral.

The E2 proteins of the five genital viruses analysed here are more conserved than the others, particularly in the DNA-binding domain (see Figure 1). This could be linked to a

particular mechanism of action of the genital E2 proteins or to a more recent divergence of these viruses.

Secondary structure predictions for the E2 proteins define three very distinct structural domains

Computer predictions of protein secondary structure can reveal interesting potential data and thus outline functionally important domains of proteins. We have used the algorithm of Garnier *et al.* (1978), based on the chemical properties of amino acids, as well as the method of Levin *et al.* (1986) which utilizes a value matrix obtained by compilation of the sequences of proteins of known secondary structures. Such a prediction was performed for the 10 E2 proteins studied and we have retained only the secondary structures obtained with these two computer programs on the 10 sequences in order to consider reliable data. The results are schematized in Figure 2. We observed a very clear division of the sequence into three regions which are remarkably overlapping with the domains defined from the homology analysis despite the low number of conserved amino acids. The structure of the domain A is extremely well defined. It begins with a very long α helix of 50 amino acids, this structure

Table I. Sequences of E2 protein mutations

Mutant name	Site	Wild-type sequence	Mutated sequences[a]
E2C A3	NcoI	E P WT L	E P **C R S A** WT L
E2C A4	NcoI	E P WT L	E P **WP R P** WT L
E2C B1	BstEII	S R S P G	S R S **Q R S S** S P G
E2C B4	SmaI	P P G R N	P P **S P R G** G R N
E2C B7	MluI	T T R T L	T T **R P R P** R T L
E2C B8	NruI	L R E L	L R **P R G P** E L
E2C C3	AflII	Q L K C	Q L K **P R L** K C
E2C C2	AflII	C L R Y	C L **S R G L** R Y
E2C C1	SphI	S G R M L	S G R **K I F** M L
E2B A10	SphI	M E T A C E	M E T **G K I D L P** A
E2B A4	NcoI	E P WS L	E P **C R S A** WS L
E2B A8	NcoI	C T M A G	C T M **Q I C M** A G
E2B A9	NcoI	C T M A G	C T M **A A A M** A G
E2B B1	StuI	A G L G	A G **E D L P** L G
E2B B2	KpnI	G T V P V	G T **S P G T** V
E2B B3	KpnI	G T V P V	G T **E D L P** V
E2B C2	BclI	I L I T F	I L I **A A A I** T F
E2H18 B1	HindII	P V N P L	P V **S P R G** N P L
E2H18 B2	HindII	P V N P L	P V **G R S S** N P L

[a]Inserted amino-acids are indicated in bold type.

being predicted in the 10 sequences even within unconserved blocks of up to 10 amino acids. This helix (A1 in Figure 2) ends with a conserved glycine and is separated from a second one of 25 residues long (A2) by a conserved proline probably surrounded by two small β sheets. The rest of domain A is mainly composed of small β sheets, often interrupted by conserved charged residues. It was named A_β region to clarify the presentation of the rest of the results. A more precise analysis of helix A1 with a computer program developed by J.M.Claverie (personal communication) gives a high amphiphilic value, meaning that charged residues are present on one side and hydrophobic residues on the other side of the helix. The major part of the helix A1 (first 40 amino acids) is negatively charged, with a net charge varying from -3 to -5 depending on the sequences, whereas two to three of the last five residues are basic, forming a small positively charged motif. The succeeding β sheet has a low positive charge whereas the A2 helix is rather acidic with a charge of -2 to -3 (see Figure 2). The overall domain A is negatively charged.

Domain B of all the E2 proteins is punctuated by numerous proline residues which have a low probability of being integrated in an organized structure. Only interspersed small β sheets were predicted in these regions. Their localization is not conserved from one sequence to another but interest-
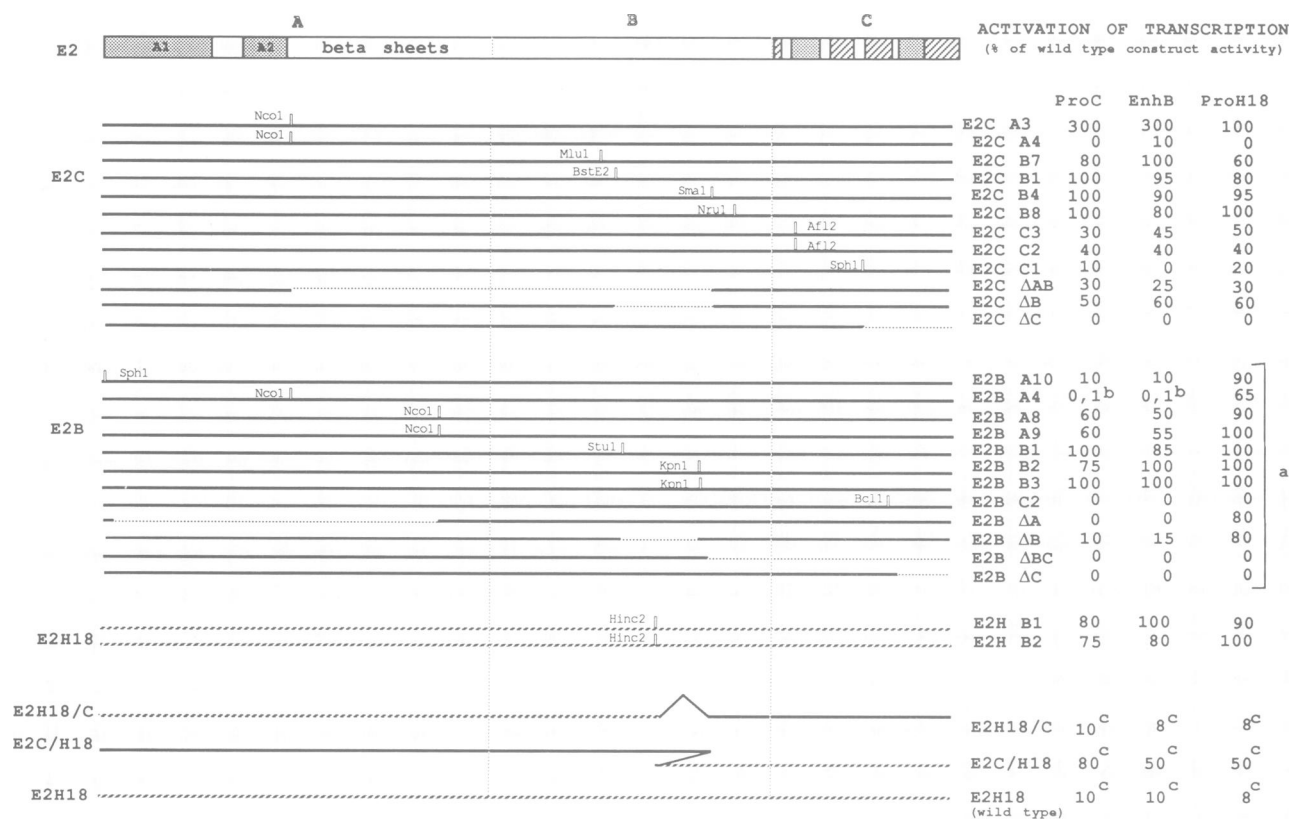

Fig. 3. Structure and activity of E2 mutants. The three structural domains are drawn at the top. Main α helix and β sheets are indicated as in Figure 2. The thin vertical lines represent the boundaries of the three domains. The position of each in-frame insertion is indicated by a white bar with the restriction site used to construct it. Deleted regions are schematized with dotted lines. The table at the right depicts the capacity of each mutant to *trans*-activate the CRPV promoter (ProC), the BPV1 enhancer (EnhB) and to *trans*-activate and/or repress (A) the HPV18 promoter (ProH18). Activities, obtained from at least two independent experiments, are expressed as percentage of the wild-type protein activity. (B) An inhibitory effect was repeatedly observed leading to a 10-fold lower activity. (C) These activities are expressed as percentage of E2C wild-type activity. HPV18 sequences are indicated with striated lines. The *trans*-activation factors obtained with wild-type E2B were between 10 and 20 in different experiments, with E2C from 5 to 12 and with E2H18 ~3.

ingly they often cover the R-X-R motifs suggesting that the arginine residues are found on the same side of the amino acid chains.

Domain C is also well structured with alternating α helices and β sheets. As in domain A, secondary structures are much more conserved than the primary amino acid sequences suggesting a functional conservation of E2 proteins during evolution of papillomaviruses. No features reminiscent of DNA-binding properties, such as cysteine−histidine fingers or helix-turn-helix, were observed.

Domain C is essential for efficient activation or repression of transcription

The high degree of conservation of structural features suggests that the conserved domains perform important functions that are common to all of the E2 proteins. To assess these functions, we constructed a set of in-frame insertion mutations of E2B, E2C and E2H18 with the use of linkers (see Materials and methods and Table I). Deleted in-frame mutants were obtained by recombination of two appropriate insertion mutants with the help of restriction sites encoded by the linkers. The mutated sequences were inserted in a eukaryotic expression vector and co-transfected with CAT test plasmids harbouring the CRPV or the HPV18 promoters or the BPV1 enhancer upstream of the SV40 promoter. Experiments with the CRPV promoter were performed in rabbit VX2 cells (Georges et al., 1985), the others in human SW13 cell line. Data obtained with the three different constructs were nearly equivalent. The results are summarized in Figure 3. All E2B and E2C proteins deleted of the C domain were no longer capable of trans-activation (Figure 3), not even the E2B ΔC construct in which only the 33 terminal residues are deleted. An attempt to restore trans-activation by increasing amounts of transfected E2 expression vector was unsuccessful (data not shown). The E2B C2 and E2C C1 mutants did not trans-activate any CAT test plasmid, whereas the E2C C2 and C3 mutants retained half the wild-type activity, suggesting that only the very end of this domain is crucial for the activity. The E2B ΔC and C2 mutants were no longer able to repress the HPV18 promoter.

A domain is essential for efficient trans-activation but not for repression

The E2B A10 mutant, in which six amino acids including a proline were inserted after the fifth residue of the protein (Table I), was unable to trans-activate papillomavirus enhancers and promoters (Figure 3). Thus the integrity of the A1 helix is strictly necessary for optimal activation of transcription. The E2B A4 mutant localized at the end of the A2 helix was unable to trans-activate and moreover strongly inhibited transcription from papillomavirus promoter as well as enhancer constructs, indicating that a stable protein probably capable of interacting with the DNA target was synthesized. Surprisingly, the E2C A3 protein which harbours exactly the same insertion at the same site was still perfectly able to trans-activate and even more efficiently than the wild-type E2C, except with HPV18 promoter. A third mutation of this conserved motif, E2C A4, which harbours two additional prolines, did not activate the reporter gene. An important structure is thus disrupted by the additional prolines. The A_β region seems to be less important for the

trans-activation function: the E2B A8 mutant located at the end of the A domain and the E2B A9, another insertion at the same point, retained about half the wild-type activity. Thus, the integrity of the N-terminal part of E2 including the two α helices appears to be strictly necessary for the trans-activation function of E2 proteins whereas mutations in the A_β region only diminished its activity. Consistent with these results, the E2C ΔAB mutant, deleted of the β part of the domain A, is still partially active, whereas the E2B ΔA construct which lacks the two α helices retains no activity.

Possible explanations as to why little or no stimulation was observed when using these mutants with deletions or insertions within the A domain are that they are unstable or unable to enter the cell nucleus, and therefore unable to interact with the DNA target sequence. However, all the E2B proteins mutated in this A domain still repressed the HPV18 promoter in SW13 cells with at least 65% of the activity of the intact E2B, indicating that functional DNA-binding proteins were synthesized from these constructs. This repression observed with the C-terminal domain recalls that occurring with the naturally truncated version of E2B (E2TR) which lacks domain A (Lambert et al., 1987). E2TR is very similar to the E2B ΔA mutant which contains only seven additional amino acids remaining from the N terminus of the protein. It is thus highly probable that domain A does not play any role in DNA binding.

The B domain is not essential for trans-activation or repression

All the insertion mutants in domain B were still capable of trans-activation, although sometimes with a lower efficiency than the wild-type proteins (from 65 to 100%, see Table I and Figure 3). The B mutants of the E2 protein of BPV1 were capable of repressing the activity of the HPV18 promoter with the same efficiency as the wild-type E2B. Addition of two negative charges in E2B B1 and B3 does not inhibit its activity. Two mutants of the HPV18 E2 protein in domain B retained the low activity level of the wild-type protein. With a large deletion (47 amino acids) in domain B of the E2C protein, the E2C ΔB mutant was still active with about half the wild-type efficiency, but a similar deletion of 36 residues in E2B strongly decreased its trans-activation function but not the repressor activity.

Swapping A domains of CRPV and HPV18

The 'domain-swap' experiments described in Figure 3 test the hypothesis that the A domain is responsible for the trans-activation properties of E2 proteins and thus for their relative activity. The idea was to substitute the domain A of E2C in the E2H18 construct and then to determine if the activity level of the E2C protein was transferred with this domain. We have used the restriction sites created in the middle of domain B of our mutants to construct these hybrids. Whereas the E2H18 protein was always poorly active, ~ 10% of the activity level of E2C, the hybrid construct E2C−H18 has half the activity level of the wild-type E2C. Thus part of the trans-activation potential of E2C has been transferred. Since the exchange was performed at non-homologous sites, the A domain of the hybrid protein could be in a conformation different from the wild type and thus slightly less active.

The reciprocal construct E2H18−C retained the activity level of E2H18 indicating that the DNA binding domain of E2H18 is not responsible for its low activity.

Discussion

The analysis of the effect of deletion and insertion mutations on the capacity of E2C and E2B proteins to *trans*-activate and *trans*-repress papillomavirus promoters and enhancers shows that E2 proteins can be divided into three functional domains overlapping the three structural domains defined by sequence comparisons. We have previously demonstrated by *in vivo* competition experiments (Giri and Yaniv, 1988) that direct interaction of the E2 proteins with papillomaviruses LCRs was necessary for the E2-mediated *trans*-activation and that the C-terminal part of the E2C protein was required for *in vitro* DNA binding. McBride *et al.* have recently established that the DNA-binding domain of the E2B protein is included in the 101 C-terminal amino acids (McBride *et al.*, 1988). The possible interdependence of the DNA-binding and enhancer functions has not been determined. We suggest here that these two functions are localized in two distinct domains of E2 proteins, the role of the third domain being still unclear. It could be simply a flexible link between the N and C termini. We propose that domain C encodes exclusively the DNA-binding part of the protein and that domain A corresponds to the activating one, the two long α helices being particularly important for this function.

Since the repressor effect of E2B on the transcriptional activity of the HPV18 promoter requires a direct interaction of the protein with the LCR (Giri and Yaniv, 1988), it is highly probable that such an inhibitory effect is due to the binding of E2B to the target sequences in HPV18 sequences and not to sequestration of cellular transcriptional factors. The fact that the E2B proteins mutated in A or B domains are still able to repress HPV18 transcription indicates that they retain a functional DNA-binding domain. Analysis of transcription activation by E2C mutants more precisely defines the border of the DNA-binding region: E2C B8 was still active whereas E2C C3 located only 21 amino acids further showed a reduced *trans*-activation effect, probably because of a decrease of DNA-binding activity. Thus, the DNA-binding domain of E2 proteins probably specifically involves the structural C domain. However, this should be directly demonstrated by the study of *in vitro* binding of E2 mutants on their target the ACCGN₄CGGT motif. Interestingly, the structural motif mutated in E2C C2 and E2C C3 corresponds to the region of maximal homology with the *mos* protooncogene (Danos and Yaniv, 1984). This sequence may play an auxiliary function in DNA binding by the protein. The DNA-binding domain of E2 proteins is rather small (∼100 amino acids) and exhibits no characteristics of known DNA-binding structures. Whereas all the E2 proteins certainly recognized the same target sequence, this region is surprisingly poorly conserved at the amino acid level (only three short blocks), suggesting that either DNA sequence specificity is conferred by only a few residues or that variable amino acid sequences can recognize the same DNA target.

At the other end of the protein, the integrity of the α part of domain A was strictly necessary for the *trans*-activation function but not for E2B repression of HPV18 promoter, suggesting that this domain is not involved in DNA binding

but is directly responsible for the activation function. The swapping experiment indicates that at least part of the activation potential of the E2C protein is encoded by this domain. Its predicted secondary structure with two long α helices separated by a proline residue is quite interesting. Acidic regions like A1 helix have been implicated in transcriptional activation by GCN4 and GAL4 proteins in yeast (Hope and Struhl, 1986; Ma and Ptashne, 1987a) and were observed in yeast *trans*-activators encoded by *Escherichia coli* genomic DNA fragments (Ma and Ptashne, 1987b). Moreover, in addition to A1, most of these charged short sequences could form amphiphilic helices, suggesting that they are localized in an accessible region of the proteins. The yeast GAL4 protein is also capable of gene activation in mammalian cells (Kakidani and Ptashne, 1988; Webster *et al.*, 1988). Thus, amphiphilic helices could also be involved in *trans*-activation functions of higher eukaryotes. It would be interesting to test the biological activity of mutations modifying the α helix charges in its acidic as well as in the small basic motif. In the E2B A10 mutant, computer analysis suggests that the N-terminal 10 amino acids are no longer in a helical conformation. In contrast, the probability of amphiphilic structure of the A2 helix of the E2C A3 mutant which is more active than the wild type is higher as predicted by the computer studies, whereas the E2C A4 construct, mutated at the same site, has no amphiphilic properties and no longer retains activity. The E2B A4 mutant, which showed inhibitory properties, has an interesting structure: the mutated end of the A2 helix presents a high amphiphilic value but is separated from the beginning by a β turn. The mutated protein probably possesses a structure able to interact with transcriptional factors but in such a way as to inhibit transcription. It is interesting to note that there are very few differences between activator and inhibitor forms of this protein. This suggests the hypothesis that a conformational change induced for example by various DNA targets or by the formation of dimers or tetramers could lead to an inhibitor form of E2 proteins. We propose that the two A1 and A2 helices are directly responsible for *trans*-activation and that this is related to their amphiphilic characteristics. The importance of amphiphilicity rather than consensus sequence has been demonstrated in other systems, such as mitochondrial presequence function (Roise *et al.*, 1988). The β part of domain A has only a secondary role in E2 *trans*-activation as the E2B A8 and A9 mutants were partially active. However, more mutations in this region would be necessary to determine if some structural features of this region play an important role in E2-mediated *trans*-activation.

The B domain, which is conserved neither in sequence nor in length among the different E2 proteins, may act as a hinge between the DNA binding and the activation domains. Its high content of prolines could confer to the protein enough flexibility to permit domain A to adjust to unknown components of the transcriptional machinery. In the E2C−H18 hybrid, domain B is 25 amino acids longer but this did not abolish the activity. Deletion of up to 47 residues in E2C ΔB did not totally inactivate the protein indicating that the length of this domain may be altered considerably without loss of function. However, deletion of 36 residues in domain B of E2B protein almost totally abolished its activity. This may be related to the particular sequence of this part of E2B which is the only acidic one among the

sequences studied. The deletion suppresses three basic residues leading to an even more acidic domain (−9 charges) which could for example inhibit the binding of the protein to its DNA target.

The structure analysed here for the papillomavirus E2 *trans*-acting proteins may be a prototype for many enhancer or promoter-binding factors: they would have a small and compact DNA-binding domain, linked by flexible sequences to a *trans*-activating domain probably relatively globular in nature. The loose connection between both functional domains may be related to one of the typical features of eukaryotic control sequences, the variability in the position of active *cis*-control elements in enhancers and promoters with relation to the transcription start site.

Materials and methods

Sequence analysis
The amino acid sequences of BPV1, HPV1 and CRPV were aligned using a computer program described by Needleman and Wunsch (1970) and the others were manually added using the conserved residues outlined by the first step. Secondary structure probabilities were determined as described by Garnier et al. (1978) and Levin et al. (1986).

Constructions of E2 mutants
The wild-type E2 expression vectors have already been published (Thierry and Yaniv, 1987; Giri and Yaniv, 1988). The E2B construct, originally named C59, was a gift of P.Howley (Spalholz et al., 1985). Insertion mutants were obtained by ligating linkers (Biolabs) harbouring a *Bgl*II site or a *Sac*II site (proline linker) of the appropriate length. The plasmid restriction sites were, if necessary, blunt-ended with T4 polymerase or Klenow polymerase according to the manufacturer's procedures (Biolabs) in such a way as to restore the frame. Deleted in-frame mutants were constructed by recombination of two appropriate insertion mutants with the use of the restriction site introduced by the linker in such a way as to maintain the frame. Plasmids from two independent clones of each mutant were prepared by two caesium chloride centrifugations and tested for their biological activity.

E2B ΔBC mutant was obtained by inserting at the *Kpn*I site a linker harbouring a stop codon in the three possible ORFs (a gift of P.Howley). E2B ΔC and E2C ΔC were constructed by blunt-ending and self-ligating the expression vectors at appropriate restriction sites (see Figure 3).

CAT assays
The plasmids harbouring the CRPV and HPV18 promoters just upstream of the *cat* gene have already been described (Thierry et al., 1987; Giri and Yaniv, 1988). The 407.1 plasmids (renamed enhB to clarify) in which the BPV1 transcriptional control region is in an enhancer configuration upstream of the SV40 promoter was provided by P.Howley (Spalholz et al., 1985). CAT enzyme assays were performed as described by Gormann et al. (1982). Briefly, subconfluent SW13 (derived from a human adenocarcinoma) or rabbit VX2 cells (Georges et al., 1985) were cotransfected by calcium phosphate precipitation with 5 μg of CAT plasmids and 5 μg of the wild-type or mutated E2B expression vectors or 15 μg of the other E2 constructs. The total amount of transfected DNA was kept constant to 20 μg with pBR322 DNA. The monolayer was washed after 24 h and extracts were prepared after 48 h. Cells were disrupted by sonication. CAT assays were performed with extracts corresponding to half a 60-mm dish and 0.1 μCi of [^{14}C]chloramphenicol (Amersham) for 3 h with VX2 cells and 30 mn with SW13 cells. Each transfection was repeated at least twice.

References

Androphy,E.J., Lowy,D.R. and Schiller,J.T. (1987) *Nature*, **325**, 70−73.
Chen,E.Y., Howley,P.M., Levinson,A.D. and Seeburg,P.H. (1982) *Nature*, **299**, 529−534.
Cole,S.T. and Streeck,R.E. (1986) *J. Virol.*, **58**, 991−995.
Cole,S.T. and Danos,O. (1987) *J. Mol. Biol.*, **193**, 599−608.
Danos,O. and Yaniv,M. (1984) *Cancer Cells*, **2**, 291−294.
Danos,O., Katinka,M. and Yaniv,M. (1982) *EMBO J.*, **1**, 231−236.
Dartmann,K., Schwartz,E., Gissmann,L. and zur Hausen,H. (1986) *Virology*, **151**, 124−130.
Fuchs,P.G., Iftner,T., Weninger,J. and Pfister,H. (1986) *J. Virol.*, **58**, 626−634.
Garnier,J., Osguthorpe,K. and Robson,B. (1978) *J. Mol. Biol.*, **120**, 97−120.
Georges,E., Breitburd,F., Jibard,N. and Orth,G. (1985) *J. Virol.*, **55**, 246−250.
Giri,I. and Danos,O. (1986) *Trends Genet.*, **2**, 227−232.
Giri,I. and Yaniv,M. (1988) *J. Virol.*, **62**, 1573−1581.
Giri,I., Danos,O. and Yaniv,M. (1985a) *Proc. Natl. Acad. Sci. USA*, **82**, 1580−1584.
Giri,I., Danos,O., Thierry,F., George,E., Orth,G. and Yaniv,M. (1985b) *UCLA Symp. Mol. Cell. Biol.*, **32**, 379−390.
Gormann,C.M., Loffat,L.F. and Howard,B.H. (1982) *Mol. Cell. Biol.*, **2**, 1044−1051.
Groff,D.E. and Lancaster,W.D. (1985) *J. Virol.*, **56**, 85−91.
Hirochika,H., Broker,T.R. and Chow,L.T. (1987) *J. Virol.*, **61**, 2599−2606.
Hope,I.A. and Struhl,K. (1986) *Cell*, **46**, 885−894.
Kakidani,H. and Ptashne,M. (1988) *Cell*, **52**, 161−167.
Lambert,P.F., Spalholz,B.A. and Howley,P.M. (1987) *Cell*, **50**, 69−78.
Levin,J.M., Robson,B. and Garnier,J. (1986) *FEBS Lett.*, **205**, 303−309.
Ma,J. and Ptashne,M. (1987a) *Cell*, **48**, 847−853.
Ma,J. and Ptashne,M. (1987b) *Cell*, **51**, 113−119.
McBride,A.A., Schlegel,R. and Howley,P.M. (1988) *EMBO J.*, **7**, 533−539.
Moskaluk,C. and Bastia,D. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 1215−1218.
Needleman,S.B. and Wunsch,C.D. (1970) *J. Mol. Biol.*, **48**, 443−453.
Phelps,W.C. and Howley,P.M. (1987) *J. Virol.*, **61**, 1630−1638.
Roise,D., Theiler,F., Horvath,S.J., Tomich,J.M., Richards,J.H., Allison,D.S. and Schatz,G. (1988) *EMBO J.*, **7**, 649−653.
Schwarz,E., Durst,M., Demankowski,C., Lattermann,O., Zech,R., Wolfsperger,E., Suhai,S. and Zur Hausen,H. (1983) *EMBO J.*, **2**, 2341−2348.
Seedorf,K., Krammer,G., Dürst,M., Suhai,S. and Röwekamp,W.G. (1985) *Virology*, **145**, 181−185.
Spalholz,B.A., Yang,Y.C. and Howley,P.M. (1985) *Cell*, **42**, 183−191.
Thierry,F. and Yaniv,M. (1987) *EMBO J.*, **6**, 3391−3397.
Thierry,F., Heard,J.M., Dartmann,K. and Yaniv,M. (1987) *J. Virol.*, **61**, 134−142.
Webster,N., Jin,J.R., Green,S., Mollis,M. and Chambon,P. (1988) *Cell*, **52**, 169−178.
Zur Hausen,H. and Schneider,A. (1987) In Salzman,N.P. and Howley,P.M. (eds), *The Papovaviridae*. Plenum Press, New York, Vol. 2, pp. 245−263.

Received on April 28, 1988