Nucleotide Sequence of Bovine Papillomavirus Type 2 Late Region

By HAROLD L. POTTER, JR AND WILLIAM J. MEINKE*

Department of Microbiology and Immunology, College of Medicine, Arizona Health Sciences Center, Tucson, Arizona 85724, U.S.A.

(Accepted 25 September 1984)

SUMMARY

The late region of bovine papillomavirus type 2 (BPV-2) DNA has been identified. The complete nucleotide sequence of the region was determined and revealed two large open reading frames. The DNA sequence results and the predicted amino acid sequence of putative polypeptides encoded by this region are presented. Comparative analysis of the BPV-2 late region and the corresponding area of BPV-1 was performed. This study demonstrates that identical genetic organization and considerable nucleotide sequence conservation exists between these two serotypes.

The bovine papillomaviruses (BPV) have been shown to be oncogenic. They are associated with alimentary cancer in cattle (Campo et al., 1980) and can induce tumour formation when injected into hamsters (Morgan & Meinke, 1980). Bovine papillomavirus type 1 (BPV-1) and type 2 (BPV-2) can transform mouse and bovine cells in vitro (Meischke, 1979; Lowy et al., 1980; Law et al., 1981; Moar et al., 1981; Lancaster & Olson, 1982). The BPV genome in these transformed cells is found in the episomal state (Law et al., 1981; Moar et al., 1981). Interestingly, no tissue culture system has been found to support the growth in vitro of any of the papillomaviruses. Thus, present studies predominantly utilize cloned viral DNA and focus on genomic organization, DNA sequence analysis, and potential products of gene transcription and translation.

Of the bovine papillomaviruses, BPV-1 has been the most extensively studied, with the entire DNA sequence reported as well as mapping of several RNA transcripts and assignment of genomic areas responsible for early and late viral functions (Lowy et al., 1980; Chen et al., 1982; Engel et al., 1983). Recently, the complete nucleotide sequence of human papillomavirus (HPV) type 1a was reported (Danos et al., 1982). Comparative sequence analysis between BPV-1 and HPV-1a revealed similarities in genomic organization and demonstrated certain regions of conserved nucleotide sequences (Danos et al., 1983). Other investigators have found analogous results in studies with BPV-1 and the European elk papillomavirus (Stenlund et al., 1983). Law et al. (1979), using stringent reassociation conditions, demonstrated homologous regions existing uniformly throughout the BPV-1 and BPV-2 genomes. Homology was not confined to certain non-contiguous regions among the two viral genomes. Therefore, it seemed reasonable to believe that some restriction enzyme sites might also be conserved throughout the DNA of both viruses. Our laboratory constructed a detailed restriction enzyme map of BPV-2 DNA and reported that if the genomic maps of BPV-2 and BPV-1 were aligned at their unique HindIII site, the order and number of cleavage sites of some restriction enzymes were preserved (Potter et al., 1981). Particularly noteworthy was the relative position of the single HindIII site and one of the three BamHI sites of BPV-2. If the correct BamHI site is cleaved, it separates BPV-2 DNA which has been opened at the unique HindIII site into 69% and 31% fragments. The single BamHI and HindIII cleavage sites of BPV-1 also divides its DNA into two fragments comprising 69% and 31% of the genome. The BPV-1 69% subgenomic fragment has been shown to contain the early region of the genome and the remaining 31% to contain nearly all of the late region (Lowy et al., 1980; Chen et al., 1982; Engel et al., 1983). These data suggest that it might be possible to deduce where the BPV-2 genome is organized into early and late regions by direct comparison with the
putative proteins. Base position 1 is the A of the proposed ATG translational initiation codon of L2. Stars are centered under all stop codons which are in the same frame as L2 and L1. The proposed translational start codon of L1 is at base position 1417. The sites of a few restriction enzymes are listed above the sequence. Other areas of interest are discussed in the text.

BPV-1 map. Heteroduplex analysis performed by Campo & Coggins (1982) also indicated alignment of the BPV genomes at their HindIII site to be correct. Additionally, their results indicated the probable early and late regions of BPV-2 to be located as described above. We present here
the location and complete nucleotide sequence of the late region of BPV-2. Additionally, a comparative analysis with the corresponding area of BPV-1 is furnished.

For these studies BPV-2 was isolated from naturally occurring bovine papillomas and supercoiled BPV-2 DNA was extracted as previously reported (Potter et al., 1981). The entire BPV-2 genome was cloned into the single HindIII restriction site of pBR322. Following transformation of Escherichia coli by the procedures of Kushner (1978), D-cycloserine enrichment was performed to increase the probability of obtaining only those clones carrying recombinant plasmids (Bolivar & Backman, 1979). Screening of ampicillin-resistant, tetracycline-sensitive clones included the use of colony hybridization (Grunstein & Wallis, 1979) and restriction enzyme analysis of miniscreen-prepared plasmid DNA (Maniatis et al., 1982). Also, AvaI-BamHI and HindIII-BamHI doubly digested BPV-2 DNA was used to shotgun clone subgenomic fragments into the large DNA restriction fragment of similarly digested pBR322. These clones were identified by restriction enzyme analysis of miniscreen-prepared plasmid DNA. Recombinant plasmids were utilized for sequencing.

In order to sequence BPV-2 DNA, it was first treated with calf intestinal phosphatase, extracted with phenol and ether, and then 5' end-labelled with T4 DNA polynucleotide kinase and [γ-32P]ATP. Uniquely end-labelled DNA fragments were produced by secondary restriction enzyme cleavage and subsequently isolated by electrophoresis in 5% polyacrylamide gels. After crush-soak elution, labelled DNA was sequenced according to the method of Maxam & Gilbert (1980). Ethanol precipitation after the piperidine reaction, rather than lyophilization, was the only modification to the procedure (Smith & Calvo, 1980). Thin 20% (0.04 x 20 x 40 cm) and 8% polyacrylamide gels (0.04 x 20 x 80 cm) were used to sequence up to 250 bases away from the labelled end. More than 94% of the area studied was either doubly sequenced or the opposite strand was sequenced. Additionally, sequencing across all restriction sites was performed.

Fig. 1 presents the 3195 bases sequenced in this study. As expected, the area in question was found to include the late region of BPV-2 and, therefore, is likely to code for viral structural proteins. Analysis of the DNA sequence revealed two large open reading frames located on the same DNA strand and separated by a single stop codon. The other DNA strand was found to contain no significant coding capacity. The smaller reading frame is referred to as L2 and the larger as L1. Beginning with their first methionine codon, each gene could code for polypeptides of 467 and 497 amino acids respectively. A potential TATA box (Goldberg, 1979) is located 21 bases upstream from the first L2 Met codon. Two polyadenylation signals (Proudfoot & Brownlee, 1976) are located 19 and 55 bases downstream from the L1 termination codon. Similarly, the BPV-1 late region contains two large open reading frames located on the same DNA strand and separated by a single stop codon (Chen et al., 1982; Engel et al., 1983). Also, regulatory sequences [TATA box, poly(A) recognition sequences] in the BPV-1 late region are located in about the same position as in BPV-2 with regard to the two open reading frames (Chen et al., 1982; Engel et al., 1983). Therefore, identical genetic organization in the late regions of the viruses is found to exist. Fig. 2 presents a schematic representation of the BPV-2 late region.

Nucleotide sequence comparison between BPV-2 and BPV-1 demonstrated the existence of considerable conservation in their late regions. Starting with the first L2 Met codon and ending with the L1 termination codon, there was 82% DNA homology between the two (Fig. 3). If examined separately, the L2 and L1 genes have about 77% and 86% DNA sequence homology.
Fig. 3. DNA sequence homology between BPV-2 and BPV-1 L2 and L1 genes. The DNA sequence begins with each proposed translational initiation codon of BPV-2 and BPV-1 L2 genes. The complete sequence of BPV-2 L2 and L1 genes is shown on the upper line. The BPV-1 sequence is listed on the lower line only at positions where base alterations occur. Proposed translation initiation codons for both genes are shaded. In four areas a shift of one codon was performed to maximize homology. The shifts are represented by stars.
Fig. 4. Demonstration of amino acid sequence homology between BPV-2 and BPV-1 L2 and L1 proteins. The protein sequence begins with the first methionine residue of BPV-2 and BPV-1 L2 proteins. Homologous regions are shown as shaded areas. In four places a shift of one amino acid was performed to maximize homology. Stars represent stop codons and dashes the 12 nucleotides separating L2 and L1 genes.

respectively. Due to the degeneracy of the genetic code, the predicted amino acid sequence homologies of their putative proteins was higher, with 84% of L2 residues and 92% of L1 residues being the same. Fig. 4 illustrates this high degree of amino acid conservation.
Several of the DNA base changes involved the last base of the codon and, accordingly, did not result in an amino acid change. In positions where changes were noted, many involved conservative amino acid substitutions. For example, several threonine residues in the BPV-1 late proteins are replaced by serines in the corresponding position of BPV-2 proteins. Both, however, are hydroxy amino acids with similar physical and chemical characteristics. Substitutions such as these probably would not alter the properties of the protein. If conserved substitutions are allowed, the observed protein homology between BPV-2 and BPV-1 late proteins is slightly higher with 90% homology for L2 proteins and 95% homology for L1 proteins. As with BPV-1, the carboxy terminus of each BPV-2 protein is quite basic (60 to 80% Arg + His + Lys), which could allow them to interact with the viral DNA (Danos et al., 1983).

The polypeptide composition of purified BPV-2 was determined by SDS–PAGE (data not shown). One major and several minor polypeptides were detected. The mol. wt. of the major component was determined to be about 55000, which is similar to the reported value in earlier communications describing BPV-2 structural proteins (Lancaster & Olson, 1978; Pfister et al., 1979). This is also close to the reported value of 53500 for the mol. wt. of the BPV-1 major capsid protein (Meinke & Meinke, 1981). Starting with the first Met codon, the BPV-2 L1 gene could code for a protein of mol. wt. 55600. Previous investigators using Northern blot analysis of poly(A)-selected BPV-1 RNA, found a 1700 base, wart-specific species of RNA whose hybridization length spans L1 (Engel et al., 1983). Subsequent elution and translation showed that this RNA species directed the synthesis of the 55 kilodalton (kDa) major capsid protein of BPV-1. Thus, we assume the 55 kDa protein of BPV-2 to be the capsomere subunit and a product of the L1 gene. Similar experiments conducted in the future should identify the origin of other papillomavirus structural polypeptides.

Several minor polypeptides of BPV-2 were also detected. Of these, one had a mol. wt. of approx. 50000. Examination of the BPV-2 L2 gene shows it to be capable of coding for this protein. Beginning with the proposed translational start codon, L2 can potentially direct synthesis of a 49.5 kDa protein. Interestingly, a protein of similar size is found in BPV-1 (Meinke & Meinke, 1981). Therefore, it seems reasonable to suggest that this protein may be a product of the BPV-2 L2 gene.

We are grateful to Dr David Mount, University of Arizona, for supplying us with a copy of his DNA sequence computer program. This investigation was supported in part by PHS Grant CA09213, awarded by the National Cancer Institute.

REFERENCES


(Received 23 July 1984)