Identification of the E5 Open Reading Frame of Human Papillomavirus Type 16

CHRISTINE L. HALBERT AND DENISE A. GALLOWAY*

Fred Hutchinson Cancer Research Center, 1124 Columbia Street, Seattle, Washington 98104, and Department of Pathology, University of Washington, Seattle, Washington 98105

Received 28 August 1987/Accepted 23 November 1987

Sequencing of the E5 open reading frame (ORF) of human papillomavirus type 16 revealed an additional nucleotide, a thymidine residue, at position 3903 compared with the original sequence (Seedorf et al., *Virology* 145:181-185, 1985). The additional T had two effects; first, in reading frame 2, in which the original E5 ORF was predicted, the additional T changed the reading frame downstream of position 3903 to create an ORF, which we designated E5*, that terminated at position 4018 and potentially encoded a 52-amino-acid polypeptide. Secondly, in reading frame 3, a new ORF was created (positions 3807 to 4097), which we propose is the authentic papillomavirus type 16 E5 ORF. It contained a methionine residue and encoded an additional 82 amino acids. Both ORFs have been cloned into bacterial expression vectors (pATH), and the fusion proteins have been used to generate polyclonal antibodies in rabbits.

Human papillomavirus (HPV) DNA has been found in many human neoplasias of squamous epithelial origin (4, 8, 11, 12), and a subset of viral genes continues to be expressed in the tumors (7, 15, 20, 22, 24, 29). In addition, both animal (27) and human (14, 16, 26, 28) papillomaviruses have been shown to transform cells in culture. Together, these observations suggest that HPVs have a role in the etiology of some cancers. HPV type 16 (HPV-16), a strain that has been shown to be associated with invasive cervical neoplasia (8, 9), has a genomic size of 7,904 base pairs (23).

Recombinant DNA technology has been useful in the study of HPV since these viruses so far have not been propagated in tissue culture. Comparative sequence analysis of HPV and animal papillomavirus genomes have shown an overall genetic relatedness (6, 21). The genome is divided into an early region that contains seven or eight open reading frames (ORFs) and a late region that contains two ORFs, L1 and L2, coding for capsid proteins. Various viral genes have been identified primarily as ORFs from nucleotide sequence data for cloned viral DNA. These ORFs have been cloned into bacterial expression vectors, and the recombinant fusion proteins have been used as immunogens to generate antibodies to detect authentic viral proteins (2, 19, 25).

We have cloned and expressed all of the ORFs of HPV-16 to look for expression of various putative viral proteins in clinical specimens and in *in vitro* studies (10). It is of interest to look for E5 expression because the E5 protein of bovine papillomavirus type 1 (BPV-1) has been shown to have transforming activity (18, 19). Although evidence suggests that the E6 and E7 ORFs of HPV have a function in transformation (1, 3, 25), a role for E5 in HPV infections has yet to be defined.

The DNA fragment of HPV-16 from position 3877 to 4284 obtained from an *Sfa*NI restriction digest was cloned into the *Sma*I site of pUC18 as shown in Fig. 1. This fragment contained the major portion of the E5 ORF (3863 to 4097) as defined by the published sequence (23). Cloning of this E5 ORF to express a fusion protein was done by isolating the DNA fragment from the pUC hybrid clone following digestion with *Eco*RI and *Hind*III, enzymes which cleave within

the polylinker of the vector. The fragment was ligated to the *Eco*RI-*Hind*III-digested pATH1 expression vector (gift of T. J. Koerner) and designated p16E5sf1. The ligated DNA was transfected into *Escherichia coli* HB101 cells. The plasmid was sequenced to verify that the fragment was inserted in the correct reading frame and to confirm the nucleotide sequence of the gene encoding the protein. Sequencing of the plasmid was done by the dideoxy nucleotide triphosphate chain termination method (17) with denatured plasmid DNA as a template (5), with a primer hybridizing to a sequence 5' of the multiple cloning site in the pATH vector. Analysis of the sequence revealed that there were four T's starting at position 3903, instead of the three reported in the published sequence (23). Figure 2A displays the four T's in the sequencing gel. A primer complementary to sequences of HPV-16 located 5' of the E5 ORF (primer 1, nucleotides 3821 to 3839) was used to directly sequence the HPV-16 DNA plasmid obtained from H. zur Hausen. This confirmed the presence of the four T's in the original pHPV16 clone (data not shown). To obtain sequence from the other strand, a primer complementary to HPV-16 sequences located 3' of the E5 ORF (primer 2, nucleotides 4107 to 4122) was synthesized. Figure 2B displays the four A's on the complementary strand.

The additional T at position 3903 had two effects. First, the original E5 ORF, with the additional T, encoded a protein of 52 amino acids rather than the originally predicted 72 amino acids. Second, the additional T created another ORF beginning at position 3807 and terminating at 4097, and it encoded a methionine codon at position 3849. This ORF overlapped the original E5 ORF and extended further upstream. The new E5 ORF was in a different reading frame and potentially encoded a polypeptide of 83 amino acids starting at the methionine residue. Figure 3 shows the size and location of the initial E5 ORF described by Seedorf et al. (23), the effect of the additional T on that ORF, designated E5*, and the new E5 ORF, which contains a methionine residue. A further effect of the additional T was that it shifted the reading frame of the L2 and L1 ORFs relative to the other ORFs. The fact that the new ORF contained a methionine residue, as do the E5 ORFs of the other papillomaviruses, while the previously designated one did not, suggests

* Corresponding author.

FIG. 1. Construction of hybrid clones p16E5sf1 and p16E5sf11. A pBR plasmid containing HPV-16 (pHPV16, originally from H. zur Hausen) was digested with *Sfa*NI restriction enzyme. The location of the DNA fragment containing E5 is indicated by its nucleotide position in the genome of HPV-16. This fragment was filled in by Klenow polymerase and inserted into the *Sma*I site of pUC18. *E. coli* JM103 was transfected with the ligation product, and white colonies were screened by restriction digests followed by nucleotide sequence analysis across the ligation junction. The pUC hybrid clone was subsequently digested with *Eco*RI and *Hind*III, and the DNA fragment containing E5 was ligated to similarly digested pATH1 (designated p16E5sf1) or pATH11 (p16E5sf11) DNA.

that the ORF located at positions 3807 to 4097 is the authentic E5 ORF. However, biochemical or genetic data are needed to confirm this prediction. It is also possible that coding sequences of the E5* ORF are utilized as part of a spliced mRNA.

Cloning of a portion of the new E5 ORF for the production of a fusion protein was done similarly to the cloning of the old E5 ORF. The DNA fragment obtained from digestion of the pUC hybrid clone was ligated to *Eco*RI-*Hind*III-digested pATH11 and designated p16E5sf11.

Induction of the *trpE* fusion proteins and the preparation of salt-insoluble fractions was done as described previously (10), and the products were analyzed on sodium dodecyl sulfate (SDS)-polyacrylamide gels. Both E5 ORF fusion proteins were expressed at very low abundance (data not shown) compared with other HPV-*trpE* fusion proteins (10). The fusion protein expressed from p16E5sf1 showed a protein which was approximately 4.0 kilodaltons (kDa) larger than the 37 kDa of the *trpE* moiety, which correlated closely with the predicted size of 5 kDa. The fusion protein expressed by p16E5sf1 corresponded to a segment of the E5* ORF (Fig. 3). The fusion protein expressed from p16E5sf11 should encode an HPV moiety of 8 kDa; however, additional proteins 2 kDa larger than *trpE* and of higher molecular weight were seen. The fusion protein expressed by p16E5sf11 corresponded to a segment of the newly designated E5 ORF (Fig. 3). To visualize the E5 fusion proteins more adequately, a Western blot (immunoblot) containing both uninduced and induced bacterial lysates as well as the salt-insoluble fractions were reacted with an antibody which detected the *trpE* moiety (a gift of J. Firzlaff) (Fig. 4). This analysis indicated that p16E5sf1 expressed a 41-kDa protein, as expected. The plasmid p16E5sf11 showed a *trpE* fusion protein migrating somewhat slower than the *trpE* protein alone and another protein of 78 kDa. The most likely explanation for the aberrant mobility of the fusion protein is that the highly hydrophobic residues encoded by E5 cause this behavior in SDS gels. The higher-molecular-

FIG. 2. Analysis of the nucleotide sequence of HPV-16 E5 ORF. Dideoxy nucleotide sequencing of hybrid clones p16E5sf1 and pHPV16 was done with 32 P-labeled primers and double-stranded DNA templates. Denaturation of the template and extension reactions were done by standard procedures (5, 17). Panels show reaction products run on an 8% polyacrylamide-SDS denaturing gel, with the nucleotide lanes indicated at the top of the gel. (A) Portion of the sequence for the sense strand from positions 3891 to 3920 sequenced with a pATH primer and p16E5sf1. (B) Complementary sequence on the reverse strand sequenced by using primer 2 and pHPV16. Asterisks designate position of the four T's and four A's on the DNA strands.

FIG. 3. The E5 ORF (3863 to 4096) described by Seedorf et al. (23) is shown below, as is the amino acid sequence it predicts. The insertion of an additional T into that reading frame generates the ORF designated E5*, with the amino acid sequence it predicts shown above. In reading frame 3 the additional T creates the new E5 ORF (3807 to 4097) which encodes a methionine at position 3849. The additional T changes the reading frame of the E5, L2, and L1 ORFs relative to the other predicted ORFs of HPV-16, as shown in the top panel. aa, Amino acids.

FIG. 4. Western blot analysis of the E5 fusion proteins with polyclonal rabbit antisera with reactivity to the *trpE* moiety. Clones p16E5sf1 and p16E5sf11 were induced to produce the E5* and E5 fusion proteins, respectively. Induction and purification of proteins were done as described previously (10). Briefly, overnight cultures of bacteria were diluted 1:10 in 50 ml of M9 medium and grown for 1 to 2 h. For induction (I), isoleucine was added and cultures were grown for another 4 h. Uninduced (UI) cultures were grown for 6 h in medium containing tryptophan. Cells were pelleted, and whole bacterial lysates (WBL; 0.1 ml equivalent of culture) and high-salt-insoluble fractions (HSIF; 1 ml equivalent of culture) were run on a 12.5% polyacrylamide-SDS gel and transferred to nitrocellulose. The blot was reacted to antisera previously absorbed with *E. coli* HB101 bacterial lysates. (A) Induced *trpE* protein from pATH1 and the induced fusion protein from p16E5sf1 (arrow). The induced protein was approximately 41 kDa, or 4 kDa larger than *trpE*. (B) Induced *trpE* protein from pATH11, which seems to be less high-salt insoluble than *trpE* from pATH1, and the induced fusion protein for p16E5sf11 (indicated with an arrow). Molecular weights are shown to the right.

BPV 1: MPNLWFLFL GLVAAVMLLL LLFLLFLV YWDHFECSCT GLPF

HPV 6b: MEVVPVQIAA GTTSTFILPV IIAFVVCVFS IILIVWISF IVYTSVLVLT LLYLLLWLL LTTPLOFLL TLLVCYCPAL YIHYYIVTQ Q

HPV16: MTNLDASTT LLACFLLCFC VLLCVCLLIR PLLSVSSTYT SLIILVLLW ITAASAFRCF IVYIIFVYIP LFLIHTHARE LIT

HPV18: MLSLIFLCF CVCMYVCCHV PLLPSVCMCA YAWVLVFVYI VVITSPATAF TVYVFCFLP MLLLHIHAIL SLQ

FIG. 5. Amino acid sequence of papillomavirus E5 ORFs. The amino acid sequences are displayed with the hydrophobic residues underlined. A comparison of amino acid residues between the E5 ORFs of BPV-1 and HPV-6B, -16, and -18 shows that the similarity resides primarily in the large number of hydrophobic residues (approximately 60%), of which a significant portion are leucine residues (20% for HPV-18, 30% for HPV-16 and -6B, and 50% for BPV-1).

weight form presumably results from incomplete denaturation in SDS.

Rabbits were immunized with the high-salt-insoluble fractions of induced bacterial cultures to produce antibodies to both E5 ORFs. The antibodies produced will be used to identify the authentic viral proteins in clinical specimens and in vivo and in vitro viral expression systems.

The E5 protein of BPV-1 has been shown to be highly hydrophobic (68%), and 34% of the amino acids are leucine (19). Figure 5 displays the amino acid sequence for the newly designated E5 ORF of HPV-16 in comparison with other papillomaviral E5 ORFs. The E5* ORF was not closely related to the other E5 ORFs. A comparison of the amino acid sequences of the E5 ORFs from HPV-6b, -16, and -18 and BPV-1 revealed that the homology resided primarily in the large number of hydrophobic residues (approximately 60%), of which many were leucines (20 to 30%). The new HPV-16 E5 ORF, like the other E5 ORFs, contained a methionine at its amino terminus and Cys-X-Cys residues interspersed in a stretch of hydrophobic amino acids. Although the new HPV-16 E5 polypeptide had these properties, the distribution of a long stretch of hydrophobic residues at the 5' end with some charged residues and a Cys-X-Cys at the 3' end did not occur in it, as they do in the BPV-1 E5 ORF. Therefore, the function of E5 in HPV-16 could be different from its transforming function in BPV-1.

In summary, DNA sequence analysis of a segment of HPV-16 DNA revealed an additional nucleotide at position 3903. This insertion created a new ORF which contained a methionine residue followed by 82 amino acids, and presumably encoded the authentic E5 ORF. The identification of the E5 ORF is extremely important in the design of fusion proteins or synthetic peptides which will be used as immunogens to create HPV-16 E5 antibodies.

We thank Margaret Swain and Janette Valentine for assistance, Juliane Firzlaff and Steve Jenison for helpful suggestions, and Marci Wright for typing this manuscript.

This work was supported by Public Health Service grants P01-CA42792 and R01-CA35568 from the National Cancer Institute.

LITERATURE CITED

1. Androphy, E. J., N. L. Hubbert, J. T. Schiller, and D. R. Lowy. 1987. Identification of the HPV-16 E6 protein from transformed mouse cells and human cervical carcinoma cell lines. *EMBO J.* **6**:989-992.
2. Androphy, E. J., J. T. Schiller, and D. R. Lowy. 1985. Identification of the protein encoded by the E6 transforming gene of bovine papillomavirus. *Science* **230**:442-445.
3. Baker, C. C., W. C. Phelps, V. Lindgren, M. J. Braun, M. A. Gonda, and P. M. Howley. 1987. Structural and transcriptional analysis of human papillomavirus type 16 sequences in cervical carcinoma cell lines. *J. Virol.* **61**:962-971.
4. Boshart, M., L. Gissmann, H. Ikenberg, A. Kleinheinz, W. Scheurlen, and H. zur Hausen. 1984. A new type of papillomavirus DNA and its presence in genital cancer biopsies and in cell lines derived from cervical cancer. *EMBO J.* **3**:1151-1157.
5. Chen, E. Y., and P. H. Seeburg. 1985. Supercil sequencing: a fast and simple method of sequencing plasmid DNA. *DNA* **4**:165-170.
6. Danos, O., L. W. Engel, E. Y. Chen, M. Yaniv, and P. M. Howley. 1983. A comparative analysis of the human type 1a and bovine type 1 papillomavirus genomes. *J. Virol.* **46**:557-566.
7. Doorbar, J., D. Campbell, R. J. A. Grand, and P. H. Gallimore. 1986. Identification of human papillomavirus-1a E4 gene products. *EMBO J.* **5**:355-362.
8. Durst, M., L. Gissmann, H. Ikenberg, and H. zur Hausen. 1983. A papillomavirus DNA from a cervical carcinoma and its prevalence in cancer biopsy samples from different geographic regions. *Proc. Natl. Acad. Sci. USA* **80**:3812-3815.
9. Durst, M., A. Kleinheinz, M. Hotz, and L. Gissmann. 1985. The physical state of human papillomavirus type 16 DNA in benign and malignant genital tumors. *J. Gen. Virol.* **66**:1515-1522.
10. Firzlaff, J. M., C. N. L. Hsia, C. L. Halbert, S. A. Jenison, and D. A. Galloway. 1987. Polyclonal antibodies to human papillomavirus type 6b and type 16 bacterially derived fusion proteins, p. 105-113. In B. M. Steinberg, J. L. Brandsma, and L. B. Taichman (ed.), *Cancer cells 5: papillomaviruses*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
11. Gissmann, L., L. Wolnik, H. Ikenberg, U. Koldovsky, G. Sch-nurch, and H. zur Hausen. 1983. Human papillomavirus types 6 and 11 DNA sequences in genital and laryngeal papillomas and in some cervical cancers. *Proc. Natl. Acad. Sci. USA* **80**:560-563.
12. Gissmann, L., and H. zur Hausen. 1980. Partial characterization of viral DNA from human genital warts (condylomata acuminata). *Int. J. Cancer* **25**:605-609.
13. Kaur, P., and J. K. McDougall. 1987. Transformation of mouse cells by HPV-6b DNA and HPV-18 DNA, p. 249-252. In B. M. Steinberg, J. L. Brandsma, and L. B. Taichman (ed.), *Cancer cells 5: papillomaviruses*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
14. Kreider, J. W., M. K. Howlett, N. L. Lill, G. L. Bartlett, R. J. Zaino, T. V. Sedlacek, and R. Mortel. 1986. In vivo transformation of human skin with human papillomavirus type 11 from condylomata acuminata. *J. Virol.* **59**:369-373.
15. Lehn, H., P. Krieg, and G. Sauer. 1985. Papillomavirus genomes in human cervical tumors: analysis of their transcriptional activity. *Proc. Natl. Acad. Sci. USA* **82**:5540-5544.
16. Piri, S., S. Yasumoto, M. Feller, J. Doniger, and J. A. DiPaolo. 1987. Transformation of human fibroblasts and keratinocytes with human papillomavirus type 16 DNA. *J. Virol.* **61**:1061-1066.
17. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**:5463-5467.
18. Schiller, J. T., W. C. Vass, K. H. Voudan, and D. R. Lowy. 1986. The E5 open reading frame of bovine papillomavirus type 1 encodes a transforming gene. *J. Virol.* **57**:1-6.
19. Schlegel, R., M. Wade-Glass, M. S. Rabson, and Y. C. Yang. 1986. The E5 transforming gene of bovine papillomavirus encodes a small hydrophobic polypeptide. *Science* **233**:464-467.

20. **Schneider-Gadicke, A., and E. Schwarz.** 1986. Different human cervical carcinoma cell lines show similar transcription patterns of human papillomavirus type 18 early genes. *EMBO J.* **5**:2285–2292.
21. **Schwarz, E., M. Durst, C. Demankowski, O. Lattermann, R. Zech, E. Wolfspurger, S. Suhai, and H. zur Hausen.** 1983. DNA sequence and genome organization of genital human papillomavirus type 6b. *EMBO J.* **2**:2341–2348.
22. **Schwarz, E., U. K. Freese, L. Gissmann, W. Mayer, B. Roggenbuck, A. Stremlau, and H. zur Hausen.** 1985. Structure and transcription of human papillomavirus sequences in cervical carcinoma cells. *Nature (London)* **314**:111–114.
23. **Seedorf, K., G. Krammer, M. Durst, S. Suhai, and W. Rowekamp.** 1985. Human papillomavirus type 16 DNA sequence. *Virology* **145**:181–185.
24. **Seedorf, K., T. Oltersdorf, G. Krammer, and W. Rowekamp.** 1987. Identification of early proteins of the human papilloma viruses type 16 (HPV 16) and type 18 (HPV 18) in cervical carcinoma cells. *EMBO J.* **6**:139–144.
25. **Smotkin, D., and F. O. Wettstein.** 1986. Transcription of human papillomavirus type 16 early genes in a cervical cancer and a cancer-derived cell line and identification of the E7 protein. *Proc. Natl. Acad. Sci. USA* **83**:4680–4684.
26. **Watts, S. L., W. C. Phelps, R. S. Ostrow, K. R. Zachow, and A. J. Faras.** 1984. Cellular transformation by human papillomavirus DNA in vitro. *Science* **225**:634–636.
27. **Yang, Y.-C., H. Okayama, and P. M. Howley.** 1985. Bovine papillomavirus contains multiple transforming genes. *Proc. Natl. Acad. Sci. USA* **82**:1030–1034.
28. **Yasumoto, S., A. L. Burkhardt, J. Doniger, and J. A. DiPaolo.** 1986. Human papillomavirus type 16 DNA-induced malignant transformation of NIH 3T3 cells. *J. Virol.* **57**:572–577.
29. **Yee, C., I. Krishnan-Hewlett, C. C. Baker, R. Schlegel, and P. M. Howley.** 1985. Presence and expression of human papillomavirus sequences in human cervical carcinoma cell lines. *Am. J. Pathol.* **119**:361–366.