The EMBO Journal vol.5 no.2 pp.355-362, 1986

# Identification of the human papilloma virus-1a E4 gene products

J.Doorbar, D.Campbell[1], R.J.A.Grand and P.H.Gallimore

Cancer Research Campaign Laboratories, Department of Cancer Studies, University of Birmingham, Birmingham B15 2TJ, and [1]Department of Biochemistry, University of Dundee, Dundee DD1 4HN, UK

Communicated by U.Pettersson

**Antibodies prepared against a human papilloma virus-1 (HPV-1) E4/β-galactosidase fusion protein identified several polypeptides in HPV-1, but not HPV-2 or 4, induced papillomas. The major E4 protein, that represented up to 30% of total cellular protein, was a 16/17-K doublet which was purified by column chromatography and analysed for amino acid content. A peptide derived by chymotryptic digestion was purified by h.p.l.c. and subjected to amino acid sequencing. The unique sequence obtained, Gly-His-Pro-Asp-Leu-Ser-Leu, identified the 16/17-K doublet as a product of the HPV-1 E4 gene region. Antibodies to both the E4/β-galactosidase fusion protein and the 16/17-K doublet identified two smaller polypeptides (10/11-K) which may represent spliced products of E4. We propose that the products of the HPV-1 E4 gene region are not classical DNA tumor virus early proteins and suggest that they play a role in virus maturation.**

*Key words:* HPV-1 E4/antibodies/papilloma proteins/fusion proteins

## Introduction

Papillomaviruses are highly host and tissue specific, and produce benign epithelial proliferations (warts) in man and in a variety of animals (for review, see Pfister, 1984). Although >25 types of human papilloma-virus (HPV) have been identified to date (Pfister, 1984; Kremsdorf *et al.*, 1984; Gassenmaier *et al.*, 1984) our understanding of the life cycle of these viruses has been restricted by the lack of an *in vitro* culture system in which the HPVs can be successfully propagated. Recent advances in molecular biology have provided some insight into the genetic organisation of these viruses, and the complete DNA sequence of five papillomavirus genomes HPV-1a, HPV-6b, HPV-16, bovine papillomavirus type 1 (BPV-1) and cotton tail rabbit papillomavirus (CRPV) has been reported and recently reviewed (Danos *et al.*, 1984; Seedorf *et al.*, 1985).

The papillomavirus genome has been divided into two regions, early and late, by analogy with polyoma virus and simian virus 40 (SV40). The 'early' region corresponds to a 69% fragment of the BPV-1 genome which is sufficient to transform mouse cells (Lowy *et al.*, 1980). To date, no virus-specific proteins have been isolated from the 'early' region although functions related to transformation and virus DNA replication have been assigned to a number of BPV-1 potential open reading frames (ORFs; Schiller *et al.*, 1984; Yang *et al.*, 1985; Sarver *et al.*, 1984; Lusky and Botchan, 1985).

The 'late' region of the papillomavirus genome corresponds to that part of the BPV-1 genome not required for *in vitro* transformation, and contains two large ORFs which are thought to encode the virion structural proteins L1 and L2 (reviewed by Danos *et al.*, 1984). By indirect evidence obtained from amino acid analysis, the major capsid protein purified from BPV-1 viral particles has been assigned to the L1 open reading frame (Meinke and Meinke, 1981), and more recently it has been shown that a biosynthetically produced L1 protein prepared in *Escherichia coli*, reacts strongly with antisera prepared against BPV-1 virions (Pilacinski *et al.*, 1984). In a similar experiment, antisera prepared against a biosynthetic BPV-1 L2 protein was shown to react with BPV-1 virions, suggesting that the L2 ORF may encode a minor structural component which has not yet been identified (Pilacinski *et al.*, 1984).

In order to identify the unknown proteins encoded by the open reading frames of the human papillomaviruses, we have raised antisera against HPV-1a gene fragments expressed in *E. coli*, and have used these antisera to identify and characterise the native gene products in naturally occurring papillomas. This approach has been used previously by others to identify the unknown protein products of several cloned genes (Reed, 1982; Greenspan *et al.*, 1985), and with the recent development of improved expression vectors (Rüther and Muller-Hill, 1983; Stanley and Luzio, 1984) it is likely to become an established technique.

In this paper we report the identification and isolation of the product of the HPV-1 E4 open reading frame utilizing antibodies made against cro-β-galactosidase/HPV-1a E4 fusion protein.

## Results

### *Constructions and identification of HPV-1a expression plasmids*

A number of DNA fragments from the major open reading frames of HPV-1a have been cloned into the pEX series of expression vectors (Stanley and Luzio, 1984) as illustrated in Figure 1.

On induction, all the HPV-1a/pEX recombinants were found to express hybrid polypeptides of the predicted size, and degradation was not observed, even when large fragments were expressed (up to 900 bp). Following a two-stage purification protocol (see Materials and methods) ~20-40 mg of fusion protein (>95% pure as determined by polyacrylamide gel electrophoresis) could be prepared from 1 litre of bacterial culture.

High titre antisera (up to 1/10 000 by ELISA) were subsequently raised against fusion proteins expressed from each of the recombinants described in Figure 1.

### *Immunological screening of wart extracts*

Wart biopsies were typed by Southern blotting, and approximately half the verrucas we obtained were found to be caused by HPV-1. The remainder were predominantly HPV-2 induced, although HPV-4 was detected in a small percentage of samples. Nearly all the common warts were found to contain HPV-2. In the first instance the fusion protein antisera were tested for their reactivity against naturally occurring papillomavirus-encoded proteins by Western blotting, after solubilization of individual wart biopsies in urea.

The anti-E4 antibody detected two major protein doublets, of approximate mol. wt 16/17 K and 10/11 K, as well as two minor

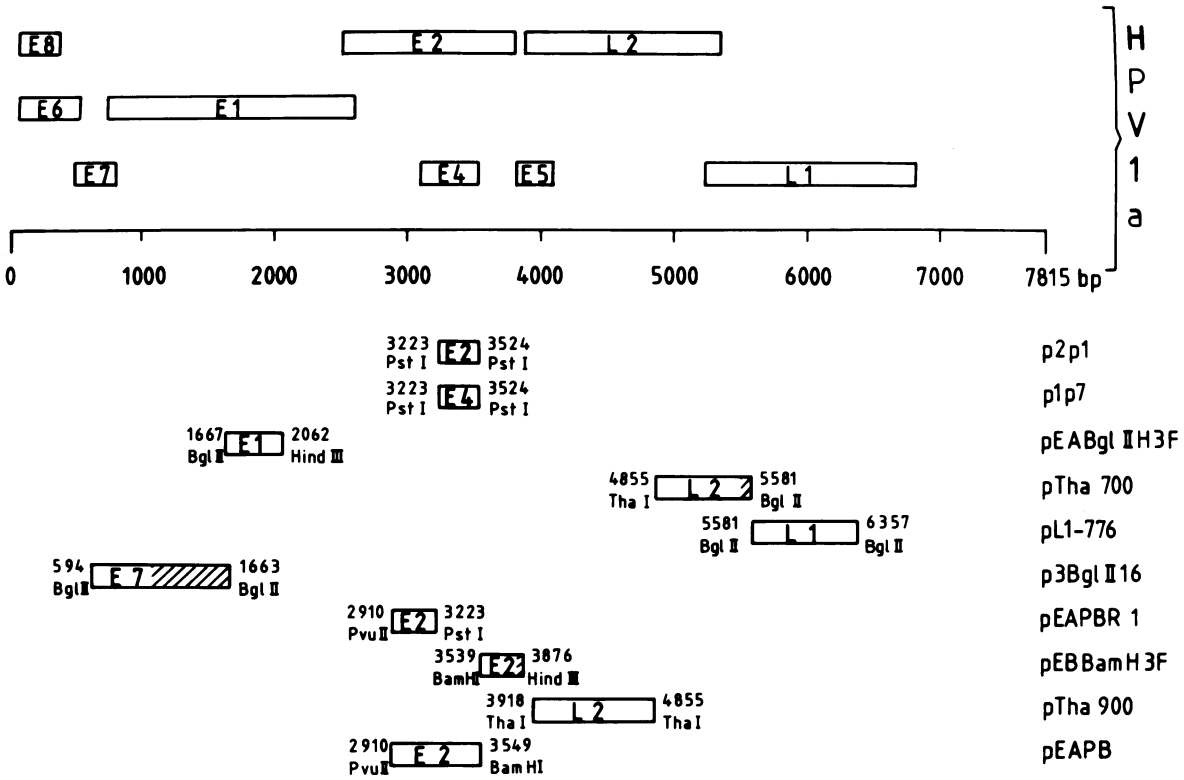



**Fig. 1.** HPV DNA fragments expressed to high level in *E. coli.* The positions of the major open reading frames of HPV type 1 are illustrated at the top of the figure. The location of fragments which have been cloned into the pEX series of expression vectors are shown beneath, along with the restriction enzyme sites which were used to isolate them. The designations of plasmids containing each region are indicated at the right of the figure.

doublets of 32/34 K and 21/23 K (Figure 2). The reaction was highly specific for HPV-1-induced wart extracts, and no equivalent bands were observed in extracts of HPV-2 (Figure 2) or HPV-4 induced papillomas (data not shown). Similarly, these protein species were completely absent in extracts prepared from normal skin biopsies, which had been removed from the dorsal and palmar surfaces of the hand, and which represented potential sites for natural infection by HPV-1. Antisera of equivalent titre raised against other HPV-1a/β-Gal fusion proteins (see Figure 1) did not detect any of the protein species described above. It seems therefore that the specific detection of the protein doublets described for the anti-p1p7 (E4) sera is due to antibodies directed against the region from the E4 ORF, rather than any part of the β-galactosidase molecule. Surprisingly, the 16/17-K protein doublet was clearly visible in HPV-1-induced wart extracts after SDS-gel electrophoresis and staining with Coomassie blue (Figure 3a), although no similar-sized protein could be visualised in HPV-2-induced warts (Figure 3b) or in normal skin samples (Figure 3c). The two proteins were usually observed together, although on some occasions one or other was depleted or was absent altogether. Two of the biopsies analysed (out of a total of 10) showed only the 16-K species. By densitometric scanning of stained gels, the two proteins were shown to represent up to 30% of the total SDS-extractable protein obtained from HPV-1-induced warts. A similar observation has been made by Croissant *et al.* (1985). On the alkaline urea gel system a characteristic staining pattern was observed, both by Coomassie blue detection and by Western blotting using anti-p1p7(E4) antibody; five distinct protein bands were apparent, probably representing charged variants of the 16/17-K proteins (Figure 3d). In some cases, the 10/11-K doublet was also visible by Coomassie blue staining and could represent up to 5%

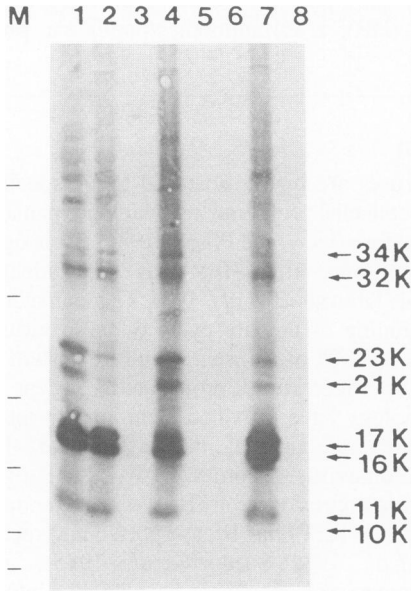
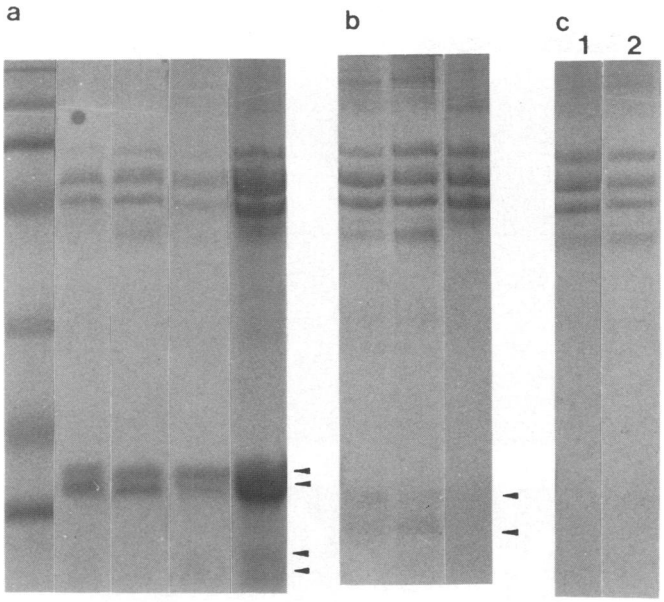



**Fig. 2.** Identification of E4-encoded proteins by Western blotting with anti-p1p7 (E4) sera. **Tracks 1, 2, 4** and **7** contain extracts of HPV-1-induced papillomas. **Tracks 3, 5** and **6** contain extracts of HPV-2-induced papillomas. **Track 8** contains an extract of normal skin obtained from the palmar surface of the hand. The positions of the major 16/17-K and 10/11-K bands are indicated, as well as the minor 21/23-K and 32/34-K proteins. Mol. wt markers (43, 25.7, 18.4, 14.3, 6.2 and 3.0 K) are shown in **track M**.



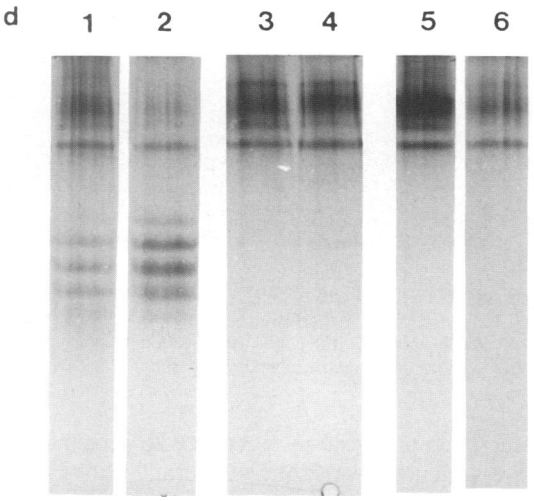




**Fig. 3.** Identification of papilloma proteins by gel electrophoresis and staining with Coomassie blue. **(a)** SDS gel of total protein extracts from four HPV-1-induced warts. The positions of the 16/17-K and 10/11-K proteins are indicated. Mol. wt markers (200, 97.4, 68, 43, 25.7, 18.4 and 14.3 K) are shown on the left. **(b)** SDS gel of total protein extracts from three HPV-2-induced warts showing the positions of the 13-K and 15-K proteins. **(c)** SDS gel of total protein extracts from two normal skin biopsies obtained from either the dorsal (**track 1**) or palmar (**track 2**) surfaces of the hand. **(d)** Alkaline urea gel of total protein extracts from: HPV-1-induced wart extracts (**tracks 1** and **2**); HPV-2-induced wart extracts (**tracks 3** and **4**); two normal skin extracts obtained from either the dorsal (**track 5**) or palmar (**track 6**) surface of the hand. The five bands identified in the centre of **tracks 1** and **2** are thought to represent charged variants of the 16/17-K doublet.

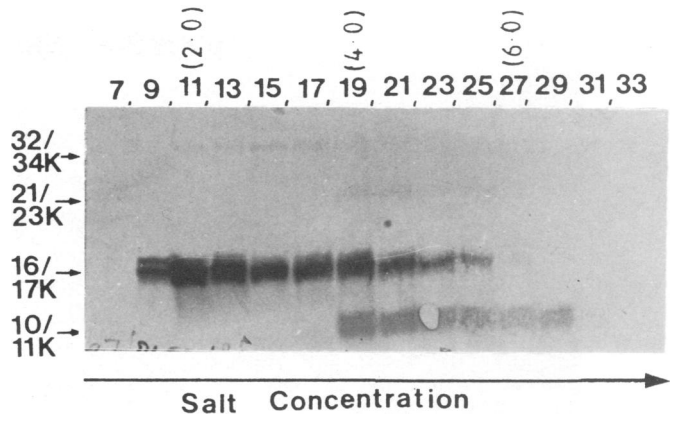



**Fig. 4.** Detection of the 16/17-K and 10/11-K proteins by Western blotting [using anti-p1p7 (E4) sera] after elution from a DEAE 52 ion exchange column. The fraction number is indicated above each track. Numbers in brackets indicate the conductivity of each fraction (ms/cm). The arrow beneath the gel indicates increasing salt concentration. The positions of the 16/17-K, 32/34-K, 10/11-K and 21/23-K proteins are indicated.

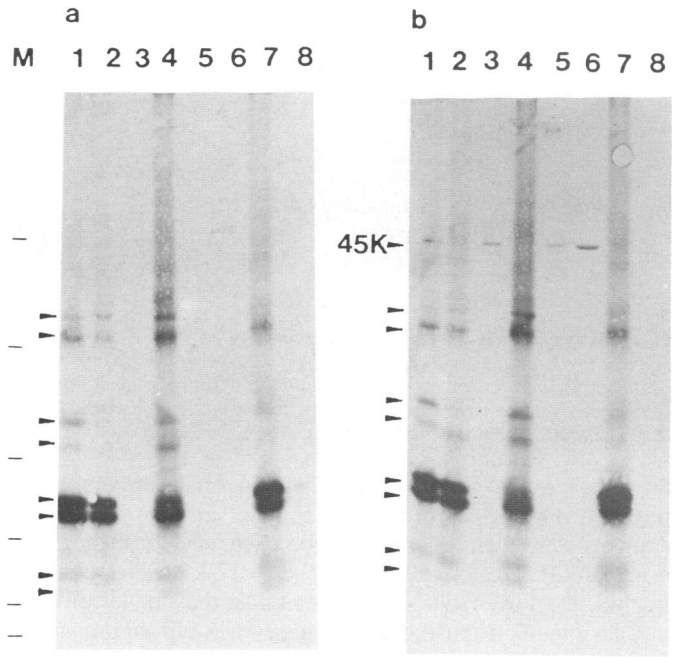



**Fig. 5.** Identification of proteins by Western blotting with anti-16/17 K and anti-10/11 K sera. **(a)** Western blot using anti-16/17 K sera; **tracks 1, 2, 4** and **7** contain extracts of HPV-1-induced papillomas. **Tracks 3, 5** and **6** contain extracts of HPV-2-induced papillomas. **Track 8** contains an extract of normal skin obtained from the palmar surface of the hand. The positions of the 10/11-K, 16/17-K, 21/23-K and 32/34-K proteins are indicated by arrows. **(b)** Western blot using anti-10/11 K sera; the contents of each track are as described above. The position of the additional 45-K protein is indicated (45 K). Mol. wt markers (43, 25.7, 18.4, 14.3, 6.2 and 3.0 K) are indicated in **track M**.

of the total SDS-extractable protein. Its presence, however, was highly variable, and often it could only be detected by Western blotting. The larger polypeptides of 32/34 K and 21/23 K were never visualised by direct staining with Coomassie blue, but were consistently detected using the anti-p1p7 (E4) antisera in Western blots. Although HPV-2-induced warts were shown not to contain these polypeptides, two faint low mol. wt (15 K and 13 K) protein bands (Figure 3b) were visible after SDS-polyacrylamide gel electrophoresis which were not apparent in the extracts of HPV-1a-induced warts (Figure 3a) or in the normal skin biopsies. Whether these proteins represent the HPV-2 equivalent of the HPV-1 16/17-K species is at present unclear.

*Purification of 16/17-K and 10/11-K proteins and preparation of antisera*

To confirm the viral origin of the 16/17-K and 10/11-K proteins, they were purified from an individual wart biopsy, and were used to prepare specific antisera. The proteins were first purified

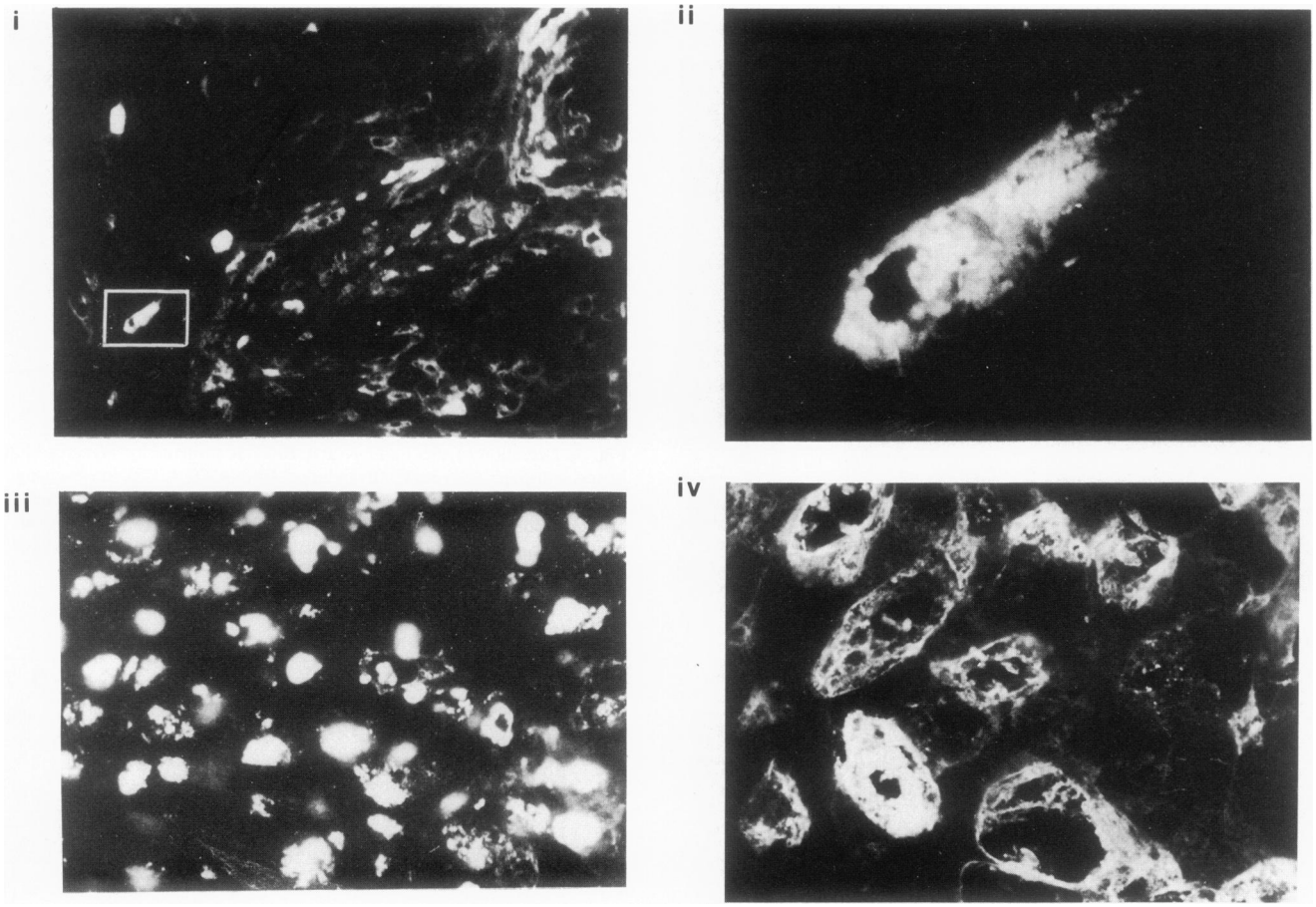



**Fig. 6.** Immunofluorescence staining of sections of an HPV-1-induced papilloma using anti-16/17 K and anti-p1p7 (E4) sera. **(i)** Vertical section through the granular layer of an HPV-1-induced papilloma, showing cytoplasmic staining with anti-16/17 K serum. Magnification 113×. **(ii)** High magnification (1025×) of immunofluorescent cell shown boxed in **(i)** to illustrate staining of the cytoplasmic inclusion granules. **(iii)** High magnification showing nuclear staining of cells in the granular layer using an anti-L1 fusion protein antiserum. Magnification 564×. **(iv)** High magnification showing staining with anti-p1p7 (E4) sera in a similar area, to that described in **(iii)**. Magnification 530×.

from their major contaminants, the epidermal keratins, by gel filtration on a Sepharose 6B-Cl column, and the two protein doublets were then separated by ion exchange chromatography using DEAE 52 anion exchanger. The 16/17-K doublet was found to elute at a lower salt concentration than the smaller proteins, thus facilitating its purification (Figure 4). A typical preparation from a single wart yielded ~500 μg of pure 16/17-K protein and ~1–2 mg of a mixture of the 10/11-K and 16/17-K doublets, which could not be separated completely by this protocol. Pure 10/11-K protein was obtained in lower amounts, and ~50 μg could be typically obtained from one preparation. The minor 32/34-K and 21/23-K proteins were found to co-purify with the 16/17-K and 10/11-K species, respectively and may represent dimers of these proteins. If this is the case then disulphide cross-links are probably not involved, as they could not be converted to their suspected monomeric form following reduction with β-mercaptoethanol. High titre antisera (>1/5000 by ELISA) were prepared against the purified 16/17-K and 10/11-K polypeptides and were screened against bacterial fusion proteins containing segments of the putative HPV-1a E2, E1, E7, L2 and L1 ORFs, as well as the E4 ORF. Both the anti-10/11 K and anti-16/17 K sera were found to react strongly with the E4 fusion protein (expressed from clone p1p7), but did not react with the expressed fragments of other HPV-1a ORFs. By Western blotting against wart extracts the anti-16/17 K and anti-10/11 K sera detected the same protein species as did the anti-p1p7(E4) serum, and produced no reaction with proteins from HPV-2 or HPV-4-induced warts, or from normal skin (Figure 5). In addition, the anti-10/11 K antiserum was found to detect an additional protein species of 45 K which was present in both HPV-1 and HPV-2 extracts, but which was absent from three adult skin biopsies. This 45 K protein is therefore antigenically related to the HPV-1a 10/11-K polypeptides, but not to the 16/17-K protein or to the p1p7 (E4) fusion protein. Furthermore this protein was not detected using antisera raised against other HPV-1a fusion proteins, which contained fragments from the putative E1, E2, E7, L2 and L1 proteins, as described in Figure 1.

By immunofluorescence analysis carried out on HPV-1-induced warts, all three antisera [anti-16/17 K, anti-10/11 K and anti-p1p7(E4)] showed diffuse cytoplasmic staining in cells of the strata germinatum and spinosum. Strong cytoplasmic fluorescence was observed in the stratum granulosum and in areas of the stratum corneum. The most intense staining was seen in those areas of the papilloma, identified by an anti-L1 serum, to be undergoing productive infection (Figure 6) and was often associated with the cytoplasmic inclusions pathognomic of HPV-1-induced warts.

*Amino acid and sequence analysis of the 16/17-K protein*

The HPV-1a DNA fragment contained in our p1p7(E4) clone was sequenced, and was found to be identical to that previously published for the E4 ORF (Danos *et al.*, 1982). This confirmed

**Table I.** Amino acid analyses of the 16/17-K and 10/11-K proteins isolated from HPV-1 warts and peptides F and C purified after chymotryptic digestion of the 16/17-K protein

| | 16/17-K protein | | 10/11-K protein | Peptide F | | Peptide C | |
|---|---|---|---|---|---|---|---|
| | Found[a] | Expected | Found[a] | Found | Expected | Found | Expected |
| Asp | 14 | 15 | 9 | 0.8 | 1 | 3.6 | 4 |
| Thr | 8 | 9 | 4 | | – | 1.0 | 1 |
| Ser | 6 | 7 | 6 | 1.1 | 1 | 1.0 | – |
| Glu | 20 | 16 | 14 | | – | 6.2 | 8 |
| Pro | 17 | 15 | 13 | 1.0 | 1 | + | 3 |
| Gly | 12 | 7 | 11 | 1.2 | 1 | 2.6 | 1 |
| Ala | 3 | 2 | 5 | | – | 0.5 | – |
| Val | 5 | 5 | 4 | | – | 1.7 | 2 |
| Met | 1 | 1 | + | | – | | – |
| Ile | 3 | 2 | 3 | | – | | – |
| Leu | 12 | 18 | 7 | 1.9 | 2 | 2.0 | 1 |
| Tyr | 2 | 4 | 1 | | – | | – |
| Phe | 2 | 1 | 2 | | – | | – |
| His | 3 | 3 | 1 | 0.6 | 1 | | – |
| Lys | 7 | 5 | 4 | | – | 1.1 | 1 |
| Arg | 9 | 13 | 4 | | – | 0.8 | 1 |
| Cys | + | 2 | + | | – | | – |
| Trp | ND | 1 | ND | ND | – | ND | – |

[a]Results (mol/mol) are the mean of duplicate samples after 24-h and 72-h hydrolyses.
Threonine and serine values calculated after extrapolation to zero time. Results are given to the nearest integer. + indicates an amino acid present on analysis but where accurate quantification was not possible. ND not determined. The expected values are those calculated from the DNA sequence of the E4 ORF (Danos *et al.*, 1982). In the peptide analyses (given as mol/mol) impurities below 0.3 mol/mol have been omitted.

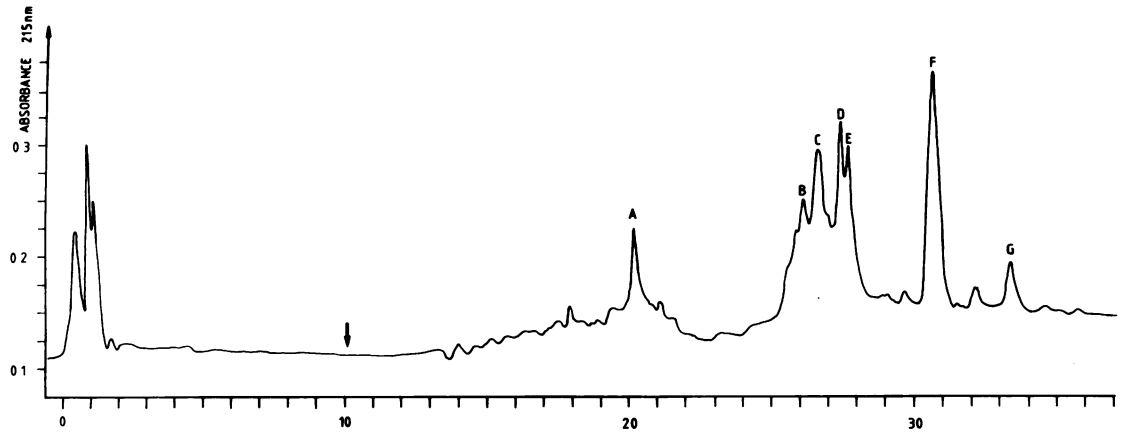



**Fig. 7.** Separation of peptides (by h.p.l.c.), obtained after chymotryptic digestion of the purified 16/17-K protein. The arrow indicates the start of the acetonitrile gradient (30% acetonitrile). Elution was complete at 50 min (80% acetonitrile). The division along the horizontal axis indicate minutes. Major peaks are indicated **A–G**.

that the antibodies which detected the 16/17-K and 10/11-K proteins were in fact directed against the putative product of the HPV-1a E4 ORF. Further evidence that these proteins are *bona fide* E4 gene products was obtained by total amino acid analysis of the purified 16/17-K polypeptides (Table I). The composition of the purified protein was found to be very similar to that predicted from the previously published HPV-1a sequence (Danos *et al.*, 1982). Certain anomalies were observed, however: glycine was found to be present in larger amounts than was expected, assuming the E4 protein to be unspliced and less leucine was detected than would have been predicted. This latter result may be partly explained by the presence of a number of Leu-Leu bonds in the proteins which are incompletely hydrolysed in 6 M HCl even after 72 h.

The 16/17-K protein was found to be blocked at its N terminus and could not be sequenced directly. Attempts to digest the protein using cyanogen bromide were unsuccessful, suggesting that it lacked an internal methionine residue. This is also true of the putative E4 amino acid sequence. Digestion with chymotrypsin, however, yielded a number of peptides which were purified by h.p.l.c. (Figure 7). The amino acid compositions determined for two of the peptides are presented in Table I together with predicted compositions for two areas of the E4 ORF which match the analyses and which might be expected as peptides after chymotryptic cleavage. The presence of extra moles of serine, glycine, alanine and leucine in the analysis of peptide C (Table I) is almost certainly due to contamination with other peptides which were not completely resolved during the h.p.l.c. run (Figure 7).

The amino acid sequence of peptide F was determined and is presented in Figure 8. A single phenylthiohydantoin (PTH)-amino acid was identified unambiguously at each step of the Edman

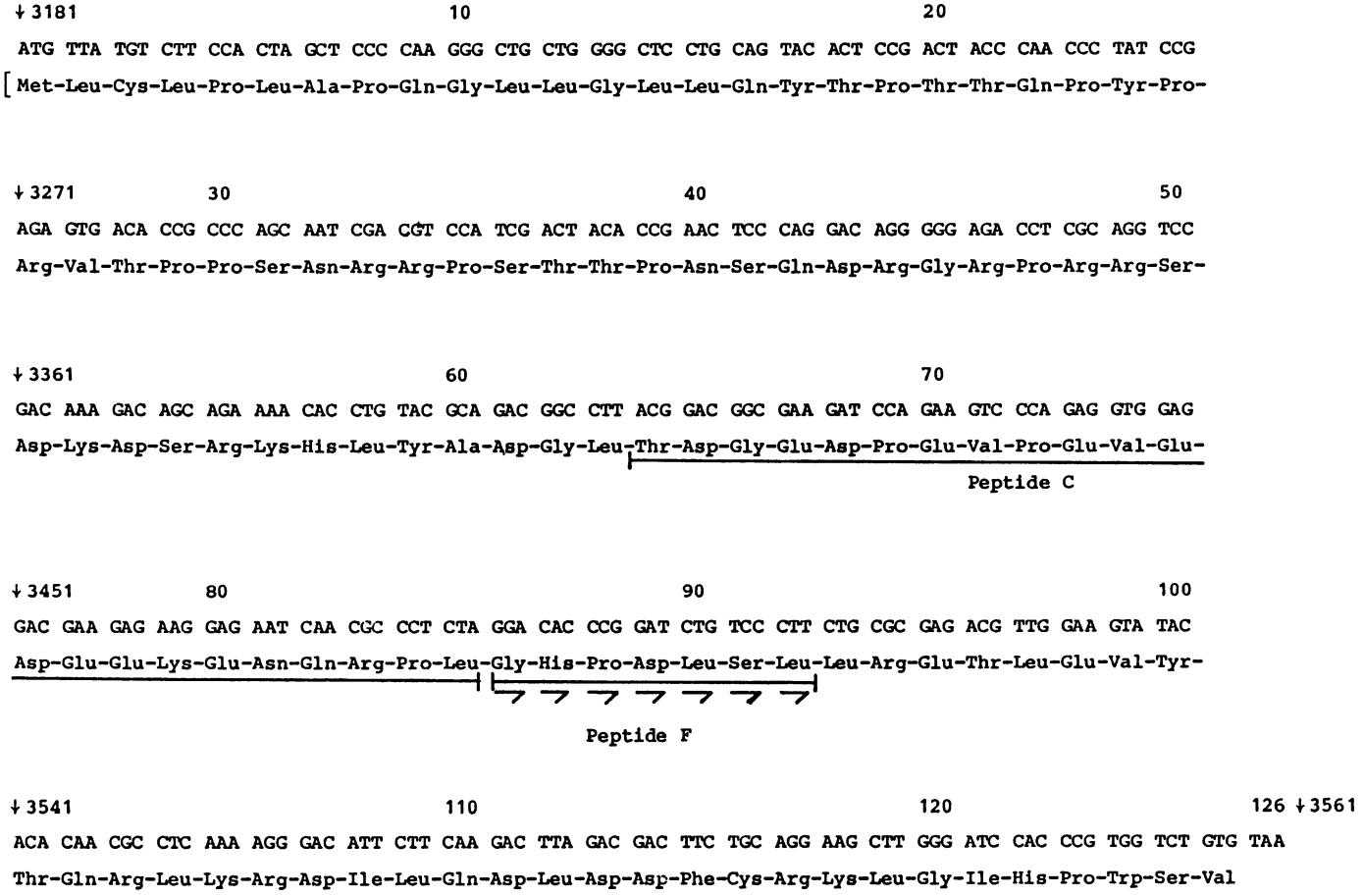



**Fig. 8.** Nucleotide and amino acid sequence of the HPV-1a E4 open reading frame. The nucleotide sequence, as determined by Danos *et al.* (1982) is given on the top line and the presumed amino acid sequence underneath. The two peptides which have been purified and analysed are shown. ⟶indicates one step of the Edman degradation. [ indicates a blocked N terminus. Nucleotide numbers are given on the left of the figure and the amino acid residue numbers above the sequence.

degradation except for cycle 2 where 0.5 nmol of isoleucine were detected as well as 0.6 nmol of histidine: no explanation for this observation can be offered at present. The sequenced peptide was a heptamer identical to a stretch of amino acids (residues 86–92) located towards the C terminus of the putative E4 protein. Assuming that each amino acid has an equal likelihood of occurring in each position, the probability of obtaining an identical sequence by chance is $\sim 8 \times 10^{-10}$ (i.e. $1/20^7$). When the individual amino acids are weighted according to their average frequency of occurrence (Dayhoff *et al.*, 1978) a slightly higher probability of $1.2 \times 10^{-9}$ is obtained. Searches of DNA and protein data bases confirmed that this 7-residue peptide had not been reported previously to occur in any nucleotide or amino acid sequence other than HPV-1a. This provided further confirmatory evidence that the purified protein was the unique product of the E4 ORF.

## Discussion

By using antibodies specific for a bacterially expressed region of the HPV-1a E4 ORF, we have identified two major proteins and a number of minor species which are consistently present in HPV-1-induced papillomas. These E4 proteins were found exclusively in warts induced by HPV-1, and were not detected in HPV-2- or HPV-4-induced papillomas, or in extracts of palmar epidermis — a natural site for infection by HPV-1 (Gissmann *et al.*, 1977; Pfister *et al.*, 1979). Amino acid analysis (Table I) and partial sequencing of the purified 16/17-K proteins confirmed them to be *bona fide* E4 gene products. The ability of the E4 ORF to encode a protein of 126 amino acids with a predicted mol. wt of 14 446 is in close agreement with our identification of the E4 product as a 16/17-K doublet. Our findings not only represent the first identification of a protein encoded by the 'early' region of HPV-1, but are in fact, the first characterisation of any such papilloma-virus-encoded 'early' protein.

Antibodies raised against expressed fragments of the HPV-1a E1, E2, E7, L1 and L2 ORFs failed to detect the 16/17-K E4 protein, suggesting that the E4 ORF is not spliced to these regions of the HPV-1a genome. Similarly, antisera prepared against the purified 16/17-K and 10/11-K proteins reacted only with the E4 fusion protein, and not with fusion proteins containing the expressed portions of the HPV-1a genome described in Figure 1. However, our studies do not rule out the possibility that other ORFs or parts of ORFs are spliced onto E4 or parts of E4. The 10/11-K E4-related protein may be a degradation product of the 16/17-K doublet, or alternatively may be the translated product from a spliced mRNA containing a small region of the E4 protein. This latter hypothesis is supported by the observation that the amino acid composition presented in Table I for the 10/11-K protein does not correspond to either the N- or C-terminal two thirds of the E4 ORF and that anti-10/11 K serum detects a 45-K protein which was not detected by either anti-16/17 K or anti-p1p7(E4) sera. At present no explanation can be offered for the presence of one protein of 17 K and one of 16 K encoded by

E4 ORF. It is possible that the amino acids present in the two peptide chains could be identical and that the separation into two components is due to some post-translational modification. However, this hypothesis awaits confirmation.

The level of E4 expression may be directly related to the quantity of virus in a given papilloma. Pertinent to this idea is our observation that the level of 16/17-K protein, as detected by our p1p7 (E4) antiserum, appears to correlate with the levels of both HPV-1a DNA and virus structural proteins (detected using antisera raised against L1 and L2 fusion proteins) in individual HPV-1-induced warts (unpublished data). Warts induced by HPV-1 generally produce large quantities of virus particles, (up to $10^{12}$ particles/wart) as compared with common warts induced by other HPV sub-types (Barrera-Oro *et al.*, 19862; and our unpublished data). The two faint Coomassie blue-stained bands of 15 K and 13 K which were frequently seen following gel electrophoresis of HPV-2-induced wart extracts may represent the HPV-2 E4 proteins.

The fact that the HPV-1a E4 16/17-K protein represents up to 30% of total wart protein indicates that it is not one of the classical DNA tumour virus (e.g. polyomavirus, SV40) early gene products which are typically expressed at very low levels both in lytically infected and transformed cells. Additionally the immunofluorescence data presented here suggests that the E4 gene products may be involved in virus maturation, a role not ascribed to 'early' gene products encoded in the transforming regions of other papova viruses. On the basis of the cytoplasmic localization of the E4 products and their abundance in the higher levels of the wart, we postulate that they might serve as scaffold, transport or structural proteins, or might have other functions, e.g. interference in keratin metabolism. In a preliminary study on one SDS-disrupted CsCl-purified virus preparation, the 16/17-K doublet was identified. This preparation revealed no contaminating epidermal keratins (by Coomassie blue staining) but the E4 doublet was clearly visible by Western blotting and by staining with Coomassie blue (our unpublished results). No such proteins have previously been reported in HPV-1a virion preparations (Favre *et al.*, 1975; Pfister and zur Hausen, 1977, 1978) although small histone-like proteins have been frequently observed. Two small proteins, reminiscent of the E4 16/17-K species have, however, been observed in 'heavy full' HPV particles (Pfister *et al.*, 1977). Further studies are required to confirm whether the 16/17-K proteins are in fact present in the virion. Whatever the outcome of these studies, evidence is accumulating from experiments in another laboratory that a mRNA transcribed from E4 is the major transcript in BPV-1 fibropapillomas (C.C. Baker and P.M. Howley, personal communication).

From our data and from the evidence that papillomavirus mRNAs have a unique polarity (Pettersson *et al.*, 1985) a clear distinction can be drawn between these viruses and SV40 and polyoma making re-classification of the papovavirus group desirable.

## Materials and methods

*Bacterial strains, plasmid vectors*

*E. coli* K12ΔHΔtrp [$Sm^R$,lac Zam,Δbio-Uvr,Δtrp EA2, (λNam 7, Nam 53, cI857, ΔH1)] (Bernard *et al.*, 1979) was the host used for the growth of all the pEX recombinants. Plasmid vectors pEX1, pEX2, and pEX3 have been described by Stanley and Luzio (1984). HPV-1a fragments were obtained from plasmids pHPV-1 (monomer) (Burnett and Gallimore, 1983) or from the following subclones prepared in this laboratory; pEA(HPV-1) comprises the HPV1a *Eco*RI−*Bam*HI fragment (967−3542) and pEB(HPV-1) the *Bam*HI−*Eco*RI fragment (3542−7778) cloned between the *Bam*HI and *Eco*RI sites of pBR322, p2(HPV-1) consists of the *Bam*HI−*Pvu*II fragment (3542−7815/0−744) cloned between the *Pvu*II and *Bam*HI sites of pBR322, and p2.1 (HPV-1) contains the HPV-1a *Eco*RI−*Pvu*II fragment (964−2910) cloned into the *Pvu*II and *Eco*RI sites of pBR322. p(HPV-2) and p(HPV-4) contain the complete genomes of either HPV-2 or HPV-4, respectively cloned into pBR322 (Burnett, 1983).

The basic growth media used were L-broth (1% bacto-tryptone, 0.5% yeast extract, 1% NaCl, pH 7.5) and L-plates (1.5% agar in L-broth). Ampicillin was used at 50 μg/ml in liquid cultures and in plates.

*Construction and identification of fusion proteins*

50 ng of gel-purified HPV-1a fragment was ligated to 25 ng of linearized and phosphatase-treated vector (Maniatis *et al.*, 1982) and the ligation mixture was used to transform competent *E. coli* K12ΔHΔtrp as described by Hanahan (1983) except that the heat shock was carried out at 32°C and subsequent incubation was at 28°C for 36 h. The colonies obtained were transferred to 1 ml of L-broth, and were grown with vigorous agitation (28°C) until they reached early/mid log phase (i.e. $O.D_{550} = 0.3 \rightarrow 0.5$, $1-2 \times 10^8$ cells/ml). The incubation temperature was then raised to 42°C, and induction was carried out for 90 min before the bacteria were pelleted, washed and resuspended in 1.5 ml of 50 mM glucose, 10 mM EDTA, 25 mM Tris-HCl (pH 8.0) containing 10 mg/ml lysozyme. After 5 min, 10 μl of 0.5 M EDTA (pH 8.0) was added, and after a further 2 min, the spheroplasts were lysed by the addition of 300 μl of 1.5% Triton, 25 mM Tris-HCl pH 8.0. Insoluble material was pelleted (10 000 *g*, 10 min), solubilised in lysis buffer [5% (w/v) SDS, 50 mM β-mercaptoethanol, 50 mM Tris-HCl (pH 8.0)] and analysed directly by SDS-polyacrylamide gel electrophoresis (7% gel) by the method of Laemmli (1970), except that the stacking gel was omitted.

Plasmid DNA was prepared from recombinants expressing cro-β-galactosidase fusion protein of increased size, and was analysed by restriction enzyme digestion (Birnboim and Doly, 1979).

*DNA sequencing of E4 expression clone*

DNA sequencing was carried out using the chain termination method of Sanger *et al.* (1977). For this purpose, the 322-bp *Bam*HI−*Hind*III fragment (containing the HPV-1a *Pst* fragment, 3227−3529) was gel purified from p1p7 and was inserted between the *Bam*HI−*Hind*III sites of M13mp9. After transformation of competent *E. coli* JM101, clear plaques were selected (Messing *et al.*, 1981) and single-stranded DNA was prepared as described by Winter and Fields (1980). The sequencing reaction was initiated from a 17-nucleotide direct sequencing primer (Amersham International).

*Large-scale preparation of fusion proteins*

A 500-ml culture was grown overnight (28°C) with gentle shaking to an $O.D_{500}$ of 0.3−0.5 ($1-2 \times 10^8$ bacteria/ml), before being induced at 42°C (vigorous shaking) for 90 min. It was important that the bacteria were in early/mid log phase at the time of induction, as growth beyond this stage led to an increase in total bacterial proteins without a significant increase in the levels of fusion protein. The cells were pelleted and were lysed by incubation with lysozyme (5 mg/ml) in 25% sucrose, 50 mM Tris-HCl (pH 8.0) for 5 min. EDTA was then added to 50 mM and, after a further 5 min, Triton X-100 was added to 1%. After 10 min incubation, the cell debris (which contained the fusion protein) was sedimented (39 000 *g*, 30 min, 4°C). The pellet was solubilised by homogenisation using a Polytron, sonication ($3 \times$ 20-s bursts), and boiling (5 min) in 5 ml of 5% (w/v) SDS, 50 mM β-mercaptoethanol, 50 mM Tris-HCl (pH 8.0) containing one drop of octanol. The clear extract was applied to a column of Sepharose 4B-Cl ($2.6 \times 100$ cm) equilibrated with 2% (w/v) SDS, 15 mM β-mercaptoethanol, 50 mM Tris-HCl (pH 8.0). Fractions containing the purified fusion protein were identified by electrophoresing aliquots on SDS-polyacrylamide gels and were pooled and stored at −20°C.

*Typing of warts and purification of HPV-1a specific 16/17-K proteins*

Total DNA was extracted from slivers obtained from individual wart biopsies after overnight digestion (37°C) in 100 mM NaCl, 1 mM EDTA, 0.5% SDS, 5 mM Tris-HCl (pH 8.0) containing 500 μg/ml proteinase K, and was typed by agarose gel electrophoresis and Southern blotting as described by Maniatis *et al.* (1982). Cloned HPV-1a, HPV-2 or HPV-4 DNA were labelled by nick translation (Rigby *et al.*, 1977) using biotin-11-dUTP and used as probes. Hybrids were visualised using a commercially available detection system (BRL).

Individual HPV-1a-induced warts (0.1−1 g wet weight) were minced up as finely as possible, and were solubilised in 0.5−2 ml of 7 M urea, 15 mM β-mercaptoethanol, 30 mM Tris-HCl (pH 8.0) using a hand-held homogenizer. Insoluble material was removed by centrifugation, and the supernatant was run on a column ($1.6 \times 100$ cm) of Sepharose 6B-Cl equilibrated with solubilization buffer (described above). Fractions containing the 16/17-K protein were detected by Western blotting (Towbin *et al.*, 1979) using anti-p1p7 (E4) fusion protein antisera and were pooled. The proteins were purified further by ion exchange chromatography on a DEAE 52 column ($1 \text{ cm} \times 5$ cm) equilibrated with the same buffer. The 16/17-K protein was eluted, with a shallow salt gradient (0−0.3 M) and was detected by Western blotting. After dialysis against distilled water (two changes), the purified protein was freeze dried and was stored at −20°C.

*Immunisation of rats*

Hooded Lister or Lou rats were immunised at multiple s.c. sites using 50 μg of purified fusion protein in 0.5 ml of Freund's complete adjuvant. The injections were repeated after 14 and 28 days using the same amount of protein in Freund's incomplete adjuvant — the injection at 28 days being given i.p. Rats were bled from the tail 1 week after the third immunisation and then every 2 weeks subsequently. The response was boosted at 56 days, by i.p. injection of 50 μg protein in a 2% suspension of Alu-Gel-S(Serva Feinbiochemica, FRG) and again at 84 days by i.m. and s.c. injection of the same amount of protein in pertussis vaccine (Wellcome Foundation, England). Antibodies against the 16/17-K protein were prepared as above, except that 20 μg of protein was used per immunisation. Antibody titres were determined by ELISA assays as described by Pfaff *et al.* (1982), except that antigen binding was carried out overnight at room temperature, in 0.1 M $Na_2CO_3/NaHCO_3$ (pH 9.5).

*Identification of specific proteins in wart biopsies*

Individual typed warts were solubilised in 7 M urea, 15 mM β-mercaptoethanol, 50 mM Tris-HCl (pH 8.0) by homogenisation (Polytron) and sonication, and the homogenates were analysed by SDS-polyacrylamide gel electrophoresis (Laemmli, 1970) or by alkaline urea gel electrophoresis (Grand *et al.*, 1979). Western blotting was carried out according to Towbin *et al.* (1979), except that nitrocellulose filters were blocked in phosphate-buffered saline (pH 7.5) containing 0.5% gelatin. Antibody binding was detected using a commercially available peroxidase-linked streptavidin-biotin system (Amersham International).

Indirect immunofluorescence staining was carried out on acetone-fixed 5 μm frozen sections using rat sera as the first antibody followed by FITC-labelled goat anti-rat IgG as the second antibody.

*Proteolytic cleavage and amino acid analysis of the purified 16/17-K protein*

The 16/17-K protein was digested with chymotrypsin for 2 h at 37°C in 50 mM ammonium bicarbonate at a substrate:enzyme ratio of 50:1. After freeze drying, the peptides were dissolved in water and fractionated by reverse-phase h.p.l.c. on a C8, S5 octyl column (Spheri Orb.) eluted with a linear acetonitrile gradient (30–80%, flow-rate 1 ml/min). Individual peaks were collected after detection at 215 nm, and their amino acid compositions determined after hydrolysis at 110°C for 16 h in 6 M HCl.

Amino acid analysis of proteins was carried out on duplicate samples after 24 h and 72 h hydrolysis in 6 M HCl.

*Amino acid sequence determination*

Automated amino acid sequencing was performed with an updated Beckman 890C sequencer using a 0.1 M Quadrol programme, with a single coupling followed by double cleavage. In order to retain the peptide (1 nmol) in the sequencer cup 3 mg of polybrene was first dried as a film. Conversion to amino acid PTH was by incubation with 20% (v/v) trifluoroacetic acid, 0.02% (w/v) dithioerythritol for 8 min at 80°C. PTH-amino acids were identified using a Waters Associates h.p.l.c. system equipped with a Dupont Zorbax column at 50°C. The PTH-amino acids were separated on a gradient formed from 27.5 mM sodium acetate, pH 4.75, 15% acetonitrile (Buffer A) and 33 mM sodium acetate pH 4.75, 55% acetonitrile (Buffer B). Curve 11, from the Waters manual, was used to give a step gradient from 22% to 84% Buffer B. PTH derivatives were dissolved in 0.1% acetic acid 30% acetonitrile prior to injection on the column.

*Protein and DNA data base searches*

To check whether the amino acid sequence determined for the peptide isolation from the 16/17-K protein had previously been observed in some other protein or DNA, a number of computer searches of known sequences were undertaken. Two protein data bases, the extended version of the PIR Protein Sequence Data Base (Pre-release 1, June 1984) and the Doolittle Data Base, were searched for amino acid sequences similar to the peptide sequence. Two DNA data bases, the EMBL Data Base (Release 5) and the Genetic Sequence Data Bank (GENBANK Release 29, 22 February 1985) were searched for nucleotide sequences similar to that coding for the heptamer from the 16/17-K protein. Searches were performed using the PEPSCAN (version of 18 March 1984) and EMBLSCAN (version 16 June 1984) programmes for the protein and DNA data bases, respectively on the Cambridge University IBM 3081 computer.

## Acknowledgements

We wish to thank the following: the staff nurse and doctors at the Sutton Coldfield Cottage Hospital for providing the biopsy samples used in this study. Dr David Pioli, ICI/University of Leicester Joint Laboratories for DNA and protein computer searches. Dr Ian Trayer, Department of Biochemistry University of Birmingham for assistance with h.p.l.c., Drs Phil Byrd and Keith Brown for helpful comments on the manuscript. Valarie Nash and Liz Fletcher for excellent technical assistance and Debbie Williams for her secretarial skills. The authors are grateful to Professor Philip Cohen and the MRC for provision of amino acid sequencing facilities and to the Cancer Research Campaign, U.K. for supporting this research. P.H.G. is a Cancer Research Campaign Fellow.

## References

Barrera-Ora,J.G., Smith,K.O. and Melnick,J.L. (1962) *J. Natl. Cancer Inst.*, **29**, 583–595.

Bernard,H.U., Remaut,E., Hershfield,M.V., Das,H.K., Helinski,D.R., Yanofsky,C. and Franklin,N. (1979) *Gene*, **5**, 59–76.

Birnboim,H.C. and Doly,J. (1979) *Nucleic Acids Res.*, **7**, 1513–1523.

Burnett,T.S., (1983) PhD Thesis, University of Birmingham, UK.

Burnett,T.S. and Gallimore,P.H. (1983) *J. Gen. Virol.*, **64**, 1509–1520.

Croissant,O., Breitburd,F. and Orth,G. (1985) *Clin. Dermatol.*, **13**, in press.

Danos,O., Katinka,M. and Yaniv,M. (1982) *EMBO J.*, **1**, 231–236.

Danos,O., Giri,I., Thierry,F. and Yaniv,M. (1984) *J. Invest. Dermatol.*, **83**, 7s–11s.

Dayhoff,M.O., Hunt,L.T. and Hurst-Calderone,S. (1978) In Dayhoff,M.O. (ed.), *Atlas of Protein Sequences and Structure*. The National Biomedical Research Foundation, MD, pp. 361–365.

Favre,M., Breitburd,F., Croissant,O. and Orth,G. (1975) *J. Virol.*, **15**, 1239–1247.

Gassenmaier,A., Lammel,M. and Pfister,H. (1984) *J. Virol.*, **52**, 1019–1023.

Gissman,L., Pfister,H. and zur Hausen,H. (1977) *Virology*, **76**, 569–580.

Grand,R.J.A., Perry,S.V. and Weeks,R.A. (1979) *Biochemistry*, **177**, 521–529.

Greenspan,D., Krystal,M., Nadaka,S., Arnheiter,M., Lyles,D.S. and Palese,P. (1985) *J. Virol.*, **54**, 833–843.

Hanahan,D. (1983) *J. Mol. Biol.*, **166**, 557–580.

Kremsdorf,D., Favre,M., Jablonska,S., Obalek,S., Rueda,L.A., Lutzner,M.A., Blanchet-Bardon,C., van voorst Vader,P.C. and Orth,G. (1984) *J. Virol.*, **52**, 1013–1018.

Laemmli,U.K. (1970) *Nature*, **227**, 680–685.

Lowy,D.R., Dvoretzky,I., Shober,R., Law,M.-F., Engel,L. and Howley,P.M. (1980) *Nature*, **287**, 72–74.

Lusky,M. and Botchan,M.R. (1985) *J. Virol.*, **53**, 955–965.

Maniatis,T., Frisch,E.F. and Sambrook,J. (1982) *Molecular Cloning. A Laboratory Manual*, Cold Spring Harbor Laboratory Press, NY.

Meinke,W. and Meinke,G.C. (1981) *J. Gen. Virol.*, **52**, 15–24.

Messing,J., Crea,R. and Seeburg,P.H. (1981) *Nucleic Acids Res.*, **9**, 309–321.

Pettersson,U., Ahola,H., Stenlund,A. and Moreno-Lopez,J. (1985) In Saltzman,N. and Howley,P.M. (eds), *Papillomaviruses*. Academic Press, in press.

Pfaff,E., Mussgay,M., Bohm,H.O., Schulze,G.E. and Schuller,H. (1982) *EMBO J.*, **1**, 869–874.

Pfister,H. (1984) *Rev. Physiol. Biochem. Pharmacol.*, **99**, 111–181.

Pfister,H. and zur Hausen,H. (1977) *Virology*, **83**, 131–137.

Pfister,H. and zur Hausen,H. (1978) *Med. Microbiol. Immunol.*, **166**, 13–19.

Pfister,H., Gissmann,L. and zur Hausen,H. (1977) *Virology*, **83**, 131–137.

Pfister,H., Huchthausen,B., Gross,G. and zur Hausen,H. (1979) *J. Natl. Cancer Inst.*, **62**, 1423–1425.

Pilacinski,W.P., Glassman,D.L., Krzyek,R.A., Sadowski,P.L. and Robbins,A.K. (1984) *Biotechnology*, **1**, 356–360.

Reed,S.I. (1982) *Gene*, **20**, 255–265.

Rigby,P.W.J., Dieckmann,M., Rhodes,C. and Berg,P. (1977) *J. Mol. Biol.*, **113**, 237–251.

Rüther,U. and Muller-Hill,B. (1983) *EMBO J.*, **2**, 1791–1794.

Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.

Sarver,N., Rabson,M.S., Yang,Y.-C., Byrne,J.C. and Howley,P.M. (1984) *J. Virol.*, **52**, 377–388.

Schiller,J.T., Vass,W.C. and Lowy,D.R. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 7880–7884.

Seedorf,K., Krammer,G., Durst,M., Sahai,S. and Rowekamp,W.G. (1985) *Virology*, **145**, 181–185.

Stanley,K.K. and Luzio,J.P. (1984) *EMBO J.*, **3**, 1429–1434.

Towbin,H., Staehelin,T. and Gordon,J. (1979) *Proc. Natl. Acad. Sci. USA*, **76**, 4350–4354.

Winter,G. and Fields,S. (1980) *Nucleic Acids Res.*, **8**, 1965–1974.

Yang,Y.-C., Okayama,H. and Howley,P.M. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 1030–1034.

*Received on 28 October 1985; revised 20 December 1985*