Journal of General Virology (1999), **80**, 595–600. Printed in Great Britain SHORT COMMUNICATION

Human papillomavirus type 16 variant lineages characterized by nucleotide sequence analysis of the E5 coding segment and the E2 hinge region

Annika Eriksson,[1]† John R. Herron,[1,2] Takashi Yamada[1]‡ and Cosette M. Wheeler[1,2]

[1] University of New Mexico, School of Medicine, Health Sciences Center, Department of Molecular Genetics and Microbiology, 915 Camino de Salud NE, Albuquerque, NM 87131-5276, USA

[2] Epidemiology and Cancer Control Program, Albuquerque, NM 87131, USA

We have previously examined 29 cervical cell isolates for human papillomavirus type 16 (HPV-16) sequence variations in the E6, L2 and L1 coding regions, and the long control region (LCR). Twenty-five of these isolates as well as 23 additional isolates are characterized here as we present the complete E5 coding segment and the E2 hinge region. Eight amino acid variations were observed in the E5 coding segment, 13 were identified in the E2 hinge region and 5 were observed in the overlapping E4 coding segment. These amino acid variations may be relevant to differences in biological functions and may result in altered humoral or cell-mediated immune responses to HPV-16 variants. The characterization of sequence variation within high-risk HPV types might be important in the search for epidemiological correlates of cervical cancer risk. This work complements and extends HPV-16 genome sequence information from specific isolates previously reported by our group.

To date, more than 70 different types of human papillomavirus (HPV) genome have been identified (de Villiers, 1994). Certain HPV types have been classified as high-risk due to their association with anogenital cancers, mainly cervical cancer. DNA from these HPVs, predominantly HPV-16 and -18, is found in approximately 93% of invasive cervical cancer cases worldwide (Bosch *et al.*, 1995). The characterization of sequence variation within high-risk HPV types represents a rational approach to identifying naturally occurring variants which may exhibit altered biological functions. The E2, E4 and E5 PV proteins are important to several virus functions including transcription and replication, interaction with the cytoskeletal network, and immortalization (Doorbar *et al.*, 1991; Ham *et al.*, 1991; McBride *et al.*, 1991; Pim *et al.*, 1992; Roberts *et al.*, 1993, 1997; McBride & Myers, 1996).

HPV intratypic variation has been most extensively studied for HPV-16 (Chan *et al.*, 1992; Eschle *et al.*, 1992; Icenogle *et al.*, 1992; Bavin *et al.*, 1993; Ho *et al.*, 1993; Pushko *et al.*, 1994; Smits *et al.*, 1994; Yamada *et al.*, 1995, 1997; Ku *et al.*, 1997; Terry *et al.*, 1997; Tornesello *et al.*, 1997). Within HPV-16, five major phylogenetic branches have been distinguished, each predominating within specific geographical regions (Ho *et al.*, 1991, 1993; Yamada *et al.*, 1995, 1997). These main HPV-16 branches were designated E (European), As (Asian), AA (Asian-American), Af1 (African-1) and Af2 (African-2). Additional minor phylogenetic branches have been identified and designated NA1 (North American 1), E-G131 and AA-G183/AA-c (Yamada *et al.*, 1995, 1997).

In the study of Yamada *et al.* (1995), HPV-16 nucleotide sequence variations in E6 (nt 104–559), parts of the L2 (nt 4272–5657) and L1 (nt 5665–7148) ORFs, and the LCR (nt 7479–7842) were established in 29 selected United States isolates. In order to extend this study, we examined the sequence variation within the complete E5 coding segment (nt 3850–4101) and the E2 hinge region (nt 3338–3571) in 25 of these isolates and 23 additional isolates (Yamada *et al.*, 1997).

Isolates labelled as OR. were HPV-16 DNA-containing cervicovaginal lavage samples from subjects enrolled in an epidemiological investigation conducted in Portland, Oregon (Schiffman *et al.*, 1993). Isolates labelled as IS. were crude DNA preparations from HPV-16 DNA-containing cervical cancer specimens obtained from the International Biological Study of Cervical Cancer (IBSCC) (Bosch *et al.*, 1995). Clinical samples

Author for correspondence: Cosette M. Wheeler (mail to address 1). Fax +1 505 277 5273. e-mail cwheeler@salud.unm.edu

†**Present address:** Karolinska Institute, Department of Cell and Molecular Biology, Medical Nobel Institute, 171 77 Stockholm, Sweden.

‡**Present address:** 4-13 Nishi-17-jo-kita-3-chome, Obihiro City, Hokkaido 080, Japan.

The GenBank accession numbers of the sequences reported in this paper are AF120674–AF120693 for the HPV-16 E2 sequences and AF120694–AF120713 for the HPV-16 E5 sequences.

HPV16 Nt Seq	E6	E5	E2	
		33333333344444	333333333333333	
	11111111111222222223345	88889999900000	333344444444555555	
	03333446778556688893503	56666799913478	678811234679112336	
	91257351683856795695503 2	80787901570279	274706519397674386	Class-
Ref.	TAGATCGCGTTGGCGGCTATCTAA	TTTGAAACTGAAAT	ACTTCGAGGGAGCTTGAT	Subclass
OR.5110	------------------------	--A----------	------------------	E-p
OR.4724	------------------------	-----C-----G--	--C-T-------------	E-p
OR.4997	--------------AA-----G--	-----G-----G--	----TA------------	E-p
OR.6311	-------------T-------G--	-----C--C--G--	----T-------------	E-p
IS. 645	---------------------G--	-----C-----G--	----T-------------	Em
OR.6170	C-----------------------	-G----------	------------------	E-C109T
OR.8329	C--------------------G--	-----C-----G-C	----T-------------	E-C109G
OR.8987	C--------------------G--	-----C-----G--	----T-------------	E-C109G
OR.9237	-G-------------------G--	-----C-----G--	--C-T-------------	E-G131G
OR.0198	-G-------------------G--	-----C-----G--	--C-T-------------	E-G131G
IS. 244	--------A---------------	----------G--	-******-----------	E-A176T
IS. 463	--------A-----------A--	-----C-----G--	----T-------------	E-A176A
IS.1032	--------A---------------	-----C-----G--	----T-------------	E-A178T
IS. 364	-----------C------------	----------	------------------	E-C188T
IS. 670	-----------C------------	----------	----T-------------	E-C188T
IS. 7	-----------C---------G--	-----C-----G--	----T-------------	E-C188G
IS. 30	-----------C---------G--	-----C-----G--	-G--T-------------	E-C188G
OR.2087	---------G--------------	-----C-----GT-	--C-T---A-----C---	As
OR.5428	---------G--------------	-----C-----GT-	--C-T---A--A-CCA--	As
OR.7574	---------G--------------	-----C-----GT-	--C-T---A-----CA--	As
OR.7587	--C--GT----------AG-T---	C--A-C-T---T-C	GG--T-C-A-C-A----G	Af1-a
OR.1905	--C--GT----------AG-T---	C--A-C-T---G-C	GG--T-C-A---A----G	Af1-a
OR.7632	--C--GT----------AG-T---	C--A-C-T---T-C	GG--T-C-A---A----G	Af1-a
OR.6106	--C--GT----------AG-T---	ND	GG--T-C-A---A----G	Af1-a
IS. 393	--C--GT----------AG-T--G	C--A-C-T---T-C	-G--T-T-A---A----G	Af1-b
IS. 398	-G---GT----------AGGTG--	C--A-C-T---T-C	-G--T---A---A----G	Af1-d
IS. 838	-G---GT----------AGGTG--	C--A-C-T---T-C	-G--T---A---A----G	Af1-d
IS. 347	-----GTG----T----AGGTG--	C--C-C-G---G-G	-G--T---A---A----G	Af1-e
OR.3473	C-T--GT----------AG-T-G-	ND	GG--T--AA---AC--CG	Af2-a
OR.3759	C-T--GT----------AG-T-G-	C----C-T---G-C	GG--T--AA---AC---G	Af2-a
OR.7145	C-T--GT----------AG-T-G-	C----C-G---G-C	GG--T--AA---AC--CG	Af2-a
IS. 812	C-T--GT----------AG-T-G-	C----C-T---G-C	-G--T---A---AC--CG	Af2-a
IS. 808	-----GT----------AG-T---	C----C-T--CG-C	-G--T--AA---AC--CG	Af2-b
IS. 825	-----GT----------AG-T---	C----C-T-----C	-G--T--AA---AC--CG	Af2-b
IS. 811	----AGT---------GAG-T---	C----C-T---C-C	-G--T--AA---AC--CG	Af2-c
IS. 815	----AGT---------GAG-T---	C----C-T---C-C	-G--T--AA---AC--CG	Af2-c
OR.3136	------T----------AG-TG--	C----C-T-A-G-C	ND	NA1
OR.4541	------T----------AG-TG-G	C----C-G-A-G-C	-G-CT---AC--AC--CG	AA-a
OR.8160	------T----------AG-TG-G	C----C-G-A-G-C	-G-CT---A---ACCACG	AA-a
OR.7908	------T----------AG-TG-G	C----C-G-A-G-C	-G-CT---A---AC--CG	AA-a
OR.4451	------T----------AG-TG-G	C----C-G-A-G-C	-G-CT---A---AC--CG	AA-a
OR.8863	------T----------AG-TG-G	ND	-G-CT---A---AC--CG	AA-a
OR.7754	------T----------AG-TG-G	C----C-G-A-G-C	-G-CT---A---AC--CG	AA-a
OR.7875	------T----------AG-TG-G	C----C-G-A-G-C	ND	AA-a
IS. 21	------T----------AG--G-G	C----CGG-A-G--	-G--T---A---AC--C*	AA-b
OR.5691	------T---G------AG-TG-G	C---GC-G-A-G-C	-G--TA--A---AC--CG	AA-c
OR.8392	------T---G------AG-TG-G	C---GC-G-A-G-C	-G--TA--A---AC--CG	AA-c
OR.1783	------T---G------AG-TG-G	C---GC-G-A-G-C	-G--TA--A---AC--CG	AA-c

Fig. 1. All E2 and E5 nucleotide sequence variations among the HPV-16-containing clinical specimens. The identification codes of the samples indicated along the left correspond to specimens obtained from Portland, Oregon (Schiffman *et al*., 1993) and the IBSCC (Bosch *et al*., 1995). Phylogenetic groupings based on sequences are indicated in the far right-hand column. Nucleotide sequences for E6 have previously been reported for these specimens. The HPV-16 revised sequence (HPV16R) is indicated as reference (Myers *et al*., 1995). The nucleotide positions at which variations were observed are written vertically across the top. For each variant, positions that do not vary relative to the HPV reference sequence are marked with dashes. The specimens from which no sequence data were obtained are indicated as not determined (ND) and an asterisk (*) indicates that nucleotide sequence information was not distinguished at the designated nucleotide position.

HPV16 AA Seq		E2		
	E5	2222222222222	E4	
	00444466	0011122234567	34567	
	47048915	3801914628401	61410	Class-
Ref.	LATILLII	NPIIPATAEGTDF	RQLQH	Subclass
OR.5110	--------	-------------	-----	E-p
OR.4724	---L---V	--T-S--------	-----	E-p
OR.4997	---V---V	----ST-------	-----	E-p
OR.6311	---L-S-V	----S--------	-----	E-p
IS. 645	---V---V	----S--------	-----	Em
OR.6170	R-------	-------------	-----	E-C109T
OR.8329	---L----	----S--------	-----	E-C109G
OR.8987	---L---V	----S--------	-----	E-C109G
OR.9237	---L---V	--T-S--------	-----	E-G131G
OR.0198	---L---V	--T-S--------	-----	E-G131G
IS. 244	-------V	-###########	%%%%%	E-A176T
IS. 463	---L---V	----S--------	-----	E-A176A
IS.1032	---L---V	----S--------	-----	E-A178T
IS. 364	--------	-------------	-----	E-C188T
IS. 670	--------	----S--------	-----	E-C188T
IS. 7	---L---V	----S--------	-----	E-C188G
IS. 30	---L---V	-A--S--------	-----	E-C188G
OR.2087	---L---V	--T-S---K----	-----	As
OR.5428	---L---V	--T-S---KR-N-	--P--	As
OR.7574	---L---V	--T-S---K--N-	-----	As
OR.7587	-T-L---L	DA--S-P-K-N-V	-HI-Q	Af1-a
OR.1905	-T-L---V	DA--S-P-K-N-V	--I-Q	Af1-a
OR.7632	-T-L---L	DA--S-P-K-N-V	--I-Q	Af1-a
OR.6106	ND	DA--S-P-K-N-V	--I-Q	Af1-a
IS. 393	-T-L---L	-A--S---K-N-V	--I-Q	Af1-b
IS. 398	-T-L---L	-A--S---K-N-V	--I-Q	Af1-d
IS. 838	-T-L---L	-A--S---K-N-V	--I-Q	Af1-d
IS. 347	-T-LV--V	-A--S---K-N-V	--I-Q	Af1-e
OR.3473	ND	DA--S--TK-N-V	--TPQ	Af2-a
OR.3759	---L---V	DA--S--TK-N-V	--T-Q	Af2-a
OR.7145	---LV--V	DA--S--TK-N-V	--TPQ	Af2-a
IS. 812	---L---V	-A--S---K-N-V	--TPQ	Af2-a
IS. 808	---L--LV	-A--S--TK-N-V	--TPQ	Af2-b
IS. 825	---L----	-A--S--TK-N-V	--TPQ	Af2-b
IS. 811	---L---L	-A--S--TK-N-V	--TPQ	Af2-c
IS. 815	---L---L	-A--S--TK-N-V	--TPQ	Af2-c
OR.3136	---L---V	ND	ND	NA1
OR.4541	---LV--V	-A-TS---K-N-V	P-TPQ	AA-a
OR.8160	---LV--V	-A-TS---K-NNV	--TPQ	AA-a
OR.7908	---LV--V	-A-TS---K-N-V	--TPQ	AA-a
OR.4451	---LV--V	-A-TS---K-N-V	--TPQ	AA-a
OR.8863	ND	-A-TS---K-N-V	--TPQ	AA-a
OR.7754	---LV--V	-A-TS---K-N-V	--TPQ	AA-a
OR.7875	---LV--V	ND	ND	AA-a
IS. 21	---LV--V	-A--S---K-N-*	--TP*	AA-b
OR.5691	--ALV--V	-A--ST--K-N-V	--TPQ	AA-c
OR.8392	--ALV--V	-A--ST--K-N-V	--TPQ	AA-c
OR.1783	--ALV--V	-A--ST--K-N-V	--TPQ	AA-c

Fig. 2. All E2, E4 and E5 amino acid sequence variations among the HPV-16-containing clinical specimens. As in Fig. 1, the identification codes of the samples are indicated along the left and phylogenetic groupings are indicated in the far right-hand column. The E2, E4 and E5 amino acid sequences based on HPV16R (Myers *et al*., 1995) are indicated as E2, E4 and E5 aa reference, respectively. The first methionine in the E2 and E5 coding regions is numbered as E2 aa 1 and E5 aa 1, respectively. The first amino acid in the E4 coding region corresponds to the first amino acid beginning at the E1 ^ E4 splice acceptor site at HPV-16 nt 3357. The amino acid positions at which variations were observed are written vertically across the top. For each variant, positions that do not vary relative to the HPV-16 reference sequence are marked with dashes. The specimens from which no data were obtained are indicated as not determined (ND) and an asterisk (*) indicates that the amino acid present at the designated position was not distinguished. In IS.244, # indicates positions in E2 which now consist of E4 residues; @ indicates that this residue in the E2 ^ E4 fusion protein corresponds with the E4 stop codon; % indicates that all E4 variant amino acid positions, as observed in all other isolates examined in this study, occur at or after a stop codon and would not be translated.

containing HPV-16 DNA were selected for this study to maximize the potential spectrum of E2, E4 and E5 sequence diversity. All discrete HPV-16 variant genomes previously identified in our studies of over 600 HPV-16-containing clinical specimens were represented.

A hemi-nested PCR system was used to amplify a 514 bp target spanning the HPV-16 E5 coding region. In the outer reaction primers HPV16F-E5 3745 (5′ TGCAATTGTTACA-CTTACATATG) and HPV16R-E5 4318 (5′ ATGTACCTG-CCTGTTTGCATG) were used. In the inner reaction primers HPV16F-E5 3745 and HPV16R-E5 4258 (5′ TTGCAGAAC-GTTTGTGTCGCA) were used. For the 314 bp PCR product spanning the HPV-16 E2 hinge region, we used a nested PCR system. The outer reaction primers were HPV16F-E2 3300 (5′ AAGTATGGGAAGTTCATGCGG) and HPV16R-E2 3697 (5′ TGCAGTATACAATGTACAATGCT). The inner reaction primers were HPV 16F-E2 3317 (5′ GCGGGTGGTCAGG-TAATATTA) and HPV16R-E2 3654 (5′ CATTTTAAAGT-ATTAGCATCACCT). HPV16R-E5 4258 and HPV16F-E2 3317 were biotinylated to facilitate subsequent purification of the single-stranded DNA templates.

The outer reactions consisted of 50 µl amplifications containing 10 mM Tris pH 8·3, 50 mM KCl, 200 µM of each dNTP (dATP, dCTP, dGTP and dTTP), 2·5 mM $MgCl_2$, 10 pmol each of the forward and reverse primers and 2·5 units of *Taq* DNA polymerase (Perkin-Elmer). A 2 µl aliquot of the crude DNA preparation was used for the first reaction. PCR amplification was conducted for 35 cycles with denaturation at 94 °C for 30 s, annealing at 55 °C for 30 s and extension at 72 °C for 1 min. Amplification cycles were preceded by a 5 min denaturation at 94 °C and followed by a 10 min final extension at 72 °C. Two µl of the outer amplification reactions were used as the template for the inner amplification reactions. The PCR conditions were identical in both the inner and outer reactions, except that the inner reaction volumes were 100 µl. Two independent nested amplification reactions were conducted for each clinical specimen. Duplicate amplification reactions were subjected to direct DNA sequence analysis.

Single-stranded DNA templates were prepared by binding biotinylated PCR DNA strands to streptavidin-coated magnetic beads (Dynabeads M-280 Streptavidin; Dynal) according to the manufacturer. Nucleotide sequences were determined using ^{35}S and Sequenase kit version 2.0 (United States Biochemical). HPV16F-E5 3771 (5′ GTGAATGGCAACGT-GACCA) was used for the E5 sequencing primer and HPV16R-E2 3624 (5′ TGTACTATGGGTGTAGTGTTAC) was used

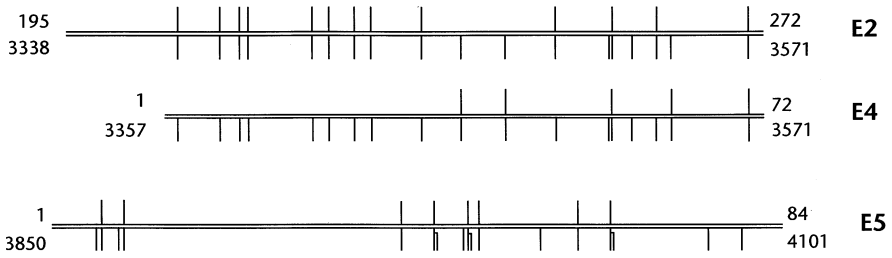

Fig. 3. Distribution of nucleotide and amino acid changes among HPV-16 variants within the E2, E4 and E5 coding sequences. The horizontal lines represent the sequenced portion of each region. The beginning and ending nucleotide positions are indicated below the lines to the left and right, respectively; the beginning and ending residues of the predicted amino acid sequences are indicated above the lines. Vertical bars below the lines represent the positions of nucleotide substitutions; vertical bars above the lines represent predicted amino acid changes.

for the E2 hinge region sequencing primer. The sequence reactions were performed according to the manufacturer. The majority sequence determined by direct DNA sequence analysis was reproduced for each set of duplicates analysed.

The nucleotide sequence data are summarized in Fig. 1 and the corresponding amino acid data are summarized in Figs 2 and 3. The E4 coding region overlaps the E2 hinge region and the amino acid variations within the corresponding E4 segment are also shown in Figs 2 and 3. Nucleotide markers were identified for the major HPV-16 phylogenetic branches E, AA, Af1 and Af2 as well as within the minor branches As and AA-G183, also designated AA-c [Fig. 1 (Ho *et al*., 1991; Chan *et al*., 1992; Yamada *et al*., 1995, 1997)] for both regions targeted.

Although samples for this study were selected based on anticipated nucleotide diversity, it is worth noting that nucleotide sequences were determined for the E2 hinge region in all 18 invasive cervical cancer specimens included. Only one of these specimens, IS.244, contained a deletion spanning nt 3374–3428. Similarly, Terry *et al*. (1997) reported amplifying the entire E2 gene in two overlapping segments for 12 of 14 HPV-16-positive invasive cervical cancers. These results are somewhat unexpected given existing data on the physical status of HPV-16 genomes within invasive cervical cancers (Cullen *et al*., 1991). Of 40 cervical cancers containing HPV-16 DNA, 29 (72%) were reported to contain integrated viral genomes (Cullen *et al*., 1991), only 8 of which contained both episomal and integrated forms. Furthermore, Vernon *et al*. (1997) recently provided detailed maps of the E1/E2 region for several integrated HPV-16 genomes Of the 19 cervical cancers examined, 14 (74%) contained potential integration sites within the region amplified in our studies. Although our data set contains relatively few invasive cancers, we successfully obtained E2 hinge region sequence in every case. In addition, the nucleotide sequences were consistent with phylogenetic designations derived from other regions of the genome (Yamada *et al*., 1995, 1997) and argue against any potential laboratory artifact. If strictly integrated genomes represented the majority of all invasive cervical cancers, our results seem unlikely. We would suggest that further studies on the physical status of HPV genomes are warranted in all levels of high-grade cervical disease and cancer.

In the HPV-16 E5 coding segment, the nucleotide variations were concentrated within the extreme 5′ region and the 3′ half of the gene. Fourteen nucleotide changes were identified in this region and resulted in 6 synonymous and 8 nonsynonymous changes (Figs 1, 2 and 3). In an earlier report, nucleotide variations were identified for 23 HPV-16 specimens in the E5 coding region and in the LCR (Chan *et al*., 1992). Variations were observed at 21 nucleotide positions within the E5 coding region, of which nine were similarly observed for the specimens analysed here. Of the 12 additional variations observed by Chan *et al*. 11 were each restricted to single, or in one case, two isolates. However, one change at nt 4059 was shared with four isolates that were identified as Af2 according to the nucleotide variations in the LCR; these isolates may belong to an additional HPV-16 African sublineage. The remaining nucleotide variations within the E5 coding segments in these specimens corresponded with the changes observed for the Af2 specimens analysed here. In our current study, five additional nucleotide variations were observed that had not been previously reported, of which four were single point mutations in three different specimens. The remaining nucleotide change (nt 3967) was shared with all specimens that belonged to the AA-G183/AA-c subclass, and thus separated these specimens within the AA lineage.

The observed E5 substitutions were generally conserved. No substitutions were noted between amino acids 11–24, a stretch of hydrophobic residues potentially representing a transmembrane helical region. This region is the most conserved segment of E5 between HPV types of the A9 HPV group (Myers *et al*., 1995; Halpern & McCance, 1996). Conservative amino acid substitutions were seen in other hydrophobic regions of the protein including a stretch spanning amino acids 46–50, a region that is relatively well-conserved between HPV types. Overall, substitutions within E5 were consistent with there being a significant selectional pressure despite the fact that E5 is generally poorly conserved between HPV types.

In the 249 bp segment (nt 3338–3571) spanning the E2 hinge region and the overlapping E4 coding region (from the E1^E4 splice acceptor site at nt 3357), 18 nucleotide variations were observed (Figs 1 and 3). In the E2 reading frame, 13 nucleotide variations resulted in nonsynonymous amino acid changes and 5 in synonymous changes (Fig. 2), the latter all situated in the last half of the 249 bp segment. In contrast, within the E4 coding region, the 18 nucleotide variations resulted in 13 synonymous changes and 5 nonsynonymous changes (Fig. 2). Here, all of the 5 nonsynonymous changes were found in the second half of the E4 domain. Considering that the E2 hinge region and the E4 gene are extremely varied in length and amino acid composition between different PVs (Lefkowitz & Broker, 1995), the number of nucleotide variations resulting in nonsynonymous E2 amino acid changes might be expected. In the E4 gene, however, 5 of the 18 nucleotide changes resulted in nonsynonymous amino acid changes, suggesting that the E4 gene may be subject to negative selection pressure that does not similarly affect the E2 hinge region. It has been suggested that the main function of the E2 hinge region is to provide for a flexible link between the transactivating and DNA binding domains (Gauthier *et al.*, 1991). The E2 hinge region might therefore allow for less constraint on amino acid conservation than the E4 coding segment.

To date, the only known function for the cytoplasmic E1^E4 fusion protein is its interaction with cytokeratins (Doorbar *et al.*, 1991; Roberts *et al.*, 1993, 1997). Two conserved motifs, MADXXA (coded from the E1 region) and LLXLL, located in the amino terminus of PV E1^E4 fusion proteins, are important for this interaction (Rogel-Gaillard *et al.*, 1992; Roberts *et al.*, 1994). The function of the amino-terminal portion of the E1^E4 protein in mediating the interaction with cytokeratins might provide an explanation for the absence of amino acid variations in this region. Recently, it was noted that the carboxy terminus of the HPV-16 E1^E4 protein is dispensable for keratin cytoskelton association but is involved in inducing disruption of the keratin filaments (Roberts *et al.*, 1997).

HPV protein sequence variations may affect virus carcinogenic potential. As for the HPV-16 E5 protein, amino acid changes might alter the transforming activity of the protein by affecting the interactions with the EGFR, the 16 kDa subunit of the H^+-ATPase or, potentially, other cellular proteins (Goldstein *et al.*, 1991; Conrad *et al.*, 1993; Straight *et al.*, 1995). Also, variations in the E2 protein might affect the transforming potential of HPV-16 due to altered affinity for cellular transcription factors or for viral DNA. To date, different specific HPV molecular variants have been suggested to be associated with risk of cervical neoplasia (Ellis *et al.*, 1995; Londesborough *et al.*, 1996; Xi *et al.*, 1997). One sequence variant distinguished in the HPV-18 E2 hinge region was reported to be associated with decreased risk of cervical neoplasia (Hecht *et al.*, 1995). Further studies are needed to examine HPV-16 variant associations with cervical disease risk as well as to characterize functional differences.

Dr Eriksson was supported by the Swedish Institute and a University of New Mexico Cancer Center postdoctoral award. This work was funded by the National Institutes of Health (RO1 A1 32917) and utilized the University of New Mexico, School of Medicine, Molecular Analysis Shared Facility. We thank Allan Hildesheim, Mark Schiffman and the Kaiser Permanente study personnel, Nubia Muñoz, Xavier Bosch, Julian Peto, Michele Manos and the International Biological Study of Cervical Cancer (IBSCC) study personnel for providing HPV-16 clinical specimens, Lee Fernando for assistance in manuscript preparation, Aaron Halpern for critical review and contributions.

References

Bavin, P. J., Walker, P. G. & Emery, V. C. (1993). Sequence microheterogeneity in the long control region of clinical isolates of human papillomavirus type 16. *Journal of Medical Virology* **39**, 267–272.

Bosch, F. X., Manos, M. M., Muñoz, N., Sherman, M., Jansen, A. M., Peto, J., Schiffman, M. H., Moreno, V., Kurman, R., Shah, K. V. & the International Biological Study of Cervical Cancer Study Group (1995). Prevalence of human papillomavirus in cervical cancer: a worldwide perspective. *Journal of the National Cancer Institute* **87**, 796–802.

Chan, S.-Y., Ho, L., Ong, C.-K., Chow, V., Drescher, B., Dürst, M., ter Meulen, J., Villa, L., Luande, J., Magaya, H. N. & Bernard, H.-U. (1992). Molecular variants of human papillomavirus type 16 from four continents suggest ancient pandemic spread of the virus and its coevolution with humankind. *Journal of Virology* **66**, 2057–2066.

Conrad, M., Bubb, V. J. & Schlegel, R. (1993). The human papillomavirus type 16 E5 proteins which associate with the 16-kilodalton pore-forming protein. *Journal of Virology* **67**, 6170–6178.

Cullen, A. P., Reid, R., Campion, M. & Lorincz, A. T. (1991). Analysis of the physical state of different human papillomavirus DNAs in intraepithelial and invasive cervical neoplasm. *Journal of Virology* **65**, 606–612.

de Villiers, E.-M. (1994). Human pathogenic papillomavirus types: an update. *Current Topics in Microbiology and Immunology* **186**, 1–12.

Doorbar, J., Ely, S., Sterling, J., McLean, C. & Crawford, L. (1991). Specific interaction between HPV-16 E1-E4 and cytokeratins results in collapse of the epithelial cell intermediate filament network. *Nature* **352**, 824–827.

Ellis, J. R. M., Keating, P. J., Baird, J., Hounsell, E. F., Renouf, D. V., Rowe, M., Hopkins, D., Duggan-Keen, M. F., Bartholomew, J. S., Young, L. S. & Stern, P. L. (1995). The association of an HPV16 oncogene variant with HLA-B7 has implications for vaccine design in cervical cancer. *Nature Medicine* **1**, 464–470.

Eschle, D., Dürst, M., ter Meulen, J., Luande, J., Eberhardt, H. C., Pawlita, M. & Gissmann, L. (1992). Geographical dependence of sequence variation in the E7 gene of human papillomavirus type 16. *Journal of General Virology* **73**, 1829–1832.

Gauthier, J.-M., Dillner, J. & Yaniv, M. (1991). Structural analysis of the human papillomavirus type 16-E2 transactivator with antipeptide antibodies reveals a high mobility region linking the transactivation and the DNA-binding domains. *Nucleic Acids Research* **19**, 7073–7079.

Goldstein, D. J., Finbow, M. E., Andresson, T., McLean, P., Smith, K., Bubb, V. & Schlegel, R. (1991). Bovine papillomavirus E5 oncoprotein binds in the 16 k component of vacuolar H(+)-ATPase. *Nature* **352**, 347–349.

Halpern, A. L. & McCance, D. J. (1996). Papillomavirus E5 proteins. In *Human Papillomaviruses. A Compilation and Analysis of Nucleic Acid and Amino Acid Sequences*, pp. III-81–III-111. Edited by G. Myers, C. Baker, C.

Wheeler, A. Halpern & J. Doorbar. Publication LA-UR 96-3007. Los Alamos, NM: Los Alamos National Laboratory.

Ham, J., Dostatni, N., Gauthier, J.-M. & Yaniv, M. (1991). The papillomavirus E2 protein: a factor with many talents. *Trends in Biochemical Science* **16**, 440–444.

Hecht, J. L., Kadish, A. S., Jiang, G. & Burk, R. D. (1995). Genetic characterization of the human papillomavirus (HPV) 18 E2 gene in clinical specimens suggests the presence of a subtype with decreased oncogenic potential. *International Journal of Cancer* **60**, 369–376.

Ho, L., Chan, S.-Y., Chow, V., Chong, T., Tay, S.-K., Villa, L. L. & Bernard, H.-U. (1991). Sequence variants of human papillomavirus type 16 in clinical samples permit verification and extension of epidemiological studies and construction of a phylogenetic tree. *Journal of Clinical Microbiology* **29**, 1765–1772.

Ho, L., Chan, S.-Y., Burk, R. D., Das, B. C., Fujinaga, K., Icenogle, J. P., Kahn, T., Kiviat, N., Lancaster, W., Mavromara-Nazos, P., Labropoulou, V., Mitrani-Rosenbaum, S., Norrild, M., Pillai, M. R., Stoerker, J., Syrjanen, K., Syrjanen, S., Tay, S.-K., Villa, L. L., Wheeler, C. M., Williamson, A.-L. & Bernard, H.-U. (1993). The genetic drift of human papillomavirus type 16 is a means of reconstructing prehistoric viral spread and the movement of ancient human populations. *Journal of Virology* **67**, 6413–6423.

Icenogle, J. P., Laga, M., Miller, D., Tucker, R. A. & Reeves, W. C. (1992). Genotypes and sequence variants of human papillomavirus DNAs from human immunodeficiency virus type 1-infected women with cervical intraepithelial neoplasia. *Journal of Infectious Disease* **166**, 1210–1216.

Ku, J.-L., Kim, W.-H., Park, H.-S., Kang, S.-B. & Park, J.-G. (1997). Establishment and characterization of 12 uterine cervical-carcinoma cell lines: common sequence variation in the E7 gene of HPV-16-positive cell lines. *International Journal of Cancer* **72**, 313–320.

Lefkowitz, E. J. & Broker, T. R. (1995). *Alignments of the Overlapping E4 and E2-Hinge Open Reading Frames and Protein Sequences for Thirty Human and Animal Papillomavirus Genotypes*. University of Alabama, Birmingham.

Londesborough, P., Ho, L., Terry, G., Cuzick, J., Wheeler, C. & Singer, A. (1996). Human papillomavirus genotype as a predictor of persistence and development of high grade lesions in women with minor cervical abnormalities. *International Journal of Cancer* **69**, 364–368.

McBride, A. & Myers, G. (1996). The E2 proteins. In *Human Papillomaviruses. A Compilation and Analysis of Nucleic Acid and Amino Acid Sequences*, pp. III-15–II-31. Edited by G. Myers, C. Baker, C. Wheeler, A. Halpern & J. Doorbar. Publication LA-UR 96-3007. Los Alamos, NM: Los Alamos National Laboratory.

McBride, A. A., Romanczuk, H. & Howley, P. M. (1991). The papillomavirus E2 regulatory proteins. *Journal of Biological Chemistry* **266**, 18411–18414.

Myers, G., Bernard, H.-U., Delius, H., Baker, C., Icenogle, J., Halpern, A. & Wheeler, C. (1995). *Human Papillomaviruses. A Compilation and Analysis of Nucleic Acid and Amino Acid Sequences*. Publication LA-UR 95-3675. Los Alamos, NM: Los Alamos National Laboratory.

Pim, D., Collins, M. & Banks, L. (1992). Human papillomavirus type 16 E5 gene stimulates the transforming activity of the epidermal growth factor receptor. *Oncogene* **7**, 27–32.

Pushko, P., Sasagawa, T., Cuzick, J. & Crawford, L. (1994). Sequence variation in the capsid protein genes of human papillomavirus type 16. *Journal of General Virology* **75**, 911–916.

Roberts, S., Ashmole, I., Johnson, G. D., Kreider, J. W. & Gallimore, P. H. (1993). Cutaneous and mucosal human papillomavirus E4 proteins form intermediate filament-like structures in epithelial cells. *Virology* **197**, 176–187.

Roberts, S., Ashmole, I., Gibson, L. J., Rookes, S. M., Barton, G. J. & Gallimore, P. H. (1994). Mutational analysis of human papillomavirus E4 proteins: identification of structural features important in the formation of cytoplasmic E4/cytokeratin networks in epithelial cells. *Journal of Virology* **68**, 6432–6445.

Roberts, S., Ashmole, I., Rookes, S. M. & Gallimore, P. H. (1997). Mutational analysis of the human papillomavirus type 16 E1 ^ E4 protein shows that the C terminus is dispensable for keratin cytoskeleton association but is involved in inducing disruption of the keratin filaments. *Journal of Virology* **71**, 3554–3562.

Rogel-Gaillard, C., Breiburd, F. & Orth, G. (1992). Human papillomavirus type 1 E4 proteins differing by their N-terminal ends have distinct cellular localizations when transiently expressed in vitro. *Journal of Virology* **66**, 816–823.

Schiffman, M. H., Bauer, H. M., Hoover, R. N., Glass, A. G., Cadell, D. M., Rush, B. B., Scott, D. R., Sherman, M. E., Kurman, R. J., Wacholer, S., Stanton, C. K. & Manos, M. M. (1993). Epidemiologic evidence showing that human papillomavirus infection causes most cervical intraepithelial neoplasia. *Journal of the National Cancer Institute* **85**, 958–964.

Smits, H. L., Traanberg, K. F., Krul, M. R. L., Prussia, P. R., Kuiken, C. L., Jebbink, M. F., Kleyne, J. A. F. W., van den Berg, R. H., Capone, B., de Bruyn, A. & ter Schegget, J. (1994). Identification of a unique group of human papillomavirus type 16 sequence variants among clinical isolates from Barbados. *Journal of General Virology* **75**, 2457–2462.

Straight, S. W., Herman, B. & McCance, D. J. (1995). The E5 oncoprotein of human papillomavirus type 16 inhibits the acidification of endosomes in human keratinocytes. *Journal of Virology* **69**, 3185–3192.

Terry, G., Ho, L. & Cuzick, J. (1997). Analysis of E2 amino acid variants of human papillomavirus types 16 and 18 and their associations with lesion grade and HLA DR/DQ type. *International Journal of Cancer* **73**, 651–655.

Tornesello, M. L., Buonaguro, F. M., Meglio, A., Buonaguro, L., Beth-Giraldo, E. & Giraldo, G. (1997). Sequence variations and viral genomic state of human papillomavirus type 16 in penile carcinomas from Ugandan patients. *Journal of General Virology* **78**, 2199–2208.

Vernon, S. D., Unger, E. R., Miller, D. L., Lee, D. R. & Reeves, W. C. (1997). Association of human papillomavirus type 16 integration in the E2 gene with poor disease-free survival from cervical cancer. *International Journal of Cancer* **74**, 50–56.

Xi, L. F., Koutsky, L. A., Galloway, D. A., Kuypers, J., Hughes, J. P., Wheeler, C. M., Holmes, K. K. & Kiviat, N. B. (1997). Genomic variation of human papillomavirus type 16 and risk for high grade cervical intraepithelial neoplasia. *Journal of the National Cancer Institute* **89**, 796–802.

Yamada, T., Wheeler, C. M., Halpern, A. L., Stewart, A.-C. M., Hildesheim, A. & Jenison, S. A. (1995). Human papillomavirus type 16 variant lineages in United States populations characterized by nucleotide sequence analysis of the E6, L2, and L1 coding segments. *Journal of Virology* **69**, 7743–7753.

Yamada, T., Manos, M. M., Peto, J., Greer, C. E., Muñoz, N., Bosch, F. X. & Wheeler, C. M. (1997). Human papillomavirus type 16 sequence variation in cervical cancers: a worldwide perspective. *Journal of Virology* **71**, 2463–2472.

Received 8 July 1998; Accepted 3 November 1998