Intratype sequence variation among clinical isolates of the human papillomavirus type 6 L1 ORF: clustering of mutations and identification of a frequent amino acid sequence variant

W. Caparrós-Wanderley,¹ N. Savage,² M. Hill-Perkins,² G. Layton,² J. Weber³ and D. H. Davies¹

¹ Infection and Immunity Research Group, Division of Life Sciences, King’s College London, Campden Hill Road, London W8 7AH, UK
² British Biotech Pharmaceuticals Ltd, Watlington Road, Oxford OX4 5LY, UK
³ St Mary’s Hospital Medical School, Jefferiss Wing, Praed Street, London W2 1NY, UK

Introduction

Human papillomaviruses (HPVs) constitute a group of viruses that cause hyperproliferative lesions in mucosal and cutaneous epithelia. The major capsid protein (L1) of all HPV types tested to date possesses the intrinsic ability to self-assemble into virus-like particles (VLPs) when expressed alone and at high levels in eukaryotic systems (Kirnbauer et al., 1992; Hofmann et al., 1995; Heino et al., 1995). These VLPs are morphologically identical to naturally occurring virions (Hagensee et al., 1994) and, in animal models of papillomavirus infection, they have been shown to induce protective immunity against challenge with infectious virus (Breitburd et al., 1995; Kirnbauer et al., 1996). Therefore, VLPs represent ideal candidates for a vaccine to prevent HPV infection.

The efficacy of VLPs as potential HPV vaccines is based on their ability to display the same conformational epitopes present on naturally occurring virions. However, this potential may be severely limited if different HPV subtypes present distinct amino acid sequences within B-cell epitopes in their L1 proteins. Furthermore, amino acid substitutions within L1 may also result in reduced yields of VLPs as described for HPV-16 L1 (Touze et al., 1998).

DNA sequencing of the L1 ORF represents the best method for the identification of amino acid substitutions in the L1 protein. However, analysis of genomic DNA sequence variation has wider applications than vaccine development. Conservative and non-conservative mutations identified in the genome of an organism are of taxonomic value because the...
DNA sequence is both inherited material suitable for phylogenetic evaluation and a phenotypic feature that can be determined reliably (Chan et al., 1995). On this basis, HPV types, subtypes and variants are currently defined by DNA sequence analysis of the viral genome, in particular of subfragments of the L1, E6 and E7 ORFs (de Villiers, 1989; Chan et al., 1995; Stewart et al., 1996).

Analysis of DNA sequence variation within given HPV types has additional applications. Databases on intratype genomic diversity can (a) aid in the development of efficient diagnostic tools, (b) be applied to epidemiological studies (sequence variations are potential markers for monitoring HPVs in defined populations) and (c) assist in establishing genotype and phenotype relationships (i.e. certain variants may be positively or negatively associated with specific lesions, depending on their different biological and functional properties) (Stewart et al., 1996).

Several groups have examined sequence variability among genomes of HPV-6, an HPV type associated with benign condyloma. However, the majority of these studies have focused on the long control region (LCR) and the E6 and E7 ORFs (Rübben et al., 1992; Roman & Brown, 1995; Heinz et al., 1995; Grassmann et al., 1996) and only limited information is available on the L1 ORF. Hofmann et al. (1995) described several nucleotide differences between the L1 ORFs of HPV-6a and the prototype HPV-6b (Schwarz et al., 1983), but found no amino acid changes in the resulting L1 protein. Icenogle et al. (1991) described several nucleotide changes within a 253 nt fragment of the L1 ORF from different geographical isolates of HPV-6, but reported no amino acid substitutions. Suzuki et al. (1997) reported, in addition to several silent nucleotide substitutions, an Asn-to-Thr substitution in aa 55 of the L1 protein. However, it is not known whether this substitution affects the stability, yield and/or morphology of the resulting VLPs compared to those derived from the prototype HPV-6b L1 sequence. Such considerations are important for the development of VLP-based vaccines.

In the present study we have analysed the DNA and amino acid sequence variability over the entire L1 ORF amongst clinical isolates of HPV-6 obtained in the London area. We have also compared this variability to that reported for HPV-6 L1 isolates from different geographical locations. Our aim was, firstly, to establish the degree and distribution of the sequence variability throughout the L1 ORF and, secondly, to study the effect of major naturally occurring amino acid substitutions on the stability and yield of VLPs.

Methods

- **Wart biopsies.** Wart biopsy tissue was obtained from 37 homosexual and heterosexual patients (24 men, 13 women, 19–40 years old) attending the Middlesex and St Mary’s Hospitals (London, UK) for the treatment of condyloma acuminata.

- **DNA isolation from wart biopsies.** Several shavings from a genital wart biopsy were placed in extraction buffer (10 mM Tris–HCl, pH 8.0; 2 mM EDTA, 10 mg/ml DTT and 0.1% SDS) containing Proteinase K (500 µg/ml) and incubated at 37°C with gentle agitation until fully dissolved. The resulting solution was then extracted with phenol–chloroform–isoamyl alcohol and the DNA precipitated in ethanol and resuspended in Tris–EDTA buffer.

- **Cloning of HPV-6 L1 isolates into pGEM-T.** Wart biopsy DNA was used as template DNA to amplify the HPV-6 L1 ORF by PCR using Taq DNA polymerase (Promega) in 35 cycles of amplification (94°C, 30 s; 58°C, 1 min; 72°C, 2 min). The sense (5′ TTA GTC AAGGTATCAG TG GTCCG GGC CTA GCG ACA GCA CA 3′) and antisense (5′ ACT GTA GGATCCT TAA GAT CTC CTT GTA GTC TTG GGC GGC CTA CG 3′) primer set used in the PCR were based on the L1 sequence of the prototype HPV-6b (Schwarz et al., 1983). Both primers contain BamHI sites (underlined). Initiation and stop codons are indicated in bold.

Following gel purification, the L1 PCR products were cloned into the pGEM-T vector (Promega) according to the manufacturer’s instructions.

- **Sequencing.** Sequencing of the different L1 isolates was primarily carried out in an ABI model 373 automated sequencer (Perkin Elmer) at the Advanced Biotechnology Centre (Charing Cross Hospital, London, UK) using the primers indicated by (a) in Table 1. Additionally, in some cases, sequencing of the L1 isolates was carried out manually using [α-32P]dATP (Amersham), the Sequenase version 2.0 DNA sequencing kit (United States Biochemical) and the primers indicated by (b) in Table 1.

- **Phylogenetic analysis.** Phylogenetic analysis was performed using the entire nucleotide sequence of the HPV-6 L1 ORF obtained from the different clinical isolates as well as the published L1 ORF sequences of HPV-6a, the prototype HPV-6b and four different HPV-6 isolates from Japan. The most parsimonious phylogenetic tree was determined with PAUP [Wisconsin Package version 9.1, Genetics Computer Group (GCC), Wisconsin, USA] using the branch-and-bound search strategy.

- **Production of HPV-6 L1 VLPs in yeast.** The HPV-6 L1 ORFs were excised from the recombinant pGEM-6L1 plasmids by BamHI digestion and subcloned into the pJC78 plasmid (British Biotech, Oxford, UK) at a unique BglII site downstream of a galactose inducible mini PGK/GAL + promoter. Each of the resulting pJC78-6L1 plasmids was used, together with the pUG41S plasmid (British Biotech), to cotransform S. cerevisiae strain MC2 (ura3-1 leu2-3; pah1-1122 brh4-3 prc1-407 gal2) using the lithium acetate method (Becker & Guarente, 1991). The pU41S plasmid encodes the GAL-4 protein. This protein mediates high-level, galactose-inducible expression of foreign genes driven by GAL promoters through highly specific interactions with nucleotide sequences in the promoter sequence (Schneider & Guarente, 1991). Transformed yeast colonies were selected on minimum medium supplemented with tryptophan.

VLPs were isolated from transformed yeast cultures grown for 48 h at 30°C in 500 ml minimum medium supplemented with galactose and tryptophan. Induced cells were washed three times in 50 ml chilled TEN (Tris–EDTA–NaCl buffer), resuspended in 3 ml chilled TEN containing 4 g of acid-washed glass beads 212–300 µm (Sigma) and lysed by vigorously vortexing the cell suspension for 10 min at 30°C with 30 s cooling in ice. After collecting the resulting supernatant, the remaining pellet was resuspended in 2 ml of chilled TEN, vortexed and a further 2 ml of supernatant collected.

The pooled supernatants were clarified by centrifugation (5000 g, 4°C, 15 min) and 2 ml of the protein solution was layered on top of a 15–45% (w/v of TEN) sucrose step gradient, left to linearize overnight at 4°C but with a freshly added 60% sucrose cushion. This gradient was centrifuged at 40000 r.p.m. (285000 g) and 4°C for 90 min in an SW-40Ti rotor (Beckman) and fractions were collected from the top of the
Intratype sequence variability in HPV-6 L1

Table 1. List of primers used for sequencing of the HPV-6 L1 ORF from recombinant pGEM-T plasmids

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>T7 promoter (a, b)</td>
<td>5’ TAA TAC GAC TCA CTA TAG GG 3’</td>
</tr>
<tr>
<td>Sp6 promoter (a, b)</td>
<td>5’ ATT TAG GTG ACA CTA TAG AAT AC 3’</td>
</tr>
<tr>
<td>FS1: nt 5984–6003 (b)</td>
<td>5’ CAA TAC AGG GTA TTT AAG G 3’</td>
</tr>
<tr>
<td>FS2: nt 6190–6205 (a, b)</td>
<td>5’ CCC TGG ACA GGA TAA C 3’</td>
</tr>
<tr>
<td>FS3: nt 6376–6393 (b)</td>
<td>5’ GGT TGA CAC AGG CTT TGG 3’</td>
</tr>
<tr>
<td>FS4: nt 6577–6592 (b)</td>
<td>5’ GGG GGA ACC TGT GCC 3’</td>
</tr>
<tr>
<td>FS5: nt 6783–796 (a, b)</td>
<td>5’ CCA CAC GCA GTA CC 3’</td>
</tr>
<tr>
<td>FS6: nt 7034–7048 (b)</td>
<td>5’ GTG CAG TCA CAG GCC 3’</td>
</tr>
</tbody>
</table>

‘a’ and ‘b’ indicate primers used for, respectively, automated and manual sequencing of the HPV-6 L1 ORF.

gradient. Samples from all fractions were analysed by Western blot (Burnette, 1981) with monoclonal antibody (MAB) Camvir-1 (McLean et al., 1990), 1/300 dilution, and HRP-conjugated anti-mouse serum (DAKO), 1/1000 dilution. Reactive fractions were dialysed overnight against PBS, concentrated to about one-fifth of their initial volume on a layer of PEG 8000, dialysed again and analysed by electron microscopy after negative staining with 2% ammonium molybdate.

Results

Cloning of the HPV-6 L1 isolates

After extraction of total DNA from 37 wart biopsies, 62% of the samples (23 patients) were found to be positive by PCR using the HPV-6b L1 primers. In all positive samples, with the exception of isolate M48364, PCR resulted in the amplification of a 1-5 kbp fragment, which is the reported size of the HPV-6b L1 ORF. In isolate M48364, PCR consistently amplified a 0-9 kbp fragment, even though several PCRs were carried out on this sample, some of them at annealing temperatures of up to 62 °C (data not shown).

After isolation of all 23 HPV-6 L1 PCR fragments, 17 of them were successfully cloned into pGEM-T.

Nucleotide and amino acid sequence variation amongst HPV-6 L1 isolates

Complete sequencing of all 17 HPV-6 L1 ORFs cloned into pGEM-T revealed 28 locations where substitutions from the prototype HPV-6b sequence occurred (summarized in Table 2).

None of the isolates analysed possessed the reported L1 nucleotide sequence of either HPV-6a (Hofmann et al., 1995) or the prototype HPV-6b (Schwarz et al., 1983). Nucleotide variation within the L1 ORF ranged from three substitutions (isolates M51033 and M44371) to 14 substitutions (isolate M49898).

The most dramatic change observed amongst the DNA sequences analysed was the deletion of a 582 bp fragment (nt 6614–7196 in the prototype HPV-6b genome) in isolate M48364. A similar deletion near the C-terminal end of the L1 ORF has been described in an HPV-6a isolate associated with recurrent laryngeal papillomatosis (Suzuki et al., 1995).

Of all the nucleotide substitutions from the prototype HPV-6b L1 sequence, eight were present in 76% (13/17) of isolates. These substitutions were all silent and involved an A-to-T substitution at nt 5923, a C-to-T substitution at nt 6052, an A-to-G substitution at nt 6073, an A-to-T substitution at nt 6217, an A-to-T substitution at nt 6598, a C-to-A substitution at nt 6625, a G-to-A substitution at nt 6661 and an A-to-C substitution at nt 7219. Six of these substitutions (nt 5923, nt 6052, nt 6073, nt 6598, nt 6661 and nt 7219) have been described in HPV-6a (Hofmann et al., 1995). Three of them (nt 6598, nt 6625 and nt 6661) have been described, after sequencing of a subfragment of the L1 ORF, in HPV-6 isolates from India, Philippines and the United States (Icenogle et al., 1991). All eight substitutions have been described previously only in some HPV-6 L1 isolates from Japan (Suzuki et al., 1997).

Some of the eight nucleotide substitutions described above were also observed in an additional 12% (2/17) of isolates (F50602 and M48364). The remaining 12% (2/17) of isolates (M44371 and M51033) did not have any of these eight specific substitutions. Furthermore, with the exception of a G-to-C substitution at nt 6661 present in both M51033 and F50602, none of the nucleotide substitutions observed in isolates M44371 and M51033 were present in any other isolate. Interestingly, and according to the phylogenetic tree shown in Fig. 1, these two DNA sequences are the most closely associated to that of the prototype HPV-6b.

Non-silent nucleotide substitutions were observed in 42% of the locations (12/28) where nucleotide substitutions had been detected (Table 2). Four of the seventeen isolates analysed (F49413, M48163, M33382 and M44371) contained the reported HPV-6b L1 amino acid sequence (Schwarz et al., 1983). Amino acid differences compared to the L1 prototype...
Table 2. Nucleotide and amino acid sequence variation amongst HPV-6 L1 isolates

Nucleotide positions at which substitutions are observed are given in the top row (the L1 ORF extends from nt 5789–7291 within the HPV-6b genome). Each row indicates, from left to right, the identification code of the isolate and the nucleotide alignment compared with the prototype sequence. The prototype HPV-6b and HPV-6a L1 sequences are indicated as ‘HPV6b’ and ‘HPV6a’, respectively. Amino acid substitutions caused by the nucleotide substitutions are given across the bottom row, indicated ‘Muta’. Positions that do not vary relative to the prototype nucleotide or amino acid sequence are marked with a dash in the alignment. Nucleotide and amino acid positions situated in the R1, R2 and R3 regions are indicated within the boxes. ‘X’ indicates the absence of nucleotide sequences at this position.

| 5940 | 5903 | 5923 | 5994 | 6013 | 6044 | 6052 | 6073 | 6094 | 6203 | 6214 | 6235 | 6394 | 6426 | 6478 | 6507 | 6598 | 6625 | 6661 | 6727 | 7027 | 7079 | 7205 | 7219 | 7227 | 7245 | 7254 | 7271 |
|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| HPV6b |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| HPV6a |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| F49413 | -   | -   | T    | A    | -    | T    | G    | -    | A    | -    | -    | -    | -    | -    | -    | -    | T    | A    | -    | -    | -    | -    | C    | -    | -    | -    | -    |
| M48163 | -   | -   | T    | -    | -    | T    | G    | -    | -    | -    | T    | -    | -    | -    | -    | -    | T    | A    | A    | -    | -    | -    | C    | -    | -    | -    | -    |
| M39010 | -   | -   | T    | -    | -    | T    | G    | -    | -    | -    | T    | -    | -    | -    | -    | -    | T    | A    | A    | -    | -    | -    | C    | -    | -    | -    | -    |
| F47718 | -   | -   | T    | -    | -    | T    | G    | -    | -    | -    | T    | -    | -    | -    | -    | -    | T    | A    | A    | -    | -    | -    | C    | -    | -    | -    | -    |
| M47875 | -   | -   | T    | -    | -    | T    | G    | -    | -    | -    | T    | -    | -    | -    | -    | -    | T    | A    | A    | -    | -    | -    | C    | -    | -    | -    | -    |
| M47769 | -   | -   | T    | -    | -    | T    | G    | -    | -    | -    | T    | -    | -    | -    | -    | -    | T    | A    | A    | -    | -    | -    | C    | -    | -    | -    | -    |
| M46769 | -   | -   | T    | -    | -    | T    | G    | -    | -    | -    | T    | -    | -    | -    | -    | -    | T    | A    | A    | -    | -    | -    | C    | -    | -    | -    | -    |
| F50528 | -   | -   | T    | Q    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | T    | A    | A    | -    | -    | -    | C    | -    | -    | -    | A    |
| F50027 | -   | -   | T    | Q    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | T    | A    | A    | -    | -    | -    | C    | -    | -    | -    | C    |
| M33382 | -   | -   | T    | Q    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | T    | A    | A    | -    | -    | -    | C    | -    | -    | -    | C    |
| M49576 | -   | -   | T    | Q    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | T    | A    | A    | -    | -    | -    | C    | -    | -    | -    | C    |
| F50602 | -   | -   | T    | Q    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | T    | A    | C    | -    | -    | -    | C    | -    | -    | -    | T    |
| M49898 | C    | C    | T    | -    | -    | T    | Q    | -    | -    | C    | T    | C    | -    | -    | -    | -    | T    | A    | A    | C    | C    | -    | C    | -    | -    | -    | -    |
| M6    | -    | -   | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | G    | -    | G    | -    | -    | -    | -    | -    | -    | -    | -    |
| M48364 | -   | -   | T    | C    | T    | Q    | -    | -    | -    | T    | -    | -    | -    | -    | -    | -    | T    | X    | X    | X    | X    | -    | C    | -    | -    | -    | -    |
| M44371 | -   | -   | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | C    | -    | C    | -    | -    | -    | G    | -    | -    | -    | -    |
| M51033 | -   | -   | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | G    | -    | G    | -    | -    | -    | -    | -    | -    | -    | -    |

| 6b | S | S | G | V | P | S | F | L | L | G | N | N | V | G | K | Y | D | T | R | P | Q | Y | E | S | T | K | K | A | R |
| Muta | P | P | - | - | P | - | - | D | - | - | - | - | G | - | - | H | - | Q | P | - | R | R | V | S |

R1  R2  R3
Intratypic sequence variability in HPV-6 L1

Intratypic sequence variability in HPV-6 L1

Fig. 1. Phylogenetic tree of HPV-6 L1 isolates developed from the nucleotide sequence data. Included are the L1 nucleotide sequences of the clinical isolates, HPV-6a, the prototype HPV-6b and five different HPV-6 isolates from Japan (C42, C43, C45, KP3 and KP5). The sequence of isolate M48364, which contains a 582 bp deletion, is not included in this tree. Analysis of the sequences was carried out with the PAUP program (Wisconsin Package version 9.1) using the branch-and-bound search strategy. The tree was rooted using the HPV-11 L1 ORF nucleotide sequence.

Fig. 2. Location of the R1, R2 and R3 regions within the HPV-6b L1 ORF. The vertical bars indicate the locations where nucleotide changes are observed. The height of each vertical bar is proportional to the number of isolates presenting nucleotide substitutions from the prototype HPV-6b L1 ORF at the indicated positions.

HPV-6b amongst the remaining isolates ranged from one substitution (isolates F47718, M39010, M47769, M49576 and M6) to four substitutions (isolate M49898).

The most common amino acid change observed amongst the isolates was a Glu-to-Gln substitution at aa 431, generated by an A-to-G substitution at nt 7079. This amino acid change, found in 41% (7/17) of isolates, does not alter significantly the hydrophilicity of the position but implies a shift of its electrical charge from negative to neutral. This particular substitution represented the characteristic and unique feature of 29% (5/17) (isolates F47718, M39010, M47769, M49576 and M6) of all L1 protein sequences analysed.

The second most common amino acid change was a Lys-to-Arg substitution at aa 480, generated by an A-to-G substitution at nt 7227. This amino acid change, found in 12% (2/17) of isolates, does not alter either hydrophilicity or the electrical charge of the position since both amino acids are acid and basic.

The remaining amino acid changes are characteristic features of individual isolates. Most of them do not significantly alter the hydrophobicity and/or the electrical charge of the position. However, the case of the Ser-to-Pro substitutions at aa 18, 39 (both in isolate M49898), 86 (isolate M48364) and 473 (isolate M47875) is different. Since the nitrogen atom of
the amino group of proline is incorporated into a ring, these substitutions might disrupt the usual organization of the L1 protein backbone, causing a sharp transition in the direction of the chain. Consequently, they might have detrimental effects on the properties of the resulting protein.

Based on the results presented here, it is evident that two different L1 amino acid sequences are predominant in the population under study. These L1 sequences are the previously reported HPV-6b L1 prototype sequence, found in 23% of isolates (4), and a novel ‘London’ variant, found in 29% of isolates (5), which is characterized by a single Glu-to-Gln substitution at aa 431.

Another important observation from this study was that sequence substitutions from the prototype HPV-6b L1 ORF among the isolates analysed were more frequently found within three regions: R₁ (nt 5920–nt 6075), R₂ (nt 6590–6670) and R₃ (nt 7070–7230) (Fig. 2). Moreover, all the substitutions described by Icenogle et al. (1991) and 80% (16/20) of those described by Suzuki et al. (1997) in HPV-6 L1 isolates from India, the United States, Philippines and Japan can be located within these three regions.

The R₁ and R₂ regions contain mainly synonymous changes whilst the R₃ region contains predominantly non-synonymous changes, including the Glu-to-Gln substitution at aa 431. The R₃ region ranks highly as a potential antibody-reactive site according to the Jameson–Wolf algorithm (Fig. 3). Furthermore, sera from patients with genital condyloma acuminata have been shown to react with a peptide (aa 417–437) spanning part of this region of the HPV-6 L1 ORF (Jenison et al., 1989). These observations suggest that the R₃ region might constitute a linear B-cell epitope on the surface of the HPV-6 capsid.

Production of HPV-6 L1 VLPs in yeast

The identification of the London variant of HPV-6 L1 as the predominant sequence in the population under study clearly implies that this new variant is infectious. Since infection cannot occur without virus assembly, it is entirely predictable that this novel variant of the HPV-6 L1 sequence would be able to self-assemble into VLPs. However, the sequence and epidemiological data presented so far cannot predict whether the stability and yield of the VLPs obtained following expression of this novel sequence will be significantly different from those obtained following expression of the prototype sequence. In order to analyse the self-assembly potential of this novel sequence, the L1 ORFs from isolates F47718 (London variant) and M48163 (prototype HPV-6b) were cloned into pJC78 and used as templates for the expression of the L1 protein.

Following cotransformation of S. cerevisiae MC2 with pUG41S and the corresponding recombinant pJC78-6L1 plasmid, cultures of two transformed yeast clones from each cotransformation were prepared and heterologous gene expression was induced. Cell lysates were prepared from mock-transformed and induced cultures and their components separated by centrifugation down sucrose gradients. Fractions were collected from all sucrose gradients and screened for L1 expression by Western blot with MAb Camvir-1.
Intratype sequence variability in HPV-6 L1

In all gradients, fractions 3–11 showed reactivity against Camvir-1 in a band at the expected molecular mass of the L1 protein (55 kDa) (typical results are shown in Fig. 4a) although clear peaks of reactivity were evident in fractions 4 and 5. No significant differences can be observed in the intensity of the bands corresponding to L1 and its degradation products between equivalent fractions of the two gradients.

The non-reactive fractions were found to be those collected from the top of the gradient (low-density fractions), where all soluble proteins are expected to accumulate. The lack of reactivity in these top fractions and its accumulation in certain higher density fractions is, therefore, consistent with self-assembly of the expressed L1 monomers into VLPs.

The reactivity present in fraction 11 (the 60% sucrose cushion), where major cellular debris is expected to accumulate, was probably due to monomers of L1 crossing the nuclear cell membrane at the time of lysis. This notion is supported by the observation that HPV VLPs produced in yeast assemble inside the nucleus of the cell (Sasagawa et al., 1995) in the same way that PV virions assemble inside the nucleus of infected human cells.

Electron microscopic analysis of the reactive fractions from both gradients revealed the presence of structures corresponding to VLPs with a diameter of 40–50 nm (typical results are shown in Fig. 4b). No VLPs or VLP-like structures were found in the equivalent fractions of mock-transformed yeast lysates.

Considered together, these results indicate that the Glu-to-Gln amino acid substitution present in aa 431 of the L1 London variant does not affect the stability, yield and/or morphology of VLPs.

Discussion

In the present study we have analysed the DNA and amino acid sequence variability over the entire L1 ORF amongst clinical isolates of HPV-6 obtained in the London (UK) area. We have also compared this variability to that reported for HPV-6 L1 isolates from different geographical locations. Our aim was, firstly, to establish the degree and distribution of the sequence variability throughout the L1 ORF and, secondly, to study the effect of major naturally occurring amino acid substitutions on the stability and yield of VLPs.

The DNA sequence analysis of the entire L1 ORF from 17 HPV-6 isolates carried out in this study revealed 28 locations where nucleotide substitutions occur compared to the prototype HPV-6b sequence. It is unlikely that any of these nucleotide changes are due to errors produced during the PCR or sequencing procedures. In all cases, sequencing was performed on multiple clones from each L1 isolate, in some cases from independent PCRs.

Specific silent nucleotide substitutions at eight fixed locations in the HPV-6 L1 ORF were observed in 76% of isolates. These eight substitutions have been previously described in either HPV-6a (Hofmann et al., 1995) or HPV-6 isolates from India, Japan, Philippines and the United States (Icenogle et al., 1991; Suzuki et al., 1997). These results indicate that the prototype HPV-6b is not the predominant subtype
found in condyloma acuminata biopsies from patients in either Asia, London (UK) or the United States.

None of the isolates analysed in this study possessed the reported nucleotide sequence of either HPV-6a or the prototype HPV-6b. Nonetheless, 23% of isolates gave rise to the same L1 amino acid sequence described for both HPV-6a and HPV-6b. The remaining isolates all possessed amino acid substitutions elsewhere in the L1 sequence.

The most frequent substitutions from the prototype HPV-6b L1 ORF sequence appeared to accumulate within three discrete regions: $R_1$ (nt 5920–6075), $R_2$ (nt 6590–6670) and $R_3$ (nt 7070–7230). The $R_1$ and $R_2$ regions contain mainly synonymous changes whilst the $R_3$ region contains predominantly non-synonymous changes. These three regions also contain the vast majority of the nucleotide substitutions from the prototype HPV-6b L1 ORF observed worldwide. However, none of these regions is fully represented in the 217 nt subfragment of the L1 ORF (nt 6860–7147 in HPV-6b) generally used in studies on intra- and inter-type sequence variation in HPVs (Chan et al., 1995; Stewart et al., 1996). This region provides only limited information on the phylogenetic relationships among the isolates analysed and HPV-6a and the prototype HPV-6b. Moreover, it is unable to differentiate between HPV-6a and the prototype HPV-6b since all the nucleotide substitutions characteristic of these subtypes fall outside the subfragment sequenced. On this basis, we suggest that sequencing of the $R_1$, $R_2$ and $R_3$ regions might be a more accurate method for the analysis of HPV-6 intratype variation.

Amongst the different amino acid substitutions observed in the different isolates analysed, two different groups are particularly interesting. The first group comprises the Ser-to-Pro substitutions in isolates M49898, M48364 and M47875, and the large deletion in isolate M48364. These amino acid changes are likely to render the L1 protein unable to acquire its correct structure and thus prevent the assembly of the viral capsid. However, since assembly-deficient mutants of the virus would be expected to become self-extinguishing, it is more likely that these amino acid changes appeared after integration of the viral genome into the host’s genome. We have no evidence of viral genome integration in any of the biopsies. Nonetheless, according to Cullen et al. (1991), mutations and deletions are common in integrated HPV genomes and previous reports indicate that HPV-6 is capable of integrating into the host’s genome (Bercovich et al., 1991; Alvarez et al., 1997).

The second, and most important group, is that comprising the most common amino acid change observed: a Glu-to-Gln substitution at aa 431. To our knowledge, this single amino acid substitution has not been described before and, although present in 41% of the isolates, represents the characteristic feature of 29% of all isolates analysed. In agreement with the phylogenetic tree drawn from the results of this study, we have defined this group of isolates as a new London variant of HPV-6. Expression in yeast of the HPV-6 L1 ORFs from both this London variant and an isolate encoding the sequence of the prototype HPV-6b L1 protein revealed no differences in the stability, yield and/or morphology of the VLPs produced.

In summary, we have identified three regions in the HPV-6 L1 ORF which contain the most frequent nucleotide substitutions from the prototype HPV-6b sequence reported worldwide. The predominant substitution in the most C-terminal of these regions causes a Glu-to-Gln change at aa 431, which constitutes the characteristic feature of a London variant of HPV-6 L1 and does not affect the self-assembly potential of L1. This L1 sequence represents a novel sequence variant of HPV-6 which is more commonly found in the population here under study (29%) than the previously described HPV-6a and prototype HPV-6L1 sequences (23%).

We thank Dr McCarthy (Middlesex Hospital) and Dr Hanna (St Mary’s Hospital) for the biopsy specimens. We also thank Dr M. A. Stanley (Dept. of Pathology, University of Cambridge) for the Camvir-1 MAb, and Mr J. Pacey (EM unit, King’s College London) for technical assistance with electron microscopy. This work was supported by a studentship to W.C.-W. from British Biotech Ltd and by a project grant from the MRC to D.H.D.

References


Received 28 September 1998; Accepted 18 November 1998