doi:10.1006/jmbi.2001.4464 available online at http://www.idealibrary.com on IDEAL® *J. Mol. Biol.* (2001) **307**, 173–182

JMB

Papillomavirus Capsid Protein Expression in *Escherichia coli*: Purification and Assembly of HPV11 and HPV16 L1

Xiaojiang S. Chen[1]*, Gregory Casini[2], Stephen C. Harrison[3] and Robert L. Garcea[2]

[1]*Department of Biochemistry & Molecular Genetics, University of Colorado School of Medicine, 4200 E. 9th Ave, Denver CO 80262, USA*

[2]*Section of Pediatric Hematology/Oncology University of Colorado School of Medicine, 4200 E. 9th Ave Denver, CO 80262, USA*

[3]*Harvard University and Howard Hughes Medical Institute, Department of Molecular and Cellular Biology 7 Divinity Ave, Cambridge MA 02138, USA*

**Corresponding author*

The L1 major capsid proteins of human papillomavirus (HPV) types 11 and 16 were purified and analyzed for structural integrity and *in vitro* self-assembly. Proteins were expressed in *Escherichia coli* as glutathione-*S*-transferase-L1 (GST-L1) fusions and purified to near homogeneity as pentamers (equivalent to viral capsomeres), after thrombin cleavage from the GST moiety and removal of tightly associated GroEL protein. Sequences at the amino and carboxy termini contributing to formation of L1 pentamers and to *in vitro* capsid assembly were identified by deletion analysis. For both HPV11 and HPV16 L1, up to at least ten residues could be deleted from the amino terminus (ΔN10) and 30 residues from the carboxy terminus (ΔC30) without affecting pentamer formation. The HPV16 pentamers assembled into relatively regular, 72-pentamer shells (''virus-like particles'' or VLPs) at low pH, with the exception of HPV16 L1 ΔN10, which assembled into a 12-pentamer, $T = 1$ capsid (small VLP) under all conditions tested. The production of large quantities of assembly-competent L1, using the expression and purification protocol described here, has been useful for crystallographic analysis, and will be valuable for studies of virus-receptor interactions and potentially for vaccine design.

Keywords: human papillomavirus; virus assembly; virus capsids; protein folding

Introduction

The Papovaviridae include two virus subfamilies: papilloma- and polyomaviruses. Polyomaviruses are DNA tumor viruses that transform cells in tissue culture and cause tumors in experimental animals. Papillomaviruses are important human pathogens, whose infection is directly related to the subsequent development of cervical cancer for high-risk subtypes such as human papillomavirus (HPV) 16 and 18 (Bosch *et al.*, 1995; Pisani *et al.*, 1993).

Both papilloma- and polyomavirus capsids have a similar overall architecture: 360 identical copies of their major capsid protein are arranged as 72 pentamers at the vertices of a $T = 7$ icosahedral lattice (Rayment *et al.*, 1982; Baker *et al.*, 1991). Surface feature details, as determined by cryoelectron microscopic image reconstructions of virions, differ noticeably, however (Belnap *et al.* , 1996; Trus *et al.*, 1997). Comparison of VP1 proteins among different polyomaviruses (e.g. SV40, murine polyoma) and L1 proteins among different papillomaviruses (e.g. bovine papilloma, HPV) shows that the primary sequences are clearly related within the two subfamilies, but that there is no detectable sequence similarity between L1s and VP1s.

The crystal structures of the VP1 major capsid proteins from murine polyomavirus and SV40 are known (Liddington *et al.*, 1991; Stehle *et al.*, 1994). In addition, the biochemistry of *in vitro* capsid assembly for recombinant VP1 capsomeres has been studied in detail (Garcea *et al.*, 1987; Salunke *et al.*, 1989). Polyoma VP1 contains a jellyroll β-barrel core, and outside the jellyroll core there are long and flexible arms at both amino and

Abbreviations used: HPV, human papillomavirus; GST, glutathione-S-transferase; VLP, virus-like particle; EM, electron microscopy; rt, room temperture.

E-mail address of corresponding author: xiaojiang.chen@uchsc.edu

0022-2836/01/010173–10 $35.00/0

carboxy termini. These arms are dispensable for pentamer formation, but parts of both are required for inter-pentamer interactions and thus essential for particle assembly. The positively charged N-terminal segment is a DNA-binding domain and likely interacts with the viral DNA. The domain organization of HPV L1 domain differs from that of polyoma VP1 in that the positively charged segment is at the extreme carboxy terminus (Li *et al*., 1997; Zhou *et al*., 1991), following a sequence of 70 residues important for inter-pentameric interactions (Chen *et al*., 2000).

Papillomaviruses are difficult to grow in cell culture, but the L1 proteins have been produced by recombinant expression in *Escherichia coli* and in insect cells using baculovirus vectors (Kirnbauer *et al*., 1992; Le Cann *et al*., 1994; Li *et al*., 1997; Rose *et al*., 1993; Xi & Banks, 1991). We report here a purification strategy for obtaining large quantities of pure, well-folded HPV11 and HPV16 L1 pentamers after recombinant expression in *E. coli*. Using deletion mutagenesis and purification of recombinant mutant L1 proteins, sequences at the amino and carboxy termini of L1 important for L1 folding and pentamer formation were identified. The carboxy-terminal sequences were important for particle assembly, and the amino terminus appeared to regulate the size of the icosahedral shell.

Results

Protein expression in *E. coli*

The L1 proteins from both HPV11 and HPV16 were analyzed in parallel in order to generalize the findings, as well as to utilize minor differences in the primary sequences to define domain boundaries. Both proteins behaved similarly, and only the differences are specifically discussed. The L1 coding sequences were cloned into pUC based vectors under the Ptac promoter for expression in *E. coli*. Using these constructs, the L1 protein was not detected by SDS-PAGE in whole cell lysates, despite efforts to improve expression by varying the DNA sequence near the L1 start codon, mutating several rare arginine codons, or changing bacterial strains (data not shown). The low level of L1 accumulation was subsequently determined to be due, in part, to protein degradation rather than low expression levels (data not shown). As an alternative approach, we fused glutathione-*S*-transferase (GST) to the L1 amino terminus using the pGEX-2T vector (Pharmacia). GST-L1 fusion proteins were detected by SDS-PAGE of whole cell lysates (Figure 1, lanes 2-5), indicating accumulation to high levels.

Purification of HPV11 and HPV16 L1

When expressed at 37 °C, the GST-L1 fusions were found in the insoluble cell pellet. When expression was induced at room temperature (rt), however, the GST-L1 fusions accumulated to high level, and partitioned into the soluble fraction of the cell lysates. When the soluble fraction was chromatographed on glutathione-Sepharose, the GST fusions were retained on the column, as shown by the 80 kDa species seen by SDS-PAGE in Figure 1, lanes 6 and 7. An additional intense 60 kDa band was present (Figure 1, lanes 6 and 7),

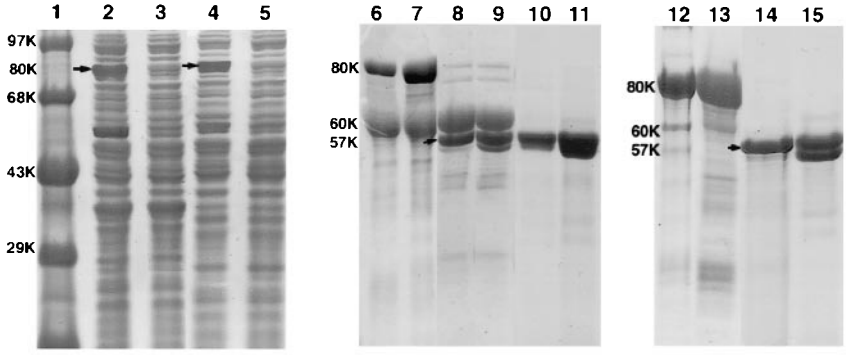

Figure 1. Purification of HPV11 and HPV16 L1 proteins after recombinant expression in *E. coli*. Shown is an SDS-PAGE analysis of GST-L1 and thrombin-released L1 proteins. Protein standards (lane 1); whole cell lysates induced and uninduced for HPV11 GST-L1 (lanes 2 and 3) and HPV16 GST-L1 (lanes 4 and 5) expression; GST-L1 fusions bound to glutathione resins for HPV11 (lane 6) and HPV16 (lane 7); glutathione affinity column eluates after thrombin cleavage for HPV11 (lane 8) and HPV16 (lane 9); L1 proteins precipitated after ATP-$MgCl_2$ treatment of the eluates in lane 8 (lane 10) and lane 9 (lane 11); glutathione affinity column eluates for HPV11 (lane 12) and HPV16 GST-L1 (lane 13) for preparations using the complete re-folding protocol (see Materials and Methods); eluates after thrombin cleavage from glutathione affinity column for HPV11 (lane 14) and HPV16 (lane 15) L1 proteins after using the re-folding purification protocol. The GST-L1 proteins of both virus types have an apparent molecular mass of 80 kDa (lanes 2, 4, 6, 7, 12, and 13). The 60 kDa species (lanes 6, 7, 8, and 9) is GroEL, as determined by N-terminal sequencing. Without using the re-folding protocol (lanes 6 to 9) the GroEL quantitatively co-eluted with L1. After using the refolding protocol, the GroEL contamination was almost eliminated (compare lanes 8 and 9 with lanes 14 and 15).

which could not be removed by extensive washing with various non-denaturing buffers. N-terminal sequencing identified the 60 kDa band as GroEL, an essential *E. coli* chaperone protein, which together with groES assists the folding of many proteins (Chen & Sigler, 1999; Weissman *et al*., 1994, 1995). Electron microscopy (EM) analysis also detected GroEL double heptamers in the column eluate (data not shown; Chen *et al*., 1994).

Thrombin digestion of the column-bound GST-L1 released L1 from the GST column, and this L1 is seen as a band migrating just below the 60 kDa GroEL band (Figure 1, lanes 8 and 9). GroEL also eluted from the column together with the cleaved L1 (Figure 1, lanes 8 and 9). The co-eluted GroEL could not be separated from L1 by gel-filtration or ion-exchange chromatography, suggesting a tight association. ATP-$MgCl_2$ releases GroEL from its substrate (Lissin *et al*., 1990; Viitanen *et al*., 1990). When ATP and $MgCl_2$ were added to the L1-GroEL eluate, a white precipitate formed. SDS-PAGE demonstrated that the precipitate was pure L1 (Figure 1, lanes 10 and 11). L1 could be separated from GroEL by treating the thrombin eluate with 2.5 M urea, but again all the L1 released from GroEL precipitated. These findings suggested that L1 was bound to GroEL in a such a way that its release led to an insoluble product.

If L1 was bound to GroEL in a ''non-native'' conformation, ATP treatment in the presence of GroES should facilitate release of ''native'' L1 protein. Therefore, the whole cell lysate, which also contains GroES, was treated with ATP and $MgCl_2$. GST-L1 was then purified by glutathione-Sepharose chromatography, yielding a GST-L1 fusion with reduced GroEL contamination (about 10-20 %). A similar reduction in the amount of GroEL was observed if the whole cell lysate was incubated with 3.5 M urea (see Materials and Methods). However, if the cell lysate was first treated with ATP and then with 3.5 M urea, the level of GroEL co-purifying with GST-L1 was reduced to approximately 1 % of the total protein (Figure 1, lanes 12 and 13), and the yield of L1 was increased two- threefold. The L1 released from the column by thrombin digestion was soluble and homogeneous (Figure 1, lanes 14 and 15) and could be concentrated to approximately 20 mg/ml^{-1}.

The purified L1 protein of HPV11 migrated on SDS-PAGE with an apparent molecular mass of 57 kDa (Figure 1, lane 14). In the case of HPV16 L1, however, there was a species of approximately 54 kDa in addition to the 57 kDa band (Figure 1, lane 15). The lower-to-upper band ratio increased with storage at 4 °C, suggesting proteolysis at either the amino or carboxy terminus. N-terminal sequencing of these two bands demonstrated an intact HPV16 L1 amino terminus (data not shown), indicating proteolysis at the carboxy terminus.

The purification protocols for the HPV16 and HPV11 L1 proteins are described in detail in Materials and Methods. The protocol is the same for both HPV11 and 16, except that the highest urea concentration for washing the glutathione-Sepharose column is 3.0 M for HPV11 L1, but only 2.3 M for HPV16 L1. A urea concentration greater than 2.3 M denatured and precipitated the HPV16 protein. A urea concentration of 3.5 M urea is optimal, however, for extraction of both HPV11 and 16 L1 from the total cell lysates. Using these conditions, approximately 3-5 mg of near homogeneous L1 was obtained from one liter of bacterial cell culture.

L1 Characterization

The purified L1 protein was analyzed by FPLC gel-filtration chromatography and negative-stain EM analysis (data not shown). Superdex-200 chromatography demonstrated a protein peak of about 300 kDa, consistent with pentamers. EM analysis showed typical pentamer ''donuts'', as previously seen for both L1 and VP1 purified after recombinant expression.

When L1 pentamers were digested with trypsin, the full-length 57 kDa L1 was reduced to a 42 kDa species on SDS-PAGE (Figure 2); the trypsin-digested pentamers appeared similar to untreated pentamers in the electron microscope (Li *et al*., 1997). N-terminal sequencing of the tryptic 42 kDa products recovered from SDS-PAGE identified an

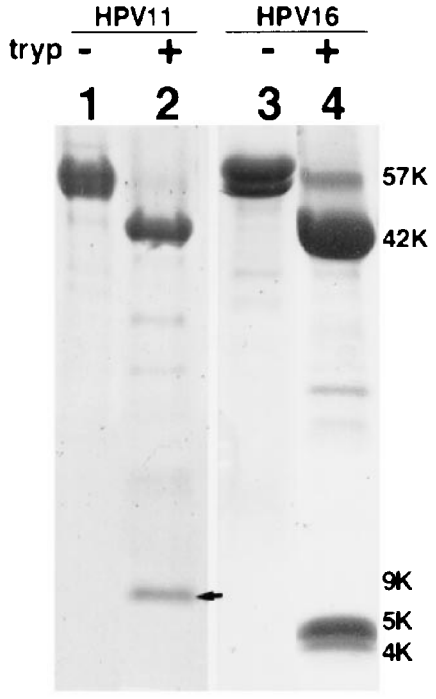

Figure 2. Trypsin digestion of purified HPV11 and HPV16 L1 proteins. Shown is an SDS-PAGE analysis of untreated (lanes 1, 3) and trypsin-digested HPV11 and HPV16 L1 proteins (lanes 2 and 4). In addition to a 42 kDa fragment common to both L1 types, trypsin digestion generated a 9 kDa product for HPV11 L1 and two species of approximately 4-5 kDa for HPV16 L1.

intact amino terminus for both for both HPV11 and HPV16 L1. N-terminal sequencing identified internal amino terminal sequences for HPV11 L1 as S_{348}-A-T-Y-T-N, and Y_{417}-V-Q-S-Q-A (corresponding to cleavage at K347 and K416), and for HPV16 L1 as N_{358}-T-N-F-K-E and F_{421}-V-T-S-Q-A (corresponding to cleavage at K357 and K420). A minor cleavage site was detected at K434 for HPV11 and K436 for HPV16.

When trypsin-digested L1 pentamers, presumably containing the 42 kDa fragments, were analyzed by Superdex-200 FPLC (15/60 column) they eluted at a position with an apparent molecular mass of 300 kDa, similar to that of the full-length L1. High-resolution SDS-PAGE analysis of the pentamer peak of trypsin-digested L1 from the Superdex-200 gel-filtration showed the 42 kDa band, and an additional 9 kDa band for HPV11 (Figure 2, lane 2), and two other bands, of approximately 4 and 5 kDa for HPV16 (Figure 2, lane 4). Thus the smaller tryptic carboxy-terminal fragments remained associated with the larger 42 kDa fragment during gel-filtration chromatography, implying that the carboxy-terminal tryptic fragments are integral parts of the L1 pentamer structure.

Carboxy-terminal sequences required for L1 pentamer formation

To determine the sequences at the carboxy terminus necessary for proper L1 folding, we generated a series of carboxy-terminal deletions for both HPV11 and HPV16 L1. HPV11 L1 deletions were made as intein fusions to the L1 carboxy terminus,

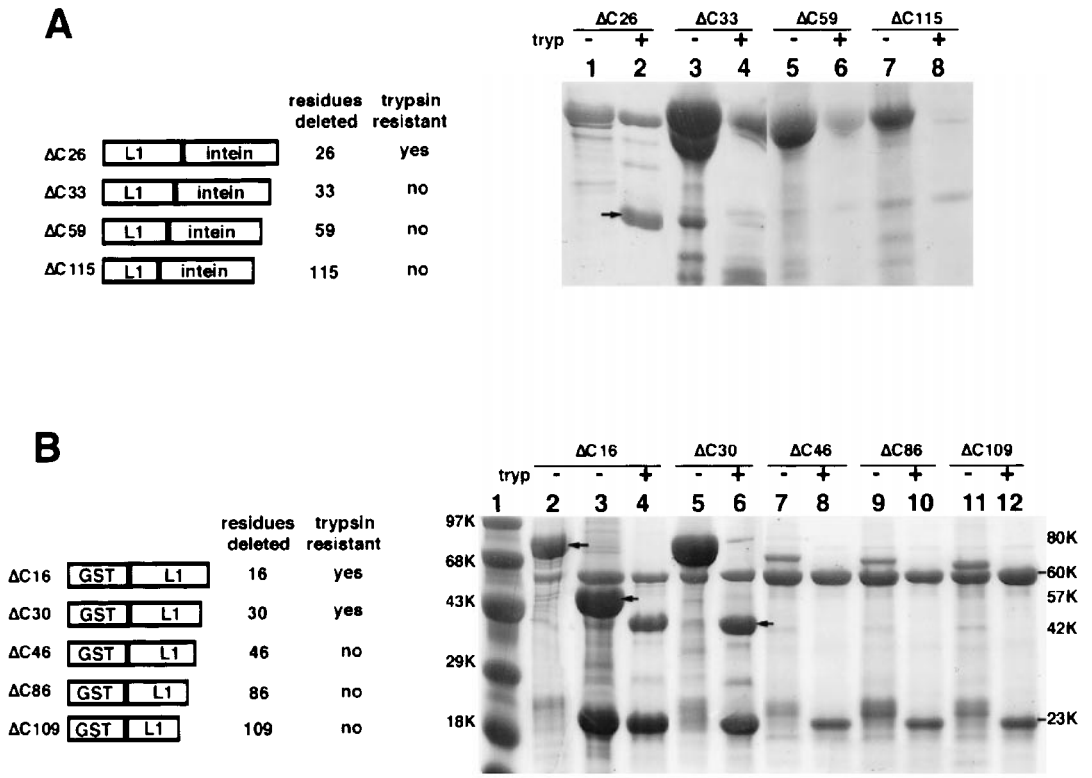

Figure 3. Pentamerization of carboxy-terminal deletion mutants of L1 proteins. The presence of a 42 kDa fragment after trypsin digestion was used as the criterion for proper folding and pentamer formation. (a) HPV11 L1 deletions were expressed as L1-intein fusions and whole cell lysates (incubated for 15 minutes with ATP-$MgCl_2$) were partially purified by chitin afffinity chromatography. SDS-PAGE analysis shows the L1 proteins (cleaved from the L1-intein fusion by incubating the chitin resin with 20 mM DTT overnight followed by elution with three bed volumes of buffer L) before and after trypsin digestion. L1 was not well separated from GroEL (lanes 1, 3, 5, and 7). L1 degradation to smaller species was seen for ΔC33 and longer deletions. After trypsin digestion, only ΔC26 yielded the 42 kDa fragment (lanes 2, 4, 6, and 8). The bands remaining at the top of the gel are GroEL. (b) HPV16 L1 deletions were expressed as GST-L1 fusions and whole cell lysates (incubated for 15 minutes with ATP-$MgCl_2$) were partially purified by glutathione affinity chromatography. The protease was added directly to the GST-L1 fusions bound to glutathione-Sepharose. SDS-PAGE analysis shows the L1 proteins (as GST-L1 fusions on the resin) before and after trypsin treatment. Before trypsin digestion, the amounts of intact GST-L1 for ΔC46, ΔC86 and ΔC109 (lanes 7, 9, and 11) were decreased in comparison with ΔC16 and ΔC30 (lanes 2 and 5). Lane 3 was ΔC16 treated with thrombin. After trypsin digestion, although free GST was seen in all cases (lanes 4, 6, 8, 10, and 12), the 42 kDa product was seen only for ΔC16 and ΔC30. The 60 kDa GroEL remained unchanged. GST bands after trypsin digestion were more intense than the intact GST-L1 bands for ΔC46, ΔC86 and ΔC109, suggesting the presence of non-intact GST-L1 fusions.

while HPV16 L1 truncations were made as GST fusions to the L1 amino terminus (Figure 3). The criteria for correct pentamer oligomerization and folding was the presence of the 42 kDa trypsin-resistant species after trypsin digestion of the purified deleted proteins.

The HPV11 L1-intein constructs included ΔC26 (26 residues deleted), ΔC33, ΔC59 and ΔC115 (Figure 3(a)). The expressed fusion proteins of each construct were partially purified by chitin affinity chromatography (Figure 3(a)), and the column eluates containing both GroEL and L1 were digested with trypsin and analyzed by SDS-PAGE (Figure 3). Trypsin digestion of ΔC26 yielded the 42 kDa trypsin-resistant band on SDS-PAGE (Figure 3(a), lane 2), indicating proper folding. For the other three carboxy-terminal deletions, however, the 42 kDa band disappeared after trypsin digestion (Figure 3(a), lanes 3-8), indicating that the products of these deletions were in a non-native conformation. EM analysis of these proteins demonstrated the presence of pentamers only for the ΔC26 protein (data not shown), consistent with the trypsin digestion results.

Carboxy-terminal deletions of the HPV16 GST-L1 fusion behaved similarly to the HPV11 L1-intein fusions. The HPV16 GST-L1 deletions included ΔC16, ΔC30, ΔC46, ΔC86 and ΔC109 (Figure 3(b)). SDS-PAGE analysis of the glutathione-Sepharose before addition of trypsin showed the presence of both the GST-L1 deletion fusion proteins and GroEL (Figure 3(b), lanes 2, 5, 7, 9 and 11). Trypsin digestion of the ΔC16 and ΔC30 fusions yielded the 42 kDa band (Figure 3(b), lanes 4 and 6), ΔC46 and longer deletions of L1 were degraded completely (Figure 3(b), lanes 7-12). Again, EM analysis revealed that only ΔC16 and ΔC30 had a pentamer appearance. Only a single band was present on SDS-PAGE for the HPV16 ΔC30 protein after purification and thrombin cleavage (data not shown), corresponding to the lower band of the doublet for full-length L1 (Figure 1, lane 15). This result implies that a protease-sensitive site on full-length HPV16 L1 lies between the carboxy-terminal 26-30 residues, where there are indeed three lysine residues.

Amino-terminal sequences required for L1 pentamer formation

The N-terminal sequencing results showed that the L1 amino terminus is unaffected by trypsin digestion. In order to define the role of amino terminus region in L1 folding and pentamer formation, we made amino-terminal deletions in both HPV11 and HPV16 L1 (Figure 4). For L1-intein fusions of HPV11, ΔN5, ΔN8, ΔN14 and ΔN19 were expressed, the fusion proteins partially purified (Figure 4(a)) and subjected to trypsin digestion. The trypsin-resistant 42 kDa band was present for ΔN5 and ΔN8 (Figure 4(a), lanes 1, 2 and 5, 6), faintly present for ΔN14, and almost absent for ΔN19 (Figure 4(a), lanes 3, 4 and 7, 8). This result suggests that ΔN5 and ΔN8 proteins fold into trypsin-resistant pentamers, whereas longer deletions result in either only small amounts of the folded pentamers (ΔN14) or none at all (ΔN19).

For HPV16 GST-L1, the deletion constructs are shown in Figure 4(b). Trypsin digestion of the partially purified GST-L1 yielded the 42 kDa trypsin-resistant band for ΔN8, ΔN10 and to some extent for ΔN13 (Figure 4(b), lanes 1, 2, 3 and 5, 6, 7). The GST-L1 fusion of ΔN13 was not efficiently digested by trypsin (Figure 4(b), lanes 3 and 7); some full-length GST-L1 fusion was present even after prolonged incubation with trypsin. For ΔN20 (and ΔN30, not shown), the GST-L1 fusion proteins and GroEL were not detected by SDS-PAGE (Figure 4(b), lane 4). Instead, a broad band with the mobility of GST was present (Figure 4(b), lanes 4 and 8). It is possible that the L1 moieties of these GST-L1 fusions were degraded after expression but before purification.

Effect of deletions on *in vitro* capsid assembly

HPV16 L1 deleted proteins capable of pentamer formation were tested for *in vitro* self-assembly into higher-order structures (Table 1). All the amino-terminal and carboxy-terminal deletions except ΔN10 assembled into $T = 7$ particles under the conditions used (Figure 5(a)), but ΔN10

Table 1. The assembly of HPV16 L1 deletion mutants under different pH values

Constructs	Size of assembled particles				
	pH 4.0	pH 5.4	pH 6.2	pH 7.5	pH 8.5
Δ N-0	T7	T7	T7	T7/pent	pent/T7
Δ N-4	NT	T7	T7	T7/pent	NT
Δ N-6	NT	T7	T7	T7/pent	NT
Δ N-8	NT	T7	T7	T7/pent	pent/T7
Δ N-10	T1	T1	T1	T1	pent
Δ N-10 + Gly	T7	T7	T7	NT	NT
Δ N-10 + Thr(ΔN-9)	T7	T7	T7	NT	NT
Δ C-16	T7	T7	T7	NT	NT
Δ C-30	T7	T7	T7	NT	NT

Assembly properties of different HPV16 L1 truncation mutants at different pH. The assembly buffers were: pH 4.0 and 5.4, 40 mM sodium acetate; pH 6.2 and 7.5, 40 mM Hepes; pH 8.5, 40 mM Tris-HCl; all contained 1 M Nacl. NT, not tested.

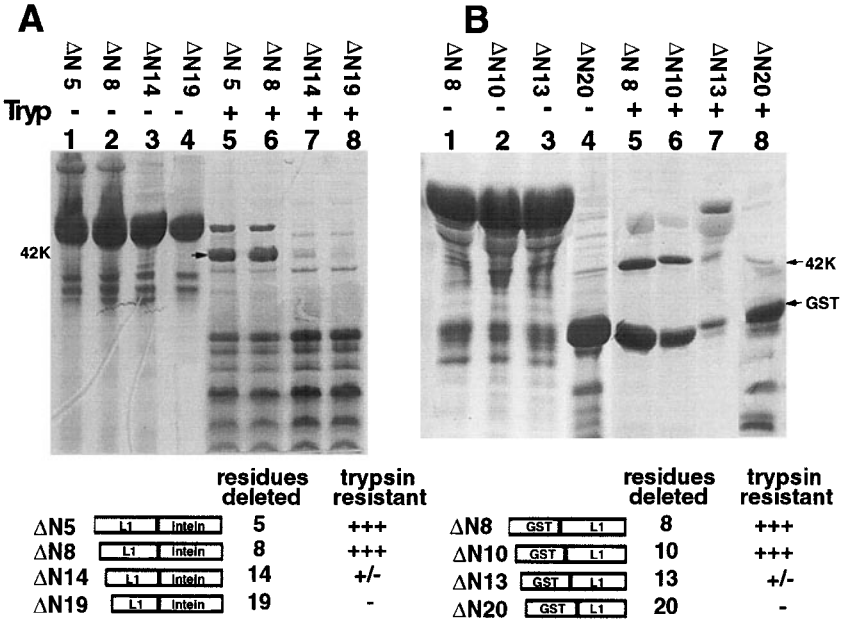

Figure 4. Pentamerization of amino-terminal deletion mutants of L1 proteins. The presence of a 42 kDa fragment after trypsin digestion was used as a criterion for proper folding and pentamer formation. (a) HPV11 L1 deletions were expressed as L1-intein fusions and whole cell lysates (incubated for 15 minutes with ATP-$MgCl_2$) were partially purified by chitin affinity chromatography. SDS-PAGE analysis shows the L1 proteins (as intein-L1 fusions on the resin) before and after trypsin treatment. Before trypsin digestion, full-length fusions are seen at the top of the gel for ΔN5 and ΔN8 (lanes 1 and 2). This band intensity decreases for ΔN14 (lane 3) and almost disappears for ΔN19 (lane 4). There is an intense band at the position of 60 kDa for each lane, which contains mostly intein-CBP (Chitin Binding Protein) and a small amount of GroEL (confirmed by N-terminal sequencing). The free intein-CBP implies extensive proteolysis of the fusion between L1 and intein. The cleaved L1 (together with some GroEL) does not bind chitin beads during purification. After trypsin treatment, the 42 kDa band is present for ΔN5 and ΔN8 (lanes 5 and 6), to some extent for ΔN14 (lane 7), but absent for ΔN19 (lane 8). Free chitin-CBP is digested to smaller fragments (lanes 5 and 8). (b) HPV16 L1 deletions were expressed as GST-L1 fusions, and whole cell lysates (incubated for 15 minutes with ATP-$MgCl_2$) were partially purified by glutathione affinity chromatography. SDS-PAGE shows the L1 proteins (as GST-L1 fusions on the resin) before and after trypsin treatment. Before trypsin digestion, full-length fusions for ΔN8, ΔN10, and ΔN13 migrate at the top of the gel (lanes 1, 2, and 3). Only free GST is present for ΔN20, possibly due to proteolytic degradation of the L1 portion during purification (lane 4). After trypsin digestion, the 42 kDa fragment is present for ΔN5 and ΔN10 (lanes 5 and 6). ΔN13 shows some 42 kDa and undigested fusion proteins (lane 7).

assembled into $T = 1$ particles under all conditions tested (Figure 5(b)). The precise deletion was critical. When even one Gly or Thr residue (the tenth L1 residue) was added to the amino terminus of ΔN10 (ΔN10 + Gly or ΔN10 + Thr), the proteins assembled into $T = 7$ particles (Figure 5, (c) and (d)).

The appearance of assembled products for each construct were generally the same in buffers with different pH (Table 1). The most consistent results were obtained at pH 5.4. Clumping of pentamers tended to become more severe at high pH, possibly due to disulfide bond formation of the free cysteine residues, and oligomeric pentamers (aggregates) were the predominant species at pH 8.5.

Recently the ΔN10 deletion of HPV16 L1 was crystallized as a $T = 1$ particle, and its X-ray structure determined (Chen *et al.*, 2000). The structure showed that the carboxy terminal segment from residue 384-446 folds into three helices with connecting loops and turns (Figure 6). These helices constitute the primary interpentamer bonding contacts in the assembled $T = 1$ particle. To test whether these helices also affect particle assembly, L1 proteins were generated with a specific deletion of helix 4 (see the structure in Figure 6) for both HPV16 (residues 408-431), and HPV11(residues 409-429). Pentamers were purified for these deleted proteins, as demonstrated by their FPLC elution profile and a donut appearance seen by EM. No assembly of particles from these pentamers was found under any condition tested (Figure 5(e)), suggesting that this carboxy-terminal helical domain is essential for particle assembly.

Discussion

Previous efforts to express HPV L1 in *E. coli* have produced only modest quantities of assembly-competent pentamers. The protocol described here gives a substantially enhanced yield of pure L1, using a relatively straightforward

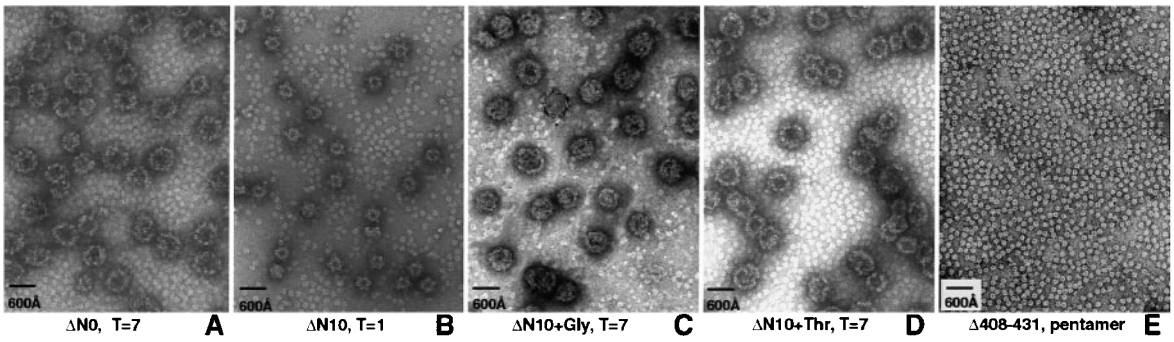

Figure 5. *In vitro* self-assembly of carboxy and amino-terminal deleted HPV16 L1 proteins. Electron micrographs of assembly reactions conducted in 1 M NaCl and 40 mM sodium acetate (pH 5.4), for purified L1 proteins. The bar length represents 600 Å, the diameter of a papillomavirus virion. (a) $T = 7$ particles (with diameter of about 600 Å) assembled from ΔN0 (full-length L1), with many unassembled pentamers in the background. (b) $T = 1$ particles (diameter of about 300 Å) assembled from ΔN10. (c) and (d) show the assembly products for ΔN10 + Gly and ΔN10 + Thr. The particles in (c) and (d) have a diameter of approximately 600 Å ($T = 7$), like those of ΔN0. (e) Free pentamers are the only products of *in vitro* assembly for Δ409-431L1.

purification strategy. Two factors appear to have been critical: (1) the use of GST-L1 fusion proteins, which greatly facilitated purification; and (2) a purification strategy that separated native L1 from the GroEL-bound complex. Despite the 27 kDa amino-terminal addition of the GST moiety, the fusion polypeptides fold and form pentamers properly, as evidenced by the release of L1 pentamers from the fusions by thrombin and their ability to be crystallized (Chen *et al.*, 2000).

GroEL release was accomplished by adding ATP-$MgCl_2$ to the cell lysate, followed by treatment with 3.5 M urea. We observed that 3.5 M urea alone released about half of the GroEL. EM analysis demonstrated that GroEL double heptamers began to disassemble into single subunits in 2.0 M urea and disassembled completely in 3.5 M urea, while L1 pentamers withstood these urea concentrations. ATP-$MgCl_2$ alone also released about half of the GroEL. Neither urea nor ATP could be used used to release GroEL from the purified L1-GroEL complex. Factors in the cell lysate presumably participated in the success of both treatments. In the case of ATP treatment, these factors may have included GroES, which is required for normal ATP-dependent release of proteins from

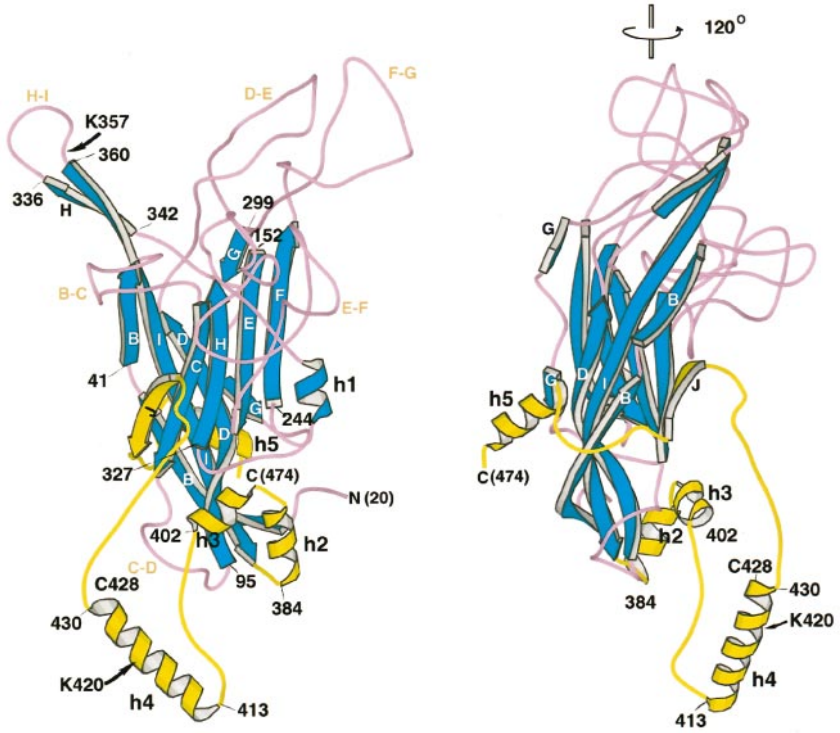

Figure 6. Three-dimensional structure of the HPV16 L1 monomer (Chen *et al.*, 2000). The orientation in (b) is rotated approximately 120° about the vertical axis from the orientation in (a). The β-strands are labeled by letter B through J. The starting residues of most β-strands and the starting and ending residue numbers for the carboxy-terminal α-helices are identified. Trypsin-sensitive sites are marked by arrows. Deletion of h4 (helix 4) of both HPV11 and HPV16 L1 abolished assembly of particles (see Figure 5 (e)).

GroEL, but why the urea treatment was unsuccessful when applied later in the purification is less evident.

How can we picture the GroEL/GST-L1 complex? The staining ratio in Figure 1 indicates about three GroEL subunits per GST-L1, or about one GroEL double heptamer per L1 pentamer. Since there is substantially less than one GroEL double heptamer per L1 monomer, the L1 must be largely folded and assembled, even with GroEL attached. The release of the GroEL by treatment with ATP-$MgCl_2$ in the presence of factors from the *E. coli* lysate may therefore proceed as a ''normal'' completion of the GroE cycle. Successful use of urea to dissociate GroEL and to produce assembly-competent L1 pentamers would require that the part of L1 interacting with GroEL be able to fold spontaneously once released.

The preparation of truncated forms of HPV11 and HPV16 L1 allowed further definition of the L1 domains required for pentamer and particle formation beyond that previously reported (Chen *et al.*, 2000). Up to 30 residues can be deleted from the carboxy terminus and at least ten residues from the amino terminus, without affecting pentamer formation. The amino-terminal 20 residues and the carboxy-terminal 30 residues are disordered in the crystal structure, and the deletion analysis thus shows that residues not part of the core of L1 (which we define as residues 21-474 for HPV16 and 19-471 for HPV11) are also not essential for pentamer assembly. A similar observation has been reported for polyomavirus VP1, where a protein with 32 residues deleted at the amino terminus and 63 residues deleted at the carboxy terminus can still form a pentamer amenable to crystallization and structure analysis (Stehle *et al.*, 1994). As in the case of L1, deletions compatible with pentamer assembly are ones that leave the protein core intact.

The carboxy-terminal sequences of L1 from different papillomaviruses are particularly variable. For example, the similarity of HPV16 and HPV11 L1 sequences ceases abruptly 30 residues from the carboxy terminus, just C-terminal to helix h5 (Chen *et al.*, 2000). Immediately preceding this boundary in the sequence are the residues that participate in the primary interpentameric contacts. Helices h2, h3, and h4 (see Figure 6) constitute a small subdomain that projects laterally from the β-jelly roll of L1, and helix h5 anchors this interaction domain by inserting back into the central cavity of the L1 pentamer. Data presented here, and related results in the literature, demonstrate that an intact h5 is essential for pentamer assembly as well as for subsequent capsid formation. For example, BPV L1 forms capsids with a 24 residue carboxy-terminal deletion, but not with a 40 residue deletion (Paintsil *et al.*, 1996); likewise, stable folding of canine oral papillomavirus (COPV) L1 tolerates a 26 residue, but not a 67 residue, deletion (Chen *et al.*, 1998). Our finding, that deletion of even three residues beyond the 30 residue boundary in HPV11 L1 (HPV11-L1 ΔC33) produces an unfolded, trypsin-sensitive endproduct, demonstrates the importance of the complete L1 core for pentamer stability, including the helical subdomain and its h5 anchor. Our observation that the 42 kDa fragment obtained by limited proteolysis is not an independently stable entity leads to the same conclusion.

Because the helical subdomain of L1 is necessary for the integrity and stability of the pentamer itself, in addition to being essential for interactions among pentamers in the capsid, we presume that it adopts its folded structure prior to viral assembly. That is, a carboxy-terminal segment does not exit the L1 monomer to form a helix 5 contact with an L1 in a neighboring pentamer. In contrast, the entire, 63 residue, carboxy-terminal arm of polyomavirus VP1 is disordered on the subunits of free pentamers, and it folds into ordered structures only when it ''invades'' a neighboring pentamer as capsids assemble (Liddington *et al.*, 1991).

How does truncation of the amino terminus of L1 affect the design of the assembled particle? In a $T = 1$ small VLP, all of the carboxy-terminal interaction domains contact neighbors in identical, symmetrically related ways. In the 72 pentamer shells of virions, there must be some variability in the way these domains interact. Moreover, the curvature of the virion shell is much smaller than that of the small VLP, so the carboxy-terminal domains must face their interaction partners at different angles from that found in the $T = 1$ particles. The amino-terminal residues of the polypeptide chain lie immediately adjacent to the carboxy-terminal interaction domain, and it is reasonable to postulate that their presence prevents the high-curvature interaction seen in the small VLPs. A structure for the 72 pentamer shell would be required to understand precisely why the switch to small-shell assembly occurs so abruptly with deletion of the tenth amino-terminal residue.

In summary, the protocol described in detail in this paper permits expression and purification of large quantities of HPV L1 proteins. We have already used this procedure to determine the structure of HPV 16 L1, and we anticipate that the same methods will be useful for crystallizing L1 proteins from additional HPV types. The resulting pentamers and *in vitro* assembled particles will provide reagents for co-crystallization efforts, new substrates for immunologic studies, and potential new vaccine immunogens.

Materials and Methods

Cloning and deletion constructs

L1 clones of HPV 11 from human laryngeal (accession no. M14119) and of an HPV16 strain from human cervical carcinoma (accession no. AF140325) were used for this study. L1 PCR products were inserted downstream of GST coding sequence of pGEX-2T vector (Pharmacia) to generate GST fusions at the amino terminus of L1. For

HPV11 L1, primer 1 (forward) 5′ ACCTAC AGATCT GGT ATGTGGCGTCCTAGCGACAGC and primer 2 (reversal) 5′ CA GAATCC TTACTTTTTGGTTTTGG-TACGTT were used for amplifying the L1 coding sequence and the PCR fragment were digested with *Bgl* II/*Eco*R I for inserting into *Bam*HI/*Eco*RI-digested pGEX2T vector. The primers for HPV16 L1, forward 5′ CTGA AGATCT GGT ATG TCC CTG TGG CTG CCT AGT GAG GCC ACT GTC and reversal 5′ CAGTTC GATATC TTAACGCTTACGTTTTTTGCGTTTAGC were used for PCR and the PCR product was digested with *Bgl*II/*Eco*RV for inserting into *Bam*HI/*Sma*I sites of pGEX2T.

The carboxy-terminal fusion constructs of HPV11 L1 with intein, were generated by cloning L1 PCR fragments into the pCYB1 vector (IMPACT I system, NEB Biolabs). The primers for PCR cloning were: forward 5′ CAATGTCA CATATG TGGCGTCCATCTGAT AGCAC TGTATACGTCCACCTCCAAACCCTGTATCC and reversal 5′ AGACTT GCTCTTCC GCAGTTTTT GGTTTTGGTACGTTTTCG. The PCR product was digested with *Nde*I/*Sap*I and then ligated into *Nde*I/*Sap*I-digested pCYB1 vector. *E. coli* strain XA90 [Δ*lacproXIII, nalA, argE(am), thi, rifr/F′lacI*^{q}*′ZY,proAB*] was used for all the cloning and expression studies. This strain consistently expressed L1 fusions to higher level than other strains tested.

Deletion mutants were constructed using PCR based on the full-length L1 fusion clones. Sets of two primers were used for PCR to amplify the whole plasmid containing L1 gene except for the nucleotides to be deleted. The PCR fragment contained a new unique restriction site on both ends (incorporated into the primers) for digestion and ligation. Clones were first identified by digestion of the newly generated site and then further confirmed by DNA sequencing. This strategy proved to be a very efficient deletion method. Plasmids of 6-8 kb were routinely amplified by PCR using the following conditions: 94 °C for 1.5 minutes, 65 °C for 1 minute, 68 °C 12 minutes, two cycles of PCR, then 94 °C for 1.5 minutes, 68 °C for 13 minutes, 23 cycles of PCR. The reaction contained five units of Taq polymerase and 0.25 unit of Pfu (Strategen). The inclusion of Pfu at the concentration of 1/20 Taq units in the PCR reaction was essential to obtain large quantities of full-length PCR products for cloning.

Protein Purification

The detailed protocols of protein expression and purification of both HPV11 and HPV16 L1 are described as follows. A 2 ml overnight cell culture was inoculated to one liter of 2× YT at 37 °C. Cells were grown untill $A_{600} = 0.1$-0.3. The flasks were transferred to 25 °C and incubated with shaking for 30 minutes. IPTG was added to 0.2 mM and protein expression induced for 16-20 hours. Cells from a one liter culture were resuspended in 30 ml of buffer L (50 mM Tris-HCl (pH 8.0), 0.2 M NaCl, 1 mM DTT, 1 mM EDTA) and lysed by sonication. After sonication, ATP and $MgCl_2$ were added to final concentrations of 2 mM and 5 mM, respectively. After one hour incubation at room temperture, urea (ultrapure grade) was slowly added to the lysate to a final concentration of 3.5 M. The mixture was incubated at room temperature for one hour with gentle shaking, and then dialyzed against three changes of buffer over 18 hours. The lysate was clarified by centrifugation at 25,000 ***g*** for 75 minutes. The GST-L1 (or L1-intein) in the supernatant was purified using a glutathione affinity column (or chitin chromatography) using 15 ml bed volume for a lysate from eight liters of bacteria. After washing the column with ten bed volumes of buffer L, residual GroEL was eluted by washing with ten bed volumes of 3.0 M urea (in buffer L) for HPV11 L1 or 2.3 M urea for HPV16 L1. L1 was cleaved from the GST fusion using thrombin (Sigma, T6634) in an approximate ratio of 100 μg GST-L1 to one NIH unit of enzyme. The digestion was carried out at 4 °C overnight. L1 was further purified by Superdex-200 (16/60 column) gel-filtration FPLC. L1-intein fusions were cleaved by incubation with DTT at 4 °C overnight (IMPACT I system of New England BioLabs).

Trypsin Digestion and *In vitro* Assembly

Limited trypsin digestion of L1 (trypsin to L1 ratio of 1:1000) was performed at room temperature, and the digestion was terminated by adding 1 mM PMSF and an equal volume of 2× SDS sample buffer followed by immediate boiling. Assembly assays were conducted using two methods. First, L1 protein was dialyzed (at a concentration of approximately 0.1 mg/ml) against the assembly buffer in a micro-dialysis button (Pierce) for 60 minutes. Second, concentrated L1 protein (10-20 mg/ml) was diluted 100-1000-fold directly into assembly buffer and incubated at room temperature for 30 minutes. Both methods gave equivalent results. For EM analysis, the L1 protein was spotted on glow-discharged, carbon-coated grids (EM Science) and stained with 2% (w/v) uranyl acetate. A JEOL 100-CX electron microscope was used for visualization.

Acknowledgments

We thank Dr Renee Finnen and Dr Yorgo Modis for generating the helix 4-deletion mutants of HPV 11 and HPV16 L1. The work was supported by NIH grants CA13202 (S.C.H.) and CA37667 (R.L.G.). S.C.H. is an investigator in the Howard Hughes Medical Institute.

References

Baker, T. S., Newcomb, W. W., Olson, N. H., Cowsert, L. M., Olson, C. & Brown, J. C. (1991). Structures of bovine and human papillomaviruses. Analysis by cryoelectron microscopy and three-dimensional image reconstruction. *Biophys. J.* **60**, 1445-1456.

Belnap, D. M., Olson, N. H., Cladel, N. M., Newcomb, W. W., Brown, J. C., Kreider, J. W., Christensen, N. D. & Baker, T. S. (1996). Conserved features in papillomavirus and polyomavirus capsids. *J. Mol. Biol.* **259**, 249-263.

Bosch, F. X., Manos, M. M., Munoz, N., Sherman, M., Jansen, A. M., Peto, J., Schiffman, M. H., Moreno, V., Kurman, R. & Shah, K. V. (1995). Prevalence of human papillomavirus in cervical cancer: a worldwide perspective. International biological study on cervical cancer (IBSCC) Study Group. *J. Natl Cancer Inst.* **87**, 796-802.

Chen, L. & Sigler, P. B. (1999). The crystal structure of a GroEL/peptide complex: plasticity as a basis for substrate diversity. *Cell,* **99**, 757-768.

Chen, S., Roseman, A. M., Hunter, A. S., Wood, S. P., Burston, S. G., Ranson, N. A., Clarke, A. R. & Saibil, H. R. (1994). Location of a folding protein

and shape changes in GroEL-GroES complexes imaged by cryo-electron microscopy. *Nature,* **371**, 261-264.

Chen, X., Garcea, R., Goldberg, I., Casini, G. & Harrison, S. C. (2000). Structure of small virus-like particles assembled from the L1 protein of human papillomavirus 16. *Mol. Cell,* **5**, 557-567.

Chen, Y., Ghim, S. J., Jenson, A. B. & Schlegel, R. (1998). Mutant canine oral papillomavirus L1 capsid proteins which form virus-like particles but lack native conformational epitopes. *J. Gen. Virol.* **79**, 2137-2146.

Garcea, R. L., Salunke, D. M. & Caspar, D. L. (1987). Site-directed mutation affecting polyomavirus capsid self-assembly *in vitro*. *Nature,* **329**, 86-87.

Kirnbauer, R., Booy, F., Cheng, N., Lowy, D. R. & Schiller, J. T. (1992). Papillomavirus L1 major capsid protein self-assembles into virus-like particles that are highly immunogenic. *Proc. Natl Acad. Sci. USA,* **89**, 12180-12184.

Le Cann, P., Coursaget, P., Iochmann, S. & Touze, A. (1994). Self-assembly of human papillomavirus type 16 capsids by expression of the L1 protein in insect cells. *FEMS Microbiol. Letters,* **117**, 269-274.

Li, M., Cripe, T. P., Estes, P. A., Lyon, M. K., Rose, R. C. & Garcea, R. L. (1997). Expression of the human papillomavirus type 11 L1 capsid protein in *Escherichia coli*: characterization of protein domains involved in DNA binding and capsid assembly. *J. Virol.* **71**, 2988-2995.

Liddington, R. C., Yan, Y., Moulai, J., Sahli, R., Benjamin, T. L. & Harrison, S. C. (1991). Structure of simian virus 40 at 3.8 Å resolution. *Nature,* **354**, 278-284.

Lissin, N. M., Venyaminov, S. & Girshovich, A. S. (1990). (Mg-ATP)-dependent self-assembly of molecular chaperone GroEL. *Nature,* **348**, 339-342.

Paintsil, J., Muller, M., Picken, M., Gissmann, L. & Zhou, J. (1996). Carboxyl terminus of bovine papillomavirus type-1 L1 protein is not required for capsid formation. *Virology,* **223**, 238-44.

Pisani, P., Parkin, D. M. & Ferlay, J. (1993). Estimates of the worldwide mortality from eighteen major cancers in 1985. Implications for prevention and projections of future burden. *Int. J. Cancer,* **55**, 891-903.

Rayment, I., Baker, T. S., Caspar, D. L. & Murakami, W. T. (1982). Polyoma virus capsid structure at 22.5 Å resolution. *Nature,* **295**, 110-115.

Rose, R. C., Bonnez, W., Reichman, R. C. & Garcea, R. L. (1993). Expression of human papillomavirus type 11 L1 protein in insect cells: *in vivo* and *in vitro* assembly of viruslike particles. *J. Virol.* **67**, 1936-44.

Salunke, D. M., Caspar, D. L. & Garcea, R. L. (1989). Polymorphism in the assembly of polyomavirus capsid protein VP1. *Biophys. J.* **56**, 887-900.

Stehle, T., Yan, Y., Benjamin, T. L. & Harrison, S. C. (1994). Structure of murine polyomavirus complexed with an oligosaccharide receptor fragment. *Nature,* **369**, 160-163.

Trus, B. L., Roden, R. B., Greenstone, H. L., Vrhel, M., Schiller, J. T. & Booy, F. P. (1997). Novel structural features of bovine papillomavirus capsid revealed by a three-dimensional reconstruction to 9 Å resolution. *Nature Struct. Biol.* **4**, 413-420.

Viitanen, P. V., Lubben, T. H., Reed, J., Goloubinoff, P., O'Keefe, D. P. & Lorimer, G. H. (1990). Chaperonin-facilitated refolding of ribulosebisphosphate carboxylase and ATP hydrolysis by chaperonin 60 (groEL) are K^+ dependent. *Biochemistry,* **29**, 5665-5671.

Weissman, J. S., Kashi, Y., Fenton, W. A. & Horwich, A. L. (1994). GroEL-mediated protein folding proceeds by multiple rounds of binding and release of nonnative forms. *Cell,* **78**, 693-702.

Weissman, J. S., Hohl, C. M., Kovalenko, O., Kashi, Y., Chen, S., Braig, K., Saibil, H. R., Fenton, W. A. & Horwich, A. L. (1995). Mechanism of GroEL action: productive release of polypeptide from a sequestered position under GroES. *Cell,* **83**, 577-587.

Xi, S. Z. & Banks, L. M. (1991). Baculovirus expression of the human papillomavirus type 16 capsid proteins: detection of L1-L2 protein complexes. *J. Gen. Virol.* **72**, 2981-2988.

Zhou, J., Doorbar, J., Sun, X. Y., Crawford, L. V., McLean, C. S. & Frazer, I. H. (1991). Identification of the nuclear localization signal of human papillomavirus type 16 L1 protein. *Virology,* **185**, 625-232.

Edited by M. Yaniv

(Received 23 October 2000; received in revised form 20 December 2000; accepted 9 January 2001)