The EMBO Journal Vol. 21 No. 18 pp. 4754–4762, 2002

Atomic model of the papillomavirus capsid

Yorgo Modis, Benes L.Trus[1] and Stephen C.Harrison[2]

Howard Hughes Medical Institute, Children's Hospital and Harvard Medical School, 320 Longwood Avenue, Boston, MA 02115 and [1]Computational Bioscience and Engineering Laboratory, Division of Computer Research and Technology, and Laboratory of Structural Biology, National Institute of Arthritis, Musculoskeletal, and Skin Diseases, National Institutes of Health, Bethesda, MD 20892-5624, USA

[2]Corresponding author
e-mail: harrison@crystal.harvard.edu

Papillomaviruses propagate in differentiating skin cells, and certain types are responsible for the onset of cervical cancer. We have combined image reconstructions from electron cryomicroscopy (cryoEM) of bovine papillomavirus at 9 Å resolution with coordinates from the crystal structure of small virus-like particles of the human papillomavirus type 16 L1 protein to generate an atomic model of the virion. The overall fit of the L1 model into the cryoEM map is excellent, but residues 402–446 in the 'C-terminal arm' must be rebuilt. We propose a detailed model for the structure of this arm, based on two constraints: the presence of an intermolecular disulfide bond linking residues 175 and 428, and the clear identification of a feature in the image reconstruction corresponding to an α-helix near the C-terminus of L1. We have confirmed the presence of the disulfide bond by mass spectrometry. Our 'invading arm' model shows that papilloma- and polyomaviruses have a conserved capsid architecture. Most of the rebuilt C-terminal arm is exposed on the viral surface; it is likely to have a role in infection and in immunogenicity.

Keywords: disulfide bond/electron microscopy/papillomavirus/vaccine design/virus assembly

Introduction

Papillomaviruses are double-stranded DNA viruses that propagate in differentiating skin cells. Over 100 human types been identified, of which a few—notably types 16, 18, 31 and 45—are known to be responsible for the onset of cervical cancer (Bosch *et al*., 1995; Parkin *et al*., 2000), the second most common cancer among women worldwide (Parkin *et al*., 2000). Papilloma virions are non-enveloped particles. Their outer shell, ~600 Å in diameter, contains 72 pentamers of the protein L1, situated at the vertices of a T = 7*dextro* icosahedral lattice (Figure 1B; Baker *et al*., 1991). Cellular histones condense the 8 kb genome into a double-stranded, covalently closed circular DNA minichromosome (Howley, 1996).

Electron microscopy (EM) of virions in vitreous ice reveals that cottontail rabbit papillomavirus, bovine papillomavirus type 1 (BPV1) and human papillomavirus type 1 (HPV1) have essentially identical structures (Baker *et al*., 1991; Hagensee *et al*., 1994; Belnap *et al*., 1996; Trus *et al*., 1997); indeed, their L1 subunits have nearly 50% sequence identity. The atomic structure of recombinant HPV16 L1 has been determined from crystals of a 12-pentamer assembly, or small virus-like particle (small VLP), stable at low pH (Figure 1C–F; Chen *et al*., 2000). Despite a lack of significant sequence similarity, the architecture of the papillomavirus L1 pentamer closely resembles that of the VP1 pentamers of simian virus 40 (SV40) and murine polyomavirus: a ring of five β-jellyroll domains, tightly linked by interacting loops between framework β-strands (Figure 1E and F; Liddington *et al*., 1991; Stehle *et al*., 1994). A C-terminal subdomain of the subunit mediates pentamer–pentamer contacts in the small VLPs of HPV16. The polypeptide chain exits the β-jellyroll, forms the protruding subdomain, and reinserts into the central channel of the pentamer from which it arises (Figure 1C; Chen *et al*., 2000). By contrast, the C-terminal arms of the VP1 subunits in polyomavirus 'invade' neighboring pentamers, thereby tying the virion shell together (Figure 1A; Liddington *et al*., 1991). The observation that there might be distinct modes of subunit association in the two related classes of viruses is not by itself a paradoxical finding. But the otherwise noteworthy similarity of the papilloma and polyoma viral pentamers, and the requirement in full-size VLPs for interpentamer contacts that are sufficiently versatile to allow either five or six pentamers to be coordinated around any given pentamer (rather than just five in the small VLPs), led us to speculate that in papilloma virions and full-size VLPs, the C-terminal subunit arms might indeed invade neighboring pentamers, in a pattern resembling that found in murine polyoma and SV40.

The detection of an interpentamer disulfide bond in virions and full-size VLPs (Sapp *et al*., 1995; Fligge *et al*., 2001) is consistent with 'arm invasion', rather than simple contact, in papilloma virions. Moreover, mutagenesis experiments suggest that the disulfides in question link the C-terminal arm of one subunit with an outer loop of another (Li *et al*., 1998; Sapp *et al*., 1998). It is worth noting that in SV40 the invading C-terminal arm is anchored to a neighboring pentamer by a disulfide, and in murine polyomavirus, the invading arm is clamped in place by a disulfide within the invaded pentamer, thus reinforcing the interpentamer linkage (Stehle *et al*., 1996).

There are no disulfide bonds, either within pentamers or between them, in the small VLPs, and the only way to account for the detection of such bonds is to postulate that the C-terminal arms fold differently in small VLPs than in virions. By fitting the HPV16 L1 structure into image reconstructions of BPV1 from electron cryomicroscopy (cryoEM) at 9 Å resolution, we have been able to verify

© European Molecular Biology Organization

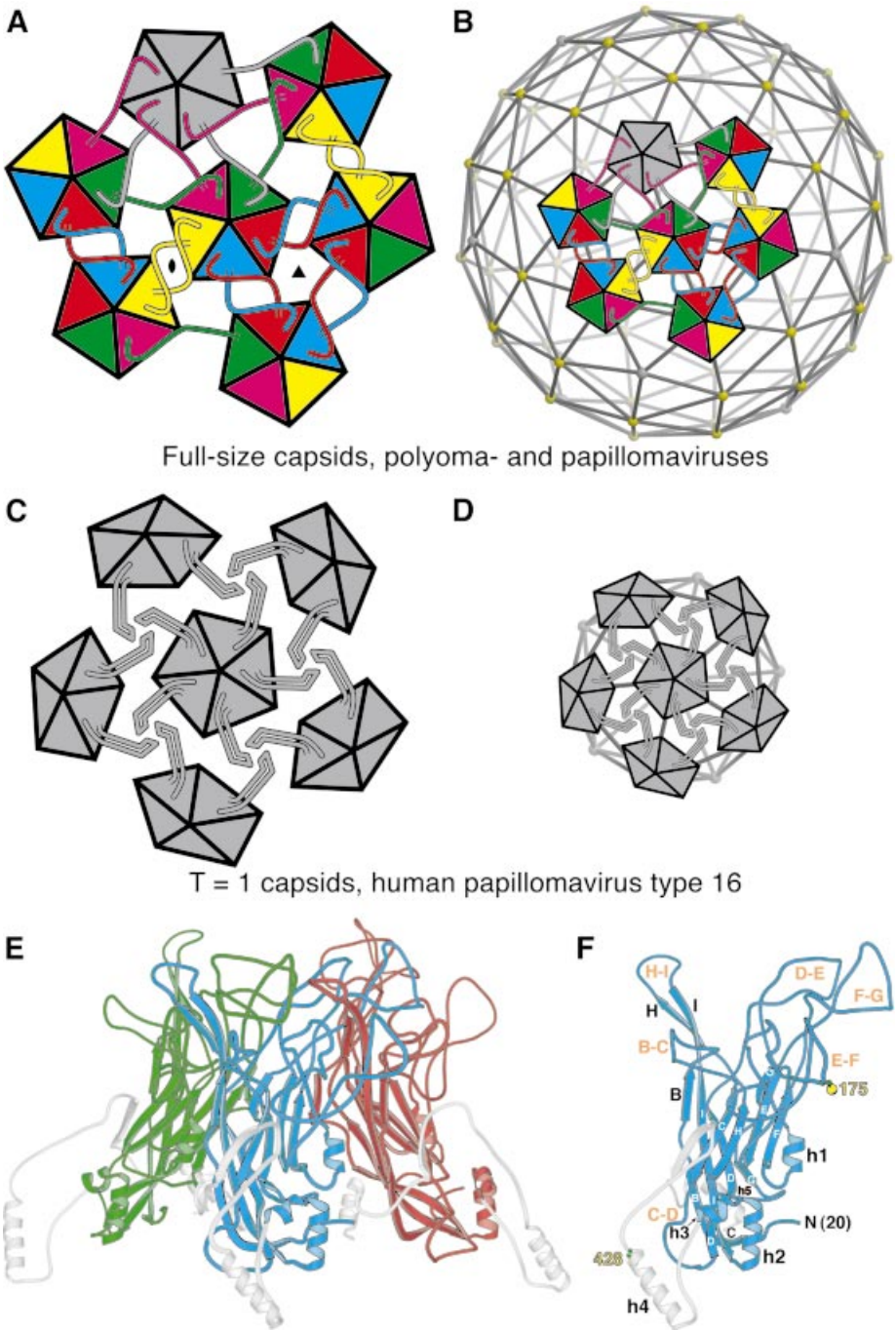

Fig. 1. Modes of interaction between pentamers in various papovavirus subunit assemblies. (**A**) Pentamer contacts in SV40 (Liddington *et al*., 1991), polyomavirus, and in full-size (T = 7) papillomavirus particles, as shown in this paper. The subunit 'body', a jellyroll β-barrel (see E and F), is represented by a triangle that forms one-fifth of a pentamer. (**B**) Diagram in (A) superimposed on the full T = 7 lattice. (**C**) Pentamer contacts in small (T = 1) VLPs derived from HPV L1 (Chen *et al*., 2000). Note that the C-terminal arms fold back into the subunit to which they belong, rather than reaching out into a neighboring pentamer. (**D**) Diagram in (C) superimposed on a T = 1 lattice. (B) and (D) are drawn to the same scale. (**E**) Ribbon diagram of an HPV16 L1 pentamer, in the conformation found in small VLPs (Chen *et al*., 2001). Three subunits are shown in green, blue and red. The C-terminal arms are in gray, to indicate that these portions of the subunit rearrange when pentamers assemble into virions or into full-sized capsids. (**F**) Ribbon diagram of a single subunit, approximately in the orientation of the blue subunit in (E). Secondary structural elements are labeled, with letters B–I for β-strands and h1–h5 for the five α-helices. The N- and C-termini of the ordered structure are marked N(20) and C, respectively. Residue 20 is the first ordered residue in the small VLP structure. Approximately the last 30 residues of the subunit (the exact number varies among types) are disordered and extend into the interior of the particle. Loops between strands are labeled B–C, C–D, etc. The two cysteines that participate in interpentamer disulfide bonding in the virion or in virion-sized particles are shown in yellow, together with their residue numbers, 175 and 428.

this postulate directly. The density shows that residues 402–445 must be rebuilt in order to fit the EM map. We have been able to construct a relatively detailed model for the course of these residues, thanks to two constraints: (i) the presence of a disulfide bond linking residue 428 in the rebuilt arm with residue 175 in the E–F loop; and

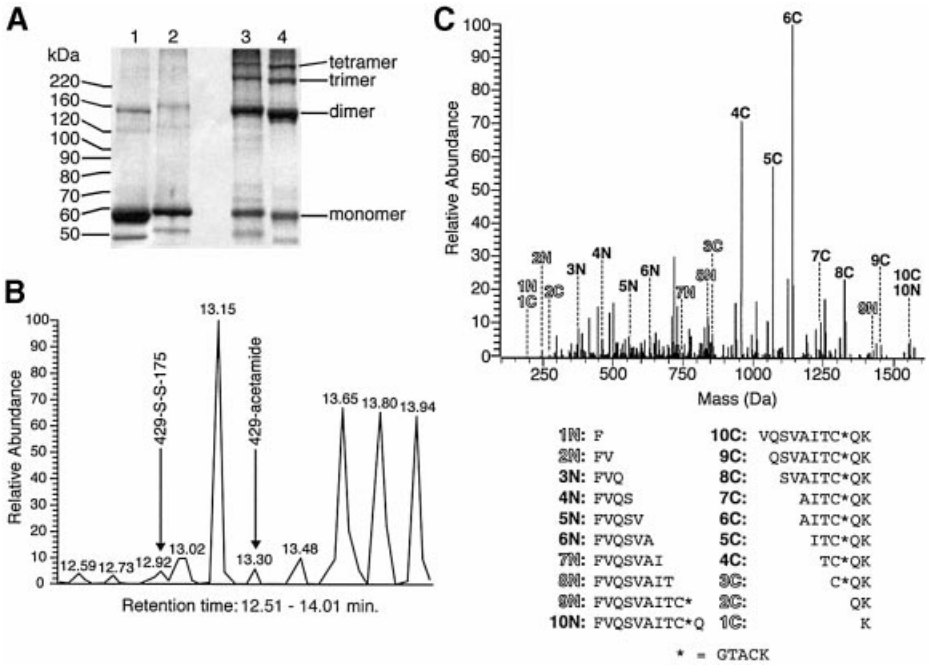

Fig. 2. Detection of intermolecular disulfide bonds in HPV. (**A**) SDS–PAGE of HPV capsids. Samples of HPV16 (lanes 1 and 4) and HPV18 (lanes 2 and 3) run under reducing (lanes 1 and 2) and non-reducing (lanes 3 and 4) conditions. In the reduced samples, the protein migrates as a monomer. In the non-reduced samples, most of the protein migrates as a dimer, but trimer, monomer and higher order bands are also visible. In all samples, the protein was denatured with 8 M urea and 1% SDS, as heating promotes the lysis and exchange of disulfide bonds. (**B**) HPLC trace of the tryptic peptides resulting from digestion of full-size (T = 7) HPV18 VLPs. (**C**) Sequencing mass spectrum of the HPV18 L1 tryptic peptides containing Cys175 and Cys429 linked by a disulfide bond. The lower panel shows the list of all possible peptide fragments that would be consistent with the peptide FVQSVAITCQK cross-linked to the peptide GTACK by a disulfide bridge. N-terminal fragments are labeled N1–10 and C-terminal fragments C1–10. Peptide fragments for which a peak is observed at the expected mass in the spectrum are labeled in solid black text; fragments for which no peak is observed in the spectrum are labeled in outlined text.

(ii) the clear identification of a feature in the EM image reconstruction corresponding to residues 462–474, a short α-helix on the inward-facing surface of the subunit. We have confirmed the validity of the former constraint by mass spectrometry of peptides from VLPs of HPV18. The EM density feature that provides the latter constraint corresponds precisely to the same α-helix in the small VLP crystal structure, but in the virion it is due to the invading arm of a subunit in a neighboring pentamer, rather than the reinserted arm of the subunit against which the helix packs (Figure 1).

VLPs are the immunogens in candidate HPV vaccines, currently in clinical trials. The detailed model for the virion presented here has significant implications for vaccine design, as most of the rebuilt portion of the C-terminal arm is exposed on the surface of the virion, near the outer rim of the L1 pentamer, and is therefore expected to be highly antigenic. In addition, the structure we find resolves the apparent discrepancy between basic assembly mechanisms in the otherwise structurally related polyoma- and papillomaviruses.

Results

Papillomavirus capsids contain an intermolecular disulfide bond

There is good biochemical evidence that intercapsomeric disulfide bonds stabilize papillomavirus capsids (Sapp *et al*., 1995; Fligge *et al*., 2001). Mutagenesis experiments performed by Li *et al*. (1998) and Sapp *et al*. (1998) have identified Cys175 and Cys428 (in HPV16), both of which are conserved among different types, as the likely partners in the disulfide cross-link. Cys175 and Cys428 are the only solvent-exposed cysteines in HPV16 L1 (Figure 1F). We have verified the assignment of Cys175 and Cys428 as follows. Samples of recombinant HPV16 L1 capsids purified from *Escherichia coli* (see Materials and methods) and HPV18 L1 capsids provided by Yang Wang (Merck Research Laboratories, Westpoint, PA) were analyzed by SDS–PAGE under both reducing and non-reducing conditions (Figure 2A), taking precautions to minimize disulfide exchange due to the exposure of reactive cysteines as the protein was denatured (see Materials and methods). In the non-reduced sample, ~70% of the L1 migrated in the 7% polyacrylamide gel as a 150 kDa species, 15% migrated as a 230 kDa species, 10% migrated as one or two larger species, and <5% migrated at 56 kDa, the expected molecular weight for the monomer. In the reduced sample (0.15 M β-mercaptoethanol), >95% of the protein migrated as a monomer (Figure 2A). The 150 kDa band is between the expected positions for a dimer or a trimer of L1. We propose that it is due to a dimer and that the band at 230 kDa is due to a trimer. These assignments are consistent with our atomic model, which contains 90 dimers and 60 trimers of L1 (see below). We note that disulfide-linked species, with internal cross-links, will tend to migrate more slowly in SDS–PAGE than a non-cross-linked molecule and will therefore tend to appear larger than their actual size. The minor band, labeled as a tetramer in Figure 2A, is most likely to have come from

disulfide formation or exchange during denaturation. HPV16 and HPV18 have 12 and 15 cysteine residues, respectively, and some cysteines may have formed disulfides or exchanged into existing disulfides before being blocked with iodoacetamide (see below).

The dimer band from HPV18 capsids was excised from the polyacrylamide gels, digested 'in-gel' with trypsin (Shevchenko *et al*., 1996) and analyzed by liquid chromatography–tandem mass spectrometry (LC-MS/MS; Materials and methods). We used HPV18 because of a favorable disposition of tryptic sites around Cys175. Unlike most HPV types, HPV18 has a very short (five-residue) tryptic peptide around Cys175. The corresponding HPV16 peptide could not be detected, due to its size (46 residues, including two cysteines) and hydrophobicity. LC-MS/MS of typsin-treated HPV18 L1 dimer revealed a species corresponding to the 11 residue tryptic fragment around the Cys428 equivalent of HPV18, Cys429 (using the residue numbering from entry VL1_HPV18 in the SwissProt database), with a 476 Da adduct on Cys429. This adduct is much larger than an iodoacetamide adduct (58 Da), but it corresponds exactly to the size of the five-residue tryptic peptide containing Cys175 (Figure 2C). The species identified was the peptide FVQSVAITCQK linked to the peptide GTACK by a disulfide bond. The same HPV18 sample was also analyzed by LC-MS/MS after reduction of the disulfide bonds with 20 mM dithiothreitol (DTT). The cross-linked peptide was found to be absent from this sample, and the peptide FVQSVAITCQK with an unmodified cysteine was found instead. This peptide and the corresponding peptide in HPV16 were also found with an iodoacetamide adduct in both reduced and unreduced samples, an observation consistent with the detection of some L1 monomers in SDS–PAGE (Figure 2A) and with other experiments, which show that recombinant HPV33 L1 capsids are less extensively cross-linked than native virions (Sapp *et al*., 1998; Fligge *et al*., 2001). LC-MS/MS does not allow a quantitative estimation of the extent of disulfide cross-linking in the capsids, but the HPLC trace does support this finding (Figure 2B). In any case, it provides compelling evidence for a disulfide bond in HPV18 between residues Cys175 and Cys429, which are equivalent to Cys175 and Cys428 in the HPV16 L1 crystal structure. Furthermore, our SDS–PAGE results suggest that at least 95% of the L1 molecules in our HPV16 and HPV18 capsids are in fact cross-linked (Figure 2A).

One other disulfide-linked tryptic peptide was identified from the LC-MS/MS analysis of HPV18: the tryptic peptide containing Cys185 (SRPLSQGDCPPLELK) cross-linked to the peptide GTACK containing Cys175. The sulfur atoms of these cysteine residues in the crystal structure of HPV16 L1 are <15 Å apart, and Cys175 is located on a flexible loop. The disulfide bond between Cys175 and Cys185 was probably formed very rapidly on denaturation, before iodoacetamide could block the cysteines.

Fit of the atomic coordinates of HPV16 L1 in EM maps of BPV1

The atomic coordinates of HPV16 L1 were positioned in EM image reconstructions of bovine papillomavirus at 9 and 13 Å resolution, as described in the Materials and methods. Residues 402–446 in the C-terminal arm were omitted from the fitting procedure (see below). The quality of the initial fit was excellent (Figure 3A–C). Certain features in the EM electron density maps allowed us to define unambiguously the positions of both pentavalent and hexavalent pentamers. In particular, the rotation of the pentamers about the 5-fold axis of each pentamer was determined by aligning protrusions near the crown of the pentamer (the B–C and E–F loops in the atomic coordinates; see Figure 1F) with similar protrusions in the EM maps. The protrusions lend papillomavirus capsomeres their star-shaped appearance. Although the fit of these two loops in the EM maps was excellent (Figure 3A), the features in the EM maps were larger than could be accounted for by the loops alone, leaving an unfilled volume of $\sim 7 \times 10^3$ Å at the tips of the protrusions. This was by far the largest single volume in the protein shell left unoccupied by the atomic coordinates. Another feature in the EM maps, which helped validate the fit, was a bulge on the inside of the protein shell. Indeed, the shape of the bulge was perfectly accounted for by the C-terminal portion of helix h5 (residues 462–473; see Figure 3B). The general agreement of the fits using the EM maps at 9 and 13 Å resolution was very good. The features discussed above were present in both maps, although they were slightly more pronounced in the 9 Å map, as might be expected. The smaller number of particles used in calculating the 9 Å map (209, versus 707 for the 13 Å map) made it substantially noisier than the 13 Å map, and therefore higher resolution did not yield any significant additional information for positioning the atomic coordinates.

The only part of the atomic coordinates that did not match the EM maps was the segment containing residues 402–446. This part of the C-terminal arm includes helix h4 (residues 414–429). It forms all the interpentamer contacts in the crystal structure of the HPV16 L1 T = 1 particles (Chen *et al*., 2000). In the conformation it adopts in small VLPs, it would protrude from the electron density and overlap with another pentamer on the virion surface. Residues 402–446 were therefore rebuilt in a new conformation into the unoccupied volume of the EM maps described above (at the tip of the protrusions near the crown of the pentamer). The conformation was further constrained by the requirement that Cys428, in the rebuilt segment of the C-terminal arm, form an intermolecular disulfide bond with Cys175, in the E–F loop. In order to satisfy the above constraints, an 'invading arm' model was adopted (Figure 3D), with interpentamer contacts similar to those in SV40 and murine polyomavirus (see below).

The 'invading arm' model

In an invading arm model, the C-terminal arms form the principal interpentamer contacts, by extending away from their subunit of origin and invading a subunit in an adjacent pentamer. Each pentamer receives five invading arms, one from each of five adjacent pentamers, and donates five arms to surrounding pentamers (Figure 1A). This pattern is identical to that observed in both SV40 (Liddington *et al*., 1991) and murine polyomavirus (Stehle *et al*., 1994). Residues 403–413 act as a flexible hinge or adapter, bridging the gaps between donor and acceptor pentamers, at the base of the protein shell (Figure 3D). The

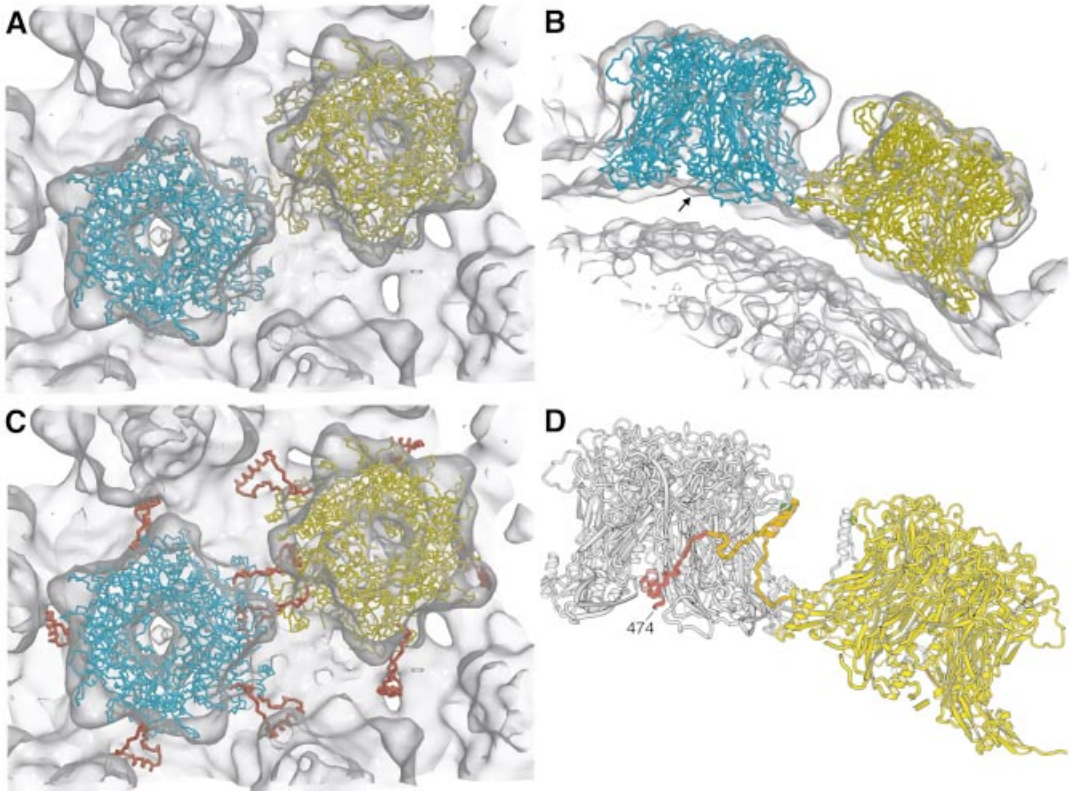

Fig. 3. Fit of the atomic coordinates of HPV16 L1 into the 13 Å EM map of BPV1. (**A**) View perpendicular to the surface. Two L1 pentamers were inserted into the density: one on a 5-fold symmetry axis (in blue) and one on an adjacent pseudo 6-fold (in yellow). Residues 402–446 were omitted from the fit and are not shown. (**B**) As in (A), but view rotated ~90° to show the outer surface of the capsid. A black arrow indicates the bulge in the EM maps on the inside of the protein shell, which helped validate the fit. This feature is accounted for by the C-terminal portion of helix h5 (residues 462–473). This figure, and Figures 4B, C and 5C, were drawn with BobScript (Kraulis, 1991; Esnouf, 1997) and Raster 3D (Merritt and Bacon, 1997). (**C**) Same view as in (A), but with the C-terminal arm (red) in the same conformation as in the HPV16 L1 crystal structure. (**D**) The 'invading arm' model. The invading arm (orange and red) from an L1 pentamer lying on a pseudo 6-fold symmetry axis (yellow) is shown invading an adjacent pentamer lying on a 5-fold symmetry axis (gray). The orange part of the molecule has been rebuilt to fit the electron density, in a position that would allow Cys175 and Cys428 (green spheres) to be linked by an intermolecular disulfide bridge. Drawn with MolScript (Kraulis, 1991).

flexibility of this proline- and glycine-rich hinge region allows a given pentamer to donate C-terminal arms either in a pentavalent environment, where the pentamer lies on an icosahedral 5-fold axis, surrounded by five other pentamers, or in a hexavalent environment, where the pentamer is surrounded by six other pentamers. For hexavalent pentamers, the relative position of the receiving pentamer is different for each C-terminal arm, and the conformation of the hinge region of the C-terminal arm is different for each subunit.

In rebuilding the rest of the C-terminal arm (residues 414–446), we assumed that residues 414–429, helix h4 in the small VLP, would remain in an α-helical conformation. Conformational changes involving rigid-body reorientations of α-helices are far more common than changes involving loss of secondary structure. Indeed, the cryoEM density can nicely accommodate h4, so positioned that Cys428 at its C-terminus lies opposite Cys175 on the invaded subunit (Figure 3B and D). The C-terminus of h4 inserts into the space between loops B–C and E–F. Residues 430–446 can then extend around the circumference of the target pentamer, linking h4 with the C-terminal segment (Figures 3D and 4A). The entire rebuilt region of the C-terminal arm fits comfortably into the EM maps at 9 and 13 Å resolution (Figure 4B and C; Trus *et al*., 1997).

The segment containing residues 447–474 has the same conformation and secondary structure as in the crystal structure of HPV16 L1 (Chen *et al*., 2000), but it originates from an adjacent pentamer. Thus, residues 447–474 have been displaced from their pentamer of origin (in the small VLP) to an equivalent position in an adjacent pentamer (in the virion), where they adopt the same conformation, including β-strand J and helix h5 (Figures 1E, F and 3D). This segment inserts into the receiving pentamer at the interface between two L1 subunits and firmly links the invading and receiving pentamers (Figure 4A). It is capped by residues 402–414 of yet another C-terminal arm (Figure 4A), thereby blocking access of helix h5 and the C-terminus to the outer viral surface. This interaction may help prevent the C-terminal arm from peeling off and thus may enhance the overall stability of the virion.

Structure and function of the N-terminal arm

The model of the HPV16 capsid generated using the coordinates of L1 with the rebuilt C-terminal arm contains gaps, 20–30 Å in diameter, at three different locations between pentamers in the base of the protein shell. The EM maps also contain gaps in the surface, at two locations on or near the icosahedral 2-fold axes, but these gaps are

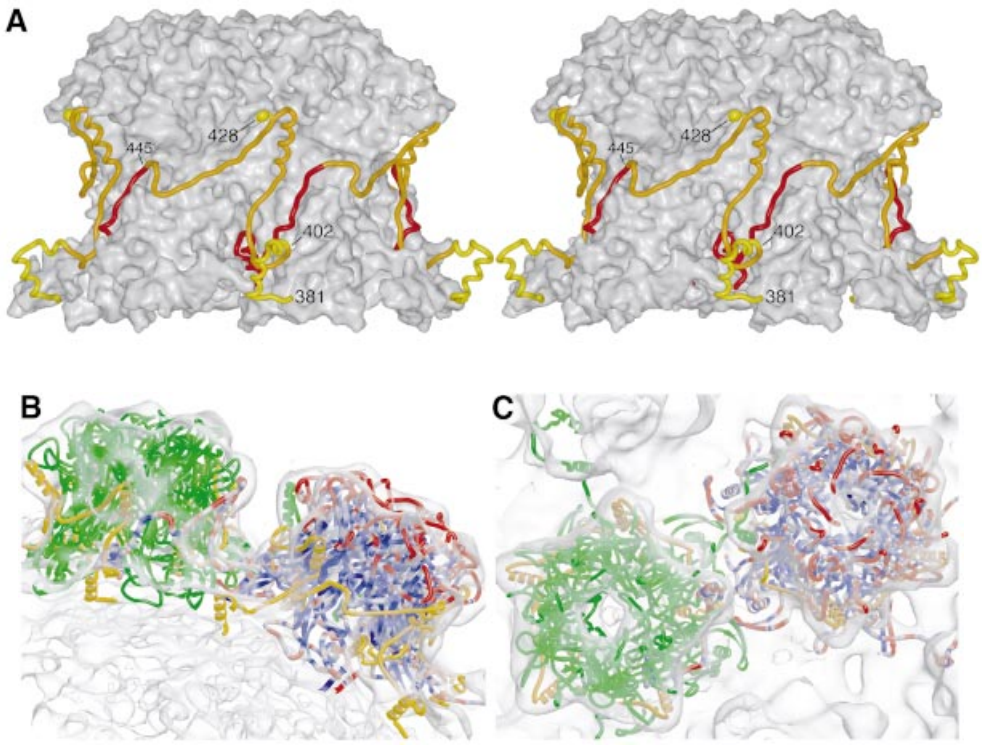

Fig. 4. (**A**) Stereo view of a HPV16 pentamer with rebuilt C-terminal arms from surrounding pentamers bound in the 'invading arm' conformation. The pentamer is shown as a gray surface, which was calculated from the atomic coordinates of a pentamer of HPV16 L1 lacking C-terminal arms. The C-terminal arms from surrounding donor subunits are shown as 'worms'. The coloring scheme is the same as in Figure 3C: the rebuilt portion of the C-terminal arm is orange, and the part of each arm that has the same conformation as in papilloma small VLPs is red. A short segment of each donor subunit leading into the C-terminal arm is shown in yellow. Drawn with SPOCK (Christopher, 1998) and Raster 3D (Merritt and Bacon, 1997). (**B**) Fit of the atomic model of HPV16 in the EM map. The 13 Å EM map of BPV1 is shown as a transparent surface. Two L1 pentamers are shown, one on a 5-fold symmetry axis (green) and one on the pseudo 6-fold. The latter is colored by sequence conservation among a set of 49 different HPV types (HPV Compendium 1997: http://hpv-web.lanl.gov). Highly variable positions are red, fully conserved positions are blue and positions of average variation are white. All the hypervariable regions lie on the outward-facing surface of the pentamer. The C-terminal arm has several exposed, hypervariable residues (red). In HPV16, these are residues 413, 421, 424, 431, 432, 443 and 444. Invading C-terminal arms originating from adjacent pentamers are shown in orange. (**C**) As in (B), but rotated ~90°. The view is looking down at the outer capsid surface, approximately parallel to the 5-fold.

only 10 × 20 Å. In the atomic model, the N-terminus (residue 20) of each L1 molecule points directly toward a gap in the surface. Part of the N-terminal arm (residues 9–20) is known to be essential for assembly of T = 7 capsids (Garcea *et al*., 1987; Chen *et al*., 2001), and residues 1–20 are conserved in all papillomaviruses, suggesting that they are indispensable. We therefore propose that the N-terminal arm is ordered in the virion (unlike in small VLPs, where it is disordered) and that it plugs gaps in the viral surface. The additional interpentamer contacts thus created may be important for capsid stability.

The conformation of the N-terminal arm in papillomavirus is unknown, and the EM maps do not allow us to infer it. In polyomaviruses, there is a β-hairpin, formed by the last 18 residues of the C-terminal arm, which plugs holes in the protein shell (Liddington *et al*., 1991). We have built a similar β-hairpin structure for the first 19 residues of HPV16 L1 (Figure 5C). In order to avoid clashes and effectively fill the holes in the shell, the β-hairpin was given three different orientations with respect to its subunit of origin: one for the β-hairpins nearest the icosahedral 2-fold axis, a second for the β-hairpins nearest the icosahedral 3-fold axis, and a third for all others. Near the 2-fold axis, the modeled β-hairpins lie roughly parallel to the viral surface (Figure 5C), on either side of a hole in the EM electron density on the 2-fold axis. Around the 3-fold axis, the modeled β-hairpins extend out towards the symmetry axis (Figure 5C), like a set of C-terminal β-hairpins in polyomaviruses, which also point straight towards the 3-fold axis. On the remaining subunits, we modeled the N-terminal β-hairpin so that it wraps snugly around helix h2 (residues 384–395).

The N-terminal hairpins can be positioned to fill most of the holes in the protein shell. In addition, some of the C–D loops (residues 82–96) may reach out and fill spaces left unoccupied even by the N-terminal segments. In our model, C–D loops in the same conformation as in the HPV16 L1 crystal structure form important contacts between subunits in pentavalent pentamers and adjacent subunits in hexavalent pentamers. The four C–D loops in the hexavalent pentamers that do not interact with pentavalent pentamers could move out to plug the remaining gaps in the protein shell. In doing so, these loops will form additional interpentameric contacts, in particular near the 3-fold icosahedral axis (Figure 5C). A flexible C–D loop may thus be another element that allows papillomaviruses to assemble into T = 7 icosahedral particles with a pentamer as the building block.

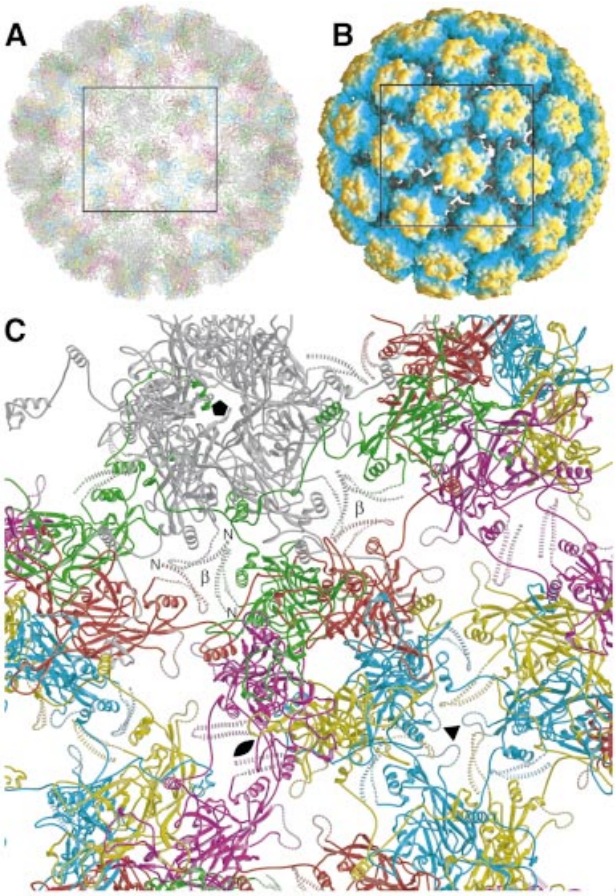

Fig. 5. Packing of capsomeres on the viral surface. The view is looking down a pseudo 6-fold symmetry axis (center), with the neighboring 5-fold above and to the left, the 2-fold below and to the left, and the 3-fold below and to the right. The symmetry axes are marked in standard notation. (**A**) View of the entire capsid, drawn as a Cα trace. The box indicates the area enlarged in (C). Drawn with O (Jones *et al*., 1991). (**B**) View of the molecular surface of the atomic model colored according to the distance from the center of the capsid. The box indicates the area enlarged in (C). Drawn with Grasp (Nicholls *et al*., 1991). (**C**) Close-up of the viral surface. The coloring scheme is the same as in (A): each of the six molecules in the icosahedral asymmetric unit is colored differently, with the subunits surrounding the 5-fold symmetry axis in gray. The most intimate intercapsomeric contacts are formed by the invading arms, but the N-terminal arm (residues 1–20) and the C–D loop (residues 82–96) are proposed to provide additional contacts and fill the gaps in the protein shell. These regions have lighter colors and dashed secondary structure, to indicate the more speculative nature of the modeling. The pseudo 3-fold axis, which is surrounded by the β-annulus formed by the N-terminal arms, is labeled 'β'. The N-termini of the β-hairpins forming the β-annulus are labeled 'N'.

Discussion

A common solution to the assembly puzzle for papovaviruses

The structure of SV40 first illustrated an elegant solution to the structural puzzle of how icosahedral particle assembly can proceed using 72 pentamers as building blocks (Liddington *et al*., 1991). From our atomic model, it appears that papillomaviruses have the same design. Long C-terminal arms of the L1 subunits extend into surrounding pentamers, where tight interactions anchor them firmly. An intermolecular disulfide bond further strengthens the attachment. The anchoring interactions are

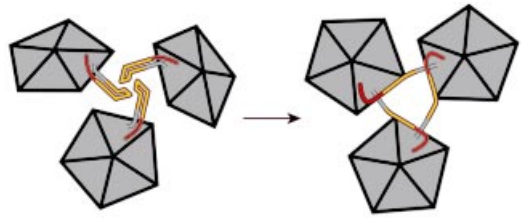

Fig. 6. Schematic representation of the conformational rearrangement required for the C-terminal arm to switch from the folded conformation, observed in the crystal structure of small VLPs (Chen *et al*., 2000), to the 'invading' conformation that we propose for full-size papillomavirus VLPs and virions.

the same for all subunits. Variability in the geometry of contacts in the virion resides almost entirely in the inherent flexibility of the first part of the C-terminal arm (residues 403–413), which allows it to extend out in the direction of the nearest neighboring pentamer. As in polyomaviruses (Liddington *et al*., 1991), the particular relative orientations of hexavalent and pentavalent pentamers lead to just three kinds of interpentamer interaction: a 3-fold cluster and two different 2-fold clusters. Each type of cluster is defined by the pattern of interchange of C-terminal arms. The 3-fold cluster includes one pentavalent pentamer and two hexavalent pentamers. The C-terminal arm invades the clockwise-neighboring pentamer. Thus, the pentamers in a 3-fold cluster exchange C-terminal arms in a cyclical fashion. Both kinds of 2-fold cluster contain two hexavalent pentamers. In one kind, pentamers related by the icosahedral 3-fold axis exchange C-terminal arms; in the other, pentamers related by an icosahedral 2-fold axis exchange arms (Figure 5C). Even the helical conformation of h4 corresponds to that of its counterpart, helix αC in SV40 and polyomavirus. The conformational rearrangement required for the C-terminal arm to switch from its folded conformation, observed in the crystal structure of small VLPs, to its 'invading' conformation is illustrated in Figure 6.

What determines whether the C-terminal arm adopts the folded, non-invading conformation of small VLPs or the invading conformation of full-size VLPs? Only pentamers of L1 truncated by 10 N-terminal residues (ΔN10L1) form small VLPs; even nine-residue truncations give full-sized (T = 7) particles (Chen *et al*., 2001). We have proposed a structural role for the N-terminal arm in the model presented here, but equally important appears to be the likelihood that residues 1–10 can block the C-terminal arm from returning to its pentamer of origin. In ΔN10L1 pentamers expressed in *E.coli*, the C-terminal arms fold back, with helix h5 docked into its site facing the interior hollow of the pentamer (Figure 1E and F; Chen *et al*., 2000). In small VLPs, the 'elbows' of these folded arms make the interpentamer contacts (Figure 1C and D). Although residues 1–10 are absent in the L1 crystal structure, our model suggests that they would be in the correct location to interfere with the C-terminal-arm foldback (Figure 5C). Thus, in pentamers of full-length L1 (or of L1 truncated by <10 residues), residues 1–10 probably prevent the C-terminal arms from folding back, leaving them available for interchange on assembly. It is possible that during assembly *in vivo*, the presence of appropriate chaperones further facilitates arm invasion.

Potential immunogenicity of the C-terminal arm

Much of the C-terminus of L1 is exposed on the viral surface in our model, suggesting that it might be immunogenic, or that it might participate in immunogenic epitopes. Most of the rebuilt portion of the C-terminal arm (residues 402–446), especially the loop at the crown of the pentamer containing residues 420–429, would be relatively accessible to antibodies or to B-cell receptors (Figures 4 and 5). While these elements are located between pentamer protrusions, simple molecular modeling reveals that an antigen-binding fragment (Fab) with dimensions of 40 × 50 × 75 Å can in fact be inserted between the L1 pentamers, in such a way that its variable loops can contact the rebuilt C-terminal arms and probably even reach to the N-terminal arms and the C–D loops of L1 (data not shown). There is also experimental evidence that antibodies can recognize epitopes located between pentamers of BPV (Booy *et al*., 1998). The extent to which the C-terminal arms (residues 402–446, but residues 420–429 in particular), the N-terminal arms (residues 1–20) and the C–D loops (residues 82–96) contribute to neutralizing epitopes will require direct experimental analysis.

Materials and methods

Expression, purification and in vitro assembly of HPV16 capsids

The pGEX-2T construct used to express L1, the major capsid protein of HPV16, was identical to the ΔN0 construct used by Chen *et al*. (2001). L1 is expressed as a glutathione *S*-transferase fusion protein, which, when cleaved by thrombin, produces a molecule spanning residues –3 to 505, where the residue numbers refer to entry U37217 in DDBJ/EMBL/GenBank. To facilitate cloning of the construct, three residues were added to the N-terminus with the sequence Gly-Ser-Gly, and residue 505 was mutated from Leu to Arg. HPV16 L1 was expressed in *E.coli* and purified to homogeneity based on the method developed by Chen *et al*. (2001). The protein was then extensively dialyzed against 50 mM sodium acetate pH 5.5, 1 M NaCl, 1 mM EDTA, to remove all traces of DTT, leading to assembly of T = 7 papillomavirus-like particles, which were visualized with a JEOL 100-CX electron microscope using carbon-coated copper grids stained with 2% (w/v) uranyl acetate (data not shown).

Recombinant T = 7 HPV18 particles were kindly provided by Yang Wang (Merck Research Laboratories). They were dialyzed against the same buffer used for HPV16 L1 assembly.

SDS–PAGE analysis of HPV capsids

The HPV16 and HPV18 particles were incubated at pH 9 for 1 h at 23°C to promote the formation of disulfide bonds. The pH was adjusted by adding 1 M Tris–HCl pH 9.0 to a final concentration of 0.2 M. Iodoacetamide was then added in a 50-fold molar excess of the cysteines in the protein, which corresponded to 67 mM, and the mixture was incubated for 1 h at 23°C to allow all exposed and reactive cysteines to be alkylated (Jocelyn, 1972). In addition, 1 mM EDTA was added to prevent any metal-catalyzed oxidation of free cysteines (Kirkpatrick and Maclaren, 1973). As a final precaution, the capsids were denatured in 8 M urea without heating, because heating to 90°C can lead to disulfide bond exchange (Kim and Kim, 2001). Solid urea was added to a final concentration of 8 M, followed by a 1 h incubation at 23°C. The denatured proteins were analyzed by SDS–PAGE under both reducing and non-reducing conditions, and stained with NOVEX Colloidal Blue (Invitrogen; Figure 2A).

LC-MS/MS analysis of disulfide-linked HPV dimers

The major band on the non-reducing SDS–PAGE gels of HPV16 and HPV18 corresponded to a dimer of L1 (Figure 2A). This band was excised and treated with trypsin (Shevchenko *et al*., 1996). The resulting peptides were eluted and analyzed on an LCQ DECA ion-trap mass spectrometer (ThermoFinnigan). The LC-MS/MS analysis (Figeys *et al*., 1998) consisted of three separation stages. First, the sample was applied to a microcapillary reversed-phase HPLC column and bound peptides were eluted with a gradient of acetonitrile. In the second stage, eluted peptides were ionized by electrospray ionization and analyzed by an ion-trap mass spectrometer, which was programmed to retain in its ion-trap chamber peptides in the 400–1700 Da range only. In the third stage, the remaining peptide(s) in the ion-trap chamber were subjected to a second mass spectrometry run, in which each peptide bond was stochastically broken by helium atom bombardment, thus allowing the amino acid sequence of each peptide to be determined. This process was facilitated by use of the Sequest algorithm (Eng *et al*., 1994), which matches acquired mass spectra to theoretical mass spectra generated from the known protein sequence.

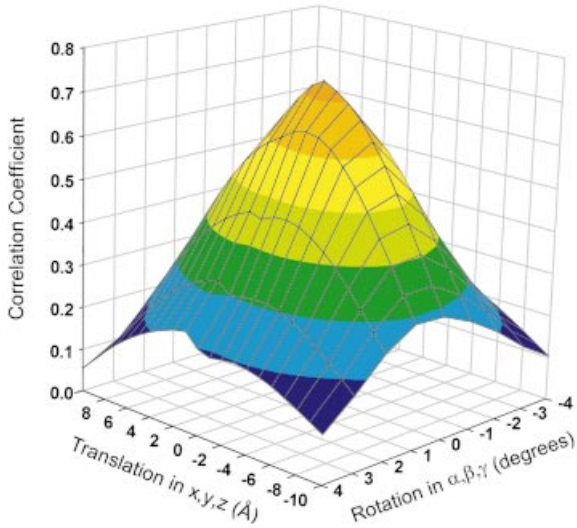

Fig. 7. The correlation coefficient between the atomic coordinates and the EM map as a function of translational and rotational shifts from the final fitted atomic model. The correlation coefficients were calculated with MAVE (Kleywegt and Read, 1997) using the 13 Å EM map of BPV1 and a 16 Å resolution map generated from the atomic model. The atomic model was incrementally translated in *x*,*y*,*z* and/or rotated in α,β,γ (Euler angles) away from its final fitted position to demonstrate that the fit corresponds to a peak in the correlation coefficient. The maximum correlation coefficient is 0.736, at zero translation and rotation from the fitted position. Without the C-terminal arm, the maximum correlation coefficient is 0.642. The fit was performed without the C-terminal arm (see text).

Fitting of the atomic coordinates of HPV16 into the EM maps of BPV1

The EM maps of the 13 and 9 Å resolution reconstructions of BPV (Trus *et al*., 1997) were prepared for fitting as follows. The maps were converted from PIF to CCP formats with the program EM2EM (Image Science Software GmbH) and then to the DSN6 format with MAPMAN (Kleywegt and Read, 1997). The coordinates of the center of the virus were determined by visual inspection, and the maps were translated with MAPROT (CCP4, 1994) so that the center of the virus was on the origin. The maps were rotated 90° about the *z*-axis with MAPROT to position them in the 2(*Z*)-3-5-(*X*)2 icosahedral convention. In this convention, a 2-fold symmetry axis lies on each of the *x*-, *y*- and *z*-axes, such that a 3-fold lies 21° from the *z*-axis and a 5-fold lies 58.3° from the *z*-axis in the positive quadrant of the *xz*-plane. Papillomaviruses have been reported to have a diameter of ~600 Å (Baker *et al*., 1991; Hagensee *et al*., 1994; Trus *et al*., 1997). The capsid diameter was 589 Å in the 13 Å map (contoured at 0.4 σ) and 606 Å in the 9 Å map (contoured at 1 σ); this apparent discrepancy was corrected after the initial fitting process (see below).

The fit was performed initially by visual inspection using the program O (Jones *et al*., 1991). The atomic coordinates of an L1 pentamer [Protein Data Bank (PDB) entry 1DZL; Chen *et al*., 2000] were placed in the 13 Å EM map so that the 5-fold symmetry axis of the pentamer coincided with an icosahedral 5-fold of the map. The pentamer was translated along its axis, and then rotated about it, to the position where it best fit the electron density. The pentamer was then moved to an adjacent hexavalent position (on a 'pseudo 6-fold' between the 2-, 3- and 5-fold symmetry axes) and adjusted similarly. The fitting procedure for the two pentamers was

performed in both the T = 7*laevo* and T = 7*dextro* lattice hands (the map was inverted with MAPMAN). As expected, the fit was substantially better in the T = 7*dextro* hand (Belnap *et al*., 1996), which was used for all subsequent fitting. When the two fitted pentamers were placed in the 9 Å map, it was apparent that the thickness of the protein shell was greater than could be accounted for by the atomic coordinates. The map was thus scaled down by 1.9% to match optimally the height of the pentamers, while avoiding clashes between their bases. The 13 Å map was scaled up 0.9% to match this diameter, and the pentamers were readjusted in the maps as described above. The diameter of the capsid in both scaled maps was 594 Å.

In the final stage of fitting, the positions of the pentavalent and hexavalent pentamers were optimized by real-space correlation refinement. Residues 402–446 of the crystal structure were omitted, as they clearly protruded from the electron density. These residues are responsible for all the interpentameric contacts in the T = 1 particle (Chen *et al*., 2000). The atomic coordinates of each pentamer were transformed into structure factors with SFALL (CCP4, 1994), and then into theoretical electron density maps at various resolution ranges with FFT (CCP4, 1994). A calculated map spanning 1000–16 Å resolution was selected as the best match for the 13 Å EM map, and a map spanning 1000–13 Å was selected for the 9 Å EM map. Masks encompassing each of the calculated maps were created with MAPMAN and improved with MAMA (Kleywegt and Jones, 1999). Finally, the masks were fitted into their respective EM maps by real-space correlation least-squares minimization with MAVE (Figure 7; Kleywegt and Read, 1997).

Atomic coordinates

The coordinates for our atomic model of HPV16 L1 were deposited in the PDB under the accession No. 1L0T.

Acknowledgements

Mass spectrometry experiments were performed by Dongmei Chen and Ross Tomaino at the Taplin Mass Spectrometry Facility at Harvard Medical School. We thank Yang Wang (Merck Research Laboratories) for providing us with the HPV18 particles used in this work; Xiaojiang Chen, for advice on expression and purification of HPV16; Robert Garcea, for useful comments and guidance, particularly on improving the homogeneity of our VLPs; and John Schiller, for his careful reading of the manuscript and useful comments. This work was supported by long-term fellowships to Y.M. from the European Molecular Biology Organization and the Human Frontier Science Program Organization, and by NIH grants CA-13202 and GM-62580 (to S.C.H.). 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. and 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. and Baker,T.S. (1996) Conserved features in papillomavirus and polyomavirus capsids. *J. Mol. Biol.*, **259**, 249–263.

Booy,F.P., Roden,R.B., Greenstone,H.L., Schiller,J.T. and Trus,B.L. (1998) Two antibodies that neutralize papillomavirus by different mechanisms show distinct binding patterns at 13 Å resolution. *J. Mol. Biol.*, **281**, 95–106.

Bosch,F.X. *et al*. (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**, 796–806.

CCP4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D*, **50**, 760–763.

Chen,X.S., Garcea,R.L., Goldberg,I., Casini,G. and 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,X.S., Casini,G., Harrison,S.C. and Garcea,R.L. (2001) Papillomavirus capsid protein expression in *Escherichia coli*: purification and assembly of HPV11 and HPV16 L1. *J. Mol. Biol.*, **307**, 173–182.

Christopher,J.A. (1998) *SPOCK: The Structural Properties Observation and Calculation Kit Program Manual*. The Center for Macromolecular Design, Texas A&M University, College Station, TX.

Eng,J., McCormack,A. and Yates,J.R. (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.*, **5**, 1579–1583.

Esnouf,R.M. (1997) An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. *J. Mol. Graph.*, **15**, 132–134.

Figeys,D., Gygi,S.P., Zhang,Y., Watts,J., Gu,M. and Aebersold,R. (1998) Electrophoresis combined with novel mass spectrometry techniques: powerful tools for the analysis of proteins and proteomes. *Electrophoresis*, **19**, 1811–1818.

Fligge,C., Schäfer,F., Selinka,H.C., Sapp,C. and Sapp,M. (2001) DNA-induced structural changes in the papillomavirus capsid. *J. Virol.*, **75**, 7727–7731.

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

Hagensee,M.E., Olson,N.H., Baker,T.S. and Galloway,D.A. (1994) Three-dimensional structure of vaccinia virus-produced human papillomavirus type 1 capsids. *J. Virol.*, **68**, 4503–4505.

Howley,P. (1996) *Papillomavirinae: The Viruses and Their Replication*. Raven Press, New York, NY.

Jocelyn,P.C. (1972) *Biochemistry of the SH Group*. Academic Press, New York, NY.

Jones,T.A., Zou,J.-Y., Cowan,S.W. and Kjeldgaard,M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A*, **47**, 110–119.

Kim,J.S. and Kim,H.J. (2001) Matrix-assisted laser desorption/ionization time-of-flight mass spectrometric observation of a peptide triplet induced by thermal cleavage of cystine. *Rapid Commun. Mass Spectrom.*, **15**, 2296–2300.

Kirkpatrick,A. and Maclaren,J.A. (1973) Thiol solutions—inhibition of autoxidation. *Anal. Biochem.*, **56**, 137–139.

Kleywegt,G.J. and Jones,T.A. (1999) Software for handling macromolecular envelopes. *Acta Crystallogr. D*, **55**, 941–944.

Kleywegt,G.J. and Read,R.J. (1997) Not your average density. *Structure*, **5**, 1557–1569.

Kraulis,P.J. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.*, **24**, 946–950.

Li,M., Beard,P., Estes,P.A., Lyon,M.K. and Garcea,R.L. (1998) Intercapsomeric disulfide bonds in papillomavirus assembly and disassembly. *J. Virol.*, **72**, 2160–2167.

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

Merritt,E.A. and Bacon,D.J. (1997) Raster 3D photorealistic molecular graphics. *Methods Enzymol.*, **277**, 505–524.

Nicholls,A., Sharp,K.A. and Honig,B. (1991) Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins*, **11**, 281–296.

Parkin,D.M., Bray,F.I. and Devesa,S.S. (2000) Cancer burden in the year 2000. The global picture. *Eur. J. Cancer*, **37**, S4–S66.

Sapp,M., Volpers,C., Müller,M. and Streeck,R.E. (1995) Organization of the major and minor capsid proteins in human papillomavirus type 33 virus-like particles. *J. Gen. Virol.*, **76**, 2407–2412.

Sapp,M., Fligge,C., Petzak,I., Harris,J.R. and Streeck,R.E. (1998) Papillomavirus assembly requires trimerization of the major capsid protein by disulfides between two highly conserved cysteines. *J. Virol.*, **72**, 6186–6189.

Shevchenko,A., Wilm,M., Vorm,O. and Mann,M. (1996) Mass spectrometric sequencing of proteins from silver-stained polyacrylamide gels. *Anal. Chem.*, **68**, 850–858.

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

Stehle,T., Gamblin,S.J., Yan,Y. and Harrison,S.C. (1996) The structure of simian virus 40 refined at 3.1 Å resolution. *Structure*, **4**, 165–182.

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

Received June 14, 2002; revised July 29, 2002; accepted July 30, 2002