Subunit interactions in bovine papillomavirus

Matthias Wolf[a], Robert L. Garcea[b], Nikolaus Grigorieff[c,1], and Stephen C. Harrison[a,d,1]

[a]Jack and Eileen Connors Structural Biology Laboratory, Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, 250 Longwood Avenue, Boston, MA 02115; [b]Department of Molecular, Cellular, and Developmental Biology, University of Colorado, Boulder, CO 80309; [c]Rosenstiel Basic Medical Research Center and Howard Hughes Medical Institute, Brandeis University, 415 South Street, Waltham, MA 02154; and [d]Howard Hughes Medical Institute, Harvard Medical School, 250 Longwood Avenue, Boston, MA 02115

Edited by John Kuriyan, University of California, Berkeley, Berkeley, CA, and approved February 22, 2010 (received for review December 22, 2009)

Papillomaviruses, members of a group of dsDNA viruses associated with epithelial growths and tumors, have compact capsids assembled from 72 pentamers of the protein L1. We have determined the structure of bovine papillomavirus by electron cryomicrosopy (cryoEM), at ∼3.6 Å resolution. The density map, obtained from single-particle analysis of ∼4,000 particle images, shows the trace of the L1 polypeptide chain and reveals how the N- and C-terminal "arms" of a subunit (extensions from its β-jelly-roll core) associate with a neighboring pentamer. Critical contacts come from the C-terminal arm, which loops out from the core of the subunit, forms contacts (including a disulfide) with two subunits in a neighboring pentamer, and reinserts into the pentamer from which it emanates. This trace corrects one feature of an earlier model. We discuss implications of the structure for virion assembly and for pathways of infectious viral entry. We suggest that it should be possible to obtain image reconstructions of comparable resolution from cryoEM images of asymmetric particles. From the work on papillomavirus described here, we estimate that such a reconstruction will require about 1.5 million images to achieve the same number of averaged asymmetric units; structural variability will increase this number substantially.

virus assembly | virus structure | icosahedral symmetry

The papillomaviruses are a group of small, DNA viruses associated with benign and malignant epithelial growths (1). Human papillomavirus (HPV) isolates fall into over 100 "types," initially classified serologically and now by DNA sequence variations that primarily reflect differences in the amino acid sequences of several prominent loops on the major structural protein L1. Infection with "high risk" types, e.g., HPV16 and HPV18, predispose to cervical cancer, prompting development of the recently introduced recombinant vaccines (2). The immunogens in these vaccines are virus-like particles (VLPs), obtained by expressing L1 in eukaryotic cells.

Seventy-two pentamers of L1 assemble to form the outer shell of a papillomavirus particle (3), which packages the closed circular dsDNA genome and associated nucleosomal histones. Twelve of the L1 pentamers lie on 5-fold axes; the other 60, at general positions in the icosahedral surface lattice, are 5-fold symmetric structures in a 6-coordinated surround (Fig. 1). Capsids of the polyomaviruses (in particular, those of SV40 and murine polyomavirus, studied crystallographically) have a closely related, 72-pentamer architecture (4–6). Crystal structures of recombinant L1 pentamers from HPV16 (7) and from three other HPV types (8) have shown as expected that they resemble the major capsid protein (VP1) pentamers from polyomaviruses. The core of the L1 subunit, like those of polyomavirus VP1, is a β-jelly roll domain, with N- and C-terminal extensions ("arms"). In the polyomavirus structures, the C-terminal arms mediate nearly all the interpentamer contacts. They "invade" a neighboring pentamer, forming part of its folded structure; bound Ca^{2+} ions stabilize the invading conformation of the arm, and interor intrapentamer disulfide bonds further lock the assembly together (5, 6).

When the HPV16 L1 structure was used to interpret a moderate-resolution electron cryomicrosopy (cryoEM) reconstruction of bovine papillomavirus (BPV), the best fit to the density suggested that the L1 C-terminal arms mediate interpentamer contacts with a pattern of invasion very similar to the one seen in SV40 and polyoma (9). That picture was consistent with protein-chemical evidence for an interpentamer disulfide between a cysteine in the C-terminal arm (BPV Cys426) and one within the jelly roll β-barrel (BPV Cys171) (10, 11). We have now obtained a near-atomic resolution cryoEM reconstruction of BPV that allows us to follow the full polypeptide chain (with the exception of a few, inward-projecting C-terminal residues). The long, C-terminal arms indeed interact with neighboring pentamers, forming disulfide bonds in the pattern previously suggested. The full model corrects a feature of the earlier one, however, and thereby alters one principal conclusion of the previous work. Unlike the invading arms of the polyomaviruses, which terminate within the target subunit, those of BPV L1 reinsert into the core of the pentamer from which they emerge. That is, they are essentially elaborate loops, which extend away from the jelly roll core to create the interpentamer contacts (including the disulfide bond), while retaining a C-terminal anchor within their pentamer of origin. The L1 C-terminal arms in the recombinant HPV16 L1 crystal structure also reinsert in this way. Our map thus shows that this reinsertion is a conserved property of the L1 pentamer rather than (as proposed in connection with the earlier cryoEM fit) a specific consequence of truncations needed to obtain the HPV16 L1 crystals.

The complete polypeptide chain trace shows that the N-terminal extensions (about 20 residues) also participate in contacts among pentamers. These interactions are incompatible with formation of the smaller T = 1 shell (12 pentamers) that is seen in the crystals of N-terminally truncated L1 (7). They appear to be one of the switching elements that govern accurate assembly of a 72-pentamer capsid. Comparison of the BPV and HPV16 L1 structures and sequences allows us to identify a number of conserved motifs, including an additional disulfide bond in a subset of papillomavirus types.

Results

Structure Determination. We have obtained a three-dimensional density map of BPV type 1 by single-particle reconstruction from electron micrographs of individual virus particles in vitreous ice. Images of about 4,000 particles from 49 different films

Author contributions: M.W., N.G., and S.C.H. designed research; M.W. performed research; R.L.G. contributed new reagents/analytic tools; M.W., N.G., and S.C.H. analyzed data; and M.W., R.L.G., N.G., and S.C.H. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Data deposition: Coordinates and maps have been submitted to EMDataBank-Research Collaboratory for Structural Bioinformatics with the deposition numbers EMD-5155 for the icosahedrally symmetric final reconstruction, EMD-5156 for the NCS-averaged subunit density, and 3IYJ for the coordinates.

[1]To whom correspondence may be addressed. E-mail: niko@brandeis.edu or harrison@crystal.harvard.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0914604107/DCSupplemental.

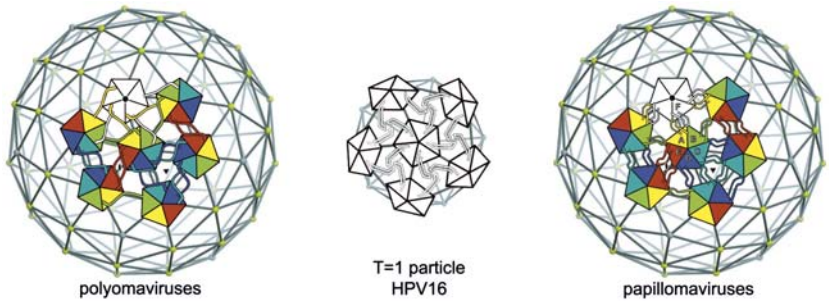


Fig. 1. Diagrams of pentamer packing and C-terminal arm interchange in polyoma and papillomaviruses. In the full-size capsids of viruses from both groups, 72 pentamers of the major capsid protein (called VP1 in the polyomaviruses and L1 in the papillomaviruses) are arrayed at the vertices of a T = 7 d icosahedral lattice. Twelve pentamers (white) are centered on 5-fold symmetry axes, whereas the 60 others (6 of which are shown, colored) are at 6-coordinated positions. In a 12-pentamer (T = 1) icosahedral particle, obtained with N-terminally truncated L1 from HPV16, all 12 pentamers are on icosahedral 5-fold symmetry axes. *Left*: In the polyomaviruses, the VP1 C-terminal arm "invades" a neighboring pentamer, in the pattern shown, and "inserts" into its structure. *Right*: In the papillomaviruses, the L1 C-terminal arm invades a neighboring pentamer but turns around and inserts back into its pentamer of origin. The T = 1 shell has arm–arm interactions distinct from those in the virion.

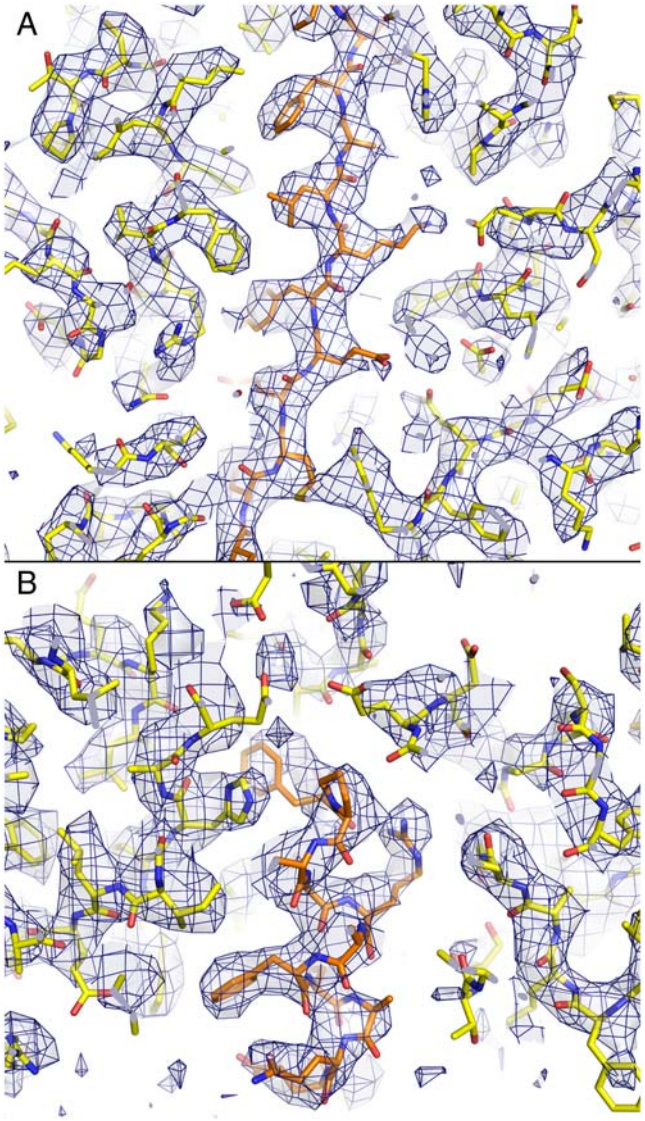


Fig. 2. Portions of the BPV cryoEM density map, in the 6-fold NCS-averaged region, with the refined molecular model superposed. The map is contoured at 3.2σ above the average density within the L1 β-barrel core. Side-chain densities in an α-helix (*A*) and a β-strand (*B*) lead to clear assignment of rotamers in many cases.

contributed to the reconstruction. The definition of features in the maps corresponds to an experimental x-ray crystallographic density map at about 3.5 Å resolution (Fig. 2). The data were acquired as described in *Materials and Methods*. We could follow nearly all the L1 chain, including previously undescribed segments, by using the icosahedrally symmetrized map. Additional 6-fold "noncapsid symmetry" (NCS) averaging, superposing the core domains of the six subunits in the icosahedral asymmetric unit, showed detail at the level of amino acid side chains. An atomic model of the icosahedral asymmetric unit (six L1 chains) has a conventional R factor, after energy-minimization refinement, of 0.385 (0.570 in the shell at 3.6 Å resolution), calculated against "Fobs" computed from the final icosahedral map (not NCS-averaged) by embedding it in a P1 cell (see *Materials and Methods*).

Interpentamer Contacts: C-Terminal Arms. Cross-bridges formed by the long C-terminal extensions (residues 404–495) of the core β-jelly roll link the L1 pentamers, in a pattern of interchange shown in Fig. 1 *Right*. The polypeptide chain emerges from the jelly-roll domain at its base and extends toward the neighboring pentamer. A segment with a conserved LxDtYR(Y/F) amino acid sequence (residues 412–418 in BPV1) augments (at the B-strand edge) the sheet with strands B, I, D, and G of one subunit in the target pentamer, and the conserved cysteine a few residues C-terminal to this segment (Cys426) forms a disulfide bond with a similarly conserved cysteine (Cys171) in the elaborate EF loop of a second subunit in the target pentamer. The interactions of residues in the LxDxYR(Y/F) motif account for their conservation and for the conservation of their partners: Asp414 has a salt bridge with Arg40; Tyr416 packs against Leu42; Arg417 has a salt bridge with Asp193 and packs against Trp444 (Fig. 3).

The charged and proline-rich segment immediately C-terminal to Cys426 returns the polypeptide chain to the core of the subunit from which it originated. This fully extended, 11-residue segment bridges the gap between pentamers, with essentially no contacts. It reinserts into the "parent" subunit as a strand (residues 444–449), which augments the sheet with strands C, H, E, and F at the C-strand edge. The tenuous density for the bridging connection was not evident in the lower-resolution map used to build the earlier model (9), and a link was instead suggested to residue 444 in the first subunit of the target pentamer. Except for this connection, which assigned residues from 438 to the C terminus to the wrong polypeptide chain (and the wrong pentamer), and for some conformational details of the segment between 405 and 427, the previous model (9) was correct. In particular, the disulfide was properly identified. The C-terminal arm is thus effectively a large loop: Residues 401–437 project away from the

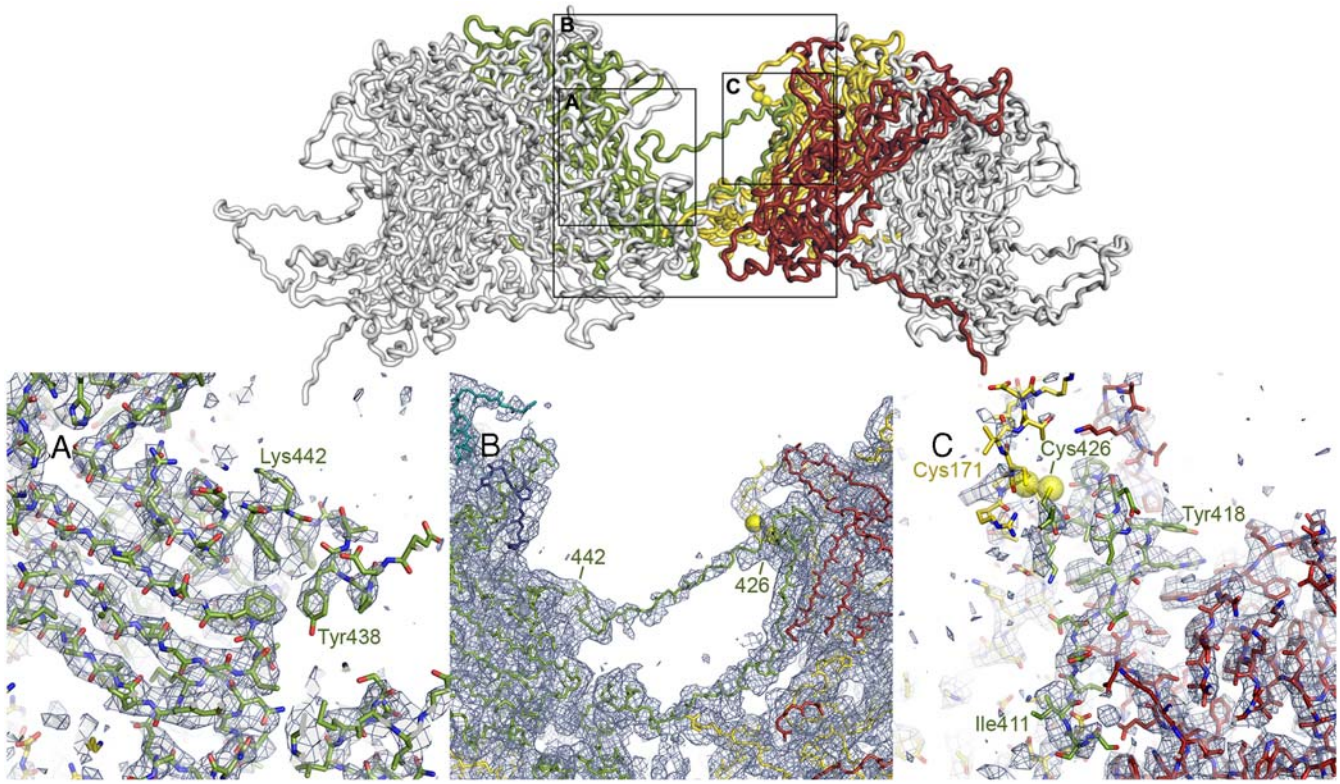


Fig. 3. The L1 C-terminal arm. *Top*: The interface between two 6-coordinated pentamers. Polypeptide chains are in "worm" representation. The arm of the "green" subunit in the left-hand pentamer (the other subunits are in gray, for clarity) invades the right-hand pentamer, where it augments a β-sheet in the red subunit, forms a disulfide bond (yellow spheres) with the yellow subunit, and extends back to reinsert into its pentamer of origin. The N-to-C direction of the green arm is left-to-right at the "lower" part of the interface shown and right-to-left in the contact-free segment that retraverses the interpentamer gap. The outlined rectangles indicate regions shown in the lower part of the figure. *Bottom*: Details of the density map to show unambiguous tracing of the L1 arm. In (*A*) and (*C*), the 6-fold NCS-averaged map is contoured at 3.0σ above mean density. In (*B*), a region in which the NCS symmetry does not hold, the map is the final icosahedrally symmetrized reconstruction, contoured at 2.5σ. (*A*) Reinsertion of the arm into the left-hand pentamer. The PYAGFKF sequence (residues 437–443) where the arm reinserts can be assigned unambiguously, as can residues following it. (*B*) Connection between the disulfide (Cys426) and the PYAGF segment in (*A*). The sequence in this bridge is proline-rich in most papillomaviruses. (*C*) Invasion of the red subunit of the target pentamer (augmentation of the BIDG sheet) and disulfide bond between Cys426 in the arm and Cys171 in the yellow subunit. Assignment of sequence to the augmenting (green) strand [LxDxYR(Y/F) motif, from 412–418] is unambiguous. Cys171 is in a surface loop, less rigid than the core β-jelly roll, and the density of the yellow chain in the vicinity of Cys171 is correspondingly weaker than the core density of the red subunit.

subunit, presenting Cys426 to Cys171 on a subunit of the neighboring pentamer, whereas residues 438–495 reinsert to hold the terminus of the loop within the pentamer from which it emerges.

The C-terminal part of the arm lies within the conical hollow around the pentamer axis (Fig. 3*A*). A short helix (residues 461–470) is followed by a strand (residues 478–484), which augments the G-strand edge of the BIDG sheet in the neighboring subunit (in a clockwise relationship, looking along the pentamer axis from outside the virion). A cysteine at position 473 pairs in a disulfide with a cysteine at neighboring-subunit position 22. The disulfide, for which there is clear density, thus links the C-terminal part of the L1 polypeptide chain with the N-terminal part of the adjacent chain. The two residues at the corresponding positions are both cysteines in several other papillomavirus types (see the HPV sequence database: http://hpv-web.lanl.gov/), but most types lack this disulfide. The disulfide is clearly not essential for anchoring the C-terminal segment, however, as there are also extensive noncovalent interactions. We can trace about ten residues following Cys473 in the polypeptide chain: They extend into a strand that augments the BIDG sheet of the neighboring subunit at the G-strand edge. The last ten residues of the chain, of which six are lysine, appear to project inward, probably contacting the minichromosome.

Conformational Variability in the C-Terminal Arm. In the assembled virion, the projecting portion of the C-terminal arm must adopt 6 distinct conformations, to adapt to the 6 distinct packing relationships within an icosahedral asymmetric unit. Fig. 4 shows the range of variation within the BPV structure. The distal part of the loop, which contributes the augmenting strand J to the BIDG sheet in one subunit of the target pentamer and Cys426 to a disulfide bond with Cys171 of another subunit, has a uniform structure, determined by those interactions, and conformational variability is restricted to the outgoing and returning segments, residues 403–411 and 427–435 (Fig. 4*A*). In the T = 1 shell that assembles at low pH from N-terminally truncated HPV16 L1 pentamers (the small VLP), the projecting C-terminal arm also mediates interpentamer contacts, but the segment that would form strand J and the disulfide coils up as an α-helix (α4) instead (7).

Interpentamer Contacts: N-Terminal Arms. The much shorter N-terminal extension (residues 1–20) also participates in a set of interpentamer contacts, largely with other N-terminal arms. One such interaction pairs the arms of chains A and C from two 6-coordinated pentamers related by an icosahedral 5-fold axis (yellow and light blue in Fig. 5). A second is a cyclic set of contacts among the three D-subunit arms related by a 3-fold axis (dark blue in Fig. 5). The third pairs the arm of one E subunit with the CD loop of another, 2-fold-related E subunit and with a short segment of its C-terminal arm (red in Fig. 5). The tip of the E arm also contributes to an A–C arm contact. Arms B and F are disordered up to residue 15. Despite the diversity of these

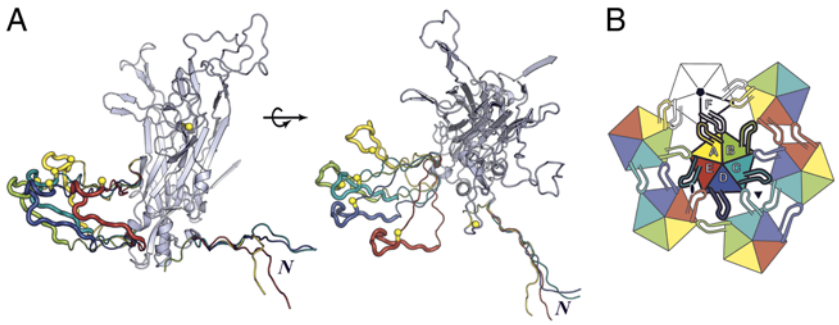


Fig. 4. Conformations of the BPV L1 arm. (*A*) Subunits A–E, superposed so that their core β-jelly rolls coincide (only one is shown, in gray), with N-terminal arms and extended part of C-terminal arms in color. The core of the subunit is in standard ribbon representation; the arms are in worm representation, in colors corresponding to those in the diagram in *B* and in Figs. 1,3, and 5. The gray arm, which has the same conformation as the yellow arm, is not shown. The parts of the C-terminal arms that contact subunits of the neighboring pentamer are shown as thicker worms; the conformation of these segments is the same for all subunits, because it is determined by conserved interactions (including the disulfide) with the target subunits. *Yellow balls*: The cysteines of the interpentamer disulfide. The N-terminal arms all have the same conformation C-terminal to residue 14 (the position at which the superposed arms splay apart). The two parts of *A* are related by a 90° rotation, as shown. (*B*) The icosahedral asymmetric unit, viewed from outside the particle, showing the pattern of C-terminal arm contacts.

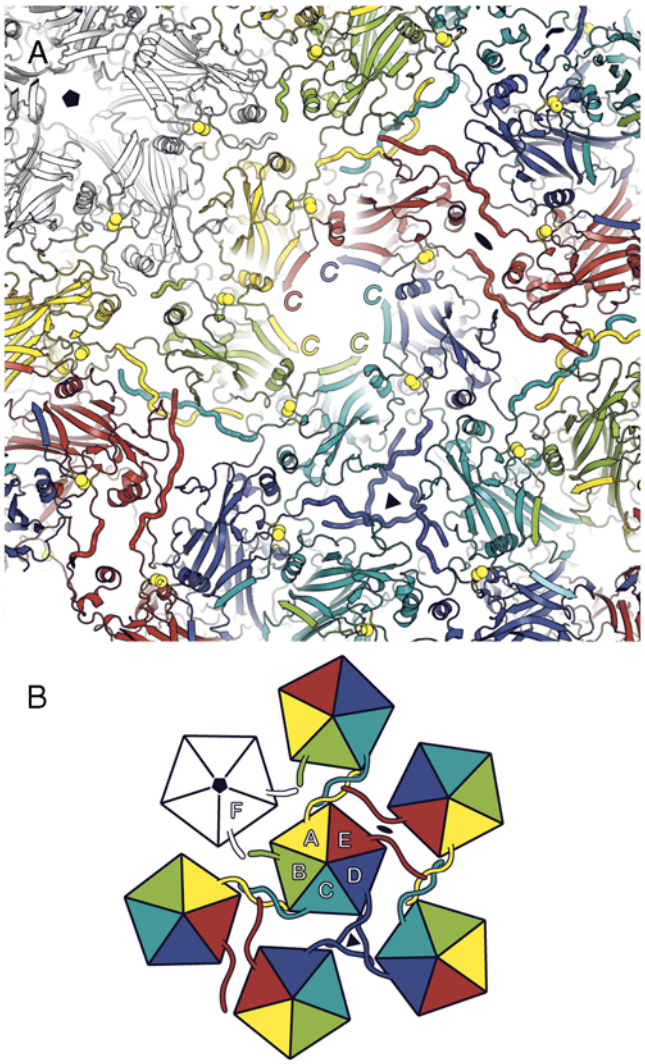


Fig. 5. View of BPV capsid from inside, looking outwards, centered along the axis of a 6-coordinated pentamer. (*A*) Ribbon representation, with N-terminal arms in worm representation, to distinguish them. Note the augmentation of a β-sheet in each subunit by the C-terminal segment of the neighboring subunit in the pentamer. (*B*) Icosahedral asymmetric unit, viewed from within the capsid. The pattern of N-terminal arm contacts is illustrated by the worm-like projections from each subunit. Note that N-terminal arms are ordered only on subunits A, C, D, and E.

interactions, the extended conformation of residues 1–14 is essentially invariant among the 4 ordered copies within an icosahedral asymmetric unit. A stretch of conserved residues (V/L-Y-L/V-P-P, 10–14 in BPV1) and a conserved tryptophan (residue 4 in BPV1) participate in these varied contacts: A similar N-terminal "scaffold" is therefore likely to be present in other papillomavirus particles.

Discussion

Virion Structure. The central problem for the molecular architecture of a 72-pentamer, icosahedral structure is how to adapt a 5-fold symmetric assembly unit to a 6-coordinated surround. The solution found in capsids of polyomaviruses and papillomaviruses depends on the properties of adaptable subunit arms emanating from a conserved pentamer core. In capsids of both groups of viruses, the long C-terminal arms can have a variety of orientations and degrees of extension, but the segment that interacts with its target site in a neighboring pentamer has an invariant conformation and invariant set of contacts. Moreover, in both cases, disulfide bonds strengthen these contacts (5, 11), requiring that for productive infection the virus pass through a compartment in which these bonds are broken. The BPV structure shows that an important difference between capsids in the two groups of viruses—and a revision to the earlier model for the papillomavirus structure (9)—is whether the C-terminal arm inserts for its entire length into the target pentamer (in the polyomaviruses) or it inserts back into its pentamer of origin (in the papillomaviruses, as shown here). The BPV C-terminal arm emerges from the core of the subunit at residue 402. Residues 411–426 contact the target pentamer, to which Cys426 contributes a disulfide; residues 427–435 bridge back to the pentamer from which the chain originated. The new chain trace provides a more plausible relationship between the contacts in the T = 1 small VLP and the contacts in the virion. The C terminus of the L1 polypeptide chain is anchored identically in both assemblies, creating a long, flexible loop between residues 402 and 435. The tip of this loop contains the conserved LxDxYR(Y/F) motif (residues 412–418) and Cys426 and docks in a uniform way against the target pentamer in all 6 NCS-related contacts in the virus. The small VLP forms only when residues 1–10 are deleted (7). The BPV structure shows that the presence of these residues, which are a large part of the N-terminal arm interaction, is incompatible with the sharp curvature needed to generate the smaller icosahedral particle.

The atomic model accounts for several sequence motifs conserved among the large number of different papillomavirus types (see HPV sequence database: http://hpv-web.lanl.gov/). One of the most unusual features of the structure—the essentially

contact-free, "strung out" conformation of residues 429–435—also correlates with conserved sequence characteristics (prolines and polar residues). These conservations are strong evidence that the same sets of L1 capsid contacts are present throughout the papillomavirus family.

Particle Assembly. L1 pentamers (like polyomavirus VP1 pentamers) have such intricate and extensive interactions between adjacent subunits that soluble monomers are not present in detectable quantity and the pentameric "capsomere" is clearly the fundamental assembly unit of the virion. One copy of the internal protein L2 can associate with an L1 pentamer, through contacts formed within its inward-facing conical hollow (as seen at low resolution in cryoEM reconstructions of pseudovirions with high L2 occupancy) (12, 13). A set of hydrophobic residues toward the C terminus of L2 is essential for L1 binding (12). The resulting picture is similar to the one derived from crystallographic studies of a polyomavirus VP1:VP2/3 complex (14): The C-terminal part of L2 plugs the conical hollow in a hairpin-like conformation, with the hydrophobic segment in contact with one or more L1 subunits and the bulk of the L2 polypeptide chain projecting inward. Estimates of the number of L2 copies per virion vary, but in any case only a small fraction of the L1 pentamers bear this inward-projecting appendage. There is no evidence in our reconstruction for L2 density (or for other connections between L1 and the minichromosome), and we can therefore conclude that L2 is unlikely to be incorporated selectively into the 5-coordinated pentamers. L2 is critical both for DNA packaging and for entry (15).

A first step in particle assembly must be formation of a pentamer–pentamer interface. The contact between a 6- and a 5-coordinated pentamer is the only one that includes interactions other than those with N- or C-terminal arms. The CD loops in the jelly-roll cores of the B and F subunits abut, in a particularly tight contact. This interface is likely to be the one that forms when two pentamers associate, as suggested for polyomaviruses (16). Subsequent steps may also conform to the general considerations outlined for polyomavirus capsid formation. The disulfide bonds between cysteines 426 in the bridging arms and cysteines 171 in their target pentamers, shown to be required for virion stability, lock the capsid structure into place. During intranuclear assembly, however, the cysteines presumably remain reduced, forming disulfides only after the cell has lysed and the virus particles have entered an oxidizing environment. Indeed, pseudovirions or capsids that assemble in cells transfected with plasmids encoding L1, L2, and (for pseudovirions) packageable DNA have few if any disulfide bridges when freshly isolated from culture supernatants, and the particles collapse and flatten when negatively stained with uranyl acetate for electron microscopy (17). If the preparation is allowed to "mature" overnight, specific disulfides form, and particle images have the familiar round, rigid appearance of purified virus particles.

Virus Entry. The primary cell-attachment receptors for many types of papillomavirus are thought to be heparan-sulfate proteoglycans (18), but specific interactions have yet to be visualized experimentally (19). Following attachment, virions undergo a conformational transition that allows the N-terminal part of L2 to become accessible to cell-surface furin (20). This transition may be related to an in vitro expansion detected by electron microscopy in samples imaged at low ionic strength or low pH (21). Furin cleavage (at a conserved site around residue 10), which is required for infectivity, in turn exposes an L2 epitope (residues 17–36 in HPV16) that can elicit neutralizing antibodies against multiple papillomavirus types (22). There is evidence that transfer to a secondary receptor follows the primary-receptor induced conformational transition and that association with the secondary receptor leads to endocytic uptake (23). A hydrophobic/basic sequence near the C terminus of the L2 polypeptide chain is essential for escape from endosomes (24).

The BPV shell has only rather restricted "holes" through which the N terminus or C terminus of L2 could emerge. The opening along the 5-fold axis of L1 has a van der Waals diameter of 10–12 Å, and a gap to either side of the 2-fold contact has about the same size. A small expansion that retained the contacts within a pentamer would leave the former opening unchanged but substantially increase the latter. For example, the 7% increase in diameter reported for the "open" form of cottontail rabbit papillomavirus (CRPV) (21) would increase the interpentamer distance by about 10 Å, essentially all of which would be an enlargement of the gaps between them. Such an expansion could be accommodated by conformational changes in the segments of the C-terminal arm that cross from one pentamer to another and back (residues 403–411 and 427–435), without breaking the Cys171-Cys426 interpentamer disulfide or disrupting its surrounding noncovalent contacts. Reduction of the disulfides in BPV also leads to expansion, accompanied by exposure of the L1 C-terminal arms (10), but reduction is probably a later event, after the particle has entered the cytosol.

High-Resolution cryoEM. The clarity and resolution of the density map, from which we have derived an atomic model for BPV, are products of many incremental improvements in experimental and computational cryoEM methods. The virus preparation itself was exceptionally homogeneous, and the BPV particles are unusually robust. Merging the data may have been facilitated by the use of images all collected within a few hours from a single grid square, without major changes in beam geometry or eucentric height. The size of the reconstruction required a 64-bit processor for *FREALIGN* computation (see *Materials and Methods*). Sampling was adequate at 1.27 Å/pixel for the resolution obtained (2.9-fold oversampling, Nyquist limit at 2.54 Å) and required a relatively large reconstruction volume (880^3 voxels). Accurate determination of defocus for each particle, on the basis of micrograph tilt and calibrated magnification, was essential (a reconstruction without individual defocus adjustment had 4.5 Å resolution). We also paid careful attention to scaling and weighting of the amplitudes for the reconstructed density, while avoiding oversharpening (see *Materials and Methods*). "Bad" particles were excluded on the basis of a projection-reference correlation criterion. The additional NCS averaging within the icosahedral asymmetric unit further improved the signal-to-noise ratio and extended the resolution from 4.2 to 3.6 Å. The reconstruction was computed from about 4,000 "good" particles. Because each view is icosahedrally equivalent to 59 others, the effective number of projections was about 240,000. The 6-fold NCS averaging implies that a better number, for comparison with required views from an asymmetric object, might be about 1.5 million. These estimates suggest that it may be possible to achieve similar resolution with asymmetric particles, if there is enough low-resolution contrast to avoid orientation ambiguities and if the objects themselves are as homogenous and rigid as BPV. Successful alignment will depend on the presence of clear particle features that help identify orientations. Structural variability of the particles can lead to misalignment and loss of resolution in the variable regions, and images will therefore need to be sorted into subsets of structurally equivalent particles. This sorting will increase the total number of required images, depending on the number of subsets.

Materials and Methods

Image Acquisition. Data from vitrified samples of BPV, prepared as described (10), were collected on an FEI Tecnai F30 electron microscope (300 kV). Forty-nine micrographs, acquired on film during a single session at 25 $e^-/Å^2$, nominal magnification 59,000× (56,588× calibrated), defoci between −1.8 to −2.9 μm, were digitized at 7 μm (1.237 Å/pixel at the specimen). For further details, see *SI Text*.

Table 1. Refinement statistics

Number of "structure factors"	16,264,110
Constraints (icosahedral operators)	60
NCS restraints (local NCS operators)	6
NCS groups	4
Polypeptide chains in icosahedral a.u.	6
Number of residues (icosahedral a.u.)	2,862
Number of atoms in full assembly	1,357,200
Resolution range (Å)	15–3.6
R factor	37.5
Ramachandran plot:	
Residues in allowed regions	100%
Residues in most favored region	80.80%
rmsd bond lengths (Å)	0.01
rmsd bond angles (°)	1.63
Average B factor, all atoms ($Å^2$)	131
Average B factor, backbone ($Å^2$)	126

Data Processing. BPV images (6,015 particles) were selected from scanned micrographs with *SIGNATURE* (25), boxed, and padded. Defocus, tilt, and astigmatism of each micrograph were determined, and individual particle defocus adjusted, with *CTFTILT* (26). Details of these steps and of the determination and refinement of particle orientations are described in *SI Text*. Ewald-sphere correction (27) enabled during the final rounds of refinement and reconstruction had only a marginal influence on the final reconstruction (3,977 particles).

Resolution. We estimated resolution by Fourier shell correlation (FSC) between two half-set reconstructions and by R factor with the final model. Resolution for the icosahedral map was 4.9 Å at FSC = 0.5 and 4.2 Å at FSC = 0.143. The R factor with the refined model reached 0.55 at a resolution of 4.1 Å, in accord with the FSC = 0.143 criterion (28). We estimated the influence of NCS on resolution by 6-fold averaging each half-set reconstruction within an asymmetric unit, by using a soft-edged spherical mask within the boundary of the binary NCS mask. The resolution after NCS averaging was 4.2 Å at FSC = 0.5 and 3.6 Å at FSC = 0.143 (*SI Text*).

Model Building, NCS Averaging, and Coordinate Refinement. Procedures for initial model building, magnification calibration, sharpening, and NCS operator determination are described in *SI Text*. Refinement in *CNSsolve* (29) was carried out by placing the reconstruction in a cubic P1 cell (512 pixels on a side; 1.237 Å sampling), calculating structure factors and figures of merit (see *SI Text*), and applying standard procedures for crystallographic refinement within the resolution range 15–3.6 Å: isotropic B-factor correction, automatic bulk-solvent parameter search, bulk-solvent correction, coordinate refinement by energy minimization, restrained individual B-factor refinement, and unrestrained group B-factor refinement. We used icosahedral constraints plus 6-fold NCS restraints for the invariant cores of the 6 L1 protomers in each icosahedral asymmetric unit. See *SI Text* for details. Refinement statistics are in Table 1. Because the refinement target was the transform of the icosahedrally symmetrized map without local NCS averaging, the calculated R factor does not reflect the enhanced resolution of the NCS-averaged density. Unlike full crystallographic refinement, our procedures adjust the model to fit the final image reconstruction, subject to restraints, but do not iteratively correct the map itself. Thus, they are roughly equivalent to an implementation in reciprocal space of a round of real-space refinement.

ACKNOWLEDGMENTS. We thank Patti Estes for technical assistance and SBGrid for computational support. The work was supported by National Institutes of Health Grants R01 CA13202 (to S.C.H.), R01 CA37667 (to R.L.G.), and P01 GM62580 (to S.C.H. and N.G.). S.C.H. and N.G. are Investigators in the Howard Hughes Medical Institute.

1. Howley PM, Lowy DR (2007) Papillomaviridae. *Fields Virology*, eds DM Knipe and PM Howley (Lippincott Williams and Wilkins, Philadelphia), pp 2299–2354.
2. Shi L, et al. (2007) GARDASIL: Prophylactic human papillomavirus vaccine development—From bench top to bed-side. *Clin Pharmacol Ther* 81:259–264.
3. Baker TS, et al. (1991) Structures of bovine and human papillomaviruses. Analysis by cryoelectron microscopy and three-dimensional image reconstruction. *Biophys J* 60:1445–1456.
4. Rayment I, Baker TS, Caspar DL, Murakami WT (1982) Polyoma virus capsid structure at 22.5 A resolution. *Nature* 295:110–115.
5. Liddington RC, et al. (1991) Structure of simian virus 40 at 3.8-A resolution. *Nature* 354:278–284.
6. Stehle T, Yan Y, Benjamin TL, Harrison SC (1994) Structure of murine polyomavirus complexed with an oligosaccharide receptor fragment. *Nature* 369:160–163.
7. Chen XS, Garcea RL, Goldberg I, Casini G, Harrison SC (2000) Structure of small virus-like particles assembled from the L1 protein of human papillomavirus 16. *Mol Cell* 5:557–567.
8. Bishop B, et al. (2007) Crystal structures of four types of human papillomavirus L1 capsid proteins: Understanding the specificity of neutralizing monoclonal antibodies. *J Biol Chem* 282:31803–31811.
9. Modis Y, Trus BL, Harrison SC (2002) Atomic model of the papillomavirus capsid. *EMBO J* 21:4754–4762.
10. Li M, Beard P, Estes PA, Lyon MK, Garcea RL (1998) Intercapsomeric disulfide bonds in papillomavirus assembly and disassembly. *J Virol* 72:2160–2167.
11. Sapp M, Fligge C, Petzak I, Harris JR, Streeck RE (1998) Papillomavirus assembly requires trimerization of the major capsid protein by disulfides between two highly conserved cysteines. *J Virol* 72:6186–6189.
12. Finnen RL, Erickson KD, Chen XS, Garcea RL (2003) Interactions between papillomavirus L1 and L2 capsid proteins. *J Virol* 77:4818–4826.
13. Buck CB, et al. (2008) Arrangement of L2 within the papillomavirus capsid. *J Virol* 82:5190–5197.
14. Chen XS, Stehle T, Harrison SC (1998) Interaction of polyomavirus internal protein VP2 with the major capsid protein VP1 and implications for participation of VP2 in viral entry. *EMBO J* 17:3233–3240.
15. Holmgren SC, Patterson NA, Ozbun MA, Lambert PF (2005) The minor capsid protein L2 contributes to two steps in the human papillomavirus type 31 life cycle. *J Virol* 79:3938–3948.
16. Stehle T, Gamblin SJ, Yan Y, Harrison SC (1996) The structure of simian virus 40 refined at 3.1 A resolution. *Structure* 4:165–182.
17. Buck CB, Thompson CD, Pang YY, Lowy DR, Schiller JT (2005) Maturation of papillomavirus capsids. *J Virol* 79:2839–2846.
18. Knappe M, et al. (2007) Surface-exposed amino acid residues of HPV16 L1 protein mediating interaction with cell surface heparan sulfate. *J Biol Chem* 282:27913–27922.
19. Sapp M, Bienkowska-Haba M (2009) Viral entry mechanisms: human papillomavirus and a long journey from extracellular matrix to the nucleus. *FEBS J* 276:7206–7216.
20. Richards RM, Lowy DR, Schiller JT, Day PM (2006) Cleavage of the papillomavirus minor capsid protein, L2, at a furin consensus site is necessary for infection. *Proc Natl Acad Sci USA* 103:1522–1527.
21. Belnap DM, et al. (1996) Conserved features in papillomavirus and polyomavirus capsids. *J Mol Biol* 259:249–263.
22. Day PM, Gambhira R, Roden RB, Lowy DR, Schiller JT (2008) Mechanisms of human papillomavirus type 16 neutralization by l2 cross-neutralizing and l1 type-specific antibodies. *J Virol* 82:4638–4646.
23. Selinka HC, et al. (2007) Inhibition of transfer to secondary receptors by heparan sulfate-binding drug or antibody induces noninfectious uptake of human papillomavirus. *J Virol* 81:10970–10980.
24. Kamper N, et al. (2006) A membrane-destabilizing peptide in capsid protein L2 is required for egress of papillomavirus genomes from endosomes. *J Virol* 80:759–768.
25. Chen JZ, Grigorieff N (2007) SIGNATURE: A single-particle selection system for molecular electron microscopy. *J Struct Biol* 157:168–173.
26. Mindell JA, Grigorieff N (2003) Accurate determination of local defocus and specimen tilt in electron microscopy. *J Struct Biol* 142:334–347.
27. Wolf M, DeRosier DJ, Grigorieff N (2006) Ewald sphere correction for single-particle electron microscopy. *Ultramicroscopy* 106:376–382.
28. Rosenthal PB, Henderson R (2003) Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J Mol Biol* 333:721–745.
29. Brunger AT, et al. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D* 54:905–921.

Supporting Information

Wolf et al. 10.1073/pnas.0914604107

SI Text

SI Materials and Methods. ***Sample preparation.*** Bovine papillomavirus (BPV) was purified from cow warts on a CsCl gradient as described (1). The 100 μL purified sample (0.5 mg/mL) was dialyzed (7,000 MW cutoff) against buffer (20 mM Tris pH 6.2, 100 mM NaCl, 0.5 mM $CaCl_2$) overnight and concentrated fourfold on a Microcon YM100 centrifugal concentrator (Millipore Corp.). A 3.5 μL concentrated virus solution was applied to a glow-discharged C-flat holey carbon grid (CF-1/2-4C, Protochips Inc.), blotted at 4 °C for 25 sec from the carbon side with Whatman no. 40 filter paper, and vitrified by plunging into liquid ethane at liquid nitrogen temperature on a manual plunger.

Image acquisition. Data were collected during a single session on a FEI Tecnai F30 electron microscope operating at 300 kV and equipped with a field emission source. The optical system was aligned by using standard procedures (beam tilt/shift, eucentric height, pivot points, rotational center, astigmatism). Coma-free alignment was not performed, but the imaging of high-resolution features was verified by visualizing a gold lattice on a test grid. The objective aperture had a resolution cutoff of 2.4 Å. Forty-nine images were acquired by using a low-dose procedure at electron dose of 25 $e^-/Å^2$ on Kodak SO163 film at a nominal magnification of 59,000× (56,588× calibrated) and with defoci covering the range −1.8 to −2.9 μm. Specimen drift on a Gatan 626 stage was limited to <3 Å/sec as measured by CCD image shift. Ice thickness was 80–120 nm. Micrographs were developed with full strength Kodak D19 for 10 min and digitized on a Zeiss SCAI scanner at 7 μm sample size (corresponding to 1.237 Å/pixel at the specimen level) using a combination of the three color channels and the Zeiss scanner software for optical density correction.

Data processing. Particles were preselected from scanned micrographs with *SIGNATURE* (2) using a single rotationally averaged reference derived from 10 manually picked particles in a micrograph at the median of the defocus range. After visual inspection, 6,015 particles were boxed with 880^2 pixel area at full resolution. A padding of 250 Å around the particle preserved high-resolution information to 3.0 Å displaced by a defocus of up to 3.0 μm. Defocus, tilt, and astigmatism of each micrograph were determined with *CTFTILT* (3). The mean tilt was 4.5 ± 4.2°. Individual particle defocus was adjusted from the average value of its micrograph, depending on tilt and particle position. The deviation of mean tilt from zero may have been the result of several factors: bent grid during cryotransfer, undulating carbon support, and so forth. It will have contributed to the defocus spread in the data, but small variations of focus during the image acquisition cycle will probably have contributed equally. It is unlikely to have had any role in obtaining the near-atomic resolution we report. The initial orientation search was performed using 2-fold binned data (2.47 Å/pixel, 384 $pixel^2$) and a previously calculated HPV11 virus-like particle reconstruction as three-dimensional reference. Handedness and calibrated magnification of the alignment reference had been established previously. A systematic orientation search to 5 Å was carried out using *FREALIGN* (mode 3) and resolution extended to 4.2 Å over five local angular and positional refinement cycles (*FREALIGN* mode 1) against the latest BPV reference reconstruction. A histogram of the particle-reference correlation (program variable PRES) was bimodal; only particles from the major peak (higher correlation) were included (66%). After five cycles of refinement using unbinned data, magnification was recalibrated with an atomic pentamer model (see *SI Model Building, NCS Averaging, and Coordinate Refinement*). All contrast-transfer-function parameters were recalculated using the improved pixel size followed by an additional three cycles of refinement and reconstruction. Ewald-sphere correction (4) was enabled during both refinement and reconstruction with the unbinned data, but structural detail in the final reconstruction (3,977 particles) was only marginally inferior without the correction.

Resolution estimate. We estimated resolution by Fourier shell correlation (FSC) between two half-set reconstructions and by R factor with the final model. The icosahedrally symmetrized reconstruction was masked inside and outside the capsid at radii of 205 and 295 Å with a soft cosine edge of half-width = 10 pixel. The resolution for the icosahedral map was 4.9 Å at FSC = 0.5 and 4.2 Å at FSC = 0.143. The R factor with the refined model reached 0.55 at a resolution of 4.1 Å, in accord with the FSC = 0.143 criterion (5). To estimate map resolution after nonicosahedral capsid symmetry (NCS) averaging, each half-set reconstruction was averaged 6-fold within the asymmetric unit and masked with a soft-edged spherical mask within the boundary of the binary NCS mask to avoid mask contributions. The measured resolution after NCS averaging was 4.2 Å at FSC = 0.5 and 3.6 Å at FSC = 0.143 (Fig. S4). The NCS-averaged map included data to 3.2 Å after treatment with a low-pass filter (cutoff with cosine edge of 3 pixels width) from a reconstruction calculated to 3.0 A resolution. We limited the data used for model refinement by specifying a resolution range of 15.0–3.6 Å in *CNS*. We used NCS to restrain coordinate refinement for selected groups within the invariant cores of the 6 L1 protomers; because the refinement target was the transform of the icosahedrally symmetrized map without local NCS averaging, the calculated R factor does not reflect the improved resolution of the NCS-averaged density.

SI Model Building, NCS Averaging, and Coordinate Refinement. Models for two pentamers of HPV16 L1 core domains (6), with the loop of the C-terminal arm deleted, were manually docked into the electron cryomicrosopy reconstruction at a 5-fold and at an adjacent general position, using the program *O* (7). We calibrated magnification with *SITUS* (8) by maximizing the real-space correlation coefficient between the atomic pentamer model and the map while varying pixel size and then recalculated the reconstruction with the refined pixel size. Average amplitudes from the raw map (in resolution shells) were scaled to those from a model-generated map (see below) using the program *DIFFMAP* (http://emlab.rose2.brandeis.edu/diffmap); the scaled amplitudes were then sharpened with a B factor of −160 combined with figure-of-merit weighting (5).

NCS operators relating the six protomers in the asymmetric unit were determined in *O* (command LSQ_EXPLICIT) and improved in *MAVE* (9, 10). Density was 6-fold averaged in *MAVE* and calculated within a masked monomer and a pentamer (five orientations of the general pentamer averaged with the 5-fold pentamer). The peptide sequence was altered from HPV16 to BPV1 using *MODELER* (mean sequence identity = 0.46; mean sequence similarity = 0.66) and manually rebuilt in *O*. C-terminal and N-terminal arms were traced *de novo* in the icosahedral (non-NCS-averaged) map relying on the fragment-based *LEGO* algorithm in *O* to impose good backbone geometry. At this stage, side chains were mostly modeled as their preferred

rotamer, but density in the core domain often allowed unambiguous assignment of side-chain rotamers. A model of the complete capsid was generated with *PDBSET* (11) and converted to a map with *SFALL* (11) and *FFT* (11).

To refine the entire capsid in *CNS* (12) while using available amplitude and phase information, we converted the real-space data to reciprocal space and generated amplitude and phase error statistics. The reconstruction was placed in a cubic cell with side dimensions of 512 pixels, the particle center shifted to the origin with *MAPMASK* (11), and structure factors calculated with *SFALL* (11). The resulting P1 lattice was sampled at 1.237 Å. Structure factors from half-set reconstructions were calculated independently and combined with *CAD* (11). Amplitude errors were approximated in *SFTOOLS* (11), by difference amplitudes between the two half-set reconstructions $[\sigma(\mathrm{f}) = |\mathrm{sqrt}(0.5) * \Delta\mathrm{f}|]$. The phase figure-of-merit (FOM) was estimated from the FSC as described for c_{ref} (5) and set to the same value for all structure factors within a given resolution shell. A separate program was written to assign FOMs to the respective Fourier terms, and phase and FOM were converted to Hendrickson–Lattman coefficients HA and HB using *SFTOOLS* ($\mathrm{HC} = 0$, $\mathrm{HD} = 0$ for monomodal distribution). After addition of an R-free column (flagging 5% of structure factors), the data were combined in *CAD* and written out to *CNS* format with *MTZ2VARIOUS* (11) in the format "H K L F sigmaF HA HB HC HD R-free."

Refinement in *CNSsolve* (12) used the standard procedures for crystallographic refinement within the resolution range 15–3.5 Å: isotropic B-factor correction, automatic bulk-solvent parameter search, bulk-solvent correction, coordinate refinement by energy minimization, restrained individual B-factor refinement, and unrestrained group B-factor refinement. Both icosahedral and non-icosahedral capsid symmetry were imposed, using icosahedral constraints and 6-fold NCS restraints within the asymmetric unit defined in five NCS groups (*CNSsolve* version 1.21, 64 bit) (12). We used default x-ray scattering factors and a maximum likelihood refinement target ("mlhl") with amplitude and phase distribution. Manual model building, taking difference maps into account, alternated three times with automatic refinement in *CNS*. We validated model geometry with *PROCHECK* (13) and *MOLPROBITY* (14). Refinement statistics are in Table 1. Unlike full crystallographic refinement, the procedures we have adopted adjust the model to fit the final image reconstruction, subject to restraints, but do not iteratively correct the map itself. Thus, they are roughly equivalent to an implementation in reciprocal space of a round of real-space refinement.

Figures were prepared with PYMOL (http://pymol.sourceforge.net/faq.html#CITE: DeLano Scientific) and GNUPLOT.

1. Li M, Beard P, Estes PA, Lyon MK, Garcea RL (1998) Intercapsomeric disulfide bonds in papillomavirus assembly and disassembly. *J Virol* 72:2160–2167.
2. Chen JZ, Grigorieff N (2007) SIGNATURE: A single-particle selection system for molecular electron microscopy. *J Struct Biol* 157:168–173.
3. Mindell JA, Grigorieff N (2003) Accurate determination of local defocus and specimen tilt in electron microscopy. *J Struct Biol* 142:334–347.
4. Wolf M, DeRosier DJ, Grigorieff N (2006) Ewald sphere correction for single-particle electron microscopy. *Ultramicroscopy* 106:376–382.
5. Rosenthal PB, Henderson R (2003) Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J Mol Biol* 333:721–745.
6. Chen XS, Garcea RL, Goldberg I, Casini G, Harrison SC (2000) Structure of small virus-like particles assembled from the L1 protein of human papillomavirus 16. *Mol Cell* 5:557–567.
7. Jones TA, Zou JY, Cown SW, 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.
8. Wriggers W, Milligan RA, McCammon JA (1999) Situs: A package for docking crystal structures into low-resolution maps from electron microscopy. *J Struct Biol* 125:185–195.
9. Jones TA (1992) A, yaap, asap, @#*? A Set of Averaging Programs. *CCP4 Proceedings*, pp 91–105.
10. Kleywegt GJ, Read RJ (1997) Not your average density. *Structure* 5:1557–1569.
11. Collaborative Computational Project (1994) The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr D* 50:760–763.
12. Brunger AT, et al. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D* 54:905–921.
13. Laskowski RA, Moss DS, Thornton JM (1993) Main-chain bond lengths and bond angles in protein structures. *J Mol Biol* 231:1049–1067.
14. Davis IW, et al. (2007) MolProbity: All-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* 35(Web Server issue):W375–383.
15. Gouet P, Courcelle E, Stuart D, Metoz F (1999) ESPript: Multiple sequence alignment in PostScript. *Bioinformatics* 15:305–308.
16. Harauz G, van Heel M (1986) Exact filters for general geometry three dimensional reconstruction. *Optik* 73:146–156.
17. van Heel M, Schatz M (2005) Fourier shell correlation threshold criteria. *J Struct Biol* 151:250–262.
18. Orlova EV, et al. (1997) Structure of keyhole limpet hemocyanin type 1 (KLH1) at 15 A resolution by electron cryomicroscopy and angular reconstitution. *J Mol Biol* 271:417–437.

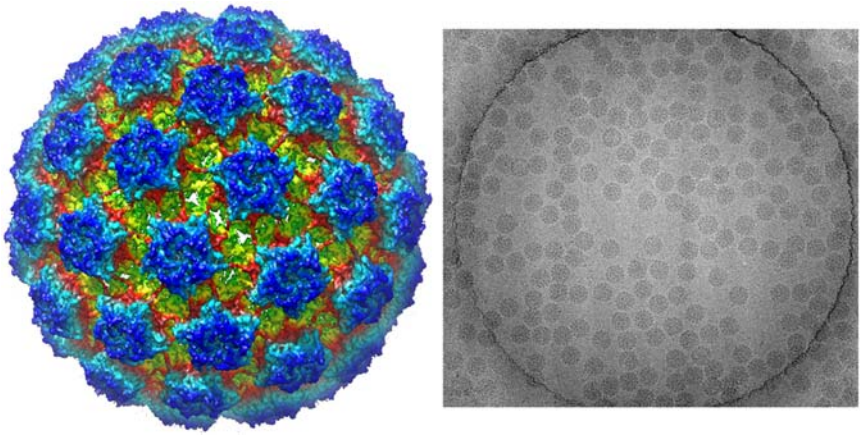


Fig. S1. Surface representation of the BPV three-dimensional image reconstruction. Radial coloring emphasizes the isolated connections formed by the C-terminal loops (*Red Layer*). *Right*: Representative electron micrograph with BPV particles suspended in amorphous ice. The diameter of the circular hole in the carbon film is 1.0 μm. The perimeter of the electron beam was slightly outside the displayed area. This image was 2.3 μm underfocused at 300 kV acceleration voltage.

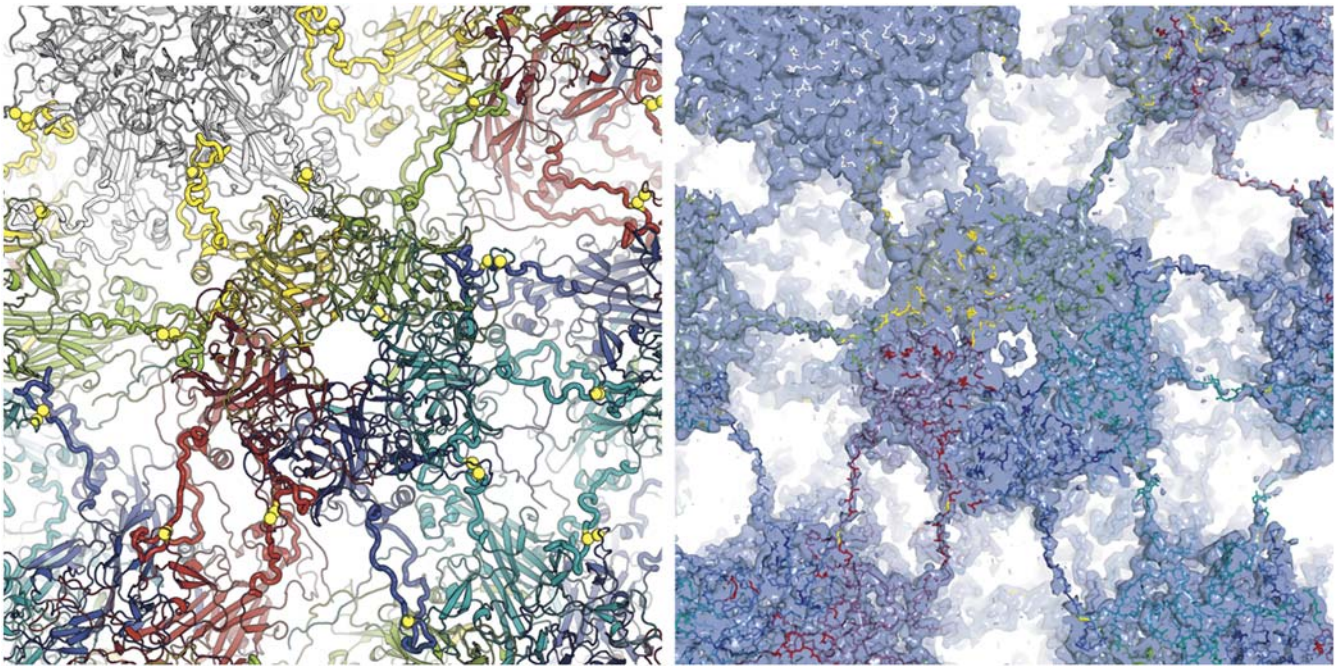


Fig. S2. Partial radial section of the particle viewed from the outside along the axis of a six-coordinated pentamer. Colors as in Fig. 1. *Left*: Ribbon representation of a pentamer and part of its surrounding pentamers. The L1 C-terminal arms are emphasized as worms, and the interpentamer disulfides are highlighted as yellow spheres. *Right*: Semitransparent, depth cued density map (contoured at 2.5σ above the average density), with embedded model.

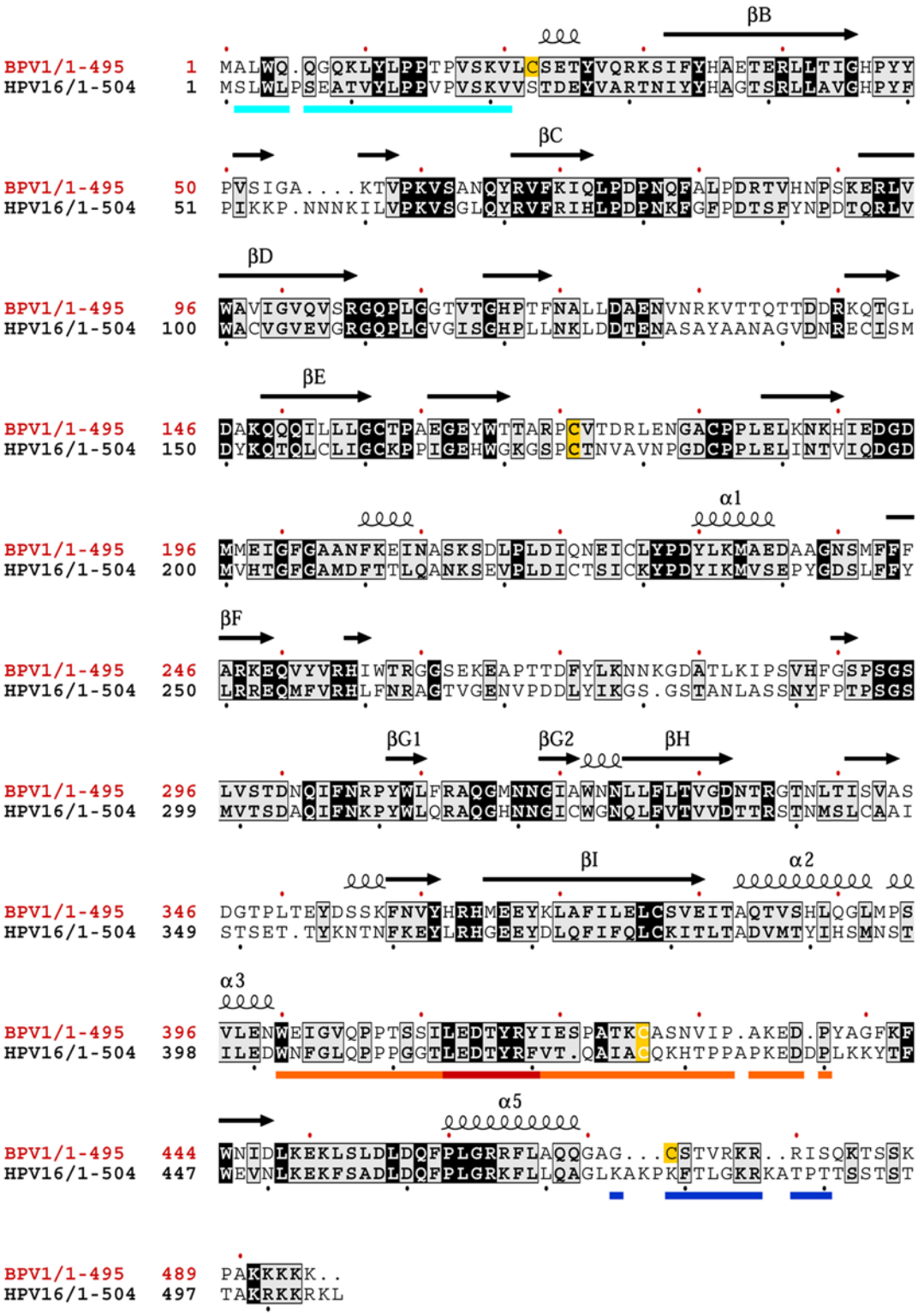
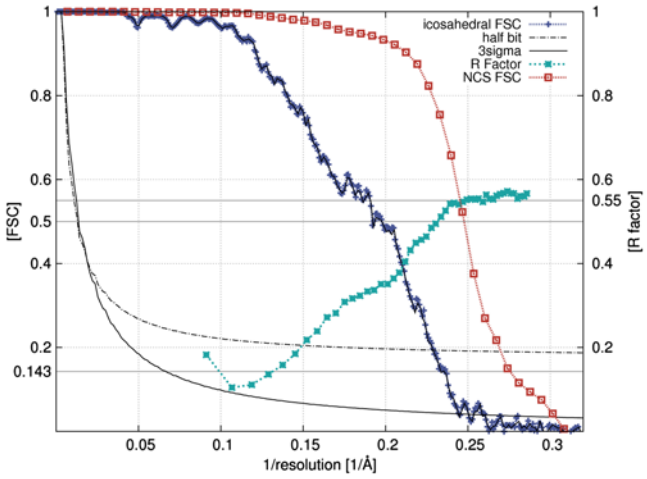


Fig. S3. Alignment of BPV1 and HPV16 L1 amino acid sequences. Boxed residues are conserved across most papillomavirus types; those strictly conserved among 25 selected types (BPV 1/2, HPV 5/6/11/16/18/31/33/35/43/44/45/53/54/56/57/58/59, COPV, CRPV, DPV, ECPV, FPV, RPV1; mean identity: 0.20, mean similarity: 0.69) are in white letters on black; those >70% conserved are in black letters on gray. The portion of the C-terminal arm that extends into a neighboring pentamer and returns is underlined in orange; the conserved tip of this extended loop, including Cys426, is underlined in red; residues not present or disordered in the HPV16 L1 structure (6) are underlined in cyan (N-terminal segment) or in blue (C-terminal segment). The cysteines that participate in the interpentamer disulfide are highlighted in yellow. Secondary structural elements are shown as coils or arrows above the sequence. The framework strands of the β-jelly roll are labeled according to convention. Helices 1–3 and 5 are numbered as in the HPV16 L1 structure. Figure drawn with ESPript (15).



Method	Intercept		Resolution [Å]	
			Icosahedral map	Additional 6-fold NCS
Fourier shell correlation	FSC=	0.500	5.05	4.05
	FSC=	**0.143**	**4.22**	**3.63**
	half-bit curve (FSC=	0.193)	4.36	
	3σ curve (FSC=	0.042)	4.10	
R Factor	R=	0.550	4.05	

Fig. S4. Image analysis statistics. *Top*: FSC (16) and R factor, plotted as functions of 1/resolution. The effective resolutions of the icosahedrally symmetrized reconstruction and of the 6-fold NCS-averaged part of the map were estimated from the FSC for two reconstructions, each containing half the number of particle images. The FSC for the 6-fold NCS-averaged map was determined using soft-masked subvolumes of the half-set reconstructions. Intercepts corresponding to FSC = 0.5 and FSC = 0.143 are shown, as are the half-bit (17) and 3-sigma (18) curves for the overall reconstruction (icosahedral symmetry only). An R factor of the computed structure factor amplitudes of the atomic model calculated against F_{obs} of the experimental map reached 55% (random amplitude agreement) at a resolution of 4.1 Å. *Bottom*: Table summarizing resolution as determined by the various criteria shown.