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.

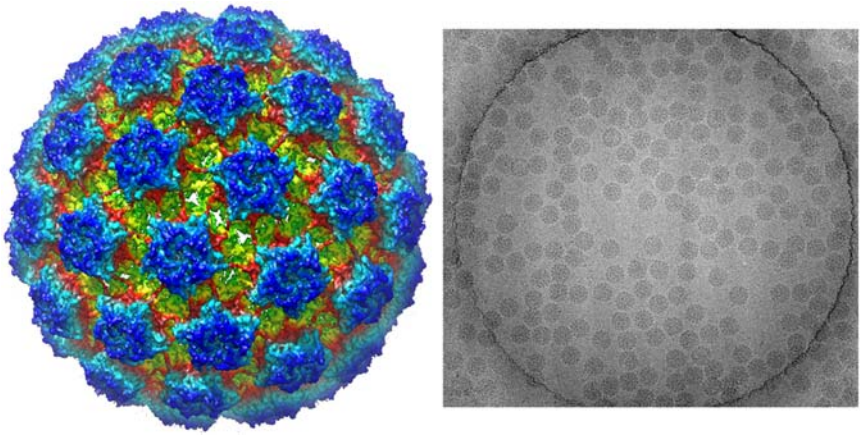


Fig. S1. Surface representation of the BPV three-dimensional image reconstruction. Radial coloring emphasizes the isolated connections formed by the C-terminal loops (*Red Layer*). *Right*: Representative electron micrograph with BPV particles suspended in amorphous ice. The diameter of the circular 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.

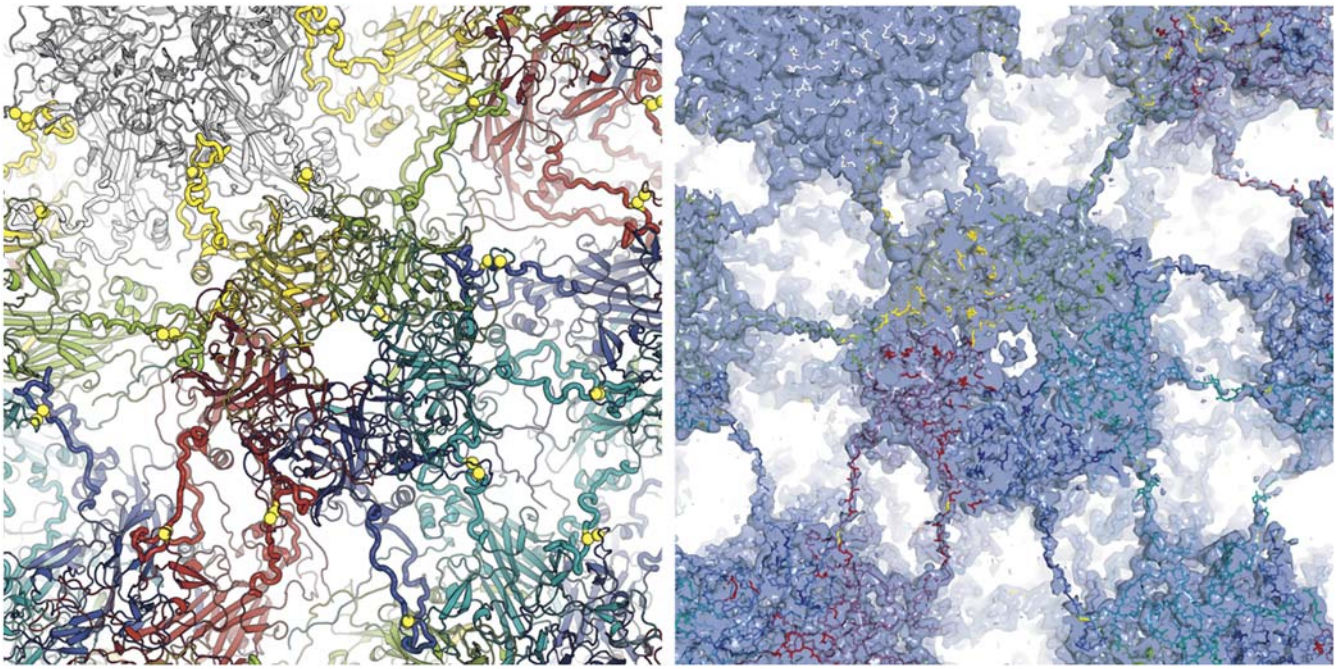


Fig. S2. Partial radial section of the particle viewed from the outside along the axis of a six-coordinated pentamer. Colors as in Fig. 1. *Left*: Ribbon representation of a pentamer and part of its surrounding pentamers. The L1 C-terminal arms are emphasized as worms, and the interpentamer disulfides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

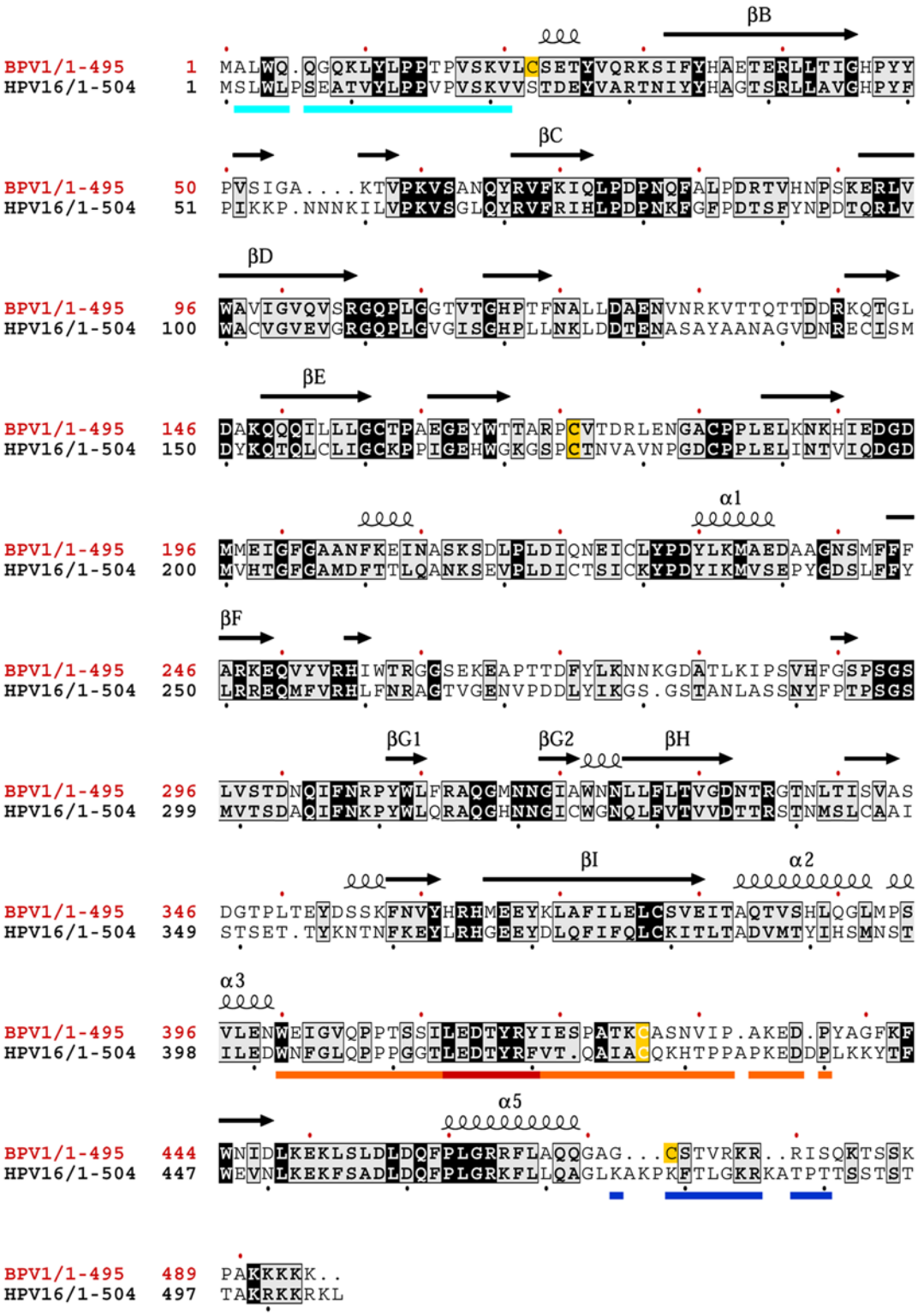
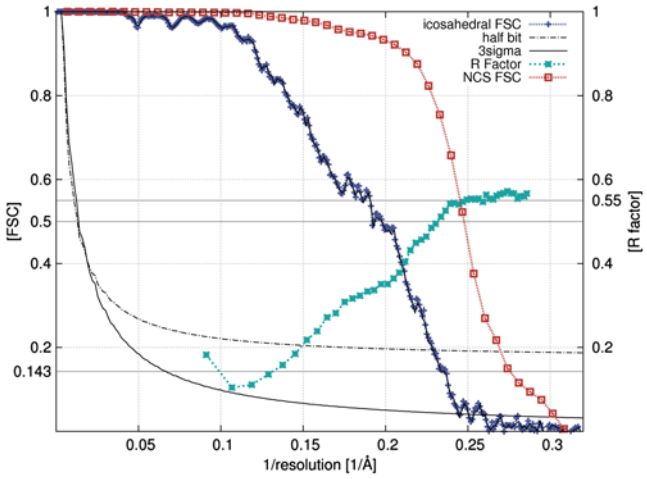


Fig. S3. Alignment of BPV1 and HPV16 L1 amino acid sequences. Boxed residues are conserved across most papillomavirus types; those strictly conserved among 25 selected types (BPV 1/2, HPV 5/6/11/16/18/31/33/35/43/44/45/53/54/56/57/58/59, COPV, CRPV, DPV, ECPV, FPV, RPV1; mean identity: 0.20, mean similarity: 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.