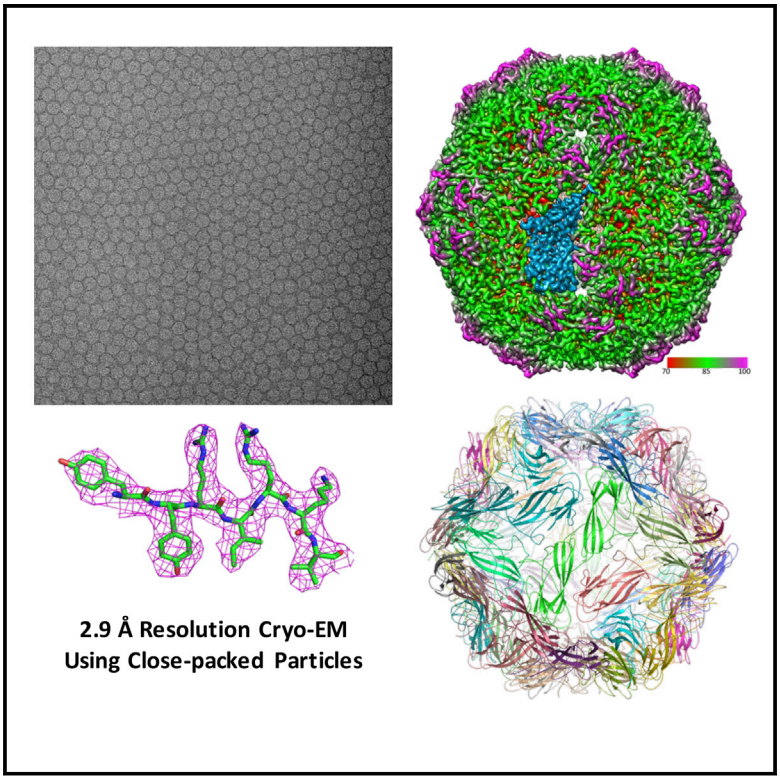
Resource

Structure

2.9 Å Resolution Cryo-EM 3D Reconstruction of Close-Packed Virus Particles

Graphical Abstract



Authors

Zheng Liu, Fei Guo, Feng Wang, Tian-Cheng Li, Wen Jiang

Correspondence

jiang12@purdue.edu

In Brief

Liu et al. developed an image processing strategy for single-particle cryo-EM images of close-packed virus particles. This strategy was used to achieve a 2.9 Å resolution reconstruction of a 1.67 MDa virus-like particle of a circovirus, PCV2, recorded on 86 photographic films in 8 hr.

Highlights

- A strategy for processing close-packed particles often regarded as unusable data
- 2.9 Å resolution 3D reconstruction based on gold standard FSC
- Cross-validation by FSC_{true} and local resolution assessment
- Cross-validation by FSC with atomic model, MolProbity and EMRinger

Accession Numbers

3JCI

Liu et al., 2016, Structure *24*, 319–328
February 2, 2016
http://dx.doi.org/10.1016/j.str.2015.12.006

Structure
Resource

2.9 Å Resolution Cryo-EM 3D Reconstruction of Close-Packed Virus Particles

Zheng Liu,[1,3] Fei Guo,[1] Feng Wang,[1] Tian-Cheng Li,[2] and Wen Jiang[1,*]
[1]Department of Biological Sciences, Markey Center for Structural Biology, Purdue University, West Lafayette, IN 47907, USA
[2]Department of Virology II, National Institute of Infectious Diseases, Gakuen 4-7-1, Musashi-murayama, Tokyo 208-0011, Japan
[3]Present address: Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY 10032, USA
*Correspondence: jiang12@purdue.edu
http://dx.doi.org/10.1016/j.str.2015.12.006

SUMMARY

Single-particle cryoelectron microscopy typically discards close-packed particle images as unusable data. Here, we report an image processing strategy and case study of obtaining near-atomic resolution 3D reconstructions from close-packed particles. Multiple independent de novo initial models were constructed to determine and cross-validate the particle parameters. The particles with consistent views were further refined including not only Euler angles and center positions but also defocus, astigmatism, beam tilt, and overall and anisotropic magnification. We demonstrated this strategy with a 2.9 Å resolution reconstruction of a 1.67 MDa virus-like particle of a circovirus, PCV2, recorded on 86 photographic films. The map resolution was further validated with a phase-randomization test and local resolution assessment, and the atomic model was validated with MolProbity and EMRinger. Close-packed virus particles were thus shown not only to be useful for high-resolution 3D reconstructions but also to allow data collection at significantly improved throughput for near-atomic resolution reconstructions.

INTRODUCTION

Single-particle cryoelectron microscopy (cryo-EM) has become a powerful method to determine the structure of large macromolecular complexes and viruses, which are often not readily accessible to X-ray crystallography and nuclear magnetic resonance spectroscopy. Tremendous progress has been made in recent years on improving the reconstructions to near-atomic resolutions (2–4 Å) at which atomic models can be reliably built (Amunts et al., 2014; Bartesaghi et al., 2015; Campbell et al., 2015; Grant and Grigorieff, 2015b; Grigorieff and Harrison, 2011; Jiang et al., 2008; Li et al., 2013; Liao et al., 2013; Ludtke et al., 2008; Zhang et al., 2008, 2010; Zhou, 2011). To overcome the high level of noise in cryo-EM images, a large amount of data (10^4–10^5 particles) is a prerequisite. The current blotting method for sample grid preparation is known for its poor control of particle density and distribution. For example, particles are often seen clustered around the edges of grid holes. Close-packed particles, though often encountered, are generally considered unusable for 3D reconstructions especially for high-resolution targets, due to interference in the determination of particle views and the fringes of contrast transfer function (CTF) from neighboring particles (Frank, 2006). Some image processing methods also implicitly assume that the particles are dispersed with sufficient background area for estimation of CTF parameters (Tang et al., 2007). There are also concerns on the view distribution constrained by interactions among neighboring particles.

Porcine circovirus 2 (PCV2) is a non-enveloped, single-stranded DNA virus of the Cicroviridae family. With a 1.7 kb genome and ~19 nm in diameter, PCV2 is one of the smallest spherical viruses (Crowther et al., 2003; Finsterbusch and Mankertz, 2009; Khayat et al., 2011). Its sole capsid protein (233 amino acids [aa]), encoded by the open reading frame 2, is capable of self-assembling into virus-like particles (VLP) of T = 1 icosahedral morphology similar to that of native PCV2 virions (Liu et al., 2008). Since its mass (~1.67 MDa) and size (~19 nm) is much smaller than virus structures previously determined at near-atomic resolutions by cryo-EM (Grigorieff and Harrison, 2011; Guo et al., 2014; Jiang et al., 2008; Yu et al., 2015; Zhang et al., 2008, 2010; Zhou, 2011), the PCV2 VLP is potentially a good test target for high-resolution cryo-EM reconstructions of small viruses.

We describe here an image processing strategy that successfully overcame difficulties in processing close-packed particles and solved the structure of PCV2 particles recorded on photographic films to 2.9 Å, the highest resolution obtained without direct electron detectors. Competitive multi-model refinements allowed robust construction of de novo initial models. Consensus solutions from refinements of independent initial models allowed robust cross-validation and identification of particles with correctly determined views for further high-resolution refinements. Refinements of higher-order microscopy parameters, including defocus, astigmatism, beam tilt, and overall and anisotropic magnifications, further improved the resolution achievable by standard refinements including only Euler angles and center positions. Thus, in contrast to the common view of these particles as unusable data, we have shown here that close-packed particles are not only useful but also allow near-atomic resolution 3D reconstructions. The large number of particles in a single image can also significantly improve the throughput of data collection.

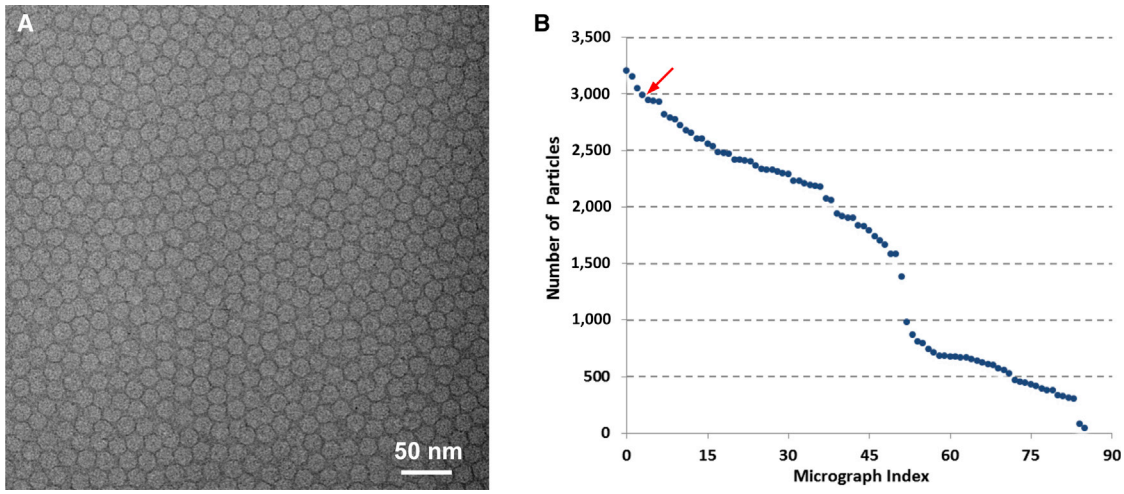


Figure 1. Cryo-EM Images

(A) Representative micrograph of close-packed PCV2 particles. There are 2,991 particles in this micrograph.

(B) Number of particles in the 86 micrographs. The arrow points to the micrograph shown in (A).

See also Figure S1.

RESULTS

Cryo-EM and Data Quality Evaluation

We overexpressed the full-length capsid protein in insect cells and purified the VLPs for structural studies using single-particle cryo-EM. Close-packed particles were encountered (Figure 1A) when the sample grids were examined on the Titan Krios microscope. Instead of preparing new sample grids to further optimize particle density/distribution and using additional imaging sessions to obtain images of more dispersed particles as in typical cryo-EM projects, we chose to continue data collection to make the best use of the microscope time but with the intention to overcome the crowding problem by improving image processing methods. Of the 168 photographic films, which were manually imaged in 8 hr, 141 were selected for digitalization. 55 of the scanned micrographs were discarded after power spectra screening due to significant sample drift. The close packing made it remarkably efficient in obtaining a large number of particles. From the remaining 86 micrographs, we selected 139,871 particles with more than 1,600 particles per micrograph on average (Figure 1B). However, almost all of the particles were packed and in close contact with multiple neighbor particles (Figure 1A). Assessments of the image quality (Figure S1) found that the average B factor of these micrographs is in the range of 300–400 $Å^2$ (Figure S1C) with the highest-resolution Thon rings visible to 6–7 Å resolution (Figures S1A and S1B). Following conventional wisdom (Grassucci et al., 2008; Liu et al., 2010), these close-packed virus particles would not be suitable for near-atomic resolution 3D reconstructions.

Image Processing and 3D Reconstruction: De Novo Initial Models

Unsurprisingly, the close packing initially posed serious challenges to image processing and 3D reconstructions, especially in building initial models and reliable determination of initial particle orientations. Nevertheless, a new image processing strategy (Figure 2) as described below demonstrated the feasibility of overcoming the challenges and reaching 2.9 Å resolution from these suboptimal images with close-packed particles.

We first randomly selected small datasets (1,000 particles in each set) and assigned them random orientations (i.e. Euler angles) followed by iterative refinements to construct de novo initial models using 4× binned particle images (Guo and Jiang, 2014; Liu et al., 2007; Yan et al., 2007). Three independent models were constructed using this random model approach for cross-validation (Figure 2). However, in contrast to consistent models typically obtained for many less challenging virus datasets (Guo and Jiang, 2014; Liu et al., 2007; Yan et al., 2007), close packing led to large discrepancies among these PCV2 de novo models (Figure 3). To reach correct initial models, these discrepant models were then used as starting references to compete in iterative multi-model refinements against the combined 3,000-particle set and to assign the particles to the best matching model (i.e. highest correlation) (Figure S2). The winning model typically attracted the majority of the particles (~60%–80%) and was chosen as the final model (Figure 3). Although the multi-model competitive refinement was often used to separate particles of heterogeneous conformations (Chen et al., 2006; Leschziner and Nogales, 2007; Scheres et al., 2007; Spahn and Penczek, 2009), its application here had effectively led to consistent initial models from independently processed datasets (Figure 3), and thus helped solve the challenges in obtaining reliable de novo initial 3D models from close-packed particles. The losing models were discarded, and only the winning model was used as good initial model for subsequent refinement against the entire dataset. As shown in Figure 3, this process generated a single winning model that was represented as one of the boxes labeled "De Novo Initial Model" in Figure 2. This process was repeated to generate multiple winning models, each represented by a different De Novo Initial Model box numbered differently (1, 2, or 3) (Figure 2).

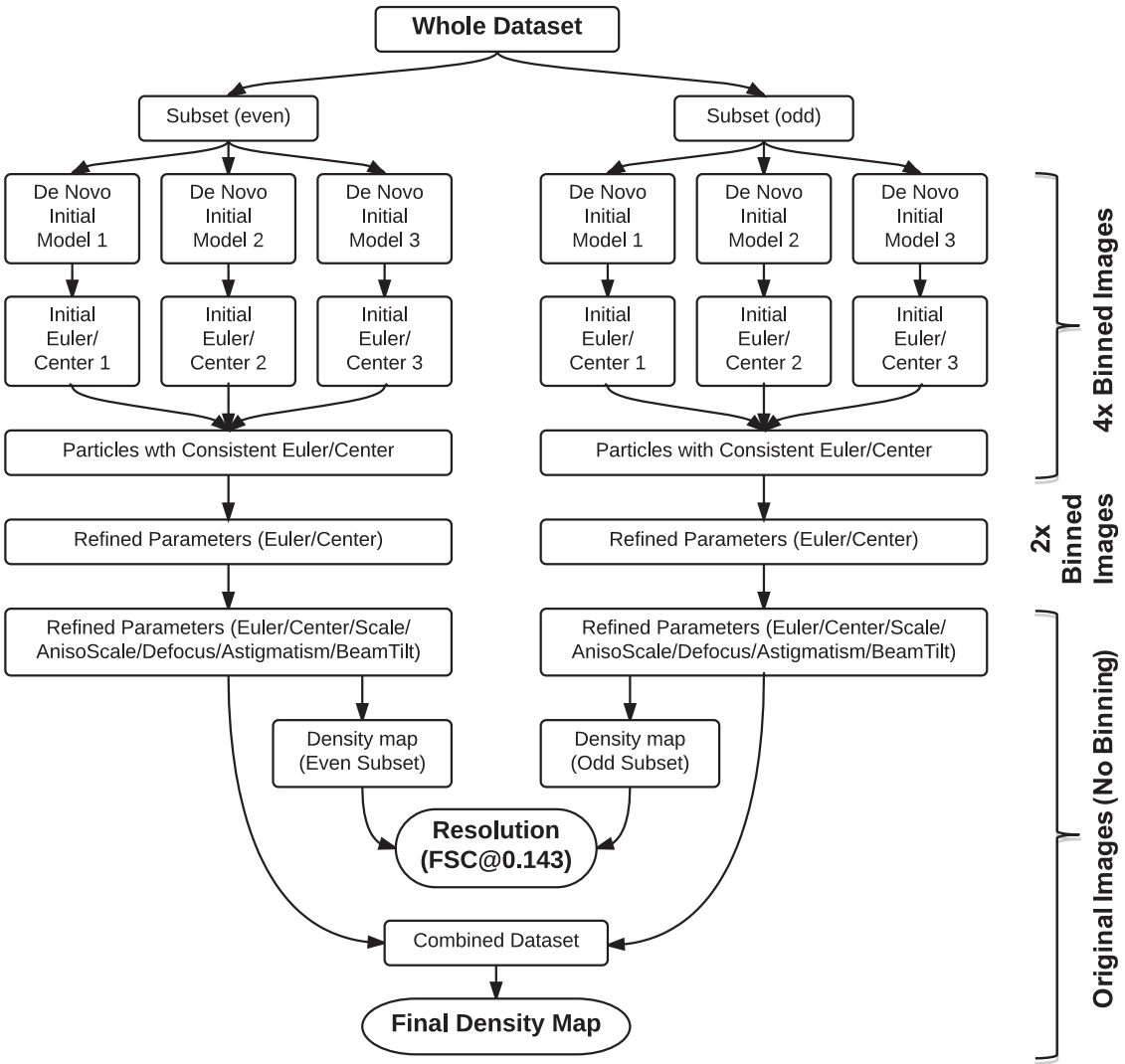


Figure 2. Flowchart of the Image Processing Strategy

The de novo initial model boxes are further expanded in Figure S2. See also Figures S2–S4.

To minimize the impact of close packing in 2D alignments, a polar Fourier transform (PFT) method (Baker and Cheng, 1996) with radial range limited to particle edge was used in conjunction with center search. The PFT method essentially uses a mask of optimal size that moves with trial particle centers during center search. This is in contrast to the self-correlation function-based 2D alignment method (Ludtke et al., 1999; Tang et al., 2007) whereby the mask is centered at the image center instead of the particle center and the mask sizes need to be significantly larger than the particle to accommodate potential offset of particles, due to imperfect centering during particle selection. To minimize the signals of neighbor particles due to CTF-induced signal spread, the particles were CTF-corrected by phase flipping before 2D alignment by PFT.

Image Processing and 3D Reconstruction: Low-Resolution Refinements

To further improve the reliability of particle views and remove particles with unreliable views, we obtained three sets of particle parameters by refining each particle with three independent de novo models, with each model being a winning model from a separate competitive multiple-model refinement using its own subset of particles (3 × 1,000 particles) (Figure 2). Only ~40% of the particles had consistent views (<2°) among all three sets of parameters and were kept for further refinements (Figure S3). This consensus approach is analogous to a popular vote and jury verdict whereby the consensus is adopted as the final solution when ground truth is not available. Using the particle parameters obtained from 4× binned images but applied to images without binning, the 3D reconstruction reached 3.66 Å, a resolution more than 2-fold better than the Nyquist resolution (8.64 Å) of the 4× binned images used for the refinements (Figure S4). This result suggests that the accuracy of 2D alignment using only low-resolution information can often be sufficient to reach significantly higher-resolution 3D reconstructions (Chen et al., 2013; Liu et al., 2007).

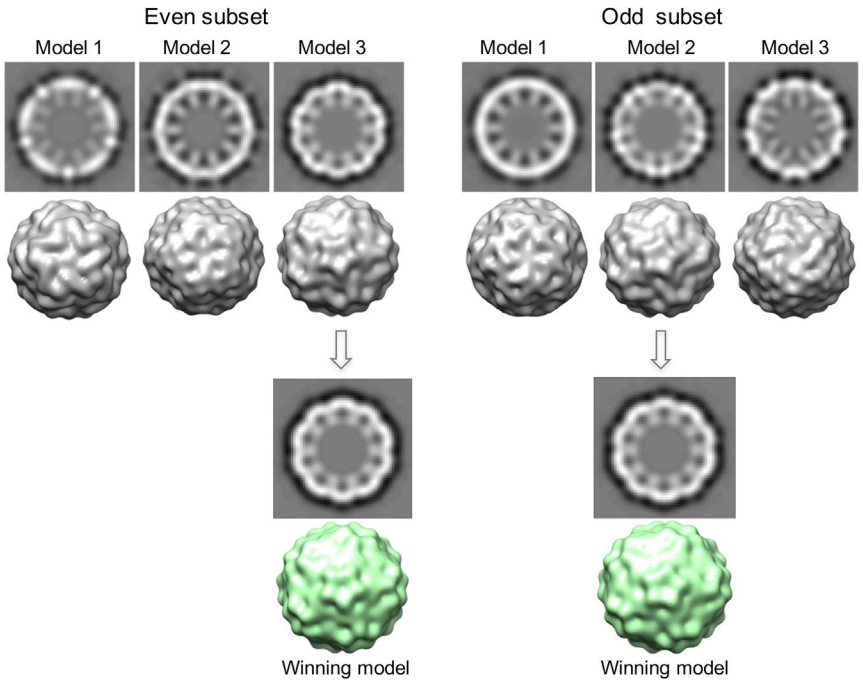


Figure 3. Competitive Multi-Model Refinement for Reliable De Novo Initial Models

Only one out of three initial models obtained for even and odd subsets of particles is shown as an example in this figure. The central-section views and the surface views of the three independent random models (each built from a unique subset containing 1,000 randomly selected particles) were significantly different, suggesting that some or all of the three models converged to wrong solutions. These three models were then used as starting models in a competitive multi-model refinement process against the combined 3,000-particle set (i.e. pooling the three subsets of particles each containing 1,000 particles). Each particle was assigned to the model with best matching score (i.e. highest correlation). This competitive multi-model refinement helped identify the correct model and further evolve the model toward the correct solution. The model winning the most particles (~60%–80% of particles) was selected as the "correct" initial model. As shown by the two independent examples for even and odd datasets, the additional competitive multi-model refinements could effectively identify and evolve significantly divergent models to nearly identical models (ignoring the difference in handedness). The consistent winning models among independent datasets and refinements were considered as the correct PCV2 model for further refinements. See also Figure S2.

Image Processing and 3D Reconstruction: High-Resolution Refinements

The particle parameters from refinements using 4× binned particles were then transferred to 2× binned particles and the original particles without binning for further refinements (Figure 2). Refinement using 2× binned particles improved the resolution to 3.43 Å (Figure S4). Extensive high-resolution refinements were performed using original particle images without binning, which started with refinements of only orientation and center parameters, and then was further extended to also include higher-order parameters (Baker et al., 2013; Guo and Jiang, 2014; Guo et al., 2014) including defocus, astigmatism, beam tilt, overall scale (i.e. magnification), and anisotropic scale (i.e. different magnifications along different directions). These further refinements improved the 3D reconstruction to 2.9 Å resolution (Figures 4, 5B, and S4). Reconstructions with only orientation and center parameters without these high-order parameters were limited at lower (3.29 Å) resolution (Figure 4). Each of these high-order refinement parameters alone only slightly improved the resolution (Figure 4B), but collectively these parameters significantly improved the resolution to 2.9 Å (Figure 4A).

The refined particle views allowed us to analyze the relative view angles among neighboring particles (Figure S5B). We found that the histogram for neighbor particles is very similar to that for control pairs of particles from different micrographs and control pairs of particles with randomized views. The lack of a peak around 0° suggests that there is little synchrony of views among neighboring particles despite the close packing. The close packing thus only constrains the particle position but not the rotational freedom. The slightly higher number of particles pairs (both neighboring particles and particles in different micrographs) at around 10° compared with the particles with randomized views reflects the more clustered views around icosahedral three-fold axes (Figure S5A).

Resolution Evaluation and Validation: Gold Standard FSC, FSC_{true}, and Local Resolutions

In this work, the stated resolutions were evaluated using the gold standard Fourier shell correlation (FSC) = 0.143 criterion (Figures 4, 5B, and S4) (Rosenthal and Henderson, 2003; Scheres and Chen, 2012). However, we adopted an even more stringent, truly independent strategy whereby not only the dataset was split into even and odd halves before iterative refinements (Figure 2) as recommended by the EM Validation Task Force (Henderson et al., 2012), but also the initial models were independently built de novo for each subset instead of using a common low-resolution starting model (Scheres and Chen, 2012). As shown in the overall image processing workflow (Figure 2), initial models were separately constructed de novo for each of the two half datasets, and the subsequent refinements were all limited to the corresponding half datasets without mixing.

To monitor potential resolution inflation that might have been caused by overfitting inadvertently included in the refinements of multiple higher-order parameters, we also performed the noise substitution test (Chen et al., 2013) in parallel with the refinements of original particles. Each of the half datasets was phase-randomized in the resolution range of 5 Å and higher. As shown in Figure 6A, the FSC curve between the two phase-randomized datasets (FSC_{noise}) showed an ignorable correlation beyond 5 Å resolution, suggesting that the refinement of higher-order parameters did not cause overfitting. The FSC_{true} curve was then calculated by subtracting FSC_{noise} from the FSC between original particles using Equation 4 in Chen et al. (2013). The FSC_{true} curve was suggested as a more reliable

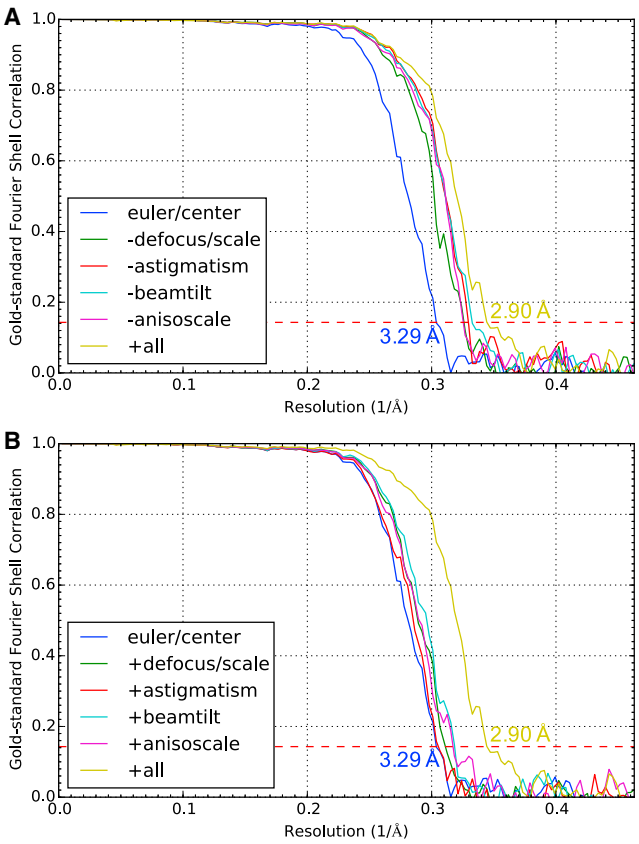


Figure 4. Gold Standard Fourier Shell Correlation Curves for Reconstructions Using Various Combinations of Refinement Parameters

(A) Reconstructions using all (+all) refinement parameters: Euler, center, defocus, scale, astigmatism, beam tilt, and anisotropic scaling; all except one of the parameter sets: without refined defocus and scale (-defocus/scale), without refined astigmatism (-astigmatism), without refined beam tilt (-beamtilt), without refined anisotropic scaling (-anisoscale); and reconstruction with only Euler and center parameters, defocus pre-fitted using power spectra (euler/center).

(B) Similar to (A), except that four of the reconstructions (+defocus/scale, +astigmatism, +beamtilt, +anisoscale) used only the single indicated parameter set in addition to Euler and center parameters.

measurement of authentic resolution of cryo-EM 3D reconstructions. It is apparent that our FSC_{true} curve essentially superimposes on the FSC curve of original particles (Figure 6A). Using the FSC_{true} = 0.143 criterion, our PCV2 structure overall is at 2.9 Å resolution.

The above two FSC curves report the map resolution based on the overall correlation of the entire capsid. To assess the local resolution distribution and further validate our map with another independent resolution measurement, we used the resmap method (Kucukelbir et al., 2014) to calculate the local resolution of our PCV2 map (Figures 6B and 6C). It can be seen that a significant portion of the capsid is colored in the resolution range of 2.4–2.7 Å, slightly better than the overall resolution at 2.9 Å. The outer surface region of the capsid is colored in the resolution range of 3.3–3.6 Å, suggesting that the surface region of the capsid protein structure has more flexible conformations (Figure 7).

Resolution Evaluation and Validation: Correlation with Atomic Models

Structural analysis suggested that the level of details resolved in the density map were consistent with the resolvable features expected at this resolution (2.9 Å) by comparing them with several X-ray crystal structures of viruses (Figure S6). With the main-chain densities easily traceable by visual examination and the clearly visible side-chain densities (Figures 5C, 5D, and S7), an atomic model (aa 42–231) of the capsid protein (Figures 5E and 7A) was manually built using Coot (Emsley et al., 2010) and then refined using the real-space refinement method in Phenix (Adams et al., 2010). The N-terminal 41 residues were not modeled as they were disordered, probably due to lack of interactions with genome in the VLP. FSC analysis indicates that this atomic model agrees with the cryo-EM reconstruction to 3.0 Å resolution at 0.5 cut-off (Rosenthal and Henderson, 2003) (Figure 5B). Extensive model geometry statistics and density fitting scores, including Ramachandran statistics, side-chain rotamers, clash score, MolProbity score (Chen et al., 2010), model/map correlation coefficient, and EMRinger score (Barad et al., 2015), were assessed (Table 1). The excellent statistics suggest that our model is physicochemically accurate and agrees well with the densities. The EMRinger score was recently developed to specifically probe high-resolution (4.5 Å and better) structural features by measuring the agreement of side-chain rotamers in the model with the densities (Barad et al., 2015). Our EMRinger score of 6.5 is exceptionally high when compared with the expected average score of 2.2 at this resolution (2.9 Å) and the best score, 3.26, for the large number of near-atomic resolution cryo-EM map/model pairs (Barad et al., 2015).

The PCV2 structure consists of predominantly β sheets with a jelly-roll fold that is prevalent in virus capsid protein structures (Rossmann and Johnson, 1989). For ~96% (183 of 190) of the residues, the side-chain densities (Figures 5D and S7) were resolved. Characteristic shapes of the side chains of some residues, for example, the benzene ring of Tyr96, branched side chain of Ile98, and bend in Arg116, were resolved (Figures 5D and S7). FSC analysis indicates that our cryo-EM reconstruction agrees with the X-ray structure (PDB: 3R0R) (Khayat et al., 2011) of the N-terminally truncated VLP of a different PCV2 strain to 3.0 Å resolution at 0.5 cut-off (Rosenthal and Henderson, 2003) (Figure 5B), despite additional 41 N-terminal residues and 14 internal different residues in the 233-aa capsid protein forming the particles studied in this work (Figure S8). Our atomic model built independently from the EM map was found to be consistent with this PCV2 crystal structure, with an overall C_α root-mean-square deviation (RMSD) of 0.31 Å. The largest discrepancies were localized mainly in the exposed loops in which most of the different residues between these two PCV2 strains were also localized (Figure S8). The densities in these loops are weaker in both our EM map and the crystal structure (Figures 7D and 7E), suggesting increased conformational flexibilities in these exposed loops, which was independently revealed by the local resolution assessment (Figure 6C).

DISCUSSION

In summary, we demonstrated that single-particle cryo-EM could obtain 2.9 Å resolution reconstructions for close-packed

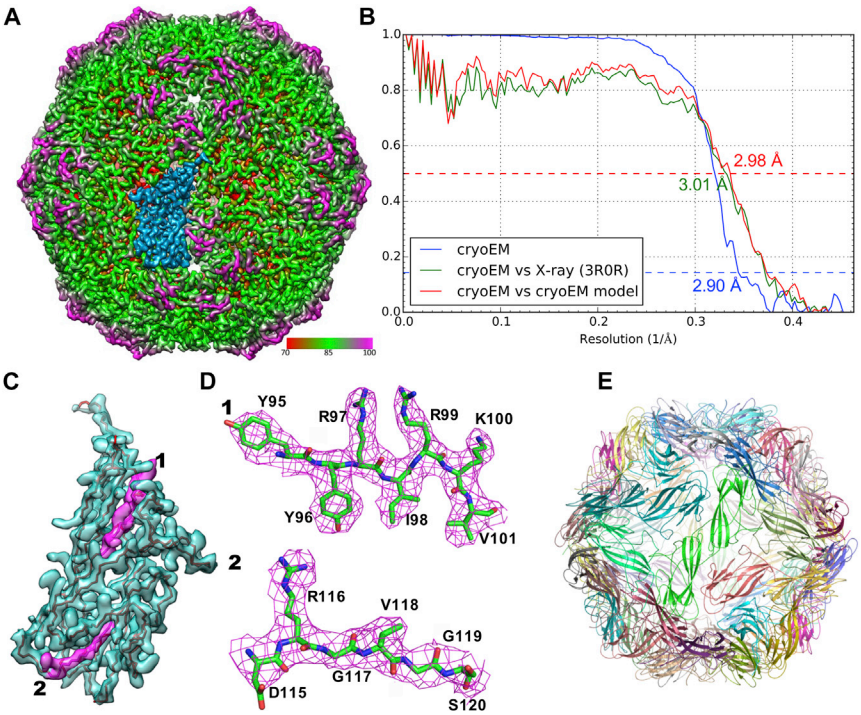


Figure 5. 3D Reconstruction and Atomic Model of PCV2

(A) Radially colored surface view of PCV2 density map at 2.9 Å resolution. An asymmetric unit is highlighted in marine color.

(B) FSC curves. The blue curve is the FSC between two completely independent cryo-EM reconstructions indicating 2.9 Å resolution using 0.143 criterion. The green curve is the FSC between our cryo-EM reconstruction of the whole dataset and the electron density map generated from the X-ray crystallographic model (PDB: 3R0R) indicating 3.01 Å resolution using 0.5 criterion. The red curve is the FSC between our cryo-EM reconstruction and the atomic model derived from this map indicating 2.98 Å resolution using 0.5 criterion.

(C) Cryo-EM density (marine color) of a monomer highlighted in (A) superimposed with its atomic model (only backbone is shown for clear display).

(D) Cryo-EM density (magenta mesh) of two highlighted regions in (C) superimposed with our atomic model (green, blue, and red sticks).

(E) Atomic model of the entire PCV2 VLP.

See also Figures S6 and S7.

particles of PCV2, the smallest virus with a relatively smooth capsid surface. This represents the highest resolution and the only sub-3 Å single-particle cryo-EM structure without using a direct electron detector. This accomplishment was enabled by an image processing strategy with several merits. The particle distributions are significantly relaxed to allow grids with close-packed virus particles, which would significantly improve the rate of usable sample grid preparation and throughput of data collection. The random model approach enhanced with competitive multi-model refinements was able to reliably construct de novo initial 3D models for challenging datasets. The consensus solution of multiple truly independent refinements presents a practical method for identification of reliable particle parameters for large datasets, which would complement the tilt pair method more suited to small datasets (Henderson et al., 2011; Rosenthal and Henderson, 2003). The refinement of high-order parameters beyond Euler/center can further determine or correct residual image imperfections to improve the resolution. The resolution obtained from the early split of dataset and truly independently refined reconstructions conforms to the gold standard that can reliably gauge the quality of single-particle cryo-EM 3D reconstructions (Henderson et al., 2012; Scheres and Chen, 2012). This strategy should be generally applicable to other single-particle cryo-EM projects of viruses, although applicability to non-virus macromolecular complexes needs to be further studied. The observation of minimal-view synchronization among neighbor particles suggests that close packing constrains particle positions but not orientations. However, more systems need to be analyzed to discover how general this observation is.

We note that this ~2.9 Å resolution cryo-EM 3D reconstruction used data with suboptimal particle distribution (i.e. close-packed particles) and detector (i.e. photographic films with consumer grade scanner). This suggests that higher-resolution (2–2.5 Å) structures might be feasible when dispersed particles are imaged on a direct electron detector with dose fractionation and correction of sample movements during imaging (Abrishami et al., 2015; Li et al., 2013; McMullan et al., 2014; Ruskin et al., 2013). The level of structural details in cryo-EM structures at these resolutions (Bartesaghi et al., 2015) would be sufficient to localize the binding site and reveal the detailed interactions of small drug ligands with virus particles to allow structure-based rational design of antivirals.

EXPERIMENTAL PROCEDURES

Sample Preparation

The VLPs of PCV2 Yamagata strain (GenBank: AB426905.1) were prepared using the previously reported method (Liu et al., 2008). In brief, viral DNA was extracted from the PCV2 virion, and the full-length ORF2 was ligated into the transfer vector pVL1393 (Pharmingen). A recombinant baculovirus (AcPCV2-ORF2) was then constructed to express the capsid protein encoded by ORF2. Insect Tn5 cells were infected with AcPCV2-ORF2. The culture medium of the infected Tn5 cells was harvested at 7 days post infection and the PCV2 VLPs were purified by CsCl gradient centrifugation.

Cryo-EM

A 3-μl aliquot of the PCV2 sample at 3 mg/ml concentration was applied to 400-mesh holey carbon grids (1.2/1.3 C-flat, Protochips). After blotting excessive sample solution, the grid was plunge-frozen in liquid ethane cooled by liquid nitrogen using the Cp3 plunge-freezing device (Gatan). Cryo-EM images were taken using an FEI Titan Krios electron cryomicroscope with field emission gun operated at 300 kV and parallel illumination. The images were recorded on Kodak SO163 negative films at low dose (25–27 e/Å^2) and at a nominal magnification of 59,000 with intended defocus range of 1.5–2.5 μm. The films without obvious ice contamination (141 of 168) were selected for digitization using a Nikon Super CoolScan 9000ED scanner with step size of 6.35 μm/pixel. The final calibrated sampling of the scanned images is 1.08 Å/pixel.

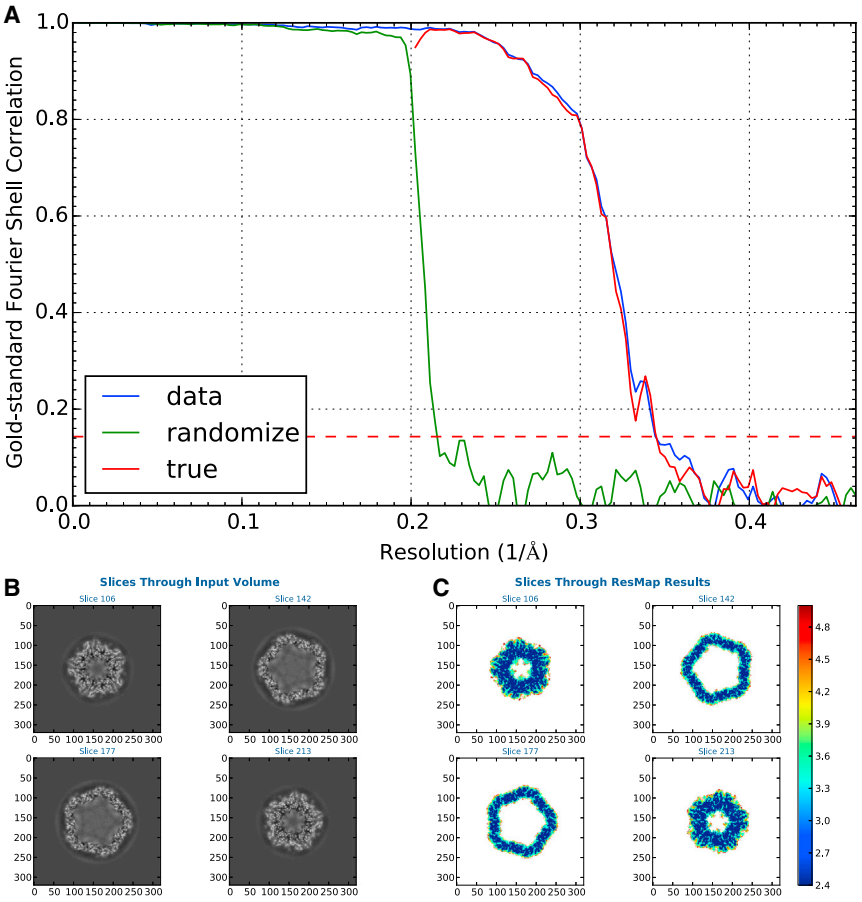


Figure 6. Validation with Phase-Randomized Images and Local Resolution Assessment

(A) The particles were phase randomized in the resolution range from 5 Å to Nyquist and independently refined in parallel with the original images using identical parameters. Shown are the gold standard FSC curves for original images (data), phase-randomized images (randomize), and the true FSC curve computed from these two FSC curves (true).

(B and C) Local resolution evaluation using resmap. Shown are outputs of the resmap software: (B) density sections and (C) resolution distributions in the same set of sections shown in (B).

See also Figure S8.

Image Processing and 3D Reconstruction

Among the 141 scanned micrographs, 55 were discarded due to significant drift/charging after power spectra screening. From the remaining 86 micrographs, 139,871 particles (320 × 320 pixels) were selected using the e2boxer.py program of the EMAN2 software package (Tang et al., 2007). No masking was performed on the particles at this stage. Masking would be performed later on the fly using the best known particle center at the 2D alignment and 3D reconstruction steps of each refinement iteration. The CTF parameters were first automatically fitted using the fitctf2.py program (Jiang et al., 2012) and then verified using the EMAN ctfit program (Ludtke et al., 1999). These fitted CTF parameters were used only as initial parameters that would be further refined in later high-resolution refinements. CTF correction was thus not performed at this stage, but on the fly using the most recently refined parameters during each iteration of 2D alignment and 3D reconstruction. The entire dataset was then divided into two halves (69,936 for the even subset and 69,935 for the odd subset) and all subsequent image processing, including construction of de novo initial models and iterative refinements, was performed on each of the two subsets independently (Figure 2) using our jspr software (Guo and Jiang, 2014) built on EMAN (Ludtke et al., 1999) and EMAN2 (Tang et al., 2007).

Initial Models and Refinements

Construction of initial icosahedral models and initial refinements were performed using 4× binned images. Three independent de novo initial icosahedral models were constructed using a random model approach (Guo and Jiang, 2014; Yan et al., 2007) enhanced with multi-model refinements for each of the subsets (Figures 2, 3, and S2). The particle orientation and center parameters were first determined using a PFT alignment method (Baker and Cheng, 1996) with center search implemented in jspr software (Guo and Jiang, 2014). The radial range for PFT was limited to 25 pixels, which is slightly larger than the particle size. 2D alignment was initially performed using exhaustive projection matching (within icosahedral asymmetric unit) with projections sampled at 3° angular step size, and refined using a gridless projection matching method employing Simplex optimization (Nelder and Mead, 1965). The initial models (three for the even subset and three for the odd subset) were used to independently determine the particle orientation and center parameters to obtain three sets of parameters for each particle. Only particles (27,131 for the even subset and 28,047 for the odd subset) with consistent orientation ($\sigma \leq 2°$) and center ($\sigma \leq 1$ pixel) parameters among the three sets (Figure S3) were kept for further high-resolution refinements.

High-Resolution Refinements and Resolution Evaluation

The particle parameters obtained from initial refinements using 4× binned particles were transferred (i.e. by adjusting center positions by 2× while reusing Euler angles) to 2× binned particles for further refinements (Figure 2). Similarly, the parameters refined using 2× binned particles were then transferred to original particles without binning for high-resolution refinements. During the iterative high-resolution refinements, the 2D alignments were extended beyond particle orientation and center to also include high-order parameters, magnification, defocus, and astigmatism, in jspr as described previously (Baker et al., 2013; Guo and Jiang, 2014; Guo et al., 2014), and beam tilt (Glaeser et al., 2011; Henderson et al., 1986; Zhang and Zhou, 2011) and anisotropic magnification (Grant and Grigorieff, 2015a; Zhao et al., 2015) implemented in jspr more recently. The CTF phase correction (phase flipping and beam tilt phase correction) was performed using the most recently refined values during 2D alignments and 3D reconstructions. After final iteration of refinement, the two completely independently refined 3D reconstructions, one from the even subset and one from the odd subset, were used to compute gold standard FSC to evaluate the resolution at FSC = 0.143 (Henderson et al., 2012; Rosenthal and Henderson, 2003; Scheres and Chen, 2012). The final 3D reconstruction of the whole dataset (50,352 particles) was built by pooling the two subsets of particles without further refinement. The final reconstruction was sharpened by first setting the map structural factor curve to a reference structural factor curve and then low pass-filtered using the FSC-derived curve (C_{ref}) as in the Henderson approach (Rosenthal and Henderson, 2003). The density values in the final density map were normalized by setting the background mean to zero and the background σ to 1, which results in the density values equivalent to signal-to-noise ratios (Guo and Jiang, 2014). The local resolution of final reconstruction was assessed using resmap (Kucukelbir et al., 2014). Each subset of the particles was also subjected to noise substitution by randomizing phases beyond 5 Å using e2proc2d.py with processor filter.lowpass.randomphase in EMAN2 (Tang et al., 2007). The phase-randomized particles were refined in parallel with the original

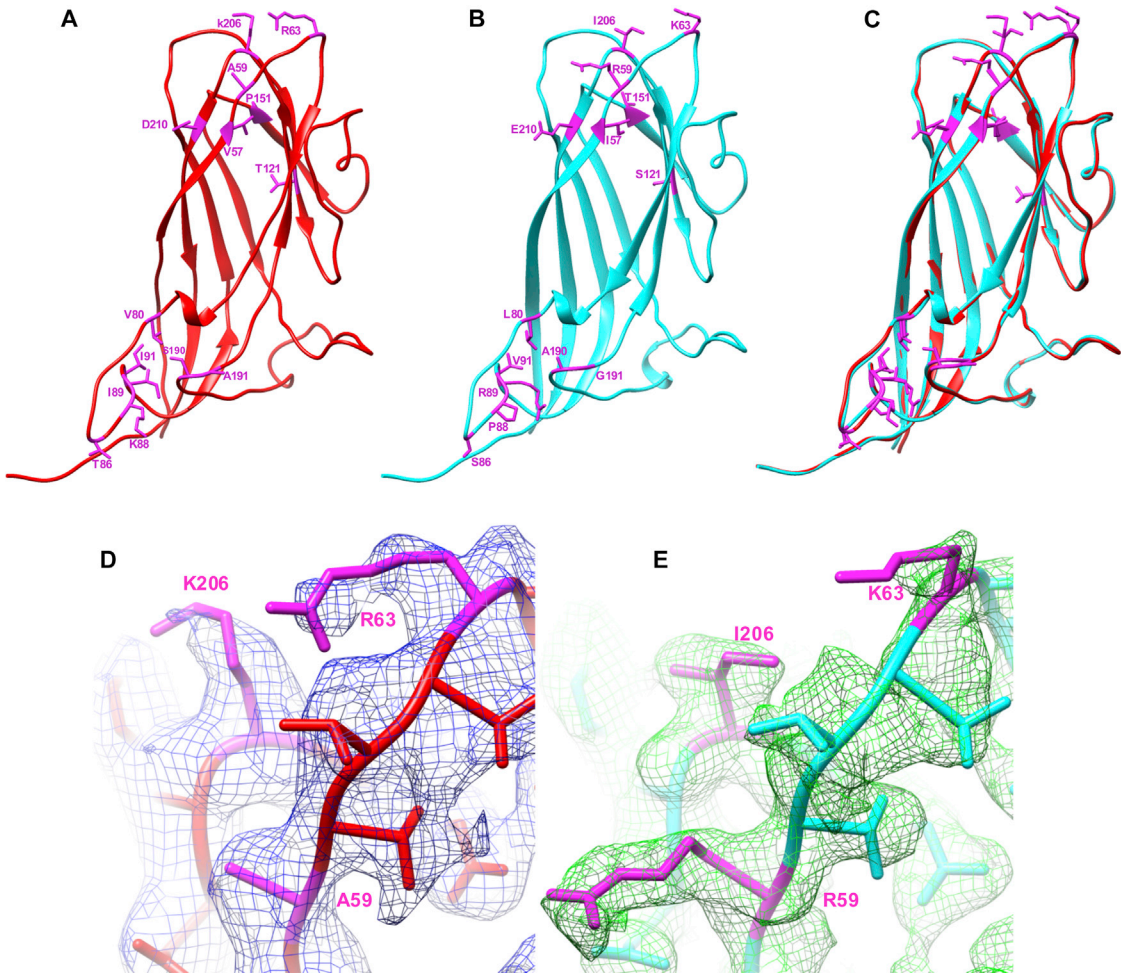


Figure 7. Comparison of PCV2 Atomic Models

(A–C) The atomic model of a single subunit of the PCV2 capsid protein is shown in ribbon diagram for (A) our atomic model, (B) the crystal structure (PDB: 3R0R), and (C) the two models superimposed. The residues different in the two virus strains are labeled in magenta, which are clustered in the loops. The overall C_α RMSD between these two atomic models is 0.31 Å as measured using the Chimera program.

(D and E) Comparison of densities of cryo-EM map (D) and crystal map (PDB: 3R0R) (E) in the flexible loop region. The cryo-EM density map (blue mesh in D) and the $2F_o - F_c$ density map of PDB: 3R0R (green mesh in E, downloaded from the Electron Density Server at http://eds.bmc.uu.se/eds) were superimposed with corresponding atomic models (colored sticks). Although the electron densities in these two loops are the least resolved regions in both density maps, distinct shapes of the side-chain densities could be seen for some of the residues different in the two virus strains. Shown are three pairs of different residues labeled in magenta: A59, R63, and K206 in (D) and R59, K63, and I206 in (E).

See also Figure S8.

particle sets (i.e. a total of four "datasets," original even subset, original odd subset, phase-randomized even subset, and phase-randomized odd subset, were refined simultaneously using identical parameters). The FSC between the phase-randomized subsets and the FSC between the original subsets were then combined to derive the FSC_{true} that would be free of resolution inflation caused by overfitting (Chen et al., 2013). To compute FSC with X-ray crystal structure (PDB: 3R0R) and our atomic model of PCV2 built from the EM density map, electron densities of the atomic models were computed using the e2pdb2mrc.py program in EMAN2 (Tang et al., 2007) from the PDB file with the icosahedral center of PDB: 3R0R adjusted to compensate for the center offset. The Python script written for interparticle angle analysis is included as Script S1.

All these iterative processing tasks were performed using our jspr software (Guo and Jiang, 2014), available on the authors' Web site (http://jiang.bio.purdue.edu/jspr). The graphics for densities were generated using Chimera (Goddard et al., 2007). The FSC curves were plotted using matplotlib software (http://matplotlib.org).

Model Building and Validation

A density block slightly larger than an asymmetric unit was extracted from the entire map to build an initial atomic model of the capsid protein. The density map was well resolved, and the clearly visible side-chain densities made interactive modeling straightforward using Coot (Emsley et al., 2010). By visually tracing the main chain, a Baton model was first constructed and then converted to a main-chain model. The "jelly-roll" topology and side-chain densities along the chain were used as landmarks to compare with the predicted secondary structure elements (mostly β strands) and the long side-chain residues in the protein sequence to assign residues to the model. The N-terminal 41 residues were disordered and thus not modeled. The Coot model (aa 42–231, 190 residues) was further refined using Phenix real-space refinement (program phenix.real_space_refine) (Adams et al., 2010). The final model was validated with good bond length, angles, Ramachandran statistics, clash score, MolProbity score (Chen et al., 2010) (from the program phenix.molprobity), and the new EMRinger score (from the program phenix.emringer). The validation statistics are compiled in Table 1. The graphics for atomic models

Table 1. Validation Statistics for PCV2 Model

Parameter	Values
Phenix Real-Space Refinement	
Map CC (around atoms)	0.828
MolProbity	
Ramachandran favored	97.34%
Ramachandran allowed	2.13%
Ramachandran outliers	0.53%
Rotamer outliers	0.00%
C_β deviations	0
Clashscore	8.15 (81%)
RMSD (bonds)	0.0103
RMSD (angles)	1.15
MolProbity score	1.57 (93%)
EMRinger Score	6.5

were generated using Chimera (Pettersen et al., 2004) and PyMOL (http://www.pymol.org).

ACCESSION NUMBERS

The final density map has been deposited in the Electron Microscopy Databank with accession code EMD-6555. The atomic model has been deposited in the PDB with accession code PDB: 3JCI.

SUPPLEMENTAL INFORMATION

Supplemental Information includes eight figures and one script and can be found with this article online at http://dx.doi.org/10.1016/j.str.2015.12.006.

AUTHOR CONTRIBUTIONS

Z.L. and W.J. conceived and performed experiments, and wrote the manuscript. T.L. provided reagents. F.G. and F.W. provided expertise.

ACKNOWLEDGMENTS

This work was supported by the NIH. The cryo-EM images were taken in the Purdue Cryo-EM Facility, and the Purdue Rosen Center for Advanced Computing provided the computational resource for the 3D reconstructions. We thank Dr. Agustin J. Avila-Sakar and Frank S. Vago for help in preparation of the manuscript.

Received: September 21, 2015
Revised: December 14, 2015
Accepted: December 14, 2015
Published: January 14, 2016

REFERENCES

Abrishami, V., Vargas, J., Li, X., Cheng, Y., Marabini, R., Sorzano, C.O., and Carazo, J.M. (2015). Alignment of direct detection device micrographs using a robust optical flow approach. J. Struct. Biol. *189*, 163–176.

Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., et al. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D Biol. Crystallogr. *66*, 213–221.

Amunts, A., Brown, A., Bai, X.C., Llacer, J.L., Hussain, T., Emsley, P., Long, F., Murshudov, G., Scheres, S.H., and Ramakrishnan, V. (2014). Structure of the yeast mitochondrial large ribosomal subunit. Science *343*, 1485–1489.

Baker, T.S., and Cheng, R.H. (1996). A model-based approach for determining orientations of biological macromolecules imaged by cryoelectron microscopy. J. Struct. Biol. *116*, 120–130.

Baker, M.L., Hryc, C.F., Zhang, Q., Wu, W., Jakana, J., Haase-Pettingell, C., Afonine, P.V., Adams, P.D., King, J.A., Jiang, W., and Chiu, W. (2013). Validated near-atomic resolution structure of bacteriophage epsilon15 derived from cryo-EM and modeling. Proc. Natl. Acad. Sci. USA *110*, 12301–12306.

Barad, B.A., Echols, N., Wang, R.Y., Cheng, Y., DiMaio, F., Adams, P.D., and Fraser, J.S. (2015). EMRinger: side chain-directed model and map validation for 3D cryo-electron microscopy. Nat. Methods *12*, 943–946.

Bartesaghi, A., Merk, A., Banerjee, S., Matthies, D., Wu, X., Milne, J.L., and Subramaniam, S. (2015). 2.2 A resolution cryo-EM structure of beta-galactosidase in complex with a cell-permeant inhibitor. Science *348*, 1147–1151.

Campbell, M.G., Veesler, D., Cheng, A., Potter, C.S., and Carragher, B. (2015). 2.8 A resolution reconstruction of the *Thermoplasma acidophilum* 20S proteasome using cryo-electron microscopy. Elife *4*, e06380.

Chen, D.H., Song, J.L., Chuang, D.T., Chiu, W., and Ludtke, S.J. (2006). An expanded conformation of single-ring GroEL-GroES complex encapsulates an 86 kDa substrate. Structure *14*, 1711–1722.

Chen, V.B., Arendall, W.B., 3rd, Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr. D Biol. Crystallogr. *66*, 12–21.

Chen, S., McMullan, G., Faruqi, A.R., Murshudov, G.N., Short, J.M., Scheres, S.H., and Henderson, R. (2013). High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy. Ultramicroscopy *135*, 24–35.

Crowther, R.A., Berriman, J.A., Curran, W.L., Allan, G.M., and Todd, D. (2003). Comparison of the structures of three circoviruses: chicken anemia virus, porcine circovirus type 2, and beak and feather disease virus. J. Virol. *77*, 13036–13041.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta Crystallogr. D Biol. Crystallogr. *66*, 486–501.

Finsterbusch, T., and Mankertz, A. (2009). Porcine circoviruses—small but powerful. Virus Res. *143*, 177–183.

Frank, J. (2006). Three-Dimensional Electron Microscopy of Macromolecular Assemblies: Visualization of Biological Molecules in Their Native State, Second Edition (Oxford University Press).

Glaeser, R.M., Typke, D., Tiemeijer, P.C., Pulokas, J., and Cheng, A. (2011). Precise beam-tilt alignment and collimation are required to minimize the phase error associated with coma in high-resolution cryo-EM. J. Struct. Biol. *174*, 1–10.

Goddard, T.D., Huang, C.C., and Ferrin, T.E. (2007). Visualizing density maps with UCSF Chimera. J. Struct. Biol. *157*, 281–287.

Grant, T., and Grigorieff, N. (2015a). Automatic estimation and correction of anisotropic magnification distortion in electron microscopes. J. Struct. Biol. *192*, 204–208.

Grant, T., and Grigorieff, N. (2015b). Measuring the optimal exposure for single particle cryo-EM using a 2.6 A reconstruction of rotavirus VP6. Elife *4*, e06980.

Grassucci, R.A., Taylor, D., and Frank, J. (2008). Visualization of macromolecular complexes using cryo-electron microscopy with FEI Tecnai transmission electron microscopes. Nat. Protoc. *3*, 330–339.

Grigorieff, N., and Harrison, S.C. (2011). Near-atomic resolution reconstructions of icosahedral viruses from electron cryo-microscopy. Curr. Opin. Struct. Biol. *21*, 265–273.

Guo, F., and Jiang, W. (2014). Single particle cryo-electron microscopy and 3-D reconstruction of viruses. Methods Mol. Biol. *1117*, 401–443.

Guo, F., Liu, Z., Fang, P.A., Zhang, Q., Wright, E.T., Wu, W., Zhang, C., Vago, F., Ren, Y., Jakana, J., et al. (2014). Capsid expansion mechanism of bacteriophage T7 revealed by multistate atomic models derived from cryo-EM reconstructions. Proc. Natl. Acad. Sci. USA *111*, E4606–E4614.

Henderson, R., Baldwin, J.M., Downing, K.H., Lepault, J., and Zemlin, F. (1986). Structure of purple membrane from *Halobacterium halobium*—recording, measurement and evaluation of electron-micrographs at 3.5 Å resolution. Ultramicroscopy *19*, 147–178.

Henderson, R., Chen, S., Chen, J.Z., Grigorieff, N., Passmore, L.A., Ciccarelli, L., Rubinstein, J.L., Crowther, R.A., Stewart, P.L., and Rosenthal, P.B. (2011). Tilt-pair analysis of images from a range of different specimens in single-particle electron cryomicroscopy. J. Mol. Biol. *413*, 1028–1046.

Henderson, R., Sali, A., Baker, M.L., Carragher, B., Devkota, B., Downing, K.H., Egelman, E.H., Feng, Z., Frank, J., Grigorieff, N., et al. (2012). Outcome of the first electron microscopy validation task force meeting. Structure *20*, 205–214.

Jiang, W., Baker, M.L., Jakana, J., Weigele, P.R., King, J., and Chiu, W. (2008). Backbone structure of the infectious epsilon15 virus capsid revealed by electron cryomicroscopy. Nature *451*, 1130–1134.

Jiang, W., Guo, F., and Liu, Z. (2012). A graph theory method for determination of cryo-EM image focuses. J. Struct. Biol. *180*, 343–351.

Khayat, R., Brunn, N., Speir, J.A., Hardham, J.M., Ankenbauer, R.G., Schneemann, A., and Johnson, J.E. (2011). The 2.3-angstrom structure of porcine circovirus 2. J. Virol. *85*, 7856–7862.

Kucukelbir, A., Sigworth, F.J., and Tagare, H.D. (2014). Quantifying the local resolution of cryo-EM density maps. Nat. Methods *11*, 63–65.

Leschziner, A.E., and Nogales, E. (2007). Visualizing flexibility at molecular resolution: analysis of heterogeneity in single-particle electron microscopy reconstructions. Annu. Rev. Biophys. Biomol. Struct. *36*, 43–62.

Li, X., Mooney, P., Zheng, S., Booth, C.R., Braunfeld, M.B., Gubbens, S., Agard, D.A., and Cheng, Y. (2013). Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. Nat. Methods *10*, 584–590.

Liao, M., Cao, E., Julius, D., and Cheng, Y. (2013). Structure of the TRPV1 ion channel determined by electron cryo-microscopy. Nature *504*, 107–112.

Liu, X., Jiang, W., Jakana, J., and Chiu, W. (2007). Averaging tens to hundreds of icosahedral particle images to resolve protein secondary structure elements using a Multi-Path Simulated Annealing optimization algorithm. J. Struct. Biol. *160*, 11–27.

Liu, L.J., Suzuki, T., Tsunemitsu, H., Kataoka, M., Ngata, N., Takeda, N., Wakita, T., Miyamura, T., and Li, T.C. (2008). Efficient production of type 2 porcine circovirus-like particles by a recombinant baculovirus. Arch. Virol. *153*, 2291–2295.

Liu, H., Jin, L., Koh, S.B., Atanasov, I., Schein, S., Wu, L., and Zhou, Z.H. (2010). Atomic structure of human adenovirus by cryo-EM reveals interactions among protein networks. Science *329*, 1038–1043.

Ludtke, S.J., Baldwin, P.R., and Chiu, W. (1999). EMAN: semiautomated software for high-resolution single-particle reconstructions. J. Struct. Biol. *128*, 82–97.

Ludtke, S.J., Baker, M.L., Chen, D.H., Song, J.L., Chuang, D.T., and Chiu, W. (2008). De novo backbone trace of GroEL from single particle electron cryomicroscopy. Structure *16*, 441–448.

McMullan, G., Faruqi, A.R., Clare, D., and Henderson, R. (2014). Comparison of optimal performance at 300keV of three direct electron detectors for use in low dose electron microscopy. Ultramicroscopy *147*, 156–163.

Nelder, J.A., and Mead, R. (1965). A simplex-method for function minimization. Comput. J. *7*, 308–313.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. J. Comput. Chem. *25*, 1605–1612.

Rosenthal, P.B., and Henderson, R. (2003). Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. J. Mol. Biol. *333*, 721–745.

Rossmann, M.G., and Johnson, J.E. (1989). Icosahedral RNA virus structure. Annu. Rev. Biochem. *58*, 533–573.

Ruskin, R.S., Yu, Z., and Grigorieff, N. (2013). Quantitative characterization of electron detectors for transmission electron microscopy. J. Struct. Biol. *184*, 385–393.

Scheres, S.H., and Chen, S. (2012). Prevention of overfitting in cryo-EM structure determination. Nat. Methods *9*, 853–854.

Scheres, S.H., Gao, H., Valle, M., Herman, G.T., Eggermont, P.P., Frank, J., and Carazo, J.M. (2007). Disentangling conformational states of macromolecules in 3D-EM through likelihood optimization. Nat. Methods *4*, 27–29.

Spahn, C.M., and Penczek, P.A. (2009). Exploring conformational modes of macromolecular assemblies by multiparticle cryo-EM. Curr. Opin. Struct. Biol. *19*, 623–631.

Tang, G., Peng, L., Baldwin, P.R., Mann, D.S., Jiang, W., Rees, I., and Ludtke, S.J. (2007). EMAN2: an extensible image processing suite for electron microscopy. J. Struct. Biol. *157*, 38–46.

Yan, X., Dryden, K.A., Tang, J., and Baker, T.S. (2007). Ab initio random model method facilitates 3D reconstruction of icosahedral particles. J. Struct. Biol. *157*, 211–225.

Yu, X., Jiang, J., Sun, J., and Zhou, Z.H. (2015). A putative ATPase mediates RNA transcription and capping in a dsRNA virus. Elife *4*, e07901.

Zhang, X., and Zhou, Z.H. (2011). Limiting factors in atomic resolution cryo electron microscopy: no simple tricks. J. Struct. Biol. *175*, 253–263.

Zhang, X., Settembre, E., Xu, C., Dormitzer, P.R., Bellamy, R., Harrison, S.C., and Grigorieff, N. (2008). Near-atomic resolution using electron cryomicroscopy and single-particle reconstruction. Proc. Natl. Acad. Sci. USA *105*, 1867–1872.

Zhang, X., Jin, L., Fang, Q., Hui, W.H., and Zhou, Z.H. (2010). 3.3 A cryo-EM structure of a nonenveloped virus reveals a priming mechanism for cell entry. Cell *141*, 472–482.

Zhao, J., Brubaker, M.A., Benlekbir, S., and Rubinstein, J.L. (2015). Description and comparison of algorithms for correcting anisotropic magnification in cryo-EM images. J. Struct. Biol. *192*, 209–215.

Zhou, Z.H. (2011). Atomic resolution cryo electron microscopy of macromolecular complexes. Adv. Protein Chem. Struct. Biol. *82*, 1–35.

