

Accurate model annotation of a near-atomic resolution cryo-EM map

Corey F. Hryc^{a,1}, Dong-Hua Chen^{b,1,2}, Pavel V. Afonine^c, Joanita Jakana^b, Zhao Wang^b, Cameron Haase-Pettingell^d, Wen Jiang^e, Paul D. Adams^c, Jonathan A. King^d, Michael F. Schmid^{a,b}, and Wah Chiu^{a,b,3}

^aGraduate Program in Structural and Computational Biology and Molecular Biophysics, Baylor College of Medicine, Houston, TX 77030; ^bNational Center for Macromolecular Imaging, Verna and Marrs McLean Department of Biochemistry and Molecular Biology, Baylor College of Medicine, Houston, TX 77030; ^cMolecular Biophysics and Integrated Bioimaging Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720; ^dDepartment of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139; and ^eDepartment of Biological Sciences, Purdue University, West Lafayette, IN 47907

Contributed by Wah Chiu, February 2, 2017 (sent for review December 7, 2016; reviewed by Terje Dokland and Jack E. Johnson)

Electron cryomicroscopy (cryo-EM) has been used to determine the atomic coordinates (models) from density maps of biological assemblies. These models can be assessed by their overall fit to the experimental data and stereochemical information. However, these models do not annotate the actual density values of the atoms nor their positional uncertainty. Here, we introduce a computational procedure to derive an atomic model from a cryo-EM map with annotated metadata. The accuracy of such a model is validated by a faithful replication of the experimental cryo-EM map computed using the coordinates and associated metadata. The functional interpretation of any structural features in the model and its utilization for future studies can be made in the context of its measure of uncertainty. We applied this protocol to the 3.3-Å map of the mature P22 bacteriophage capsid, a large and complex macromolecular assembly. With this protocol, we identify and annotate previously undescribed molecular interactions between capsid subunits that are crucial to maintain stability in the absence of cementing proteins or cross-linking, as occur in other bacteriophages.

cryo-EM | P22 | model | structure | annotation

Recently, cryo-EM maps with associated atomic coordinates have been reported at resolutions better than 4 Å (1, 2). However, few have been subjected to rigorous evaluation of the reliability of the observed features or of the correlation between the experimental map and its corresponding model at the residue level. Typically, such correlation is reported in terms of a curve known as the Fourier shell correlation (FSC), which is a function of spatial frequency (3, 4). Although informative, it does not assure the authenticity of local features, nor does it indicate which features in the model agree or disagree with observed density. Ideally, a molecular model can be used to generate a map that replicates the experimental map in most or all of its details, and thus constitutes a trustworthy and informative model for the specimen's structure at the reported resolution. This study examines the agreement and/or disagreement between the model and the experimental map density, determined at near-atomic resolution. These efforts establish the groundwork for a quantitative assessment of a cryo-EM structure. The methods described here were applied to the capsid of P22 bacteriophage, which infects *Salmonella* and has been extensively studied through biochemistry, genetics, and biophysics (5–8).

Results

Cryo-EM Images and Reconstructions. To study the P22 structure, we used a 300-keV electron cryomicroscope (JEM-3200FSC; JEOL Ltd.) and a Direct Electron detector (DE-20; operated in integrating mode) to collect frozen, hydrated P22 bacteriophage images (Fig. 1A and Table S1). Signal was detectable out to 3-Å resolution (Fig. S1 and Movie S1). A total exposure of 37.5 e⁻/Å² was fractionated into 24 frames during a 1.5-s exposure. All frames were dose-weighted (4) and used to refine particle orientation

parameters; empirically, we found that using image frames 1 to 6 (a cumulative exposure of ~10 e⁻/Å²) resulted in the best experimental map (Fig. 1B), with a resolution of 3.3 Å (Fig. S1 and Movie S2).

Model Generation. Using this experimental map, a molecular model of the P22 capsid shell was built de novo (Movie S3) and optimized. The well-resolved densities allowed us to easily segment and independently model each of the seven capsid subunits within the asymmetric unit (ASU) of the *T* = 7 icosahedral lattice (Fig. 1C). Side-chain densities of all 33 aromatic residues from each polypeptide were clearly visible and used as anchor points to confirm the sequence registration between the map and the model (Movie S4). All 430 amino acids of the major capsid protein gp5 were modeled and had strong density, except for the first and last amino acids (Met1 and Ala430, respectively), which had disordered density, likely due to flexibility and/or being surface-exposed. The capsid fold is homologous to that

Significance

Electron cryomicroscopy is a rapidly growing field for macromolecular structure determination. We establish a computational protocol to construct a de novo atomic model from a cryo-EM density map, along with associated metadata that describe coordinate uncertainty and the density at each atom. This model faithfully replicates experimental map densities, as evidenced by cross-correlation and other metrics. Our method of annotation will be especially informative for macromolecular assemblies that exhibit resolvability variations in different parts of their structure. This procedure was applied to a 3.3-Å resolution structure of the P22 bacteriophage to delineate interactions that stabilize the neighboring subunits in a *T* = 7 icosahedral capsid.

Author contributions: J.A.K., M.F.S., and W.C. designed research; C.F.H., D.-H.C., J.J., and Z.W. performed research; C.F.H., P.V.A., C.H.-P., W.J., and M.F.S. contributed new reagents/analytic tools; C.F.H., D.-H.C., J.A.K., and M.F.S. analyzed data; and C.F.H., D.-H.C., W.J., P.D.A., J.A.K., M.F.S., and W.C. wrote the paper.

Reviewers: T.D., University of Alabama at Birmingham; and J.E.J., The Scripps Research Institute.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

Data deposition: The cryo-EM maps and models of the P22 capsid reported in this paper have been deposited in the Electron Microscopy Data Bank (EMDB) (accession no. EMD-8606) and Protein Data Bank (ID code 5UU5). The complete EM dataset of boxed-out particles used for the final 3D reconstruction is available for download through the EMDB Electron Microscopy Pilot Image Archive (EMPIAR), www.ebi.ac.uk/pdbe/emdb/empiar/ (accession no. EMPIAR-10083).

¹C.F.H. and D.-H.C. contributed equally to this work.

²Present address: Department of Structural Biology, Stanford University, Stanford, CA 94305.

³To whom correspondence should be addressed. Email: wah@bcm.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1621152114/-DCSupplemental.

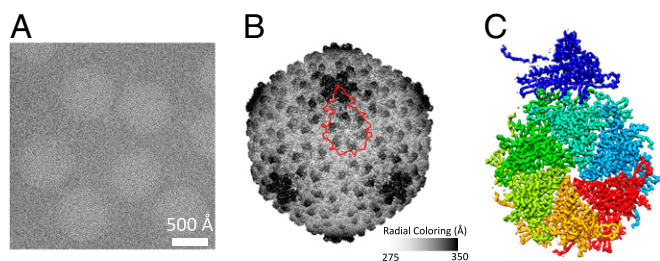


Fig. 1. Cryo-EM data and map. (A) Micrograph of P22 mature virion particles after motion correction and radiation damage compensation. (B) Complete 3.3-Å density map with an asymmetric unit outlined in red. (C) An asymmetric unit from the cryo-EM density map has been segmented from the complete capsid, and the seven individual capsid proteins comprising the asymmetric unit are colored differently.

of the classic HK97 phage capsid protein (9). Each subunit is composed of distinct domains (Fig. 2A), as previously suggested from other phage studies (10–12).

The model optimization process (Movie S5) was performed iteratively, building up the complex from one subunit (Fig. 2A), to one asymmetric unit, and finally an asymmetric unit surrounded by the adjacent subunits in neighboring asymmetric units (24 capsid protein subunits in all) to include all interaction interfaces. The model for each of the seven subunits in an asymmetric unit is slightly different, particularly in the N termini and some loop regions (Fig. 2B and Fig. S2). Inclusion of the surrounding subunits was critical to ensure accurate modeling without clashes at the molecular interfaces, both within and across asymmetric units. A final measure of model quality derived from the full image dataset was based on MolProbity (13), which indicated excellent stereochemistry of the model, the cross-correlation between model and map, and the EMRinger score (14), revealing good side-chain placement within the density (Table S1). Model validation was further performed by building independent versions of the seven models from the two independent maps that resulted from the gold-standard protocol (4) (Fig. 2C). The resulting root-mean-square deviation (RMSD) of 0.56 Å between the C α -atom positions of the two independent model sets indicates the reproducibility of the data (Fig. S3). Unsurprisingly, these two independent models also agree well

with the model built using the full dataset (~ 0.6 Å for both datasets).

Negative/Zero Density. As shown in Fig. 2A, the modeled atoms' locations and the experimental map appear at first glance to be in excellent agreement. However, upon close inspection, densities were not always in agreement with the atomic coordinates that were placed into the experimental map (e.g., Arg109 or Glu116 in the spine helix). This discrepancy prompted us to examine the actual density values of all the atoms in more detail (Fig. 3 and Fig. S4A). Strong positive density was apparent for almost all side chains and provided experimental constraints for the model. However, the side chains of negatively charged amino acids typically did not show this strong positive density, a phenomenon that has been previously documented (15–17). We observed that for aspartate and glutamate, atoms beyond the C β -atom had the weakest “positive” density values. Furthermore, a large proportion of their carboxyl oxygen atoms had density values close to zero or, indeed, even negative at some locations. Map values for negatively charged residues (364 per asymmetric unit, representing 12% of all residues) were statistically significant ($P < 0.01$) for their consistently weak-to-negative densities, except compared with each other (Fig. S4B). Negative densities tended to preferentially cluster and surround the negatively charged side chains in a hemispherical shell (Fig. 4A and B). Such density features were reproducibly observed in both the even and odd maps (Fig. S5). In contrast, although negative density could also be found around other types of amino acids, this density appeared to be randomly distributed; that is, no consistent trend in the distribution of negative density was found around either positively charged (Fig. 3) or neutral amino acid residues.

It has been suggested that the negatively charged residues might be more sensitive to electron radiation damage (16, 18) or have inherently different scattering properties in different local ionic environments (15, 17). To further emphasize the importance of the ionic environment, it has been shown that electron scattering differs between free atoms and chemically bonded atoms for inorganic compounds (19). Therefore, the surrounding environment of various atoms may play a role in the density values found around the residues in question. However, regardless of the physical cause of the observed weak and/or negative densities around the negatively charged residues, the modeling of such residues must properly account for it.

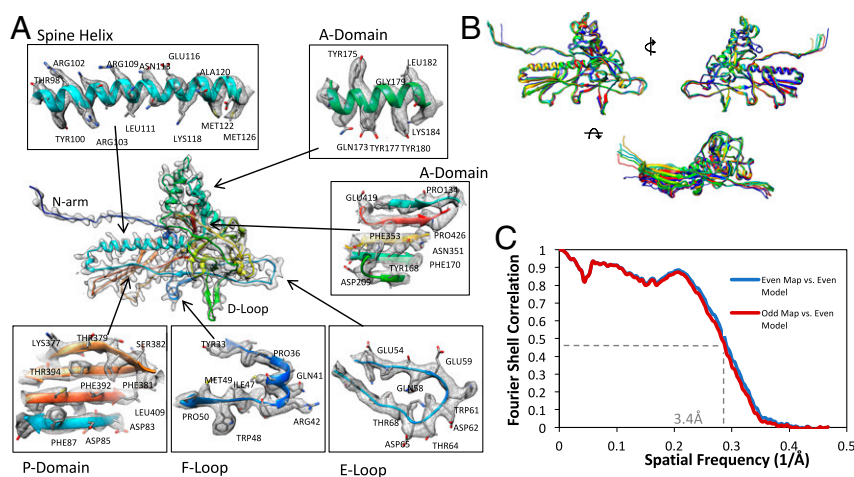


Fig. 2. Cryo-EM map-derived models and model validation. (A) Domains from a hexon subunit revealing atomic-level details of the cryo-EM density map and its corresponding molecular model. (B) Overlapping the seven individual models reveals the small nuances and similarities between the capsid proteins. (C) Two FSC curves are computed for the even and odd density maps and the even model. These curves show that overfitting did not occur, as the odd map and even model are slightly worse than the even map and its corresponding model.

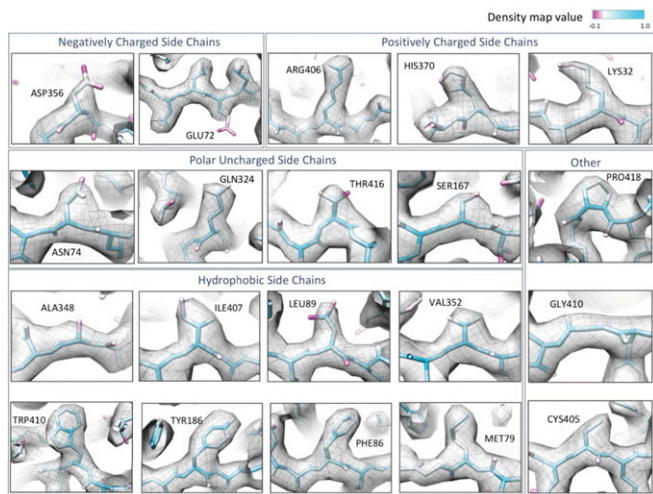


Fig. 3. Density map values for atomic positions for instances of the 20 amino acids. An optimized molecular model is colored by the corresponding map value. The map is rendered at a threshold of 0.22 sigma, which corresponds to white on the model. Atoms that lie in strong density are shown in cyan, whereas weak/negative density is shown in magenta.

Correspondence Between Model and Map. When generating a model-derived map, it has been common practice in cryo-EM to consider all atoms equally weighted (20–22). Because our model strives to be a faithful representation of the original experimental density map, it is important that we correctly account for the map densities present at individual atom locations. To do so, we included atomic displacement parameters (ADPs; also often referred to as B-factors) per atom in the model optimization process (Fig. S6). These ADPs were found to correlate well with

the variation between the two independent models (derived from “odd” and “even” maps) used to determine the level of agreement (Fig. S3). Such correlation of high ADPs with large positional uncertainty of entire loops is demonstrated in the D loop and the N- and C-terminal residues. By including ADPs, we are able to adjust the amount of attenuation and delocalization of the atoms in the model; the optimization routines of Phenix accomplished this, as it does for models optimized against crystallographic data (23).

To account for negative and weak density, ideally highly accurate electron scattering factors should be used. Although Phenix uses electron scattering factors, accurate form factors for all atoms at the electron cryomicroscope’s beam energy (300 keV) are still unavailable. Therefore, we took an entirely empirical approach in our modeling of the cryo-EM density. We can model these negative or weak densities by adjusting the weight of each carboxyl oxygen atom in our model using the Phenix suite (*Materials and Methods*) (24). It should be noted that we only included carboxyl oxygens when assessing negative density, whereas ADPs were assessed for every atom. This was based on our observations of the distribution of negative density (Fig. 4A and B), and the necessary balance between the parameters of weighting and ADPs, to ensure that our calculated map optimally matches the experimental map. The cartoon in Fig. S7A illustrates our interpretation of the calculated density for two atoms, taking into account ADPs and resolution in combination with the appearance of positive/negative/weak density.

The success of this procedure can be seen with a segment of the original experimental map (Fig. 4C), compared with a model-derived density map where all atoms are treated equally (Fig. 4D) and another model-derived map that is calculated from a model with observed densities of the atoms, by using more accurate ADPs and density weighting to account for negative or zero density in the experimental map (Fig. 4E). This proper weighting scheme can now reproduce the negative density around the negatively charged residues (Fig. S7B). This model-

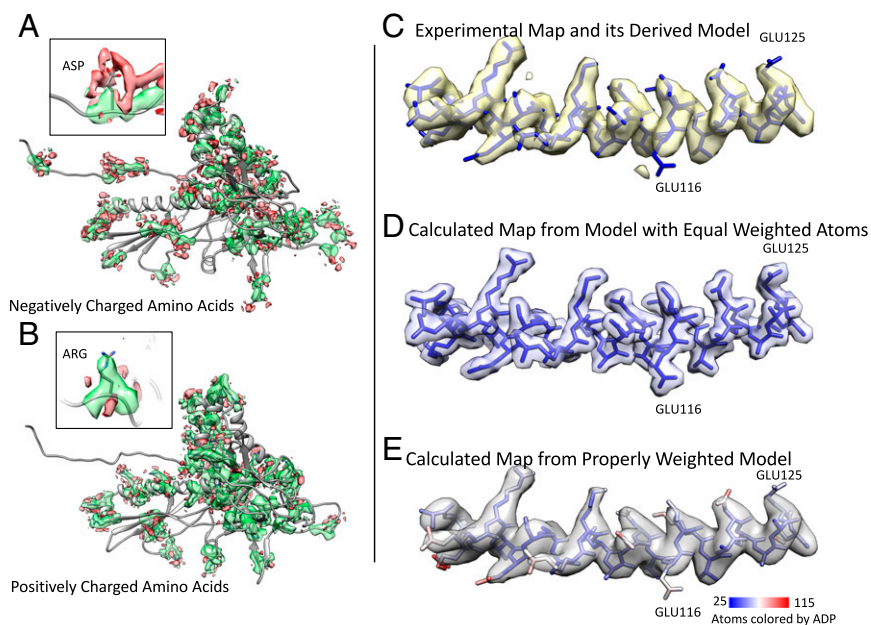


Fig. 4. Assessing the experimental map and corresponding model, and proper representation of the experimental data derived from the molecular model itself. (A) Density surrounding negatively charged amino acids is shown, with green representing strong positive density and red representing weak, negative density. Note the negative cloud-like density surrounding the negatively charged residues. (Inset) A specific Asp residue. (B) Density surrounding positively charged amino acids is highlighted with the same threshold as used in A. (Inset) A specific Arg residue. (C) Experimental map density and model for the spine helix are shown. (D) Currently, when creating a map from a model, all atoms are weighted equally, as shown; however, this is not a proper representation of the experimental density map. (E) The model-derived map of this helix, with proper ADPs and density weights. It faithfully recreates the experimental density map, including uncertainty/weak density in the map, and also negative map values (Fig. S7) that exist at the individual atoms themselves.

derived map places no more and no less density at or near the atomic positions than was present in the experimental map created from the images. The importance of representing the model along with ADPs and density weighting is substantiated by the improvements in FSC between map and model and per amino acid correlations (Fig. S8). This properly weighted model now allows us to correctly represent the experimental data that were used to derive it.

P22 Capsid–Protein Interactions. The resulting molecular model of an asymmetric unit of the $T = 7$ icosahedral lattice of the P22 bacteriophage reveals variations among the six capsid subunits within a hexon (RMSD 0.86 Å for all 430 amino acids) (Fig. 2*B* and Fig. S2). The N arm and the E loop show the highest amount of conformational variability. Furthermore, when comparing the penton subunit with any hexon subunit, a significant amount of conformational variation was observed (RMSD 3.79 Å for all 430 amino acids). This variation is most likely due to subunits at the capsid vertex having different intersubunit interactions. In addition, 1/12 of the pentamers are replaced by a portal protein complex, which affects the densities attributed to the pentamer in this icosahedrally enforced reconstruction (11). The capsid protein domains shown here are generally conserved among all HK97-like bacteriophages, even with large sequence variation (25).

Based on the molecular model, we identified two distinct classes of interactions that stabilize the capsid, the first being the N-terminal arms that extend from one asymmetric unit to its neighbor (Movie S6). N-terminal arms (residues 1 to 24) of every subunit form a β -sheet (residues 3 and 4, and 10 to 12) between adjacent asymmetric units (Fig. 5*A*). In addition, the N-terminal arm (specifically Glu5 and Glu15) also makes contact via salt bridges (Lys31 and Arg42, respectively) with a neighboring subunit in the same asymmetric unit. As shown in our lower-resolution map of the procapsid (11), the N arm structurally rearranges during maturation, and may be one of the driving conformational signals for capsid maturation (26).

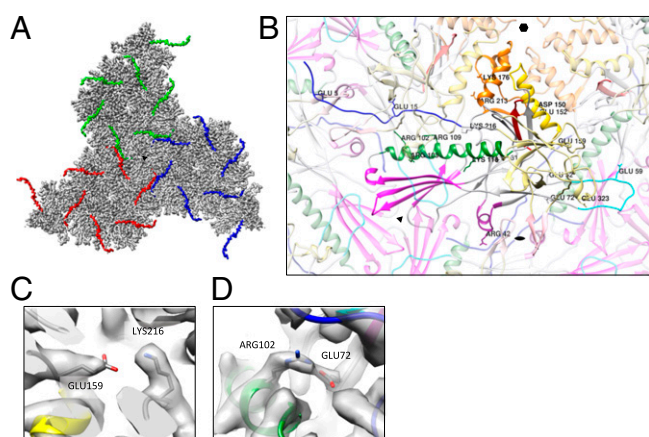


Fig. 5. Capsid–protein interactions critical for capsid stabilization. (A) N-terminal arms from three asymmetric units stretch across to neighboring ASUs. Every one of these N-terminal arms makes an antiparallel β -strand pair with one from a neighboring subunit, even for subunits not shown. All three ASUs are tied together through potential hydrogen bonds using the N-terminal arms at the threefold axis. (B) A vast array of potential salt bridges for one subunit is highlighted. These salt bridges are commonly found between individual subunits and their neighboring subunits. (C) A representative salt bridge between Glu159 and Lys216. This salt bridge is between neighboring subunits at the base of the A domain. (D) Another salt bridge between Arg102, of the spine helix, and Glu72 in the E loop from the neighboring subunit.

The second set of specific interactions is based on potential salt bridges between neighboring subunits within an individual hexon/penton (Fig. 5*B*, Fig. S9, and Movie S7). As shown in our analysis above, the density for negatively charged residues generally has weak or negative density, and thus it is hard to directly visualize such interactions in the density map itself. However, the distances between the nitrogens and oxygens of these side chains range from 2.5 to 4 Å. Moreover, these salt bridges are identified from the model using two distinct structure analysis methods, the PISA server and the PDBSum server (27, 28). The capsid subunits form intimate and extensive interfaces with each other around the sixfold and fivefold axes and across the twofold and threefold axes. We have identified specific residues on adjacent helices (Fig. 5*B*) that may be vital to closing the hexon/penton opening in the capsid maturation process (11). One potential amino acid pairing at this location is Arg213–Glu152, which sits just inside the capsid between neighboring A-domain helices. In addition, another salt bridge was identified between Glu159 and Lys216 at the A-domain base. The Lys density is strong whereas Glu density also shows positive density (Fig. 5*B*), most likely due to its close proximity to the positively charged residue.

The E loop, which stretches over the adjacent subunit, has multiple residues that play a role in stabilizing neighboring subunits. One specific residue–ion pair occurs between the spine helix and the E loop (Arg102–Glu72) (Fig. 5*C*). It is likely that one or both salt bridges involving Arg109–Glu52 and/or Lys118–Glu59 lock the E loop to the neighboring capsid subunit. Moreover, there is another ionic interaction between Arg103 of the spine helix and Glu323 in the small helix in the I domain. This may act as a pin, and possibly aid in the interaction of neighboring subunits. When assessing map values for negatively charged residues participating in a salt bridge, we note that their average side-chain map values are slightly more positive (0.159) compared with negatively charged residues not participating in a salt bridge (0.151). This difference is not statistically significant, but is reasonable, given that negative side chains stay negative, even in a salt bridge (Fig. S7).

Discussion

In complex structures such as virus capsids, a large number of amino acids are in contact with each other. In fact, it is likely that only a subset of these are essential for the assembly of the subunits into the virion. Interestingly, there are no known P22 capsid mutants that correspond to the residues in our observed ion pairs. A large majority of known P22 capsid mutations are conditional and can assemble into viable virions under different experimental conditions (6, 8, 29). For instance, temperature-sensitive mutants can form particles (7). After many decades of failing to identify the residues directly involved in subunit interactions through genetic analysis, the reliably determined model presented here has identified an essential set of residues—forming ion pairs between subunits—needed to build a stable icosahedral shell. These interactions could be partially attributable to maintaining the stability of the P22 capsid particle without cross-linking between capsid proteins or additional cementing proteins.

In addition to producing an atomic model of the P22 bacteriophage, our method of analysis of a cryo-EM map and model provides a protocol for assuring the reliability and correspondence of various structural features down to side-chain density levels. It is commonly observed that cryo-EM maps have variable resolutions in different regions, and thus it is problematic to justify whether the interpretations of structural features in those regions can be made. Our methods of model optimization accompanied with metadata (i.e., ADPs and actual density values) provide a quantitative indication of the level of confidence to interpret and annotate side-chain densities, water

molecules, and ligands in cryo-EM density maps. It is particularly useful to include such information in the Protein Data Bank (PDB)-deposited structures from cryo-EM studies, analogous to the annotated X-ray crystallography data in the PDB. These annotations will have an even greater impact on data that exhibit dynamic changes in resolvability, because the model can now represent less resolved features accurately and explicitly. Furthermore, a potential utility of our density-weighted model is to compute a map for further structure feature validation and possibly new rounds of structure refinement. It should be noted that, in any map, caution should be taken during refinement to ensure that bias is not introduced to invalidate the resulting structure. In practice, multiple models with careful designs could be used to reveal an accurate map of a complex biological assembly.

Finally, the time is ripe for cryo-EM atomic coordinates to be reported in the Worldwide PDB (30) along with a quantitative description of their local map density uncertainties. This allows for the map and the model to be studied together, giving investigators richer information content and a higher level of confidence in its structure assessment and biophysical interpretations.

Materials and Methods

Cryo-EM Imaging and Data Processing. The P22 virion specimen was purified according to an established procedure (11). The cryo-EM data collection and processing are summarized in Table S1.

All of the selected 45,150 particle images were first shrunk by a factor of 4 to a box size of 216×216 to accelerate the data processing at low resolutions. About 2,100 shrunk particle images with the largest defocusses were selected from each subset to build the initial template, again using the program JSPR (31): Five sets of 300 particle images were randomly selected from the highly defocused 2,100 particle images of each subset, and then the global orientation search was performed using JSPR for 20 iterations. The maps from each set were visually examined and one of the converged maps was selected from the last iterations of each subset. This was then used as the initial template for the global orientation search for all particle images shrunk by a factor of four. Several global orientation searches were carried out for the fourfold-shrunk data until the resolutions converged, as judged by the Fourier shell correlation curve of two independent datasets (32, 33) with the best $\sim 11,000$ particles each. The subsequent local orientation determination was performed using data up to a resolution slightly lower than the resolution assessed by the gold-standard FSC 0.143 criterion in the previous iteration, until the resolution had no improvement. The orientations and centers for the data shrunk by a factor of four were then migrated to the full-size (864×864) particle images for additional orientation determination.

It should be noted that the first frame was removed from all the images and that orientation determination was done with all remaining 23 frames. We then experimented with different sets of subframes of the same particle dataset and assessed the density connectivity and resolvability in these different maps. Once complete, we found empirically that using frames 1 to 6 (dose of $\sim 10 \text{ e}^-/\text{\AA}^2$), with both motion and damage corrections, yielded the best-resolved density map (Movie S2) with a resolution of 3.3 \AA based on the gold-standard estimate (32, 33) (Fig. S1). The final reconstruction was from the best $\sim 50\%$ particle images of the total number (Table S1). The amplitude of all cryo-EM density maps for visualization was scaled to the X-ray structure of the bacteriophage HK97 mature capsid (PDB ID code 1OHG) and low-pass-filtered to $\sim 3.0\text{-}\text{\AA}$ resolution.

Atomic Modeling and Optimization. To generate the atomic model, we fit our old model (PDB ID code 2XYZ) into subunit A, specifically the hexon capsid protein that sits across the twofold axis with a penton subunit. To segment the capsid protein, a $30\text{-}\text{\AA}$ color zone in Chimera was used to separate the density (34). This ensured that any alteration in protein fold between the previous and current models would not be missed during the segmentation process. The segmented map was then imported into Coot (35) and amino acid Pro25 was easily identified with a kink, similar to where it was previously located. This residue is located at the end of the N arm and is predicted to be in a small helix that can be identified in the map. From there, baton building was done to the C terminus and back to the N terminus, completing the N arm. Once complete, amino acids were mutated computationally, one at a time and registration errors were adjusted based on the

visible density. All aromatic amino acids were visible and used as anchor points for other amino acids that lacked strong positive density, such as negatively charged residues. The model was then optimized using the density as a constraint using phenix.real_space_refine (23) with default parameters plus simulated annealing to encourage fit to density. Coot was then used to adjust various regions of the model that did not converge into the density, such as a portion of the N arm (weak density) and D loop (containing several negatively charged amino acids with weak side-chain density). Real-space model optimization again followed, this time without simulated annealing.

This model has two domains, the insertion domain and the P domain/N arm (Fig. 2A), that were interpreted differently from previous models derived from a lower-resolution map (11). We expected that improved resolution would alter the protein fold in the insertion domain; however, we did not anticipate any alterations in other regions of the model. When assessing the P domain, we noted improved connectivity of the β -sheets, in addition to a fourth strand, which was previously modeled incorrectly as the N-terminal helix due to poor resolvability. This fourth strand has never been seen in capsid proteins of the bacteriophage, and thus modeling it differently (11) was understandable. When adjusting the threshold, a strand of density extending toward the neighboring twofold axis became visible from our anchor point of Pro25.

The resulting model was placed into the density of the other six subunits in an ASU, and loops with large variations (E loop and a loop in the A domain) were adjusted manually with Coot. Again, real-space model optimization using default parameters. The refined ASU was then surrounded by neighboring asymmetric units, ensuring that, when optimized, clashes were avoided and interactions were optimized. Coot was then used to manually adjust any rotamers or regions with poor geometry. Moreover, phenix.molprobity was run to generate a Coot import file, allowing for the removal of extreme clash and Ramachandran outliers. Finally, a second round of real-space model optimization was completed with simulated annealing and morphing applied. This allowed for a greater freedom of model movement. A final MolProbity check was completed to assess stereochemistry (13). To assess fit to density, cross-correlation was computed during phenix.real_space_refine (4) and an EMRinger was computed (14) (Table S1).

Model to Calculated Map. To generate a weighted map from the model that would represent the experimental density map, both occupancies and ADPs had to be refined against the experimental map. ADPs were first set to 50 \AA^2 for all amino acids in our ASU with surrounding subunits and refined with phenix.real_space_refine (run = adp). Two iterations were done to ensure convergence. Occupancies were then changed to -0.5 for all carboxyl oxygens in the refined complex. This negative value was needed to produce negative density in the calculated map from the model. These occupancies do not refer to the absence of atoms but are used as weighting values due to the lack of a proper form factor. Occupancies were then refined with phenix.refine (23) in reciprocal space, and resulted in some values reverting to positive. An additional iteration of ADPs and occupancy refinement then followed. With atom positions, ADPs, and occupancies all refined, a map was calculated from the model at $3.3\text{-}\text{\AA}$ resolution using phenix.fmodel (23) and converted to a ccp4 format map using phenix.mtz2map.

Model Validation/Uncertainty. Generating two independent models, optimized for the two half-maps (4), is a validation practice that assures that agreement is consistent with the claimed $3.3\text{-}\text{\AA}$ resolution (Fig. 2), as reflected from the FSC 0.5 numerical value suggested previously (32, 33). Moreover, assessing independent models provides an understanding of the level of uncertainty within the map (Fig. S3). Both half-datasets ($\sim 3.4 \text{ \AA}$ in resolution) were modeled independently. Model variation to assess uncertainty was computed in Chimera (34), and the FSC was computed with EMAN2 (36).

ACKNOWLEDGMENTS. We thank Xueming Li at Tsinghua University for the use of his map-sharpening program ampcorrect. We are grateful for the computing resources at the Texas Advanced Computer Center at the University of Texas at Austin, and the Computational and Integrative Biomedical Research Center at Baylor College of Medicine. We thank Drs. Matthew L. Baker and Steve J. Ludtke at Baylor College of Medicine and Dr. Benjamin Bammes at Direct Electron for their helpful discussions. This work was supported by the National Institutes of Health (P41GM103832, R01GM079429, and PN2EY016525) and the Robert Welch Foundation (Q1242). C.F.H. was supported by a predoctoral fellowship under the National Library of Medicine Training Program in Biomedical Informatics (Grant T15LM007093) awarded to the Keck Center of the Gulf Coast Consortia. P.D.A. and P.V.A. acknowledge support by the National Institutes of Health under Grant P01GM063210.

1. Kühlbrandt W (2014) Cryo-EM enters a new era. *eLife* 3:e03678.
2. Henderson R (2015) Overview and future of single particle electron cryomicroscopy. *Arch Biochem Biophys* 581:19–24.
3. Chen S, et al. (2013) High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy. *Ultramicroscopy* 135:24–35.
4. Wang Z, et al. (2014) An atomic model of bromo mosaic virus using direct electron detection and real-space optimization. *Nat Commun* 5:4808.
5. King J, et al. (1976) Structure and assembly of the capsid of bacteriophage P22. *Philos Trans R Soc Lond B Biol Sci* 276(943):37–49.
6. Gordon CL, King J (1993) Temperature-sensitive mutations in the phage P22 coat protein which interfere with polypeptide chain folding. *J Biol Chem* 268(13):9358–9368.
7. Teschke CM, King J (1995) In vitro folding of phage P22 coat protein with amino acid substitutions that confer in vivo temperature sensitivity. *Biochemistry* 34(20):6815–6826.
8. King J, Chiu W (1997) The procapsid-to-capsid transition in double-stranded DNA bacteriophages. *Structural Biology of Viruses*, eds Chiu W, Burnett RM, Garcea RL (Oxford Univ Press, New York), pp 288–311.
9. Wikoff WR, et al. (2000) Topologically linked protein rings in the bacteriophage HK97 capsid. *Science* 289(5487):2129–2133.
10. Helgstrand C, et al. (2003) The refined structure of a protein catenane: The HK97 bacteriophage capsid at 3.44 Å resolution. *J Mol Biol* 334(5):885–899.
11. Chen DH, et al. (2011) Structural basis for scaffolding-mediated assembly and maturation of a dsDNA virus. *Proc Natl Acad Sci USA* 108(4):1355–1360.
12. Rizzo AA, et al. (2014) Multiple functional roles of the accessory I-domain of bacteriophage P22 coat protein revealed by NMR structure and cryoEM modeling. *Structure* 22(6):830–841.
13. Chen VB, et al. (2010) MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* 66(Pt 1):12–21.
14. Barad BA, et al. (2015) EMRinger: Side chain-directed model and map validation for 3D cryo-electron microscopy. *Nat Methods* 12(10):943–946.
15. Mitsuoka K, et al. (1999) The structure of bacteriorhodopsin at 3.0 Å resolution based on electron crystallography: Implication of the charge distribution. *J Mol Biol* 286(3):861–882.
16. Bartesaghi A, Matthies D, Banerjee S, Merk A, Subramaniam S (2014) Structure of β-galactosidase at 3.2-Å resolution obtained by cryo-electron microscopy. *Proc Natl Acad Sci USA* 111(32):11709–11714.
17. Yonekura K, Kato K, Ogasawara M, Tomita M, Toyoshima C (2015) Electron crystallography of ultrathin 3D protein crystals: Atomic model with charges. *Proc Natl Acad Sci USA* 112(11):3368–3373.
18. Glaeser RM (1971) Limitations to significant information in biological electron microscopy as a result of radiation damage. *J Ultrastruct Res* 36(3):466–482.
19. Zhong S, Dadarlat VM, Glaeser RM, Head-Gordon T, Downing KH (2002) Modeling chemical bonding effects for protein electron crystallography: The transferable fragmental electrostatic potential (TFESP) method. *Acta Crystallogr A* 58(Pt 2):162–170.
20. Wu W, et al. (2013) Specificity of an anti-capsid antibody associated with hepatitis B virus-related acute liver failure. *J Struct Biol* 181(1):53–60.
21. Zhang X, et al. (2015) 3D structural fluctuation of IgG1 antibody revealed by individual particle electron tomography. *Sci Rep* 5:9803.
22. Liu Z, Guo F, Wang F, Li TC, Jiang W (2016) 2.9 Å resolution cryo-EM 3D reconstruction of close-packed virus particles. *Structure* 24(2):319–328.
23. Adams PD, et al. (2010) PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66(Pt 2):213–221.
24. Afonine PV, et al. (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D Biol Crystallogr* 68(Pt 4):352–367.
25. Baker ML, et al. (2013) Validated near-atomic resolution structure of bacteriophage epsilon15 derived from cryo-EM and modeling. *Proc Natl Acad Sci USA* 110(30):12301–12306.
26. Guo F, 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(43):E4606–E4614.
27. Krissinel E, Henrick K (2007) Inference of macromolecular assemblies from crystalline state. *J Mol Biol* 372(3):774–797.
28. Laskowski RA, et al. (1997) PDBsum: A web-based database of summaries and analyses of all PDB structures. *Trends Biochem Sci* 22(12):488–490.
29. Teschke CM, McGough A, Thuman-Commike PA (2003) Penton release from P22 heat-expanded capsids suggests importance of stabilizing penton-hexon interactions during capsid maturation. *Biophys J* 84(4):2585–2592.
30. Berman HM, et al. (2000) The Protein Data Bank. *Nucleic Acids Res* 28(1):235–242.
31. Guo F, Jiang W (2014) Single particle cryo-electron microscopy and 3-D reconstruction of viruses. *Methods Mol Biol* 1117:401–443.
32. Henderson R, et al. (2012) Outcome of the first Electron Microscopy Validation Task Force meeting. *Structure* 20(2):205–214.
33. Scheres SH, Chen S (2012) Prevention of overfitting in cryo-EM structure determination. *Nat Methods* 9(9):853–854.
34. Pettersen EF, et al. (2004) UCSF Chimera—A visualization system for exploratory research and analysis. *J Comput Chem* 25(13):1605–1612.
35. Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66(Pt 4):486–501.
36. Tang G, et al. (2007) EMAN2: An extensible image processing suite for electron microscopy. *J Struct Biol* 157(1):38–46.
37. Mindell JA, Grigorieff N (2003) Accurate determination of local defocus and specimen tilt in electron microscopy. *J Struct Biol* 142(3):334–347.
38. Kivioja T, Ravanti J, Verkhovsky A, Ukkonen E, Bamford D (2000) Local average intensity-based method for identifying spherical particles in electron micrographs. *J Struct Biol* 131(2):126–134.
39. Bammes BE, Jakana J, Schmid MF, Chiu W (2010) Radiation damage effects at four specimen temperatures from 4 to 100 K. *J Struct Biol* 169(3):331–341.
40. Liu X, Jiang W, Jakana J, 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(1):11–27.