A Conformational Switch in Bacteriophage P22 Portal Protein Primes Genome Injection

Hongjin Zheng, Adam S. Olia, Melissa Gonen, Simeon Andrews, Gino Cingolani, and Tamir Gonen

INTRODUCTION

The icosahedral capsid represents a formidable example of a biological nanocage that withstands tremendous internal pressures generated by the tightly packaged viral genome (Chemla et al., 2005; Ivanovska et al., 2004; Johnson and Chiu, 2007). In many bacteriophages and herpesviruses, the capsid displays extraordinary structural plasticity and undergoes dramatic conformational changes during viral maturation (Jiang et al., 2003; Johnson and Chiu, 2007; Knopf, 2000; Mettenleiter et al., 2006; Steven et al., 1997; Zhang et al., 2000). In the morphogenesis of the double-stranded DNA (dsDNA) bacteriophage P22, a well-characterized member of the Podovirus family, the coat protein (gp5), together with scaffolding (gp8), portal protein (gp1), and injection proteins (gp7, gp17, gp20), self-assembles to form a 600 Å spherical intermediate known as the procapsid (Israel, 1977; Jiang et al., 2003; Johnson and Chiu, 2007; Thuman-Commike et al., 1996; Zhang et al., 2000). The portal protein occupies a unique vertex of the icosahedron, where it forms a gateway through the proteinaceous capsid as well as the attachment site for external tail accessory factors. The terminase complex (gp2, gp3) temporarily associates with portal protein to form the “genome packaging motor,” a powerful molecular machine that uses ATP to package dsDNA into the capsid (Casjens and Huang, 1982; Jackson et al., 1982). While the viral genome is packaged inside the virion (Casjens and Huang, 1982; Earnshaw and Harrison, 1977; Earnshaw et al., 1978; Jackson et al., 1982), the capsid volume increases 2-fold through major conformational changes in coat proteins (Jiang et al., 2003; Thuman-Commike et al., 1996; Zhang et al., 2000) and becomes an angularized icosahedron more than 700 Å in diameter (Chang et al., 2006). Finally, the terminase complex dissociates from portal protein, and the “genome injection machinery” begins to assemble in preparation for a new cycle of viral infection (Johnson and Chiu, 2007).

During infection, the viral genome is injected through the portal protein channel into the host. In P22, the genome injection machinery is a 2.8 mDa molecular machine composed of a dodecamer of gp1 (Chang et al., 2006; Lander et al., 2006) forming the portal protein, in complex with several copies of tail accessory factors gp4, gp10, and gp26, which assemble sequentially onto portal protein in this order (Bhardwaj et al., 2007; Strauss and King, 1984). This is followed by the attachment of the tailspike adhesive gp9, which completes the assembly of the genome injection machinery, rendering the virus fully infectious (Chang et al., 2006; Johnson and Chiu, 2007; Lander et al., 2006).

Using electron cryomicroscopy, we show that portal protein undergoes dramatic quaternary structure rearrangements during virus maturation. We propose that, similarly to the coat protein, the portal protein adopts a procapsid conformation and a mature phage conformation. The mature phage conformation, which is induced by gp4 binding, enables portal protein to bind viral DNA and likely recruits the injection proteins inside the virion in preparation for a new cycle of viral infection.

RESULTS AND DISCUSSION

Procapsid Form of Dodecameric Portal Protein

Portal protein from bacteriophage P22 is significantly larger than in most phages (~1 mDa) and is therefore more similar to portals found in the herpesvirus family (Newcomb et al., 2001). Ectopically expressed P22 portal protein is polymorphic in solution, consisting of rings of 11- and 12-fold symmetry (Cingolani et al., 2002; Lander et al., 2006; Orlova et al., 2003; Poliakov
properties, and biochemical binding activity identical to the full-length portal (Figure S2). Native mass spectrometry confirms that portalh202 assembles exclusively into stable 12-mers (Lorenzen et al., 2008). In vivo, the carboxy-terminal end of portal protein is dispensable for oligomerization (Lorenzen et al., 2008), assembly into the virion, and virus infectivity (Bazinet et al., 1990). A 3D reconstruction of portalh202 is shown in Figure 2. The quaternary structure of dodecameric portalh202 resembles a hollow mushroom, 155 Å in diameter and 110 Å tall with a central pore 35 Å in diameter, large enough to accommodate fully hydrated dsDNA. Similar overall architecture has been reported for portal proteins from other viruses such as herpesvirus, Phi 29, SPP1, and T7 (Agirrezabala et al., 2005; Guasch et al., 1998; Orlova et al., 2003; Simpson et al., 2000; Trus et al., 2004). Three distinct domains are identified in side view, annotated from top to bottom as the collar, crown, and funnel domains (Figure 2A). The “top” of the portal protein is the region that resides inside the virion (Chang et al., 2006; Lander et al., 2006), while the “bottom” is the region that protrudes outwards. Directly below the collar domain, the crown domain consists of 12 clearly defined spokes emanating from the central DNA channel in portalh202 (Figure S1), which appear truncated in portalh202 (Figures 1A, 1B, and 2C and Figure S1E). The spokes have a periodicity of 30° and adopt a slightly anticlockwise orientation when viewed from the top (Figure S1). The funnel domain is 45 Å long, is 80 Å in diameter, and is situated directly below the crown domain. It protrudes away from the viral capsid and forms the only possible interface between portal protein and adaptor proteins such as the terminase complex or the tail accessory factor gp4 (Chang et al., 2006; Johnson and Chiu, 2007; Lander et al., 2006).
**Mature Capsid Form of Dodecameric Portal\(_{602}\) in Complex with gp4**

The funnel domain of the portal protein protrudes out from the viral capsid, leaving it as the only possible binding site for gp4 (Chang et al., 2006; Lander et al., 2006). In our cryomicroscopy reconstructions, gp4 binds to the funnel domain of portal\(_{602}\) in a one-to-one ratio (Figure S2), in full agreement with previously published data (Olia et al., 2006). The density corresponding to gp4 is represented by two 12-fold symmetric concentric rings, which are consistent with a total mass of ~220 kDa and correspond to 12 copies of gp4 (Figure 3A, dashed box). In the gp4-bound form, portal\(_{602}\) undergoes dramatic conformational changes spanning the entire length of the channel (Figure 3A). The funnel domain increases in diameter from 85 to 110 Å while the crown domain reconfigures, now displaying 12 well-defined spokes (Figures 3B and 3C). The collar domain decreases its diameter from 85 to 70 Å in the portal\(_{602}\)-gp4 complex and exposes 12 elongated rod-like densities pointing straight up toward a large dome, 40 Å in diameter, that is situated directly above the DNA channel (Figure 3A).

The density of the dome domain may originate from regions of portal\(_{602}\) that are disordered in the procapsid form of portal protein, and therefore invisible in our 3D reconstructions, but become stabilized following gp4 binding. Although the density

**Figure 2. Structure of 12-Fold Symmetric Portal\(_{602}\) at 10 Å Resolution**

Electron cryomicroscopy structure of portal\(_{602}\) in the procapsid form. For all figures, we define the “top” as the region of portal that would reside inside the P22 virion, while the “bottom” is the region that would protrude outward.

(A) In side view, the portal assembly adopts a mushroom-like conformation of about 110 Å in height, which presents three domains, annotated from top to bottom as the collar, crown, and funnel domains.

(B and C) Bottom and top views of dodecameric portal\(_{602}\) reveal a central pore large enough to accommodate hydrated dsDNA. The outer diameters of the crown and funnel domains are approximately 155 and 80 Å, respectively. In all panels, the density map was normalized using MAPMAN (Kleywegt and Jones, 1996) and displayed at a contour level of 4.6σ above background corresponding to the molecular mass as calculated in SPIDER (Frank et al., 1996).

**Figure 3. Structure of 12-Fold Symmetric Portal\(_{602}\)-gp4 Complex at 10 Å Resolution**

Electron cryomicroscopy structure of portal\(_{602}\) bound to twelve copies of gp4.

(A) (Side View) The portal\(_{602}\)-gp4 complex is 160 Å in height. The gp4 density is visible directly below the funnel domain and contributes 30 Å to the total height of the complex (enlarged box). A large dome-shaped domain is seen above the collar domain, directly over the pore of the DNA channel.

(B and C) Bottom and top views of portal\(_{602}\)-gp4 complex reveal that, following gp4 binding, the central DNA channel remains hollow but is occluded by a large dome-shaped domain at the very top. The overall diameter of the crown domain remains unchanged from portal\(_{602}\) at 155 Å but now presents 12 clearly defined spokes. The funnel domain increases its diameter to 110 Å. As for Figure 2, the density map for portal\(_{602}\)-gp4 complex was filtered to 10 Å, normalized using MAPMAN (Kleywegt and Jones, 1996), and displayed at 4.1σ above background corresponding to the molecular mass as calculated in SPIDER (Frank et al., 1996).
of the dome seems discontinuous from the portal* complex (Figure 3A), at a lower contour level (2.9σ above background), the density is continuous, suggesting that the dome is connected to the collar domain of portal* by flexible loops. Overall, the binding of gp4 to the funnel domain of portal* induces a conformational change that is transmitted over 100 Å away, from a location outside of the viral capsid to the inner DNA-filled cavity (Chang et al., 2006; Lander et al., 2006).

The internal structure of portal* undergoes dramatic conformational changes upon gp4 binding. As seen in a side-view slice through the DNA channel of portal protein, the channel diameter (“vestibule”) changes, and pore-lining densities reorient (Figure 4). Prior to gp4 binding, the funnel domain forms the constriction site in the channel with a diameter of 35 Å (Figure 4B, asterisk). Directly above, the vestibule opens up in the crown domain to a diameter of 70 Å and finally narrows again at the collar to 40 Å. Pore-lining densities in portal* are elongated, appear rod-like in shape, and are tilted with respect to the vertical axis of the portal protein (Figure 4B, yellow highlights). Following gp4 binding to portal*, both the channel diameter and the orientation of the pore-lining densities change. The channel diameter widens to 40 Å at the constriction site (Figure 4C, asterisk), 80 Å at the crown domain, and 45 Å at the collar. Most dramatically, all pore-lining densities reorient perpendicularly to the vertical axis of the pore and appear to point straight towards the center of the DNA channel (Figure 4C, yellow highlights).

The Dome Is a DNA Scaffolding Domain

Incubation of portal*;gp4 with a 38-mer dsDNA oligonucleotide (∼130 Å in length) clearly revealed DNA bound to the dome domain in negatively stained preparations (Figure 5A, arrows). Portal* did not bind DNA in the absence of gp4 (data not shown), suggesting that conformational changes induced by gp4 binding not only stabilize the dome domain but also prime this molecular machinery for binding DNA.

How does the dome bind DNA? Sequence alignment of nine portal proteins identified a single conserved motif of positively charged residues, RRR, common to all members of the Podo- viridae family of bacteriophages (Figure 5B, yellow). This region is disordered in the structure of Phi 29 portal protein, where it was thought to form part of the collar domain and project into the DNA channel (Guasch et al., 1998; Simpson et al., 2001; Simpson et al., 2000). Secondary structure prediction for the P22 portal protein suggests that this motif falls in a region (residues 277–293) with high propensity to fold into a helix, which we will refer to as DNA-binding helix or “db helix.” Overall, the db helix contains seven positively charged residues that may be important for DNA binding (Figure 5B). A helical wheel representation of the db helix reveals diagnostically opposed positively charged and hydrophobic faces (Figures 5C and 5D). Side-by-side packing of 12 such db helices within the dome domain would create a funnel lined with strong positive surface charge that could be important for binding DNA. In this scenario, the positively charged residues are important for binding DNA and could be used to chaperone DNA through the channel, while the hydrophobic face could be used for helix-helix packing interactions that stabilize the dome domain.

We hypothesize that the db helix is disordered in the procapsid conformation of portal protein but becomes structured in the mature capsid form following gp4 binding. Polyarginine repeats have been previously reported to have an intrinsic structural plasticity. For instance, the arginine-rich RRE-binding element of HIV-1 Rev protein is fully helical in complex with a 35-mer RRE RNA aptamer I (Ye et al., 1996), whereas it adopts an extended conformation (Ye et al., 1999) in complex with the 27-mer RRE RNA aptamer II. The transition between the two structural forms presumably occurs upon interaction with the acidic RNA pocket. Similarly, the IIB domain of importin α also presents a basic stretch, RRR, which adopts a random
Figure 5. DNA Binding to Portal602 Dome Domain

(A) Negatively stained preparation of portal602:gp4 in complex with DNA. An elongated density (~100 Å) is clearly visible projecting outward from the dome domain of portal602:gp4 in the presence of DNA (arrows). Inset projection averages: portal602:gp4 on the left, DNA-bound portal602:gp4 on the right. (B) Sequence alignment of nine related portal proteins identifies a possible DNA-binding helix (db helix). A stretch of conserved positively charged residues is identified in yellow. Surrounding glycine residues (highlighted in blue) suggest this region may be flexible. (C and D) When folded as a helix, one face is positively charged and may be important for DNA binding, while the other face is hydrophobic and may be important for packing interactions. Multiple sequence alignment (Altschul et al., 1990) was done in Clustal W (Thompson et al., 1994); secondary structure prediction was performed with PREDATOR (Frishman and Argos, 1996) and helical wheel representation according to Schiffer and Edmundson (1967).

coiled conformation in solution (Cingolani et al., 2000) as well as in the autoinhibited structure of importin α (Kobe, 1999) but becomes folded into a helix upon binding to the acidic surface of the receptor importin β (Cingolani et al., 1999).

Virus Morphogenesis and Initiation of Infection

The 17 Å electron cryomicroscopy reconstruction of the mature P22 virion (Chang et al., 2006; Lander et al., 2006) shows a large DNA-filled icosahedral capsid (Figure 6A, gray), of which a unique 5-fold vertex is occupied by the genome injection machinery (Figure 6A, green). Four donut-like densities were identified directly above the portal domain and were postulated to be injection proteins (Lander et al., 2006). We extracted the density for the genome injection machinery together with the four densities and fit our portal602:gp4 reconstruction (filtered to 17 Å resolution to facilitate the following comparison) into the full-length portal protein reconstruction from the mature virion (Lander et al., 2006) (Figures 6B and 6C). An overlay of the two reconstructions matches well: the dome domain seen in the portal602:gp4 reconstruction clearly exists in the mature virion, and the C-terminal residues lacking in portal602 clearly localize to these external spokes (Figure 6C, arrows). Although the dome forms a hollow donut surrounding genomic dsDNA in the mature virion (Lander et al., 2006), the very same region of the dome is solid in our portal602:gp4 reconstruction (Figure 3).

In addition to binding DNA, the dome domain may be important in recruiting the injection proteins. This is supported by the cryomicroscopy reconstruction of the mature phage P22, in which three additional rings of density are visible above the dome domain, which are likely the injection proteins gp7, gp16, and gp20 (Lander et al., 2006). It is possible that, as genome packaging nears completion and the procapsid undergoes conformational changes to the mature capsid form, the terminase complex is displaced by gp4. The binding-induced oligomerization of gp4 (Olia et al., 2006) onto the funnel domain of the portal ring would initiate a global conformational switch in the portal protein that everts the dome domain directly above the DNA channel. This domain may function as a clamp to grasp
the dsDNA that still lines the channel and to recruit the injection proteins from within the virion, as seen in the mature phage cryoreconstruction (Lander et al., 2006). The arginine-rich stretch of the dome domain folds upon binding to the acidic DNA, forming a highly positively charged surface in the dome, used to prime DNA in preparation for a new round of viral infection.

Closing Remarks
Using electron cryomicroscopy, we have characterized a dramatic conformational change in the large portal protein of bacteriophage P22. Free portal\textsubscript{sp2}, in its procapsid conformation, adopts a quaternary structure significantly different from that seen in the mature virion. We demonstrate that the conformational switch from the procapsid form to a mature phage conformation is specifically triggered by the assembly of tail factor gp4. When DNA packaging is complete, the terminase is disassembled from portal protein, likely via direct competition with the tail accessory factor gp4, which oligomerizes upon binding to the portal ring (Olia et al., 2006). This binding event, which marks the beginning of the tail assembly, leads to a dramatic reorganization of the DNA channel into its mature phage conformation, in which a dome-shaped domain is stabilized directly above the DNA channel. This domain functions as a DNA-binding scaffold, possibly used to clamp onto the viral genome and recruit the injection proteins at the portal protein vertex from inside the virion.

Our work provides an important snapshot of how the end of virus morphogenesis may be coupled to a new round of infection. In the context of the mature phage reconstruction (Chang et al., 2006; Lander et al., 2006), the gp4-induced conformational switch in portal protein may serve to recruit both viral DNA and the injection proteins in preparation for a new round of viral infection. Future studies must delve into the properties of the dome domain and its association with viral DNA and injection proteins. Similarly, investigation into maturation events in portal proteins from members of the herpesvirus family is critical for understanding how these viruses specifically release their genome into the cell nucleus upon interaction with the nuclear pore complex.

EXPERIMENTAL PROCEDURES
Expression, Purification, and Assembly of Dodecameric Portal Protein Complexes
The gene-encoding phage P22 portal protein (gp1) was cloned in a pET-21b vector. The C-terminal histidine tag was removed by site-directed mutagenesis by introducing a stop codon immediately after residue 725. C-terminally truncated portal protein (1–602) was generated by introducing an amber stop codon at position 603. Both untagged portal protein constructs were expressed in E. coli strain BL21 cells for 4–5 h at 30°C and purified from the soluble fraction using 30% ammonium sulfate. Precipitated portal monomer
was dialyzed and concentrated to ~200 mg/mL using a Millipore 100 kDa concentrator. To enrich the sample for fully assembled dodecameric portal rings, 0.5 M EDTA was added to a final concentration of 60 mM, and portal protein samples were incubated at 37 °C for 2–5 hr. The heat shock resulted in massive precipitation, which was pelleted by centrifuging at 100,000 g for 35 min. The resulting supernatant, which contained only fully oligomerized dodecameric portal protein (Lorenzen et al., 2008), was further purified on a Sephacryl S-300 gel filtration column equilibrated with phosphate-buffered saline. Recombinant gp4 was expressed in E. coli strain BL21 cells and purified as described before (Olia et al., 2006). The complex portalGp2-Gp4 was formed by adding a 3-fold molar excess of gp4 to dodecameric portal protein followed by purification of the complex on a Sephacryl 300 gel filtration column. Native gel electrophoresis on agarose gel was performed according to established procedures (Olia et al., 2006). For DNA binding, portalGp2 or portalGp2-Gp4 complexes were incubated with a 28 bp segment of DNA (sequence ACGGTTTTGCCGAAAATGT ACGGATTTCCGCAAATTG). Preparations were negatively stained according to established protocols (Ohl et al., 2004) and viewed on a transmission electron microscope operated at 100 kV (Morgagni, FEI, Hillsboro, Oregon). Images were recorded using a 2k x 2k Gatan charge coupled device camera. Projection averages for both portalGp2-Gp4 and DNA-bound portalGp2-Gp4 were calculated in Spider (Frank et al., 1998).

### Electron Cryomicroscopy and Data Processing

Full-length portal protein (portalGp2-Gp4 complexes were prepared for electron cryomicroscopy as follows. A 2 µl drop of sample at 10 mg/ml was applied to a Quantifoil holey carbon grid (Quantifoil, Germany), blotted with filter paper, and immediately frozen in liquid ethane using a Vitrobot (FEI, Hillsboro, Oregon). Grids were mounted onto a Gatan high-resolution cryo holder and inserted into an FEI Tecnai F20 microscope equipped with an ion thinning gun and operated at 200 kV. Images of the vitrified specimen were recorded under low-dose conditions at a nominal magnification of 50,000 x and defocus values ranging from 2.5 to 5 μm. All micrographs were visually examined using an optical diffractometer to select only drift-free images with no significant astigmatism. The film was digitized with a Nikon Super Cool Scan 9000 using a step size of 6.9 μm corresponding to 1.27 Å/pixel. Digitized micrographs were binned twice, yielding a pixel size of 2.54 Å. Approximately 20,000 individual particles were selected for portalGp2-Gp4 and 36,000 for portalGp2-Gp4 using WEB, and processed in SPIDER to generate an initial low-resolution 3D density map following established protocols (Frank et al., 1996). Image refinement using FREALIGN version 7.00 (Grigorieff, 2007) was performed as described previously (Fotin et al., 2004), Briefly, the contrast transfer function (CTF) parameters were determined for each micrograph using the program CTFITL (Mandell and Grigorieff, 2003), and only those images where the CTF parameters could be accurately determined were used for further refinement. CTFITL was used to calculate the defocus, astigmatism angle, position of tilt angle and tilt axis, and x and y coordinates for each individual particle visible in each micrograph. Search and refine protocols were used in FREALIGN (Grigorieff, 2007) and were limited initially to a resolution range of 500–40 Å. Euler angles and x, y shifts for each individual particle relative to the initial search model were determined and further refined in consecutive rounds of refinement in FREALIGN (Grigorieff, 2007). The nominal resolution of the final reconstruction was estimated from the spatial frequency at which the FSC fell to 0.143 (Rosenthal and Henderson, 2003) was 8 Å for portalGp2-Gp4 complex (Figure 1D). Both density maps were filtered to 10 Å resolution to facilitate the comparison, and a negative B factor of 1000 Å2 was applied to the final reconstructions to restore high-resolution contrast. Density maps were normalized in MAPMAN version 7.8 (Kleywegt and Jones, 1996). The Contour sigma threshold was determined in Spider (Frank et al., 1996) and was 4.6σ above background for portalGp2Gp4 (760 kDa) and 4.1σ above background for portalGp2-Gp4 (1.1 mA). Density maps were visualized in UCSF Chimera (Petterson et al., 2004) and all figures prepared using the above sigma levels. PortalGp2 was reconstructed to 44 Å resolution and could not be further refined because of the lack of side views (De Rosier and Klug, 1968) (Figure 1A). PortalGp2 and portalGp2-Gp4 were rotationally and translationally aligned in UCSF Chimera (Petterson et al., 2004) until density features common to both lined up (Figure S3). Figure S4 shows the electron cryomicroscopy reconstructions for both portalGp2 and portalGp2-Gp4, filtered to 12.5 and 12 Å, respectively, and corresponding to the spatial frequency at which the FSC fell to 0.5 (Stewart et al., 1999) (Figure S4). All features described in this manuscript are clearly visible using both 0.143 (Rosenthal and Henderson, 2003) and 0.5 (Stewart et al., 1999) FSC criteria.

### Supplemental Data

Supplemental Data include four figures and can be found with this article online at http://www.molecular.org/cgi/content/full/29/3/376/DC1/.

### ACKNOWLEDGMENTS

The authors thank Bridget Carragher and Clinton Potter (Scripps, La Jolla, CA) for allowing us to use their electron cryomicroscopes for data collection. Electron micrographs were collected at the National Resource for Automated Molecular Microscopy (NARMM), which is supported by the National Institutes of Health through the National Center for Research Resources’ P41 program (RR17573). We would also like to thank Thomas Waiz (Harvard Medical School) for the use of his electron cryomicroscope in the early stages of this study. The authors thank Yifan Cheng (University of California, San Francisco) for helpful discussions about single-particle reconstructions and Sherwood Casjens (University of Utah) and Carlos Catalano (University of Washington) for helpful discussions on bacteriophage P22. The authors declare that none have financial interests related to this work. H.Z., M.G., S.A., and T.G. collected the electron cryomicroscopy data, did all the data processing and subsequent analysis, and prepared manuscript figures. A.S.O. and G.C. prepared the portalGp2, portalGp2-Gp4, and portalGp2-Gp4 complexes and did the gel shift assay. G.C. and T.G. wrote the manuscript.

Received: July 30, 2007

Revised: October 8, 2007

Accepted: November 16, 2007

Published: February 14, 2008

### REFERENCES


Accession Numbers
Density maps portal-72S, portal-62S, and portal-50S/gp4 have been deposited at the Macromolecular Structure Database under reference numbers 5631, 5632, and 5633, respectively.
LETTERS

Backbone structure of the infectious ε15 virus capsid revealed by electron cryomicroscopy

Wen Jiang1*, Matthew L. Baker2*, Joanita Jakana3, Peter R. Weigele3, Jonathan King3 & Wah Chiu2

A half-century after the determination of the first three-dimensional crystal structure of a protein1, more than 40,000 structures ranging from single polypeptides to large assemblies have been reported2. The challenge for crystallographers, however, remains the growing of a diffracting crystal. Here we report the 4.5-Å resolution structure of a 22-MDa macromolecular assembly, the capsid of the infectious epsilon15 (ε15) particle, by single-particle electron cryomicroscopy. From this density map we constructed a complete backbone trace of its major capsid protein, gene product 7 (gp7). The structure reveals a similar protein architecture to that of other tailed double-stranded DNA viruses, even in the absence of detectable sequence similarity3,4. However, the connectivity of the secondary structure elements (topology) in gp7 is unique. Protruding densities are observed around the two-fold axes that cannot be accounted for by gp7. A subsequent protometric analysis of the whole virus identifies these densities as gp10, a 12-kDa protein. Its structure, location and high binding affinity to the capsid indicate that the gp10 dimer functions as a molecular staple to prevent viral assembly.

Beyond ε15, this method potentially offers a new approach for modelling the backbone conformations of the protein subunits in other macromolecular assemblies at near-native solution states.

Bacteriophages have been valuable model systems for studying virus structures and assembly5–8, as well as protein folding and maturation9,10. These viruses also have a critical function in transferring genes between bacterial host cells, influencing host pathogenicity11 and the population ecology of environmental microorganisms12. With an estimated 1011 particles, tailed double-stranded DNA (dsDNA) bacteriophages are likely to be the most abundant life forms in the biosphere13. Despite repeated efforts, no infectious tailed dsDNA phage has been crystallized, although the crystal structure of a recombinant empty capsid of HK97 has been determined14. The great variation and diversity in the amino acid sequences among these tailed dsDNA bacteriophages has prevented the prediction of the structural organization of the capsid from the genome sequence alone.

Single-particle electron cryomicroscopy (cryo-EM) was used about ten years ago to determine the first structures oficosahedral viruses to subnanometre resolutions (7.4–9 Å), revealing long α-helices15,16. Here we report a 4.5-Å resolution icosahedral capsid structure of the infectious tailed dsDNA bacteriophage ε15 of Salmonella anatum17 with the use of single-particle cryo-EM.

Purified infectious ε15 particles were imaged with a liquid-helium 300-kV electron cryomicroscope16,17. Typical images (Fig. 1a) clearly demonstrated signals beyond 5 Å in their power spectra (Supplementary Fig. 1a, b). From about 20,000 individual particle images, the density map of ε15 was reconstructed to a resolution of 4.5 Å (Fig. 1b, Supplementary Fig. 1c and Supplementary Movie 1). This resolution is the result of advances and accumulated experience in instrumentation, the collection of large data sets, and image processing, and also in the availability of large-scale distributed computing (see Methods). The map clearly shows protruding densities around all the icosahedral and local two-fold positions (Fig. 1b and Supplementary Movie 1) that connect the underlying masses, forming the capsid shell. The density map was sufficiently well resolved to allow each of the seven subunits in the asymmetric unit of the T = 7 icosahedron to be computationally extracted (Fig. 1c and Supplementary Movie 1). At this resolution it was evident that each subunit consists of two distinct lobes connected by weak densities. Each of the lobes was segmented individually, yielding a small protruding, mainly hollow, globular density and a larger, more extended density.

In the larger of the two density lobes, eight α-helices and two β-sheets were identified with the feature recognition program SEHunter18 (Supplementary Fig. 2a). The spatial dispositions of these secondary structure elements seemed to be similar to those of the major capsid proteins of other tailed dsDNA phages and herpesviruses15,18,19. The large density lobe was therefore assigned to the major capsid protein, gp7 (335 amino-acid residues).

Using computational modelling methods (Supplementary Information) together with sequence analysis, we constructed the ε15 backbone models (residues 1–335) for each of the seven gp7 monomers (Fig. 2a, Supplementary Fig. 2 and Supplementary Movie 2). The pitch of the α-helices, the separation of the β-strands, and several bulky side-chains were visible in the density and corresponding models (Fig. 2b). The gp7 model could be roughly divided into four domains: an extended amino terminus (N-arm), a central, triangular domain (A-domain), an elongated protrusion domain (P-domain) and a long, extended loop (E-loop) (Fig. 2c). As noted previously15, the architecture of gp7, including the overall shape and the arrangement of its domains and secondary structure elements, was similar to that of the HK97 capsid protein14, although no sequence similarity was evident. However, the connectivity of the secondary structure elements (that is, their topology) was different because of the permutation of the order of the two carboxy-terminal sequence segments (residues 158–250 and 251–335) of ε15 gp7 (Supplementary Information and Supplementary Fig. 8a). Further sequence analysis indicates that this type of sequence permutation might be a general feature of the major capsid proteins of tailed dsDNA phages (Supplementary Information). Similar types of structural conservation in the presence of sequence and topology permutation have been reported in several protein families19,20,21.

This model building was possible primarily because of the ability to distinguish unique secondary structure elements accurately in

1Markey Center for Structural Biology, Department of Biological Sciences, Purdue University, West Lafayette, Indiana 47907, USA. 2National Center for Macromolecular Imaging, Verna and Marrs McLean Department of Biochemistry and Molecular Biology, Baylor College of Medicine, Houston, Texas 77030, USA. 3Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. *These authors contributed equally to this work.
Figure 1 | Cryo-EM image and three-dimensional map of bacteriophage \( \varepsilon 15 \). a, Typical 300-kV image of ice-embedded \( \varepsilon 15 \) kept at 4.2 K, recorded in a JEM3000SFF electron cryomicroscope. Fingerprint signature features of viral DNA are clearly seen in the projection images. b, The 4.5-Å resolution three-dimensional reconstruction of \( \varepsilon 15 \) phage, about 700 Å in diameter, radially coloured. c, The seven subunits in an asymmetric unit, annotated in different colours. Each subunit contains one copy of gp7 and one copy of gp10.

Figure 2 | Model of gp7 monomer. a, A gp7 model superimposed on the corresponding segmented density map. The model is coloured from the N terminus (blue) to the C terminus (red) and several landmark residues are annotated. b, Regions of the density map and model of gp7, illustrating model features such as the 3.4-Å pitch in a helix (top left, middle left and bottom right), the roughly 4.8-Å β-strand separation (top right) and bulky side chains.

c, Two views of the seven gp7 models and an average density map from the asymmetric unit, illustrating the conformational variations among the seven gp7 monomers. The seven models are coloured with the colour scheme in Fig. 1c. The four domains of a gp7 monomer are also annotated: N, N-arm; P, P-domain; A, A-domain; E, E-loop.
both the sequence and the cryo-EM map (Supplementary Information). Failure to correlate these structural elements accurately may limit model accuracy and the general applicability of this approach to other data at a similar resolution. As a result of low sequence identity, neither the HK97 nor the T4 capsid proteins was used explicitly as a structural template during the model building process (Supplementary Information). However, the quality of the map and our method of modelling were adequate for a backbone trace of gp7. The accuracy of our backbone trace was subsequently verified by a root mean squared deviation of 3.765 Å of the 198 structurally conserved residues throughout the structures of 15 gp7 and HK97 gp5 (Supplementary Information and Supplementary Fig. 8b, c).

Although mostly similar, the seven individual gp7 subunit models within the asymmetric unit do contain some structural variations as expected for the quasi-equivalent packing of chemically identical subunits in an icosahedral lattice22 (Fig. 2c and Supplementary Fig. 3). On average, the models have a root mean squared deviation of about 2.3 Å (2.1–3.5 Å), with the major differences occurring at the N terminus and the E-loop. In particular, the E-loops of the gp7 subunits in the pentameric capsomere (penton) and the hexameric capsomere (hexon) most proximal to the penton show a significant bend (about 20°) towards the interior of the capsid (Fig. 2c). This bending is needed to pack the subunit into a five-fold vertex and to form the icosahedral lattice structure.

Applying icosahedral symmetry to the seven gp7 models of an asymmetric unit, a model of the T = 7 capsid containing 420 copies of gp7 was constructed (Fig. 3a and Supplementary Movie 3). Although our Cα models lack side-chain conformations and are therefore not sufficient for inferences to be made about the exact nature of inner surface charge and its influences on the packing of the encapsulated dsDNA genome, the capsid model clearly shows the arrangement and interaction sites of gp7 throughout the capsid (Supplementary Fig. 4a, b). The triangular A-domain interacts laterally around the centre of the penton/hexagon ring (Fig. 3a) through interactions between two nearly parallel helices (residues 256–269 and 284–296, respectively), one from each of the two adjacent subunits (Supplementary Fig. 4a). In contrast, the N-arm, P-domain and E-loop are located at the periphery of the penton/hexon capsomers. In particular, the extended E-loop crosses over the middle region of the P-domain of the adjacent subunit in the same capsomere (Supplementary Fig. 4b) and contacts the distal end ‘hook’ of a P-domain from a neighbouring capsomere (Fig. 3b). Two sequential positively charged arginine residues (R66, R67) at the end of the E-loop of one subunit are proximal to the highly negatively charged hook (E200, D201, D204, D205) of a subunit in an adjacent capsomere at the icosahedral and local three-fold axes. These interactions across capsomeres probably stabilize the E-loop and the entire capsid. These interactions occur at the equivalent locations in HK97 where two subunits are covalently crosslinked to stabilize the capsid6.

In this 4.5-Å map it became clear that gp7 could not account for the entire capsid shell as had previously been assumed2. We re-examined the protein composition of the phage particles with a polyacrylamide gel at a higher percentage, which resolved a second, small, high-copy-number capsid protein (Fig. 4a). Mass spectrometry analysis identified this protein as gp10 (about 12 kDa; 111 residues) and it was therefore assigned to the small protruding density lobe.

Feature detection with SSEHunter18 revealed that the relatively hollow gp10 consists mainly of β-sheets, although two short α-helices (about three turns each) were observed; a similar set of secondary structure elements was predicted computationally from the sequence (Supplementary Fig. 5a, b). Sequence searches revealed no known homologues. However, several structure-based searches did consistently return PDZ domains as potential structural templates. Unlike gp7, constructing a reliable model for gp10 from the density map was not possible owing to the relatively high percentage of β-sheet and only two short, similar-sized α-helices. Nevertheless, the PDZ-like domain model seems to have a size and shape similar to those of the small-lobe density of the subunit (Supplementary Fig. 5c).

Despite the lack of a high-resolution model for gp10, a potential role for this protein can be derived from its relative position on the capsid. A back-to-back dimer of gp10 sits on the surface of the virus at the two-fold axes and makes contact with six gp7 subunits (Fig. 4b and Supplementary Movies 1 and 3). Each gp10 molecule contacts four gp7 subunits, namely two subunits contributed from each of the two neighbouring capsomeres (Fig. 4c and Supplementary Fig. 4a). In the gp10 dimer, the two N-arms of gp7 buttress the gp10 dimer laterally (Fig. 4d), and the insertion loop and the E-loop line the groove in which the dimer rests. These interactions not only join neighbouring gp7 subunits, but they also link the neighbouring capsomeres. Thus, gp10 ‘staples’ the underlying gp7 capsomeres into a robust cage that can withstand the pressure from the densely packed dsDNA genome in the infectious phage particles.

In support of this potential role, a significant amount of gp10 was subsequently found to remain bound to the capsid in the presence of 5 M guanidinium chloride (GuHCl) (Fig. 4a and Supplementary Fig. 6), far beyond the concentration of denaturant needed to remove the tail and portal proteins in infectious phage particles. Capsids are disrupted by 6 M GuHCl, indicating that the gp7–gp10 interactions are of the same order of magnitude as those of gp7–gp7 interactions.

Figure 3 | Capsid model and interactions of gp7. a, A full capsid model of gp7, with the same colour scheme as in Fig. 1c. b, The charged amino acids of gp7, viewed slightly offset from the three-fold axis. The residue labels are shown in blue and green, corresponding to the gp7 monomer colour scheme in a.
Similar stabilizing roles have been ascribed to other phage decorating proteins23-24, as well as being consistent with the role of PDZ domains in mediating protein interactions in large protein complexes25. The ε15 gp10 dimer interacts with the underlying gp7 shell around two-fold axes, whereas the decorating proteins in phage λ (ref. 24) and L (ref. 26) are located at the three-fold axes. However, in bacteriophage HK97, another tailed dsDNA bacteriophage with a similar structure, no accessory protein is needed because interactions between capsomeres are mediated through a covalently crosslinked chain-mail network4. As such, the ‘molecular staple’ role of gp10 is probably an alternative strategy taken by ε15 to maintain capsid stability and indicates the evolution of additional surface proteins in maintaining capsid stability in tailed dsDNA bacteriophages. Taken together, gp7 and gp10 form a complex network of interactions to protect the viral genome while maintaining an intact viral capsid in the face of various stressful conditions present in natural environments.

Beyond the insight into the structure and function of the ε15 bacteriophage capsid shell afforded by these models, this work signifies a new stage in the structural research of complex biological nanomachines. These biological nanomachines can now be studied in near-native solution conditions by cryo-EM without crystals, at a level of detail close to X-ray crystallography and nuclear magnetic resonance.

METHODS SUMMARY

Infectious bacteriophage ε15 was purified from cultures of Salmonella anatum4. The purified phage particles were rapidly plunge-frozen and imaged on a JEM3000SFF electron cryo-microscope operating at liquid-helium specimen temperature46-67. Of about 3,000 digitized photographic film images, about 40% preserved clear signals beyond a resolution of 6 Å and were used for further image processing and three-dimensional reconstruction. Individual particles (36,259) were manually selected from these micrographs and were then processed with EMAN2. These particle images were refined iteratively and reconstructed with a large-scale distributed Condor computing resource at Purdue University (http://web.cac.purdue.edu/condorview). Icosahedral symmetry was imposed during refinement and reconstruction, resulting in the final 4.5-Å resolution structure of the icosahedral shell as assessed by the 0.5 Fourier shell correlation criteria48. Individual subunits were segmented from the reconstruction with Amira (http://www.amira.com) and Chimera29 and then analysed with SSEHunter49. Coupling the cryo-EM densities with secondary structure elements predicted from the sequence, the Cx backbone models for each of the seven gp7 subunits were constructed individually. A full capsid model for gp7 was then generated by applying icosahedral symmetry to the gp7-based asymmetric unit. Analysis of the gp10 sequence was performed with a meta-prediction server (http://meta.biosisinfo.pl), ALIGN2 (http://workbench.sdsc.edu) and the NCBI psi-blast server (http://www.ncbi.nlm.nih.gov/blast). Structural alignments were calculated with MmRMS20 and displayed with Chimera29.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 23 September 2007; accepted 3 January 2008.


Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank D. Braun, P. Smith and B. Loftis; D. Sierkowska, P. Mikeal and S. Wilson for their support with the Condor computing system; and R. H. Goradia and M. Dougherty for digitizing the image data and preparing the Supplementary Movies respectively. This work was supported by grants from the National Institutes of Health and the National Science Foundation.

Author Contributions M.L.B. and W.J. conducted the structural analysis and contributed equally to this work. W.J. performed the image processing and reconstructions. J.J. collected the cryo-EM image data. P.R.W. performed the biochemical purification and characterization of εIl. M.L.B., W.J., P.R.W., J.K. and W.C. interpreted the results and wrote the manuscript.

Author Information The three-dimensional density map has been deposited into the EBI–EMD database with accession number EMD-5003. The backbone model has also been deposited in the Protein Data Bank with accession number 3CSB. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to W.C. (wah@bcm.edu) or W.J. (jiang12@purdue.edu).
METHODS

Infectious bacteriophage c15 was purified as described previously. In brief, mid-exponential-phase cultures of Salmonella anatum were infected with c15 at a multiplicity of infection between 0.01 and 0.001 and further incubated until lysis. Virions were precipitated from clarified lysates by the addition of poly(ethylene glycol) (average molecular mass 8,000 kDa) and NaCl to 10% and 0.5 M final concentrations, respectively. The phage particles were pelleted and resuspended in 50 mM Tris–HCl pH 7.5, 25 mM NaCl, 10 mM MgCl2, and further purified by CsCl gradient centrifugation.

Identification of virion structural proteins. In a similar manner to that described previously, c15 virion proteins were resolved by SDS–PAGE but at a higher polyacrylamide concentration (15%). The bands were excised and digested with trypsin at the MIT Biopolymers core facility. Peptide mixtures were resolved by HPLC and fed into a linear trap quadrupole (LTQ) mass spectrometer (Thermo Electron Corporation), where the peptides were ionized. Spectra of the resulting peptide B- and Y-ions were compared with a database of translated open reading frames from the c15 genome (GenBank accession number NC_004775) and assigned a putative peptide sequence. Peptide matches were confirmed by visual inspection and their Sequest Database Search Software scores (XC, Delta Cn, Sp, RSp)15.

Purified virions were incubated for 2 h in buffered 0.1, 1, 3, 4, 5 and 6 M GuHCl and sedimented through a layer of 20% sucrose containing the same concentration of GuHCl, on to a 1.6 g ml–1 layer of buffered CsCl. The buffer used was 50 mM Tris–HCl pH 7.5, 25 mM NaCl, 5 mM MgCl2. The GuHCl-treated virion material formed a visible bluish band at the interface and was harvested through the side of the centrifuge tube with a hypodermic needle. The material was dialysed against three changes of the same buffer, and the proteins were resolved by standard SDS–PAGE under reducing conditions.

The gels were stained with Coomassie blue to reveal the proteins, and their quantities in the gel were determined by densitometric analysis by scanning the gel using a flatbed transparency scanner with a transparency lid (Microtek). The scanner was calibrated with a no. 2 optical density step tablet (Kodak). Protein bands were measured with NIH-Image software, and their total optical density was calculated from the area of their corresponding peaks.

Electron cryomicroscopy. The purified infectious phage particles were prepared for cryo-EM by rapid plunge freezing. The images (Fig. 1a) were taken at a dose of about 25 e–Å–2 in a JEM3000SFF electron cryomicroscope operated at 300kV and at liquid helium specimen temperature16,17. These images were recorded on Kodak SO163 films at ×60,000 nominal magnification. The films were digitized with a Nikon Super CoolScan 9000 ED scanner at 6.35 μm per pixel and resulted in 1.06 A per pixel sampling for the scanned images.

Image processing. Individual particle images (768 pixels × 768 pixels) were first selected automatically with the ethan method19 followed by manual screening with the EMAN20 boxer program to remove contamination and to keep only the well-separated particles. The selected particles within a micrograph were incoherently averaged to generate both a two-dimensional power spectrum and a one-dimensional scattering curve for the evaluation of image quality and the determination of the contrast transfer function (CTF) parameter. A total of 1,235 micrographs, about 40% of all digitized micrographs, with CTF peaks visible beyond 6 Å and isotropic CTF rings (that is, negligible charging, drift and astigmatism) were used for further processing. The CTF parameters of these ‘good’ micrographs were first fitted automatically by using a constrained nonlinear optimization method for CTF fitting (C. Yang, W.J., D. Chen, U. Adiga, W.C. and E. Ng, unpublished observations) and then verified visually with the EMAN cift program20. The defocuses of these micrographs had a distribution of 0.83 μm ± 0.33 μm. The experimental B factors24 of the particles used in the subsequent analysis were 46 ± 9 Å2.

A total of 36,259 particle images were selected and then refined iteratively with the relatively standard EMAN protocols25. First, these particles were classified with the EMAN projection matching program classesbynum with an angular projection step size of 0.4°. The particles within each class were averaged to generate class averages with full CTF correction with a Wiener filter and signal-to-noise ratio weighting. The class averages were merged into a three-dimensional map using the make3d program with isoschedral symmetry imposed. The make3d program was parallelized with Message Passing Interface (MPI) to speed the computation for a large number of large images. These steps were iterated until the refinement converged. The resolution of the final map was evaluated by using the 0.5 threshold criterion of Fourier shell correlation (FSC)26. Only the isoschedral shell region was included in this FSC analysis by masking out the external background noises and the internal DNA densities with soft masks. Because the isoschedral shell is polyhedral and the masks applied were spherical, some DNA densities were still retained around the isoschedral five-fold vertices in the density maps used for FSC analysis.

Because of the large number of particle images, the large size for each particle image and the large number of projections, the refinement, especially the projection matching step, was computationally very intensive and required about 105 CPU hours per refinement iteration. To meet these computational requirements we developed a high-throughput image-processing strategy with the Condor job management system (http://www.cs.wisc.edu/condor) to use CPUs simultaneously from multiple clusters and scattered desktop Linux computers at Purdue University (http://web.rcac.purdue.edu/condorview) (W. Wu, T. Chachiyo and W. J., unpublished observations). Typically, about 300–500 CPUs and occasionally about 1,000 CPUs were harvested by Condor for the refinement, reducing the overall time for each refinement iteration to about ten days. Ten iterations (about 100 days total) were used in computing the final density map from about 20,000 particle images.

Density interpretation and model building. The individual protein monomers (seven gp7 subunits and seven gp10 subunits) were visually segmented by using a combination of Amira (http://www.amira.com) and Chimera27. Identification of secondary structure elements was accomplished with SSEHunter28. Sequence-based prediction of secondary structure elements was performed by using a consensus-based prediction described in detail in Supplementary Information for both gp7 and gp10 (UniProtKB accession numbers Q858G8 and Q858G5, respectively). These secondary structure elements were combined in the context of the density and its associated skeleton for construction of the gp7 models. The models were subsequently constructed and refined with the Coot toolkit29 (Supplementary Information). Icosahedral symmetry was applied to the seven models in an asymmetric unit to generate a full viral capsid structure.