Journal of Virology®

STRUCTURE AND ASSEMBLY

Cryo-EM Structure of Gokushovirus ΦEC6098 Reveals a Novel Capsid Architecture for a Single-Scaffolding Protein, Microvirus Assembly System

Hyunwook Lee,[a] Alexis J. Baxter,[a] Carol M. Bator,[b] Bentley A. Fane,[c] Susan L. Hafenstein[a,b,d]

[a]Department of Biochemistry and Molecular Biology, Pennsylvania State University, University Park, Pennsylvania, USA
[b]Huck Institute of the Life Sciences, Pennsylvania State University, University Park, Pennsylvania, USA
[c]The BIO5 Institute, Keating Building, University of Arizona, Tucson, Arizona, USA
[d]Department of Medicine, Pennsylvania State University College of Medicine, Hershey, Pennsylvania, USA

ABSTRACT Ubiquitous and abundant in ecosystems and microbiomes, gokushoviruses constitute a *Microviridae* subfamily, distantly related to bacteriophages ΦX174, α3, and G4. A high-resolution cryo-EM structure of gokushovirus ΦEC6098 was determined, and the atomic model was built *de novo*. Although gokushoviruses lack external scaffolding and spike proteins, which extensively interact with the ΦX174 capsid protein, the core of the ΦEC6098 coat protein (VP1) displayed a similar structure. There are, however, key differences. At each ΦEC6098 icosahedral 3-fold axis, a long insertion loop formed mushroom-like protrusions, which have been noted in lower-resolution gokushovirus structures. Hydrophobic interfaces at the bottom of these protrusions may confer stability to the capsid shell. In ΦX174, the N-terminus of the capsid protein resides directly atop the 3-fold axes of symmetry; however, the ΦEC6098 N-terminus stretched across the inner surface of the capsid shell, reaching nearly to the 5-fold axis of the neighboring pentamer. Thus, this extended N-terminus interconnected pentamers on the inside of the capsid shell, presumably promoting capsid assembly, a function performed by the ΦX174 external scaffolding protein. There were also key differences between the ΦX174-like DNA-binding J proteins and its ΦEC6098 homologue VP8. As seen with the J proteins, C-terminal VP8 residues were bound into a pocket within the major capsid protein; however, its N-terminal residues were disordered, likely due to flexibility. We show that the combined location and interaction of VP8's C-terminus and a portion of VP1's N-terminus are reminiscent of those seen with the ΦX174 and α3 J proteins.

IMPORTANCE There is a dramatic structural and morphogenetic divide within the *Microviridae*. The well-studied ΦX174-like viruses have prominent spikes at their icosahedral vertices, which are absent in gokushoviruses. Instead, gokushovirus major coat proteins form extensive mushroom-like protrusions at the 3-fold axes of symmetry. In addition, gokushoviruses lack an external scaffolding protein, the more critical of the two ΦX174 assembly proteins, but retain an internal scaffolding protein. The ΦEC6098 virion suggests that key external scaffolding functions are likely performed by coat protein domains unique to gokushoviruses. Thus, within one family, different assembly paths have been taken, demonstrating how a two-scaffolding protein system can evolve into a one-scaffolding protein system, or vice versa.

KEYWORDS cryo-EM, DNA-binding protein, *Microviridae*, gokushovirus, microvirus, mushroom spike

Gokushoviruses constitute a subfamily within the *Microviridae*, which encompasses the icosahedral, single-stranded (ss) DNA bacteriophages. Atomic structures of several ΦX174-like capsids and procapsids are available (1 to 4); however, there is only

Editor Lori Frappier, University of Toronto

Address correspondence to Susan L. Hafenstein, shafenstein@psu.edu, or Bentley A. Fane, bfane@email.arizona.edu.
The authors declare no conflict of interest.

Received 27 June 2022
Accepted 21 September 2022
Published 18 October 2022

a low-resolution gokushovirus map and a predicted structure (5, 6). Like other family members, gokushoviruses have small ~25 nm in diameter, nonenveloped, T = 1 icosahedral capsids (1, 5, 7). Sequence homologies between gokushovirus coat proteins provide the strongest data that interconnect members across the entire family (8, 9). Unlike ΦX174-like phages, gokushovirus coat proteins form protrusions, consisting of a globular head and stalk domain, that extend ~54 Å above the icosahedral 3-fold symmetry axes (5). The head domain's distal tip contains a predicted hydrophobic depression hypothesized to bind a cellular receptor.

Although convincing sequence homologies with other ΦX174 proteins are not apparent, bioinformatic analyses suggest that gokushoviruses encode a DNA pilot protein (VP2), the equivalent of the ΦX174 H protein, which mediates penetration. These proteins exhibit a predicted α-helical fold with periodic placement of DNA-interacting side chains. The purified ΦX174 DNA pilot protein H (10, 11) forms a decameric tube. The DNA-interacting side chains, amide and guanidinium groups, point toward the center of the tube and mediate genome translocation (12). Lastly, gokushoviruses encode a small basic protein (VP8) that resembles the ΦX174 DNA-binding protein J, which is a structural protein and essential for genome packaging.

Unlike many small viruses ($T \leq 3$), ΦX174 morphogenesis is mediated by two scaffolding proteins, an internal and external species. These scaffolding proteins temporally divide the pathway into two discernible phases. During early morphogenesis, the internal scaffolding protein B facilitates the formation of pentameric intermediates, consisting of coat F, spike G, internal scaffolding B, and DNA pilot H proteins, in a respective 5:5:5:1 stoichiometry (13 to 15). During late morphogenesis, 240 copies of the external scaffolding protein D organize 12 pentameric intermediates into the procapsid (14, 16). In contrast, gokushovirus assembly appears to be mediated by a single internal scaffolding protein, VP3, which is present in purified procapsids but not in virions (17).

Here, we report a 2.6 Å-resolution cryo-EM structure of ΦEC6098, a gokushovirus first detected as a prophage in a metagenomic database of *Escherichia marmotae* and elegantly resurrected as a biological isolate (18, 19). The structure suggests an assembly mechanism in which the viral coat protein performs, at least in part, the functions associated with the ΦX174 external scaffolding protein D. In addition, the structure confirms that VP8 is a structural protein and performs the functions associated with the ΦX174 DNA-binding protein J.

RESULTS

Cryo-EM structure of gokushovirus ΦEC6098. Purified gokushovirus ΦEC6098 was vitrified and imaged on a Titan Krios equipped with a Falcon 3EC camera (Table 1). In the micrographs, many of the ΦEC6098 particles were positioned in 2D regular arrays (Fig. 1A), seemingly connected by the hydrophobic heads of the 3-fold protrusions (5). An icosahedral single-particle 3D reconstruction was performed to produce a 2.6-Å-resolution map, which depicted a 27-nm diameter capsid with mushroomlike protrusions extending from the each icosahedral 3-fold axis of symmetry (Fig. 1B). The density corresponding to these trimeric protrusions was weaker than the density corresponding to the capsid shell, dwindling as the protrusion extended radially away from the capsid (Fig. 1C). However, at lower contour levels, the ~5-nm protrusions were fully visible (Fig. 1D) and exhibited a globular head domain connected to the capsid by a narrow stalk, as previously noted (5). Other features include an elevated ridge of density running between the icosahedral 5-fold and 3-fold symmetry axes. Each icosahedral 2-fold symmetry axis had a 20-Å-deep depression, and a shallow depression encircled a raised bump at each 5-fold vertex (Fig. 1E). The central section of the cryo-EM map showed strong amorphous density corresponding to the viral genome (Fig. 1C).

Structure of the ΦEC6098 capsid protein. The atomic model of gokushovirus ΦEC6098 was built *de novo* into the cryo-EM density map (Fig. 2A and Table 1). The capsid map quality was sufficient to unambiguously model most regions (residues 4 to 221 and 310 to 566) of the major capsid coat protein, VP1, which is 566 amino acids in length. The atomic coordinates for 88 amino acids (residues 222 to 309), corresponding

TABLE 1 Cryo-EM data collection, refinement, and validation statistics

Parameter	Value
Data collection	
Microscope	FEI Titan Krios
Camera	Falcon 3EC camera
Magnification	59,000
Voltage (kV)	300
Electron dose (e-/Å^2)	50
Defocus range (μM)	0.5 to 3.3
Pixel size (Å)	1.1
Initial micrographs (#)	2,152
Final micrographs (#)	2,110
Initial particles (#)	53,021
Final particles (#)	32,308
Refinement	
Resolution (Å)	2.6
Model-data correlation coefficient (mask)	0.86
Model-data correlation coefficient (volume)	0.86
Map-sharpening *B* factor (Å^2)	–100.514
Validation	
MolProbity score	0.63
Clashscore	0.0
Poor rotamers (%)	0.24
Ramachandran plot	
Favored (%)	97.3
Allowed (%)	2.7
Disallowed (%)	0.00

to the apically distal head domain of the 3-fold protrusion, were undetermined due to the poorly resolved cryo-EM density. This may indicate head-domain flexibility. The major structural protein, VP1, contained the conventional jelly-roll structure comprising eight antiparallel β-barrels (BIDG and CHEF) connected by loops (Fig. 2B). The EF (residues 92 to 363) and HI (residues 425 to 539) insertions constituted ~68% of VP1, contained 12 α-helices, and formed most of the capsid outer surface, including the 3-fold protrusions (Fig. 2B).

Notably, the VP1 structure included a long N-terminal extension, residues 4 to 28 (Fig. 2B), that was not observed in any of the ΦX174-like structures (1 to 4). These residues extended across the inner surface of the capsid shell from a corner of the β-barrel motif toward the 5-fold axis of the 2-fold related pentameric unit (Fig. 2B). In addition, a small region of well-ordered density was found inside a small pocket at the center of each coat protein (Fig. 2C), which was identified as VP8, the DNA-binding protein. Out of a total of 39 residues, 10 C-terminal VP8 residues ARSMRGGIRL were built into the density. The arginine-rich N-terminal residues, analogous to J protein's DNA-binding domain, were not modeled due to the lack of ordered density, suggesting conformational flexibility.

The ΦEC6098 capsid surface was mostly neutral or negatively charged (Fig. 2D), whereas in the ΦX174 and α3 structures, the capsid surface has more positively charged surface regions (Fig. S1 in the supplemental material) (1, 2).The differences in the surface charges were mainly due to a negatively charged patch on the EF-insertion loop that consists of acidic residues (DEEE, residue numbers 178 to 181) and is exposed to the outer surface, where ΦX174 and α3 have positively charged residues instead (Fig. S2). This electrostatic difference may reflect coat-spike (F and G) protein interactions present in ΦX174-like phages, but not in ΦEC6098 and other gokushoviruses.

VP1: the 3-fold protrusion and the hydrophobic core. Above the capsid surface at the 3-fold symmetry axes, three symmetry-related VP1 proteins came together to form a

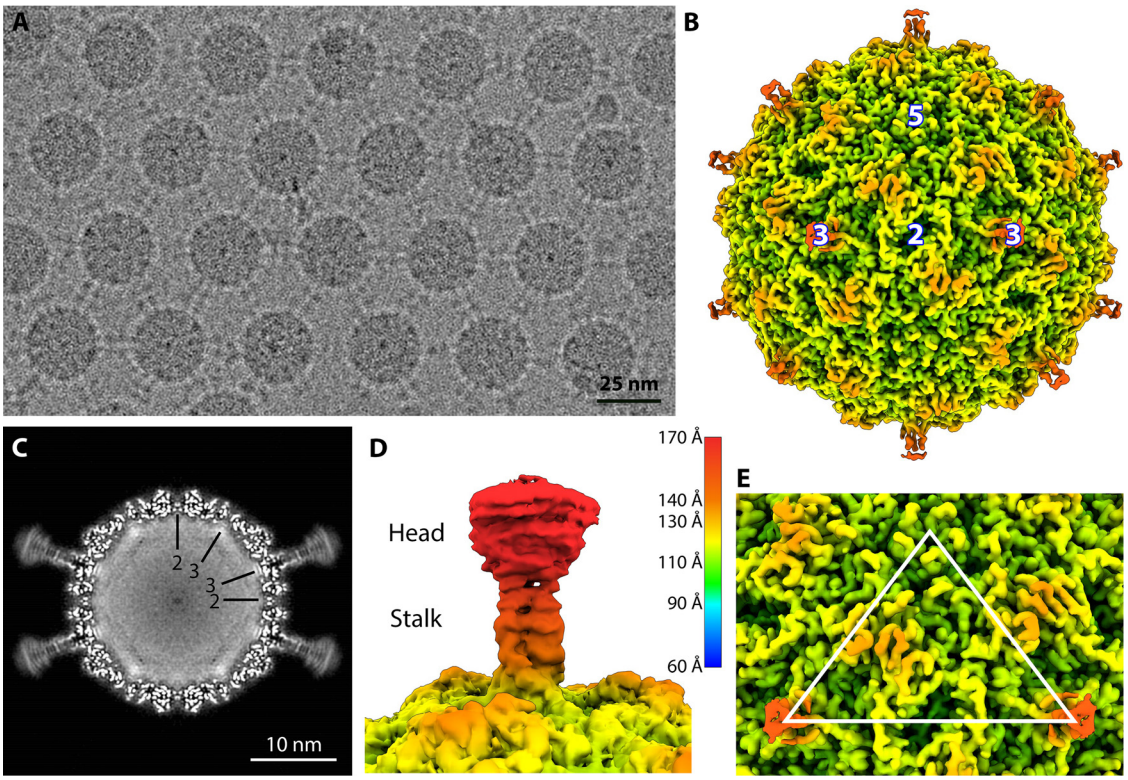

FIG 1 Cryo-EM Analysis of φEC6098. (A) A representative micrograph of ΦEC6098 from the data collection shows an array of capsids with adjoining protrusions (scale bar, 25 nm). (B) A surface-rendered icosahedrally averaged map of ΦEC6098 (colored radially according to key) was reconstructed to 2.6-Å resolution. The 3-fold protrusions were visible only at low contour due to weaker density. Icosahedral symmetry axes are marked (white numbers). (C) The central section of the ΦEC6098 capsid shows the 3-fold protrusions. The capsid is filled with amorphous cryo-EM density corresponding to the icosahedrally averaged ssDNA genome. Icosahedral symmetry axes are shown by black lines. (D) A surface rendering of the 3-fold protrusion shown at low contour illustrates the globular head domain and narrow stalk domain, colored according to key. (E) The asymmetric unit of the icosahedral cryo-EM map is shown, marked by a white triangle.

stalk. Specifically, EF insertion residues extend upward away from the 3-fold surface, form the globular domain, and return to the capsid. Residues corresponding to V212 to S221 could be built into the stalk density rising away from the capsid. However, the density after residue 221 was too poor to continue the build. Consequently, VP1 residues could not be built into the upper stalk and globular domain, but as the stalk returned to the capsid surface, the build could be resumed. Residues E310 to I320 were built into the stalk leading back to the 3-fold axis (Fig. 3A). The residues in the lower stem formed a large hydrophobic core (Fig. 3B and C), which was strengthened by three I563 residues from the C-termini of three capsid proteins. These hydrophobic core residues are conserved among gokushoviruses (Fig. 3D).

The top portion of the protrusion, including the globular head domain, could only be visualized at a low contour level (Fig. 1D). The cryo-EM density volume for the poorly resolved region, residues 222 to 309, was measured to be ~35,100 Å^3 at a 1.36-σ contour level, which was close to the predicted 34,434-Å^3 volume based on the number of residues. The low quality of the head domain reconstruction could indicate flexibility, heterogeneous structural conformations, or structures that do not conform to the applied icosahedral symmetry. As shown in the cryo-EM images, the particles in regular 2D arrays were connected via globular head interactions, possibly mediated by hydrophobic interactions (Fig. 1A). If contact-forming head domains have different structural conformations compared to free ones, this may also have contributed to heterogeneity.

We performed 3D classification of subvolumes that encompassed the head domain to improve the corresponding cryo-EM density; however, this approach did not improve either the resolution or map quality. The lack of improvement is likely due to

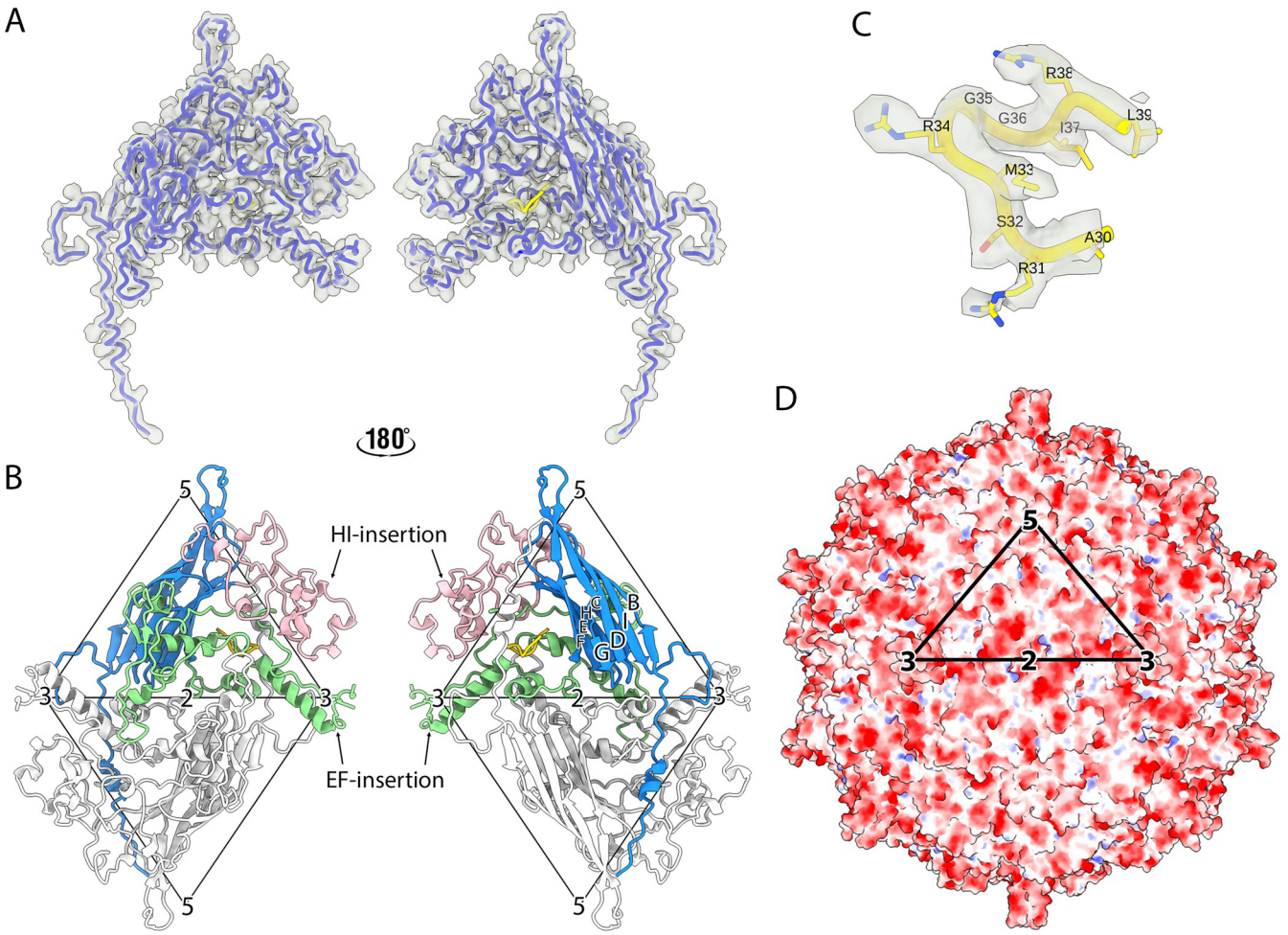

FIG 2 Atomic model of ΦEC6098. (A) The protein structure built into one asymmetric unit, which consists of one copy each of VP1 (dodger blue) and VP8 (yellow), shown in a licorice representation within the cryo-EM density map (transparent surface). The left and right images show the external and internal surfaces flipped by 180°. (B) The atomic model of VP1 has a core anti-parallel β-barrel motif (BIDG and CHEF) with an N-terminal extension (blue). The EF- (green) and HI-insertion loops (pink) are also highlighted. The adjacent asymmetric unit across the 2-fold symmetry axis is colored in light gray. The left and right images show the external and internal surfaces of the coat protein dimer to illustrate the position of the VP1 N-terminal extension (blue) and VP8 (yellow). The two icosahedral asymmetric units are marked by triangles, and the symmetry axes are labeled. (C) The ordered region of VP8 encompassing residues 30 to 39 (yellow) was built into the map density (transparent gray). The VP8 region prior to residue 30 was too disordered to build. (D) The electrostatic potential is shown for the icosahedral capsid surface of ΦEC6098 and colored red and blue for negative and positive charges, respectively. The black triangle designates the asymmetric unit.

the small molecular weight of the head domain density and limited particle numbers. We cannot rule out the effect of collecting data in linear mode instead of electron counting mode at higher detector quantum efficiency. Nevertheless, the final 2.6-Å map resolution is near the theoretical maximum, indicating that the Falcon 3 linear mode provides high-quality data with much shorter collection time and can be beneficial for large biological molecules like virus particles.

N-terminal extensions of VP1 interconnect the pentamers. The ΦEC6098 capsid protein contained a long N-terminal extension, residues 4 to 28 (Fig. 2A and B), not seen in the ΦX174-like phages. Underneath the capsid shell, the VP1 N-terminal residues passed through a cleft formed in the junction between two VP1s, interacting heavily with the two coat proteins (Fig. 4A). Specifically, residues 11 to 19 formed a short α-helix that filled the cleft in the VP1 junction. This short helix was adjacent to VP8 and contacted VP8 residues 16 and 19 to 21. The VP1 N-terminal residues prior to the short helix interacted with a patch of negatively charged surface area around the 5-fold symmetry axis (Fig. 4B and C). The most N-terminal VP1 residues interacted with hydrophobic patches formed by the VP1 residues beneath the 5-fold symmetry axis

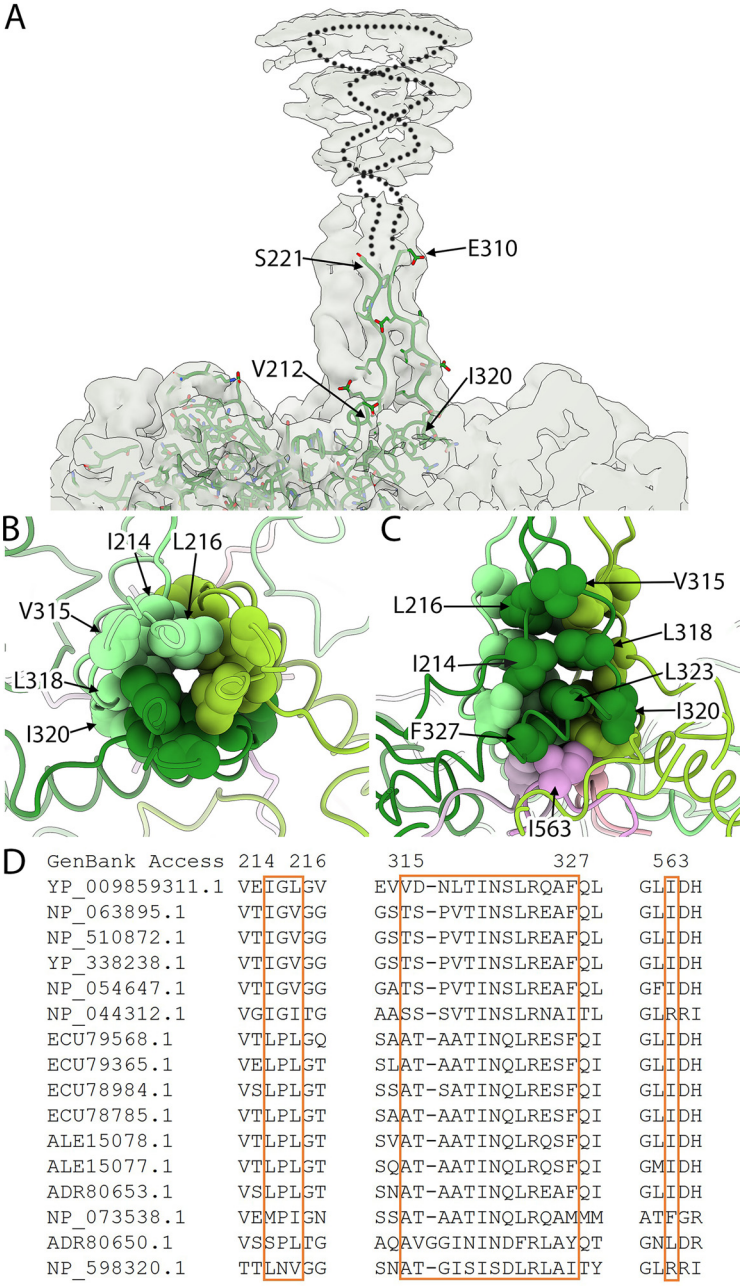

FIG 3 The 3-fold protrusion and hydrophobic core. (A) The 3-fold protrusion is depicted in a transparent surface rendering. A single VP1 built into one of the three copies of VP1 that form the trimeric stalk is shown in a stick representation (green). Rising from the capsid, residues were modeled up to residue 221, where the density became too disordered to continue. Extending from residue 221, a dotted line suggests a path to indicate that VP1 continues into the globular domain before tracing back toward the capsid surface. At residue 310, the build was resumed into the ordered stalk density. (B and C) Top and side views of the hydrophobic core at the center of the stalk region. (B) Three copies of VP1 (light, medium, and dark green) comprising the trimeric stalk are shown where the stalk adjoins the capsid surface. The residues contributing to the hydrophobic core are rendered as spheres and labeled. (C) At the bottom of the core are three I563 residues (light, medium, and dark pink). These residues are contributed to by three different, 2-fold symmetry related VP1 proteins, according to color. For example, the light green and light pink VP1 proteins are related by 2-fold symmetry. (D) Sequence comparison of the hydrophobic core (boxed residues) illustrates conservation among gokushoviruses. YP_009859311.1 is the accession code for the ΦEC6098 strain.

(Fig. S3). The N-terminal VP1 extensions made connections that linked the virion's 12 pentamers, forming a capsid-stabilizing network (Fig. 4D) that is not seen in ΦX174 and related phages G4 and α3. In this network, each VP1 N-terminal extension reached across to the neighboring pentamer and interacted with two of the five VP1s. In

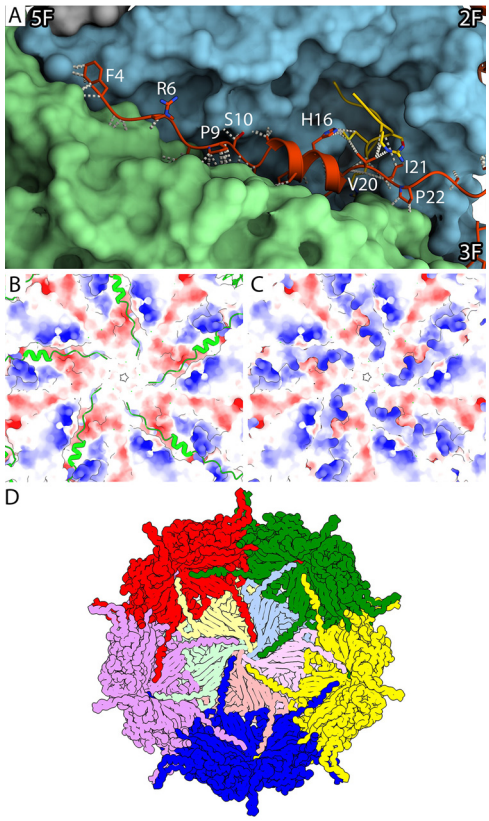

FIG 4 Characterization of the N-terminal extension of VP1. (A) A closeup view along a cleft in the capsid's inner surface where each VP1 N-terminal extension interacts with two neighboring pentamers. The long N-terminal extension (red) and VP8 (yellow) are depicted in a ribbon diagram with side chains. The neighboring pentamers (blue and green) are shown surface rendered to visualize the cleft. Representative VP1 residues are labeled, and icosahedral symmetry axes are indicated. (B and C) The electrostatic potential of the inner capsid's surface around the 5-fold symmetry axis is shown to illustrate that the positively charged N-terminal extensions (green) bind across the negatively charged clefts in the capsid inner surface (C). (D) A cartoon of the inner surface of the capsid demonstrates the N-terminal arm-interconnecting network. Coat protein pentamers (red, green, yellow, blue, purple) surround a pentamer illustrated with its five VP1s colored in tints (yellow, blue, pink, salmon, and green). Each N-terminal extension crosses over to an adjacent pentamer, where it makes connections in the cleft between two VP1 proteins, extending toward the 5-fold axis.

addition, the interconnecting extensions likely create a more stable environment for the packaged genome by neutralizing the acidic patches around the 5-fold axes.

DNA-binding protein VP8. Microvirus J proteins are divided into two domains. Their N-termini are highly basic and mediate DNA binding, whereas the hydrophobic C-termini mediate J-coat protein interactions. Unlike the C-termini of the ΦX174 and α3 J proteins, which are very hydrophobic and contain no basic residues, the ΦEC6098 VP8 C-terminus is less hydrophobic and contains three arginine residues (Fig. 5A). When the coat and DNA-binding proteins of ΦX174, α3, and ΦEC6098 were superimposed, the J and VP8 C-terminal residues overlapped (Fig. 5B). All three DNA-binding proteins interacted with their respective coat protein within a pocket formed by the EF and HI insertions (Fig. 5C and D). However, the interactions were quite different: VP1 and VP8 associated with 11 H-bonds (Fig. 5C), whereas the ΦX174-coat–J interactions made only five H-bonds (Fig. 5D), suggesting that a stronger interaction occurs in ΦEC6098.

VP1 N-terminal extension functionally replaces part of the J protein. The N-terminal DNA-binding domains of the ΦX174 and α3 J proteins were fully visualized in virion structures (1, 2, 7, 20). In contrast, VP8's arginine-rich region, presumably a DNA-binding domain, was not resolved. Instead, a portion of VP1, its most N-terminal α-helix, occupied this position (Fig. 4A, 6A, and 6B). No icosahedrally ordered DNA was found in our cryo-EM reconstruction, whereas the other bacteriophage structures resolved multiple ordered DNA nucleotides (1, 2, 7, 20). Instead, the ΦEC6098 icosahedral reconstruction exhibited

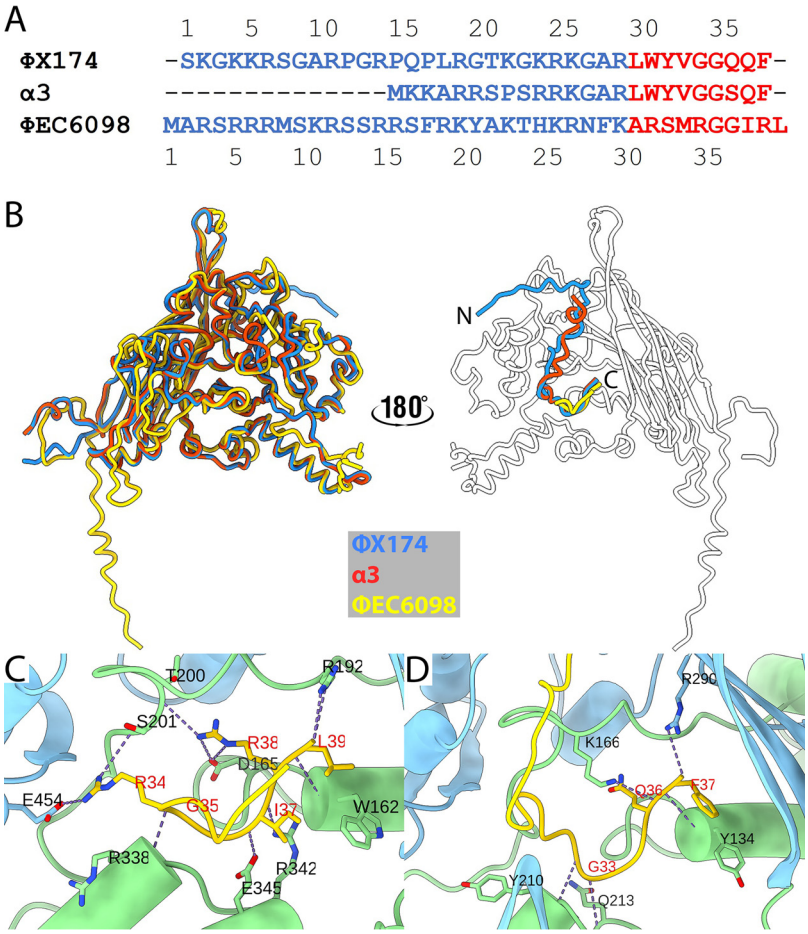

FIG 5 Comparison of the DNA-binding proteins. (A) Amino acid sequence alignment for the DNA-binding proteins from ΦX174, α3, and ΦEC6098. The residues for the N- and C-terminal domains are shown in blue and red, respectively. (B) Left, seen from the outside of the capsid, the coat and DNA-binding proteins are shown superimposed as ribbon diagrams colored according to the key. On the right, the ΦX174 (blue) and α3 (red) J proteins are superimposed onto VP8 (yellow) on the inner surface of the ΦEC6098 capsid protein (white). (C and D) ΦEC6098 VP8 (C) and ΦX174 J protein (D) bound into coat protein pockets formed by portions of the EF- (light green) and HI-insertion loops (light blue). H-bonds between the coat protein and DNA-binding protein are shown in dots, and the residues forming the H-bonds are labeled.

strong but irregular-shaped non-VP1 densities adjacent to the ordered region of VP8 near the capsid shell (Fig. 6C and D). This unfilled density may include VP8's DNA binding domain as well as parts of the DNA genome. The icosahedral averaged amorphous densities extended toward the 5-fold axis, where the symmetry-related densities were merged (Fig. 6C). The internal volumes for ΦEC9068 and ΦX174 were estimated to be 4.123×10^6 Å^3 and 4.253×10^6 Å^3, respectively (Materials and Methods section). ΦEC6098 has considerably less DNA to package into a capsid of similar volume (estimated DNA densities: 0.339 and 0.391 Da/Å^3 for ΦEC9068 and ΦX174, respectively); thus, the ΦEC6098 genome would be less constrained, allowing for greater flexibility within the capsid.

DISCUSSION

Two versus one scaffolding protein microvirus assembly system. The two ΦX174 scaffolding proteins divide the assembly pathway into two distinct phases. The internal scaffolding protein B mediates early assembly, resulting in the pentameric intermediates, which are, in turn, organized into the procapsid by the external scaffolding protein D. Gokushoviruses encode an internal scaffolding protein, VP3, which likely performs similar functions as the ΦX174 internal scaffolding protein B (17, 21). In the ΦX174 system, 240 copies of the external

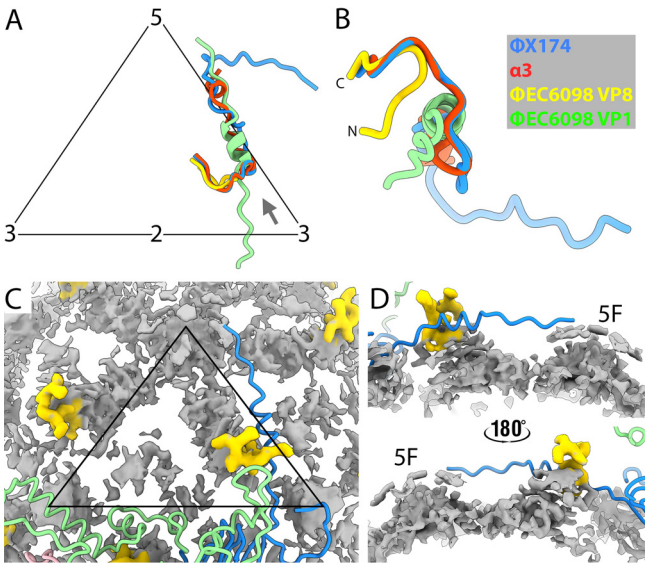

FIG 6 Superimposition of DNA-binding proteins and local disordered density. (A) Shown within the labeled asymmetric unit (black), ΦX174 and α3 J proteins are superimposed on the ΦEC6098 DNA-binding protein. The helical domain of the ΦEC6098 VP1 N-terminal extension (green) also maps to this region, where DNA is tethered to the inner surface of the capsid. (B) The superimposed DNA-binding proteins and N-terminal extension in panel A are rotated to view from the vantage point indicated by the gray arrow in panel A. The N-terminus of VP8 bends away from the superimposed J proteins because of the location of the VP1 N-terminal extension. (C) VP1 densities were subtracted from the cryo-EM map to show non-VP1 density in the interior (gray). Most of the density is amorphous and corresponds to genomic DNA plus the disordered region of VP8 that could not be built. However, the C-terminal VP8-ordered region (residues 30 to 39, yellow) was distinct. The icosahedral 2-fold symmetry related VP1 is shown in a ribbon diagram and colored as in Fig. 2B. Icosahedral symmetry axes are marked by a black triangle. (D) Side views of the amorphous density connected to the VP8 C-terminal domain. The 5-fold symmetry axis is labeled. The extra densities are likely the disordered portion of VP8 and associated ssDNA that do not conform to icosahedral symmetry.

scaffolding protein organize 12 pentameric intermediates into procapsids. Since there are little or no interactions between coat protein pentamers (14) across the 2-fold axes of symmetry (2, 22), the external scaffolding protein is critically important to ΦX174 procapsid stability. Gokushovirus does not have an external scaffold to organize or stabilize pentameric assembly subunits. In the absence of a Gokushovirus procapsid structure, insights on assembly are limited. However, it is possible that the N-terminus of the ΦEC6098 coat protein performs a stabilizing function to the Gokushovirus procapsid. In the virus structure, each N-terminus extends well across the 2-fold axis of symmetry (Fig. 2B), where it interacts with two subunits in the neighboring pentamer, forming a network beneath the capsid shell (Fig. 4D). This does not occur in ΦX174, where the N-terminus of the coat protein resides directly atop the 3-fold axis. Thus, it is likely that gokushovirus coat proteins, more specifically their N-termini, have assumed the function of the external scaffolding protein, by providing stability across the 2-fold. This may be further facilitated by an adjacent loop extending across the 2-fold axis that is absent in ΦX147.

DNA packaging and the 3-fold axes of symmetry. In the ΦX174 and α3 procapsid cryo-EM structures, there are pores at the 3-fold axes of symmetry that are not found in the virion. These pores are the only openings within the structure; thus, they are likely the site through which the genome is translocated. Genetic and biochemical data support this hypothesis (23, 24) and indicate that pore formation and stability are mediated by interactions between the external scaffolding and coat protein residues. As there is currently no published gokushovirus procapsid structure, it is unknown whether their procapsids would display similar pores. However, if the site of DNA packaging is conserved, either a region of the gokushovirus coat protein or the internal scaffolding protein would likely be involved in pore formation.

ΦX174 packaging requires the DNA-binding protein J, which is a small highly basic

peptide. The J protein C-terminus, which is strongly conserved among the ΦX174-like viruses (25), interacts with the coat protein, whereas the highly basic N-terminus interacts with DNA. As with the J proteins, VP8 C-termini are strongly conserved among gokushoviruses, whereas the highly basic N-termini display much variation, especially in the distribution of the basic amino acid residues. Although little or no homology exists between ΦX174-like J proteins and the VP8s, the overall architectures of these proteins are very similar and suggest similar functions. In ΦEC6098, the C-terminus of VP8 interacts with inner-surface VP1 grooves, a location reminiscent of J proteins. Compared to other microviruses, for which complete DNA-binding protein structures can be built into virion structures (1, 2, 7, 20), VP8's N-terminus was not ordered in the ΦEC6098 structure. In addition, no icosahedrally ordered DNA density was observed, whereas ~10% of the ΦX174 and α3 phosphate backbone is icosahedrally ordered. Although we identified non-VP1 density within the ΦEC6098 capsid that might include DNA along with the N-terminus of VP8 (Fig. 6C and D), the density is amorphous. Genome organization may be less critical in gokushoviruses. Their genomes are considerably smaller. The ΦEC6098 genome is 4.53 kb in length compared to 5.38 kb for ΦX174, which has the smallest genome of the ΦX174-like viruses. In contrast, the internal volumes of their capsids are quite similar, differing by only ~3.0%. It is possible that the tighter interaction between VP8 and the pocket is needed to counterbalance the increased flexibility of the DNA-attached VP8.

DNA penetration and the 5- and 3-fold axes of symmetry. The discussion herein and in the above paragraph presumes that the location of packaging and penetration are spatially conserved throughout the *Microviridae*. Thus, it should be regarded as somewhat speculative. For the ΦX174-like viruses, lipopolysaccharide is both necessary and sufficient for host-cell attachment and penetration (26). Reversible attachment is likely mediated by sugar-binding sites found in the coat protein (1, 27). Five such interactions would anchor one vertex to the host-cell membrane. Interactions with the negatively charged lipopolysaccharide (LPS) jettison the major spike complex at the infecting vertex, a penetration intermediate that has been directly observed (28). The ability of LPS to interact with the coat–spike interface appears to be the major extracellular host range determinant (29). Afterwards, the coat proteins at the infecting vertex rearrange, forming a pore through which the DNA pilot protein emerges to transport the genome (10, 28).

Although the ΦEC6098 receptor is unknown, gokushoviruses infecting chlamydia utilize a proteinaceous receptor (30), and this difference in receptor utilization may be reflected in the electrostatic surface charge of the ΦEC6098, which is mostly negatively charged or neutral. In contrast, ΦX174, G4, and α3 surfaces are more positively charged (1, 2, 4), which would facilitate LPS interactions.

The 3-fold protrusions in the ΦEC6098 structure seem flexible, a property that could facilitate host–virus interactions. Indeed, cell attachment studies with chlamydiaphages (30) and limited host range studies conducted with chimeric genomes between ΦEC6098 and closely related phages are consistent with this supposition (31). Receptor interactions with five protrusions surrounding an icosahedral vertex would bring the central vertex in contact with the host cell membrane, ultimately leading to the interactions that prompt pore formation and genome delivery. Considering the structural similarities observed between the ΦEC6098 and ΦX174 capsid protein at the 5-fold axes of symmetry, the model is feasible, albeit speculative.

MATERIALS AND METHODS

Generation and purification of ΦEC6098. Two liters of TKY media (1.0% tryptone, 0.5% KCl, 0.5% yeast extract) supplemented with 0.5 mM $CaCl_2$ and 10 mM $MgCl_2$ were preheated to 37°C then inoculated with 20 mL of a fresh overnight culture of *E. coli* strain JW2862-1 [F-, Δ(araD-araB)567, ΔlacZ4787 (::rrnB-3), λ-, ΔxerD745::kan, rph-1, Δ(rhaD-rhaB)568, hsdR514]. Cells were infected with ΦEC6098 at a multiplicity of infection of 10^{-3} and incubated overnight, shaking, at 37°C. Afterwards, the culture was chilled at 4°C for an additional 24 h, before the cellular debris, with attached virions, was concentrated by centrifugation. The pellet was resuspended in 30 mL of BE buffer (50 mM $Na_2B_4O_7$/3.0 mM ethylenediaminetetraacetic [EDTA]) and shaken for 18 h at 4°C. Cellular debris was removed by centrifugation. We layered 8.0 mL of supernatant atop CsCl step gradients (1.5 mL steps of 1.7 g/cm^3, 1.5 g/cm^3, and 1.3 g/cm^3 CsCl in BE buffer), which were spun for 2 h at 140,000 × *g* in a Beckman SW40.1 at 18°C. Material

with a density between 1.30 and 1.45 g/cm^3 was pulled from the gradients (approximately 2.0 mL) and dialyzed against 1 L BE buffer overnight at 4.0C° to remove the CsCl and then dialyzed twice against 1 L of sucrose gradient buffer (SGB: 100 mM NaCl, 5.0 mM EDTA, 6.4 mM Na_2HPO_4, and 3.3 mM KH_2PO_4 [pH 7.0]). After dialysis, the sample was concentrated to 250 μL in a Corning Spin-X column (molecular weight cutoff: 100 kDa). The concentrated sample was layered atop a 5 to 30% (wt/vol) sucrose gradient and spun for 1.0 h at 240,000 $\times$ *g* in an SW50.1 rotor at 4°C. Gradients were divided into ~40 125-μL fractions, and virions were detected by UV spectroscopy (A_{260} and A_{280}).

Negative stain TEM and cryo-EM. For negative staining, particle concentration was adjusted appropriately prior to applying 3.5 μL of sample to a glow-discharged, continuous carbon-coated copper grid. The grids were stained with 6 μL of freshly made 1.0% uranyl formate and imaged on a FEI (Hillsboro, OR) Tecnai G2 Spirit Biotwin (Huck Institute of the Life Sciences, Pennsylvania State University). For cryo electron microscopy, 3.5 μL of sample was applied to freshly glow-discharged 2/1 Quantifoil EM grids (Quantifoil Micro Tools, Jena, Germany), blotted for 1.0 s in 95% humidity at 4°C, and plunged into liquid ethane using a Thermo Fisher Scientific (Waltham, MA) Vitrobot. Vitrified grids were imaged in a Titan Krios microscope equipped with a Cs corrector (Huck Institute of the Life Sciences, Pennsylvania State University). Movies were collected using the Falcon-3 direct electron detector in linear mode at 54 frames, each lasting 1.35 s at a defocus range of 0.5 μm to 3.3 μm. Automated single-particle data acquisition was performed with a magnification of $\times$59,000. The absolute, or actual, pixel size of 1.1 Å/pixel was verified by scaling to a crystal structure in experiments with apoferritin.

Image processing and 3D reconstruction. A total of 2,152 movies were collected. All image processing was performed in RELION Version 3.1.beta (32) using icosahedral symmetry averaging. Frame alignment and dose weighting were performed with MotionCorr 2.0 (33), and contrast transfer function (CTF) parameter estimation was done with CTFFIND4.1 (34). Particles were initially picked by hand and then automatically selected using RELION's auto-picking algorithm. After 3D classification, 32,308 particles were used to generate a 2.6-Å resolution icosahedral map. DeepEMhancer was used to perform automatic postprocessing of the cryo-EM map (35).

Subparticle extraction and classification. Icosahedral subparticle extraction and correlative classification (ISECC) (36) was used to extract subparticle images corresponding to the 20 3-fold protrusions on each capsid. Specifically, ISECC_subparticle_extract with a vector for the 3-fold symmetry axis was used after icosahedral refinement to designate subparticles. Different values for the subparticle center and box size were tested. The extracted subparticle images were subjected to multiple rounds of 3D classification in RELION with various number of classes in an unsuccessful attempt to improve the density of the 3-fold protrusion (32).

Atomic model building. The initial model for the gokushovirus was built *de novo* using the Phenix tool phenix.map_to_model (37). The automated model-building procedure was performed against the 2.6-Å resolution capsid density map (unsharpened) with the viral capsid protein amino sequence (18). The icosahedral symmetry type was used to generate a partial model for an asymmetric unit, which contained 17 segments of polypeptides, including most of the beta strands and alpha helices. Although most of the Cα chains were traced unambiguously, the sequence assignment was not reliable except for that of two α helices, amino acids (aa) 149 to 163 (α-helix 2) and aa 320 to 339 (α-helix 4). All of the problematic residues were reassigned during model building and carried out in Coot (38) and ChimeraX for ISOLDE (39). The asymmetric unit was duplicated for 60 icosahedral units and refined with noncrystallographic symmetry using PHENIX real space refine (37). The refined model was visually inspected in Coot (38) and validated by MolProbity (40).

Structural display and analysis. Map visualization and images were generated in ChimeraX (41) and Bshow (42). The volume of the unresolved head domain of the 3-fold protrusion was estimated from the molecular weight of the peptide and an average protein partial specific volume by using the web server http://biotools.nubic.northwestern.edu/proteincalc.html. To measure the volume enclosed by virus capsids, the atomic models of the coat proteins from the ΦEC6098 and ΦX174 structures were used. Specifically, 8-Å resolution density maps were generated from the atomic models and the threshold was set where no holes were connecting the inside and outside. After splitting the inside surface from the outside surface by using the "pick blob" option in ChimeraX, the volume enclosed by the inside surface was measured. The DNA densities were estimated from the measured inner volumes and the molecular weights of the ssDNA genomes of ΦEC6098 (GenBank: MT185428.1) and ΦX174 (GenBank: NC_001422.1).

Data availability. The cryo-EM map of the gokushovirus ΦEC6098 has been deposited in the EM database (EMDB: EMD-27397) (https://www.ebi.ac.uk/emdb/), and the coordinates for the atomic model of the asymmetric unit have been deposited in the protein data bank (PDB: 8DES) (https://www.rcsb.org).

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

SUPPLEMENTAL FILE 1, PDF file, 1.5 MB.

ACKNOWLEDGMENTS

We kindly thank Paul Kirchberger for supplying the ΦEC6098 bacteriophage and cell lines. Research reported in this publication was supported by the Office of the Director, NIH, under award number S10OD026822-01 (S.L.H.), and the National Science Foundation, MCB

2013653 (B.A.F.). Funding was also provided by the Pennsylvania Department of Health Commonwealth Universal Research Enhancement (CURE) funds.

REFERENCES

1. McKenna R, Xia D, Willingmann P, Ilag LL, Krishnaswamy S, Rossmann MG, Olson NH, Baker TS, Incardona NL. 1992. Atomic structure of single-stranded DNA bacteriophage phi X174 and its functional implications. Nature 355:137–143. https://doi.org/10.1038/355137a0.
2. Bernal RA, Hafenstein S, Olson NH, Bowman VD, Chipman PR, Baker TS, Fane BA, Rossmann MG. 2003. Structural studies of bacteriophage α3 assembly. J Mol Biol 325:11–24. https://doi.org/10.1016/s0022-2836(02)01201-9.
3. Dokland T, McKenna R, Ilag LL, Bowman BR, Incardona NL, Fane BA, Rossmann MG. 1997. Structure of a viral procapsid with molecular scaffolding. Nature 389:308–313. https://doi.org/10.1038/38537.
4. McKenna R, Bowman BR, Ilag LL, Rossmann MG, Fane BA. 1996. Atomic structure of the degraded procapsid particle of the bacteriophage G4: induced structural changes in the presence of calcium ions and functional implications. J Mol Biol 256:736–750. https://doi.org/10.1006/jmbi.1996.0121.
5. Chipman PR, Agbandje-McKenna M, Renaudin J, Baker TS, McKenna R. 1998. Structural analysis of the spiroplasma virus, SpV4: implications for evolutionary variation to obtain host diversity among the *Microviridae*. Structure 6:135–145. https://doi.org/10.1016/s0969-2126(98)00016-1.
6. Diemer GS, Stedman KM. 2016. Modeling microvirus capsid protein evolution utilizing metagenomic sequence data. J Mol Evol 83:38–49. https://doi.org/10.1007/s00239-016-9751-y.
7. Bernal RA, Hafenstein S, Esmeralda R, Fane BA, Rossmann MG. 2004. The φX174 protein J mediates DNA packaging and viral attachment to host cells. J Mol Biol 337:1109–1122. https://doi.org/10.1016/j.jmb.2004.02.033.
8. Kirchberger PC, Martinez ZA, Ochman H. 2022. Organizing the global diversity of microviruses. mBio 13:e00588-22. https://doi.org/10.1128/mbio.00588-22.
9. Sawaya NA, Baran N, Mahank S, Varsani A, Lindell D, Breitbart M. 2021. Adaptation of the polony technique to quantify *Gokushovirinae*, a diverse group of single-stranded DNA phage. Environ Microbiol 23:6622–6636. https://doi.org/10.1111/1462-2920.15805.
10. Sun L, Young LN, Zhang X, Boudko SP, Fokine A, Zbornik E, Roznowski AP, Molineux IJ, Rossmann MG, Fane BA. 2014. Icosahedral bacteriophage ΦX174 forms a tail for DNA transport during infection. Nature 505:432–435. https://doi.org/10.1038/nature12816.
11. Sun L, Rossmann MG, Fane BA. 2014. High-resolution structure of a virally encoded DNA-translocating conduit and the mechanism of DNA penetration. J Virol 88:10276–10279. https://doi.org/10.1128/JVI.00291-14.
12. Roznowski AP, Fisher JM, Fane BA. 2020. Mutagenic analysis of a DNA translocating tube's interior surface. Viruses 12:670. https://doi.org/10.3390/v12060670.
13. Cherwa JE, Uchiyama A, Fane BA. 2008. Scaffolding proteins altered in the ability to perform a conformational switch confer dominant lethal assembly defects. J Virol 82:5774–5780. https://doi.org/10.1128/JVI.02758-07.
14. Cherwa JE, Organtini LJ, Ashley RE, Hafenstein SL, Fane BA. 2011. In vitro assembly of the øX174 procapsid from external scaffolding protein oligomers and early pentameric assembly intermediates. J Mol Biol 412:387–396. https://doi.org/10.1016/j.jmb.2011.07.070.
15. Siden EJ, Hayashi M. 1974. Role of the gene B product in bacteriophage φX174 development. J Mol Biol 89:1–16. https://doi.org/10.1016/0022-2836(74)90159-4.
16. Mukai R, Hamatake RK, Hayashi M. 1979. Isolation and identification of bacteriophage phi X174 prohead. Proc Natl Acad Sci U S A 76:4877–4881. https://doi.org/10.1073/pnas.76.10.4877.
17. Clarke IN, Cutcliffe LT, Everson JS, Garner SA, Lambden PR, Pead PJ, Pickett MA, Brentlinger KL, Fane BA. 2004. Chlamydiaphage Chp2, a skeleton in the φX174 closet: scaffolding protein and procapsid identification. J Bacteriol 186:7571–7574. https://doi.org/10.1128/JB.186.22.7571-7574.2004.
18. Kirchberger PC, Ochman H. 2020. Resurrection of a global, metagenomically defined gokushovirus. Elife 9:e51599. https://doi.org/10.7554/eLife.51599.
19. Liu S, Jin D, Lan R, Wang Y, Meng Q, Dai H, Lu S, Hu S, Xu J. 2015. Escherichia marmotae sp. nov., isolated from faeces of *Marmota himalayana*. Int J Syst Evol Microbiol 65:2130–2134. https://doi.org/10.1099/ijs.0.000228.
20. McKenna R, Ilag LL, Rossmann MG. 1994. Analysis of the single-stranded DNA bacteriophage φX174, refined at a resolution of 3·0 Å. J Mol Biol 237:517–543. https://doi.org/10.1006/jmbi.1994.1253.
21. Liu BL, Everson JS, Fane B, Giannikopoulou P, Vretou E, Lambden PR, Clarke IN. 2000. Molecular characterization of a bacteriophage (Chp2) from *Chlamydia psittaci*. J Virol 74:3464–3469. https://doi.org/10.1128/jvi.74.8.3464-3469.2000.
22. Ilag LL, Olson NH, Dokland T, Music CL, Cheng RH, Bowen Z, McKenna R, Rossmann MG, Baker TS, Incardona NL. 1995. DNA packaging intermediates of bacteriophage φX174. Structure 3:353–363. https://doi.org/10.1016/S0969-2126(01)00167-8.
23. Uchiyama A, Fane BA. 2005. Identification of an interacting coat-external scaffolding protein domain required for both the initiation of φX174 procapsid morphogenesis and the completion of DNA packaging. J Virol 79:6751–6756. https://doi.org/10.1128/JVI.79.11.6751-6756.2005.
24. Uchiyama A, Chen M, Fane BA. 2007. Characterization and function of putative substrate specificity domain in microvirus external scaffolding proteins. J Virol 81:8587–8592. https://doi.org/10.1128/JVI.00301-07.
25. Rokyta DR, Burch CL, Caudle SB, Wichman HA. 2006. Horizontal gene transfer and the evolution of microvirid coliphage genomes. J Bacteriol 188:1134–1142. https://doi.org/10.1128/JB.188.3.1134-1142.2006.
26. Michel A, Clermont O, Denamur E, Tenaillon O. 2010. Bacteriophage PhiX174's Ecological niche and the flexibility of its *Escherichia coli* lipopolysaccharide receptor. Appl Environ Microbiol 76:7310–7313. https://doi.org/10.1128/AEM.02721-09.
27. Ilag LL, McKenna R, Yadav MP, BeMiller JN, Incardona NL, Rossmann MG. 1994. Calcium ion-induced structural changes in bacteriophage φX174. J Mol Biol 244:291–300. https://doi.org/10.1006/jmbi.1994.1730.
28. Sun Y, Roznowski AP, Tokuda JM, Klose T, Mauney A, Pollack L, Fane BA, Rossmann MG. 2017. Structural changes of tailless bacteriophage ΦX174 during penetration of bacterial cell walls. Proc Natl Acad Sci U S A 114:13708–13713. https://doi.org/10.1073/pnas.1716614114.
29. Roznowski AP, Young RJ, Love SD, Andromita AA, Guzman VA, Wilch MH, Block A, McGill A, Lavelle M, Romanova A, Sekiguchi A, Wang M, Burch AD, Fane BA. 2019. Recessive host range mutants and unsusceptible cells that inactivate virions without genome penetration: ecological and technical implications. J Virol 93:e01767-18. https://doi.org/10.1128/JVI.01767-18.
30. Everson JS, Garner SA, Lambden PR, Fane BA, Clarke IN. 2003. Host range of chlamydiaphages φCPAR39 and Chp3. J Bacteriol 185:6490–6492. https://doi.org/10.1128/JB.185.21.6490-6492.2003.
31. Kirchberger PC, Martinez ZA, Luker LJ, Ochman H. 2021. Defensive hypervariable regions confer superinfection exclusion in microviruses. Proc Natl Acad Sci U S A 118:e2102786118. https://doi.org/10.1073/pnas.2102786118.
32. Zivanov J, Nakane T, Forsberg BO, Kimanius D, Hagen WJ, Lindahl E, Scheres SH. 2018. New tools for automated high-resolution cryo-EM structure determination in RELION-3. Elife 7:e42166. https://doi.org/10.7554/eLife.42166.
33. Zheng SQ, Palovcak E, Armache J-P, Verba KA, Cheng Y, Agard DA. 2017. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. Nat Methods 14:331–332. https://doi.org/10.1038/nmeth.4193.
34. Rohou A, Grigorieff N. 2015. CTFFIND4: fast and accurate defocus estimation from electron micrographs. J Struct Biol 192:216–221. https://doi.org/10.1016/j.jsb.2015.08.008.
35. Sanchez-Garcia R, Gomez-Blanco J, Cuervo A, Carazo JM, Sorzano COS, Vargas J. 2021. DeepEMhancer: a deep learning solution for cryo-EM volume post-processing. Commun Biol 4:874. https://doi.org/10.1038/s42003-021-02399-1.
36. Goetschius DJ, Hartmann SR, Subramanian S, Bator CM, Christensen ND, Hafenstein SL. 2021. High resolution cryo EM analysis of HPV16 identifies minor structural protein L2 and describes capsid flexibility. Sci Rep 11:3498. https://doi.org/10.1038/s41598-021-83076-5.
37. Afonine PV, Poon BK, Read RJ, Sobolev OV, Terwilliger TC, Urzhumtsev A, Adams PD. 2018. Real-space refinement in PHENIX for cryo-EM and crystallography. Acta Crystallogr D Struct Biol 74:531–544. https://doi.org/10.1107/S2059798318006551.
38. Emsley P, Lohkamp B, Scott WG, Cowtan K. 2010. Features and development of *Coot*. Acta Crystallogr D Biol Crystallogr 66:486–501. https://doi.org/10.1107/S0907444910007493.

39. Croll TI. 2018. ISOLDE: a physically realistic environment for model building into low-resolution electron-density maps. Acta Crystallogr D Struct Biol 74:519–530. https://doi.org/10.1107/S2059798318002425.
40. Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC. 2010. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr 66:12–21. https://doi.org/10.1107/S0907444909042073.
41. Pettersen EF, Goddard TD, Huang CC, Meng EC, Couch GS, Croll TI, Morris JH, Ferrin TE. 2021. UCSF ChimeraX: structure visualization for researchers, educators, and developers. Protein Sci 30:70–82. https://doi.org/10.1002/pro.3943.
42. Heymann JB, Belnap DM. 2007. Bsoft: image processing and molecular modeling for electron microscopy. J Struct Biol 157:3–18. https://doi.org/10.1016/j.jsb.2006.06.006.