The structure of a thermostable phage's Portal Vertex and Neck Complex illuminates its maturation process

Emma L. Sedivy[a], Emily Agnello[a,b], Julia E. Hobaugh[a], Rakeyah Ahsan[a], Kangkang Song[a], Chen Xu[a], and Brian A. Kelch[a¶]

[a] Department of Biochemistry and Molecular Biotechnology, University of Massachusetts Chan Medical School, Worcester MA
[b] Current affiliation: Institute for Applied Life Sciences, University of Massachusetts, Amherst MA

¶ Correspondence: brian.kelch@umassmed.edu, address: 364 Plantation St., Worcester MA 01605, phone: 508-856-8322

SUMMARY

Nearly all viruses use a proteinaceous capsid shell to protect their genome. In tailed bacteriophages and herpesviruses, the genome is packaged through the conserved Portal complex, a dodecameric ring embedded in the capsid at a single vertex. After genome packaging is complete, the Portal complex connects to the Tail, which is required for host recognition and infection. Portal and the proteins that connect it to the Tail form the Neck complex, which controls genome retention in the capsid. We determined the Neck and Portal Vertex structure of the hyperthermophilic phage P74-26 using cryo-Electron Microscopy (cryo-EM). Because this phage inhabits an extremely harsh environment, it is an excellent model to study principles of thermostability and viral maturation. We describe a thermostability mechanism wherein the Portal Vertex is stabilized by highly intertwined loop structures, like in a wicker basket. Moreover, by comparing the mature virion Neck to published structures of procapsids and empty mature capsids, we observe that the Portal protrudes from the capsid upon completion of genome packaging. We propose that this repositioning of the Portal underlies a pressure-sensitive switch that terminates genome packaging and triggers Tail attachment in many phages.

Keywords: viruses, bacteriophages, capsid, cryo-EM, thermophile

INTRODUCTION

Bacteriophages (phages), viruses that infect bacteria, are the most abundant biological entities on Earth[1]. Phages abound in every environment inhabited by bacteria, from human microbiomes to scalding geothermal vents to open oceans[2]. They have vital impacts on our lives because they shape bacterial populations as vectors of horizontal gene transfer, drivers of microbial evolution and population dynamics, and bactericidal agents[3]. Since their discovery in 1915, phages have been used as a model to understand the fundamental rules of molecular biology and as a source of therapeutic agents and molecular tools[4,5]. Nevertheless, the mechanisms powering many of their intricate components are still poorly understood. Closing these gaps in our understanding will enable biotechnology to take greater advantage of the opportunities that phages offer.

Most phages encapsulate their double-stranded DNA genome inside an icosahedral capsid and use a tail for host recognition and as a conduit for DNA passage into the cell[2]. The Neck Complex attaches the filled capsid to the tail. The Neck is situated in the capsid at a five-fold symmetric vertex, where a single pentamer of capsid proteins is replaced with the Neck proteins that provide access to the interior of the capsid. The most conserved component of the Neck is the homo-dodecameric Portal complex, which nucleates assembly of the initial, smaller procapsid form of the capsid[2,6,7]. The DNA packaging motor, Terminase, binds to the Portal Vertex and Portal provides a conduit for DNA to be packaged by Terminase into the capsid[8–11]. During packaging, the procapsid also expands to form a mature capsid in the shape of an icosahedron with the Portal at one vertex[12,13]. Double-stranded DNA phages generally replicate their genomes as concatemers that are subsequently cleaved by Terminase after a full genome is packaged into the capsid. There are several strategies the packaging motor can use to recognize the appropriate point at which to cleave the phage DNA. Some Terminases recognize a fixed sequence or position in the genome at which to cleave, cutting at the same position in each phage. In contrast, in headful phages, Terminase cuts when the capsid is full, cutting at a slightly different position each time and generating phages with circularly permuted genomes[12]. It is not known how headful phages signal to Terminase that the capsid is full and DNA should be cleaved, although it is clear that Portal plays a major role in the headful signal[14]. After packaging is complete, Terminase dissociates and Portal changes conformation to retain packaged DNA[8,15]. It then binds to the tail, creating the Neck complex that joins the icosahedral capsid head to the tubular tail. Because the capsid and tail have different symmetries, the Neck must accommodate the mismatch between these components[8]. The phage genome must be retained inside the capsid until the host is found, as premature genome release renders the virion inert. However, the packaged genome generates extremely high pressure inside the capsid, which must be contained by both the capsid and the Neck[16,17]. Attachment to the host triggers irreversible rearrangements in the virion and drives genome ejection through the Neck and the tail into the host, where the phage can replicate[18,19].

Phages following this basic architecture and life cycle make up most known viruses and are found in every environment, including the human microbiome[2]. Furthermore, capsid and Portal components are conserved among human pathogens in the Herpesvirus order, including those that cause chickenpox, shingles, infectious mononucleosis, and Kaposi's sarcoma[20,21]. In contrast to the prodigious diversity of phages, phage biology has historically focused on a handful of model phages that infect lab strains of bacteria[2]. Examining a greater diversity of phages, especially those that inhabit extreme niches, may help to uncover general principles of phage biology. The nearly identical phages P74-26, P23-45, and G20C infect the bacterium *Thermus thermophilus* HB8 in geothermal hot springs at 65-75 ˚C and have the longest tails of all known phages[22,23]. These phages have recently been developed as a model system for revealing basic phage assembly mechanisms because of the high thermostability and novel mechanisms these phages utilize to overcome the challenges of their extreme environment[15,24–31]. We and others described features of the P74-26 capsid and tail tube proteins that enhance their stability[28,29,31], and others described the procapsid Portal Vertex of P23-45[15].

Their extreme habitat intensifies a problem faced by all viruses: protecting the genome until the appropriate time to release it. The challenge of retaining the genome is even greater at elevated temperatures, and the Portal Vertex creates a unique vulnerability in the capsid shell. We wondered how a long-tailed thermophilic phage makes a durable Portal Vertex and Neck. Furthermore, it remains unknown how the Portal Vertex rearranges during the viral maturation process. We can now reveal these mechanisms by determining the structure of the Neck in the mature virion and comparing it with the structures of the Portal Vertex in the procapsid and empty mature capsid states[15,28].

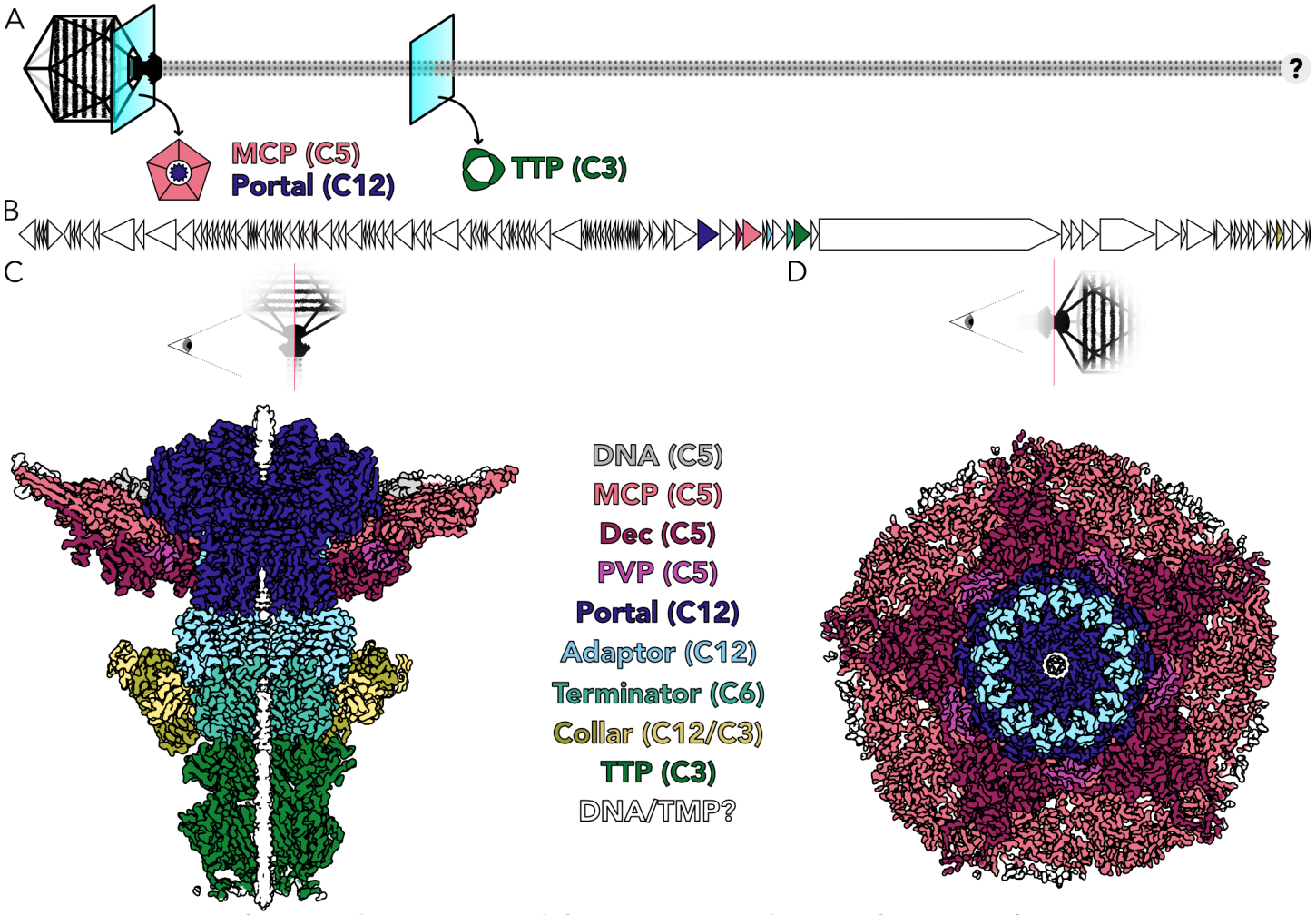

Figure 1: Reconstruction of P74-26 Neck components revealed proteins not previously annotated. A. Diagram of the P74-26 virion. Cross-sections show the symmetry of previously studied phage components. **B. Linear map of the P74-26 genome, with the location of structural genes highlighted.** Each ORF in the 82kb genome is displayed as a triangle proportional to its length. Genes are colored according to the key in C. **C. Cross-section of a composite reconstruction of the P74-26 Neck complex.** The C5 and C3 maps of the Neck complex were aligned by fitting into the asymmetric map and are colored according to protein identity. The symmetry of each component is noted in parentheses. C5 DNA is found inside the capsid, wrapping around Portal.The Portal Vertex Protein (PVP), Adaptor, Terminator, and Collar were not previously annotated. Unmodeled density (white) threading through Portal likely represents DNA, while unmodeled density threading through Terminator and TTP likely represents Tape Measure Protein (TMP). **D. View of the portal vertex from outside the capsid.** The composite reconstruction in C was rotated 90° and sliced to show the portal vertex.

We report the structure of the P74-26 Neck at high resolution using cryo-EM and identify its constituent proteins. Analysis of this structure provides insight into mechanisms of thermal stability of large complexes and how pressure regulates the maturation of headful bacteriophages.

RESULTS

The P74-26 Neck includes regions of C12, C6, C5, and C3 symmetry

We determined the structures of the P74-26 Neck by cryo-EM. The organization of the Neck was not known, but the P74-26 tail tube has 3-fold symmetry[31] and the Portal complex is dodecameric[15,28]. The capsid vertex at which the Neck is located has five-fold symmetry[8] (Fig. 1A). Known components of the Neck and Portal Vertex include the Portal, Decoration Protein (Dec), Major Capsid Protein (MCP), and Tail Tube Protein (TTP), which are located near each other in the P74-26 genome along with many unannotated open reading frames that may include Neck proteins (Fig. 1B)[23,29].

We first reconstructed the P74-26 Neck with 12-fold symmetry imposed (Fig. S1-3). There was significant variation in local resolution, and regions that exhibit other symmetries (such as capsid, packaged genome, and tail tube) were blurry, as expected (Fig. S4A). This reconstruction reached 2.3-Å Gold Standard Fourier Shell Correlation (GSFSC) resolution with a mask excluding off-symmetry regions of the map (Fig. S4A). The Neck forms a hollow tube with a central channel ~45 Å in diameter, similar to the lumen of the tail tube[31]. There is amorphous density running through the central pore, likely representing the Tape Measure Protein (TMP) and/or packaged DNA (Fig. 1C). The N-terminal region of TMP is predicted to reside in the Neck of similar long-tailed phages[19]. However, this density was not clearly resolved even after subtracting the signal of other parts of the complex, likely because its symmetry is not aligned with that of the complex overall and/or because it is too small for proper alignment.

The C12 P74-26 Neck reconstruction includes density consistent with the mushroom-shaped Portal that sits in the aperture of the capsid opening[15,28]. In P74-26, the Portal is ~120 Å diameter at its widest point and is connected to the tail tube by a region that is ~80 Å in length and has 12 lobes protruding to 200 Å wide. We termed this ring around the Neck the collar region. The lobes of the collar did not achieve high resolution with 12-fold symmetry imposed. To separate virions with conformational or compositional heterogeneity in this region, we 3D classified the particle stack with a mask applied to the region (Fig. S3). We reconstructed two classes, one lacking the lobes entirely and another with a complete set of lobes in the collar region. The

resolution of the lobes was still weak relative to the rest of the Neck (Fig. 4A). Because there were no other structural changes between Necks with and without the collar lobes, we concluded that they are not required for structural integrity of the virion but may be required or beneficial for the phage during infection (Fig. S4B).

To visualize the region of the Neck that contacts the C3 tail tube, we performed symmetry expansion of the particle stack around the C12 symmetry group followed by 3D classification with a mask applied to the collar region (Fig. S5). Duplicate particles were removed from the largest class, which was then refined with three-fold symmetry to 2.6-Å GSFSC resolution (Fig. S7A). With three-fold symmetry imposed, it became clear that while the core of the collar has 12-fold symmetry, the region that interacts with the first ring of the tail tube has 3-fold symmetry (Fig. S5, S16). In addition, some of the lobes of the collar region may be flexible and/or unstructured, explaining their poor resolution (Fig. S16). We also applied this workflow to a stack of particles that were selected because they resembled the C12-refined Neck without density for the capsid (Fig. S6). This particle stack yielded a 2.7-Å resolution reconstruction that was identical to the complete Neck map generated in Fig. S5, except for the absence of the capsid itself (Fig. S7B). We hypothesize that these particles were formed by removal or fragmentation of the capsid and conclude that the survival of the Neck highlights its durability.

To visualize how the capsid accommodates the aperture in which the Portal sits, we refined the same particle stack with five-fold symmetry imposed to 2.9-Å GSFSC resolution (Fig. 1D, S8-9). In addition to the capsid vertex, we also resolved a pseudo-circle of DNA that adopts apparent five-fold symmetry around the portal. To visualize how the C5 capsid interacts with the C12 portal, we performed symmetry expansion of the particle stack around the C5 symmetry group followed by 3D classification with a mask applied to the Neck (Fig. S10-11). Duplicate particles were removed from the largest resulting class, and it was refined asymmetrically. This reconstruction reached 3.5-Å GSFSC resolution (Fig. S11).

P74-26 Neck proteins have structural similarity to Neck proteins of other phages

Other than Portal, Dec, MCP, and TTP, we did not know the identities of the proteins that compose the Neck region. P74-26 has low sequence similarity to other phages, and the genome does not contain any hypothetical proteins with sequence similarity to known Neck proteins. To identify the proteins in our reconstructions, we leveraged the automated cryo-EM atomic model-building software ModelAngelo[32]. After performing unbiased model-building (without sequence input), we used the NCBI-BLAST server to search for sequence similarity between built peptides and all known proteins in the non-redundant RefSeq database (Table S1). Using this approach, we were able to identify all Neck proteins in our reconstructions. In the C3 map, ModelAngelo built peptides corresponding to the Portal Protein and Tail Tube Protein (TTP), as expected, as well as gp90, gp92, and gp112 of P74-26, previously annotated as hypothetical proteins (Fig. 1B, C). Into the C5 map, ModelAngelo built peptides corresponding to the Major Capsid Protein (MCP) and Decoration Protein (Dec), as expected, as well as hypothetical protein gp34 (Fig. 1B,D). To confirm our assignments, we used mass spectrometry to identify the proteins that are present in the mature virions (Table S2). All putative Neck proteins are present in the mature virion samples as measured by mass spectrometry.

We then performed ModelAngelo model-building, this time including the consensus sequences of identified P74-26 proteins as input. These models were manually adjusted and refined (Table S3). We used the DALI structural comparison server to search for proteins with similarity to these models. We found that gp34, which we named Portal Vertex Protein (PVP), is similar only to the gp2 RNA Polymerase inhibitor of phage T7[33] and an auxiliary capsid protein of phage ΦcrAss001 (Fig. S12A, S13)[34]. We found that gp90 is structurally similar to Adaptor proteins present in the Neck of a wide variety of tailed phages, and therefore named it Adaptor protein (Fig. S12B, S14)[35]. We found that gp92 closely resembles the phage Lambda tail terminator protein gpU and therefore named it Terminator protein (Fig. S12C)[36,37]. gp112, which we named Collar protein, consists of two Immunoglobulin (Ig) domains, common in phage Necks and Tails[38], connected by a flexible linker (Fig. S12D, E, S15). The ModelAngelo results, mass spectrometry data, and structural similarity to known Neck components give us high confidence in our protein assignments.

The Portal Vertex of the capsid is bordered by five hexameric capsomers of MCP, with trimers of Dec at the junctions between capsomers. Five subunits of PVP also border the aperture, bridging Dec trimers (Fig. 1D). The Portal complex sits inside this aperture. A dodecamer of Adaptor forms a ring that joins the Portal dodecamer to the hexameric ring of Terminator, which in turn connects to the first trimeric ring of TTP in the Tail Tube. The twelve lobes of the collar region encircle the Terminator ring, each formed by a dimer of Collar proteins. Within the Collar dimer, the two N-terminal domains interact with each other, leaving the two C-terminal domains free (Fig. S16). One of these C-terminal domains forms the distal part of each lobe in the upper collar, extending perpendicular from the main tube of the Neck. The other C-terminal domain of each dimer extends towards the first ring of TTP, although only six of them can be resolved. The resolved six C-terminal domains form the lower collar and interact with TTP in three pairs arranged with three-fold symmetry around the ring of TTP, which also has three-fold symmetry. Although the remaining six C-terminal domains are not clearly resolved, when the contour threshold is relaxed, three areas of density appear in the gaps where C-terminal domains are missing. This density likely represents the missing domains of the lower collar, which have been displaced by steric clashes with TTP and are either flexibly tethered or unstructured (Fig. S16).

Portal is in the closed conformation

The Portal complex plays vital roles in the Necks of tailed phages and herpesviruses, providing a conduit for DNA to be translocated into the capsid. P74-26 Portal protein, like other Portals, is dodecameric and forms a mushroom-shaped complex with a central pore. The crown domain forms the top of the mushroom, with the Wing domain forming the outer rim of the mushroom cap and the clip domain forming the stem (Fig. 2A). The Portal has been proposed to act as a gatekeeper during DNA packaging and genome ejection, with its central pore changing conformation to either allow passage of DNA or to retain it in the capsid[15,39,40]. Previous studies of closely related thermophilic phages have reported structures of the Portal complex in procapsids and in crystals (PDB 6QJT and 6IBG, respectively). The authors hypothesized that the procapsid Portal is in an open conformation, allowing passage of DNA through the pore, while the crystallized form of Portal represents the conformation in mature virions that retains DNA inside the capsid[15,28]. On the other hand, studies of T7 phage suggest that the Portal is in the open conformation and that the Nozzle protein at the tip of the tail is the gate that retains DNA[41]. Thus, it is unclear whether Portal is the gatekeeper that retains DNA in long-tailed thermophilic phage.

Supporting Portal's role as a gatekeeper, the conformation that we observe in mature P74-26 virions matches the closed form of the crystal structure (C_a RMSD = 1.028 Å; Fig. 2B). In contrast, the Portal conformation in mature virions is inconsistent with that found in Procapsids (PDB 6QJT), with substantial differences in the crown domain and the pore loops and smaller alterations in the clip domain (Fig. 2B). As reported by Bayfield *et al.*, the procapsid form of Portal has a wider and more hydrophilic pore, while the crystal and mature Portal's pore is narrower and more hydrophobic (Fig. 2C). We were surprised to find that none of the other Neck proteins formed constrictions in the Neck pore that could contribute to genome retention, as have been found in other similar phages[42,43]. Furthermore, we fortuitously isolated a form of phage P74-26 from a band that migrates lower than intact virions in a Cesium Chloride gradient in which the tail is broken along the length of the tail. Negative stain EM and Cryo-EM imaging of these particles clearly shows DNA retained in the head, even though the Tail Tip Complex is missing (Fig. 2D, S17). These particles indicate that the Portal complex is the main gate for

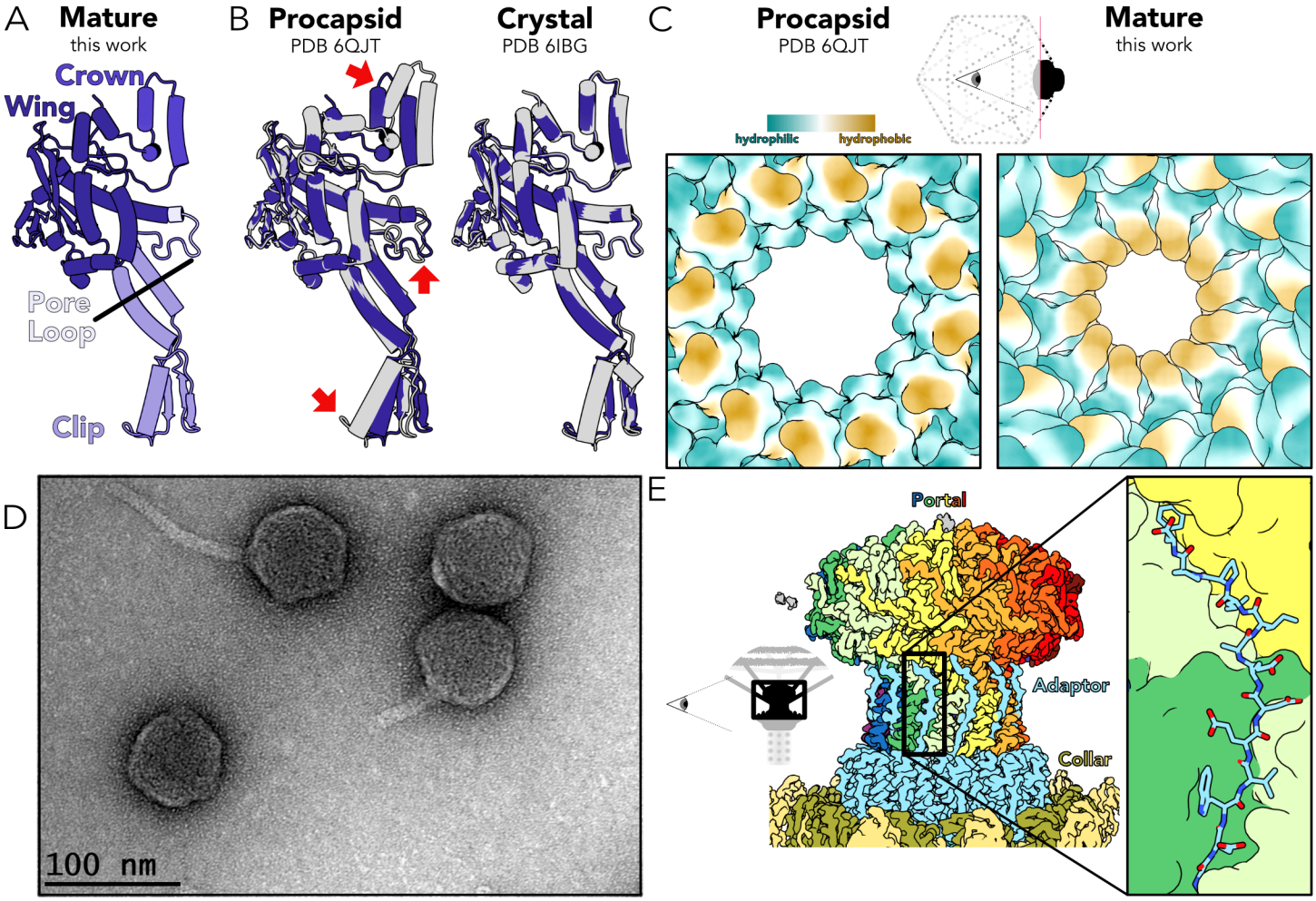

Figure 2: Portal protein is in the closed conformation and forms strong interactions with the C-terminal tail of Adaptor. A. One subunit of the mature virion Portal protein. Significant domains of the portal protein are noted. **B. Comparison of mature Portal protein (dark blue) with subunits of the Portal protein crystal and procapsid Portal protein (grey).** Regions with significant differences (clip, crown, and pore loop) are noted with red arrows. **C. Comparison of the Portal Pore Loop constriction in procapsids and mature virions.** Surface view of the models of the procapsid (open) and mature (closed) Portals colored according to hydrophobicity. **D. Virions with broken tails imaged by negative stain EM.** A band migrating below mature virions in a Cesium Chloride gradient was purified and stained with Uranyl Acetate. **E. EM density of the Neck complex with Portal in rainbow.** Adaptor and collar proteins are colored as in Figure 1. Inset shows the atomic model of the Adaptor tail (sticks) interacting with the clip region of three different Portal subunits (surfaces).

retaining DNA inside the capsid. It remains unknown how the Portal transitions from an open conformation to a closed conformation, and what provides the signal to trigger this change.

Consecutive rings transition from twelve-fold symmetry of Portal to the three-fold symmetry of the tail tube

The Adaptor protein forms a dodecameric ring and uses its C-terminal tail to attach to the Portal clip region (Fig. 2E). Adaptor tails bind in the clefts formed in the clip domain where adjacent Portal subunits meet. Each Adaptor tail contacts three subunits of Portal (Fig. 2E). The interface between the Adaptor tail and the Portal clip exhibits complementary hydrophobic (Fig. S18B) and electrostatic interactions (Fig. S18C). For example, the C-terminal residue (Phe130) inserts its sidechain into a hydrophobic pocket between the clip and wing domains, while the C-terminal carboxylate moiety forms electrostatic interactions with basic residues of the clip domain. Thus, the Adaptor interaction appears strong, which makes sense given its vital role in anchoring the 0.9 µm Tail onto the capsid.

The hexameric Terminator bridges the three-fold symmetry of the Tail Tube to the twelve-fold symmetry of Adaptor and Portal. Two Adaptor subunits interact with the "top" loops of one Terminator subunit through electrostatic and hydrophobic interactions (Fig S19). Meanwhile, a long loop at the "bottom" of each Terminator subunit binds to the sockets of the three-fold symmetric TTP ring. This arrangement mimics the loop-and-socket geometry that TTP rings use to stack into the 0.9 µm long Tail Tube (Fig. S20)[31]. Each TTP subunit contains two sockets that both need to be occupied, for a total of six sockets per TTP ring. Each of the six Terminator subunits contains a single loop, which can insert into either of the two TTP sockets. In contrast, in a typical stack of rings of the tail tube, the two loops from a TTP in the ring above insert into the sockets of a TTP subunit of the ring below[31]. The two loops of TTP are each specific to one socket, with Loop1 binding to Socket1 and Loop2 binding to Socket2. In contrast, the loop of Terminator is promiscuous and occupies either socket of TTP. Likewise, the sockets of TTP can be occupied by loops either from Terminator or from another TTP subunit. To accomplish this promiscuity, all loop-and-socket interactions use similar anchor points that consist of complementary hydrophobic and electrostatic sites (Fig. S20). The overall loop configuration varies widely between Terminator and TTP (Fig S20D), but these anchor points are found in all interaction pairs and likely account for a large proportion of the binding energy between these components.

Alterations to capsid vertex allow it to accommodate the Neck

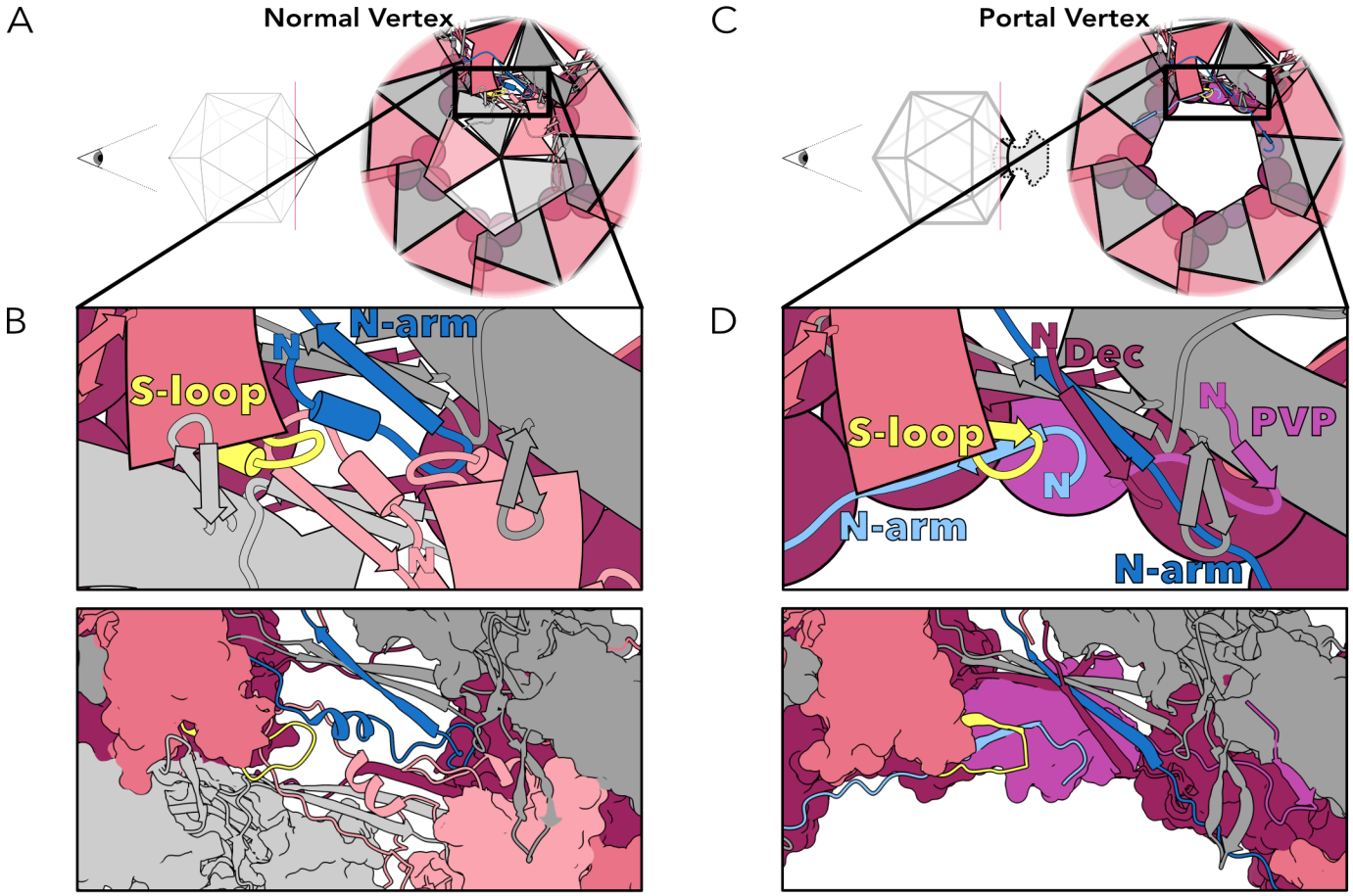

Figure 3: The Portal vertex is strengthened by topological linkages between MCP, Dec, and PVP. A,B. Interactions between a hexon MCP and the neigboring penton at a normal vertex (PDB 6O3H). Hexon MCPs (pink and grey) and penton MCPs (light pink and light grey) interact through their S-loops and N-arms. The hexon MCP S-loop (yellow) and N-arm (blue) interact with the penton N-arm (light pink), while the N-terminus forms an intra-subunit interaction with its P-domain. The N-arms of Dec (maroon) interact with the outer surface of the capsid. **C,D. Interactions between MCP, Dec, and PVP at the unique Portal vertex (Portal not shown).** Colors are as in A, except as noted. Rather than folding back to interact with the penton, the MCP N-arm (blue) extends into the hexamer to the right. The MCP N-arm from the hexamer to the left (light blue) interacts with the S-loop (yellow). While both the N-arm and S-loop form helices in the icosohedral MCP, they form β-strands that interact with each other in the unique vertex. The N-arms of one Dec monomer (maroon) and PVP (magenta) fold into the pore to interact with the inner surface of the capsid.

The phage capsid must accommodate the Neck at one vertex without sacrificing integrity. The capsid is composed of sub-assemblies of MCP called capsomers, which can be either pentameric (pentons) or hexameric (hexons), and which are held together through specific interactions of MCP (Fig. S21). The capsid contains eleven pentons, each at one of the twelve vertices of the icosahedron. Each penton is secured to five hexons using helical elements in the N-arms and S-loops of MCP (Fig. 3A,B). Dec protein strengthens the inter-capsomer interface, with the N-arms of two Dec subunits interacting with the E-loops of one penton MCP and one hexon MCP on the exterior surface of the capsid. The twelfth vertex, the Portal vertex, replaces the penton with the aperture in which the Neck sits (Fig. 3C), leaving the hexons to compensate for these important interactions that normally occur with a penton.

The hexons are rearranged in the Portal vertex to maintain stability without a penton (Fig. 3C,D). Instead of folding back to interact with its own P-domain, the MCP N-arm stretches into the neighboring hexon to interact with an S-loop there. Where the MCP N-arm and S-loop both form helices in typical hexons, in the Portal vertex they interact through two β-strands, illustrating the conformational flexibility of these regions. In addition, the Dec N-arm threads through the aperture to the interior of the capsid, where it interacts with a hexon's E-loop (Fig. 3D, S22B). In contrast, it interacts with an E-loop of the penton on the exterior of the capsid in a typical vertex.

The Portal vertex is further stabilized by the addition of PVP at the edge of the aperture. While in a normal vertex the E-loop forms a four-stranded β-sheet with strands from the neighboring MCP N-arm and Dec N-arm (Fig. S22A,C), in the Portal vertex this sheet becomes eight-stranded with the addition of three strands of PVP on the exterior of capsid and one strand of the Dec N-arm folded into the interior (Fig. S22B,D). Though the globular domain of PVP sits on the exterior of the pore, it also has an N-arm that folds through the aperture to join a β-sheet with the P-domain of MCP on the interior of the capsid (Fig. 3D). Thus, all three components of the Portal vertex are intricately interwoven, forming multivalent interactions that prevent the aperture from widening and maintaining structural integrity of the capsid. The intertwined topology of proteins stabilizing the Portal Vertex is reminiscent of the topological links that link the rest of the capsomers[29].

Portal is repositioned in response to DNA packaging

Our C5 symmetric reconstruction of the Neck revealed a loop of DNA around the Portal wing region, with clearly visible major and minor grooves (Fig. 4A). All three capsid proteins (MCP, Dec, and PVP) directly interact with the DNA, thus imparting pseudo-five-fold symmetry to this segment of DNA. Lys2 of the N-arm of PVP inserts into the minor groove (Fig. 4C), while the T-loop of MCP inserts into the major groove (Fig. 4D), and the Dec N-arm contacts the DNA backbone (Fig. 4E). Because the DNA interacts so closely with the capsid and is positioned directly around the

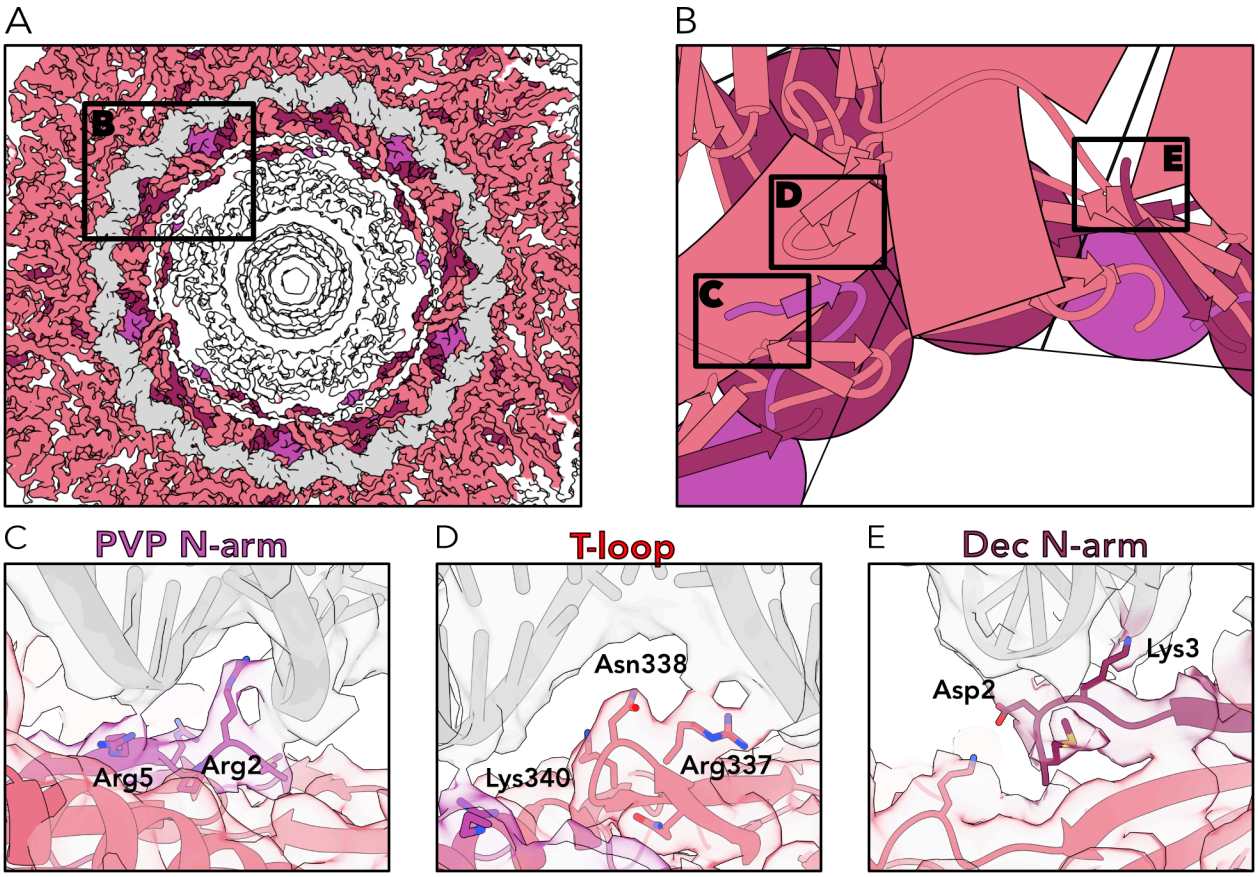

Figure 4: One loop of packaged DNA surrounds Portal with 5-fold symmetry. A. The C5 map, viewed from inside the capsid, colored by protein identity as in Figure 1 with DNA in grey. The area enlarged and diagrammed in B is marked. **B. Diagram of DNA-contacting capsid regions (DNA not shown).** The areas enlarged in panels C-E are marked. **C. Interactions between the N-arm of PVP and DNA.** Arg2 inserts into the minor groove, while Arg5 interacts with the DNA backbone. **D. Interactions between the T-loop of MCP and DNA.** Arg337 interacts with the DNA backbone, while Asn338 inserts into the major groove. **E. Interactions between the N-arm of Dec and DNA.** Asp2 and Lys3 interact with the DNA backbone.

Portal wing, it could be influencing the conformation of the Portal Vertex.

Our structure of the P74-26 Neck reveals that the Portal drastically alters how it contacts MCP during viral maturation. In the procapsid, the Portal wing domain interacts with a T-loop (T-loop1) of MCP, while the Portal clip domain interacts with the MCP S-loop and P-domain rather than a T-loop (Fig. 5A,B, S23)[15]. In contrast, in the mature virion the Portal wing still contacts a T-loop of MCP, but it contacts the T-loop of a different MCP subunit within the same hexon (T-loop2). This movement allows T-loop1 to bind the pseudo-5-fold symmetric ring of packaged DNA wrapped around the Portal (Fig. 5B, S23C). Because T-loop2 is lower in the capsid than T-loop1, Portal must move toward the outside of capsid pore, pushing the clip domain outside of the capsid so that it does not bind MCP at all but is instead exposed to bind the Adaptor C-terminal tail. When aligned on T-loop1, the geometric center of Portal moves downward by over 22 Å (Fig. 5C). This movement is not due to the internal motions of Portal, because the geometric centers of the procapsid Portal and mature virion Portal change minimally when the two are aligned against each other (~1 Å) (Fig. S24). Thus, we identified a dramatic repositioning of the Portal relative to the capsid shell, which is coupled to the packaging of DNA.

To determine whether this repositioning is due to the changes in the capsid conformation or to the pressure of packaged DNA, we compared the mature Portal and Portal Vertex to those of expanded empty capsids, which have a capsid conformation similar to the mature virion but lack tails and packaged DNA[28]. Empty expanded capsids can be isolated from phage preparations and most likely represent an assembly or disassembly byproduct. To compare the expanded empty capsid structure to the mature virion, we fit our model of the mature Portal and capsid shell into the C5-symmetric map of expanded empty capsids (EMD4446)[28] (Fig. S25A).The mature capsid hexons fit the corresponding region in the map of expanded empty capsids, but the Portal does not fit into the map, indicating that Portal changes position between mature virions and expanded empty capsids. Compared to mature virions, the density for Portal moves into the capsid lumen and less of the clip protrudes on the exterior of the capsid in capsids lacking packaged DNA.

Because the expanded empty capsid was reconstructed with C5 symmetry, it was not possible to determine the precise interactions between Portal and MCP. However, Portal is not close enough to T-loop1 to interact with it, as it does in procapsids. To measure the position of Portal in expanded empty capsids, we fit both the mature and procapsid Portals (without MCP) into the expanded empty map (Fig. S25B). The geometric center of Portal is at an intermediate position in expanded empty capsids, ~9 Å more inward than in mature virions, and ~14 Å more outward than in procapsids. The geometric centers of the mature and procapsid Portals in the expanded empty map are only minimally different from each other (~1 Å). Thus, in empty expanded capsids, we find that the Portal takes an intermediate position, with more of the clip protruding from the capsid than in procapsids, but less than in mature capsids (Fig. S25). This result suggests that the packaged DNA is a major driver of the Portal repositioning during phage maturation.

DISCUSSION

We have visualized the Neck of thermophage P74-26 at high resolution using cryoEM. Though many of the components of this phage were previously unknown, we were able to identify them from our EM maps using the computational model building tool

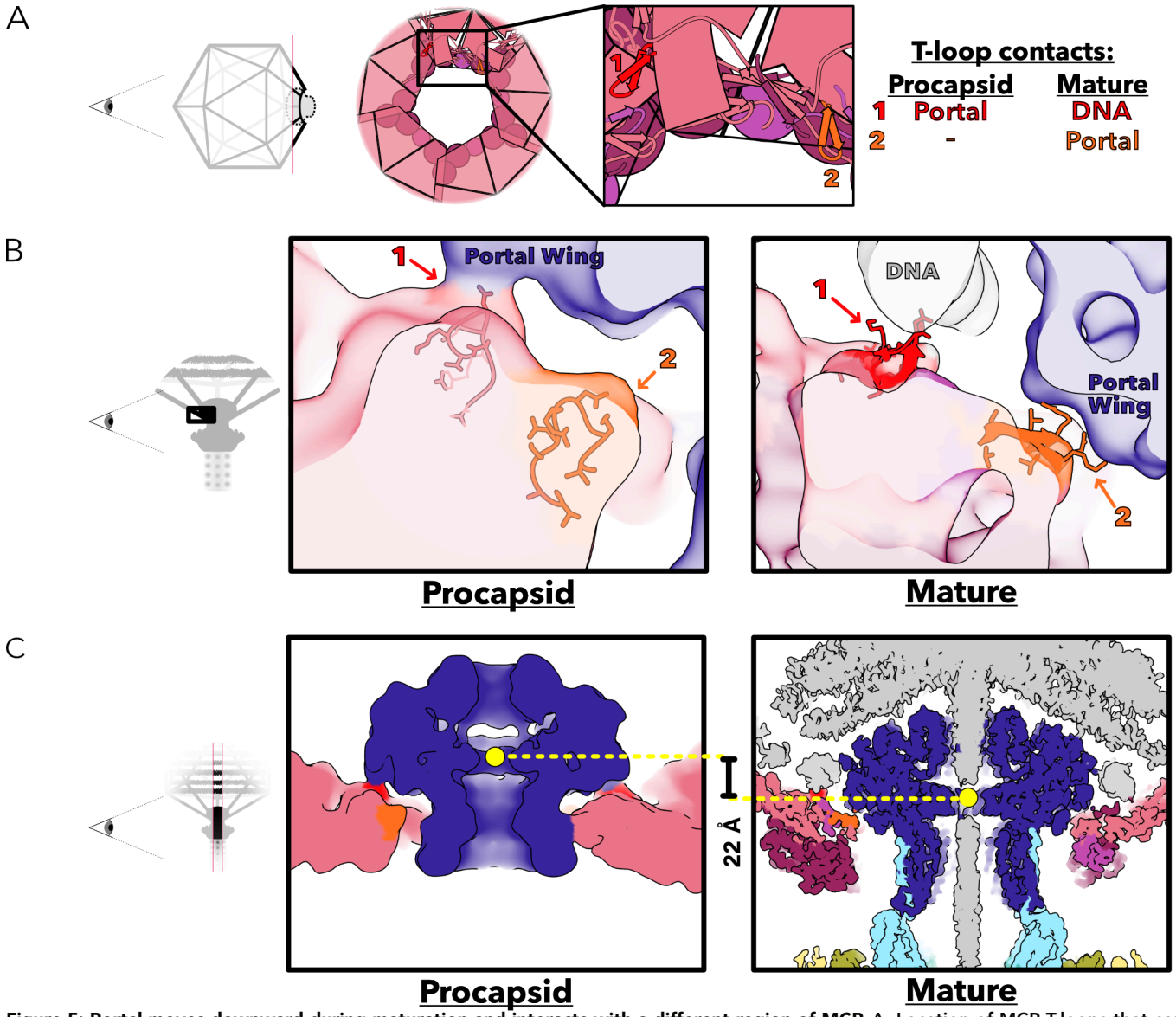

Figure 5: Portal moves downward during maturation and interacts with a different region of MCP. A. Location of MCP T-loops that contact the Portal in procapsids and mature virions. MCP, except for its T-loops, is shown in salmon. T-loop1, which contacts Portal in the procapsid, is shown in red. T-loop2, which contacts Portal in mature virions, is shown in orange. **B. Regions of procapsids (left) and mature virions (right) that contact the Portal.** T-loops 1 and 2 are shown with stick sidechains and colored as in A. The EM density for the procapsid and its Portal (EMDB: 4445, left) and of the C1 mature Portal and capsid (Gaussian filtered, right) are shown as transparent surfaces colored by protein identity. **C. Cross sections of Portal in procapsids (left) and mature virions (right) with their geometric centers in yellow.** The EM density for the procapsid and its portal (EMDB: 4445, left) and of the C1 mature Portal and capsid (right) are colored by protein identity. The procapsid and mature virion models were aligned on T-loop1. The geometric center of Portal is ~22 Å lower in mature virions than in procapsids.

ModelAngelo. This achievement was made possible by the enormous strides in cryoEM methodology and demonstrates the utility of ModelAngelo for future studies of complexes with unknown components. Analysis of these proteins revealed homology to previously studied phage proteins with some surprising connections, such as the similarity between the Portal Vertex Protein found in the capsid to an RNA Polymerase inhibitor from phage T7. These relationships underscore the modular construction of phages and the high degree of horizontal gene transfer, as well as the extent to which the diversity of phage proteins is still undersampled. Our analysis provides significant insights into how the DNA genome and the Portal are retained within the capsid at high temperatures and how bacteriophage maturation is coordinated.

Stabilization of the Portal Vertex

The Portal vertex, found in the vast majority of phages as well as in pathogens of the Herpesvirus order, presents a conundrum for the viruses. The Portal and the aperture in the capsid in which it sits are required to translocate the viral genome across the capsid shell, which occurs during both genome packaging and infection of the host cell. However, creating an opening in the capsid introduces a weak point that could compromise capsid stability, which is crucial for ensuring that the virus particle can persist long enough to infect a new host. The viral genome is packed tightly into the capsid, creating enormous pressure that the capsid must contain in order for the virus to survive. This pressure is exacerbated at high temperatures, which can also destabilize or even denature viral proteins. We previously described how the capsid proteins of a thermophilic phage have evolved to withstand high temperature by increasing topological connections between subunits[29]. We therefore wondered how this phage would strengthen its Portal Vertex to withstand both high temperature and the pressure of its own packaged genome.

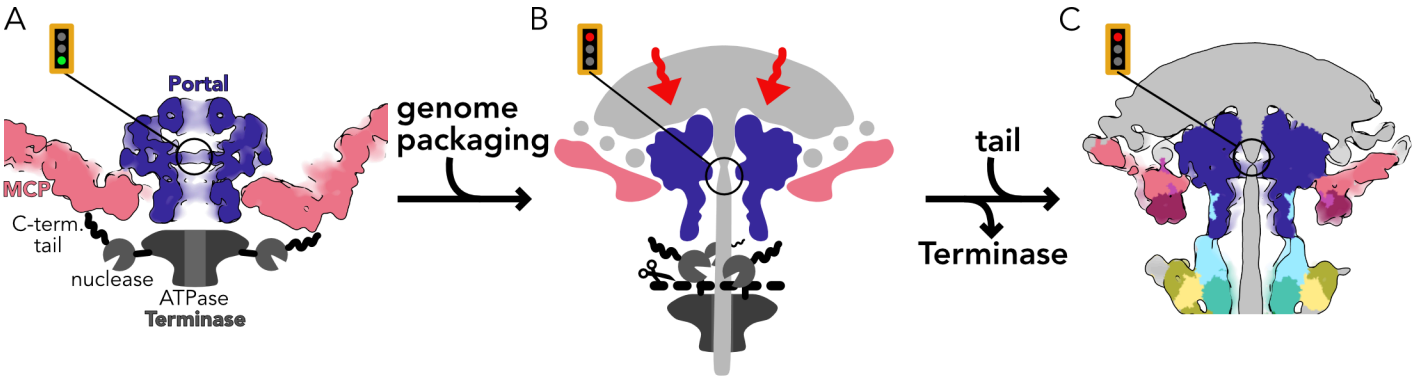

Fig. 6: Speculative model of the phage Neck maturation pathway. A. In order for genome packaging to occur, the Terminase motor must attach to the Portal in the vertex of a procapsid. EM density (EMDB: 4445) of the Portal and procapsid is cross-sectioned and colored according to protein identity as in Figure 1. The Terminase is schematized in grey and black. Scaffold protein and DNA are not shown. In the procapsid, the Portal pore is open to allow DNA packaging. The clip region of Portal, to which Adaptor tails will eventually bind, is occluded by the procapsid. The Terminase is pentameric, which we propose allows it to interact with the 5-fold vertex of the procapsid through its C-terminal tails. This tethers the nuclease domains away from the central channel through which DNA will be translocated. The Terminase can then package DNA into the capsid. During packaging, the capsid also expands. **B. We propose that pressure of the packaged DNA changes the position of Portal, switching it from genome packaging to allow Tail attachment.** When the capsid is full, the enormous pressure of packaged DNA (light grey) presses the Portal downwards like a piston. We propose that this disengages the C-terminal tails of Terminase from the capsid, freeing the nuclease domain so that it can access DNA. Once DNA has been cleaved, the Terminase and unpackaged DNA dissociate from the packaged capsid. After downward motion of the Portal and dissociation of the Terminase, the clip region is now available to bind Adaptor tails. **C. The tail is attached to the fully packaged capsid.** EM density (C1 reconstruction, gaussian filtered) of the mature Portal vertex and Neck is cross-sectioned and colored according to protein identity as in Figure 1.

Naively, we expected that interaction between Portal and MCP would be strong to prevent the Portal from popping out of the capsid. Interestingly, the interactions between Portal and MCP do not appear particularly numerous or high strength. This may be necessary to preserve the flexibility of the Portal-capsid interface, which must resolve a mismatch of C12 to C5 symmetries. Rotation of the Portal by only 6° creates an equivalent angular register of the Portal with respect to the capsid[15]. Thus, there are multiple energetically equivalent registers of Portal and capsid, which may facilitate their assembly and stability. The flexibility of this interaction enables the shift in position we observed upon maturation, with the Portal moving ~22 Å towards the outside of the capsid shell. Such flexibility would also allow rotation of the Portal relative to capsid with relatively little energetic penalty. Rotation of the Portal does not appear to play a role in DNA packaging[44], but could potentially occur during genome ejection.

Instead, we observed that Portal is held in the capsid by topological constraints between hexons that prevent dilation of the Portal vertex aperture. If the aperture were to widen even slightly, it would allow the Portal wing to pass through and thereby release the phage genome. To prevent this, P74-26 uses a highly intertwined network of interactions that lock the aperture width. The N-arm of MCP transitions from an intra-subunit interaction in a normal vertex to an inter-subunit interaction with a neighboring capsomer, moving the C_a of the N-terminal residue over 100 Å. The N-arm of one subunit of Dec wraps through the aperture to interact with the interior of MCP. PVP is added to the exterior of Portal vertex, and its N-arm also wraps under the aperture to interact with the interior of MCP. The formation of an 8-stranded sheet including strands from two MCP subunits, two Dec N-arms, and PVP highlights the interconnectedness of these interactions. Thus, these subunits are woven together as in a wicker basket, strengthening the aperture through the number and topology of interactions rather than strengthening individual interactions. This is similar to the topological linkages between MCP subunits described in Stone *et al.* and we hypothesize that this mechanism of topological intertwinement could play a significant role in generating thermostability for other large assemblies that must withstand high temperatures. We further note that this strategy of interweaving structural elements is distinct from the mechanisms for generating thermostability in individual protein domains, which tends toward increased hydrophobic and electrostatic interactions or decreased conformational entropy[45–49]. We propose that incorporating a topological component into future protein design efforts might allow for engineering of large protein assemblies with enhanced thermostability.

Gating of DNA by the pore of Portal

The Portal is the gatekeeper of DNA retention in phage P74-26. We observe that viral particles with broken tails still retain DNA in the capsid despite the loss of the tail tip complex, indicating that the Neck is the primary determinant of genome retention. Moreover, our structure shows that the Portal adopts a closed pore conformation that would restrict the movement of DNA through the pore. This observation confirms a previous hypothesis that the closed pore conformation would be found in mature, DNA-filled virions[15]. No other constrictions are observed along the length of the Neck region. Thus, we demonstrate that P74-26 retains its DNA primarily using the Portal and without the aid of other components that some phages use to retain their genome, such as the Stopper protein[50] or the Tail Tip Complex[41]. Our work indicates that the signal for genome release must somehow transmit across the nearly micron-long tail of P74-26 to the Portal, a remarkable feat of long-range communication. How this signal is transmitted will be the subject of future studies.

DNA wrapping is coupled with Portal repositioning

We observe pseudo-five-fold symmetric DNA wrapped around Portal. The interactions in P74-26 do not appear to be sequence-specific, with primarily minor groove and backbone interactions. The sole major groove interaction occurs with insertion of an Asparagine and Lysine, which have limited potential for sequence-specific interactions. P74-26 uses a headful packaging mechanism, so individual P74-26 virions have circularly permuted genome termini such that the sequence of this stretch of DNA is different in each virion[28]. By binding to the T-loop1, the DNA displaces the Portal wing, which interacts with this T-loop1 in the procapsid. In the mature capsid, Portal now interacts with T-loop2 and moves outward in the aperture by ~22Å. This outward motion is likely caused by DNA pressing down on the Portal crown and displacing the Portal wing from T-loop1. In support of this hypothesis, the Portal moves outward by ~ 9 Å in mature capsids that are empty compared to mature virions that have packaged DNA, suggesting that ~40% of Portal repositioning is due to direct pressure from packaged DNA. We propose that the rest of Portal repositioning is due to the expansion of the capsid, an event that is also triggered by DNA packaging[51]. Our structural comparison indicates that movement of the Portal wing from T-loop1 to T-loop2 is coupled to capsid expansion, since the T-loops only assume these positions when MCP undergoes conformational changes to transition from the procapsid to expanded capsid.

Thus, our analysis suggests that the pressure from packaged DNA plays a major role in altering the positioning of Portal, both through direct pressure of DNA onto Portal, as well as through the expansion of capsid. We propose that the Portal repositioning due to capsid expansion is irreversible, while the Portal repositioning from direct DNA pressure is reversible. Repositioning of the Portal complex with DNA packaging has also been observed in Herpes Simplex Virus-2[52], Varicella-Zoster Virus[53], and Human Cytomegalovirus[54], but the mechanism for this movement is not clear.

Speculative mechanism of the headful switch

Using the procapsid structure as the starting point and the mature virion structure as the end, we hypothesize a pathway for the headful switch. The procapsid consists of a compacted icosohedron of MCP, without Dec or PVP, and with Portal at a single vertex. Scaffold protein is also present inside the procapsid, as well as any other co-assembling proteins. Over the course of maturation, several major changes take place, some of which may occur concomitantly. These include the expansion of the capsid, dissociation of the packaging motor, removal of Scaffold protein, addition of Dec and PVP to the capsid, conformational and positional changes in the Portal, and attachment of the virion tail.

Crucially, in the procapsid state the clip domain of Portal is bound by MCP, which blocks interactions with the Adaptor C-terminal tail and prevents further assembly of the Neck. To fill the procapsid with the viral genome, the Terminase packaging motor binds to the Portal Vertex and translocates DNA through Portal. The Terminase is a pentamer with each subunit containing an ATPase domain, a Nuclease domain, and a flexible C-terminal tail (Fig. 6A). The ATPase domains oligomerize and translocate DNA through both Terminase and Portal. The Nuclease domains are required to cleave DNA when packaging is complete.

Though no structures of this complex exist, we speculate that the five C-terminal tails each bind to the five-fold symmetric hexons arrayed around the Portal Vertex. This would have two benefits: securing the pentameric Terminase to the dodecameric Portal by tethering it to the C5 capsid, and anchoring the Nuclease domains away from DNA to prevent premature cleavage during translocation. We previously hypothesized that the headful signal facilitates dissociation of the Terminase from Portal, releasing the nuclease domain from a restricted conformation and allowing the nuclease domains to reorient and cut DNA[26].

Here, we propose that the pressure-sensitive outward motion of Portal is the headful signal that induces the switch from ATP-dependent DNA translocation to DNA cleavage. Capsid expansion and the pressure of packaged DNA pushes the Portal outward and shifts its interaction from T-loop1 to T-loop2 (Fig. 6B). This large motion pushes the clip domain of Portal out of the aperture so that it now protrudes past the capsid. We propose that this motion dislodges the Terminase tails from the hexons, freeing the nuclease domains. The nuclease domains then reposition to cleave DNA, dissociating the Terminase and unpackaged DNA from the filled capsid, such that they can bind to another procapsid for genome packaging. Likewise, the Portal clip is now free for Adaptor tails to bind, attaching the filled head to the tail (Fig 6C). Thus, we propose that the pressure-sensitive motions reported in this study underscore the headful packaging signal in a broad class of phages.

Additional mechanisms have been proposed for the headful signal, which are not mutually exclusive with that proposed here. Studies of phage T4 have led to the hypothesis that motions between the TerL ATPase and nuclease domains can switch between these two activities[55]. Furthermore, numerous studies indicate that internal conformational changes within Portal impact the headful signal[40,56]. It is likely that multiple redundant mechanisms are used regulate the headful switch in order to prevent premature cleavage of the genome during packaging. Future studies will test these mechanistic models.

MATERIALS AND METHODS

Virus purification

P74-26 phage was propagated in the host strain *T. thermophilus* HB8 (ATCC 27634) using fresh overnight cultures grown at 65 °C in *Thermus* growth medium (4 g/L yeast extract, 8 g/L tryptone, 3 g/L NaCl, 1 mM $MgCl_2$, 0.5 mM $CaCl_2$). Four milliliters of P74-26 phage stock at 1×10^6 plaque-forming units per mL (PFU/mL) was combined with 6 mL of fresh *T. thermophilus* HB8 and incubated at 65 °C for 10 min. The mixture was then inoculated into 1 L of *Thermus* Medium and incubated at 65 °C while shaking for 4–5 h. The culture containing lysed cells and phage was then spun at 4000 ×g for 20 min, and supernatant was incubated with DNase I (2 U/mL) and RNase A (1 µg/mL) for 1 h at 30 °C. To precipitate virions, solid NaCl was added to 1 M final concentration and Polyethylene Glycol MW 8000 was gradually added to a final concentration of 10% (w/v). The mixture was incubated on ice overnight. The next day, precipitated phage stock was spun in 50 ml aliquots at 11,000 ×g for 20 min at 4 °C. The resulting phage pellet was resuspended in 2 mL of 50 mM Tris pH 7.5, 100 mM NaCl, and 1 mM $MgSO_4$. The resuspension was supplemented with 0.4 g solid CsCl and added to a CsCl step gradient (2 mL steps each of 1.2, 1.3, 1.4, 1.5 g/mL CsCl and 1mL cushion of 1.7 g/mL CsCl, in 50mM Tris pH 7.5, 100mM NaCl, 1mM $MgSO_4$). Gradients were prepared in 12 mL ultracentrifuge tubes and spun in a Beckman SW40-Ti rotor at 38,000 RPM for 18 h at 4°C. The virion band was isolated by fractionation and dialyzed twice overnight into 2L of 50 mM Tris pH 8.0, 10 mM NaCl, 10 mM $MgCl_2$ at 4°C. P74-26 virions were then concentrated to 1×10^{12} PFU/mL for subsequent use in electron microscopy.

Electron microscopy

Grids were glow discharged on a PELCO easiGlow (Ted Pella) at 25 mA for 60 s (negative polarity). 3.5 µL of purified virions at 1×10^{10} PFU/mL was applied to a 400 mesh Lacey Carbon grid (Electron Microscopy Sciences) at 10 ˚C with 90% humidity in a Vitrobot Mark IV (FEI). Sample was blotted from both sides for 10 s after a wait time of 15 s, then immediately vitrified by plunging into liquid ethane. Micrographs were collected on a 300 kV Titan Krios electron microscope (FEI) equipped with a K3 Summit direct electron detector (Gatan) and a Gatan Bioquantum Imaging Filter set to a slit width of 20 eV. Images were collected at a magnification of 81000x in super-resolution mode with an unbinned pixel size of 0.53 Å per pixel and a total dose of 49.0571 $e^-/Å^2$ per micrograph, with a target defocus range set from -0.4 to -1.5 µm. In total, 16,184 micrographs, each 30 frames, were collected.

Broken phage samples were applied to either a 200 mesh Lacey Carbon grid (Heavy Band 1) or a 200 mesh UltrAuFoil grid (Heavy band 2) and imaged on a 200kV Glacios electron microscope (FEI) equipped with a Falcon4 (Thermo Fisher) direct electron detector and a Selectris energy filter operated at 10 eV. The Falcon4 camera was operated in electron event representation (EER) mode and micrographs were collected at 79,000x magnification, a pixel size of 1.53 Å, a target defocus range set from -1.0 to -2.0 µm, and a total dose of 49.2668 $e^-/Å^2$. Movies were collected using SerialEM.

Negative Stain EM: Carbon-coated 200 mesh copper grids (Electron Microscopy Sciences) were glow discharged on a PELCO easiGlow (Ted Pella) at 20 mA for 60 s (negative polarity). 4 µL of sample was applied to the grid and incubated for 1 min. Excess sample was blotted, then the grid was washed with water followed by staining with 1% uranyl acetate (pH 4.5). Grids were viewed with a Philips CM120 electron microscope at 120 kV on a Gatan Orius SC1000 camera.

Data processing and reconstruction

All processing was performed in CryoSPARC4[57]. Micrographs were preprocessed using Patch Motion Correction and Patch CTF Estimation. We first reconstructed the entire capsid with icosahedral symmetry. An initial stack of 208,224 particles was extracted using Blob Picker with a diameter of 77-87 nm (the P74-26 capsid has diameter 82 nm) and cropped in Fourier Space (F-cropped) from a box size of 1260 pixels (px) to 256 px to reduce computational load. After 2 rounds of 2D classification, we selected 31,014 particles and used their 2D classes as a template. Template Picker extracted a stack of 677,203 particles, which yielded 38,732 capsid particles after 2 rounds of 2D classification. This stack was reconstructed and refined with icosahedral

symmetry. We then used symmetry expansion around the icosahedral space group and 3 rounds of 3D classification with a spherical focus mask over one vertex of the icosohedron to select particles with density protruding from the capsid at this vertex. After removing duplicates, 34,452 particles were reconstructed to show the capsid and Neck. Particles were then re-centered and re-extracted with a box size of 360 px to isolate the Neck and exclude most of the capsid density. These particles were then used to train a convolutional neural network via the Topaz wrapper available in CryoSPARC to pick Necks[58]. We then used the model to pick 116,158 particles from the entire dataset, which were F-cropped from a box size of 360 px to 180 px to reduce computational load. 2D classification was performed on the particle stacks to remove false positives and low-quality particles, yielding a stack of 41,024 Neck particles.

To visualize the Neck, we reconstructed and refined this stack with C12 symmetry using CryoSPARC Non-Uniform Refinement and re-extracted the particles with a box size of 360 px (uncropped)[59]. 3D classification into 3 classes to remove junk left 39,678 particles. Subsequent 3D classification into 2 classes with a focus mask over the collar region separated Neck particles with and without the collar region. To resolve the transition between C12 and C3 symmetry, we performed symmetry expansion of the particle stack around the C12 symmetry group followed by 3D classification into 10 classes with a focus mask applied to the collar region. Four resulting classes contained the P74-26 Neck, while the remaining six contained junk. After duplicate removal, the largest class had 30,748 particles and was refined with C3 symmetry using CryoSPARC Non-Uniform Refinement.

To visualize the Portal vertex, we reconstructed and refined the particle stack with C5 symmetry using CryoSPARC Non-Uniform Refinement and re-extracted the particles with a box size of 360 px (uncropped). 3D classification with 3 classes to remove junk left 40,773 particles. To resolve the transition between C12 and C5 symmetry, we performed symmetry expansion of the particle stack around the C12 symmetry group followed by 3D classification into 10 classes with a focus mask applied to the Neck. Four resulting classes contained the P74-26 Neck and Portal vertex, while the remaining six contained junk. After duplicate removal, the largest class had 39,382 particles and was asymmetrically refined using CryoSPARC Non-Uniform Refinement.

We also noticed particles in our micrographs that resembled the C12 Neck without any capsid. We manually picked 303 of these particles from micrographs collected at a range of defocus values. These particles were used to train a Topaz model which was then used to extract 163,113 particles, of which 12,754 contained the capsid-less Neck after 2 rounds of 2D classification. This particle stack was used to train a second, independent Topaz model which then was used to extract 280,700 particles, of which 26,019 contained the capsid-less Neck after 3 rounds of 2D classification. The above refinement and symmetry-breaking pipeline was applied to this stack of particles to yield C12 and C3 reconstructions of the capsid-less Neck.

Identification of Neck proteins

To identify the unknown proteins comprising the Neck, we used the automated cryo-EM atomic model-building software ModelAngelo[32]. We ran ModelAngelo without sequence inputs on C12, C3, and C5 maps of the Neck, and extracted the sequences of all built peptides >10 residues in length in FASTA format. These peptide sequences were used to search the non-redundant RefSeq protein database of the NCBI BLAST server. Hits are reported in Supplementary Table 1.

Mass spectrometry

Mature P74-26 virion samples were precipitated out of solution using Trichloroacetic acid (TCA) precipitation. 1 volume of TCA was added to 4 volumes of protein sample and incubated for 10 min at 4 °C. Samples were spun down at 13,000 RPM in a microcentrifuge for 5 min and the resulting pellet was washed 3 times with 200 µL of ice-cold acetone. The pellet was air dried until it was white and cloudy. Pellets were lyophilized and shipped on dry ice for mass spectrometry analysis at the Taplin Biological Mass Spectrometry Facility.

50 µL of 50 mM ammonium bicarbonate with 10% acetonitrile was added to the dry tubes containing the TCA precipitated protein and gently vortexed. 10 µL of 20 ng/µL modified sequencing-grade trypsin (Promega, Madison, WI) was spiked into the solutions and the samples were incubated at 37 °C overnight. Samples were acidified by adding 5 µL 20% formic acid solution and then desalted by STAGE tip[60]. Before analysis, the samples were reconstituted in 10 µL of HPLC solvent A. A nano-scale reverse-phase HPLC capillary column was created by packing 2.6 µm C18 spherical silica beads into a fused silica capillary (100 µm inner diameter x ~30 cm length) with a flame-drawn tip[61]. After equilibrating the column each sample was loaded via a Famos auto sampler (LC Packings, San Francisco CA) onto the column. A gradient was formed, and peptides were eluted with increasing concentrations of solvent B (97.5% acetonitrile, 0.1% formic acid). As peptides eluted, they were subjected to electrospray ionization and then entered into an LTQ Orbitrap Velos Elite ion-trap mass spectrometer (Thermo Fisher Scientific, Waltham, MA). Peptides were detected, isolated, and fragmented to produce a tandem mass spectrum of specific fragment ions for each peptide. Peptide sequences were determined by matching protein databases with the acquired fragmentation pattern by the software program, Sequest (Thermo Fisher Scientific, Waltham, MA)[62]. All databases include a reversed version of all the sequences and the data was filtered to between a one and two percent peptide false discovery rate. The number of sequenced peptides mapping to each protein is reported in Supplementary Table 2.

Model building and refinement

We ran ModelAngelo using the consensus sequence of each identified protein on the C12, C3, and/or C5 maps of the Neck. These models were then Real-Space Refined in Phenix and used as the starting point for manual refinement in Coot and ISOLDE[63–65]. ISOLDE was accessed as a plug-in to ChimeraX[66,67]. We selected a single subunit of each protein for manual refinement based on the number of residues built and the geometry statistics. Each subunit was manually refined in Coot in the highest-resolution map available. For example, Portal was built in the C12 map rather than the C3 map. Though most of Adaptor has C12 symmetry, regions that contact Terminator have C6 symmetry. After Adaptor was built in the C12 map, it was duplicated, rotated around the symmetry axis, and rigid body fit into the density for the neighboring subunit so that the differing loops could be rebuilt into their density. The same workflow was used to rebuild the loops of Terminator where they contact TTP with C3 symmetry. ISOLDE simulations were also used to fix phi/psi angles and to resolve clashes of selected loops[65]. Once the asymmetric unit of each component was fully built, it was Real-Space Refined in Phenix for a second time. The asymmetric units were then duplicated, rotated around the symmetry axis, and fit into the density for neighboring units to generate the complete complex. The asymmetric map was used to align the C12, C3, and C5 maps and models relative to each other. Loops of MCP and Portal that contact each other were rebuilt into the asymmetric map in ISOLDE. The entire complex was Real-Space Refined in Phenix for a final time in the asymmetric map, with reference model restraints applied to the entire model except for the loops that were rebuilt in the asymmetric map. Final refinement statistics are reported in Supplementary Table 3.

Modeling the procapsid Portal and Portal Vertex

5 hexamers, without the pentamer they encircle, were extracted from the published model of the icosahedral P23-45 procapsid (PDB 6IBC). This model was fit into the asymmetric procapsid reconstruction (EMD-4445) around the density for Portal, along with the C12 model of the procapsid Portal (PDB 6QJT), to create a composite model of the procapsid Portal and Portal Vertex. Asymmetric loops of MCP and Portal were adjusted in ISOLDE to resolve clashes and fit EMD-4445.

Data Availability

EM volumes and molecular models have been deposited to the Electron Microscopy Data Bank (EMDB) and Protein Data Bank

(PDB) respectively and will be released after peer review. In the interim, models are provided as supplemental files and volumes are available upon request.

ACKNOWLEDGMENTS

We thank Dr. J. Pajak and members of the Kelch, Grigorieff, and Schiffer labs for helpful discussions. This work was funded by National Science Foundation awards (MCB-1817338 and MCB-2342927) and National Institutes of Health (R35GM156361) awards to B. Kelch. We thank Dr. C. Ouch for assistance with data collection and gratefully acknowledge a grant from the Massachusetts Life Sciences Center for funding the Titan Krios and Glacios at the UMass Chan Medical School Cryo-EM core facility, which were used to collect the data cited in this study. Molecular graphics and analyses performed with UCSF ChimeraX, developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from National Institutes of Health R01-GM129325 and the Office of Cyber Infrastructure and Computational Biology, National Institute of Allergy and Infectious Diseases.

AUTHOR CONTRIBUTIONS

E.L.S., E.A., and B.A.K. conceptualized research and analyzed data. E.L.S., E.A, J.E.H., R.A, X.C., and K.S. performed experiments. E.L.S. and B.A.K. wrote the manuscript.

REFERENCES

1. Salmond, G. P. C. & Fineran, P. C. A century of the phage: past, present and future. *Nat Rev Microbiol* **13**, 777–786 (2015).
2. Dion, M. B., Oechslin, F. & Moineau, S. Phage diversity, genomics and phylogeny. *Nat Rev Microbiol* **18**, 125–138 (2020).
3. Penadés, J. R., Chen, J., Quiles-Puchalt, N., Carpena, N. & Novick, R. P. Bacteriophage-mediated spread of bacterial virulence genes. *Current Opinion in Microbiology* **23**, 171–178 (2015).
4. Letarov, A. V. History of Early Bacteriophage Research and Emergence of Key Concepts in Virology. *Biochemistry Moscow* **85**, 1093–1112 (2020).
5. Sharma, S. *et al.* Bacteriophages and its applications: an overview. *Folia Microbiol* **62**, 17–55 (2017).
6. Prevelige, P. E. & Cortines, J. R. Phage assembly and the special role of the portal protein. *Curr Opin Virol* **31**, 66–73 (2018).
7. Motwani, T. & Teschke, C. M. Architect of Virus Assembly: the Portal Protein Nucleates Procapsid Assembly in Bacteriophage P22. *J Virol* **93**, e00187-19 (2019).
8. Rao, V. B. & Feiss, M. Mechanisms of DNA Packaging by Large Double-Stranded DNA Viruses. *Annu. Rev. Virol.* **2**, 351–378 (2015).
9. Simpson, A. A. *et al.* Structure of the bacteriophage phi29 DNA packaging motor. *Nature* **408**, 745–750 (2000).
10. Sun, S., Kondabagil, K., Gentz, P. M., Rossmann, M. G. & Rao, V. B. The Structure of the ATPase that Powers DNA Packaging into Bacteriophage T4 Procapsids. *Molecular Cell* **25**, 943–949 (2007).
11. Hawkins, D. E. D. P. *et al.* Insights into a viral motor: the structure of the HK97 packaging termination assembly. *Nucleic Acids Res* **51**, 7025–7035 (2023).
12. Rao, V. B. & Feiss, M. The bacteriophage DNA packaging motor. *Annu Rev Genet* **42**, 647–681 (2008).
13. Hendrix, R. W. & Johnson, J. E. Bacteriophage HK97 Capsid Assembly and Maturation. *Viral Molecular Machines* **726**, 351–363 (2012).
14. Casjens, S. *et al.* Bacteriophage P22 portal protein is part of the gauge that regulates packing density of intravirion DNA. *Journal of Molecular Biology* **224**, 1055–1074 (1992).
15. Bayfield, O. W., Steven, A. C. & Antson, A. A. Cryo-EM structure in situ reveals a molecular switch that safeguards virus against genome loss. *eLife* **9**, e55517 (2020).
16. Molineux, I. J. & Panja, D. Popping the cork: mechanisms of phage genome ejection. *Nat Rev Microbiol* **11**, 194–204 (2013).
17. Evilevitch, A., Castelnovo, M., Knobler, C. M. & Gelbart, W. M. Measuring the Force Ejecting DNA from Phage. *J. Phys. Chem. B* **108**, 6838–6843 (2004).
18. Taylor, N. M. I. *et al.* Structure of the T4 baseplate and its function in triggering sheath contraction. *Nature* **533**, 346–352 (2016).
19. Linares, R. *et al.* Structural basis of bacteriophage T5 infection trigger and E. coli cell wall perforation. *Sci Adv* **9**, eade9674 (2023).
20. Baker, M. L., Jiang, W., Rixon, F. J. & Chiu, W. Common Ancestry of Herpesviruses and Tailed DNA Bacteriophages. *J Virol* **79**, 14967–14970 (2005).
21. Rixon, F. J. & Schmid, M. F. Structural similarities in DNA packaging and delivery apparatuses in Herpesvirus and dsDNA bacteriophages. *Curr Opin Virol* **5**, 105–110 (2014).
22. Yu, M. X., Slater, M. R. & Ackermann, H.-W. Isolation and characterization of Thermus bacteriophages. *Arch Virol* **151**, 663–679 (2006).
23. Minakhin, L. *et al.* Genome Comparison and Proteomic Characterization of Thermus thermophilus Bacteriophages P23-45 and P74-26: Siphoviruses with Triplex-forming Sequences and the Longest Known Tails. *Journal of Molecular Biology* **378**, 468–480 (2008).
24. Hilbert, B. J. *et al.* Structure and mechanism of the ATPase that powers viral genome packaging. *Proc Natl Acad Sci USA* **112**, E3792–E3799 (2015).
25. Xu, R.-G., Jenkins, H. T., Antson, A. A. & Greive, S. J. Structure of the large terminase from a hyperthermophilic virus reveals a unique mechanism for oligomerization and ATP hydrolysis. *Nucleic Acids Res* **45**, 13029–13042 (2017).
26. Hilbert, B. J., Hayes, J. A., Stone, N. P., Xu, R.-G. & Kelch, B. A. The large terminase DNA packaging motor grips DNA with its ATPase domain for cleavage by the flexible nuclease domain. *Nucleic Acids Res* gkw1356 (2017) doi:10.1093/nar/gkw1356.
27. Stone, N. P. *et al.* A Hyperthermophilic Phage Decoration Protein Suggests Common Evolutionary Origin with Herpesvirus Triplex Proteins and an Anti-CRISPR Protein. *Structure* **26**, 936-947.e3 (2018).
28. Bayfield, O. W. *et al.* Cryo-EM structure and in vitro DNA packaging of a thermophilic virus with supersized T=7 capsids. *Proc Natl Acad Sci USA* **116**, 3556–3561 (2019).
29. Stone, N. P., Demo, G., Agnello, E. & Kelch, B. A. Principles for enhancing virus capsid capacity and stability from a thermophilic virus capsid structure. *Nat Commun* **10**, 4471 (2019).
30. Hayes, J. A., Hilbert, B. J., Gaubitz, C., Stone, N. P. & Kelch, B. A. A thermophilic phage uses a small terminase protein with a fixed helix-turn-helix geometry. *J Biol Chem* **295**, 3783–3793 (2020).
31. Agnello, E., Pajak, J., Liu, X. & Kelch, B. A. Conformational dynamics control assembly of an extremely long bacteriophage tail tube. *Journal of Biological Chemistry* **299**, 103021 (2023).
32. Jamali, K. *et al.* Automated model building and protein identification in cryo-EM maps. *Nature* **628**, 450–457 (2024).
33. Bae, B. *et al.* Phage T7 Gp2 inhibition of Escherichia coli RNA polymerase involves misappropriation of σ70 domain 1.1. *Proceedings of the National Academy of Sciences* **110**, 19772–19777 (2013).
34. Bayfield, O. W. *et al.* Structural atlas of a human gut crassvirus. *Nature* **617**, 409–416 (2023).
35. Iwasaki, T. *et al.* Three-dimensional structures of bacteriophage neck subunits are shared in *Podoviridae* , *Siphoviridae* and *Myoviridae*. *Genes to Cells* **23**, 528–536 (2018).
36. Pell, L. G. *et al.* The X-Ray Crystal Structure of the Phage λ Tail Terminator Protein Reveals the Biologically Relevant Hexameric Ring Structure and Demonstrates a Conserved

Mechanism of Tail Termination among Diverse Long-Tailed Phages. *Journal of Molecular Biology* **389**, 938–951 (2009).
37. Xiao, H. *et al.* Structure of the siphophage neck–Tail complex suggests that conserved tail tip proteins facilitate receptor binding and tail assembly. *PLOS Biology* **21**, e3002441 (2023).
38. Fraser, J. S., Yu, Z., Maxwell, K. L. & Davidson, A. R. Ig-like domains on bacteriophages: a tale of promiscuity and deceit. *J Mol Biol* **359**, 496–507 (2006).
39. Lhuillier, S. *et al.* Structure of bacteriophage SPP1 head-to-tail connection reveals mechanism for viral DNA gating. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 8507–8512 (2009).
40. Lokareddy, R. K. *et al.* Portal protein functions akin to a DNA-sensor that couples genome-packaging to icosahedral capsid maturation. *Nat Commun* **8**, 14310 (2017).
41. Cuervo, A. *et al.* Structures of T7 bacteriophage portal and tail suggest a viral DNA retention and ejection mechanism. *Nat Commun* **10**, 3746 (2019).
42. Orlov, I. *et al.* CryoEM structure and assembly mechanism of a bacterial virus genome gatekeeper. *Nat Commun* **13**, 7283 (2022).
43. Gu, Z., Wu, K. & Wang, J. Structural morphing in the viral portal vertex of bacteriophage lambda. *J Virol* **98**, e00068-24 (2024).
44. Baumann, R. G., Mullaney, J. & Black, L. W. Portal fusion protein constraints on function in DNA packaging of bacteriophage T4. *Molecular Microbiology* **61**, 16–32 (2006).
45. Imanaka, T., Shibazaki, M. & Takagi, M. A new way of enhancing the thermostability of proteases. *Nature* **324**, 695–697 (1986).
46. Matthews, B. W., Nicholson, H. & Becktel, W. J. Enhanced protein thermostability from site-directed mutations that decrease the entropy of unfolding. *Proc Natl Acad Sci U S A* **84**, 6663–6667 (1987).
47. Nicholson, H., Tronrud, D. E., Becktel, W. J. & Matthews, B. W. Analysis of the effectiveness of proline substitutions and glycine replacements in increasing the stability of phage T4 lysozyme. *Biopolymers* **32**, 1431–1441 (1992).
48. Yip, K. S. *et al.* Insights into the molecular basis of thermal stability from the analysis of ion-pair networks in the glutamate dehydrogenase family. *Eur J Biochem* **255**, 336–346 (1998).
49. Kannan, N. & Vishveshwara, S. Aromatic clusters: a determinant of thermal stability of thermophilic proteins. *Protein Eng* **13**, 753–761 (2000).
50. Chaban, Y. *et al.* Structural rearrangements in the phage head-to-tail interface during assembly and infection. *Proc Natl Acad Sci U S A* **112**, 7009–7014 (2015).
51. Jardine, P. J. & Coombs, D. H. Capsid expansion follows the initiation of DNA packaging in bacteriophage T4. *Journal of Molecular Biology* **284**, 661–672 (1998).
52. Wang, N. *et al.* Structures of the portal vertex reveal essential protein-protein interactions for Herpesvirus assembly and maturation. *Protein & Cell* **11**, 366–373 (2020).
53. Cao, L. *et al.* Insights into varicella-zoster virus assembly from the B- and C-capsid at near-atomic resolution structures. *hLife* **2**, 64–74 (2024).
54. Li, Z. *et al.* Cryo-electron microscopy structures of capsids and in situ portals of DNA-devoid capsids of human cytomegalovirus. *Nat Commun* **14**, 2025 (2023).
55. Ghosh-Kumar, M., Alam, T. I., Draper, B., Stack, J. D. & Rao, V. B. Regulation by interdomain communication of a headful packaging nuclease from bacteriophage T4. *Nucleic Acids Research* **39**, 2742–2755 (2011).
56. Lander, G. C. *et al.* The structure of an infectious P22 virion shows the signal for headful DNA packaging. *Science* **312**, 1791–1795 (2006).
57. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods* **14**, 290–296 (2017).
58. Bepler, T. *et al.* Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nat Methods* **16**, 1153–1160 (2019).
59. Punjani, A., Zhang, H. & Fleet, D. J. Non-uniform refinement: adaptive regularization improves single-particle cryo-EM reconstruction. *Nat Methods* **17**, 1214–1221 (2020).
60. Rappsilber, J., Ishihama, Y. & Mann, M. Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal Chem* **75**, 663–670 (2003).
61. Peng, J. & Gygi, S. P. Proteomics: the move to mixtures. *Journal of Mass Spectrometry* **36**, 1083–1091 (2001).
62. Eng, J. K., McCormack, A. L. & Yates, J. R. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectrom* **5**, 976–989 (1994).
63. Liebschner, D. *et al.* Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Cryst D* **75**, 861–877 (2019).
64. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of *Coot*. *Acta Crystallogr D Biol Crystallogr* **66**, 486–501 (2010).
65. Croll, T. I. *ISOLDE* : a physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallogr D Struct Biol* **74**, 519–530 (2018).
66. Goddard, T. D. *et al.* UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci* **27**, 14–25 (2018).
67. Meng, E. C. *et al.* UCSF ChimeraX: Tools for structure building and analysis. *Protein Science* **32**, e4792 (2023).