Principles for enhancing virus capsid capacity and stability from a thermophilic virus capsid structure

Nicholas P. Stone[a], Gabriel Demo[b], Emily Agnello[c], Brian A. Kelch[a,1]

[a] Department of Biochemistry and Molecular Pharmacology,
[b] RNA Therapeutics Institute,
[c] Graduate School in Biomedical Sciences,
University of Massachusetts Medical School, Worcester, MA 01655, USA
[1] Correspondence: brian.kelch@umassmed.edu, 508-856-8322

Keywords: capsid, thermophile, stability, virus, phage

SUMMARY

The capsids of double-stranded DNA viruses protect the viral genome from the harsh extracellular environment, while maintaining stability against the high internal pressure of packaged DNA. To elucidate how capsids maintain stability in an extreme environment, we used cryoelectron microscopy to determine the capsid structure of the thermostable phage P74-26. We find the P74-26 capsid exhibits an overall architecture that is very similar to those of other tailed bacteriophages, allowing us to directly compare structures to derive the structural basis for enhanced stability. Our structure reveals 'lasso'-like interactions that appear to function like catch bonds. This architecture allows the capsid to expand during genome packaging, yet maintain structural stability. The P74-26 capsid has T=7 geometry despite being twice as large as mesophilic homologs. Capsid capacity is increased through a novel mechanism with a larger, flatter major capsid protein. Our results suggest that decreased icosahedral complexity (i.e. lower T number) leads to a more stable capsid assembly.

INTRODUCTION

Capsids are protein shells that surround and protect the viral genome. Capsid proteins often self-assemble into icosahedral structures with a quasi-equivalent arrangement of individual subunits. In quasi-equivalence theory (CASPAR and KLUG, 1962), the capsid subunits use similar interactions throughout the assembly, but are arranged in slightly different ways in non-symmetry related environments.

Icosahedral capsid proteins form substructures called capsomers, with each capsomer comprised of either five or six subunits (pentons and hexons, respectively). Typical icosahedral capsids can be described by the triangulation number (T), which describes the complexity of the icosahedral symmetry. Capsids are composed of 60T protein subunits. This results in an assembly of twelve pentons and a variable number of hexons (10 x (T-1)). Although capsid structure has been intensely studied for over a half century, the principles underlying capsid size and stability remain elusive.

Capsid size is typically controlled by the underlying capsid geometry: the triangulation number and whether the capsid is prolate or isometric. Most frequently, the triangulation number is altered to control capsid size. Higher triangulation number results in more hexons and, therefore, a larger capsid; lowering the triangulation number will result in a smaller capsid. Another common mechanism for increasing capsid capacity is to convert a spherical, isometric capsid into an elongated icosahedron, referred to as a prolate capsid. A prolate capsid consists of a cylinder with two quasi-icosahedral caps at the ends, resulting in increased interior capacity. Small alterations to capsid protein primary sequence can change T number or can convert a capsid between isometric and prolate (Eiserling et al., 1970; Fiedler et al., 2012). Thus, the evolutionary barriers for altering capsid size through these mechanisms are quite low. It remains unknown if there are other mechanisms for capsid size control and whether these mechanisms have any ramifications for virus fitness, such as capsid stability.

Capsid stability is vital for virion survival in an external fluctuating environment until it finds a new host cell. Across the viral world, there is a wide spectrum of capsid stabilities (Mateu, 2013). Tailed phages and similar viruses utilize more stable capsids because of the high internal pressure of their tightly packaged genome (Evilevitch et al., 2003; 2004) and because the capsid never disassembles during the infection cycle (Casjens and Molineux, 2012; Davidson et al., 2012; Leiman and Shneider, 2012). Tailed bacteriophages, as well as herpesviruses, use Major Capsid Proteins of the conserved Johnson fold, also known as the HK97 fold (Suhanovsky and Teschke, 2015). (Although it is more common to refer to the fold as the HK97 fold, we use the term 'Johnson fold' to avoid confusion between the HK97 fold and the HK97 phage.) Capsids of these viruses first form as spherical procapsids, and then convert to an icosahedral shape concomitant with genome packaging, release or cleavage of auxiliary proteins, and conformational rearrangement of the Johnson folds (Hendrix and Johnson, 2011; Prevelige and Fane, 2012).

Thermophilic virus capsids are amongst the strongest because they survive in an especially harsh environment. Previous studies of thermophilic viruses have focused on capsids of various shapes including icosahedral (Veesler et al., 2013), filamentous (Dimaio et al., 2015; Kasson et al., 2017; Liu et al., 2018), and lemon-shaped (Hochstein et al., 2018; Hong et al., 2015). However, for most of these viruses, close mesophilic homologs are not available, which makes it challenging to identify the structural mechanisms that underlie thermostability.

We use the thermophilic, tailed bacteriophage P74-26 to elucidate the structural mechanisms of thermostability (Hilbert et al., 2015; 2017; Stone et al., 2018). P74-26 is an especially long-

tailed siphovirus that infects *T. thermophilus* (Minakhin et al., 2008; Yu et al., 2006) (Fig. 1A). We use P74-26 to compare with similar lambdoid phages and other mesophilic Caudoviruses. Because stability mechanisms of lambdoid phages have been studied in great detail (Bauer and Evilevitch, 2015; Bauer et al., 2015; Hernando-Pérez et al., 2014; Lander et al., 2008; Qiu, 2012), this comparison presents a unique opportunity to identify mechanisms of capsid stability.

Here, we report the structure of the P74-26 capsid. A series of lassos topologically tether subunits together to stabilize the capsid through a novel mechanism. The decoration protein forms a unique cage to lock the mature capsid in place. Finally, the structure reveals T=7 geometry despite a capsid capacity about twice that of typical T=7 Caudoviruses. P74-26 uses larger and flatter capsomers to achieve this larger capsid. Our work suggests that capsid geometry plays a critical role in virus stability.

RESULTS

Thermostability of P74-26 virions

We sought to elucidate the principles underlying the thermal stability of the P74-26 virus particle. Although it is clear that P74-26 is more stable than its mesophilic counterparts, the thermostability of P74-26 particles has never been directly measured. To address this, we heated samples of P74-26 virions to 80 °C and measured phage infectivity as a proxy of virus particle integrity (Fig. 1B). This is a very stringent test for capsid stability, as disruption of many other structures within the virion (e.g. tail, neck, baseplate, etc.) can lead to loss of infectivity. P74-26 virions remain stable and infectious when incubated at 80 °C for long periods ($t_{1/2}$ = 49.5 ± 0.7 minutes). In contrast, mesophilic lambda phage rapidly loses infectivity when incubated at high temperatures in a similar buffer ($t_{1/2}$ = 48 sec at 75 °C and extrapolated $t_{1/2}$ ~ 5-7 seconds at 80 °C (Bauer and Evilevitch, 2015)). We note that the stability observed here is probably not representative of virion stability in an ideal buffer; we observe that small changes in salt concentration and pH result in large changes in infectivity (data not shown). We are currently pursuing experiments to identify optimal buffer conditions to maximize particle stability, as well as identifying the basis for P74-26's sensitivity to salt and pH. Nonetheless, these experiments demonstrate that P74-26 is, to our knowledge, the most stable Caudovirus known.

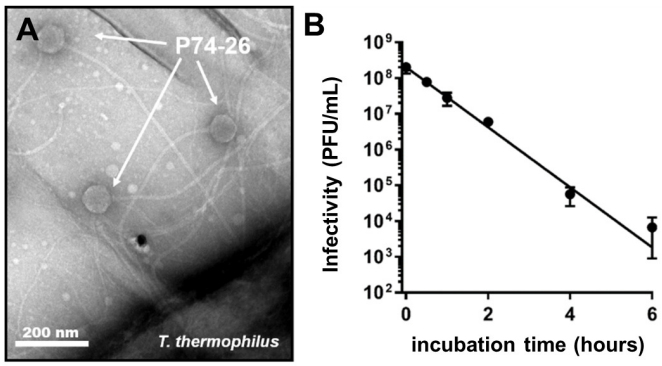

Figure 1. P74-26 is a thermostable virus.
A) Negative stain electron micrograph of phage P74-26 infecting *T. thermophilus* HB8.
B) Purified P74-26 virions retain significant infectivity after 80 °C incubation.

Overall capsid structure

To determine the structural mechanism of P74-26 stability, we used single particle cryo-EM to determine the high-resolution structure of the P74-26 capsid. In the raw images, the capsids are clearly isometric, icosahedral particles filled with DNA (Fig. 2A). The ~0.9 µm long, flexible tails emanate from a five-fold vertex. We extracted 23,178 capsid particles from the raw images and performed single particle reconstruction imposing icosahedral symmetry. The resolution of the reconstruction is 3.3Å or 3.6 Å according to gold standard 0.143 or 0.5 Fourier Shell Correlation (FSC) criteria (Fig. S1; Fig. S2A,S2B). (However, we note that the disordered DNA in the interior can adversely impact these calculations.) The P74-26 capsid is 824 Å in diameter from vertex to vertex, and 770 Å from face to face. The capsid exhibits T=7 (*laevo*) symmetry, which is surprising for a capsid of this size (see Discussion; Fig. 2B,C,D). We could easily identify density consistent with the major capsid protein (gene product 86, hereafter referred to as MCP) and the decoration protein (gene product 87, hereafter referred to as Dec^{P74-26}). We had previously determined the crystal structure of the trimeric globular regions of Dec^{P74-26}, which consists of a β-tulip domain followed by a mixed α/β subdomain (Stone et al., 2018). The Dec^{P74-26} crystal structure was easily placed into the cryo-EM maps with minor adjustments. Furthermore, we clearly resolve the entire ~23 residue N-terminal arm of Dec^{P74-26} (hereafter called the Dec-arm), most of which was missing from the crystal structure. Likewise, the entire chain of MCP is clearly represented in the reconstruction maps. Thus, we have determined the complete structures, from N- to C-termini, of the two major components of the P74-26 capsid (see Table S1 for reconstruction and model statistics).

MCP adopts the expected Johnson fold (Fig. 2E, Fig. S2C; Fig. S3). The Johnson fold has two globular domains: the rectangular P-domain (peripheral domain; residues 107 through 186, and 326 through 373) and the triangle-shaped A-domain (axial domain; residues 192 through 320, and 384 through 391). Attached to these domains are a series of loops and extended arms that facilitate protein-protein interactions to stabilize the capsid. MCP contains an especially long N-terminal region (the N-arm; residues 1 through 52) and β-hairpin called the E-loop (extended loop; residues 60 through 106). MCP also contains four non-classical elements that are either unique to P74-26 or are not found in most Johnson folds: *1)* a G-loop region (residues 166-188) of the P-domain that forms a small beta-hairpin; *2)* a P-latch loop (residues 123-133) that lies at the N-terminal end of the P-domain spine helix; *3)* an F'-loop (residues 330 through 346) that forms a flap off the bottom of the P-domain; and *4)* an extended C-terminal arm (the C-arm; residues 392 through 409) that extends away from the A-domain.

We observe conformational heterogeneity amongst the seven copies of MCP in the asymmetric unit that is restricted to the E-loop and N-arm (Fig. S4A). The MCP conformer found in the penton exhibits larger conformational changes than the hexon subunits. The E-loop displays the largest variability in conformation with a ~27 Å movement of the tip (comparing the penton subunit to the 'D' subunit of the hexon). The N-arm is the only other region that displays large heterogeneity across MCP conformers, with a ~19 Å movement of the N-arm tip. Both conformational changes can be described as rigid body rotations at hinge regions that lie at the bases of the E-loop and N-arm.

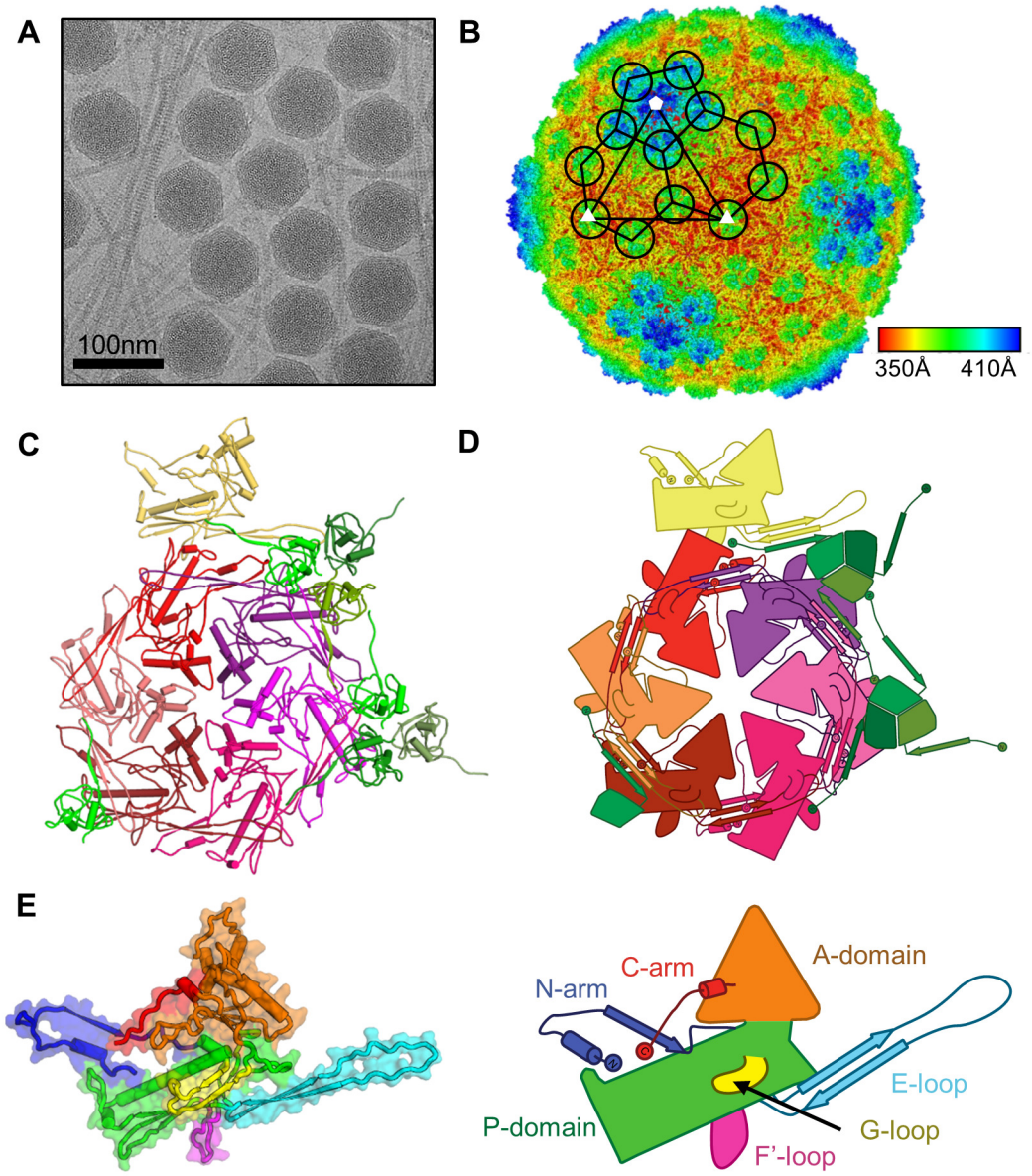

Figure 2. Determination of P74-26 capsid structure by cryo-EM.
A) Representative micrograph of purified P74-26 virions from high-resolution data collection.
B) Icosahedral reconstruction of P74-26 mature capsid, colored radially from the center of the volume (red to blue). One penton and two hexons of MCP are outlined on the volume, and decoration protein trimers are circled. The black triangle indicates the icosahedral asymmetric unit, with three- and five-fold axes labeled as small white triangles and pentagons, respectively.
C and D) Atomic model of P74-26 icosahedral asymmetric unit (C) and corresponding cartoon schematic (D). In the asymmetric unit, the Major Capsid Protein (MCP) comprises one complete hexon (shown in shades of red) and a single penton subunit (yellow). The Decoration protein subunits (gp87) are shown in shades of green.
E) Structure of P74-26 MCP colored by domain (left) with corresponding labeled schematic (right).

There are minimal conformational differences throughout the rest of the Johnson fold. This is different than observed in other Caudoviruses such as phage Sf6, where there is conformational variability throughout the Johnson fold (Zhao et al., 2017).

The P74-26 capsid is exceptionally large for a T=7 virus

To our knowledge, the P74-26 capsid is far larger than observed in all other structures of T=7 viruses. Compared to other T=7 Caudoviruses, the P74-26 capsid inner diameter is longer by ~140 Å than average, and 115 Å longer than the next biggest capsid (Fig. 3A, Table S2). This results in a capsid capacity that is about twice as large as normal for T=7 Caudoviruses. The large capsid holds the genome of P74-26, which is nearly twice as long as average for T=7 Caudoviruses. Thus, P74-26 has presumably evolved a bigger capsid to hold its large genome.

The larger size of the T=7 capsid is due to much larger and flatter capsomers than found in other Caudoviruses. The P74-26 pentons and hexons are 157 □ and 197 □ in diameter in the longest dimension, larger than normal capsomers (average of 129 □ and 158 Å, respectively). The P74-26 capsomers are also less curved than in typical Caudoviruses (Fig. 3B; Fig. S4C).

P74-26 MCP occupies much more surface area than observed in other Johnson folds. The MCP of P74-26 contains a typical number of residues (Table S3), so this is not simply due to lengthening the MCP primary sequence. Instead, P74-26 efficiently uses extended loop architecture to create a structure with a large surface area, as we describe below.

The largest changes in surface area result from the N- and C-arms and the E-loop, which are much longer than observed in any other Johnson fold. The P74-26 E-loop β-hairpin is 44 residues in length, whereas the range of E-loops in other Johnson folds is 32 to 34 residues. The greater number of residues allows the E-loop to extend much further (~70 Å vs ~53 Å for P74-26 and HK97 phage, respectively). The N-arm is 52 residues long and extends ~44 Å from the Johnson fold, opposed to an average length of 36 residues and ~37 Å extension for other Caudoviruses. The N-arm of P74-26 also loops back to directly contact the P-domain. This N-arm architecture is unique to P74-26 and covers much more surface area than other N-arms. Finally, the P74-26 C-arm forms a unique extended structure that runs along the outside of the A-domain. The C-arm binds two adjacent A-domains within a

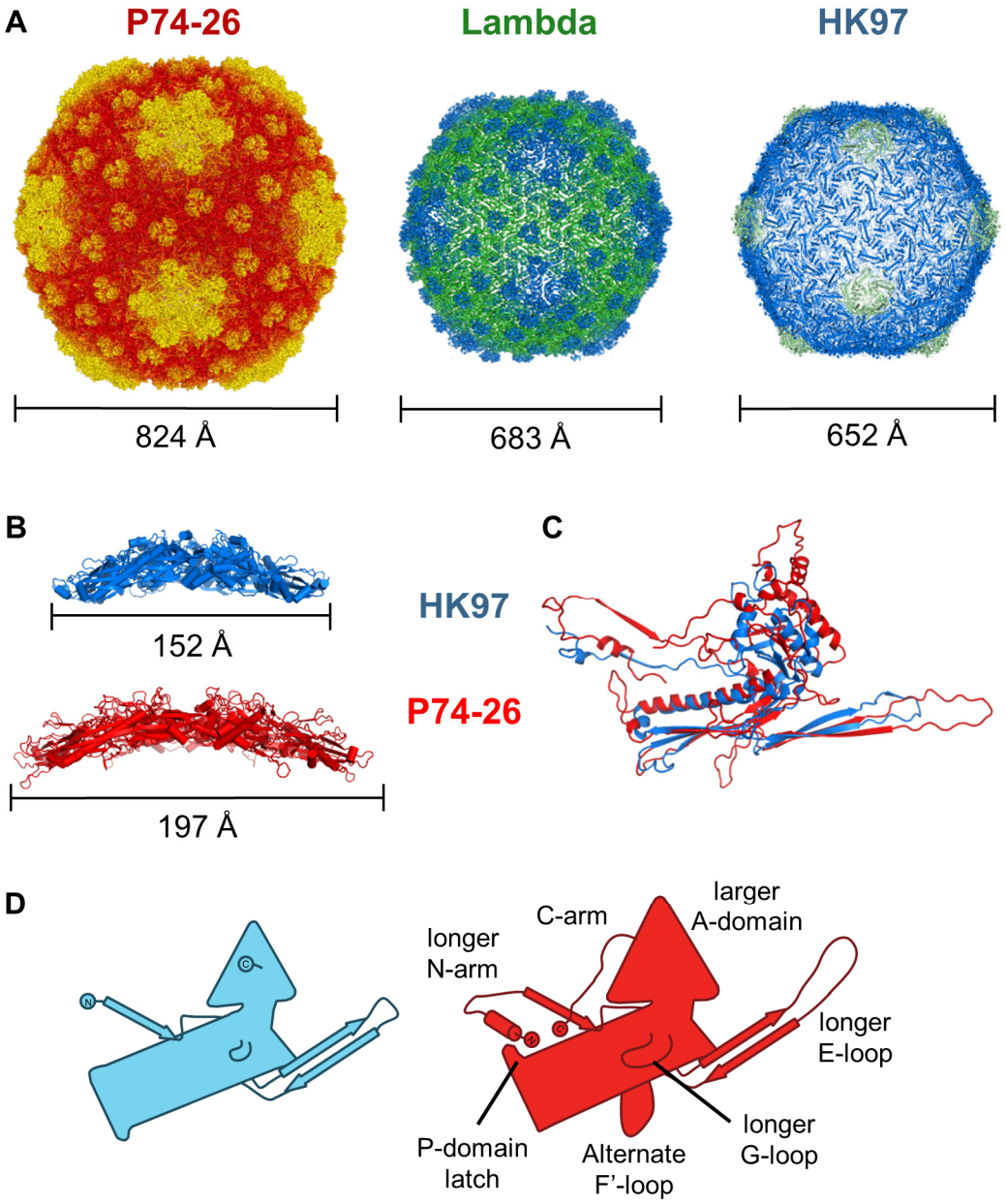

Figure 3. P74-26 has significantly increased capsid capacity compared to T=7 mesophilic viruses

A) Size comparison of P74-26 and homologous mesophilic T=7 phages Lambda and HK97. Corresponding capsid diameters (longest outer diameter measured from 5-fold axes) are listed below each structure.

B) Size comparison of hexons from HK97 (blue) and P74-26 (red); longest hexon diameters are listed below each structure.

C) Overlay of HK97 and P74-26 MCP, aligned on the Johnson fold P-domain.

D) Schematic representations of MCP from HK97 (left, blue) and P74-26 (right, red). Labels indicate features unique to P74-26 MCP.

capsomer such that the axial region of the capsomer is enlarged (Fig. 2D; Fig. S5A).

The P74-26 A- and P-domains also cover more surface area than typically seen in Johnson folds. At the tip of the A-domain, P74-26 contains a unique helix, which extends the axial dimension of MCP by ~6 Å relative to other Johnson folds. Furthermore, there is an extended loop region (residues 295 through 311) that widens the A-domain (Fig. 3C,D). The P74-26 P-domain contains three loops (the G-loop, P-latch, and F'-loop) that emanate from the globular region, which increases the P-domain surface area (Fig. 2E). Each of these three regions of the P-domain play critical roles in intra- or inter-capsomer interactions, which we delineate below.

MCP forms rings, 'lassos', and flaps to topologically link subunit-subunit interactions

The MCP-MCP interactions within a capsomer are much more extensive and intricate than seen in other Caudoviruses (Fig. 4A). The interaction area between two adjacent MCP proteins is ~3150 Å^2, which is larger than most other T=7 Johnson folds. Only phage Sf6 buries more surface area (~3277 Å^2), much of which is contributed by the large insertion domain that makes intra-capsomer interactions (Zhao et al., 2017). (We note that phage Sf6 lacks a decoration protein. Each Dec^{P74-26} subunit contributes an additional ~4100Å^2 of interaction area. See below for more details.) The hydrophobic contribution to interface stability is greater for P74-26 than other T=7 Caudoviruses. The estimated hydrophobic interaction free energy between two MCP subunits is -34.1 kcal/mol for P74-26 versus an average of -25.0 kcal/mol for other viruses (Table S3). However, the number of hydrogen bonds and salt bridges stabilizing intra-capsomer interactions are typical (34 versus 31 hydrogen bonds, and 4 versus 7 salt bridges for P74-26 and others, respectively).

The architecture of the interfaces between MCP subunits is substantially different for P74-26, with loops and extended arms that are intertwined to provide stability against internal pressure. Like other Johnson folds, the E-loop forms important subunit-subunit interactions within the capsomer. However, there are two unique networks of interactions that form topological links to establish tight E-loop binding.

The Dec-arm, N-arm, and C-arms all collaborate to form a ring that completely surrounds the β sheet region of the E-loop (Fig. 4A). The Dec-arm forms an anti-parallel β-sheet with the E-loop of a neighbor, forming the outer surface of the capsid (Fig. 4A). This

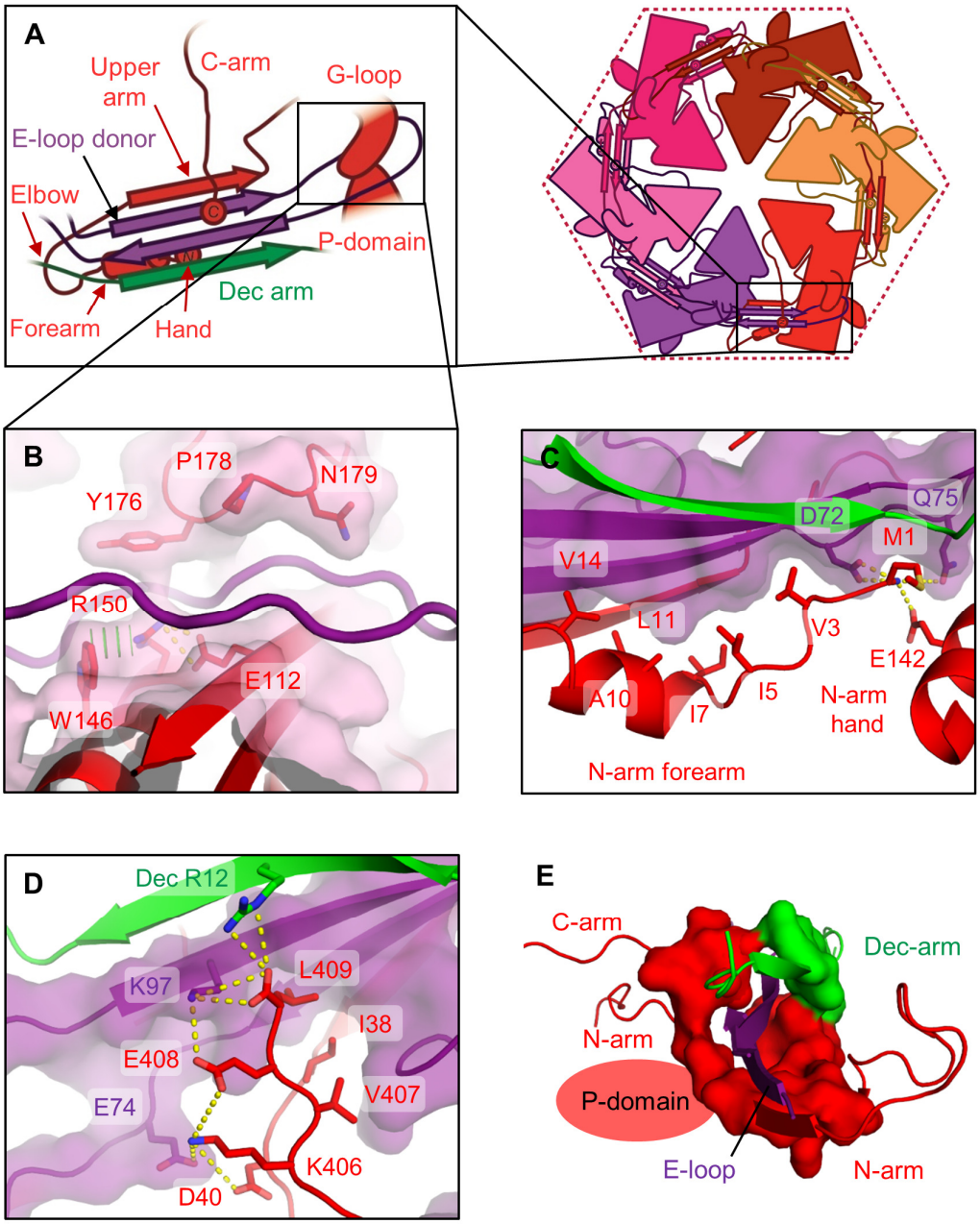

Figure 4. Lasso interactions stabilizing the E-loop.
A) Diagrams depicting interactions of the MCP E-loop (purple) with a neighboring MCP subunit (red) and Dec-arm (green).
B) The E-loop lasso (purple) is stabilized by the G-loop and P-domain of a neighboring subunit (red).
C) The N-arm 'forearm' and 'hand' bind underneath an adjacent E-loop.
D) The C-arm binds on top of all four strands of the E-loop β-sheet.
E) The β-hairpin of the E-loop (purple) is stabilized by a closed 'ring' consisting of the MCP N- and C- arms (red), and the Dec-arm (green).

architecture is similar to that observed for lambda (Lander et al., 2008). The N-arm consists of four sections: 1) a 'hand' that comprises the N-terminal four residues, 2) a helical 'forearm', 3) a turn that forms the 'elbow', and 4) the 'upper arm' that forms a β-strand conformation. The hand, forearm, and upper arm all form intimate contacts with the E-loop of an adjacent MCP. The upper arm forms a parallel β-sheet with the E-loop, forming the inner surface of the capsid shell. This conformation is similar to that observed in nearly all other Johnson folds (Suhanovsky and Teschke, 2015). However, the P74-26 N-arm is unique in that the elbow interacts with a neighboring Dec protein (see below for more details) and makes a sharp turn to orient the forearm and hand underneath the E-loop β-sheet. The forearm helix and hand interact with the E-loop using tightly packed hydrophobic interactions, with the N-terminus amine group forming salt bridges with the P-domain (Glu142) and the E-loop of the neighbor (Asp72) (Fig. 4C).

P74-26 MCP contains a C-arm that stabilizes inter-subunit interactions and, to our knowledge, is absent in other Johnson folds. In other Caudovirus MCPs, the C-terminus is tucked within the A-domain, but P74-26 has an extended, 18-residue arm that emanates from the A-domain. This C-arm runs down the length of the A domain and ends at the E-loop of the neighboring MCP. The C-arm uses hydrophobic interactions to glue the A-domains of two adjacent MCP subunits together. The C-terminal four residues bind directly on top of the E-loop of a neighbor, interacting with all four strands in the sheet (Fig. 4D). The C-terminal carboxylate forms a salt bridge with Arg12 of the Dec-arm and Lys97 of the E-loop, while the four sidechains form a series of hydrophobic and salt bridge interactions to stabilize the C-arm over the sheet. Thus, both N- and C-arms grip the neighboring E-loop in a pincer architecture, with the C-arm binding from above, while the N-arm binds from below (Fig. S5C,D). All three arms act to completely surround the E-loop of each neighbor in a ring of tight interactions, stabilizing the inter-subunit assembly (Fig. 4E). To our knowledge, this type of architecture is unique to the thermophilic phage.

The end of the E-loop forms a lasso that attaches to the neighboring MCP. The E-loop is intercalated by several residues from the neighboring subunit (Glu112, Trp146, and Arg150 of the P-domain and Tyr176, Pro178, and Asn179 of the G-loop), preventing it from forming a canonical beta-hairpin structure (Fig. 4B). The long G-loop is critical for this interaction, as it allows these residues to interdigitate between the E-loop strands. The interaction between the P-domain and G-loop of one subunit with the E-loop of its neighbor is quite tight, with numerous

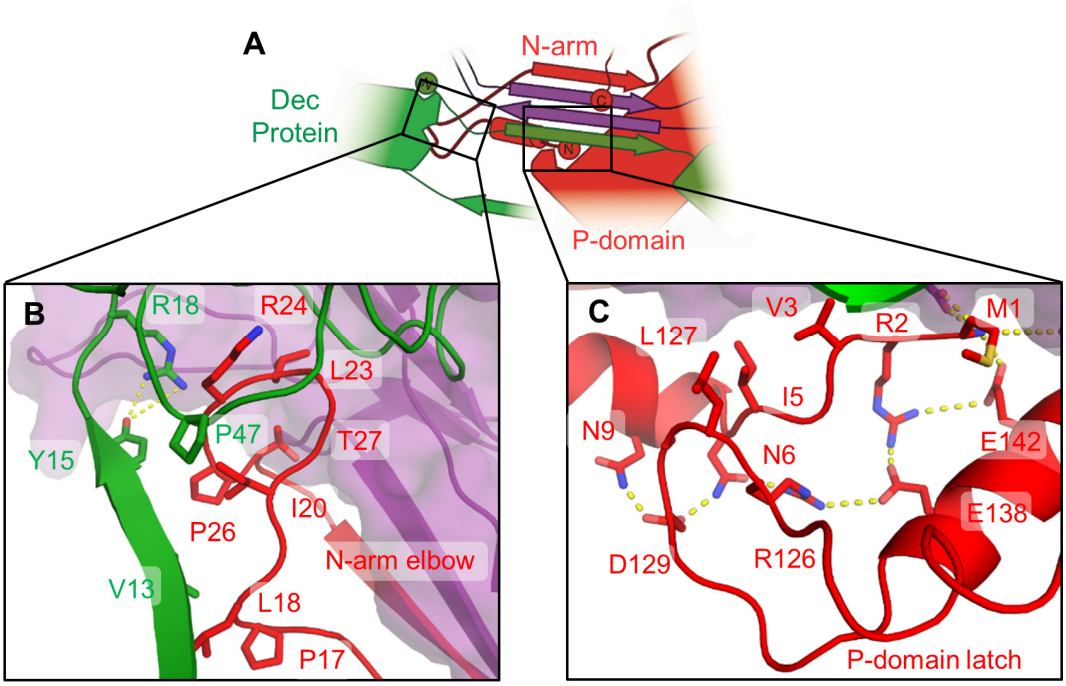

Figure 5. Lasso interactions stabilizing the N-arm.
A) Diagram depicting the N-arm lasso interactions with the P-domain (red) and adjacent Dec^{P74-26} protein (green).
B) The N-arm 'elbow' (red) extends to the globular domain of a Dec^{P74-26} protein (green) at an adjacent three-fold axis.
C) The N-arm 'hand' binds to the P-domain latch, creating a closed loop lasso.

hydrophobic interactions, hydrogen bonds, and salt bridges (Fig. S5B). In this manner, the E-loop acts as a lasso to attach to the next subunit within the capsomer.

The N-arm also forms a lasso to tightly link multiple proteins within the capsid assembly (Fig. 5A). The elbow interacts tightly with the P-domain of a neighboring MCP, and the globular domain and Dec-arm of a nearby Dec^{P74-26} (Fig. 5B). The elbow to Dec interaction is critical for stabilizing inter-capsomer interactions (described in more detail in the following section). The curve of the elbow directs the forearm and hand back toward the same Johnson fold from which the N-arm emanates (Fig. 5A). The P-domain contains the P-latch, a small loop of 11 residues that inserts into a groove between the forearm helix and the hand residues. These structural elements are locked together using a complicated web of hydrogen bonds, salt bridges, and hydrophobic interactions that serve to lock the forearm and hand in place (Fig. 5C). In this manner, the P-latch effectively closes the N-arm into a long, closed loop. Taken together, we observe two separate lasso interactions emanating from each side of MCP: one consisting of the E-loop, and the other comprised of the N-arm/P-latch combination.

The N-arm forms a 'double-lasso' structure that stabilizes the assembly at the two-fold and quasi-two-fold axes of symmetry between capsomers (Fig 6C). The N-arm elbow region of an MCP subunit in one capsomer interacts with the P-latch, forearm, and hand of the MCP subunit in the capsomer across the two-fold/quasi-two-fold axis. The interaction is primarily mediated through hydrophobic interactions. Importantly, the sidechain of Leu127 in the P-latch sits inside the opening of N-arm such that it acts as a 'hitching post' that the N-arm is lassoed around (Fig. S6B). The elbow is placed on the outside face of the capsid, while the forearm and hand form the inside surface of a neighboring capsomer. Thus, both MCP proteins provide two inter-locking lassoing interactions across the two-fold and quasi-two-fold axes.

MCP uses the F'-loop to stabilize the interaction where three capsomers meet (i.e. the three-fold and quasi-three-fold axes) (Fig. 6A). (We define these capsomers as α, β, and γ for clarity.) The F'-loop forms the inside surface of the capsid, where it binds the P-domain of an MCP in a neighboring capsomer in a 'tongue-in-groove' fashion (Fig. 6B; Fig. S6A). The interactions are reciprocated such that the F'-loop of capsomer α binds the P-domain of capsomer β, and so on. Thus, the elbows form a double-lasso on the capsid exterior to stabilize the two-fold and quasi-two-fold axes, while the F'-loop forms tongue-in-groove interactions on the capsid interior to stabilize the three-fold and quasi-three-fold axes.

The decoration protein forms an inter-capsomer cage

The Dec^{P74-26} trimer binds the capsid at the icosahedral three-fold and quasi-three-fold axes, with the Dec-arm pointing to the neighboring three-fold/quasi-three-fold axis. Dec^{P74-26} forms numerous interactions throughout the capsid. Each Dec^{P74-26} subunit contacts nine different proteins: six MCP subunits in three different capsomers, two other Dec^{P74-26} subunits within the same trimer, and one Dec^{P74-26} subunit at the neighboring three-fold/quasi-three-fold axis (Fig. 7A). The total interaction area for a single 146 residue Dec^{P74-26} subunit is ~4100 Å^2. This is vastly more extensive than seen for mesophilic decoration proteins. For example, phage TW1 has a decoration protein of similar size (149 residues), but each Dec^{TW1} subunit buries only ~60% of the equivalent surface area (~2670 Å^2) and interacts with seven other protein chains. Additionally, Dec^{P74-26} uses much greater hydrophobic interactions than mesophilic homologs (estimated free energy of hydrophobic interactions is -38 versus -9 kcal/mol for P74-26 and TW1, respectively; Table S4).

The majority of Dec^{P74-26} interactions are with capsomer α, where it contacts two separate MCP chains, termed α1 and α2. The Dec^{P74-26} globular domain primarily makes contact with the P-domain and G-loop of the α1 subunit. The Dec^{P74-26} to P-domain interaction surface is comprised of a series of hydrophobic interactions and salt bridges (Fig. S8A). The interface with the α2 subunit is the most substantial, accounting for ~1300 Å^2 of interaction area. Most of this interaction is due to the previously mentioned interaction of the Dec-arm with the E-loop (Fig. 5A). However, additional contacts are formed between the Dec-arm and the α2 G-loop (Fig. S8B).

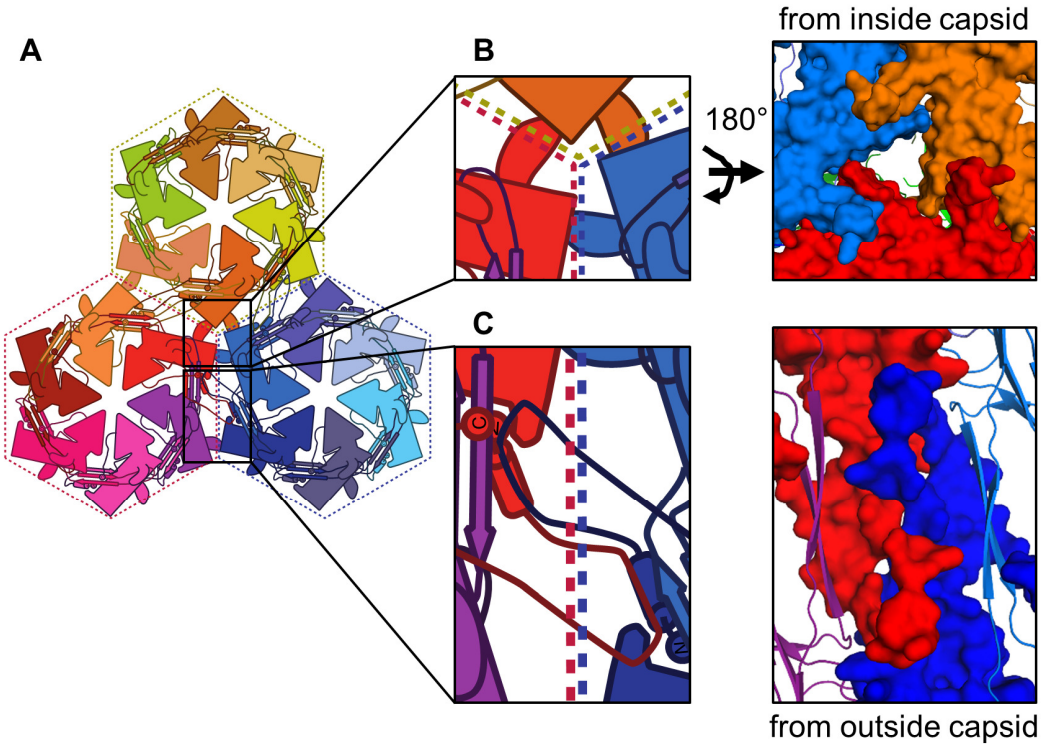

Figure 6. Inter-capsomer stabilization by extended flaps in P74-26 MCP.
A) Diagram of the three-fold axis where three capsomers (α-red, β-blue, and γ-yellow) meet.
B) The F'-loop flap interacts with the P-domain of a neighboring subunit, forming a ring around the quasi-three-fold/three-fold axes. Right, tongue-in-groove interactions of F'-loops and adjacent P-domains, as viewed from the capsid interior.
C) The N-arms of two adjacent subunits form interleaving flaps where the N-arm binds to the P-domain latch of the neighboring subunit across the quasi-two-fold/two-fold axes.

The Dec-arm reaches to a separate Dec^{P74-26} trimer at the neighboring three-fold/quasi-three-fold axis. The N-terminal seven residues of the Dec-arm use hydrogen bonds and salt bridges to bind the neighboring Dec^{P74-26} globular domain in an induced-fit mechanism (Fig. 7C; Fig. S7). This same Dec subunit reciprocates this interaction using its Dec-arm, to create two proteins embracing each other 'at arm's length'. These interactions effectively link each of the Dec^{P74-26} trimers throughout the capsid surface into one interlocked cage (Fig. 7B). To our knowledge, this interconnected architecture of decoration proteins has not been observed in mesophilic viruses.

Each Dec^{P74-26} subunit also contacts three subunits in capsomer β and one in capsomer γ, all through the Dec^{P74-26} globular domain. The β1 MCP subunit binds Dec^{P74-26} using its elbow region, burying 449 Å^2 of area. The G-loop of the β2 MCP subunit binds the side of the Dec^{P74-26} globular domain, using hydrogen bonds and hydrophobic interactions, burying 257 Å^2 (Fig. S9A). The β3 MCP subunit binds underneath the Dec^{P74-26} globular domain, using the tip of the E-loop to make hydrogen bonds (Fig. S9B). Finally, γ1 MCP makes a few hydrogen bonds to Dec^{P74-26} using its P-domain (Fig. S9C).

DISCUSSION

Architectures for enhancing stability: lassos, rings, and flaps

We observe enhanced hydrophobic interactions at the subunit-subunit interfaces within the capsid. Hydrophobic interactions are estimated to be ~2 fold higher for P74-26 than for other mesophilic homologs (Table S3,S4). This observation can partially explain the enhanced thermostability of the P74-26 capsid, as the hydrophobic effect increases in strength at high temperature (Huang and Chandler, 2000). In contrast, we observe no significant alteration in the number of hydrogen bonds or salt bridges (Table S3,S4), other interactions that have been seen to confer thermostability for some globular proteins (Petsko, 2001; Szilágyi and Závodszky, 2000; Vogt et al., 1997; Zhou and Dong, 2003).

We were not particularly surprised to find increased hydrophobic interactions in the P74-26 capsid. Numerous studies of thermophilic globular proteins show increased hydrophobic forces as a major contributor to thermal stability (Hennig et al., 1997; Knapp et al., 1999; Lim et al., 1997; Petsko, 2001; Voorhorst et al., 1997). However, there are two things that make the P74-26 capsid a unique model system: 1) high internal pressure from tightly packaged DNA induces mechanical stress in the capsid(Bauer et al., 2015; Evilevitch et al., 2003; 2004), and 2) it is a self-assembling structure in which inter-subunit architecture and topology plays an important role in overall stability. Because we can easily compare the P74-26 structure with those of numerous mesophilic homologs, we elucidate the architectural principles for capsid stabilization.

We find that the P74-26 capsid is stabilized by several loops and extensions that form lassos. The E-loop is a particularly striking example of a lasso. Towards its tip, it widens considerably, allowing residues from the neighboring MCP subunit to intercalate between the two strands. Widening of the E-loop has been seen in other Caudovirus structures (Hryc et al., 2017). However, P74-26 is unique in that the neighboring MCP G-loop and P-domain clamp from both the top and the bottom to insert into the E-loop gap (Fig. 4B; 4B). Thus, these two structural elements act as a 'hitching post' to which the E-loop lasso attaches as a topologically-linked tether. Furthermore, the N- and C-arms, together with the Dec-arm, completely ring the E-loop β-strands, another architectural element unique to P74-26 (Fig. 4E). Thus, the E-loop is ringed toward the base and lassoed towards the tip.

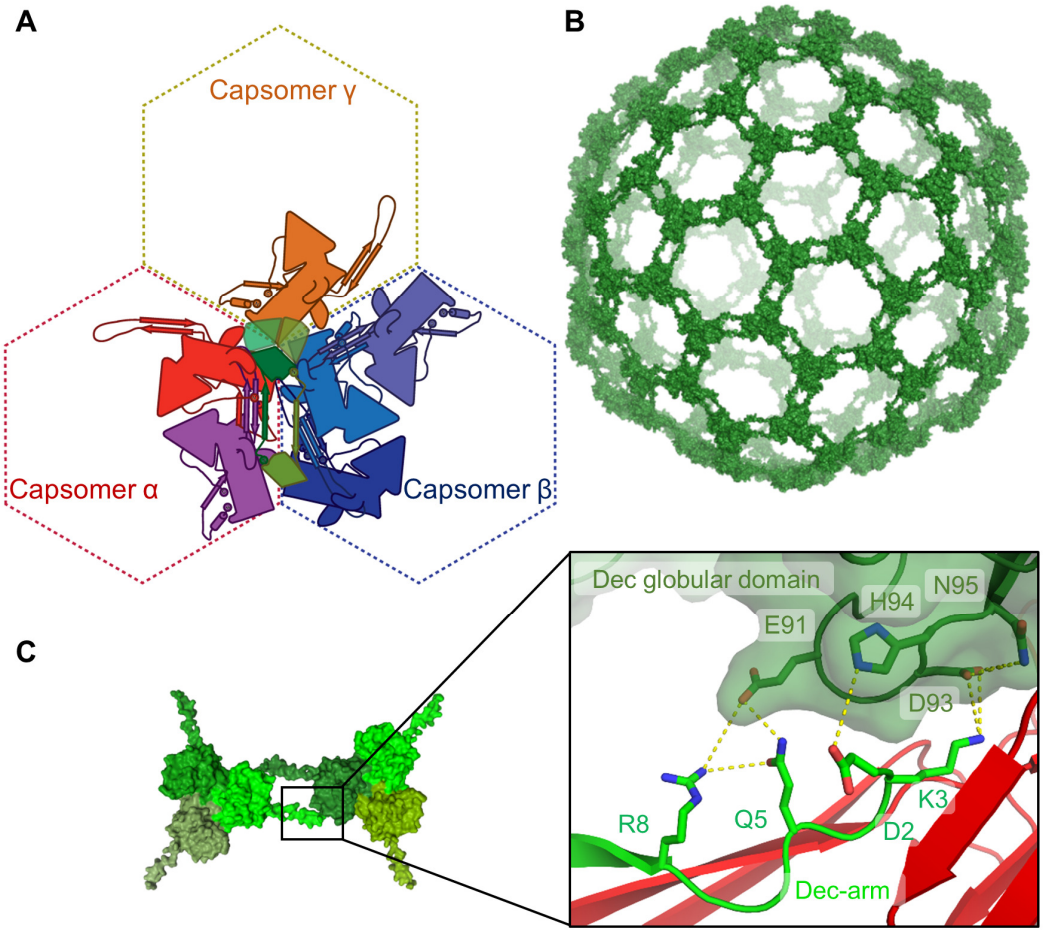

Figure 7. Stabilization of the P74-26 capsid by Dec^{P74-26}.
A) A single Dec^{P74-26} protein monomer (dark green) stabilizes capsomer interfaces by binding nine different proteins: three Dec^{P74-26} subunits (green) and six MCP subunits from three capsomers.
B) The Dec^{P74-26} proteins form an interconnected 'cage' that surrounds the capsid.
C) The Dec-arm from one subunit (light green) interacts with the globular domain of a Dec^{P74-26} protein (dark green) positioned at an adjacent quasi-three-fold/three-fold axis.

A second lasso is formed by the N-arm, which interacts with various parts of the capsid. While the N-arm is technically not a closed loop, P74-26 effectively closes the loop using the unique P-latch to fix the position of the forearm and hand regions (Fig. 5C). The N-arm lasso provides both intra- and inter-capsomer stabilization. It binds to the E-loop of a neighboring MCP within a capsomer through the upper-arm strand, the forearm helix, and hand region (Fig. 4A,C). Furthermore, the elbow, forearm, and hand stabilize capsomer-capsomer interactions by binding with Dec^{P74-26} and an MCP subunit that lies across the two-fold/quasi-two-fold axes (Fig. 5B). We find no similar N-arm lasso architecture in mesophilic Caudoviruses (Chen et al., 2017; Guo et al., 2014; Hryc et al., 2017; Wang et al., 2017; Wikoff et al., 2000; Zhang et al., 2013; Zhao et al., 2017), which suggests that this architecture is important for enhancing capsid stability.

Although lassos are not found in other Caudoviruses, the distantly related herpesviruses contain analogous lasso architecture in the N-arm of the Johnson fold (Dai and Zhou, 2018; Dai et al., 2018; Yu et al., 2017). Much like the N-arm of P74-26, the herpesvirus N-arm lassos are not true closed loops. Despite this seeming similarity, the herpesvirus lassos function differently. They exclusively stabilize inter-capsomer interactions, whereas the P74-26 N-arm lasso stabilizes both intra-capsomer and inter-capsomer interactions. Moreover, the pentons of herpesvirus capsids do not exhibit lasso interactions and the interactions are variable within hexon subunits, while P74-26 has nearly identical lasso interactions throughout both pentons and hexons. These observations indicate that the lasso architecture likely evolved independently and highlights the mechanistic flexibility of lasso-mediated stabilization.

We hypothesize that the extendable arm architecture facilitates evolution of stronger interactions within self-assembling systems such as capsids. These open-ended loops can be easily enhanced through serial single-residue extension. Perhaps this ease of evolution is why extended arm lassos are found in both P74-26 and herpesviruses. We anticipate that these types of lassos can be useful for engineering more stable capsids and other self-assembling particles.

Another advantage of the lasso architecture is that it can adopt a less extended conformation. P74-26 MCP has two lassos on either end of the Johnson fold, both of which are presumably present in the much smaller procapsid. By using these lasso structures, P74-26 can retain high stability, while providing conformational flexibility to expand during maturation. We hypothesize that the lassos are less extended in the procapsid; upon capsid expansion, the lassos reach their full extension, where they lock into place. The full extension observed in the mature capsid would provide tensional integrity, as we discuss below.

Another architectural element that stabilizes subunit-subunit interactions is the tongue-in-groove interaction of the F'-loop. The F'-loop stabilizes inter-capsomer interactions by inserting into a groove on the P-domain of an MCP subunit in a neighboring capsomer. These F'-loop interactions are found ringing the three-fold/quasi-three-fold axes along the inside face of the capsid (Fig. 6B; Fig. S6A). This architecture resembles the interleaved arrangement of flaps in the top of a moving box. Similarly, the inter-capsomer two-fold/quasi-two-fold interactions are stabilized on the outside face of the capsid by the interleaved arrangement of the N-arms (Fig. 6C; Fig. S6B). In this manner, the outside and inside faces of the capsid are stabilized by two separate interleaving flap interactions. We propose the 'moving box' arrangements seen at

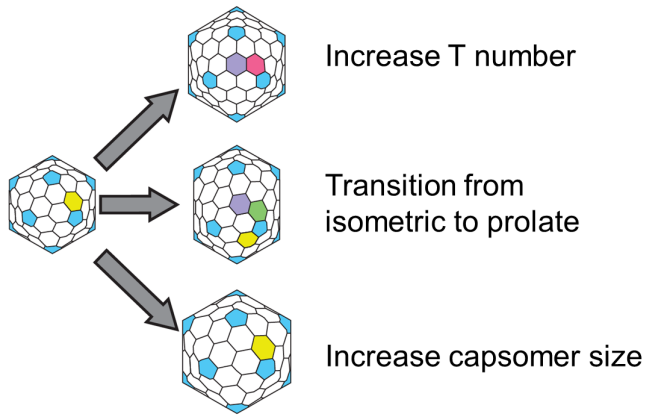

Figure 8. Mechanisms for increasing capsid capacity.
P74-26 adopts a novel mechanism for enhancing capsid capacity by increasing the size of capsomers, while retaining T=7 geometry.

symmetry and quasi-symmetry axes greatly strengthens the capsid against internal pressure because they are topologically challenging to disrupt. However, these arrangements are also presumably challenging to assemble, which raises the important question of how the P74-26 capsid assembles with an interleaved architecture. One possibility is that these elements are not interleaved in the procapsid, and only intercalate into neighboring subunits upon maturation. Future experiments will investigate this and other possibilities.

A cage of decoration proteins stabilizes the mature capsid

The three-fold/quasi-three-fold axes are further stabilized by the Dec^{P74-26} trimer. Compared to mesophilic Caudoviruses, the Dec^{P74-26} trimer interacts with more subunits across a much larger interaction area (Fig. 7A). The total interaction area per Dec^{P74-26} subunit is quite remarkable: ~4100 Å^2 for a 146-residue protein. Our previous study showed that Dec^{P74-26} is substantially more stable than its mesophilic homologs, and this stabilization is primarily through formation of an extraordinarily tight trimer (Stone et al., 2018). However, the trimerization interactions only account for a small fraction of the total Dec^{P74-26} interaction area (~18% of the total interaction area per Dec^{P74-26} subunit). This suggests that the Dec^{P74-26} interaction with the capsid contributes a substantial amount of stability.

Interactions between Dec^{P74-26} trimers forms a 'cage' holding the capsid together (Fig. 7B). Two adjacent Dec^{P74-26} trimers are linked together by reciprocating interactions mediated by Dec-arms. The N-terminus of one Dec-arm is bound in the globular domain of a Dec^{P74-26} subunit in an adjacent trimer in a two-fold or quasi-two-fold symmetric fashion (Fig. 7C). This arrangement resembles a 'middle-school slow dance', due to its similarity to two individuals embracing each other but at complete arm's length. The result of the middle-school slow dance architecture is a cage surrounding the entire capsid. This arrangement is unique to P74-26. For example, phages lambda and TW1 use a very similar Decoration Protein fold (Stone et al., 2018), but the interaction of their Dec-arm with other capsid proteins is much more limited (Lander et al., 2008; Wang et al., 2017). Furthermore, the unrelated decoration protein of phage L does not connect with neighboring trimers, and in fact is missing at the quasi-three-fold axes (Newcomer et al., 2018). T4 phage is decorated with the Soc protein that interacts with neighboring Soc subunits at the three-fold and quasi-three-fold axes; however, Soc is present in relatively low occupancy (~50%), so the cage is incomplete (Chen et al., 2017). Because decoration proteins are usually absent in the procapsid (Lander et al., 2008), we expect that the Dec^{P74-26} cage would assemble cooperatively upon capsid expansion to stabilize the capsid. Future experiments will interrogate the role of cooperativity in assembly and stability.

A tensegrity mechanism for enhancing capsid stability

How do forces from internal pressure act on the capsid, and how does the capsid architecture resist these stresses? If we assume that the pressure from encapsulated DNA is distributed uniformly around the isometric capsid, then all points on the capsid are experiencing a vector of force that is perpendicular to the surface of the capsid. By analogy, the capsid is experiencing forces that are similar to those of a balloon. Thus, the internal pressure causes lateral stresses on capsid interactions.

The architecture of the P74-26 capsid is constructed to handle lateral stress through tensional integrity. Tensional integrity, or tensegrity, is a generalized mechanism for architectural stability that involves structured regions held together by a network of flexible elements that are in continuous tension (Fuller, 1961). In the P74-26 capsid, the A- and P-domains are the structured regions, and the lassos and extended arms are the flexible elements that transmit tension. For example, the E-loop lasso will become taut against the 'hitching post' formed by the G-loop and the P-domain of the neighbor. Likewise, the N-arm forms a lasso whose end is held in place by the P-latch that locks into the groove between the forearm and hand (Fig. 5C). Thus, we predict that the P-latch will exhibit the hallmarks of a 'catch bond', a non-covalent bond that becomes stronger under tension (Dembo et al., 1988). Furthermore, the P74-26 capsid contains several flaps that interleave with each other. These interactions would topologically resist the lateral and longitudinal stresses of internal pressure. As a whole, these lasso and flap elements use tension to resist structural failure of the capsid. The tensegrity mechanism observed here is simply a more elaborate example of capsid tensegrity suggested by Caspar many years ago (CASPAR, 1980).

The lasso, flap, and arm interactions are positioned such that the internal pressure distributes the stress across multiple bonds. For example, the E-loop β-sheet experiences forces along the axis of the sheet. Thus, all bonds holding the sheet together are under stress rather than the orthogonal geometry in which the stress is just on the bonds at the end of the sheet. Capsid failure would require simultaneous disruption of many bonds (a 'shearing' geometry), rather than an unzipping geometry in which the bonds rupture one at a time (Rohs et al., 1999). Pioneering single molecule studies have shown that a shearing geometry requires much higher forces to disrupt than when forces act in an unzipping geometry (Guinn et al., 2015; Jagannathan et al., 2012; Wang et al., 2013). Thus, the P74-26 capsid exhibits cooperative interaction in which lateral forces act in a shearing geometry, resulting in high tensegrity.

Mechanisms for altering capsid capacity

The capsid of P74-26 is larger than in most Caudoviruses, which is due to its genome being abnormally large. Most T=7 Caudoviruses have genome sizes between 30 and 50 kb (Table

S2), while phage P74-26's genome is nearly twice as long at 83 kb (Minakhin et al., 2008). Based on genome size we had predicted that the capsid would be T=12 (average genome size ~80kb (Lander et al., 2012)), although T=9 or T=13 would have been possible (average genome size ~70kb or ~120kb, respectively). The capsid attains this larger size by significantly increasing the capsomer size rather than changing the icosahedral complexity. The capsomer is larger because the P74-26 MCP covers more surface area than normal, despite a typical protein length. Consequently, the capsomer is slightly thinner than normal (Fig. 2B). Thus, the number of residues in MCP does not predict total area covered, and genome size does not predict triangulation number.

To our knowledge, this is a new mechanism for increasing capsid capacity. There are two classic mechanisms for enlarging a capsid: 1) increasing the triangulation number, and 2) conversion of an isometric to prolate head. In the first case, hexons are added across all faces of the capsid, whereas in the second, hexons are added across ten of the icosahedral faces such that the capsid is elongated in one dimension (Fig. 8). In both cases, the capsomers remain the same size. Here, we have identified a third mechanism for evolution of a larger capsid: increasing the size of the capsomer.

These three mechanisms have very different evolutionary barriers. The two classical mechanisms can be implemented through simple mutations and have been observed numerous times. In many viruses, simple point mutations modify the triangulation number (Butan et al., 2010; Ferreira et al., 2003; Fiedler et al., 2012). Furthermore, the triangulation number of some capsids can be altered without changing MCP sequence (Allison et al., 2003; Crowther et al., 1994; Dearborn et al., 2017). Likewise, single point mutations in T4 phage convert the capsid from prolate to isometric or generate 'giant' heads in which the long axis of the prolate head is lengthened (Doermann et al., 1973; Eiserling et al., 1970). Therefore, the evolutionary barriers for altering capsid volume through the two classical mechanisms appear to be quite low.

In contrast, the enlarged capsomer strategy identified here requires multiple, extensive alterations to the sequences of capsid proteins. This mechanism requires large changes to the E-loop, N-arm, C-arm, and A-domain. Moreover, the Dec protein must coevolve with MCP to maintain tight binding. On the other hand, increasing the triangulation number or prolate/isometric transitions would not require significant simultaneous modification of MCP and Dec sequences. For example, T4 phage decoration proteins bind in the isometric head geometry (Chen et al., 2017). This begs the question: why did phage P74-26 utilize this seemingly more challenging evolutionary strategy rather than the 'easier', classical strategies? What constraints prevented evolution of a larger capsid through the classical routes?

Our first hypothesis is that the lassos, flaps, and arms stabilizing the P74-26 capsid require a larger capsomer for function. It is possible that the lassos need extra space to open enough for a 'hitching post' structure to insert. Likewise, the flaps and arms may require a certain length to elicit their stabilizing activity. If this were the case, then the architectural elements stabilizing the capsid require larger than normal capsomers. In this scenario, the larger capsomer is the selected structural feature and the T=7 geometry is a spandrel: a biological structure that is a byproduct of evolution rather than a result of direct selection (Gould and Lewontin, 1979). However, we do not favor this hypothesis because lassos are found in herpesviruses, in which the Johnson fold is a typical size. (Herpesvirus MCPs have several other 'tower' domains that increase size, but these domains are not part of the main capsid 'floor' and do not contribute to the Johnson fold (Yu et al., 2017).) Furthermore, other Caudovirus MCPs contain long N-arms (e.g. Sf6 phage (Zhao et al., 2017)) or E-loops that are opened nearly as wide as P74-26 (e.g. P22 phage (Hryc et al., 2017)), but these proteins are of typical size. Nonetheless, this hypothesis remains untested.

Our favored hypothesis is that capsid geometry has a direct effect on overall capsid stability. We hypothesize that the T=7 geometry is inherently more stable than higher triangulation numbers due to variable conformations of hexons. All capsids that are T=9 or higher have more than one type of hexon present, while all capsids T≤7 have exactly one type of hexon (except T=1, which has no hexons; (Mannige and Brooks, 2009; 2010)). For example, T=7 has 'skewed' hexon conformation, while T=9 has both winged and flat hexons (Fig. S10A,B). We also note that prolate capsids have multiple types of hexons (generally three or more hexon conformations; Fig. 8). Thus, major capsid proteins in T>7 viruses must accommodate the hexon conformational heterogeneity, which may adversely affect stability.

We hypothesize that T=7 geometry is the highest complexity (i.e. largest size) that is inherently stable. More complex geometries would introduce instability through variation in hexon conformation. This inherent instability may require extra stabilization mechanisms to mitigate, such as extra decoration proteins to cement the structure in place. We envision two non-mutually exclusive disadvantages of T>7 geometry. First, the larger number of MCP conformations must remain functional and stable, which would constrain evolution of MCP proteins for greater stability. The second benefit is that a lower triangulation number results in fewer subunit-subunit interfaces, thus minimizing the number of weak points in the capsid. In support of this hypothesis, all T>7 capsids use decoration proteins (to our knowledge), while many T=7 viruses lack decoration proteins.

If various triangulation numbers have different inherent stability, this would suggest that each geometry exhibits weak points at different regions of the capsid, as has been predicted from theoretical work (Wilson, 2016). We hypothesize that the three-fold/quasi-three-fold axes represent the weak points in a T=7 lattice. In support of this hypothesis, decoration proteins of T=7 Caudoviruses are commonly found at the three-fold/quasi-three-fold axes (Table S2) (Lander et al., 2008; Newcomer et al., 2018; Shen et al., 2012; Wang et al., 2017). Furthermore, these axes are stabilized by covalent cross-links in HK97 phage (Wikoff et al., 2000) and F'-loop flaps in P74-26 (Fig. 6B). To examine this idea further, we note that T=9 phage also use decoration proteins at the three-fold axes (Choi et al., 2008; Grose et al., 2014), while T=12 and T=13 phage use decoration proteins at the centers of capsomers (Effantin et al., 2006; Fokine et al., 2004; Lander et al., 2012).

We note that all of our analysis has primarily focused on Caudoviruses. These viruses do not generally break down their capsids as part of their lifecycle, so the capsid has no selective pressure to be labile. In fact, the high pressure of packaged DNA presents a high selective pressure to evolve stable capsids. We anticipate that these principles may be modified for viruses that disassemble their capsids as part of their lifecycle. Future research will delineate the differences between the two classes of viruses.

Conclusions

The P74-26 structure reveals novel mechanisms for increasing capsid size and stability. We anticipate that these principles will be useful for design of new capsids and capsid-like structures for a variety of applications. The lasso and extended arm geometry that we have identified here may be of particular utility in capsid engineering, as the new design only needs to satisfy a few interactions along the length of the arm. The P74-26 capsid itself could serve as the basis for engineering of novel materials for applications as diverse as vaccine delivery and functional nanomaterials. Finally, our work also points to an unanticipated role of capsid geometry in controlling stability, an idea that needs to be explored in more detail.

MATERIALS AND METHODS

Growth/Purification of P74-26 virions:

Phage stock preparation and virion purification were performed as previously described (Minakhin et al., 2008; Stone et al., 2018). P74-26 phage was propagated in the host strain *Thermus thermophilus* HB8 using fresh overnight cultures grown at 65º C in Thermus Medium (4 g/L yeast extract, 8g/L tryptone, 3g/L NaCl, 1 mM $MgCl_2$, 0.5mM $CaCl_2$). 4 mL of P74-26 phage stock at $1x10^6$ plaque forming units per mL (PFU/mL) was combined with 6 mL of fresh *T. thermophilus* HB8 and incubated at 65º C for 10 minutes. Reaction mixture was then inoculated into 1 L of Thermus Medium, and incubated at 65º C while shaking for 4-5 hours. Lysate was then spun at 4,000 x g for 20 minutes, and supernatant was incubated with DNase I (2 Units/mL) and RNase A (1 µg/mL) for 1 hour at 30º C. To precipitate virions, solid NaCl was added to 1M final concentration and Polyethylene Glycol MW 8,000 was added to a final concentration of 10% (w/v) while stirring. Phage stock was then incubated on ice overnight. The next day, precipitated phage stock was spun at 11,000 x g for 20 minutes at 4º C. The resulting phage pellet was resuspended in 2 mL of 50 mM Tris pH 7.5, 100 mM NaCl, and 1 mM $MgSO_4$. 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 1 mL cushion of 1.7 g/mL CsCl, in 50 mM Tris pH 7.5, 100 mM NaCl, 1 mM $MgSO_4$). Gradients were prepared in 12 mL ultracentrifuge tubes and spun in a Beckman SW40-Ti rotor at 38,000 RPM for 18 hours at 4º C. Sedimented virions were isolated and dialyzed twice overnight into 2L of 50 mM Tris pH 8.0, 10 mM NaCl, 10 mM $MgCl_2$ at 4º C. P74-26 virions were then concentrated to $1x10^{12}$ PFU/mL for subsequent use in electron microscopy.

Thermal Stability Assay:

Purified P74-26 virions were diluted in sample buffer (50 mM Tris pH 8.0, 10 mM NaCl, 10 mM $MgCl_2$) to a final concentration of ~$1x10^8$ PFU/mL and incubated at 80º C for 0, 0.5, 1, 2, 4, or 6 hours in triplicate. Incubated virions were serially diluted and 100 µL aliquots of several dilutions were each mixed with 150 µL of fresh *T. thermophilus* HB8 (OD_{600} = 1.0). The adsorption reactions were then incubated at 65º C for 10 minutes and mixed with 2.5 mL molten Thermus Medium supplemented with 0.75% agar. The molten reaction mixture was then spread evenly on top of TR Gellan Gum plates (2g/L Yeast Extract, 4g/L Tryptone, 1g/L NaCl, 15g/L Gellan gum, 1.5 mM $MgCl_2$, 1.5 mM $CaCl_2$, pH 7.5). Once the soft agar solidified, plates were inverted and incubated overnight at 65º C yielding a bacterial lawn with discernible phage plaques. Plaques were then counted from three plates for each incubation.

Electron Microscopy:

Negative-Staining EM

Infections of *Thermus thermophilus* HB8 were performed as described above (growth/purification of P74-26 virions). Aliquots of the infection culture were diluted 1:1 in thermus growth medium and applied to carbon-coated 400-mesh grids one hour after infection. Excess sample was then blotted from the grid surface, and grids were finally stained with 1% uranyl acetate. Data was then collected on a 120kV Philips CM-120 microscope fitted with a Gatan Orius SC1000 detector.

Cryo-EM Specimen Preparation

400 copper mesh lacey carbon support film grids (Electron Microscopy Sciences) were glow discharged using a Pelco easiGlow (Pelco) for 45 seconds at 20 mA (negative polarity) prior to sample application. 3.5 µL of P74-26 virions at $1x10^{12}$ PFU/mL was applied to a cleaned grid at 22º C and 90% humidity in a Vitrobot Mark IV (FEI). Samples were then blotted for 8 seconds after a wait time of 15 seconds. Sample coated grids were then vitrified by plunging into liquid ethane. Prepared grids were then stored submerged in liquid nitrogen prior to data collection.

Data Collection

Micrographs were collected on a 300kV Titan Krios electron microscope (FEI) equipped with a K2 Summit direct electron detector (Gatan). Images were collected at a magnification of 130,000 X in superresolution mode at a pixel size of 0.529 Angstroms per pixel and a total dose of 48 $e^-/Å^2$ per micrograph. Micrographs were collected with a target defocus of -0.2 to -1.2 um. For each micrograph, 32 frames were motion-corrected with 2x binning yielding a final pixel size of 1.059 Å/pixel. 4,611 total micrographs were obtained from a single 3-day collection.

Data Processing

CTFFind4 within the *cis*TEM software package (Grant et al., 2018) was used to determine defocus values for each resulting frame. An initial particle set of virion capsids was then picked using the *Find Particles* action in *cis*TEM with a specified maximum particle radius of 390 Å. This led to the selection of P74-26 capsids without selecting virion tail structures. The resulting particle stack was extracted using a box size of 1024 pixels, and then analyzed using 2D classification in *cis*TEM. Particles from 2D class averages containing complete icosahedral capsids were extracted, yielding a total of 28,880 particles. The particle stack was then used to generate an *ab initio* model for subsequent refinement using the *Ab-initio 3D* action in *cis*TEM. Finally, particles and particle parameters were exported to the Frealign software package (Frealign version 9.11; (Grigorieff, 2007)).

To speed up initial data processing, the particle stack was binned by a factor of 4 using the *resample* tool within the Frealign package and used for 5 cycles of global alignment to a high-resolution limit of 8 Å in Frealign mode 3 using the *ab initio* generated model as a template. The 2x binned particle stack was then subjected to 12 rounds of refinement and reconstruction in Frealign mode 1, increasing the high-resolution limit each 3 cycles from 12 to 6 Å. The 2x binned particle stack was then subjected to 20 rounds of 3D classification using 4 total classes and a high-resolution limit of 6 Å. Two classes containing high quality particles were combined using the *merge_classes*.exe in Frealign, yielding

a particle substack containing 23,178 particles. The unbinned particle substack (1.059 □/pixel) was then subjected to 8 cycles of refinement and reconstruction in Frealign mode 1, yielding a final resolution of 3.3 □ (0.143 FSC cutoff; 3.6 □ at 0.5 FSC cutoff). FSC curves were calculated in Frealign using odd and even particle half-sets. The map was sharpened using automatically calculated B-factors (-98 $Å^2$) using *bfactor.exe* in Frealign. The final sharpened map was used for subsequent model building and structure refinement.

Model Building

For the P74-26 MCP atomic model, a poly-alanine chain was initially built into the density map and manually refined using Coot (Emsley and Cowtan, 2004). Then, side chains were assigned, and a complete model of MCP was built and refined manually. For the P74-26 decoration protein, the previously determined atomic model of the decoration protein trimer (PDB: 6BL5) was rigid-body fit into a corresponding portion of the density map. The N-terminal 16 residues of the decoration protein were then built into the density map, and the protein trimer was refined manually in Coot. The models of the MCP monomer and decoration protein trimer were then refined into the P74-26 density map using the Phenix real-space refine procedure (Adams et al., 2010). These models were then used to generate the P74-26 icosahedral asymmetric unit. The MCP P- and A-domains were first rigid-body fit into the density corresponding to each MCP subunit of the asymmetric unit (six hexon subunits and one penton subunit). Then, the N-arm, C-arm, and E-loop of each subunit were individually rigid-body fit into the corresponding density to account for variability in the orientation of these loops throughout the asymmetric unit. The refined decoration protein model was then fit into the corresponding density of the asymmetric unit, comprising two decoration protein trimers and an additional monomer. The complete asymmetric unit was then manually refined in Coot, and final real-space refinement of the asymmetric model was performed using Phenix. The final real-space refinement statistics are listed in Table S1.

*Database Accession Numbers

Model coordinates and EM density map available upon request.

Author Contributions and Notes

NPS and BAK designed research, NPS and EA performed research, NPS, GD and BAK analyzed data; and NPS and BAK wrote the paper.
The authors declare no conflict of interest.
This article contains supporting information online.

Acknowledgments

The authors thank Drs. C. Xu, KK Song, and K. Lee for assistance with data collection, and Drs. C. Xu, A. Korostelev, C. Gaubitz, M. Herzik, and Ms. J. Hayes and A. Jecrois for advice on data processing. We thank members of the Kelch, Royer, and Schiffer labs for helpful discussions. We thank C. Gaubitz, J. Hayes, and J. Magrino for critical reading of the manuscript. This work was funded by the Pew Charitable Trusts and the National Science Foundation (grant number 1817338).

References

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

Allison, S.L., Tao, Y.J., O'Riordain, G., Mandl, C.W., Harrison, S.C., and Heinz, F.X. (2003). Two distinct size classes of immature and mature subviral particles from tick-borne encephalitis virus. J Virol *77*, 11357–11366.

Bauer, D.W., and Evilevitch, A. (2015). Influence of Internal DNA Pressure on Stability and Infectivity of Phage λ. J Mol Biol *427*, 3189–3200.

Bauer, D.W., Li, D., Huffman, J., Homa, F.L., Wilson, K., Leavitt, J.C., Casjens, S.R., Baines, J., and Evilevitch, A. (2015). Exploring the Balance between DNA Pressure and Capsid Stability in Herpesviruses and Phages. J Virol *89*, 9288–9298.

Butan, C., Lokhandwala, P.M., Purdy, J.G., Cardone, G., Craven, R.C., and Steven, A.C. (2010). Suppression of a morphogenic mutant in Rous sarcoma virus capsid protein by a second-site mutation: a cryoelectron tomography study. J Virol *84*, 6377–6386.

Casjens, S.R., and Molineux, I.J. (2012). Short noncontractile tail machines: adsorption and DNA delivery by podoviruses. Adv Exp Med Biol *726*, 143–179.

CASPAR, D.L. (1980). Movement and self-control in protein assemblies. Quasi-equivalence revisited. Biophys J *32*, 103–138.

CASPAR, D.L., and KLUG, A. (1962). Physical principles in the construction of regular viruses. Cold Spring Harb Symp Quant Biol *27*, 1–24.

Chen, Z., Sun, L., Zhang, Z., Fokine, A., Padilla-Sanchez, V., Hanein, D., Jiang, W., Rossmann, M.G., and Rao, V.B. (2017). Cryo-EM structure of the bacteriophage T4 isometric head at 3.3-Å resolution and its relevance to the assembly of icosahedral viruses. Proc Natl Acad Sci USA *114*, E8184–E8193.

Choi, K.H., McPartland, J., Kaganman, I., Bowman, V.D., Rothman-Denes, L.B., and Rossmann, M.G. (2008). Insight into DNA and protein transport in double-stranded DNA viruses: the structure of bacteriophage N4. J Mol Biol *378*, 726–736.

Crowther, R.A., Kiselev, N.A., Böttcher, B., Berriman, J.A., Borisova, G.P., Ose, V., and Pumpens, P. (1994). Three-dimensional structure of hepatitis B virus core particles determined by electron cryomicroscopy. Cell *77*, 943–950.

Dai, X., and Zhou, Z.H. (2018). Structure of the herpes simplex virus 1 capsid with associated tegument protein complexes. Science *360*.

Dai, X., Gong, D., Lim, H., Jih, J., Wu, T.-T., Sun, R., and Zhou, Z.H. (2018). Structure and mutagenesis reveal essential capsid protein interactions for KSHV replication. Nature *553*, 521–525.

Davidson, A.R., Cardarelli, L., Pell, L.G., Radford, D.R., and Maxwell, K.L. (2012). Long noncontractile tail machines of bacteriophages. Adv Exp Med Biol *726*, 115–142.

Dearborn, A.D., Wall, E.A., Kizziah, J.L., Klenow, L., Parker, L.K., Manning, K.A., Spilman, M.S., Spear, J.M., Christie, G.E., and Dokland, T. (2017). Competing scaffolding proteins determine capsid size during mobilization of Staphylococcus aureus pathogenicity islands. Elife *6*.

Dembo, M., Torney, D.C., Saxman, K., and Hammer, D. (1988). The reaction-limited kinetics of membrane-to-surface adhesion and detachment. Proc. R. Soc. Lond., B, Biol. Sci. *234*, 55–83.

Dimaio, F., Yu, X., Rensen, E., Krupovic, M., Prangishvili, D., and Egelman, E.H. (2015). Virology. A virus that infects a hyperthermophile encapsidates A-form DNA. Science *348*, 914–917.

Doermann, A.H., Eiserling, F.A., and Boehner, L. (1973). Genetic control of capsid length in bacteriophage T4. I. Isolation and preliminary description of four new mutants. J Virol *12*, 374–385.

Effantin, G., Boulanger, P., Neumann, E., Letellier, L., and Conway, J.F. (2006). Bacteriophage T5 structure reveals similarities with HK97 and T4 suggesting evolutionary relationships. J Mol Biol *361*, 993–1002.

Eiserling, F.A., Geiduschek, E.P., Epstein, R.H., and Metter, E.J. (1970). Capsid size and deoxyribonucleic acid length: the petite variant of bacteriophage T4. J Virol *6*, 865–876.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr *60*, 2126–2132.

Evilevitch, A., Lavelle, L., and Knobler, C.M. (2003). Osmotic pressure inhibition of DNA ejection from phage.

Evilevitch, A., Castelnovo, M., Knobler, C.M., and Gelbart, W.M. (2004). Measuring the Force Ejecting DNA from Phage †. The Journal of Physical Chemistry B *108*, 6838–6843.

Ferreira, D., Hernandez, R., Horton, M., and Brown, D.T. (2003). Morphological variants of Sindbis virus produced by a mutation in the capsid protein. Virology *307*, 54–66.

Fiedler, J.D., Higginson, C., Hovlid, M.L., Kislukhin, A.A., Castillejos, A., Manzenrieder, F., Campbell, M.G., Voss, N.R., Potter, C.S., Carragher, B., et al. (2012). Engineered mutations change the structure and stability of a virus-like particle. Biomacromolecules *13*, 2339–2348.

Fokine, A., Chipman, P.R., Leiman, P.G., Mesyanzhinov, V.V., Rao, V.B., and Rossmann, M.G. (2004). Molecular architecture of the prolate head of bacteriophage T4. Proc Natl Acad Sci USA *101*, 6003–6008.

Fuller, R.B. (1961). Tensegrity. Portfolio and Art News Annual.

Gould, S.J., and Lewontin, R.C. (1979). The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. Proc. R. Soc. Lond., B, Biol. Sci. *205*, 581–598.

Grant, T., Rohou, A., and Grigorieff, N. (2018). cisTEM, user-friendly software for single-particle image processing. Elife *7*.

Grigorieff, N. (2007). FREALIGN: high-resolution refinement of single particle structures. J Struct Biol *157*, 117–125.

Grose, J.H., Belnap, D.M., Jensen, J.D., Mathis, A.D., Prince, J.T., Merrill, B.D., Burnett, S.H., and Breakwell, D.P. (2014). The genomes, proteomes, and structures of three novel phages that infect the Bacillus cereus group and carry putative virulence factors. J Virol *88*, 11846–11860.

Guinn, E.J., Jagannathan, B., and Marqusee, S. (2015). Single-molecule chemo-mechanical unfolding reveals multiple transition state barriers in a small single-domain protein. Nat Commun *6*, 6861.

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

Hendrix, R.W., and Johnson, J.E. (2011). Bacteriophage HK97 Capsid Assembly and Maturation. In Viral Molecular Machines, (Boston, MA: Springer US), pp. 351–363.

Hennig, M., Sterner, R., Kirschner, K., and Jansonius, J.N. (1997). Crystal structure at 2.0 A resolution of phosphoribosyl anthranilate isomerase from the hyperthermophile Thermotoga maritima: possible determinants of protein stability. Biochemistry *36*, 6009–6016.

Hernando-Pérez, M., Lambert, S., Nakatani-Webster, E., Catalano, C.E., and de Pablo, P.J. (2014). Cementing proteins provide extra mechanical stabilization to viral cages. Nat Commun *5*, 4520.

Hilbert, B.J., Hayes, J.A., Stone, N.P., Duffy, C.M., Sankaran, B., and Kelch, B.A. (2015). Structure and mechanism

of the ATPase that powers viral genome packaging. Proc Natl Acad Sci USA *112*, E3792–E3799.

Hilbert, B.J., Hayes, J.A., Stone, N.P., Xu, R.-G., and Kelch, B.A. (2017). The large terminase DNA packaging motor grips DNA with its ATPase domain for cleavage by the flexible nuclease domain. Nucleic Acids Res *45*, 3591–3605.

Hochstein, R., Bollschweiler, D., Dharmavaram, S., Lintner, N.G., Plitzko, J.M., Bruinsma, R., Engelhardt, H., Young, M.J., Klug, W.S., and Lawrence, C.M. (2018). Structural studies of Acidianus tailed spindle virus reveal a structural paradigm used in the assembly of spindle-shaped viruses. Proc Natl Acad Sci USA *115*, 2120–2125.

Hong, C., Pietilä, M.K., Fu, C.J., Schmid, M.F., Bamford, D.H., and Chiu, W. (2015). Lemon-shaped halo archaeal virus His1 with uniform tail but variable capsid structure. Proc Natl Acad Sci USA *112*, 2449–2454.

Hryc, C.F., Chen, D.-H., Afonine, P.V., Jakana, J., Wang, Z., Haase-Pettingell, C., Jiang, W., Adams, P.D., King, J.A., Schmid, M.F., et al. (2017). Accurate model annotation of a near-atomic resolution cryo-EM map. Proc Natl Acad Sci USA *114*, 3103–3108.

Huang, D.M., and Chandler, D. (2000). Temperature and length scale dependence of hydrophobic effects and their possible implications for protein folding. Proc Natl Acad Sci USA *97*, 8324–8327.

Jagannathan, B., Elms, P.J., Bustamante, C., and Marqusee, S. (2012). Direct observation of a force-induced switch in the anisotropic mechanical unfolding pathway of a protein. Proc Natl Acad Sci USA.

Kasson, P., Dimaio, F., Yu, X., Lucas-Staat, S., Krupovic, M., Schouten, S., Prangishvili, D., and Egelman, E.H. (2017). Model for a novel membrane envelope in a filamentous hyperthermophilic virus. Elife *6*.

Knapp, S., Kardinahl, S., Hellgren, N., Tibbelin, G., Schäfer, G., and Ladenstein, R. (1999). Refined crystal structure of a superoxide dismutase from the hyperthermophilic archaeon Sulfolobus acidocaldarius at 2.2 A resolution. J Mol Biol *285*, 689–702.

Lander, G.C., Baudoux, A.-C., Azam, F., Potter, C.S., Carragher, B., and Johnson, J.E. (2012). Capsomer dynamics and stabilization in the T = 12 marine bacteriophage SIO-2 and its procapsid studied by CryoEM. Structure *20*, 498–503.

Lander, G.C., Evilevitch, A., Jeembaeva, M., Potter, C.S., Carragher, B., and Johnson, J.E. (2008). Bacteriophage lambda stabilization by auxiliary protein gpD: timing, location, and mechanism of attachment determined by cryo-EM. Structure *16*, 1399–1406.

Leiman, P.G., and Shneider, M.M. (2012). Contractile tail machines of bacteriophages. Adv Exp Med Biol *726*, 93–114.

Lim, J.H., Yu, Y.G., Han, Y.S., Cho, S., Ahn, B.Y., Kim, S.H., and Cho, Y. (1997). The crystal structure of an Fe-superoxide dismutase from the hyperthermophile Aquifex pyrophilus at 1.9 A resolution: structural basis for thermostability. J Mol Biol *270*, 259–274.

Liu, Y., Osinski, T., Wang, F., Krupovic, M., Schouten, S., Kasson, P., Prangishvili, D., and Egelman, E.H. (2018). Structural conservation in a membrane-enveloped filamentous virus infecting a hyperthermophilic acidophile. Nat Commun *9*, 3360.

Mannige, R.V., and Brooks, C.L. (2009). Geometric considerations in virus capsid size specificity, auxiliary requirements, and buckling. Proc Natl Acad Sci USA *106*, 8531–8536.

Mannige, R.V., and Brooks, C.L. (2010). Periodic table of virus capsids: implications for natural selection and design. PLoS ONE *5*, e9423.

Mateu, M.G. (2013). Assembly, stability and dynamics of virus capsids. Arch Biochem Biophys *531*, 65–79.

Minakhin, L., Goel, M., Berdygulova, Z., Ramanculov, E., Florens, L., Glazko, G., Karamychev, V.N., Slesarev, A.I., Kozyavkin, S.A., Khromov, I., et al. (2008). Genome comparison and proteomic characterization of Thermus thermophilus bacteriophages P23-45 and P74-26: siphoviruses with triplex-forming sequences and the longest known

tails. J Mol Biol *378*, 468–480.

Newcomer, R., Schrad, J., Gilcrease, E.B., Casjens, S.R., Feig, M., Teschke, C.M., Alexandrescu, A.T., and Parent, K.N. (2018). How phage L capsid decoration protein distinguishes between nearly identical binding sites on an icosahedral virion

. bioRxiv *http://dx.doi.org/10.1101/420992*.

Petsko, G.A. (2001). Structural basis of thermostability in hyperthermophilic proteins, or "there's more than one way to skin a cat". Methods in Enzymology *334*, 469–478.

Prevelige, P.E., and Fane, B.A. (2012). Building the machines: scaffolding protein functions during bacteriophage morphogenesis. Adv Exp Med Biol *726*, 325–350.

Qiu, X. (2012). Heat induced capsid disassembly and DNA release of bacteriophage λ. PLoS ONE *7*, e39793.

Rohs, R., Etchebest, C., and Lavery, R. (1999). Unraveling proteins: a molecular mechanics study. Biophys J *76*, 2760–2768.

Shen, P.S., Domek, M.J., Sanz-García, E., Makaju, A., Taylor, R.M., Hoggan, R., Culumber, M.D., Oberg, C.J., Breakwell, D.P., Prince, J.T., et al. (2012). Sequence and structural characterization of great salt lake bacteriophage CW02, a member of the T7-like supergroup. J Virol *86*, 7907–7917.

Stone, N.P., Hilbert, B.J., Hidalgo, D., Halloran, K.T., Lee, J., Sontheimer, E.J., and Kelch, B.A. (2018). A Hyperthermophilic Phage Decoration Protein Suggests Common Evolutionary Origin with Herpesvirus Triplex Proteins and an Anti-CRISPR Protein. Structure.

Suhanovsky, M.M., and Teschke, C.M. (2015). Nature′s favorite building block: Deciphering folding and capsid assembly of proteins with the HK97-fold. Virology.

Szilágyi, A., and Závodszky, P. (2000). Structural differences between mesophilic, moderately thermophilic and extremely thermophilic protein subunits: results of a comprehensive survey. Structure *8*, 493–504.

Veesler, D., Ng, T.-S., Sendamarai, A.K., Eilers, B.J., Lawrence, C.M., Lok, S.-M., Young, M.J., Johnson, J.E., and Fu, C.-Y. (2013). Atomic structure of the 75 MDa extremophile Sulfolobus turreted icosahedral virus determined by CryoEM and X-ray crystallography. Proc Natl Acad Sci USA *110*, 5504–5509.

Vogt, G., Woell, S., and Argos, P. (1997). Protein thermal stability, hydrogen bonds, and ion pairs. J Mol Biol *269*, 631–643.

Voorhorst, W.G., Warner, A., de Vos, W.M., and Siezen, R.J. (1997). Homology modelling of two subtilisin-like proteases from the hyperthermophilic archaea Pyrococcus furiosus and Thermococcus stetteri. Protein Eng *10*, 905–914.

Wang, L., Xu, X., Kumar, R., Maiti, B., Liu, C.T., Ivanov, I., Lee, T.-H., and Benkovic, S.J. (2013). Probing DNA clamps with single-molecule force spectroscopy. Nucleic Acids Res *41*, 7804–7814.

Wang, Z., Hardies, S.C., Fokine, A., Klose, T., Jiang, W., Cho, B.C., and Rossmann, M.G. (2017). Structure of the Marine Siphovirus TW1: Evolution of Capsid-Stabilizing Proteins and Tail Spikes. Structure.

Wikoff, W.R., Liljas, L., Duda, R.L., Tsuruta, H., Hendrix, R.W., and Johnson, J.E. (2000). Topologically linked protein rings in the bacteriophage HK97 capsid. Science *289*, 2129–2133.

Wilson, D.P. (2016). Protruding Features of Viral Capsids Are Clustered on Icosahedral Great Circles. PLoS ONE *11*, e0152319.

Yu, M.X., Slater, M.R., and Ackermann, H.-W. (2006). Isolation and characterization of Thermus bacteriophages. Arch. Virol. *151*, 663–679.

Yu, X., Jih, J., Jiang, J., and Zhou, Z.H. (2017). Atomic structure of the human cytomegalovirus capsid with its securing tegument layer of pp150. Science *356*.

Zhang, X., Guo, H., Jin, L., Czornyj, E., Hodes, A., Hui, W.H., Nieh, A.W., Miller, J.F., and Zhou, Z.H. (2013). A new topology of the HK97-like fold revealed in Bordetella bacteriophage by cryoEM at 3.5 A resolution. Elife *2*, e01299.

Zhao, H., Li, K., Lynn, A.Y., Aron, K.E., Yu, G., Jiang, W., and Tang, L. (2017). Structure of a headful DNA-packaging bacterial virus at 2.9 Å resolution by electron cryo-microscopy. Proc Natl Acad Sci USA *114*, 3601–3606.

Zhou, H.-X., and Dong, F. (2003). Electrostatic contributions to the stability of a thermophilic cold shock protein. Biophys J *84*, 2216–2222.