PNAS RESEARCH ARTICLE | MICROBIOLOGY

Structures of a large prolate virus capsid in unexpanded and expanded states generate insights into the icosahedral virus assembly

Qianglin Fang[a,b,1], Wei-Chun Tang[c,1], Andrei Fokine[a,1], Marthandan Mahalingam[c], Qianqian Shao[b], Michael G. Rossmann[a], and Venigalla B. Rao[c,2]

Edited by Elena Orlova, Birkbeck, University of London, London, United Kingdom; received February 27, 2022; accepted August 31, 2022 by Editorial Board Member Richard P. Novick

Many icosahedral viruses assemble proteinaceous precursors called proheads or procapsids. Proheads are metastable structures that undergo a profound structural transition known as expansion that transforms an immature unexpanded head into a mature genome-packaging head. Bacteriophage T4 is a model virus, well studied genetically and biochemically, but its structure determination has been challenging because of its large size and unusually prolate-shaped, ~1,200-Å-long and ~860-Å-wide capsid. Here, we report the cryogenic electron microscopy (cryo-EM) structures of T4 capsid in both of its major conformational states: unexpanded at a resolution of 5.1 Å and expanded at a resolution of 3.4 Å. These are among the largest structures deposited in Protein Data Bank to date and provide insights into virus assembly, head length determination, and shell expansion. First, the structures illustrate major domain movements and ~70% additional gain in inner capsid volume, an essential transformation to contain the entire viral genome. Second, intricate intracapsomer interactions involving a unique insertion domain dramatically change, allowing the capsid subunits to rotate and twist while the capsomers remain fastened at quasi-threefold axes. Third, high-affinity binding sites emerge for a capsid decoration protein that clamps adjacent capsomers, imparting extraordinary structural stability. Fourth, subtle conformational changes at capsomers' periphery modulate intercapsomer angles between capsomer planes that control capsid length. Finally, conformational changes were observed at the symmetry-mismatched portal vertex, which might be involved in triggering head expansion. These analyses illustrate how small changes in local capsid subunit interactions lead to profound shifts in viral capsid morphology, stability, and volume.

bacteriophage T4 j virus assembly j capsid expansion j capsid length control j prolate virus structure

Significance

Virus assembly occurs by repeated association of tiny building blocks, the capsid protein subunits. The assembled structures are highly dynamic and undergo maturation transformations, ultimately generating an infectious virion particle. We describe, at near-atomic resolution, a major "expansion" transformation in a large, prolate, icosahedral virus, the bacteriophage T4. It involves large-scale movements and remodeling of capsid protein domains that lead to extraordinary capsid stability and additional 70% increase in volume. Small changes in capsid subunit interactions lead to profound alterations in capsid size and volume, allowing evolution of viral variants with survival advantages. These basic principles are relevant to all viruses, particularly the herpesviruses that share common structural features and assembly mechanisms with T4.

Tailed double-stranded DNA bacteriophages (phages), constituting the order Caudovirales, are probably the most abundant and widely distributed microorganisms on Earth, with a total number estimated to be on the order of 10^{31} (1). Bacteriophage T4, a member of the Myoviridae family, possessing a contractile tail and infecting *Escherichia coli*, has long served as an extraordinary model to elucidate the basic mechanisms of virus assembly (2–4) and genome packaging (5, 6). In addition, T4 phage has emerged as an excellent platform for various biomedical applications, including vaccine design and gene therapy (7–9).

The T4 virion has a 1,200-Å-long and 860-Å-wide prolate head (or capsid) (2, 10), encapsidating a ~170-kbp linear DNA genome (11), and a ~1,400-Å-long contractile tail (12–15) to which six ~1,600-Å-long tail fibers are attached (4). The prolate architecture of T4 capsid is unique among the well-characterized large phages that are generally constituted by isometric capsids (16). The T4 capsid shell is composed of the major capsid protein, gp23* ("*" represents cleaved form), organized into a hexagonal lattice characterized by the triangulation numbers $T_{end} = 13$ *laevo* for the icosahedral ends and $T_{mid} = 20$ for the midsection (10). Eleven of the 12 capsid vertices are occupied by pentamers of the vertex protein, gp24* (17), whereas the 12th vertex is formed by the dodecameric protein, gp20, which functions as a portal for entry of DNA into the capsid during genome packaging and its exit during infection (18, 19).

In Caudovirales phages and herpesviruses the capsid assembly begins with the formation of a proteinaceous DNA-free precursor, or prohead, composed of an inner scaffolding core and an outer shell (2). In T4, the prohead assembly is initiated on the *E. coli* inner membrane by the dodecameric portal protein, gp20, which nucleates the assembly of the major capsid protein gp23 and the major scaffolding protein gp22. The outer shell, made of 155 hexameric capsomers of gp23 and 11 pentameric capsomers of the vertex

Author contributions: V.B.R. and M.G.R. designed and supervised research; Q.F., W.-C.T., and M.M. performed research; Q.F., A.F., Q.S., and V.B.R. analyzed data; and Q.F., A.F., Q.S., and V.B.R. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission. E.O. is a guest editor invited by the Editorial Board.

[1]Q.F., W.-C.T., and A.F. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: rao@cua.edu.

This article contains supporting information online at http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2203272119/-/DCSupplemental.

Published September 26, 2022.

protein gp24, coassembles with the inner scaffolding core composed of the major core protein gp22 (~580 copies); the prohead protease gp21 (~55 copies); internal proteins IPI (~360 copies), IPII (~360 copies), and IPIII (~370 copies); gp67 (~340 copies); gp68 (~240 copies); and gpAlt (~40 copies) (2, 20).

Completion of prohead assembly activates the gp21 protease, which then carries out the maturation cleavages of the shell and core proteins (21). The 65 N-terminal residues of gp23 and 10 N-terminal residues of gp24 are cleaved off, producing gp23* and gp24*, respectively (2). On the other hand, the core structure is degraded into small peptides that escape from the prohead, liberating space for genomic DNA. This "unexpanded" head is released into the cytosol and serves as a precursor for genome encapsidation driven by a pentameric packaging motor (gp17) assembled on the portal vertex (3, 22). During DNA packaging, a profound conformational change occurs in the unexpanded head shell, which "expands" gp23* capsomers, resulting in a large increase in the capsid volume and stabilization of the capsid ("expanded" head) (2, 20).

Although the above basic head assembly pathway is common to all the well-studied tailed phages, there are significant differences in the core composition and maturation processes. Although many phages [P22 (23), phi29 (24), T7 (25), SPP1 (26)], like T4 (2, 3), encode dedicated scaffolding proteins, some phages such as HK97 (27) and T5 (28, 29) have the N-terminal region of the major capsid protein acting as a scaffolding protein. Furthermore, the scaffolding core of phage T4 (and phiKZ) is a complex structure consisting of a large series of core proteins as listed above (30). Another significant difference is the mechanism of scaffold removal. While T4 (21), HK97 (27), phiKZ (30), and λ (31) employ maturation proteases that get activated upon head assembly and digest the core, phages such as P22 (23), T7 (25), and SPP1 (26) expel the scaffolding proteins through the openings located at the centers of capsomers. The coreless unexpanded heads then undergo expansion during packaging, resulting in a large increase in capsid volume, with the exception of phage phi29 (32), which shows only small conformational changes without a significant increase in capsid volume.

While the coreless, unexpanded head is conformationally stable in most phages, the T4 unexpanded head expands spontaneously, both in vitro and in vivo, even in the absence of DNA packaging (33). Additionally, it creates binding sites for two special outer capsid proteins Soc (small outer capsid protein) (34, 35) and Hoc (highly antigenic outer capsid protein) (36). Consequently, the expanded T4 capsid is decorated with up to 870 copies of Soc, which binds at the interfaces between adjacent gp23* capsomers, and up to 155 copies of Hoc, which binds to the centers of the gp23* hexamers. Soc stabilizes the capsid against extremes of pH and temperature (34, 37), whereas Hoc has only a marginal effect on capsid stability. The mechanism that triggers capsid expansion, and how the expansion transformation creates binding sites with exquisite specificity for Soc and Hoc, remain poorly understood. Furthermore, the structural features of so large a prolate capsid and how the capsid length is precisely controlled such that a cylindrical T = 20 midsection is assembled also remain unknown and elusive.

Previously we reported the X-ray structures of the vertex protein gp24 (17), Soc (35), and Hoc (36), a medium-resolution cryogenic electron microscopy (cryo-EM) structure of the expanded head (18), and, most recently, a near-atomic-resolution, icosahedrally averaged, cryo-EM structure of a mutant isometric head (38). These also provided the in situ structures of the gp23*, gp24*, and Soc proteins. However, the atomic structure of the prolate wild-type (WT) head, in either unexpanded or expanded form, could not be determined thus far. One of the major bottlenecks in generating high-resolution structures of prolate viruses is that it requires a large amount of cryo-EM data due to the large size of the capsid and low symmetry (C5) of the prolate particles. The defocus gradient in such large particles is detrimental to high-resolution reconstruction. Furthermore, the unexpanded T4 heads are extremely difficult to produce because of their high fragility and spontaneous expansion.

Here, by producing stable unexpanded head particles and collecting large cryo-EM datasets with a direct electron detector and adopting a block-based reconstruction method (39), we generated near-atomic-level resolution structures of both the unexpanded and expanded prolate heads of phage T4. The structures illustrate the movements that govern head expansion, by rotation and twisting of the gp23* subunits around the anchor points located near the quasi-threefold axes. Dramatic local conformational rearrangements of the capsid protein subunits, such as helix unfolding and locking the conformations of flexible regions, create new intercapsomer and intracapsomer interface interactions that greatly stabilize the capsid shell after expansion. These rearrangements create new binding sites for assembly of the capsid decoration proteins Soc and Hoc in the expanded state, further locking and reinforcing the expanded structure. The structures further show that the capsid length is controlled by subtle changes in intercapsomer interactions that modulate the angles between capsomer planes. Finally, our analyses suggest a plausible mechanism for capsid expansion triggered by a conformational switch at the symmetry-mismatched portal vertex that propagates as a massive "wave" through the entire capsid structure.

Results

Structure of the Unexpanded T4 Head. It has been extremely difficult to generate stable unexpanded T4 heads for structure determination because these particles expand spontaneously both in vivo and in vitro (40). In addition, the prolate T4 unexpanded heads are fragile, particularly in the flexible midsection, resulting in damaged or broken particles. We were able to overcome these problems based on a serendipitous observation that unexpanded heads assembled with a heterologously expressed gp20 portal from a recombinant plasmid remained largely in the unexpanded state and do not spontaneously expand.

To generate the unexpanded heads (or capsids) for structure determination, gp20 was briefly pre-expressed in *E. coli* from a T7 expression plasmid, and the bacteria were then infected with a T4 phage mutant that is deficient in portal protein, genome packaging, neck, and tail assembly (*10 AM13 AM17 AM20 AM*). Heads isolated from these infections had undergone normal maturation cleavages by the gp21 protease, generating cleaved capsid proteins gp23* (*SI Appendix*, Fig. S1 *A* and *B*) and gp24*. The inner scaffolding core was digested to small peptides by gp21 protease, most of which escape the heads. However, unlike the heads from a typical phage T4 infection, these coreless heads do not undergo spontaneous expansion (*SI Appendix*, Fig. S1*C*). When purified by ion exchange chromatography, only a single capsid peak corresponding to unexpanded heads was obtained (*SI Appendix*, Fig. S1 *A* and *B*) (40, 41). Therefore, it appears that the pre-expressed gp20 portal protein in the absence of its interacting components, such as gp23 (major capsid protein), gp22 (scaffolding core protein), and gp40 (membrane chaperone), assembles into a dodecameric structure

whose conformation is different from that in a natural phage infection.

Single-particle cryo-EM analysis was applied for structure determination of these unexpanded head particles (*SI Appendix*, Fig. S1*C*). Head particles extracted from 7,023 motion-corrected micrograph movies were subjected to 2D classification (42), which resulted in selection of 28,706 particles for 3D reconstruction. These were then used to reconstruct a 5.1-Å-resolution, fivefold averaged, cryo-EM map of the head by combining conventional methods (42–44) and a block-based reconstruction method (39) that accounted for the defocus gradient and flexibility of the head particles (*SI Appendix*, Fig. S2*A*).

The T4 unexpanded head is a ∼950-Å-long, ∼700-Å-wide prolate icosahedron made of the dodecameric portal protein gp20 at one vertex, pentameric capsomers of gp24* at the other 11 vertices, and 155 hexameric capsomers of gp23* major capsid protein (Fig. 1*A*). There are central openings with ∼14-Å and ∼17-Å diameter in each gp24* and gp23* capsomer, respectively (Fig. 1 *B* and *C*). These openings might allow the escape of the digested polypeptides of the core (and capsid) proteins out of the unexpanded head. Like the mature phage capsid, the unexpanded head is characterized by the triangulation numbers T_{end} = 13 for the caps and T_{mid} = 20 for the midsection (Fig. 1*D*) (10). The unexpanded head, as expected, lacked the Soc and Hoc decoration proteins, which can bind only to expanded capsids (34–36). The cryo-EM density allowed resolution of polypeptide chains in all parts of the head (*SI Appendix*, Fig. S2 *C* and *D* and Table S1), except for the 12-fold-symmetric portal protein disobeying the applied fivefold symmetry and the five gp23* capsomers that are in interaction with this portal vertex (Fig. 1 *B* and *D*). The density in the portal vertex region is weak, suggesting that this part of the unexpanded head is flexible.

The structure reveals a striking feature of the prolate capsid. Unlike the isometric capsids, the prolate structure has a different distribution of intercapsomer angles in the midsection when compared to the icosahedral caps (Fig. 1*D* and *SI Appendix*, Table S2). Calculation of the angles between the planes of adjacent gp23* capsomers shows that the angular values occurring in the icosahedral caps can be subdivided into three groups: values of ∼38° observed near the pentameric vertices; values of 17–21°, which prevail in the faces of icosahedral caps; and less frequent small angles of 1–2° (Fig. 1*D* and *SI Appendix*, Table S2). On the other hand, the elongated midsection has a broader spectrum of angles between adjacent capsomers. Most significantly, angles ranging between 3° and 14° are observed only in the midsection but not in the icosahedral caps (Fig. 1*D* and *SI Appendix*, Table S2).

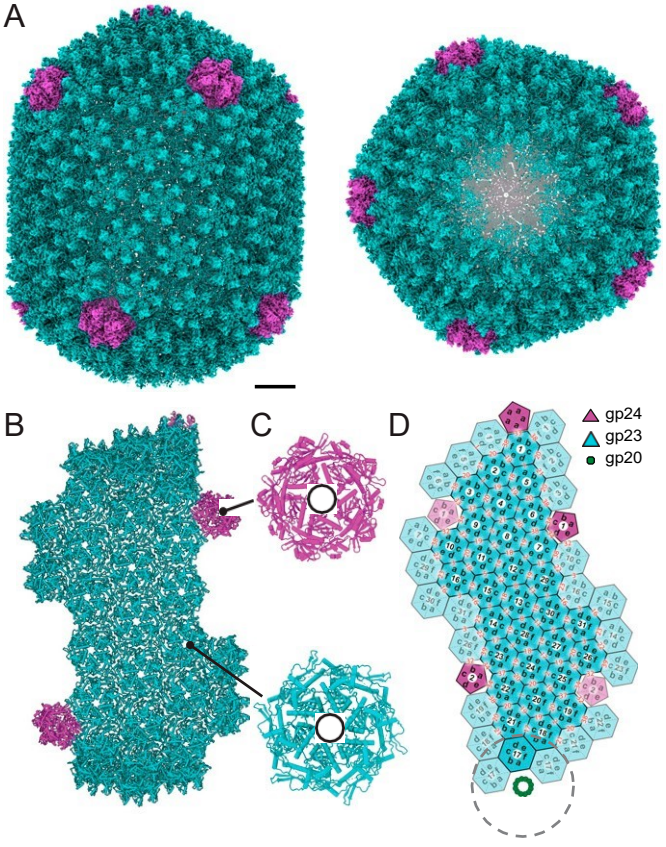

Fig. 1. Structure of the T4 unexpanded head. (*A*) Surface representation of the cryo-EM reconstruction of the T4 unexpanded head showing a side view (*Left*) and an end-on view looking down the fivefold axis of the head (*Right*). Gp23*, gp24*, and uninterpreted density around the portal vertex are colored blue, magenta, and gray, respectively. The scale bar represents 100 Å. (*B*) Atomic model of the T4 unexpanded head. Only molecules within one fivefold asymmetric unit are shown. Gp23* and gp24* are colored as in panel *A*. (*C*) Atomic models of a hexameric capsomer and a pentameric capsomer. The central opening of each capsomer is outlined with a black circle. (*D*) Diagrammatic organization of the T4 unexpanded head viewed from outside the virus. The hexameric capsomers are outlined with hexagons. The capsomers are labeled with numbers. The angles between neighboring capsomers are indicated in the boundaries between neighboring capsomers. The hexameric capsomers and portal protein assembly that are not modeled are indicated with a gray, dashed circle.

Structure of the Major Capsid Protein gp23* in Unexpanded Conformation. The gp23* protein has a polypeptide fold similar to the capsid protein of bacteriophage HK97 (45), consisting of the peripheral (P) domain, located in capsomer's periphery, axial (A) domain, located near capsomer's axis, and an additional insertion (I) domain, which forms characteristic protrusions on the T4 capsid surface (Fig. 2*A*). The insertion domain is connected to the other parts of the structure via long linkers (I-domain linkers) (Fig. 2*A*). The I-domain linkers, which are functionally equivalent to the E loop of HK97 capsid protein (45), participate in both intracapsomer and intercapsomer interactions in the capsid shell (see below). The I domain sits on top of a neighboring gp23* subunit located within the same hexameric capsomer. With an interacting surface area of ∼460 $Å^2$ this domain greatly reinforces the capsomer structure. Notably, the binding interface between the I-domain and the neighboring subunit are radically different in the unexpanded and expanded capsid structures (see below). This extra I domain, inserted via the E-loop-equivalent I-linkers and sitting on top of a neighboring subunit, is a unique capsomer stabilization feature, observed so far only in phage T4. Capsid proteins of other phages such as P22 and phi29 also have extra domains compared to the classical HK97 fold (23, 46–48). However, these domains are inserted into the HK97 fold differently from T4, and significantly, they do not sit on top of the neighboring subunits.

Some parts of the unexpanded gp23* molecule including the N-terminal region; amino acids (aa) 66–105; one of the I-domain linkers, aa 132–160; and one of the linkers connecting the A and P domains (AP linker), aa 486–502, are flexible and not resolved in the cryo-EM map of the unexpanded head. However, these regions are well resolved in the expanded head structure as they are stabilized by interactions with the neighboring subunits (see below).

The asymmetric unit of the unexpanded head reconstruction contains 186 gp23* subunits, organized into 31 hexameric capsomers (Fig. 1*D*). We were able to build the atomic structures

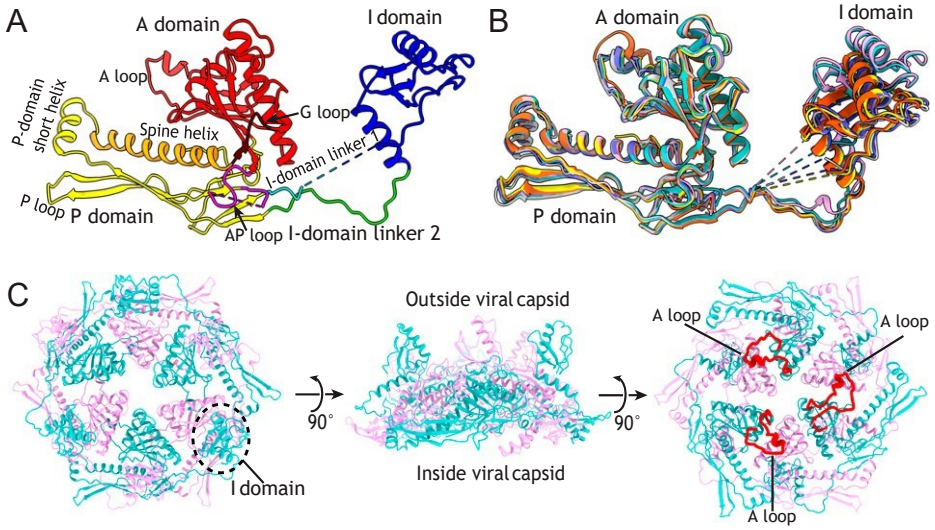

Fig. 2. Gp23* monomer and hexamer structures in the unexpanded head. (*A*) Gp23* monomer structure. The A domain (red), spine helix (orange), AP loop (magenta), P domain (yellow), I-domain linker 1 (cyan), I-domain linker 2 (green), I domain (blue) and G loop (brown) are shown. (*B*) Structural variations of gp23* monomers. The six monomers of hexameric capsomer 10 are superimposed onto each other. (*C*) Structure of hexameric capsomer 10. Alternative gp23* subunits are colored violet and cyan. The I domain of one gp23* subunit is indicated by a dotted oval (*Left*). The A loops of three gp23* subunits are highlighted by red cartoons (*Right*), showing the contacts between the A loop from one gp23* subunit and the A domain of its adjacent gp23* subunit.

of all symmetry-independent gp23* subunits except for those belonging to the capsomers adjacent to the portal proteins, which are probably flexible. All built gp23* subunits have similar overall structures. Significant local conformational variations occur in the P loop, short helix of the P domain, A loop, I domain, and I-domain linkers (Fig. 2 *A* and *B*). The P loop and short helix of the P domain are located in the periphery of each hexameric capsomer and make contacts with neighboring capsomers. Thus, the conformational variations in these peripheral structural elements probably reflect adaptation of gp23* subunits to different capsomer environments, namely to diverse angles between neighboring capsomers (Fig. 1*D* and *SI Appendix*, Table S2). The I domain and I-domain linkers mediate intracapsomer interactions by extending from one gp23* subunit to the outer surface of an adjacent subunit (Fig. 2*C*), suggesting that their conformational variations might be caused by the different relative orientations and distances between neighboring gp23* subunits within each hexameric capsomer. The A loop from one gp23* subunit lies under the A domain of an adjacent subunit with an averaged interface area of ~340 Å^2, dominating the interactions between the A domains in the core of each hexameric capsomer (Fig. 2*C*). This extended A loop, strongly interacting with another gp23* subunit, is also a unique characteristic of the T4 major capsid protein that is not observed in other tailed phages (48). Conformational variations are also observed in the A loops due to different relative orientations and distances between neighboring A domains. The most conformationally different gp23* subunits are those adjacent to the pentameric vertices of the capsid, because these subunits have the most distinct environments by virtue of their association with gp24* pentamers.

Each hexameric capsomer adjacent to a pentameric vertex is skewed into two gp23* trimers roughly related by a twofold axis (*SI Appendix*, Fig. S2*C* and Table S3). Such a hexamer skewness was also observed in the unexpanded heads of other phages such as HK97 (27, 49), T5 (29), P22 (23), and T7 (25), but not in the thermophilic phage P23-45 (50). Vertex-adjacent T4 capsomers 10 and 26 (*SI Appendix*, Table S3), showing the largest skewness, have locations corresponding to the largest capsid curvatures, compared to the other vertex-adjacent capsomers (Fig. 1*D* and *SI Appendix*, Table S3). However, the gp23* hexamers that are not adjacent to pentameric vertices are not skewed and roughly sixfold symmetric (*SI Appendix*, Table S3). This observation suggests that the skewness of the vertex-adjacent gp23* capsomers might be needed to adapt to the large curvatures near the unexpanded head vertices.

Structure of the Vertex Protein gp24* in Unexpanded Conformation. The pentameric capsomers each constituted by five gp24* subunits have similar structures (Fig. 1*C*) and obey the fivefold symmetry (*SI Appendix*, Table S4). Like gp23*, the gp24* subunits have the HK97 fold (45), consisting of the P and A domains, and the additional I domain, connected to the rest of the structure via two I-linkers (Fig. 3*A*). The structures of all 55 gp24* molecules in the unexpanded head were determined, but they do not show significant conformational variations among themselves (Fig. 3*B*) and are overall similar to the previously reported X-ray structure of the *E. coli*-expressed gp24 in uncleaved form (Cα RMSD: 3.1 Å) (17). However, there are several notable differences between the structures of vertex-associated gp24* (vertex-gp24*) and isolated gp24 (Fig. 3*C*): The long spine helix of the P domain is straight in the vertex-gp24* structures, whereas it is twisted in the isolated gp24 structure (Fig. 3*D*); the 20 N-terminal residues are resolved in the isolated gp24 structure, whereas they are disordered in the vertex-gp24* structures, probably because of the cleavage of the 10 N-terminal residues and flexibility of residues 11–20 (Fig. 3*E*); residues 348–359 in the vertex-gp24* structure are disordered, whereas they are ordered in the isolated gp24 structure (Fig. 3*F*); residues 143–154 of I-domain linker 2 that are disordered in the isolated gp24 structure are fully resolved in the vertex-gp24* structures, because they are stabilized by interactions with the nearby capsomers (Fig. 3*G*); and more of the I-domain linker 1 is resolved in the vertex-gp24* structures because it is stabilized by intracapsomer interactions (namely, residues 63–81 are disordered in the isolated gp24 structure, whereas only residues 60–68 are disordered in the vertex-gp24* structures) (Fig. 3*G*).

Sets of five gp24* molecules assemble into pentameric capsomers (Fig. 3*H*), with the P domains and A domains forming

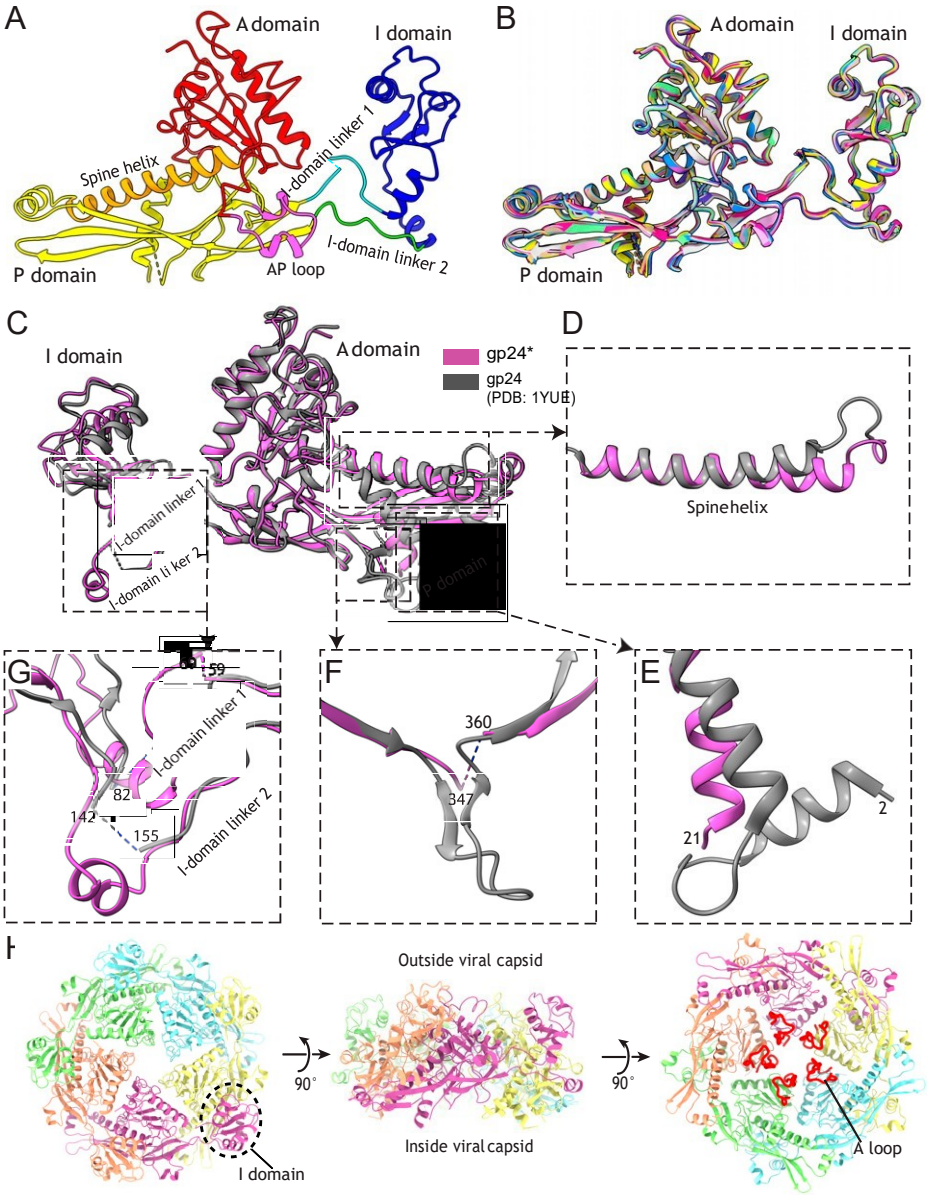

Fig. 3. Gp24* monomer and pentamer structures in the unexpanded head. (*A*) Gp24* monomer structure. The A domain (red), spine helix (orange), AP loop (magenta), P domain (yellow), I-domain linker 1 (cyan), I-domain linker 2 (green), I domain (blue) and G loop (brown) are shown. (*B*) Superimposition of the 11 gp24* monomer structures onto each other within each asymmetric unit of the unexpanded head. (*C*) Superimposition of cryo-EM gp24* structure (magenta) and gp24 crystal structure (dark gray; PDB ID: 1YUE) (17). (*D*–*G*) Zoomed-in views of the corresponding regions in panel *C*, showing the detailed differences between these two structures. The first N-terminal residues are labeled in panel *E*. The disordered residues in panels *F* and *G* are indicated by dashed lines. (*H*) Structure of pentameric capsomer 2. The I domain of one gp24* subunit is indicated by a dotted oval (*Left*). The A loops of gp24* subunits are highlighted by red cartoons (*Right*).

the periphery and core of the capsomers, respectively. As in the hexameric gp23* capsomers, the I domain and I-domain linkers of gp24* participate in intracapsomer interactions. However, unlike in the gp23* capsomers, the A loop of each gp24* molecule folds back to the bottom of its own A domain, making limited intersubunit contacts within each capsomer (Fig. 3*H*).

Intercapsomer Interactions in Unexpanded Head. Neighboring capsomers associate with each other mainly near the quasi-threefold axes of the capsid shell via potential hydrophobic interactions between the gp23*/gp24* P loops and via potential electrostatic interactions between the short, negatively charged α-helix in the P domains and a positively charged region in the P-domain β-strands of gp23* and gp24* (Fig. 4). The negatively charged residues in the P-domain helices involved in these charge interactions are conserved in gp23* and gp24* (Fig. 4).

The association between neighboring capsomers is reinforced by the gp23* I-domain linker 2 and the gp24* I-domain linker 2 (Fig. 4 *B* and *D*). In addition, the AP loop of gp24* (Fig. 3*D*) interacts with the I-domain linker 2 of gp23* from a neighboring hexameric capsomer (Fig. 4*D*). However, the corresponding region (aa residues 384–400) in the AP loop of gp23* is disordered and probably does not make contacts with its neighboring capsomer.

Structures of the Expanded Head and the Expansion-Induced Conformational Transitions. The fivefold-averaged cryo-EM structure of the expanded T4 head has been determined to 3.4-Å resolution (Fig. 5*A* and *SI Appendix*, Figs. S1*D* and S2*B*), which allowed building of an atomic model consisting of all 930 gp23* molecules, 55 gp24* molecules, and 860 of 870 Soc molecules (Fig. 5 *B*–*E*). The two Soc molecules between hexameric

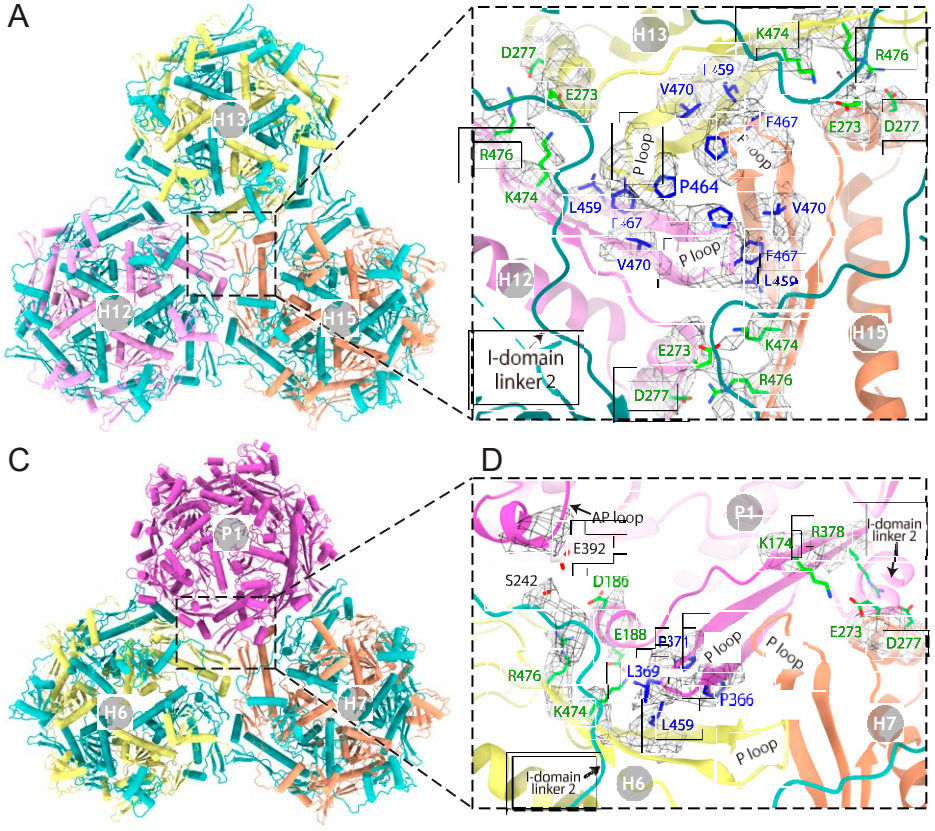

Fig. 4. Intercapsomer interactions in the unexpanded head. (*A*) Structures of hexameric capsomers 12 (H12), 13 (H13), and 15 (H15). (*B*) Zoomed-in view of the boxed region in panel *A*, showing the intercapsomer interactions between hexameric capsomers, including the potential hydrophobic interactions and the potential charge interactions around the quasi-threefold axis. Sidechains involved in the hydrophobic and charge interactions are colored blue and green, respectively. Cryo-EM densities of the residues involved in these interactions are shown in mesh. (*C*) Structures of pentameric capsomer 1 (P1) and hexameric capsomers 6 (H6) and 7 (H7). (*D*) Zoomed-in view of the boxed region in panel *C*, showing the intercapsomer interactions near one of the vertices, including the interactions between the AP loop of gp24* and I-domain linker 2 of a neighboring gp23*, as well as potential hydrophobic interactions and potential charge interactions around the quasi-threefold axis. Cryo-EM densities of the residues involved in these interactions are shown in mesh. Because the interactions between hexamer 6 and 7 are similar to what has been shown in panel *B*, they are not shown in this panel for clarity.

capsomers 16 and 31 within each fivefold asymmetric unit could not be built because the density is too weak due to either low or no occupancy of Soc at these sites. During expansion, in addition to a major reorganization of the networks of interactions, the T4 head gains a ~70% increase in volume, from ~2.1×10^8 Å^3 to ~3.5×10^8 Å^3 (Fig. 5*A*), the capsid wall gets thinned, the central openings of hexameric capsomers shrink from ~17 Å to ~14 Å in diameter, and binding sites emerge for the capsid decoration proteins Hoc and Soc (Movies S1 and S2).

The expanded head structure shows that some of the intercapsomer interactions observed in the unexpanded head are no longer maintained after capsid expansion. Although the I-domain linkers 2 of gp24* and gp23* remain involved in intercapsomer interactions, the former maintains similar contacts with its neighboring capsomer, whereas the latter binds to its neighboring capsomer by using a different region after expansion (Figs. 4 and 6 and *SI Appendix*, Fig. S3). In addition, the AP loop of gp24* still interacts with its neighboring gp23* capsomer, but its binding interface changes during expansion (Figs. 4*D* and 6*D* and *SI Appendix*, Fig. S3 *D–F*).

Strikingly, however, the distribution of the intercapsomer angles does not change significantly (*SI Appendix*, Table S2), and the overall prolate shape of the capsid is faithfully recapitulated during the expansion (Figs. 1*A* and 5*A*). This is because most of the intercapsomer interactions around the quasi-threefold axes remain unchanged, including the hydrophobic interactions between the P loops and the charge interactions between the P-domain short α-helix and β-strand (Fig. 4*B* and 6*B*), suggesting that these interactions are important for maintaining the capsid integrity. Furthermore, structures of these regions around the quasi-threefold axes also remain nearly unchanged, suggesting that they act as anchor points for rotations occurring during capsid expansion (Movies S1 and S2). In other phages, the intercapsomer interactions near the quasi-threefold axes are also largely preserved during capsid expansion (25, 27, 29, 49, 50) suggesting that this is a general characteristic of the expansion process.

Rotations of the A domain and spine helix and twists of the β-strands in the P domain and P-domain-proximal region of the I-domain linkers are observed in both gp23* and gp24* subunits, relative to the anchor points located near quasi-threefold axes (Movies S1 and S2). These rotations and twists make the gp23* and gp24* structures oriented more tangential to the capsid surface. Large displacement (>16 Å) occurs in the I domain of each gp23* subunit (Movie S1). As a consequence, different binding modes between the gp23* I domain and the adjacent gp23* subunit are observed in the unexpanded and expanded structures. The I domain lies on the junction between the spine helix and the A domain of the adjacent gp23* subunit in the expanded head, while it stands near the G loop of the adjacent gp23* subunit in the unexpanded head structure (Fig. 2 *A* and *C*). As a result, the interface area between the I domain of gp23* and its neighboring gp23* subunit from the same capsomer increases from ~460 Å^2

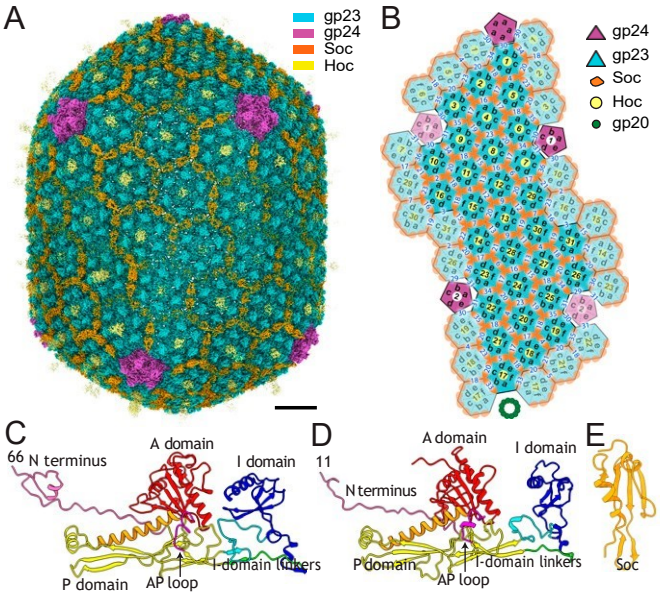

Fig. 5. Structure of the T4 expanded head. (*A*) Surface representation of the cryo-EM reconstruction of the T4 expanded head. The scale bar represents 100 Å. (*B*) Diagrammatic organization of the T4 expanded head viewed from outside the virus. The hexameric capsomers are outlined using hexagons. The capsomers are labeled with numbers. The angles between neighboring capsomers are indicated in the boundaries between neighboring capsomers. (*C*) Gp23* structure. (*D*) Gp24* structure. The N termini of these two structures are colored pink; while the remaining subdomains are color-coded as follows: A domain (red), spine helix (orange), AP loop (magenta), P domain (yellow), I-domain linker 1 (cyan), I-domain linker 2 (green), I domain (blue) and G loop (brown). (*E*) Soc monomer structure.

to ~650 Å^2 after expansion. In contrast, much smaller displacement (~8 Å) occurs for the I domain of gp24* (Movies S1 and S2), which lies on the junction between the spine helix and the A domain of the adjacent gp24* molecule before and after expansion. The A loop of gp23*, which plays critical roles in A domain–A domain interactions in the core of each hexameric capsomer in the unexpanded head, folds back to the bottom of its own A domain and makes much fewer intracapsomer contacts in the expanded head.

The I-domain linkers of both gp23* and gp24* are stretched and undergo significant conformational changes during expansion. The α-helix of the I-domain linker 2 of gp23* partially unfolds, moves to the periphery of its capsomer, and makes extensive intercapsomer contacts (38) (Movie S1). Partial unfolding also occurs in the α-helix of the I-domain linker 1 of each gp24* subunit (Movie S2).

The N-terminal regions of gp23* and gp24* undergo large conformational rearrangements and become involved in both intracapsomer and intercapsomer interactions. The short α-helix of gp24* (aa 22–29) unfolds, and a part of it (aa 26–29) becomes involved in parallel β-structure interactions with an I-domain linker of an adjacent gp24* subunit from the same pentameric capsomer (Fig. 6*E*). The previously disor- dered N-terminal region of gp24* (aa 11–20) extends to the peripheral outer surface of a neighboring hexameric capsomer and participates in extensive intercapsomer interactions (Fig. 6*E*). The disordered N terminus of gp23* becomes resolved after the capsid expansion as well (Fig. 5*C*). In the expanded capsid, this N-terminal region forms an extended arm and *N*-fist (38), which are involved in extensive intracapsomer and intercapsomer interactions (Fig. 6*D*). In other phage systems, the N-terminal regions of the capsid proteins also undergo refolding during expansion (25–27, 49, 50) and become involved in extensive intracapsomer and intercapsomer interactions in the mature heads. Furthermore, all flexible regions in the I-domain linkers of gp23* and gp24* are stabilized by intracapsomer or intercapsomer interactions and become resolved in the expanded capsid structure (Fig. 5 *C* and *D*).

In addition to interactions with other major capsid protein subunits that stabilize the expanded capsid, the *N*-fist and I-domain linkers of the expanded gp23* form binding sites for Soc molecules on the capsid surface (*SI Appendix*, Fig. S4 *A*–*C*). Each Soc molecule attached to the capsid interacts with two adjacent gp23* capsomers. With a binding surface area of ~1,160 Å^2, these interactions clamp capsomers together (*SI Appendix*, Fig. S4 *B* and *C*). At the same time, three capsomers are tied together through trimerization of Soc subunits by Soc–Soc interactions at the quasi-threefold axes. All these intricate interactions greatly increase capsid stability and further reinforce capsid structure by forming an interactive cage around the shell (*SI Appendix*, Fig. S4*A*). This type of capsid stabilization and the fold used by the tadpole-shaped T4 Soc protein (35) are distinct from those of other capsid decoration proteins found in other phages and viruses (51, 52).

Curiously, the expanded capsid reconstruction shows additional helical densities below the P domains of the gp24* vertex protein subunits. These densities may correspond to the N-terminal 10-aa peptide pieces of gp24, which are cleaved off by the maturation protease but remained attached to gp24*. Alternatively, these densities may correspond to a part of another digested protein of the inner scaffolding core.

Structure–Function Analysis of the Capsid Shell. A large collection of mutants in the major capsid protein gp23 have been well characterized by genetic and biochemical studies (2, 53). Many of these are classic mutants that led to a detailed morphogenetic pathway for phage T4 head assembly and length control. The present unexpanded and expanded capsid reconstructions allowed detailed analyses of the structural and functional consequences of these mutations.

Gp31 chaperonin bypass mutations. T4 encodes a unique cochaperonin protein, gp31, a structural and functional analog of the *E. coli* small chaperonin GroES. It replaces GroES in the GroEL–GroES chaperone complex that is essential for proper folding of the gp23 major capsid protein (but not gp24) (2). However, a single-substitution mutant A455V and a triple-substitution mutant G292S/V306I/V307I completely eliminate the requirement for the gp31 cochaperonin (54). We found that all these residues are part of the P-domain, residue 455 in the P-domain β-sheet, and residues 292, 306, and 307 in the P-domain spine helix (*SI Appendix*, Fig. S5, cyan dots), all making this domain more hydrophobic, which probably helps gp23 to fold correctly and with appropriate kinetics without the assistance of the gp31 chaperonin.

Gp24-bypass mutations. There are single-substitution mutations in gene 23, which eliminate the necessity for the vertex protein gp24 by forming pentamers in addition to hexamers that occupy the pentameric vertices (2, 55). Such mutations, N381S, A387T, and K420R, are located in the A domain of gp23, which forms the core of the capsomers. In both the unexpanded and expanded head structures, residues 381 and 387 are in a short α-helix, whereas residue 420 is in a loop region (*SI Appendix*, Fig. S5, black dots). Superposition of the gp23* structure onto the pentameric gp24* vertex of the unexpanded head shows that all these residues are located in the interface of neighboring

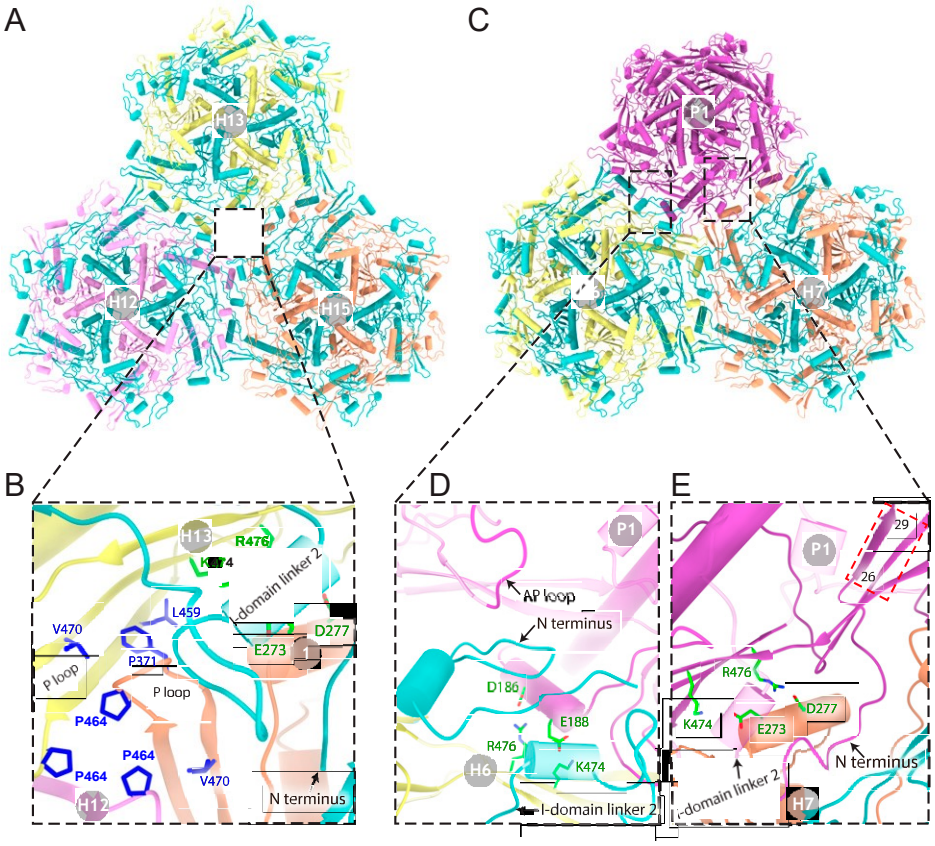

Fig. 6. Gp23* and gp24* monomer structures and intercapsomer interactions in the expanded head. (*A*) Structures of hexameric capsomer 12 (H12), 13 (H13), and 15 (H15), showing the interactions between neighboring hexameric capsomers. (*B*) Zoomed-in view of the boxed region in panel *A*, showing the hydrophobic interactions and the potential charge interactions that are maintained after capsid expansion. (*C*) Structures of pentameric capsomer 1 (P1) and two adjacent hexameric capsomers 6 (H6) and 7 (H7), showing the intercapsomer interaction near one of the capsid vertices. (*D*, *E*) Zoomed-in views of the boxed regions in panel *C*. The AP loop of gp24* contacts with the N terminus of a neighboring gp23*. A red dotted line box in panel *E* indicates the parallel β-structure interactions between two adjacent gp24* subunits. Sidechains involved in the hydrophobic and potential charge interactions in all of these panels are colored blue and green, respectively.

subunits within the capsomer. Therefore, these bypass mutations modulate the intersubunit interactions such that gp23 can form both pentamers and hexamers.

Capsid length mutations. A series of point mutations in the gp23 major capsid protein alter the capsid length and produce isometric or giant heads and various amounts of intermediate-length heads, in addition to native prolate heads (2, 53, 56, 57). Remarkably, all the well-characterized length-changing mutations are clustered near the quasi-threefold axes (*SI Appendix*, Fig. S5), consistent with the present structures that showed no significant movements or conformational changes at the quasi-threefold axes during expansion.

One group of the head-length-affecting mutations, A268V, A275T, D285V, D287V, and D287N, is located in a small region of gp23*, close to the short P-domain helix (*SI Appendix*, Fig. S5; red dots). This negatively charged helix is located near the quasi-threefold axes, relating adjacent gp23* capsomers, and is involved in charge interactions with neighboring capsomers. Although the sidechains of the residues 268, 275, 285, and 287 do not interact directly with neighboring capsomers, the mutations in these sites probably alter the position of the short P-domain helix, which in turn affects the intercapsomer interactions and angles between capsomers.

Another group of length-altering mutations, T457A, R460C, and G461S, is also located near the quasi-threefold axes, relating adjacent gp23* capsomers (*SI Appendix*, Fig. S5; yellow dots). In these quasi-threefold regions, three P loops belonging to adjacent capsomers interact with each other (Fig. 4*B*). Therefore, these mutations probably affect the P loop conformations and therefore intercapsomer interactions.

The third group of mutations, A66D, A94V, and G97S, is located in the N terminus of gp23*, which is flexible in the unexpanded head reconstruction. The first gp23* residue built in the unexpanded atomic model (L106) points to the inside of the capsid, suggesting that the N-terminal part of gp23* was originally located inside the capsid shell and probably interacted with the inner scaffolding core of the immature unexpanded head (58). After expansion, this region gets externalized and remodeled into an extended arm and the *N*-fist, which clearly form extensive intercapsomer interactions in the expanded head as described above. Thus, interactions between the N-terminal domain of gp23 and the scaffolding core (gp22) also affect the intercapsomer angles and play an important role in the capsid length determination. Interactions between the N-terminal regions of the major capsid proteins and the scaffolding proteins is a common feature found previously in the unexpanded heads of phages P22 (23), T7 (25), SPP1 (26), and α80 (59).

A Conformational Switch in Portal–Capsomer Interactions. Our structures show that the capsid protein subunits display distinct conformational states in the unexpended and expanded heads. In the in situ asymmetric reconstruction of the dodecameric portal vertex in the expanded state (19), the N-terminal six amino acids of the portal protein form a "whisker," and in 2 out of 12 portal subunits (#1 and #6), it is anchored to the capsid shell. The N-terminal methionines of these two portal

subunits form potential metal-binding clusters with Met98, Met284, and His282 residues of the gp23* molecules facing these portal subunits. Mutational studies showed that the length of the whisker is critical for head assembly (19). The gp23* residues Met284 and His282 are located in the P-domain of the protein, whereas Met98 is located in the N-terminal gp23* region. However, in the unexpanded head, the N-terminal region is disordered while it forms an extended *N*-arm in the expanded head (see above). Although the portal vertex is not resolved in the unexpanded head reconstruction, because it shows significant flexibility, it is reasonable to assume that similar anchoring of the portal's *N*-whiskers to the capsid shell is also present in the immature unexpanded T4 head. Consistently, assuming the orientations and structures of the gp23* subunits in the portal vertex are similar to those observed in the other 11 capsid vertices in the unexpanded head, the methionine and histidine residues of gp23* subunits are oriented in a similar fashion relative to the portal vertex in both the unexpanded and expanded head structures. However, strikingly, the composition of the gp23* molecules that form the metal cluster is different. In the unexpanded structure, the Met98 residue of gp23* is replaced by the Met444 residue belonging to the P-domain of a neighboring gp23* subunit rather than that of the same subunit in the expanded head structure (*SI Appendix*, Fig. S6). This might account in part for the flexibility of the portal in the unexpanded state and indicates an expansion-associated conformational switch in the capsid subunits at the capsid–portal interface. Therefore, combined with previous observations (60–62), a compelling case can be made for this conformational switch being a trigger for capsid expansion. This then leads to a domino effect, propagating the expansion conformational switch to the neighboring subunits and then to the entire capsid structure as a wave, with the portal at the epicenter (see *Discussion*).

Discussion

The atomic models of the unexpanded and expanded T4 head, containing ~71,000 and ~103,000 residues within each five-fold asymmetric unit, respectively, are among the largest available structures deposited in the Protein Data Bank. These structures, at near-atomic resolution, reveal dramatic conformational transitions in a large prolate viral capsid, generating insights into the mechanisms of capsid assembly and stabilization, head length determination, and shell expansion.

Capsid Assembly and Stabilization. Some of the structural features of the phage T4 capsid and changes during assembly and maturation are similar to those observed in the other phages, whereas others are specific to T4. For instance, similar to other phages such as HK97 and T5 (27, 29, 45), adjacent capsomers in the T4 capsid are reinforced by strong electrostatic interactions between the P domains of the capsid protein subunits located near the quasi-threefold axes and are conserved in both the unexpanded and expanded heads.

As in the other phages such as HK97, the N-terminal region of the T4 major capsid protein, gp23, is probably in interaction with the inner scaffolding core and is cleaved off by a prohead maturation protease that is also part of the core (21). However, unlike the other phages, the T4 protease also cleaves off the N-terminal region of the vertex protein gp24 and degrades the scaffolding core into small peptides (2, 20). The T4 unexpanded head structure shows 14–17 Å diameter openings in each capsomer center, through which the degraded scaffolding peptides probably escape, creating space in the interior of the capsid for viral genome encapsidation. These openings become closed during the capsid expansion. Such openings are also present in phages such as P22 (23) and SPP1 (26) but are larger because these phages do not encode a protease and hence require larger space for the uncleaved scaffolding protein to escape (23).

Similar to the other phages (29, 48), the N-terminal regions of the cleaved gp23* and gp24* capsid proteins undergo substantial structural remodeling during capsid expansion. These regions migrate from the capsid interior to the outer surface and are involved in extensive intercapsomer and intracapsomer interactions, stabilizing the capsid.

Phage T4 exhibits distinctive features, however. An unusual structural element of the T4 major capsid protein is the elongated A loop, which dominates interactions between the axial A domains of adjacent subunits in the unexpanded head. In the expanded capsid, it changes its conformation and makes fewer contacts with its neighboring subunit. Another distinctive feature is that, unlike most other phages, T4 has evolved a special vertex-specific capsid protein gp24* which forms the pentameric vertices, affording maximum curvature for the capsid. Although gp24* and gp23* proteins have sequence and structure similarity, the gp24* capsomers, unlike the gp23* capsomers, undergo the least conformational rearrangement during expansion.

Phage T4 has also evolved special capsid stabilization mechanisms. One such mechanism involves the large I domain with long linkers. The I-domain linkers are functionally equivalent to the E loop of the classical HK97 fold (45) but structurally distinct because they provide an enormous interacting surface. Furthermore, each I domain crosses over and lays on top of a neighboring subunit belonging to the same capsomer. With its ~650 $Å^2$ interacting surface area in the expanded state, the I domains firmly interlock the subunits, greatly reinforcing the capsomer structure (Fig. 6).

Second, the long I-domain linkers are dynamic, undergoing substantial remodeling during unexpanded to expanded transition (Movies S1 and S2). In the unexpanded capsid, the linkers are disordered, whereas in the expanded capsid they are ordered and form numerous intercapsomer and intracapsomer interactions (Fig. 6). Furthermore, this remodeling creates high-affinity binding sites for the assembly of Soc subunits, which, through trimerization and clamping of adjacent capsomers, and with an interacting surface area of ~3,480 $Å^2$, tie together all three capsomers at the quasi-threefold axes. Repetition of these reinforcements across the prolate capsid generate a cage-like structure locking the capsid within it (*SI Appendix*, Fig. S4*A*). Together, the above features impart extraordinary stability to T4 phage, allowing it to survive in hostile environments such as the human gut.

Capsid Length Control. The T4 head assembly is initiated by the portal vertex (3, 63). The portal protein, gp20, recruits the capsid protein gp23, and the core protein gp22, forming a nucleation complex. This facilitates the assembly of the portal-proximal icosahedral cap, followed by the assembly of the midsection, and then the distal cap. Unlike most other isometric phages and viruses, the elongated prolate head of T4 poses additional complexity as it requires precise control of the capsid length, the length of the nonicosahedral midsection. Analysis of the capsid length-altering mutations in the unexpanded and expanded structural states suggests that the capsid length is governed primarily by intercapsomer interactions that modulate the angles between capsomer planes.

In the unexpanded head caps, the angles between adjacent gp23* capsomers belong to three groups: angles of ∼38° surrounding the vertices; angles of 17–21°, which abound in the caps; and a few small angles of 1° and 2° (Fig. 1*D* and *SI Appendix*, Table S2). Therefore, it is reasonable to assume that the triangulation number of the caps (T = 13) is predetermined by the angles around the vertices and by the most common angles of 17–21°.

The elongated midsection shows a broader spectrum of intercapsomer angles. For example, angles of 8°, 11°, and 14° are observed only in the midsection but not in the caps. Furthermore, small angles of 2–5° occur frequently in the midsection (Fig. 1*D* and *SI Appendix*, Table S2). Therefore, during the midsection assembly, smaller angles of 2–14° are favored instead of the common angles of 17–21°, resulting in a smaller curvature of the capsid and, consequently, assembly of an elongated midsection.

Mutational analyses show that alteration of intercapsomer interactions at the quasi-threefold axes make certain ranges of angles less or more favorable. For example, if a mutation makes smaller (2–14°) intercapsomer angles less favorable and common angles of 17–21° more favorable, the gp23 capsomers would assemble into isometric heads instead of elongated prolate (WT) heads. On the other hand, if a mutation makes the small intercapsomer angles favorable, the midsection elongates further, generating "giant" heads that can be up to 10 times the length of WT prolate heads (2, 56, 57). This modulation is not very precise and varies between different mutants and different individual capsid assemblies, hence the position where a pentagonal hole is created in a hexagonal capsomer array for gp24 pentamer incorporation is not fixed, generating a distribution of preferred sites. Consequently, mixtures of different-length capsids, WT prolates, isometrics, intermediates, and giants, with different distributions, are produced in various mutants (2, 53, 56, 57). Even in the WT phage infections, up to ∼5% of the phages are found to contain isometric heads and ∼1% contain intermediate-length heads (64).

Our analyses further suggest that the intercapsomer angles are influenced by interactions between the capsid protein (gp23) and the scaffolding core protein (gp22). In particular, the cluster of N-terminal of length mutants (A66D, A94V, and G97S) are probably part of a region that is originally localized inside the immature capsid and interacts with the scaffolding core. These interactions, and mutations in this region that alter these interactions, can also modulate the intercapsomer angles. In fact, mutations have been identified in the gp22 scaffolding core protein that either alter capsid length or suppress the alterations in the capsid length induced by the capsid protein mutations (65, 66). This was also observed in phage phi29 where mutations in the scaffolding protein give rise to isometric icosahedral particles rather than the prolate WT particles (47, 67).

Thus, small variations in the intercapsomer interactions at the quasi-threefold axes or the capsid–scaffolding core interactions can alter the intercapsomer angles of the midsection. Furthermore, it was reported that mutations that can alter these angles in the icosahedral caps resulted in "wide" heads (68). Therefore, an important lesson emerging from analyses of a large collection of length mutations is that subtle changes, often a slight increase in hydrophobicity (e.g., A268V) that allow similar but slightly different intercapsomer contacts, lead to profound changes in the capsid length and the size of genome it can accommodate. It is tempting to speculate that the T4 phage originally was an isometric virus accommodating ∼120-kbp genome, but a small change in the intercapsomer contacts (e.g., A275T mutation) changed it to a prolate virus that can then accommodate an additional 50-kbp genome, providing home for dozens of additional genes that can confer survival advantages. In fact, ∼75% of the ∼300 open reading frames present in the phage T4 genome are considered to be nonessential under controlled growth conditions but might be essential in the natural environment or for infection of different WT *E. coli* strains. Soc and Hoc and internal proteins are good examples of such nonessential genes that are dispensable under laboratory conditions but impart survival advantages for T4 phage in its natural environment (8).

In this context, it is worth noting an unusual strategy used by the thermophilic phage P23-45 to increase the size of the capsid (50). Instead of increasing the triangulation number (T = 7), the P23-45 phage evolved larger capsomers by increasing the size of the A domain and the E loop of the major capsid protein, which resulted in about twice the larger capsid volume and twice the amount of genome packaged when compared to other T = 7 phages (50).

Capsid Shell Expansion. The unexpanded head of icosahedral phages corresponds to a metastable state of the capsid shell, which is converted to a stable expanded state as the genomic DNA gets packaged. Evidence suggests that, in vivo, head expansion occurs after ∼10–30% of genome is packaged in different phages (62, 69, 70), although the T4 unexpanded heads can expand spontaneously without any DNA being packaged, both in vivo and in vitro (40). However, mutant T4 heads in which some of the portal protein subunits were replaced by a recombinantly expressed green fluorescent protein fusion at the C terminus of the portal protein were found to remain in the unexpanded conformation (62). Our results show that the green fluorescent protein fusion is not essential, and it is probably the pre-expression of the portal protein from a recombinant plasmid and its assembly into a dodecamer in the absence of portal-interacting components such as gp40 (membrane chaperone), gp23 (capsid protein), and gp22 (core protein) that stabilized a pre-expansion conformation.

Furthermore, we discovered that the capsid subunits facing the portal subunits at the symmetry-mismatched T4 portal vertex are in a different conformation in the unexpanded state. In the unexpanded head, the portal anchoring metal cluster is formed by two gp23* subunits molecules, Met284 and His282 of one subunit and Met444 of an adjacent subunit (*SI Appendix*, Fig. S6 *A* and *B*). However, in the expanded head, the Met444 of adjacent subunit is replaced by Met98 of the same gp23* subunit, which is also part of the remodeled *N*-arm conformation in the expanded state (*SI Appendix*, Fig. S6 *C*–*E*). To our knowledge, such a conformational switch at the portal vertex has not been observed in the symmetry averaged head structures of other phages (29, 32, 71).

This conformational switch could be either a consequence of capsid expansion or a cause of capsid expansion. Evidence points to the latter. First, the portal's unique location in the capsid involves special portal–capsid interactions that undergo structural morphing in order to accommodate the symmetry mismatch between the fivefold capsid and the 12-fold portal (19). This means that these interactions are dynamic and can be modulated in response to head maturation processes, such as genome packaging. When the packaging motor assembles on the portal and packaging is initiated, a conformational change in the portal protein subunits probably triggers a conformational change in the capsid protein subunits with which they interact (72–75). Second, when the first packaged genomic

DNA enters the unexpanded head, it is not distributed uniformly in the capsid interior. It settles near the portal vertex (76, 77), imposing a degree of pressure at the unique portal vertex, and the pressure keeps building as packaging continues (78). When it reaches a threshold, in T4 after about 5–10 kb genome is packaged, or about 5 s after packaging is initiated (62, 79), this pressure would push down on the portal or cause displacement of the portal relative to the capsid (32, 71, 72). Since the dodecameric portal is anchored to the fivefold capsid vertex through two N-terminal whiskers of the portal, which form metal-binding clusters with gp23* subunits (19), this would cause a pull on these gp23* molecules, inducing a conformational change. Consequently, a remodeling of the metal-binding cluster would occur, leading to an unexpanded to expanded transition. This conformational transition of the portal-anchoring capsid subunits then triggers energetically favorable conformational changes in their immediate neighbors because of the altered intersubunit interactions, causing a domino effect. This expansion wave thus propagates through the portal-proximal icosahedral cap, then through the capsid midsection to the distal cap, while the capsomers remain fastened at the quasi-threefold axes through the intercapsomer network of electrostatic interactions. The quasi-threefold axes thus act as anchor points around which the gp23* subunits rotate and twist as the expansion wave spreads through the entire capsid (Movie S2) and are further locked in by clamping the capsomers with Soc molecules (*SI Appendix*, Fig. S3). Finally, electron micrographs of T4 polyheads and giant heads showed the presence of rare hybrid expanded heads in which expansion was interrupted midway (2, 3, 60, 61, 80). One half of the head was expanded while the other half remained unexpanded, with a sharp boundary in the middle. No patches of unexpanded or expanded regions were observed within the expanded or unexpanded capsid lattices, respectively, and the expansion front is always perpendicular to the head axis, leading to the conclusion that expansion is polar and spreads rapidly as a wave, with the symmetry-mismatched portal vertex at the epicenter.

Conclusions

In conclusion, our structures and analyses reveal near-atomic-level descriptions of a large prolate virus as it transitions from an immature unexpanded head to an expanded mature head. This transition occurs at the right moment, when the capsid is being loaded with the genome, a signal transmitted to capsid through the dynamic symmetry-mismatched portal vertex. Besides increasing the inner volume by 1.7 times, expansion leads to dramatic movements and remodeling of structural elements that impart extraordinary stability to the capsid and create new sites for attachment of decoration proteins for further structural reinforcement. The external as well as internal surface properties of the capsid are altered not only to house the entire viral genome but also to withstand the internal pressure, survive in a hostile extracellular environment (human gut in the case of bacteriophage T4), and allow efficient delivery of the genetic payload into a new host cell. Yet the process retains remarkable plasticity and allows for rapid evolution of capsids with profound alterations in size by small local changes in intercapsomer interactions. These basic principles are relevant to virus assembly and evolution of viruses in general and can potentially be exploited for biomedical applications (7–9, 81).

Materials and Methods

For preparation of unexpanded heads, *E. coli* BL21 (DE3) RIPL cells containing the WT gp20 expression plasmid were induced with 1 mM isopropyl-β-D-thiogalactoside at 37 °C for 20 min. The cells were then infected with 10 AM13 AM17 AM20 AM mutant phage. For preparation of expanded heads, *E. coli* P301 cells were infected with 10 AM13 AM mutant phage. The infected cells were lysed and the heads were purified by differential centrifugation and ion exchange chromatography. About 3-μL aliquots of the purified heads were frozen onto Lacey carbon EM grids with a Gatan CP3 freezer, and the grids were loaded on an FEI Titan Krios electron microscope operated at 300 kV and equipped with a Gatan K2 Summit detector for data collection.

More detailed procedures are provided in *SI Appendix, Methods*.

Data, Materials, and Software Availability. Cryo-EM reconstructions of the unexpanded head and the expanded head have been deposited in the Electron Microscopy Data Bank under accession numbers EMD-32103 and EMD-32109, respectively. Atomic models of the unexpanded head and the expanded head have been deposited in the Protein Data Bank under the accession numbers 7VRT and 7VS5, respectively.

ACKNOWLEDGMENTS. We thank the Purdue Cryo-EM facility for their help and support and Yingyuan Sun and Thomas Klose for their help with cryo-EM data collection. This work was supported by the NIH (grant AI081726 to M.G.R. and V.B.R.) and NSF (grant MCB-0923873 to V.B.R.). Funding for open access charge: NSF.

Author affiliations: [a]Department of Biological Sciences, Purdue University, West Lafayette, IN 47907; [b]School of Public Health (Shenzhen), Sun Yat-sen University, Shenzhen, Guangdong 518107, China; and [c]Bacteriophage Medical Research Center, Department of Biology, The Catholic University of America, Washington, DC 20064

1. R. W. Hendrix, Bacteriophage genomics. *Curr. Opin. Microbiol.* 6, 506–511 (2003).
2. L. W. Black, M. K. Showe, A. C. Steven, "Morphogenesis of the T4 head" in *Molecular Biology of Bacteriophage T4*, Jim D. Karam, Ed. (American Society for Microbiology, Washington, D.C., 1994), pp. 218–258.
3. L. W. Black, V. B. Rao, Structure, assembly, and DNA packaging of the bacteriophage T4 head. *Adv. Virus Res.* 82, 119–153 (2012).
4. P. G. Leiman *et al.*, Morphogenesis of the T4 tail and tail fibers. *Virol. J.* 7, 355 (2010).
5. S. R. Casjens, The DNA-packaging nanomotor of tailed bacteriophages. *Nat. Rev. Microbiol.* 9, 647–657 (2011).
6. V. B. Rao, M. Feiss, Mechanisms of DNA packaging by large double-stranded DNA viruses. *Annu. Rev. Virol.* 2, 351–378 (2015).
7. P. Tao, J. Zhu, M. Mahalingam, H. Batra, V. B. Rao, Bacteriophage T4 nanoparticles for vaccine delivery against infectious diseases. *Adv. Drug Deliv. Rev.* 145, 57–72 (2019).
8. P. Tao *et al.*, In vitro and in vivo delivery of genes and proteins using the bacteriophage T4 DNA packaging machine. *Proc. Natl. Acad. Sci. U.S.A.* 110, 5846–5851 (2013).
9. J. Zhu *et al.*, A universal bacteriophage T4 nanoparticle platform to design multiplex SARS-CoV-2 vaccine candidates by CRISPR engineering. *Sci. Adv.* 7, eabh1547 (2021).
10. A. Fokine *et al.*, Molecular architecture of the prolate head of bacteriophage T4. *Proc. Natl. Acad. Sci. U.S.A.* 101, 6003–6008 (2004).
11. E. S. Miller *et al.*, Bacteriophage T4 genome. *Microbiol. Mol. Biol. Rev.* 67, 86–156, table of contents (2003).
12. N. M. Taylor *et al.*, Structure of the T4 baseplate and its function in triggering sheath contraction. *Nature* 533, 346–352 (2016).
13. V. A. Kostyuchenko *et al.*, The tail structure of bacteriophage T4 and its mechanism of contraction. *Nat. Struct. Mol. Biol.* 12, 810–813 (2005).
14. P. G. Leiman, P. R. Chipman, V. A. Kostyuchenko, V. V. Mesyanzhinov, M. G. Rossmann, Three-dimensional rearrangement of proteins in the tail of bacteriophage T4 on infection of its host. *Cell* 118, 419–429 (2004).
15. M. L. Yap *et al.*, Role of bacteriophage T4 baseplate in regulating assembly and infection. *Proc. Natl. Acad. Sci. U.S.A.* 113, 2654–2659 (2016).
16. A. Fokine, M. G. Rossmann, Molecular architecture of tailed double-stranded DNA phages. *Bacteriophage* 4, e28281 (2014).
17. A. Fokine *et al.*, Structural and functional similarities between the capsid proteins of bacteriophages T4 and HK97 point to a common ancestry. *Proc. Natl. Acad. Sci. U.S.A.* 102, 7163–7168 (2005).
18. L. Sun *et al.*, Cryo-EM structure of the bacteriophage T4 portal protein assembly at near-atomic resolution. *Nat. Commun.* 6, 7548 (2015).
19. Q. Fang *et al.*, Structural morphing in a symmetry-mismatched viral vertex. *Nat. Commun.* 11, 1713 (2020).
20. P. G. Leiman, S. Kanamaru, V. V. Mesyanzhinov, F. Arisaka, M. G. Rossmann, Structure and morphogenesis of bacteriophage T4. *Cell. Mol. Life Sci.* 60, 2356–2370 (2003).
21. A. Fokine, M. G. Rossmann, Common evolutionary origin of procapsid proteases, phage tail tubes, and tubes of bacterial type VI secretion systems. *Structure* 24, 1928–1935 (2016).
22. S. Sun *et al.*, The structure of the phage T4 DNA packaging motor suggests a mechanism dependent on electrostatic forces. *Cell* 135, 1251–1262 (2008).

23. D. H. Chen *et al.*, Structural basis for scaffolding-mediated assembly and maturation of a dsDNA virus. *Proc. Natl. Acad. Sci. U.S.A.* 108, 1355–1360 (2011).
24. M. C. Morais *et al.*, Bacteriophage phi29 scaffolding protein gp7 before and after prohead assembly. *Nat. Struct. Biol.* 10, 572–576 (2003).
25. F. Guo *et al.*, Capsid expansion mechanism of bacteriophage T7 revealed by multistate atomic models derived from cryo-EM reconstructions. *Proc. Natl. Acad. Sci. U.S.A.* 111, E4606–E4614 (2014).
26. A. Ignatiou *et al.*, Structural transitions during the scaffolding-driven assembly of a viral capsid. *Nat. Commun.* 10, 4840 (2019).
27. R. W. Hendrix, J. E. Johnson, Bacteriophage HK97 capsid assembly and maturation. *Adv. Exp. Med. Biol.* 726, 351–363 (2012).
28. O. Preux *et al.*, A two-state cooperative expansion converts the procapsid shell of bacteriophage T5 into a highly stable capsid isomorphous to the final virion head. *J. Mol. Biol.* 425, 1999–2014 (2013).
29. A. Huet, R. L. Duda, P. Boulanger, J. F. Conway, Capsid expansion of bacteriophage T5 revealed by high resolution cryoelectron microscopy. *Proc. Natl. Acad. Sci. U.S.A.* 116, 21037–21046 (2019).
30. J. A. Thomas *et al.*, Extensive proteolysis of head and inner body proteins by a morphogenetic protease in the giant Pseudomonas aeruginosa phage φKZ. *Mol. Microbiol.* 84, 324–339 (2012).
31. E. Medina *et al.*, Assembly and maturation of the bacteriophage lambda procapsid: gpC is the viral protease. *J. Mol. Biol.* 401, 813–830 (2010).
32. J. Xu, D. Wang, M. Gui, Y. Xiang, Structural assembly of the tailed bacteriophage ϕ29. *Nat. Commun.* 10, 2366 (2019).
33. V. B. Rao, L. W. Black, DNA packaging of bacteriophage T4 proheads in vitro. Evidence that prohead expansion is not coupled to DNA packaging. *J. Mol. Biol.* 185, 565–578 (1985).
34. T. Ishii, M. Yanagida, The two dispensable structural proteins (soc and hoc) of the T4 phage capsid; their purification and properties, isolation and characterization of the defective mutants, and their binding with the defective heads in vitro. *J. Mol. Biol.* 109, 487–514 (1977).
35. L. Qin, A. Fokine, E. O'Donnell, V. B. Rao, M. G. Rossmann, Structure of the small outer capsid protein, Soc: A clamp for stabilizing capsids of T4-like phages. *J. Mol. Biol.* 395, 728–741 (2010).
36. A. Fokine *et al.*, Structure of the three N-terminal immunoglobulin domains of the highly immunogenic outer capsid protein from a T4-like bacteriophage. *J. Virol.* 85, 8141–8148 (2011).
37. P. D. Ross, L. W. Black, M. E. Bisher, A. C. Steven, Assembly-dependent conformational changes in a viral capsid protein. Calorimetric comparison of successive conformational states of the gp23 surface lattice of bacteriophage T4. *J. Mol. Biol.* 183, 353–364 (1985).
38. Z. Chen *et al.*, 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. U.S.A.* 114, E8184–E8193 (2017).
39. D. Zhu *et al.*, Pushing the resolution limit by correcting the Ewald sphere effect in single-particle Cryo-EM reconstructions. *Nat. Commun.* 9, 1552 (2018).
40. V. B. Rao, L. W. Black, Evidence that a phage T4 DNA packaging enzyme is a processed form of the major capsid gene product. *Cell* 42, 967–977 (1985).
41. Z. Zhang *et al.*, A promiscuous DNA packaging machine from bacteriophage T4. *PLoS Biol.* 9, e1000592 (2011).
42. T. Grant, A. Rohou, N. Grigorieff, *cis*TEM, user-friendly software for single-particle image processing. *eLife* 7, e35383 (2018).
43. F. Guo, W. Jiang, Single particle cryo-electron microscopy and 3-D reconstruction of viruses. *Methods Mol. Biol.* 1117, 401–443 (2014).
44. A. Punjani, J. L. Rubinstein, D. J. Fleet, M. A. Brubaker, cryoSPARC: Algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* 14, 290–296 (2017).
45. W. R. Wikoff *et al.*, Topologically linked protein rings in the bacteriophage HK97 capsid. *Science* 289, 2129–2133 (2000).
46. K. N. Parent *et al.*, P22 coat protein structures reveal a novel mechanism for capsid maturation: Stability without auxiliary proteins or chemical crosslinks. *Structure* 18, 390–401 (2010).
47. M. C. Morais *et al.*, Conservation of the capsid structure in tailed dsDNA bacteriophages: The pseudoatomic structure of phi29. *Mol. Cell* 18, 149–159 (2005).
48. R. L. Duda, C. M. Teschke, The amazing HK97 fold: Versatile results of modest differences. *Curr. Opin. Virol.* 36, 9–16 (2019).
49. I. Gertsman *et al.*, An unexpected twist in viral capsid maturation. *Nature* 458, 646–650 (2009).
50. O. W. Bayfield *et al.*, Cryo-EM structure and in vitro DNA packaging of a thermophilic virus with supersized T=7 capsids. *Proc. Natl. Acad. Sci. U.S.A.* 116, 3556–3561 (2019).
51. G. C. Lander *et al.*, Bacteriophage lambda stabilization by auxiliary protein gpD: Timing, location, and mechanism of attachment determined by cryo-EM. *Structure* 16, 1399–1406 (2008).
52. Z. Wang *et al.*, Structure of the marine siphovirus TW1: Evolution of capsid-stabilizing proteins and tail spikes. *Structure* 26, 238–248.e3 (2018).
53. T. Lane, F. Eiserling, Genetic control of capsid length in bacteriophage T4. VII. A model of length regulation based on DNA size. *J. Struct. Biol.* 104, 9–23 (1990).
54. J. D. Andreadis, L. W. Black, Substrate mutations that bypass a specific Cpn10 chaperonin requirement for protein folding. *J. Biol. Chem.* 273, 34075–34086 (1998).
55. K. Johnson, B. Condie, D. T. Mooney, A. H. Doermann, Mutations that eliminate the requirement for the vertex protein in bacteriophage T4 capsid assembly. *J. Mol. Biol.* 224, 601–611 (1992).
56. A. H. Doermann, A. Pao, P. Jackson, Genetic control of capsid length in bacteriophage T4: Clustering of ptg mutations in gene 23. *J. Virol.* 61, 2823–2827 (1987).
57. D. T. Mooney, J. Stockard, M. L. Parker, A. H. Doermann, Genetic control of capsid length in bacteriophage T4: DNA sequence analysis of petite and petite/giant mutants. *J. Virol.* 61, 2828–2834 (1987).
58. A. C. Steven, A. C. Bauer, M. E. Bisher, F. A. Robey, L. W. Black, The maturation-dependent conformational change of phage T4 capsid involves the translocation of specific epitopes between the inner and the outer capsid surfaces. *J. Struct. Biol.* 106, 221–236 (1991).
59. A. D. Dearborn *et al.*, Competing scaffolding proteins determine capsid size during mobilization of *Staphylococcus aureus* pathogenicity islands. *eLife* 6, e30822 (2017).
60. A. C. Steven, J. L. Carrascosa, Proteolytic cleavage and structural transformation: Their relationship in bacteriophage T4 capsid maturation. *J. Supramol. Struct.* 10, 1–11 (1979).
61. R. Lata *et al.*, Maturation dynamics of a viral capsid: Visualization of transitional intermediate states. *Cell* 100, 253–263 (2000).
62. K. Ray, M. Oram, J. Ma, L. W. Black, Portal control of viral prohead expansion and DNA packaging. *Virology* 391, 44–50 (2009).
63. R. van Driel, E. Couture, Assembly of bacteriophage T4 head-related strucutres. II. In vitro assembly of prehead-like structures. *J. Mol. Biol.* 123, 115–128 (1978).
64. G. Mosig *et al.*, Coordinate variation in lengths of deoxyribonucleic acid molecules and head lengths in morphological variants of bacteriophage T4. *J. Virol.* 9, 857–871 (1972).
65. J. R. Paulson, S. Lazaroff, U. K. Laemmli, Head length determination in bacteriophage T4: The role of the core protein P22. *J. Mol. Biol.* 103, 155–174 (1976).
66. D. H. Doherty, Genetic studies on capsid-length determination in bacteriophage T4. II. Genetic evidence that specific protein-protein interactions are involved. *J. Virol.* 43, 655–663 (1982).
67. K. H. Choi, M. C. Morais, D. L. Anderson, M. G. Rossmann, Determinants of bacteriophage phi29 head morphology. *Structure* 14, 1723–1727 (2006).
68. B. Keller *et al.*, Length and shape variants of the bacteriophage T4 head: Mutations in the scaffolding core genes 68 and 22. *J. Virol.* 62, 2960–2969 (1988).
69. T. Hohn, M. Wurtz, B. Hohn, Capsid transformation during packaging of bacteriophage lambda DNA. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 276, 51–61 (1976).
70. D. N. Fuller *et al.*, Measurements of single DNA molecule packaging dynamics in bacteriophage lambda reveal high forces, high motor processivity, and capsid transformations. *J. Mol. Biol.* 373, 1113–1122 (2007).
71. W. Chen *et al.*, Structural changes of a bacteriophage upon DNA packaging and maturation. *Protein Cell* 11, 374–379 (2020).
72. O. W. Bayfield, A. C. Steven, A. A. Antson, Cryo-EM structure in situ reveals a molecular switch that safeguards virus against genome loss. *eLife* 9, e55517 (2020).
73. L. Oliveira, A. O. Henriques, P. Tavares, Modulation of the viral ATPase activity by the portal protein correlates with DNA packaging efficiency. *J. Biol. Chem.* 281, 21914–21923 (2006).
74. M. Woodson *et al.*, A viral genome packaging motor transitions between cyclic and helical symmetry to translocate dsDNA. *Sci. Adv.* 7, eabc1955 (2021).
75. E. Nurmemmedov, M. Castelnovo, E. Medina, C. E. Catalano, A. Evilevitch, Challenging packaging limits and infectivity of phage λ. *J. Mol. Biol.* 415, 263–273 (2012).
76. K. Ray, J. Ma, M. Oram, J. R. Lakowicz, L. W. Black, Single-molecule and FRET fluorescence correlation spectroscopy analyses of phage DNA packaging: Colocalization of packaged phage T4 DNA ends within the capsid. *J. Mol. Biol.* 395, 1102–1113 (2010).
77. A. Zachary, L. W. Black, Isolation and characterization of a portal protein-DNA complex from dsDNA bacteriophage. *Intervirology* 33, 6–16 (1992).
78. D. E. Smith *et al.*, The bacteriophage straight phi29 portal motor can package DNA against a large internal force. *Nature* 413, 748–752 (2001).
79. D. N. Fuller, D. M. Raymer, V. I. Kottadiel, V. B. Rao, D. E. Smith, Single phage T4 DNA packaging motors exhibit large force generation, high velocity, and dynamic variability. *Proc. Natl. Acad. Sci. U.S.A.* 104, 16868–16873 (2007).
80. M. Müller, V. V. Mesyanzhinov, U. Aebi, In vitro maturation of prehead-like bacteriophage T4 polyheads: Structural changes accompanying proteolytic cleavage and lattice expansion. *J. Struct. Biol.* 112, 199–215 (1994).
81. C. E. Catalano, Bacteriophage lambda: The path from biology to theranostic agent. *Wiley Interdiscip. Rev. Nanomed. Nanobiotechnol.* 10, e1517 (2018).