# Nature Communications

https://doi.org/10.1038/s41467-026-69106-8

## Article in Press

# In situ structures of the portal-neck-tail complex of bacteriophage T4 inform a viral genome positioning mechanism

Andrei Fokine, Jingen Zhu, Thomas Klose, Frank Vago, Charles-Adrien Arnaud, Zhiqing Wang, Baldeep Khare, Michael G. Rossmann, Zhenguo Chen, Lei Sun, Qianglin Fang, Richard J. Kuhn & Venigalla B. Rao

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

# In situ structures of the portal-neck-tail complex of bacteriophage T4 inform a viral genome positioning mechanism

Andrei Fokine[1,2*], Jingen Zhu[2], Thomas Klose[1], Frank Vago[1], Charles-Adrien Arnaud[1], Zhiqing Wang[1], Baldeep Khare[1], Michael G. Rossmann[1], Zhenguo Chen[3], Lei Sun[3], Qianglin Fang[3], Richard J. Kuhn[1], Venigalla B. Rao[2*]

[1]Department of Biological Sciences, Purdue University, West Lafayette, IN 47907, USA
[2]Bacteriophage Medical Research Center, Department of Biology, The Catholic University of America, Washington, DC 20064, USA
[3]Shanghai Fifth People's Hospital and Institutes of Biomedical Sciences, Fudan University, Shanghai, 200032, China
[4]School of Public Health (Shenzhen), Sun Yat-sen University, Shenzhen, Guangdong, 518107, China.

[*]**Corresponding authors:** Venigalla B. Rao (rao@cua.edu); Andrei Fokine (afokine@purdue.edu).

## Abstract

The post-genome packaging mechanisms that govern the assembly of an infectious virion are poorly understood in bacteriophages and viruses. Here, our near-atomic resolution cryo-EM structural analyses uncovered an assembly- and conformation-driven genome positioning mechanism in the tailed bacteriophage T4. We show that following headful packaging, which generates a pressurized head, a global conformational change occurs in the portal structure, probably triggering packaging termination and ejection of the packaging motor. Our high-resolution structures of the neck of the virion further show that the neck undergoes conformational changes upon docking of a pre-assembled tail onto the sealed neck, which then opens a genome-gate. Driven by the pressure of the packaged DNA, the genome travels through open neck channels, binds and compresses the resident tape-measure protein, and halts at the bottom of the second topmost disk of the tail tube. Pressure-suspended within the virion's innermost tunnel and secured by a baseplate plug, the genome remains poised to flow through the tunnel into a host cell upon receiving the host receptor recognition signal.

## Introduction

After encapsidating the genome into a proteinaceous capsid to near crystalline density[1], icosahedral bacteriophages (phages) and viruses must safely deliver the entire genetic material into a new host cell. Loss of even a small segment of genome could be detrimental and result in a noninfectious virion. We recently discovered a genome retention mechanism in the tailed myophage T4 that seals off the pressurized capsid with a "neck" assembly attached to the unique dodecameric portal vertex[2]. Our high-resolution structures revealed a closed "genome-gate" formed by two sets of "stoppers", one from the hexameric neck protein gp14, and another from a nucleic acid-binding *E. coli* host protein Hfq, which is also a hexamer[2].

A pre-assembled tail would then bind to this neck connector followed by attachment of long tail fibers to the baseplate to complete the infectious virion assembly[3, 4, 5, 6]. What, if anything, happens to the genome during this process and up until delivery into a new host cell remains unknown. It was generally believed that the genome seal might remain in place until delivery, perhaps reinforced by the attachment of the tail to the neck. The seal would then be broken upon receiving a signal from a baseplate conformational change and tail sheath contraction upon host receptor recognition by the tail fibers[7, 8, 9]. To investigate this mechanism, we have determined the cryo-electron microscopy (cryo-EM) structures *in situ* of the infectious phage T4 virion.

Here, we describe the high-resolution asymmetric reconstruction of the portal-neck-tail connector complex, and reconstructions of specific segments of the phage tail. The cryo-EM maps allowed near atomic resolution of many of the virion components *in situ*. These include: the dodecameric portal protein gp20; the neck proteins gp13 (dodecamer) and gp14 (hexamer); the hexameric tail terminator protein gp15; the hexameric tail tube terminator protein gp3; the

tail terminator-tail sheath (gp18) junction, the tube terminator-tail tube (gp19) junction; segments of the tape-measure protein (TMP) gp29; and the TMP-baseplate hub protein (gp27) junction. Importantly, besides the atomic descriptions, these reconstructions revealed a series of structural and conformational transitions that lead to an entirely unanticipated genome positioning mechanism.

First, the *in situ* structures showed a global structural transition in the dodecameric portal assembly of the DNA-full mature virion when compared to the empty tail-less capsid[10] or the *in vitro*-assembled portal-neck complexes. The portal dodecamer changed from its flying saucer-shaped structure in the empty capsid to a mushroom-shaped structure in the DNA-full capsid of the mature virion. Moreover, the entire portal dodecamer moves "down" by ~10Å with respect to the capsid shell, presumably caused by the ~25-35 atm internal pressure due to tight "headful" packing of the highly negatively charged genome[11, 12, 13, 14, 15, 16, 17]. This transition most probably triggers termination of packaging, ejection of packaging motor, and exposure of binding sites for the neck assembly.

Second, major conformational transitions occur in the neck assembly. A domain in gp13 swings upwards embracing the portal while the gp14 loops that form the genome-gate swing downwards interacting with the tail. Strikingly, unlike in the *in vitro*-assembled portal-neck intermediate where the neck channel is closed by a double genome-gate, in the mature virion, the channel is completely open. Moreover, the genome end descended further down into the tail tube and the tape-measure protein (TMP) gp29 is compressed. Our reconstructions show extended coiled-coil segments of TMP as a narrow "tube" within the ~40Å-diameter gp19 tail tube. The TMP spans the tail tube all the way down to the baseplate hub where it is bound by the pseudo-six-fold-symmetric gp27 trimer of the cell-puncturing device, the gp27-gp5-gp5.4 complex.

These structural analyses lead to a conformation-driven genome positioning mechanism in which the pressurized genome in the finished virion is no longer restrained by the neck assembly's genome-gate. Attachment of the pre-assembled tail to the neck triggers dramatic conformational changes in the neck structure causing expulsion of the Hfq stopper and opening of the gp14 genome-gate. The DNA pressure then pushes the genome through the neck-gp15-gp3 connector channels where it is captured by the resident TMP's N-terminal DNA-binding tip. The DNA-TMP complex is then further pushed down by the pressure into the gp19 tail tube, compressing the TMP tube. The genome thus is pressure-suspended in the innermost tunnel of the finished virion, extending from the head into the portal, the neck, and the tail tube, showing no discernable interactions with any of the components of this long tunnel. In total, the DNA travels ~170 Å after tail attachment, starting from the genome-gate down to precisely the second topmost disk of the tail tube, where it remains poised to flow into a new host cell upon receiving the host cell receptor recognition signal.

## Results

### Cryo-EM of the phage T4 virion

Wild-type (WT) phage T4 virions twice-purified by CsCl gradient centrifugation were frozen onto cryo-EM grids and examined with a Titan Krios electron microscope equipped with a K3 Summit detector (Supplementary Fig. 1a). Cryo-EM data were collected comprising 12,176 micrograph movies and three-dimensional reconstructions focused on different parts of the T4 virion were calculated (Table 1 and Supplementary Fig. 2). These include the asymmetric and 6-fold symmetric reconstructions focused on the phage neck which have resolutions of 3.8Å and 3.3Å, respectively. The reconstructions are comprised of the portal, the neck proteins, the tail

sheath terminator, the tail tube terminator, and the respective interfaces at the top region of the tail. In addition, we report the 3-fold and 6-fold-symmetric reconstructions of the middle part of the tail which had resolutions of 3.4Å and 3.3Å, respectively; and 3-fold and 6-fold-symmetric reconstructions focused on the phage baseplate with resolutions of 3.6Å and 3.4Å, respectively. The maps showed continuous density for the polypeptide chains and recognizable side-chain densities (Supplementary Fig. 1). The atomic structures were built and refined against the cryo-EM maps (Table 1). The structures were then combined to generate the whole portal-neck-tail complex (Supplementary Fig. 2d).

**A global structural transition in the portal assembly**

Our *in situ* asymmetric structure of the portal assembly in T4 virion consists of twelve molecules of gp20 (525 residues; 61 kDa) arranged in a mushroom-like structure with a central channel (Fig. 1). Each subunit consists of the four classic domains found in our previously reported T4 portal structures[10, 18] and in many other phage/viral portals[19]: the clip domain, the stem domain, the wing domain, and the crown domain (Fig. 1a).

The central channel has a diameter of ~40Å in which resides a single thread of DNA double helical genome that has a diameter of ~20Å. The DNA is freely suspended in the channel showing no significant interactions with the portal lumen (Fig. 1b and Supplementary Fig. 3) including, notably, where the tunnel loops (residues 378-396) project into the channel. This suggests that the portal channel morphed into an "open" state in the infectious virion. Significantly, previous genetic and biochemical studies [11, 19, 20, 21, 22] indicate that the tunnel loops bind to the last packaged DNA and stabilize it while awaiting neck assembly (see below).

The shape of the virion portal dodecamer in the DNA-full capsid is remarkably different from that in the empty neck/tail-less capsid[10], or the *in vitro*-assembled portal-neck complexes[2] (Supplementary Fig. 2). While the DNA-empty portal structure has the shape of a flying saucer, the DNA-full virion portal has a mushroom-like shape (Fig. 1c, d). Moreover, in the DNA-full virion, the center of mass of portal dodecamer is pushed down by ~10Å, and the clip domain by ~15Å with respect to the capsid shell, presumably by the pressure of the highly condensed genome to near crystalline density (~500 mg/ml) [11, 12, 14, 23]. These differences reflect a global conformational transition involving a major structural reorganization of the portal (Supplementary Movies 1-3; Fig. 1c-h).

Morphing between the portal structures from the empty and full heads shows concerted movements of portal's structural elements that reposition and/or remodel the domains (Supplementary Movie 3). The crown and clip domains change their relative orientations with respect to the wing and stem domains (Supplementary Movie 3, Fig. 1g, h). Consequently, the channel diameter widens at the clip region from ~30Å to ~40Å, while it constricts at the crown region from ~44Å to ~38Å (Supplementary Movies 2, 3). The portal tunnel loops, which are flexible and exposed into the channel constricting its diameter in the empty heads, assume a rigid and flattened conformation in the full heads through interactions with residues from the stem and crown domains, causing the channel to be completely open. The stem helices forming the channel change angles from ~55º to ~40º relative to the central axis, while the clip domain is remodeled exposing new binding sites for the neck protein gp13 (Supplementary Movie 3).

Most probably, this structural transition is induced by the tightly packed acidic DNA ("headful packaging") that creates ~25-35 atm pressure inside the capsid [11, 13, 17, 24], which then pushes on the wing and crown domains causing the portal's displacement relative to the capsid

shell. Consequently, the DNA around the portal is stabilized showing well-resolved rings of

density in the cryo-EM map (Supplementary Fig. 3). These DNA rings would be in steric clashes

with the portal structure observed in the empty capsid, which is also consistent with the above

model that headful packaging, probably near the end of the process, induces the portal

conformational change. This portal transition then transmits the headful packaging signal from

the capsid interior to the clip domain at the capsid exterior resulting in ejection of the packaging

motor, and exposure of binding sites for assembly of the neck protein gp13. The neck assembly

then may cause additional conformational changes in the portal and the neck to further stabilize

the structure (see below).

The above analyses demonstrate a switch in the portal conformation from a symmetry-

mismatched motor-assembling platform to a symmetry-matched neck-assembling platform.


**Conformational transitions in the neck assembly**

*Gp13 dodecamer*

The 309-aa phage T4 gp13 neck protein binds to the portal protein at 1:1 ratio after the

packaging motor's departure (Fig. 2a-d). Each gp13 subunit in the virion structure comprise

three domains and an extended C-terminal region (Fig. 2e and Supplementary Figure 4), all well-

resolved and exhibiting distinct interactions with different components; portal, head, and Wac

(*w*hisker *a*ntigen *c*ontrol) fibers (Fig. 2, Supplementary Figs. 5, 6). The four-helix bundle domain

I consisting of the two first and the two last α-helices of the gp13 chain (Fig. 2e and

Supplementary Fig. 4) forms the core, having a fold similar to much smaller, usually single-

domain ortholog proteins from other phages, such as SPP1 [25, 26], HK97 [27], XM1 [28] and P22 [29].

Twelve domains I form a dodecameric ring that extends the portal tunnel (Fig. 2a, d) and its

inner wall is lined by the C-terminal negatively charged α-helices (residues 280-300). This core gp13 dodecamer, the structure and interactions of the gp14-binding loops emanating from the core (residues 266-279 and 17-31), and the C-terminal clip domain-binding extension (residues 301-309) which inserts in between the helices (Supplementary Fig. 5a), are all similar in both the virion and the *in vitro*-assembled neck complexes.

On the other hand, the "swing" domain II (residues 54-188) shows remarkable conformational changes in the virion when compared to the *in vitro*-assembled neck intermediates (Figs. 2e, f; Supplementary Movie 4; Supplementary Fig. 6). Most notable is the very different orientations this domain adopts in the two structures (Supplementary Fig. 6c, d; Supplementary Movie 4). While it forms the peripheral ring of the dodecamer with no interactions with the portal protein in the *in vitro*-assembled structure (Supplementary Fig. 6), in the virion the domain swings upwards by ~90º (Supplementary Movie 4), with one side of the β-sandwich formed by two β-sheets interacting with the portal's clip domain α-helix 295-309 and the hairpin 311-325 (Supplementary Fig. 5b). Furthermore, the long flexible and partly α-helical "anchor" region 90-157 (Fig. 2e, g) inserts into the gap between the portal and the capsid shell and interacts with the stem and wing domains of the portal and the P (periphery) domains of the gp23* major capsid protein (Fig. 2f). Notably, the side chains of cysteines 125 of gp13 are located in the proximity of cysteines 245 from the portal's wing domains potentially forming an S-S bond (Supplementary Fig. 5c). Because of the symmetry mismatch between the gp13 dodecamer and five-fold-symmetric capsid shell, the flexible anchor regions of the gp13 subunits adopt different conformations (Supplementary Fig. 4b) adjusting to different capsid environments. However, the rest of the gp13 polypeptide chain has similar conformation in all gp13 subunits and obeys strict twelve-fold symmetry of the dodecamer.

The T4 gp13 has also has a special domain III (residues 190-239) not found in other phages (Fig. 2e and Supplementary Fig. 7), which is not resolved in the *in vitro*-assembled neck structure due to its flexibility, but in the virion, it models into a binding site for interaction with the N-terminal domains of fibritin molecules, trimers of gpWac. Twelve fibritin molecules attached to gp13 dodecamer decorate the virion's neck, with six fibritins forming a propeller shaped collar and the other six fibritins oriented downwards as whiskers[30].

### *Gp14 hexamer*

The 256-residue neck protein gp14 assembles on the dodecameric gp13 as a hexamer (Fig. 3a-e). Gp14 completes the neck assembly and serves as an attachment point for the independently assembled phage tail[30, 31]. As was predicted previously[30], the core fold of the gp14 neck protein has some similarity with analogous neck proteins from other phages as well as the tail tube proteins such as T4 gp19.

While the basic networks of interactions between the gp14 hexamer and the gp13 dodecamer are nearly the same in the finished virion and the *in vitro*-assembled neck intermediate, the virion gp14 structure shows key conformational changes that are linked to attachment of the tail (Supplementary Movies 5, 6, Supplementary Fig.8).

First, most remarkable, the ~40 Å diameter central channel of gp14 ring is completely open (Fig. 3c, f) in the virion while housing the double-helical DNA. The stopper loops (residues 94-112) that close the genome-gate in the *in vitro*-assembled intermediate (Supplementary Fig. 8a) undergo a ~90° rotation downward. Furthermore, parts of these loops (residues 100-111) interact with the hexameric tail terminator protein gp15 in the assembled virion (Fig. 3g). Second, unlike in the *in vitro*-assembled intermediate neck structure, the gp14 hexamer is no

longer bound to the Hfq protein. Third, in the full virion, the gp14 structure shows an extended negatively charged C-terminal segment (212-245) containing an α-helix and a "hook" region (residues 231-245), which extensively interacts with the hexameric tail terminator protein, gp15 (Fig. 3d-h; see below). On the other hand, in the *in vitro*-assembled intermediate, this region is disordered and not resolved. These changes reflect key conformational transitions in gp14 (Supplementary Movies 5, 6) when it encounters gp15 present at the top region of the ~20 MDa pre-assembled tail, leading to Hfq ejection, opening of the genome-gate, and movement of DNA down into the tail structure (see below).

**Structures of the tail sheath-terminating gp15 hexamer and the neck-tail interface**

During tail assembly, a hexameric ring of gp15 (Fig. 4a-c) attaches to the top of the tail and terminates the tail growth[4], hence the name "tail terminator". Like gp14, gp15 has fold similarity to the phage tube proteins[30]. Although the core *in situ* cryo-EM structure of the gp15 ring is similar to our previously reported crystal structure of the recombinantly expressed gp15[30], there are important conformational changes and interactions with gp14 and the gp18 sheath.

Each gp15 molecule contains two loop regions 28-74 and 176-184, located at the top of the gp15 ring (Fig. 4d) that extensively interact with the gp14 regions 65-76 and 100-111 (part of the stopper loop) located at the bottom of the gp14 hexamer (Fig. 3h). Additionally, the helix and hook segments of the gp14 C-terminal region embrace the gp15 ring forming five salt-bridges between negatively charged D212, E220, D222, E227 and E233 of gp14, and positively charged R74, K46, R38, R11 and K160 of gp15 (Fig. 3d). These interactions repeated six times between the gp14 and gp15 hexamers generate a massive ~27,000 Å$^2$ interface and a strong neck-tail connection.

Phage T4 gp15 has an unusual 64-residue C-terminal domain (residues 209-272) which is essential for interactions with the tail sheath and probably is the key to terminate the tail growth (Fig. 4). This domain, containing a β-sheet formed with three anti-parallel β-strands, was not resolved in the previously reported crystal structure probably because of its flexibility in the absence of the sheath interactions. Each C-terminal domain from the gp15 hexamer attaches to the C-terminal domain of a gp18 subunit from the topmost ring of the contractile sheath.

Furthermore, our cryo-EM map allowed building of an atomic model for the entire gp18 protein including its C-terminal domain (Fig. 4e). In the previously resolved truncated mutant structure (gp18M)[32], this key C-terminal domain was missing. This domain (residues 548-631) contains two anti-parallel β-strands and two α-helices, which are also conserved in the sheath proteins of other *myoviridae* phages and contractile injection systems[5, 33].

The β-strands from both the gp15 and gp18 C-terminal domains interact forming a shared five-stranded β-sheet (Fig. 4e). In addition to reinforcing the tail structure[34], these interactions likely terminate sheath polymerization during the tail assembly by capping the last gp18 ring. Additionally, later in the infection mechanism, these would also stabilize the contracted sheath-neck connection when the sheath detaches from the tail tube.

Of note is that the tail terminator proteins of other *myoviridae* phages also have similar C-terminal segments which interact with the sheath, suggesting a common sheath termination mechanism[28, 35, 36, 37]. However, in other phage orthologs, the C-terminal regions are much smaller with fewer interactions when compared to T4 gp15. For example, the phage XM1 tail terminator protein has a C-terminal region containing only 13 residues and one β-strand which interacts with the sheath and augments a β-sheet in the C-terminal domain of its sheath protein [28].

Thus, it appears that during evolution T4 gp15 acquired a larger C-terminal domain to generate a stronger and more reinforced neck-sheath connector structure.

**Structures of the tail tube-terminating gp3 hexamer and its interfaces with the gp19 tube, the gp15 tail terminator, and the gp18 sheath**

Phage T4 has a special hexameric tube terminator protein, gp3, which attaches to the top of the tube during the tail assembly and terminates the polymerization of the tail tube protein gp19[34]. In the assembled tail, the gp3 ring is sandwiched between the gp19 tube (at the bottom) and the gp15 hexamer (on top) (Fig. 5a-c). The gp3 structure is very similar to that of the gp19 tube protein despite a low sequence identity, ~22%. The core of a gp3 subunit consists of a β-sandwich flanked by an α-helix (Fig. 5d). The hexameric ring of gp3 contains flexible loops on its top and bottom surfaces, which interact with gp15 and gp19, respectively (Fig. 5d-f). Each gp3 subunit interacts with three subunits of gp15. Namely, the loop region 145-150 of gp3 interacts with two gp15 subunits, whereas the loop region 117-121 and the elongated N-terminal segment of gp3 (residues 2-15) extensively interact with a third gp15 subunit (Fig. 5e).

At the bottom of the ring, gp3 contains a long loop region (residues 45-65), which attaches to three subunits of the gp19 tube, and a second loop region (residues 97-105), which interacts with two other gp19 subunits. Thus, each gp3 subunit interacts with five gp19 chains creating strong binding networks with the tube (Fig. 5f).

Additionally, the gp3 hexamer interacts with the C-terminal domains of gp18 from the topmost ring of the contractile sheath. Specifically, the β-sheet forming the outer surface of the gp3 hexamer interacts with the α-helix from the gp18 C-terminal domain, stabilizing the sheath (Fig. 5g). Such an intricate network of interactions between gp3, gp19, gp15 and gp18 terminate

the polymerization of the tail tube as well as generate a stable pre-assembled tail structure for attachment to the neck.

**The tape-measure protein anchors the terminal region of the genome inside the tail tube**

It has been well-documented through previous genetic studies that gp29 is a tape-measure protein (TMP), acting as a ruler that determines the tail length[4, 38, 39, 40, 41, 42]. During the tail assembly, one end of TMP attaches to the baseplate hub and helps the polymerization of the tube protein gp19, which assembles into a stack of hexameric rings around TMP. The tube growth continues until the end of the ruler is reached, where the hexameric tube terminator protein gp3 attaches to the last ring of the tube (and also probably to the ruler) terminating gp19 polymerization. Therefore, in the pre-assembled tail structure, the tip of the TMP ruler is expected to locate near the end (tip) of the tube and gp3, an architecture well-supported by genetic and structural studies in several *myoviridae* and *siphoviridae* phages[6, 43, 44, 45]. Surprisingly, however, our cryo-EM maps show that in the mature T4 phage virion, the TMP is no longer associated with the tip of the tube. Instead, the upper end of the TMP is now associated with the end of the genomic DNA and has moved down into the tube.

The DNA shows a continuous rod-like density inside the portal-neck-tail connector region beginning at the portal's crown, then extending down to the bottom of the second topmost ring of the gp19 tube (~320 Å) (Fig. 6). We attribute this density to the terminal segment of genomic DNA, showing no interaction with the portal or neck proteins. Clearly, the portal and neck channels are open, which is consistent with the conformational changes observed in the neck causing the genome-gate to open.

Moving down into the tail structure, reconstructions focused on the middle part of the tail and the baseplate show a stretch of density inside the gp19 tail tube which spans from the baseplate to the DNA terminus (Fig. 6a). This density is attributed to the tape-measure protein (TMP, gp29). At the top is the N-terminus of TMP in association with the DNA, the DNA-TMP junction (Fig. 6b), based on the genetic evidence[46], while the C-terminus is at the bottom. The 20-residue C-terminal region of TMP is well-resolved in our 3-fold-symmetric reconstruction focused on the baseplate. The cryo-EM map shows three C-terminal regions of TMP interacting with the trimeric (pseudo-hexameric) baseplate hub protein, gp27, located at the center of the baseplate (Supplementary Fig. 9). These baseplate hub-TMP interactions may form an initiation complex that leads to the initiation of the tail assembly[4].

In the middle of the tube, our cryo-EM reconstructions show two tube-like density segments (~150Å and ~200Å-long) which, most probably, correspond to coiled-coil regions of the TMP. These segments showed six-fold-symmetric features even when the reconstruction symmetry was released to C3. Models composed of hexameric coiled-coils show reasonable agreement with the cryo-EM densities (Supplementary Fig. 10), and are also consistent with a similar hexameric coiled-coil structure observed in a genome delivery intermediate we have recently discovered[46] and the direct mass determination by scanning electron microscopy[47].

Together, these cryo-EM reconstructions suggest that the TMP gets compressed by the descending DNA when the genome-gate opens. In total, the pressurized DNA travels ~170Å distance, first by passing through the open neck connector (gp14, gp15, and gp3) channels and then by compressing the TMP into the tail tube. This means that the tail attachment involves not mere docking of the tail tip on the sealed head, but by dynamic conformational transitions that pressure-suspend the DNA-TMP complex in the innermost tunnel of the virion in a metastable

state. The genome, thus, remains poised for rapid and smooth release into a new host cell during infection[46].

## Discussion

In this study, a series of focused cryo-EM reconstructions of portal, neck, and tail of the T4 virion generated several new high-resolution structures of the phage infection machinery, while revealing dramatic conformational transitions, together leading to an entirely unanticipated genome positioning mechanism in T4 and possibly other phages (Fig. 7).

First, we observed a global conformational transition in the portal structure (Supplementary Movies 1-3; Fig. 1). It appears that the highly compacted acidic viral genome and the associated internal pressure, estimated to be as high as 25-35 atm[11, 13, 15, 17, 24], pushes the portal down by ~10Å and causes restructuring of the portal dodecamer while exposing new binding sites for the neck proteins (Figs. 2, 5a). Not surprisingly, this occurs at a precise point when the head becomes full, and when the portal must switch from genome packaging to neck/tail assembly[48, 49]. Consequently, genome packaging is terminated, and the packaging motor is ejected from the clip domains (Supplementary Movie 7, Fig. 7). Previous genetic and biochemical studies indicate that the portal channel is probably closed transiently during this event and the DNA is restrained by the tunnel loops. Deletion of parts of the tunnel loops resulted in leakage of the packaged genome generating noninfectious virus particles[20, 21, 22]. But since the portal channel is open in the T4 virion, the neck attachment likely induces additional conformational changes in the portal, opening the channel. There is structural evidence that the portal channel exists in "open" and "closed" states[19, 50].

Second, dramatic conformational changes occur when a pre-assembled neck, the gp13 (dodecamer)-gp14 (hexamer)-Hfq (hexamer) complex attaches to the now well-exposed and remodeled portal clip domains to seal off the pressurized head (Supplementary Movie 7, Fig. 7). Binding and insertion of the C-terminal gp13 arm between the helices of adjacent clip domains causes the swing domain to flip upwards by 90º, allowing the anchor region to embrace the portal (Fig. 2e-g) and form a network of interactions with the stem and wing domains, and also with the P domains of the major capsid protein gp23*. Furthermore, domain III of gp13 remodels into a fibritin binding structure, which, by capturing the trimeric fibritin molecules creates the collar and whisker decorations around the neck. These networks of interactions further reinforce the head-portal-neck connections. The portal channel opens and the genome is stopped near the bottom of the neck where the channel diameter is constricted to ~10Å by the closed double genome-gate. The latter is a reinforced gate structure consisting of two stoppers, one formed by six symmetrically arranged loops of gp14 projecting into the channel lumen and the second by the Hfq hexamer attached to the bottom of the gp14 ring[2] (Supplementary Movie 7).

A third series of conformational transitions occur when a pre-assembled tail docks on to the bottom of this newly assembled neck (Supplementary Movies 5-7, Fig. 7). The unstructured and negatively charged C-terminal region of gp14 exposed at the bottom now locks into the positively charged side surface of the gp15 ring exposed at the top of the tail (Fig. 3g, h). Thirty-salt-bridges (five per subunit) form and a remarkably large ~27,000 $Å^2$ gp14-gp15 neck-tail interface is created. We posit that the tail might approach the neck from a side allowing the gp14-gp15 salt-bridges to form, in turn inducing similar interactions in the adjacent subunits around the perimeter of the neck, which then lead to expulsion of Hfq hexamer from the neck. The gp14 stopper loops remodel by rotating ~90º downward and interacting with gp15 (Supplementary

Movies 6, 7). Consequently, both the gp14 and Hfq stoppers disappear, and the genome-gate is now completely open allowing the DNA to descend further down through the open channels to the tip of the tail tube.

A final structural transition awaits while the neck-tail junction undergoes conformational changes. Our structures show that the C-terminal region of the TMP is bound to the baseplate hub protein gp27 (Supplementary Fig. 9c), whereas the N-terminal tip interacts with the genomic DNA. Previous studies demonstrate that the growth of the tail tube and the tail sheath continues until it reaches the end of the TMP ruler at which point further growth is terminated[4, 38, 39, 40]. Consistently, truncated versions of the ruler produced correspondingly shorter tails [38, 41, 42, 47]. When the genome-gate opens and the DNA drops through the gp14, gp15, and gp3 channels, it would be captured by the N-terminal region of the TMP residing at the tip of the tube. In Shao et al.[46], we show that a cluster of positively charged lysines and arginines in this region form a hopper-like structure bound to the DNA terminus. These are essential interactions because mutation of positively charged residues to alanines resulted in a lethal phenotype.

At this point, the newly formed DNA-TMP junction is expected to be inside the gp3 channel. However, our structures from independent reconstructions show that in the virion, this junction has moved further down and stopped precisely at the bottom of the second ring from the top of the gp19 tube (Fig. 6). This means that the internal pressure in the full capsid pushed the DNA further down, by ~100 Å, compressing the otherwise stretched TMP. Since the bottom of the tail is sealed by the baseplate proteins and the TMP is attached to the baseplate hub, the compressed TMP bound to the tip of the genome is now pressure-suspended in the innermost core of the tube. In total, the DNA travels ~170 Å down from the point where it was previously stopped by the genome-gate (Supplementary Movie 7).

Together, the above post-packaging conformational transitions constitute an intricate and dynamic pressure-suspended genome positioning mechanism triggered by neck and tail assembly. The genomic DNA in the virion remains in a "spring-loaded" metastable state, poised for rapid and efficient delivery during infection. What appears to hold the ~25-35 atm DNA pressure in this state is the baseplate hub to which the TMP is attached. We know from previous studies[4, 7, 51] that the baseplate structure is well-designed to unseal this plug when the long and short tail fibers attached to the baseplate send a receptor recognition signal upon irreversible attachment to the host cell. Then, powered by the pressure, the DNA attached to the TMP is smoothly piloted through the tail tube tunnel, and then through a new *E. coli* channel formed by the extruded and remodeled tube/hub/TMP proteins into the cytosol of the bacterial cell[46].

To our knowledge, no precedent exists for the genome positioning mechanism uncovered here for phage T4. The basic mechanism described here, both to contain genome pressure and to utilize it for genome positioning and delivery, is an innovation likely evolved by other tailed phages, although the details might vary. In fact, a recent high-resolution structure describes the insertion of a "tail completion protein" (TCP) in the innermost tunnel of the tailed phage 80α and potentially other *siphoviridae* phages[43]. The TCP forms a cork-like structure by linking the genome end to the TMP, seemingly to contain the DNA pressure and to enable efficient genome delivery.

## Methods

### Cryo-EM sample preparation and data collection

WT T4 phage particles were prepared from *E. coli* infected lysates using standard protocols [52, 53]. To produce particles with high purity, the phage was subjected to two CsCl

gradient centrifugations. The final phage preparation was flash-frozen and stored in PBS at -80ºC.

The cryo-EM grids were prepared using the Vitrobot Mark IV (Thermo Fisher Scientific) operated at 4°C with 100% humidity. Aliquots (3.5 μl) of the purified phage sample were applied onto glow discharged ultra-thin lacey carbon 400 mesh grids (Ted Pella catalog No: 01824) that had been glow-discharged using a PELCO easiGlow system for 60 sec at 25 mA. Excess sample was blotted away after a 15 sec wait time using a blotting force of 2 and a blotting time of 5 sec. The grids were plunge-frozen in liquid ethane and stored in liquid nitrogen until data collection.

Cryo-EM data were collected at the Purdue University Cryo-EM Facility using a 300 kV Titan Krios G4 cryo-electron microscope equipped with a Gatan BioQuantum K3 detector (Thermo Fisher Scientific). A total of 12,176 micrograph movies (each composed of 60 frames) were recorded using the EPU software with a nominal magnification of 42k, corresponding to a pixel size of 1.026 Å, and a defocus range of -0.6 to -2.0 μm. The total electron dose was 38 e/Å$^2$. The motion correction and calculation of the dose weighted micrographs were performed using CryoSPARC [54]. The CTF parameters were estimated using the ctffind4 program [55].

**3D reconstructions**

Cryo-EM reconstructions were calculated using the Relion software package [56, 57]. First, 5,390 phage capsids were manually selected from the micrographs and the capsid particles (1280 pixels) were extracted and subjected to 2D classification. The class averages were used for the Relion automatic picking procedure. A total of 540,975 boxes, produced by auto-picking, were subjected to 2D classification, and 195,670 capsid particles, belonging to good classes, were selected for calculation of the 3D capsid reconstruction. To speed up calculations, the particles

were rescaled to the pixel size of 3.648 Å (360 pixels). The 3D reconstruction was calculated

using the Relion auto-refine procedure, using the previously reported cryo-EM structure of the

capsid [58] filtered to 40 Å resolution as an initial reference map. The D5 symmetry was applied

during the reconstruction process. The reconstruction had a resolution of 7.3 Å.

Then the relion_particle_symmetry_expand program was used to expand each orientation

entry in the data file generated by capsid reconstruction procedure into ten D5 related entries in a

new STAR file. The new file was used to extract particles moving the box centers 175 pixels

(638.4 Å) along the Z-axis of the map (from the center of the capsid to the neck). Initially 256 x

256 boxes with the pixel size of 1.603 Å were extracted, and an initial reference was generated

using the relion_reconstruct program using the particle's orientations obtained in the previous

reconstruction. The extracted boxes were subjected to 3D classification imposing the C6

symmetry without alignment, and 164453 good neck-centered particles were selected. For those

particles, larger boxes of 400 x 400 pixels with the pixel size of 1.616 Å were extracted. The 3D

reconstruction was performed using the Relion auto-refine procedure imposing the C6 symmetry.

The map had a resolution of 3.3 Å based on the "gold-standard" Fourier shell correlation (FSC)

criterion using the 0.143 cutoff [59]. This map included the portal vertex of the capsid with

surrounding capsomers, the neck proteins and the top part of the tail. In this map, the major

capsid protein and the anchor region of gp13 were washed out because they do not follow the

imposed symmetry.

Then the box centers were moved 472 Å along the Z-axis, from the neck to the middle

region of the tail, and the new 400 x 400 boxes with the pixel size of 1.616 Å were extracted.

Particles which had the centers outside the micrographs were removed. An initial reference was

generated by the relion_reconstruct program using the particles' orientations obtained in the

previous (neck reconstruction) step. A total of 157,789 particles were used to calculate the 3D reconstruction of the middle part of the tail using the Relion auto-refine procedure. The C6 symmetry was imposed during the reconstruction process and the particles' orientations were locally refined. The final map had a resolution of 3.3 Å.

Subsequently the box centers were further moved 540 Å along the Z-axis, from the middle region of the tail to the baseplate, and new 400 x 400 boxes with the pixel size of 1.616 Å were extracted. Particles which had the centers outside the micrographs were discarded. An initial reference was generated by the relion_reconstruct program using the particles' orientations obtained in the previous reconstruction. A total of 151,499 particles were then used to produce the reconstruction of the baseplate region using the Relion auto-refine procedure. The C6 symmetry was imposed during the reconstruction and the particles' orientations were locally refined. The final map had a resolution of 3.4 Å.

Next the C3-symmetric reconstruction of the baseplate was calculated. For that, the relion_particle_symmetry_expand was used to generate a particle stack expanded with C6 symmetry. The particles were subjected to the masked 3D classification without alignment. The mask covered only the cell-puncturing device gp5-gp27, which has the 3-fold symmetry. A total of 113,841 particles were selected after the classification. The relion_reconstruct program was used to generate a new reference map. Then the reconstruction was calculated using the Relion auto-refine procedure. The particles' orientations were locally refined and the C3 symmetry was imposed during the reconstruction process. The final map had a resolution of 3.6 Å.

Subsequently, the box centers were moved back to the middle part of the tail, by -540 Å along the Z-axis to calculate the 3-fold symmetric reconstruction of the middle part of the tail. The relion_reconstruct program was used to generate a reference map. Then the reconstruction

was calculated using the Relion auto-refine procedure. The particles' orientations were locally refined and the C3 symmetry was imposed during the reconstruction process. The final map had a resolution of 3.4 Å.

Finally, the asymmetric (C1) reconstruction centered on the phage neck region was calculated. The relion_particle_symmetry_expand program was used to expand each orientation entry in the data file generated by the 6-fold symmetric neck reconstruction procedure into six C6 related entries in a new STAR file. The particles were subjected to the masked 3D classification without alignment using the C5 symmetry. The mask covered the major capsid protein capsomers (the 5-fold-symmetric part of the particle), and the previous neck-centered reconstruction was used as an initial reference map. A total of 69,107 particles were selected after the classification. The relion_reconstruct program was used to generate a new reference map. Then the asymmetric reconstruction was calculated using the Relion auto-refine procedure. The final map had a resolution of 3.8 Å.

The cryo-EM maps colored by local resolution are shown in Supplementary Fig. 11, and the Fourier shell correlation curves for the reconstructions are shown in Supplementary Fig. 12. The workflows for the reconstructions are presented in Supplementary Figs. 13 and 14. The angular distribution of the particles used in the reconstructions are shown in Supplementary Fig. 15.

**Model building and refinement**

The atomic structures were built using Coot [60]. To build the portal protein and surrounding major capsid protein capsomers, the previously reported structures from the empty T4 capsid were used as a starting model [10]. The gp13 and gp14 neck protein structures were built

*de novo*. For the gp15 tail terminator, the previously reported crystal structure [30] was used as a staring model. The gp3 tube terminator structure was built *de novo*. The C-terminal domain of the sheath protein gp18 was built *de novo*. For the rest of the gp18 structure, the crystal structure of the gp18C mutant was used as a starting model [32]. For the fibritin N-terminal regions, the crystal structure of the N-terminal domain [61] was used as a model. For the tail tube and the baseplate, the previously reported cryo-EM structure of the baseplate-tube complex was used as a model [8]. The atomic structures were refined in real space using Phenix [62, 63] against the asymmetric reconstruction centered on the neck and the 3-fold symmetric reconstructions of the baseplate and the middle part of the tail. The refinement statistics are summarized in Table 1.

The overlapping regions of the reconstructions were aligned, and the reconstructions were combined using ChimeraX [64, 65] to generate the map of the entire portal-neck-tail complex. The atomic structures were fitted into the combined map to create the combined model. The figures and movies were generated using ChimeraX [64, 65].

## Data availability

The C1 and C6 symmetric cryo-EM reconstructions focused on the phage neck region have been deposited in the EM Data Bank with the accession codes EMD-48458 [https://www.ebi.ac.uk/emdb/EMD-48458] and EMD-48462 [https://www.ebi.ac.uk/emdb/EMD-48462], respectively. The C3 and C6 symmetric reconstructions of the middle part of the tail have been deposited with the accession codes EMD-48460 [https://www.ebi.ac.uk/emdb/EMD-48460] and EMD-48463 [https://www.ebi.ac.uk/emdb/EMD-48463], respectively; and the C3 and C6 symmetric reconstructions of the distal part of the tail have been deposited with the accession codes EMD-48459 [https://www.ebi.ac.uk/emdb/EMD-48459] and EMD-48464 [https://www.ebi.ac.uk/emdb/EMD-48464], respectively. The composite map of the whole portal-neck-tail complex has been deposited with the accession code EMD-48324 [https://www.ebi.ac.uk/emdb/EMD-48324]. The asymmetric atomic structure of the neck has been deposited in the Protein Data Bank (PDB) with the accession code 9MOF [http://doi.org/10.2210/pdb9MOF/pdb]. The three-fold-symmetric structures of the middle and distal parts of the tail have been deposited with the PDB accession codes 9MOH [http://doi.org/10.2210/pdb9MOH/pdb] and 9MOG [http://doi.org/10.2210/pdb9MOG/pdb], respectively. The composite structure of the whole portal-neck-tail complex has been deposited with the PDB accession code 9MKB [http://doi.org/10.2210/pdb9MKB/pdb].

# References

1.    Earnshaw WC, Casjens SR. DNA packaging by the double-stranded DNA bacteriophages. *Cell* **21**, 319-331 (1980).

2.    Han L*, et al.* Cryo-EM structures of bacteriophage T4 portal-neck connector complexes reveal a viral genome retention mechanism involving a host component. *Res Sq*, rs.3.rs-6853951, doi: 10.21203/rs.3.rs-6853951/v1 (2025).

3.    Leiman PG, Kanamaru S, Mesyanzhinov VV, Arisaka F, Rossmann MG. Structure and morphogenesis of bacteriophage T4. *Cell Mol Life Sci* **60**, 2356-2370 (2003).

4.    Leiman PG*, et al.* Morphogenesis of the T4 tail and tail fibers. *Virol J* **7**, 355 (2010).

5.    Leiman PG, Shneider MM. Contractile tail machines of bacteriophages. *Adv Exp Med Biol* **726**, 93-114 (2012).

6.    Fokine A, Rossmann MG. Molecular architecture of tailed double-stranded DNA phages. *Bacteriophage* **4**, e28281 (2014).

7.    Leiman PG, Chipman PR, Kostyuchenko VA, Mesyanzhinov VV, Rossmann MG. Three-dimensional rearrangement of proteins in the tail of bacteriophage T4 on infection of its host. *Cell* **118**, 419-429 (2004).

8.    Taylor NM*, et al.* Structure of the T4 baseplate and its function in triggering sheath contraction. *Nature* **533**, 346-352 (2016).

9.    Yap ML*, et al.* Role of bacteriophage T4 baseplate in regulating assembly and infection. *Proc Natl Acad Sci U S A* **113**, 2654-2659 (2016).

10.   Fang Q*, et al.* Structural morphing in a symmetry-mismatched viral vertex. *Nat Commun* **11**, 1713 (2020).

11.   Rao VB, Fokine A, Fang Q, Shao Q. Bacteriophage T4 Head: Structure, Assembly, and Genome Packaging. *Viruses* **15**,  (2023).

12.   Black LW, Rao VB. Structure, assembly, and DNA packaging of the bacteriophage T4 head. *Adv Virus Res* **82**, 119-153 (2012).

13.   Rao VB, Feiss M. Mechanisms of DNA Packaging by Large Double-Stranded DNA Viruses. *Annu Rev Virol* **2**, 351-378 (2015).

14.   Black LW, Thomas JA. Condensed genome structure. *Adv Exp Med Biol* **726**, 469-487 (2012).

15. Liu S*, et al.* A viral packaging motor varies its DNA rotation and step size to preserve subunit coordination as the capsid fills. *Cell* **157**, 702-713 (2014).

16. Berndsen ZT, Keller N, Smith DE. Continuous allosteric regulation of a viral packaging motor by a sensor that detects the density and conformation of packaged DNA. *Biophys J* **108**, 315-324 (2015).

17. Villanueva Valencia JR, Li D, Casjens SR, Evilevitch A. 'SAXS-osmometer' method provides measurement of DNA pressure in viral capsids and delivers an empirical equation of state. *Nucleic Acids Res* **51**, 11415-11427 (2023).

18. Sun L*, et al.* Cryo-EM structure of the bacteriophage T4 portal protein assembly at near-atomic resolution. *Nat Commun* **6**, 7548 (2015).

19. Rao VB, Fokine A, Fang Q. The remarkable viral portal vertex: structure and a plausible model for mechanism. *Curr Opin Virol* **51**, 65-73 (2021).

20. Padilla-Sanchez V*, et al.* Structure-function analysis of the DNA translocating portal of the bacteriophage T4 packaging machine. *J Mol Biol* **426**, 1019-1038 (2014).

21. Grimes S, Ma S, Gao J, Atz R, Jardine PJ. Role of φ29 connector channel loops in late-stage DNA packaging. *J Mol Biol* **410**, 50-59 (2011).

22. Zhang Z*, et al.* A promiscuous DNA packaging machine from bacteriophage T4. *PLoS Biol* **9**, e1000592 (2011).

23. Prevo B, Earnshaw WC. DNA packaging by molecular motors: from bacteriophage to human chromosomes. *Nat Rev Genet* **25**, 785-802 (2024).

24. Ivanovska I, Wuite G, Jönsson B, Evilevitch A. Internal DNA pressure modifies stability of WT phage. *Proc Natl Acad Sci U S A* **104**, 9603-9608 (2007).

25. Lhuillier S*, et al.* Structure of bacteriophage SPP1 head-to-tail connection reveals mechanism for viral DNA gating. *Proc Natl Acad Sci U S A* **106**, 8507-8512 (2009).

26. Orlov I*, et al.* CryoEM structure and assembly mechanism of a bacterial virus genome gatekeeper. *Nat Commun* **13**, 7283 (2022).

27. Cardarelli L*, et al.* The crystal structure of bacteriophage HK97 gp6: defining a large family of head-tail connector proteins. *J Mol Biol* **395**, 754-768 (2010).

28. Wang Z*, et al.* Structure of Vibrio Phage XM1, a Simple Contractile DNA Injection Machine. *Viruses* **15**,  (2023).

29. Olia AS, Prevelige PE, Johnson JE, Cingolani G. Three-dimensional structure of a viral genome-delivery portal vertex. *Nat Struct Mol Biol* **18**, 597-603 (2011).

30. Fokine A*, et al.* The molecular architecture of the bacteriophage T4 neck. *J Mol Biol* **425**, 1731-1744 (2013).

31. Akhter T, Zhao L, Kohda A, Mio K, Kanamaru S, Arisaka F. The neck of bacteriophage T4 is a ring-like structure formed by a hetero-oligomer of gp13 and gp14. *Biochim Biophys Acta* **1774**, 1036-1043 (2007).

32. Aksyuk AA*, et al.* The tail sheath structure of bacteriophage T4: a molecular machine for infecting bacteria. *EMBO J* **28**, 821-829 (2009).

33. Taylor NMI, van Raaij MJ, Leiman PG. Contractile injection systems of bacteriophages and related systems. *Mol Microbiol* **108**, 6-15 (2018).

34. Zhao L, Kanamaru S, Chaidirek C, Arisaka F. P15 and P3, the tail completion proteins of bacteriophage T4, both form hexameric rings. *J Bacteriol* **185**, 1693-1700 (2003).

35. Sonani RR*, et al.* Neck and capsid architecture of the robust Agrobacterium phage Milano. *Commun Biol* **6**, 921 (2023).

36. Li F*, et al.* High-resolution cryo-EM structure of the Pseudomonas bacteriophage E217. *Nat Commun* **14**, 4052 (2023).

37. Yang F, Wang L, Zhou J, Xiao H, Liu H. In Situ Structures of the Ultra-Long Extended and Contracted Tail of Myoviridae Phage P1. *Viruses* **15**,  (2023).

38. Abuladze NK, Gingery M, Tsai J, Eiserling FA. Tail length determination in bacteriophage T4. *Virology* **199**, 301-310 (1994).

39. Duda RL, Gingery M, Ishimoto LK, Eiserling FA. Expression of plasmid-encoded structural proteins permits engineering of bacteriophage T4 assembly. *Virology* **179**, 728-737 (1990).

40. Arisaka F, Yap ML, Kanamaru S, Rossmann MG. Molecular assembly and structure of the bacteriophage T4 tail. *Biophys Rev* **8**, 385-396 (2016).

41. Mahony J*, et al.* Functional and structural dissection of the tape measure protein of lactococcal phage TP901-1. *Sci Rep* **6**, 36667 (2016).

42. Katsura I, Hendrix RW. Length determination in bacteriophage lambda tails. *Cell* **39**, 691-698 (1984).

43. Kizziah JL, Mukherjee A, Parker LK, Dokland T. Structure of the. *bioRxiv*,  (2024).

44. Xiao H*, et al.* Structure of the siphophage neck-Tail complex suggests that conserved tail tip proteins facilitate receptor binding and tail assembly. *PLoS Biol* **21**, e3002441 (2023).

45.    Linares R, Breyton C. About bacteriophage tail terminator and tail completion proteins: structure of the proximal extremity of siphophage T5 tail. *J Virol* **99**, e0137624 (2025).

46.    S Lei Sun, Lin Han, Qiyu Mao et al. Cryo-EM structures of bacteriophage T4 portal-neck connector complexes reveal a viral genome retention mechanism involving a host component, 27 June 2025, PREPRINT (Version 1) available at Research Square [https://doi.org/10.21203/rs.3.rs-6853951/v1]

47.    Duda RL, Wall JS, Hainfeld JF, Sweet RM, Eiserling FA. Mass distribution of a probable tail-length-determining protein in bacteriophage T4. *Proc Natl Acad Sci U S A* **82**, 5550-5554 (1985).

48.    Rao VB. A virus DNA gate: zipping and unzipping the packed viral genome. *Proc Natl Acad Sci U S A* **106**, 8403-8404 (2009).

49.    Oliveira L, Tavares P, Alonso JC. Headful DNA packaging: bacteriophage SPP1 as a model system. *Virus Res* **173**, 247-259 (2013).

50.    Bayfield OW, Steven AC, Antson AA. Cryo-EM structure in situ reveals a molecular switch that safeguards virus against genome loss. *eLife* **9**, e55517 (2020).

51.    Hu B, Margolin W, Molineux IJ, Liu J. Structural remodeling of bacteriophage T4 and host membranes during infection initiation. *Proc Natl Acad Sci U S A* **112**, E4919-4928 (2015).

52.    Tao P, Mahalingam M, Rao VB. Highly Effective Soluble and Bacteriophage T4 Nanoparticle Plague Vaccines Against Yersinia pestis. *Methods Mol Biol* **1403**, 499-518 (2016).

53.    Tao P, Li Q, Shivachandra SB, Rao VB. Bacteriophage T4 as a Nanoparticle Platform to Display and Deliver Pathogen Antigens: Construction of an Effective Anthrax Vaccine. *Methods Mol Biol* **1581**, 255-267 (2017).

54.    Punjani A, Rubinstein JL, Fleet DJ, Brubaker MA. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods* **14**, 290-296 (2017).

55.    Rohou A, Grigorieff N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J Struct Biol* **192**, 216-221 (2015).

56.    Scheres SH. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J Struct Biol* **180**, 519-530 (2012).

57.    Kimanius D, Dong L, Sharov G, Nakane T, Scheres SHW. New tools for automated cryo-EM single-particle analysis in RELION-4.0. *Biochem J* **478**, 4169-4185 (2021).

58. Fokine A, Chipman PR, Leiman PG, Mesyanzhinov VV, Rao VB, Rossmann MG. Molecular architecture of the prolate head of bacteriophage T4. *Proc Natl Acad Sci U S A* **101**, 6003-6008 (2004).

59. Henderson R*, et al.* Outcome of the first electron microscopy validation task force meeting. *Structure* **20**, 205-214 (2012).

60. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501 (2010).

61. Boudko SP, Strelkov SV, Engel J, Stetefeld J. Design and crystal structure of bacteriophage T4 mini-fibritin NCCF. *J Mol Biol* **339**, 927-935 (2004).

62. Afonine PV*, et al.* Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr D Struct Biol* **74**, 531-544 (2018).

63. Liebschner D*, et al.* Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr D Struct Biol* **75**, 861-877 (2019).

64. Goddard TD*, et al.* UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci* **27**, 14-25 (2018).

65. Pettersen EF*, et al.* UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci* **30**, 70-82 (2021).

## Acknowledgements

## Author Contributions

V.B.R., M.G.R. and A.F. conceived the project. A.F., J.Z., T.K., F.V., C-A.A., Z.W., B.K. and V.B.R. performed research. A.F., J.Z., C-A.A., M.G.R., Z.C., L.S., Q.F., R.J.K. and V.B.R. analyzed data. V.B.R., R.J.K. and M.G.R. supervised the project. A.F. and V.B.R. prepared the manuscript, with additional edits from all authors.

## Competing interests

The authors declare no competing interests.

**Table 1 Cryo-EM data and refinement statistics**

| Data collection | | | | | | |
|---|---|---|---|---|---|---|
| Magnification | 42000 | | | | | |
| Voltage (kV) | 300 | | | | | |
| Electron exposure (e–/Å$^2$) | 38 | | | | | |
| Defocus range (µm) | 0.6 – 2.0 | | | | | |
| Pixel size (Å) | 1.026 | | | | | |
| Number of micrograph movies | 12176 | | | | | |
| **Single particle reconstructions** | | | | | | |
| Reconstruction Name | Neck-centered C6 | Neck-centered C1 | Middle tail region C6 | Middle tail region C3 | Baseplate C6 | Baseplate C3 |
| Map resolution (Å) using FSC threshold of 0.143 (gold standard) | 3.3 | 3.8 | 3.3 | 3.4 | 3.4 | 3.6 |
| Map resolution (Å) using FSC threshold of 0.5 | 3.6 | 4.3 | 3.6 | 3.8 | 3.8 | 4.0 |
| Number of particles used | 164453 | 69107 | 157789 | 113841 | 151499 | 113841 |
| Map size (pixels) | 400$^3$ | 400$^3$ | 400$^3$ | 400$^3$ | 400$^3$ | 400$^3$ |
| Pixel size (Å) | 1.616 | 1.616 | 1.616 | 1.616 | 1.616 | 1.616 |
| Symmetry imposed | 6-fold | None | 6-fold | 3-fold | 6-fold | 3-fold |
| Map sharpening $B$ factor (Å$^2$) | 132 | 127 | 235 | 129 | 150 | 157 |
| **Structure refinement** | | | | | | |
| Name | Asymmetric structure of the portal-neck-tail connector complex | | 3-fold symmetric structure of the middle part of the tail | | 3-fold symmetric structure of the baseplate region | |
| Number of Protein Residues | 55817 | | 67410 | | 83614 | |
| Map used for refinement | Neck-centered C1 | | Middle tail region C3 | | Baseplate C3 | |
| Model map FSC (masked) | 0.76 | | 0.87 | | 0.68 | |
| R.m.s. deviations<br> Bond lengths (Å)<br> Bond angles (°) | 0.006<br>0.689 | | 0.009<br>0.787 | | 0.005<br>0.654 | |
| Ramachandran plot<br> Favored (%)<br> Allowed (%)<br> Disallowed (%)<br>MolProbity Clashscore<br>Poor rotamers (%) | 87.32<br>11.89<br>0.86<br>15.58<br>12.89 | | 85.59<br>12.77<br>1.64<br>13.50<br>15.99 | | 89.36<br>9.46<br>1.18<br>15.22<br>10.46 | |

# Figure Legends

**Fig. 1: *In situ* structure of the dodecameric portal protein gp20 in the bacteriophage T4 virion.**

(a) One subunit of the portal protein depicting the crown, wing, stem, and clip domains in red, green, blue, and magenta, respectively. (b) Cutoff view of the portal vertex. The capsid protein gp23* is colored cyan. The DNA model inside the portal channel is shown in gold. (c-h) Comparison of the portal structures in the empty capsid [10] and in the infectious virion. (c) and (d) show side views of the portal dodecamers in the empty capsid and the virion, respectively. (e) and (f) show the top views of the portals in the empty capsid and the virion, respectively. (g) and (h) show two opposite portal subunits from the empty capsid and the virion, respectively.

See Supplementary Movies 1-3 for morphing between the portal structures. See Supplementary Figs. 1, 2 and 3.

**Fig. 2: Structure of the dodecameric neck protein gp13 and its interactions with the portal.**

The portal's crown, wing, stem and clip domains are colored red, green, blue and magenta respectively. The domains I, II and III of gp13 are colored lime, gold and grey, respectively. The capsid protein is colored cyan. (a) Side view of the portal and gp13 surfaces. (b), (c) and (d) The side, top and bottom surfaces of the gp13 dodecamer. (e) Ribbon view of one gp13 subunit. (f) Interaction of one gp13 subunit with the portal and the capsid. Only one major protein capsomer is shown for clarity. (g) One gp13 subunit interacting with the portal subunits.

See Supplementary Movie 4 and Supplementary Data Figs. 4, 5 and 6.

**Fig. 3: Structures of the T4 gp14 neck protein hexamer and its interface with the hexameric tail terminator protein gp15.**

(a) side view of the head-tail neck connector complex. The portal protein is shown in green, gp13 in gold, gp14 in red, and gp15 in blue. The sheath protein, gp18, is colored lime. The N-terminal domain of fibritin, is shown in pink. Fibritin molecules attached to the front gp13 subunits were removed for clarity. (b) Side view of the gp14 hexamer surface. (c) Top view of the gp14 hexamer with DNA in the channel. (d) Interaction of the gp14 C-terminal region (red) with the gp15 hexamer (blue). Side chains of the residues involved in the charged interactions are shown. (e) One gp14 subunit rainbow colored from the N-terminus (blue) to the C-terminus (red). (f) Two opposite subunits of the gp14 hexamer are shown with a model of DNA inside the channel. (g) One gp14 subunit (rainbow colored) on top of the two gp15 subunits colored blue and cyan. (h) One gp14 subunit (rainbow colored) sitting on top of the gp15 hexamer (blue). See Supplementary Movies 5, 6 and Supplementary Fig. 2.

**Fig. 4: Structure of tail sheath-terminator gp15 hexamer and its interactions with gp18 sheath.** The gp15 chains are rainbow colored from the N-terminus (blue) to the C-terminus (red). The gp3 protein is shown in magenta and gp18 is shown in green. (a) Side view showing the gp15 hexamer sitting at the top of the tail. (b) Side view of the gp15 hexamer. (c) Top view of the gp15 hexamer. (d) One gp15 subunit. (e) Gp15 subunit interacting with the gp18 subunit from the top ring of the sheath.

**Fig. 5: Structure of tail tube-terminator gp3 hexamer and its interfaces with the gp15, gp19, and gp18 proteins in the top region of the tail.**

(a) Cutoff view showing the neck and the top region of the tail. The gp14 surface is shown in red, the gp15 surface in blue and the gp3 surface in magenta. The gp18 sheath is colored lime and gp19 tube is colored grey. The DNA model inside the channel is colored yellow. (b) and (c) show the side and top views of the gp3 hexamer, respectively. The gp3 chains are rainbow colored from the N-terminus (blue) to the C-terminus (red). (d) Structure of one gp3 subunit in rainbow colors. (e) The gp3 subunit interacting with the gp15 hexamer. Adjacent gp15 subunits are alternately colored blue and violet. (f) The gp3 subunit sitting on top of gp19 hexamer. Adjacent gp19 subunits in the hexamer are alternately colored grey and brown. (g) Interaction of gp3 with the C-terminal domain of the sheath protein gp18 colored green.

**Fig. 6: Structure of the T4 portal-neck-tail complex.**

(a) Central section of the combined map including the entire tail, the neck, and the portal vertex. The positions of the DNA and the tape measure protein are indicated by arrows. (b) Cutoff view of the portal-neck-tail complex structure. The molecular surface of the portal is shown in green, the gp13 surface in gold, the gp14 surface in red, the gp15 surface in blue, and the gp3 surface in magenta. Part of the capsid is colored cyan and N-terminal regions of the fibritin molecules are shown in pink. The terminal part of the genomic DNA is depicted by the yellow ribbon. The gp18 tail sheath is shown in lime, and the hexameric disks of the tube protein gp19 are shown in grey and light grey. The cell puncturing needle (gp5-gp5.4) is shown in red, and the gp27 protein in the center of the baseplate is shown in dark orange. The gp48 protein is shown in dodger blue and gp54 in khaki. The rest of the baseplate is shown in magenta. Two coiled-coil regions of the

tape measure proteins are shown as cyan ribbons. The rest of the tape measure protein is depicted

as a cyan semi-transparent cylinder.

See Supplementary Movie 7 and Supplementary Figs. 1, 2 and 7.


**Fig. 7: Mechanism of genome positioning in bacteriophage T4.** After completion of headful

genome packaging, the packaging motor is ejected from the portal vertex (a). Then the pre-

assembled gp13-gp14-Hfq complex[2] (b) attaches to the portal protein (c). The attachment is

accompanied by large conformational changes of the neck protein gp13 (green), strengthening

portal-neck interactions and exposing binding sites for Wac fibers assembly (purple). (d). The

gp13-gp14-Hfq complex seals the packaged head with a double genome-gate formed by gp14

(yellow) and Hfq (blue) hexamers. Next, a pre-assembled tail docks on the neck (e) causing the

ejection of Hfq and opening of the genome-gate (f). The DNA travels through the neck connector

channels and gets captured by the TMP at the neck-tail junction (g). The DNA pressure then

pushes the DNA further down into the tail tube to the bottom of the second ring from the top

compressing the TMP (g, h). The genome is pressure-suspended in the innermost tunnel of the

infectious virion (h).

**Editorial Summary:**
Icosahedral viruses package their genomes into a proteinaceous capsid to near crystalline density.
Here, the authors report cryo-EM-based insights into the post-packaging genome positioning,
virion assembly and maturation mechanisms in the tailed bacteriophage T4.

**Peer review information:** *Nature Communications* thanks Alfred Antson who co-reviewed with
Brianna Woodbury, and the other, anonymous, reviewer(s) for their contribution to the peer
review of this work. A peer review file is available.

a — crown, 478, 507, wing, 120, 451, 396, 53, tunnel loop, 378, 259, stem, 356, 278, 331, clip, 340

b — tunnel loop

**portal in the empty capsid** (c, e, g)

**portal in the assembled virus** (d, f, h)

h — tunnel loops

**a** gp20 portal
gp13
capsid

**b** gp13

**c** top
DNA

**d** bottom
DNA

**e** gp13 subunit
125
122
138
108
anchor
domain II
309
C
157  90
177
54
β-sandwich
300
188
244
239
210
228
31
280
domain III
domain I
272

**f** anchor
domain II
domain III
gp13
domain I

**g** portal
anchor
domain II
β-sandwich
gp13
domain I

a gp15 gp3 gp18

b gp15

c DNA

d gp15 subunit

neck binding loops

56
46
70
74
4
11
203
90
180
184
176
28
128
128
142

265
272
228
209
246

C-terminal sheath binding domain

e gp15

C-terminal domain

gp18

TMP compressed

DNA pushed down
and captured by TMP

Genome-gate opens