Article

Capsid structure of bacteriophage ΦKZ provides insights into assembly and stabilization of jumbo phages

Yashan Yang[1,2,6], Qianqian Shao[1,2,6], Mingcheng Guo[1,2,6], Lin Han[3,6], Xinyue Zhao[1,2], Aohan Wang[1,2], Xiangyun Li[1,2], Bo Wang[4], Ji-An Pan[4], Zhenguo Chen[3], Andrei Fokine[5], Lei Sun[3] ✉ & Qianglin Fang[1,2] ✉

Jumbo phages are a group of tailed bacteriophages with large genomes and capsids. As a prototype of jumbo phage, ΦKZ infects *Pseudomonas aeruginosa*, a multi-drug-resistant (MDR) opportunistic pathogen leading to acute or chronic infection in immunocompromised individuals. It holds potential to be used as an antimicrobial agent and as a model for uncovering basic phage biology. Although previous low-resolution structural studies have indicated that jumbo phages may have more complicated capsid structures than smaller phages such as HK97, the detailed structures and the assembly mechanism of their capsids remain largely unknown. Here, we report a 3.5-Å-resolution cryo-EM structure of the ΦKZ capsid. The structure unveiled ten minor capsid proteins, with some decorating the outer surface of the capsid and the others forming a complex network attached to the capsid's inner surface. This network seems to play roles in driving capsid assembly and capsid stabilization. Similar mechanisms of capsid assembly and stabilization are probably employed by many other jumbo viruses.

Bacteriophages, or phages, are likely the most prevalent microorganisms on Earth[1]. While most phages have genomes smaller than 175 kb, a specific group called jumbo phages stands out due to their larger genome size (exceeding 200 kb)[2,3]. Jumbo phages can infect a diverse range of hosts (including both Gram-negative and Gram-positive bacteria)[4] and mostly undergo a lytic life cycle. In addition, some of them have been found to encode nucleus-like structures that protect their genomes from the DNA-targeting bacterial defenses[5–7]. Thus, they hold great promise in the field of phage therapy against bacterial infections[8,9].

In contrast to smaller phages like HK97, the structure and assembly of the jumbo phage capsids remain largely unknown. This is mainly due to their huge capsid sizes, which pose challenges for determining their structures to high resolution[10,11]. Previously reported low- or medium-resolution cryo-electron microscopy (cryo-EM) structures of the capsid of jumbo phages include those of phages G (6.1 Å)[12], ΦRSL1 (9.0 Å)[13], ΦRSL2 (16 Å)[14], ΦXacN1 (9.1 Å)[14], 121Q (9.0 Å)[15], N3 (9.0 Å)[15], PAU (9.0 Å)[15], PBS1 (10.0 Å)[15] and Bellamy (30.0 Å)[15]. Only the capsid structure of phage ΦKp24 has been determined to a resolution of 4.1 Å[16], which currently represents the highest known resolution for the capsid structures of jumbo phages.

Jumbo phages have much more complex capsid structures compared to smaller phages such as HK97[17]. Cryo-EM studies of jumbo phage ΦRSL1 revealed a complex head structure formed by at least five

[1]School of Public Health (Shenzhen), Shenzhen Campus of Sun Yat-sen University, Shenzhen, Guangdong, China. [2]Shenzhen Key Laboratory of Pathogenic Microbes and Biosafety, Shenzhen Campus of Sun Yat-sen University, Shenzhen, Guangdong, China. [3]Shanghai Fifth People's Hospital, Shanghai Institute of Infectious Disease and Biosecurity, Institutes of Biomedical Sciences, Fudan University, Shanghai, China. [4]The Center for Infection and Immunity Study and Molecular Cancer Research Center, School of Medicine, Shenzhen Campus of Sun Yat-sen University, Shenzhen, Guangdong, China. [5]Department of Biological Sciences, Purdue University, West Lafayette, IN, USA. [6]These authors contributed equally: Yashan Yang, Qianqian Shao, Mingcheng Guo, Lin Han. ✉e-mail: llsun@fudan.edu.cn; fangqlin3@mail.sysu.edu.cn

different proteins[13]. Many jumbo phages possess numerous decoration proteins that bind to the outer and/or inner surface of the capsid[13,15]. However, limited by the achieved resolutions of the previously reported jumbo phage capsid structures, the structures, identities and functions of these minor capsid proteins mostly remain unknown.

Phage ΦKZ, with a large double-stranded DNA genome of ~280,300 base pairs (bp) that possesses 306 open reading frames[18], is the prototype of jumbo phages[9]. Its host, *Pseudomonas aeruginosa* (*P. aeruginosa*), is a pathogen that can cause infections in various parts of the body such as the respiratory system, digestive system, urinary tract, skin, bones, and joints[19]. *P. aeruginosa* is well known for its resistance to many antibiotics, posing a significant challenge to clinical practice[20,21]. Notably, phage ΦKZ has demonstrated effectiveness against *P. aeruginosa*[8,22–24].

Previously reported low-resolution cryo-EM structures of the ΦKZ capsid (at a resolution of 18 Å) showed that the ΦKZ capsid exhibits roughly icosahedral symmetry with a diameter of ~146 nm along the five-fold axis[25,26]. The major capsid proteins (MCPs), gp120, form a surface lattice of capsomers with $T = 27$ triangulation. Underneath the capsid vertex, cryo-EM densities of minor capsid proteins were observed, but the identities of these proteins remain unknown. Inside the capsid, there is a cylindrical structure called the "inner body" around which the viral genomic DNA may be wrapped like a spool[25,27]. Previous proteomic studies showed that at least 30 proteins are present in the ΦKZ capsid[28,29]. However, how many of these proteins act as structural components in forming the capsid shell remained undetermined.

In this study, we have extended the resolution of the icosahedrally averaged cryo-EM reconstruction of the phage ΦKZ capsid to ~3.5 Å. This cryo-EM reconstruction allowed us to build the atomic model of the capsid, containing 2520 polypeptide chains. The structure revealed up to ten minor capsid proteins that are associated with the major capsid shell. Two out of these ten proteins, namely gp35 and gp244, decorate the outer surface of the capsid vertices. The remaining eight proteins, together with several minor capsid proteins with poor cryo-EM densities, form a complex network attached to the inner surface of the viral capsid. The structure shows that these minor capsid proteins play important roles in maintaining stability and potentially in driving the assembly of the capsid, which can be applicable in many other jumbo phages.

Results

Overall structure of ΦKZ capsid

Both genome-full and genome-empty virus particles were observed in the cryo-EM micrographs (Supplementary Fig. 1). As most empty viral particles were severely broken, only the intact and full particles (~15,000 particles) were selected for data processing. The block-based reconstruction strategy[10] was used to determine the capsid structure. The 3D reconstructions of both of the two viral capsid blocks achieved resolutions of ~3.5 Å (Supplementary Fig. 2a–c and Supplementary Table 1).

The ΦKZ capsid has a diameter of ~146 nm along the five-fold axis and a triangulation number of $T = 27$ ($h = 3$, $k = 3$) (Fig. 1a), which aligns with the previously reported low-resolution cryo-EM structures[25,26]. The present near-atomic-resolution cryo-EM map allowed us for the de novo building of the atomic model of the ΦKZ capsid. The atomic model contains 1620 polypeptide chains of the MCPs (gp120), 900 polypeptide chains of ten minor capsid proteins and 330 polypeptide chains of some minor capsid proteins with unknown identities.

The ΦKZ capsid is predominantly built up by a lattice of hexameric capsomers in each facet and pentameric capsomers at each vertex (Fig. 1a–c). Both kinds of capsomers are assembled from the MCPs. The outer capsid surface is decorated by the minor capsid proteins gp35 and gp244, whereas the inner surface of the capsid shell is adorned with the minor capsid proteins gp28, gp85, gp86, gp91,

gp93, gp119, gp162, gp184 and the minor capsid proteins with unknown identities (Fig. 1a–c). These minor capsid proteins play important roles in stabilizing the viral capsid and/or facilitating capsid assembly (see below).

Major capsid protein

Each icosahedral asymmetric unit of the capsid shell contains 27 copies of the MCP, which construct four hexameric capsomers (designated as hexameric capsomer 1, 2, 3 and 4), one-third of another hexameric capsomer (designated as hexameric capsomer 5), and one-fifth of a pentameric capsomer (Fig. 1c). The MCP subunits within the pentameric capsomer and hexameric capsomers are labeled as A, and a, b, ..., z, respectively (Fig. 1c). No densities for the N-terminal 163 residues of the MCP were observed in the cryo-EM map. This observation is consistent with a previous biochemical study, which showed that the MCP underwent proteolytic cleavage between residues Glu163 and Asn164 (Supplementary Fig. 3)[28].

Like the MCP of phage HK97, the MCP of ΦKZ contains an N-terminal arm, a peripheral domain (P domain), and an axial domain (A domain)[30,31] (Fig. 2a). However, it features an additional insertion domain (I domain), which can be seen as an extension of the "E-loop" of the HK97 MCP (Fig. 2a). This is the most significant difference between the MCP structures of ΦKZ and HK97. The A domain and the P domain of the ΦKZ MCP can be roughly aligned with the corresponding domains of HK97 (Supplementary Fig. 4a–c). The A domain of the ΦKZ MCP has ~50 more residues than that of HK97, resulting in a slightly larger size.

Aligning the A domains of all the MCP subunits within an icosahedral asymmetric unit shows remarkable structural differences in the I domain and minor variances in the N-terminal arm, the P domain's tip, and the C-terminal arm (Supplementary Fig. 5a). The MCP subunits within hexameric capsomer 1 (subunits a to e), proximal to the pentameric capsomer, display a more significant structure variability with each other, when compared to the subunits in other hexameric capsomers (Supplementary Fig. 5b–f). Like phage T4, this variability is presumably due to adaptation to diverse curvatures within the capsid[32,33]. The MCP subunits in the pentameric capsomer exhibit an identical domain arrangement to those present in the hexameric capsomer but the most curved conformation (Fig. 2b and Supplementary Fig. 5a). In addition, the C-terminal arm of each MCP subunit in the hexameric capsomers forms an α-helix, while it forms a lengthy loop in the pentameric capsomer (Fig. 2b and Supplementary Fig. 5a).

The protein sequences of the MCPs of phages ΦKZ and ΦKp24 share a similarity of 43% and their structures exhibit a remarkable resemblance (Fig. 2c). The major difference is observed in conformations of their C-terminal arms (Fig. 2c). In addition to the different orientations, the C-terminal arm of ΦKZ MCP in the hexameric capsomers contains an α-helix, while the entire C-terminal arm of ΦKp24 MCP has a loop-like conformation. Furthermore, the N-terminal arm of the ΦKp24 MCP was not observed in the reported atomic model, probably due to the limited resolution of the cryo-EM map[16]. The extended N-terminal arm of the ΦKZ MCP is involved in extensive inter-molecular interactions with both MCP and minor capsid proteins, critically contributing to the stability of the viral capsid (see below).

Hexameric and pentameric capsomers and intra-capsomer interactions

The A domain, P domain and N-terminal arm of the MCPs within the hexameric capsomers of ΦKZ and HK97 share similar overall arrangements. Each MCP subunit of ΦKZ interacts with four of the other five subunits (Supplementary Fig. 6a, b). All domains of the MCP are involved in intra-capsomer interactions (Supplementary Fig. 6c). However, distinct features in intra-capsomer interactions are observed in ΦKZ. Specifically, the I domain sits on top of the junction of the A domain, P domain and I-domain linker from an adjacent MCP subunit
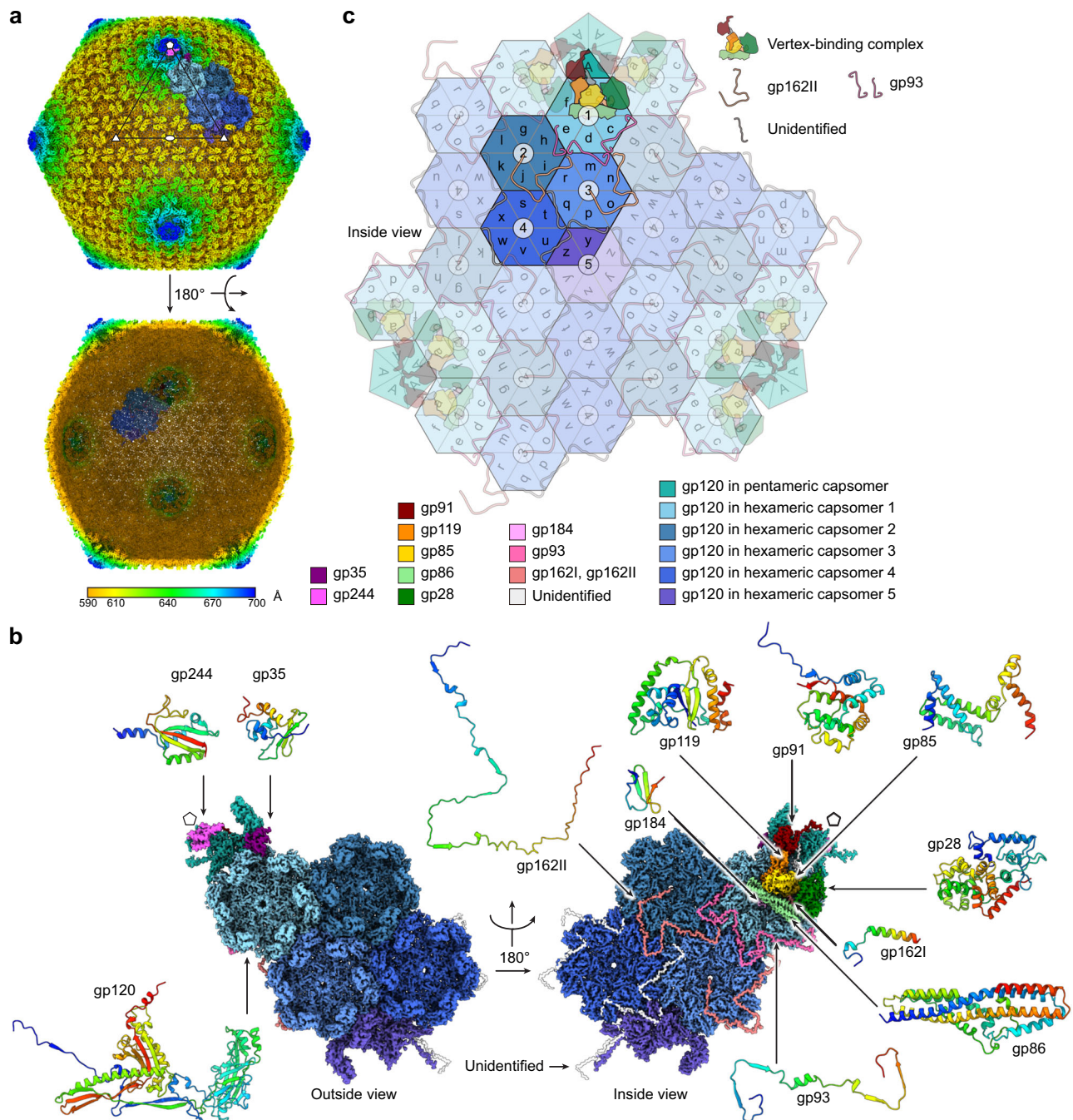
Fig. 1 | Overall Structure of the ΦKZ capsid. a Icosahedrally averaged cryo-EM reconstruction of the viral capsid, colored according to the radial distance from the center of the virus, with one asymmetry unit colored according to different protein subunits. The icosahedral five-fold, three-fold and two-fold axes are shown as white pentagon, triangle and oval, respectively. Below is the cross-section of the capsid after a 180° rotation. **b** Cryo-EM density map of an icosahedral asymmetric unit viewed from outside (left) and inside (right) the capsid. The map is colored according to different protein subunits. Rainbow-colored ribbon diagrams of representative protein structures are also shown (blue (N terminus) to red (C terminus)). **c** Diagrammatic organization of capsomers and minor capsid proteins viewed from inside the viral capsid. One asymmetric unit is highlighted, and the others are semi-transparent. The major capsid protein (MCP), gp120, and minor capsid proteins are shown as distinct shapes and colored differently, as indicated by the color keys. The hexameric capsomers within one asymmetric unit are labeled in numeric order (1, 2, 3, 4 and 5). The MCPs are labeled in alphabetic order.

(Fig. 3a). The I-domain linker also makes β-strand interactions (forming an inter-molecular β-sheet) with the N-terminal arm of an adjacent MCP subunit (Fig. 3b). In addition, the six C-terminal arms within each hexameric capsomer are assembled into a central six-α-helix bundle, forming a ~9-Å-wide pore (Fig. 3a, c) which is also observed in phage ΦKp24[16], but not in HK97[30,31]. Residues in the six-α-helix bundle engage in a combination of charge interactions and hydrophobic interactions (Fig. 3c).

The intra-capsomer interactions between the MCP subunits within the pentameric capsomer exhibit similarities to those observed in the hexameric capsomers (Supplementary Fig. 6d, e), but have two significant differences. First, the MCP subunits within the hexameric capsomer do not interact with the subunits located on the diagonal of the hexagon. In contrast, each MCP subunit of the pentameric capsomer interacts with all four other subunits (Supplementary Fig. 6b, e). Second, rather than forming a central six-α-helix bundle, the
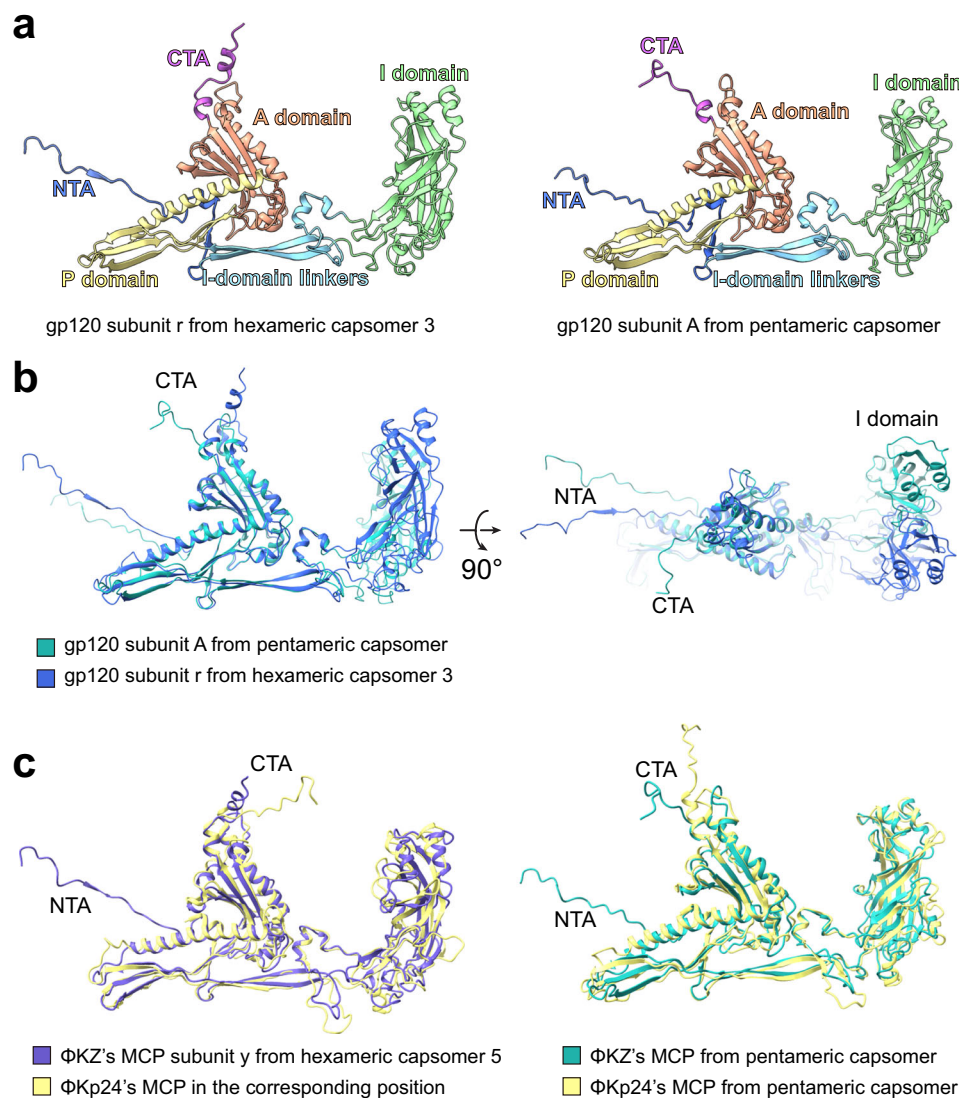
Fig. 2 | Structures of ΦKZ's MCP monomers. a Ribbon diagrams of the representative MCP subunit from the hexameric capsomer 3 (subunit r, left) and the subunit from the pentameric capsomer (subunit A, right). The N-terminal arm (blue, residues 164–204), I-domain linkers (cyan, residues 205–247 and residues 473–484), I domain (green, residues 248–472), P domain (yellow, residues 485–539 and 680–717), A domain (orange, residues 540–679 and 718–727) and C-terminal arm (magenta, residues 728–747) are labeled. **b** Structural alignment of the MCP subunit r and subunit A based on their A domains shows that the MCP from the pentameric capsomer (subunit A) adopted a more curved conformation. Additional structural alignment of all the 27 MCP subunits within one icosahedral asymmetric unit is shown in Supplementary Fig. 5a. **c** Structural alignments of ΦKZ's MCP subunits and ΦKp24's MCP subunits (PDB: 8BFL, 8BFP) (left: alignment of subunits y from hexameric capsomer 5, right: alignment of subunits A from the pentameric capsomer). NTA, N-terminal arm. CTA, C-terminal arm.

C-terminal arms of the MCPs in the pentameric capsomer shift towards the capsomer periphery, creating binding sites for the minor capsid protein gp244 (Supplementary Fig. 6f) (see below).

Inter-capsomer interactions

The inter-capsomer interactions in ΦKZ have significant differences from phage HK97. In ΦKZ, each MCP subunit in hexameric capsomer or pentameric capsomer performs inter-capsomer interactions with eight neighboring MCP subunits (Supplementary Fig. 7a), while each MCP subunit interacts with nine neighboring MCP subunits in HK97[30,31]. The ΦKZ MCP subunits, encircling the two-fold or quasi-two-fold axis, do not interact with each other (e.g., subunit e from hexameric capsomer 1 and subunit h from hexameric capsomer 2, and subunit A from pentameric capsomer and subunit a from hexameric capsomer 1) (Supplementary Fig. 7b). Instead, the two clockwise subunits beside them (e.g., subunits d and g, and subunits A and f) use

their I domains to form extensive interactions at the two-fold or quasi-two-fold axis (Fig. 3d and Supplementary Fig. 7c). The I domains of MCP subunits d and g interact with each other, forming an "I-domain dimer" at the quasi-two-fold axis (Fig. 3d). Residues Arg296, Asp366, Lys415 and Asp424 may establish charge interactions that enhance the interactions between these two I domains (Fig. 3e). In addition, this I-domain dimer further interacts with loops Lys218–Gln232 of subunits h and e and loops Pro352–Arg359 of subunits d and g (Fig. 3f).

Around the three-fold or quasi-three-fold axis, isopeptide bonds are formed between P domain and E-loop of neighboring MCP subunits to stabilize the capsid in HK97[30,31] In ΦKZ, however, only non-covalent interactions are observed around these regions (Fig. 3g). Specifically, the P domain loop Leu693–Thr703 of each MCP subunit (for example, subunit e from hexameric capsomer 1) interacts with the I-domain linker of an MCP subunit from a neighboring capsomer (for example, subunit m from hexameric capsomer 3) and two P domain
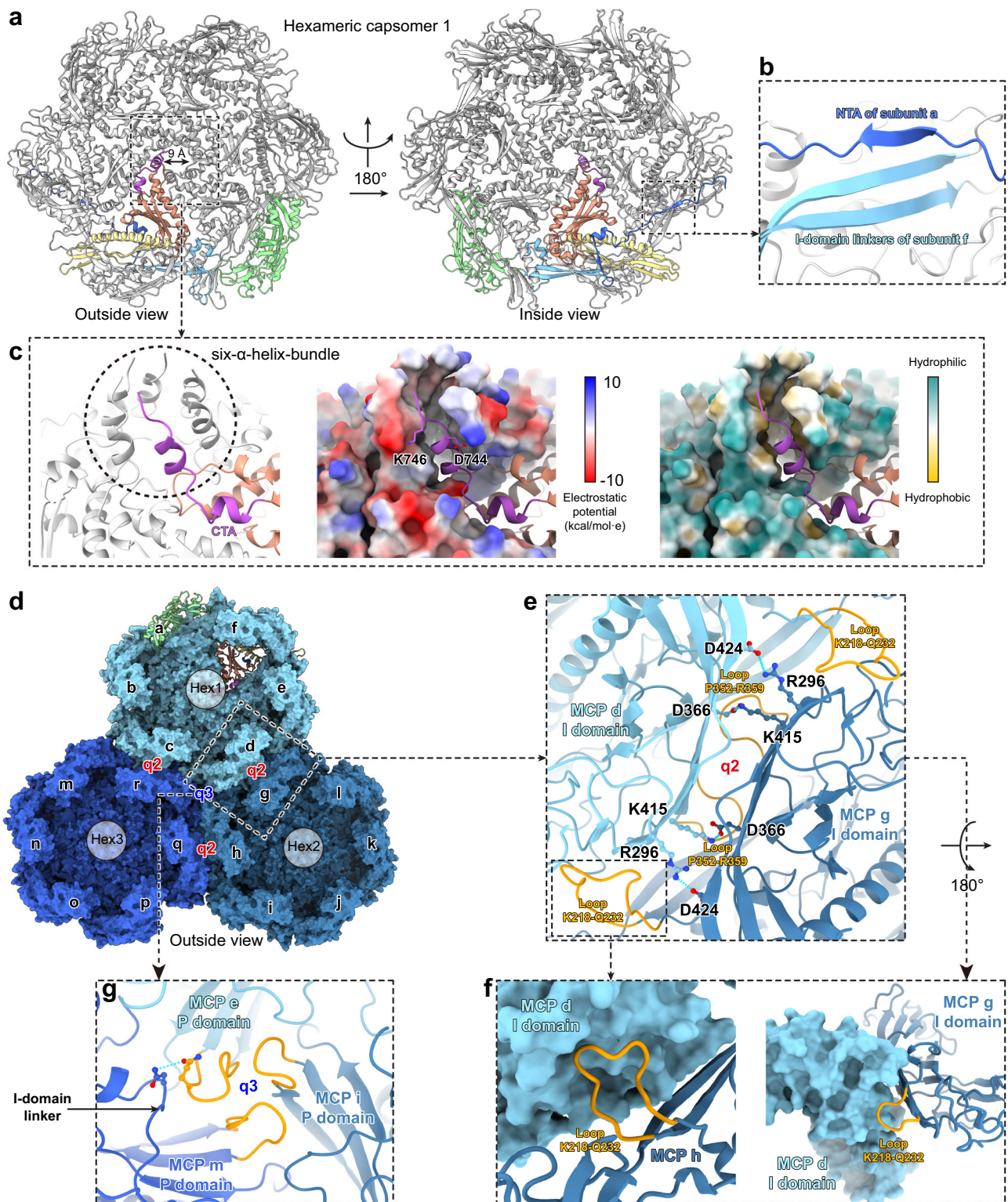
loops from two neighboring capsomers (Fig. 3g). Probably to compensate for the lack of the covalent inter-capsomer interactions in HK97, minor capsid proteins are used to enhance the associations between neighboring hexameric capsomers near the three-fold or quasi-three-fold axis of the ΦKZ capsid (see below).

Minor capsid proteins on the outer surface of the capsid vertices
The outer surface of the viral capsid is decorated with minor capsid proteins gp244 and gp35 (Fig. 4a). Gp244 is positioned on the top of

each pentameric capsomer, while gp35 is located at the quasi-three-fold axis formed by each pentameric capsomer and its two neighboring hexameric capsomers 1 (Fig. 4a).

Each set of five copies of the gp244 molecules forms a ring-like structure, capping and enhancing the stability of a pentameric capsomer (Fig. 4b). The gp244 structure consists of a core domain and an N-terminal arm. The core domain mainly consists of three α-helices, a four-stranded β-sheet and a long loop (Fig. 4c). The N-terminal arm of each gp244 subunit extends towards an adjacent

Fig. 3 | Intra- and inter-capsomer interactions. a Cartoon representation of the hexamer capsomer 1 viewed from outside (left) and inside (right) the viral capsid. One of the six MCP subunits is colored as in Fig. 2a. **b** Zoomed-in view of the boxed region in **a**, showing the β-strand interactions formed by two neighboring MCP subunits. **c** Zoomed-in view of the six-α-helix bundle. The structures are shown as ribbon diagram (left), molecular surface colored according to the electrostatic potential (middle) and molecular surface colored according to hydrophobicity (right). **d** Structure of hexameric capsomers 1, 2 and 3. Subunit a is shown as ribbon diagram while the other subunits (b to r) are shown as molecular surfaces. q2 and q3 indicate a quasi-two-fold axis and a quasi-three-fold axis, respectively. **e** Zoomed-in view of the boxed region in (**d**) showing the ribbon diagrams of the I-domain dimer

and surrounding components. Blue dashed lines indicate putative salt bridges. Loops K218–Q232 and P352–R359 are highlighted in orange. **f** The left panel is a zoomed-in view showing the interaction between the loop K218–Q232 of subunit h (highlighted in orange) and the I domain of subunit b (shown as molecular surface). The right panel shows the interaction between the loop P352–R359 of subunit g (highlighted in orange) and the I domain of subunit d. **g** Zoomed-in view of the area around the quasi-three-fold axis showing: (1) the interactions between the three adjacent P-domain loop L693–T703 (highlighted in orange) and (2) the interactions between the P-domain loop L693–T703 and the I-domain linker from a neighboring capsomer. Hex, hexameric capsomer. NTA, N-terminal arm. CTA, C-terminal arm.
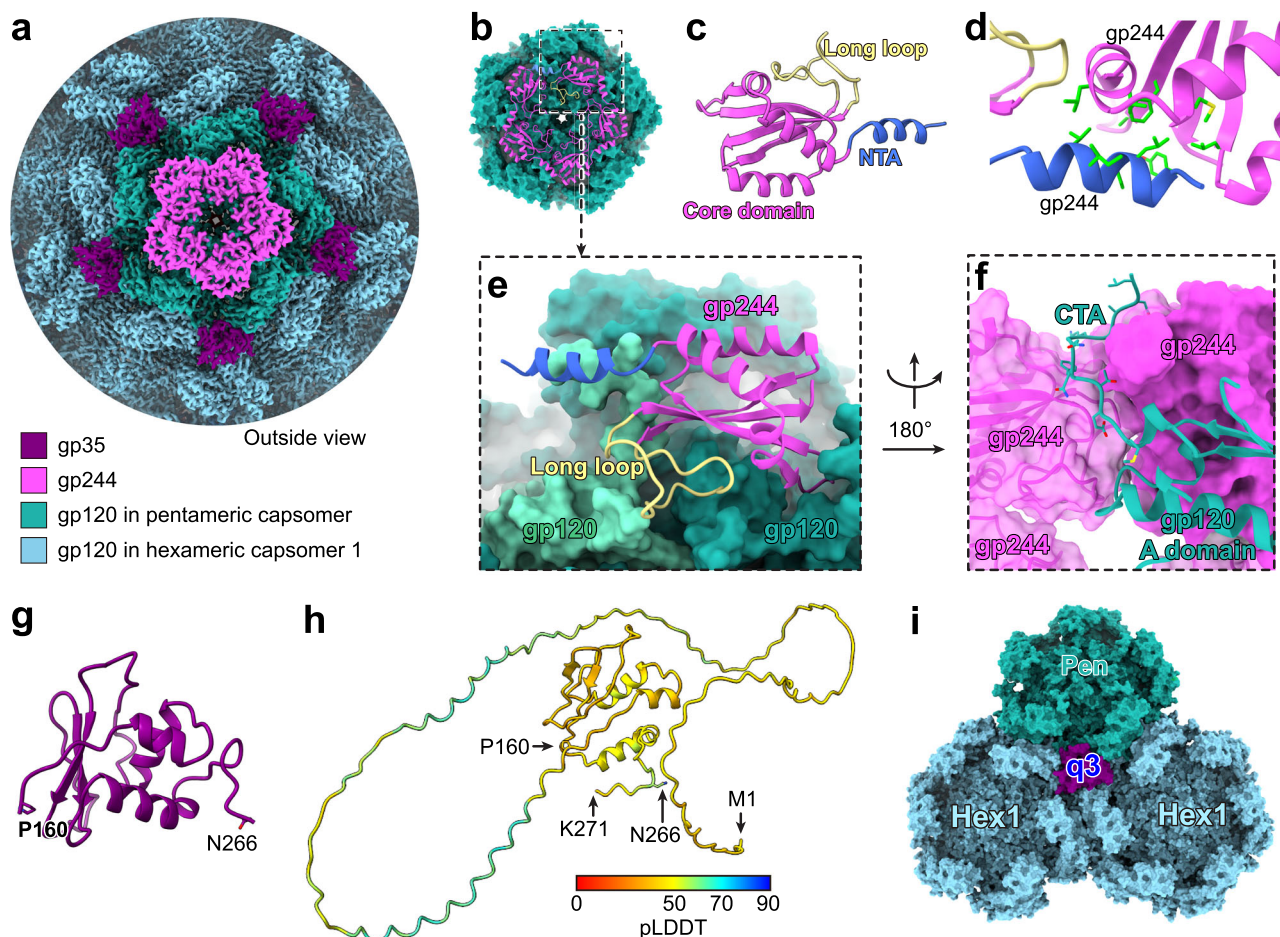

Fig. 4 | Minor capsid proteins on the outer surface of the capsid. a Cryo-EM map showing the outer minor capsid proteins of ΦKZ capsid colored according to Fig. 1. **b** Structures of the five copies of gp244 and the pentameric capsomer shown as ribbon diagrams and molecular surface, respectively. **c** Ribbon diagram of one gp244 subunit colored based on different domains. The core domain (residues 17–116), N-terminal arm (residues 3–16) and the long loop (residues 84–106) are labeled. **d** Interaction between two neighboring gp244 molecules. Sidechains of hydrophobic residues involved in the interactions are shown and colored green. **e** Zoomed-in view of the boxed region in (**d**) showing the interactions between

gp244 and its two neighboring MCP subunits. **f** The interactions between the C-terminal arm of one gp120 subunit A and its two neighboring gp244 subunits. **g** Ribbon diagram of gp35. Only the C-terminal portion of gp35 (residues 160–266) was modeled. **h** The in-silico structure of full-length gp35 predicted by AlphaFold2[34]. The ribbon diagram of the in-silico structure of gp35 is colored according to pLDDT. **i** The molecular surfaces of gp35 (purple) and three surrounding capsomers (the pentameric capsomer and two hexameric capsomer 1) show that gp35 is located at the quasi-three-fold axis (q3). Hex, hexameric capsomer. Pen, pentameric capsomer. NTA, N-terminal arm.

gp244 subunit, engaging in favorable hydrophobic interactions (Fig. 4d).

Each gp244 subunit makes contacts with two of the five MCP subunits within the pentameric capsomer (Fig. 4e). The long loop in the core domain of gp244 is positioned near the boundary between two MCP subunits of the pentameric capsomer, establishing extensive interactions with both MCP subunits (Fig. 4e). Additionally, the C-terminal arm of each MCP subunit extends to the junction between

two gp244 molecules, interacting with both gp244 molecules (Fig. 4f).

Gp35 decorates the outer surface of the viral capsid near the vertices (Fig. 4a). Only the C-terminal portion of gp35 was modeled (Fig. 4g and Supplementary Fig. 3), as the cryo-EM densities of the N-terminal portion were not observed. As predicted by AlphaFold2[34], this missing portion mainly forms a lengthy loop, which could exhibit significant flexibility (Fig. 4h). The resolved portion of gp35 is
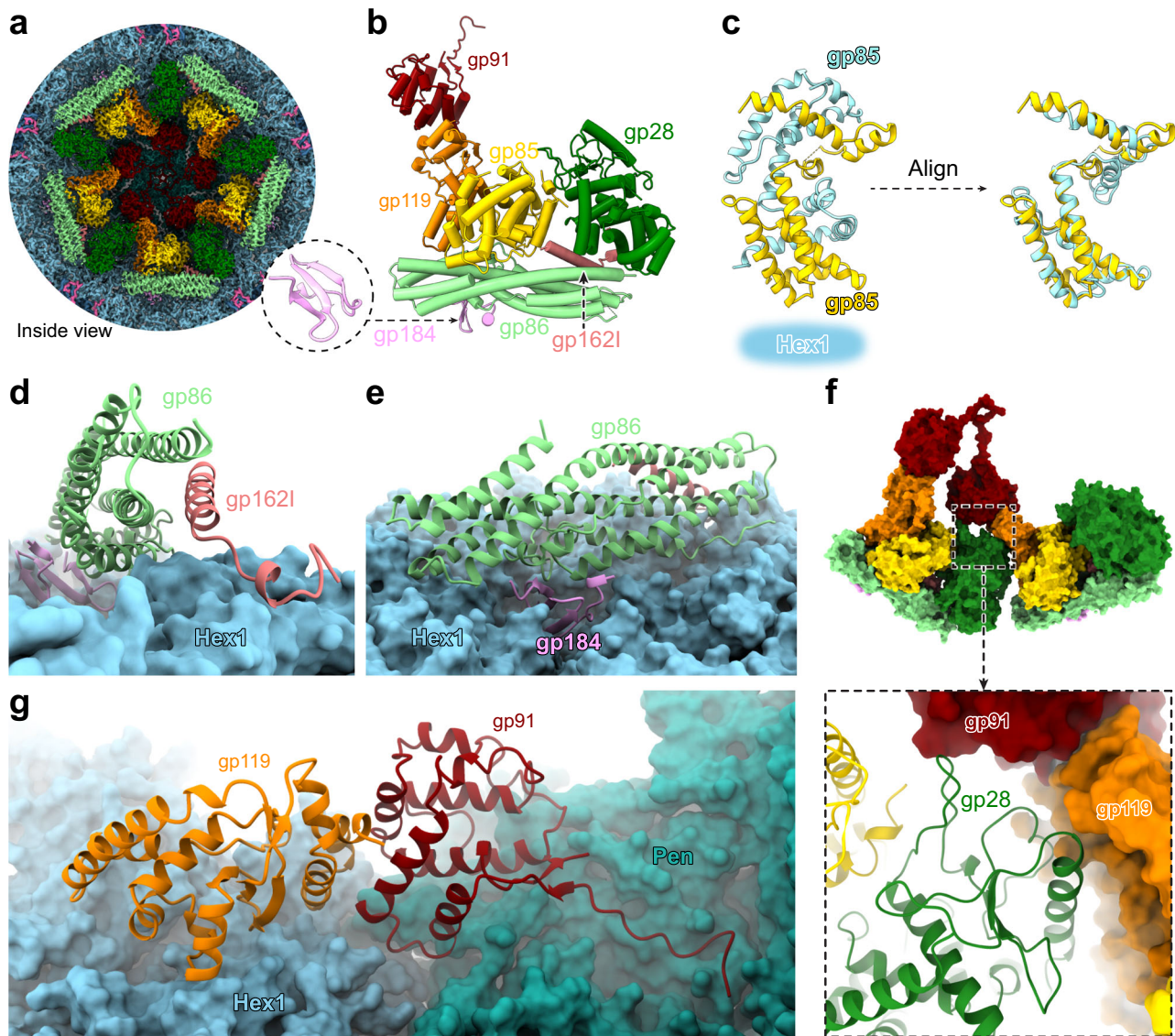
Fig. 5 | Vertex-binding complex attached to the inner surface of the capsid vertex. a Cryo-EM densities beneath the capsid vertex colored according to Fig. 1. **b** Ribbon diagram of the vertex-binding complex within one asymmetric unit. The inset shows the structure of gp184. **c** Structure of the gp85 dimer (left) and the structure alignment of the two subunits within the gp85 dimer (right). The relative position of the hexameric capsomer 1 and gp85 dimer is indicated. **d, e** Structures of gp86, gp162I, gp184 and hexameric capsomer 1 (shown as molecular surface) show the interactions among these proteins. **f** Structures of gp91, gp119, hexameric capsomer 1 and the pentameric capsomer show the interactions among these proteins. The capsomers are shown as molecular surfaces. **g** The junction between two vertex-binding complexes shows the interactions between gp28 and the minor capsid proteins (gp91 and gp119) from a neighboring vertex-binding complex. Hex, hexameric capsomer. Pen, pentameric capsomer.

positioned at the junction between each pentameric capsomer and its two adjacent hexameric capsomers 1 (Fig. 4i). This placement allows gp35 to interact with these three capsomers simultaneously, thereby reinforcing the stability of the viral capsid near the capsid vertices (Fig. 4a, i). It's worth noting that while trimeric decoration proteins at the quasi-three-fold axis are commonly observed in other phages (e.g., P74-26[35], TW1[36]), gp35 exhibits a distinct characteristic by adopting a monomeric structure.

Vertex-binding complex on the inner surface of the capsid vertices

The viral capsid has five "vertex-binding" complexes attached to the inner surface of each vertex. Each complex is composed of seven kinds of minor capsid proteins, namely gp28, gp85 (existing as homodimer), gp86, gp91, gp119, gp162 and gp184 (Fig. 1c, Fig. 5a, b). All these minor capsid proteins are rich in α-helix, except for gp184, which mainly

consists of β-strands (Fig. 5b). Since more than one fragment of gp162 is observed in the viral capsid, we designated the fragment in the vertex-binding complex as gp162I and the others as gp162II (see below). Gp86 is a product of proteolytic cleavage, as the N terminus of the modeled part is located near proteolytic cleavage sites verified by semi-tryptic peptides[28] (Supplementary Fig. 3). Gp162I and gp184 also have previously verified cleavage sites[28], but the modeled parts of gp162I and gp184 are shorter than the products of proteolytic cleavage, which are likely due to flexibility (Supplementary Fig. 3).

The center of the vertex-binding complex is occupied by a "sandwich" formed by gp28, the gp85 dimer, and gp119 (Fig. 5b). The two subunits within the gp85 dimer share nearly identical structures (Fig. 5c). The gp85 dimer interacts with almost all members in one vertex-binding complex, except for gp91 and gp184 (Fig. 5b and Supplementary Fig. 8). Gp91 extends towards the capsid vertex, while gp86, which has an elongated shape, forms the peripheral edge of the

complex on the other side (Fig. 5a, b). Gp86 is further supported by two minor capsid proteins, namely gp162I and gp184. Gp162I interacts with gp86 by joining its α-helix (residues 163-181) with the α-helix-bundle of gp86 (Fig. 5d). Gp184 inserts into the hollow between gp86 and the center of hexameric capsomer 1 (Fig. 5e). Furthermore, neighboring vertex-binding complexes are connected to each other via the interactions between gp28 from one vertex binding complex and gp91, gp119 from an adjacent vertex-binding complex, thus forming a larger complex (Fig. 5f and Supplementary Fig. 8).

The larger complex sticks to the inner surface of the capsid vertex and stabilizes the capsid vertex via interactions between all its protein components and the capsomers around the capsid vertex. Specifically, all components of the vertex-binding complex but gp91 mainly make contacts with hexameric capsomer 1. Gp91 primarily interacts with the pentameric capsomer via hydrophobic interactions and several salt bridges (Fig. 5a, g). Gp119 interacts with one MCP subunit in hexameric capsomer 1 and with the pentameric capsomer-binding protein, gp91, bridging the pentameric capsomer and hexameric capsomer 1 (Fig. 5g). Gp86 interacts with all MCP subunits from hexameric capsomer 1, except for subunit d. The gap between gp86 and MCP subunit d is filled by gp184 (Fig. 5e). Additionally, gp28 and the gp85 dimer interact with one and two of the MCP subunits in hexameric capsomer 1, respectively.

Fiber-like minor capsid proteins

The internal surface of all the hexameric capsomers of the viral capsid reveals the presence of numerous fiber-like structures forming a network that strengthens the connections between neighboring capsomers (Fig. 6a). Among this network, two minor capsid proteins, namely gp93 and gp162II, have been identified. In addition, there remain fiber-like structures near the three- and two-fold axes of the capsid whose identities have not been determined due to insufficient quality of their cryo-EM densities.

In each icosahedral asymmetric unit, there are three gp93 molecules (hereafter designated as gp93A, gp93B and gp93C), which show nearly identical structures (Fig. 6b). The first modeled N-terminal residue of all these three copies is Ser14, which aligns with one of the proteolytic cleavage sites of gp93 between Glu13 and Ser14[28] (Supplementary Fig. 3). However, less of one of the three copies of gp93 (gp93A, located between the MCP subunits e and h) was modeled compared to the other two copies (gp93B and gp93C) (Fig. 6b). The cryo-EM densities for residues 85–435 of all the three copies of gp93 are missing, probably due to flexibility and/or potential proteolytic cleavages[28].

The structures of gp93B and gp93C consist of an N-terminal arm, a helix-loop-helix motif and a C-terminal arm, exhibiting an "S-shaped" structure with roughly C2 symmetry (Fig. 6b). This symmetry likely arises due to the influence of neighboring MCPs, which create a quasi-C2-symmetrical environment at the binding site of gp93 (Fig. 6c). Each gp93 molecule mainly makes contacts with two MCP subunits from two neighboring hexameric capsomers (Fig. 6d). The N-terminal arm of gp93 forms anti-parallel β-strand interactions with the P domain and N-terminal arm of a MCP subunit from one hexameric capsomer (Fig. 6e), while the C-terminal arm of the same gp93 molecule forms parallel β-strand interactions with the N-terminal arm and P domain of a MCP subunit from the other hexameric capsomer (Fig. 6f). The helix-loop-helix motif of gp93 is deeply inserted into the gap between these two hexameric capsomers near the quasi-two-fold axis (Fig. 6d). Furthermore, the three gp93 molecules within each icosahedral asymmetric unit are connected head-to-tail via interactions between the tips of their N- and C- terminal arms (Supplementary Fig. 9). Thereby, the associations among hexameric capsomers 1, 2 and 3 are significantly reinforced.

Each icosahedral asymmetric unit has two copies of gp162II molecules (hereafter designated as gp162IIA and gp162IIB), which

show similar structures (Fig. 6g). The modeled part of each gp162II molecule contains residues 19 to 134. The gp162II corresponds exactly to the previously designated segment "gp162N"[28], which is a product of proteolytic cleavages between Glu17 and Arg18, and between Glu137 and Asp138 (Supplementary Fig. 3).

The gp162II structure can be divided into two halves, specifically, an N-terminal half and a C-terminal half (Fig. 6g). Each half adopts a conformation that resembles the "S-shaped" structure of gp93 (Fig. 6h). However, it's worth noting that the C-terminal residues 118–134 of gp162II do not entirely follow the "S shape". Instead, these residues extend towards the center of a hexameric capsomer (Fig. 6i). Despite limited sequence similarities, both halves of gp162II occupy a similar position on the hexameric capsomers as gp93. Similar to gp93, β-strand interactions were observed between gp162II and the MCPs (including the N-terminal arms and P domains of the MCPs) (Fig. 6i). This arrangement allows each gp162II molecule to associate a set of three adjacent hexameric capsomers (hexameric capsomers 2, 3 and 4) together (Fig. 6i).

Within each icosahedral asymmetry unit, there are several additional fiber-like structures that exhibit a similar "S shape" to gp93. Like gp93, each of these structures associates two neighboring hexameric capsomers (Supplementary Fig. 10a). However, the cryo-EM densities of most sidechains in these structures appear blurry, and for some regions of these structures it was even challenging to track the backbone. As a result, we were unable to determine the specific identities of these remaining fiber-like structures. One of these structures is precisely centered at the icosahedral two-fold axis (located between two hexameric capsomers 4) (Fig. 1c), whose cryo-EM density undergoes the C2 averaging during the 3D reconstruction process. The other structures closely resemble the C2-averaged one, exhibiting cross-correlation coefficients ranging from 0.91 to 0.94 (Supplementary Fig. 10b). This indicates two important points: (1) these structures likely share the same identities, and (2) this protein may bind to the capsid using two orientations that are related by either the icosahedral two-fold axis or the quasi-two-fold axis between neighboring hexameric capsomers (Supplementary Fig. 10b).

Discussion

The present capsid structure of phage ΦKZ reveals an incredibly intricate network of minor capsid proteins, consisting of numerous types of proteins and hundreds of polypeptide chains. As far as we know, there is no reported high-resolution structural information on such a complex network of minor capsid proteins in other tailed bacteriophages.

The network of minor capsid proteins appears to be essential for the assembly and stability of the ΦKZ capsid. The fiber-like minor capsid proteins (gp93, gp162II and the protein with unknown identity) extend across nearly all the boundaries between adjacent hexameric capsomers (Fig. 1a), effectively cementing these capsomers together. Proteins gp28, gp85, gp86, gp91, gp119, gp162I and gp184 form vertex binding complexes that attach to the inner surface of the capsid vertex. These complexes serve as a bridge between the pentameric capsomer and the five hexameric capsomers encircling the pentameric vertex (Fig. 1c). Additional stability to the capsid vertex is provided by the external minor capsid proteins gp35 and gp244 (Fig. 4a).

This arrangement of the minor capsid proteins is reminiscent of the minor capsid protein networks observed in other large dsDNA viruses, such as Paramecium bursaria chlorella virus 1 (PBCV-1)[37,38]. Like phage ΦKZ, PBCV-1 has a very large capsid (diameter: -190 nm) and houses up to eighteen different minor capsid proteins on the inner surface of its viral capsid. The majority of these minor capsid proteins are situated at the boundaries between neighboring capsomers. The tape-measure protein of PBCV-1, akin to gp93 and gp162II in phage ΦKZ, forms a fiber-like structure, and is thought to determine the size of the viral capsid. However, such complex networks of minor capsid
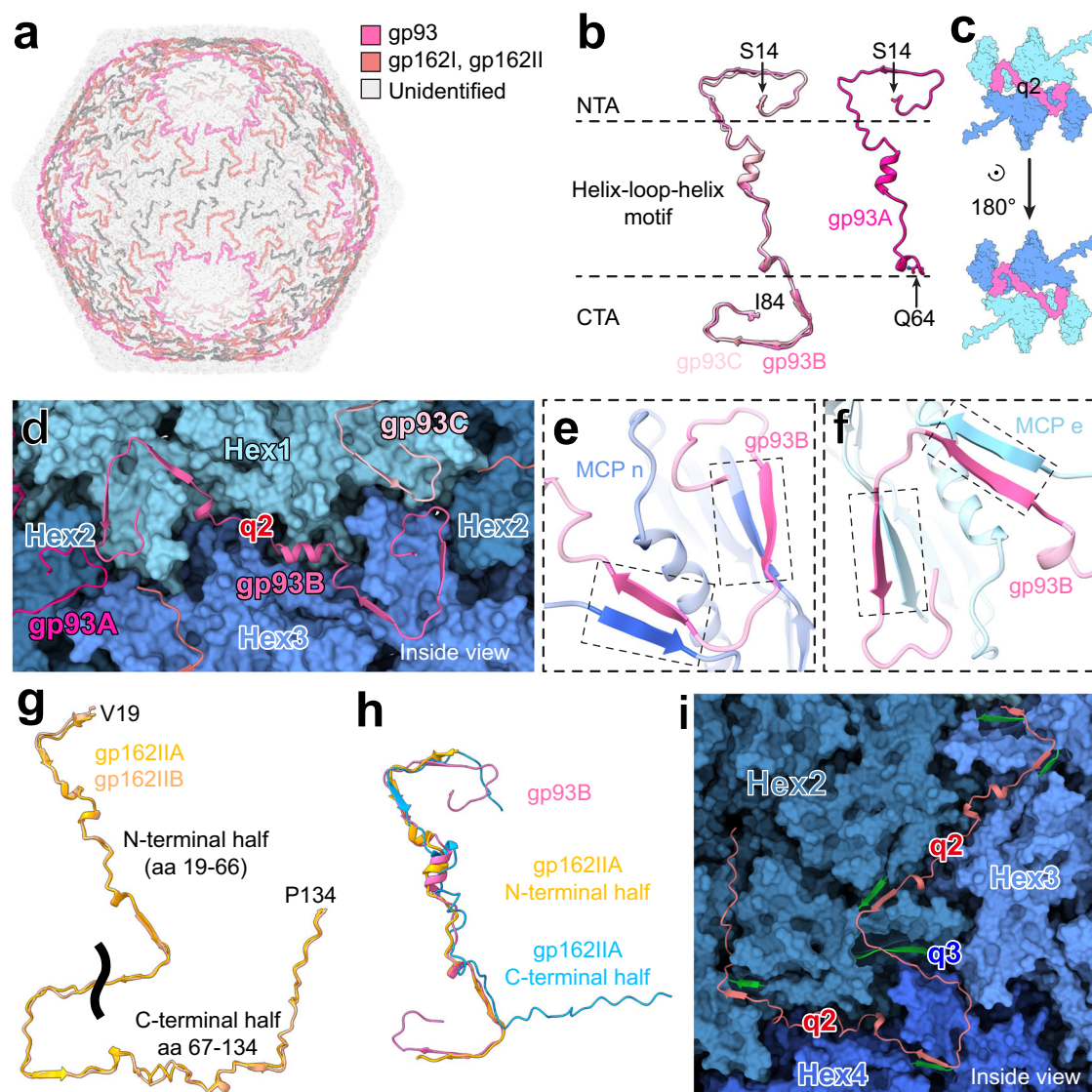
Fig. 6 | Fiber-like minor capsid proteins beneath the hexameric capsomers.
a Cryo-EM densities of the fiber-like minor capsid proteins colored according to the color key, with the cryo-EM densities of the major capsid proteins displayed semi-transparently. **b** Structural comparison of the three copies of the gp93 molecules within one icosahedral asymmetric unit. Less of gp93A (residues 14–64) was modeled compared to the other copies (gp93B and gp93C, residues 14–84). The model of gp93 consists of an N-terminal arm (residues 14–42), a helix-loop-helix motif (residues 43–62 and a C-terminal arm (residues 63–84). **c** A diagram illustrates that the structure of gp93 exhibits a quasi-C2-symmetry due to the influence of neighboring MCP subunits. The two MCP subunits are colored blue and cyan, respectively. Gp93 is colored pink. **d** Structures of gp93 (shown as ribbon diagrams) and its neighboring MCP subunits (shown as molecular surfaces). Zoomed-

in views show the anti-parallel β-strand interactions (**e**) and parallel β-strand interactions (**f**) formed between the gp93 molecules and neighboring MCP subunits. **g** Structural comparison between the two copies of gp162 molecules within one asymmetric unit. Gp162IIA and gp162IIB are colored orange and coral, respectively. The boundary of the N-terminal half (residues 19–66) and C-terminal half (residues 67–134) is indicated by a black wave line. **h** Superimposition of the N-terminal half of gp162IIA (orange), C-terminal half of gp162IIA (blue) and gp93 (pink). **i** Structures of gp162IIA and neighboring MCP subunits from the hexameric capsomers 2, 3 and 4. β-strands from the MCP subunits forming β-strand interactions with gp162II are shown as ribbon and highlighted in green. Hex, hexameric capsomer. NTA, N-terminal arm. CTA, C-terminal arm.

proteins are seldom seen in tailed bacteriophages that have smaller capsids, such as phage HK97[30,31], phage lambda[39], phage T7[40] and phage SPP1[41]. This suggests that sophisticated networks of minor capsid proteins are essential for establishing and stabilizing giant capsids characteristic of jumbo phages.

The ΦKZ virion has an "inner body" enclosed within the viral genomic DNA that may play roles related to DNA packaging and DNA ejection[17,42,43]. According to a previous proteomic study, four of these proteins, namely gp93, gp95, gp97 and gp162, are potential major components of the inner body[28]. Among the predicted structures of all ΦKZ gene products generated using AlphaFold2[34], gp93, gp95, gp97 and gp162, along with gp92, gp94, gp96 and

gp163, contain similar α-helix-rich folds (Supplementary Fig. 11). This suggests that all these eight proteins may play roles in the formation of the inner body or other related functions. Thus, gp93 and gp162 may function not only in the capsid assembly, but also in the inner body formation.

The MCP and minor capsid proteins of ΦKZ, such as gp86, gp93, gp162 and gp184, undergo proteolytic cleavage performed by the head maturation protease gp175[28]. Large portions of minor capsid proteins gp35, gp86, gp93, gp162 and gp184 were not resolved in the cryo-EM map. Except for gp35, at least one terminus of the resolved portion is close to a previously identified gp175 cleavage site[28] (Supplementary Fig. 3). The unresolved N-terminal region of gp35, according to its

predicted structure, mainly consists of loops, which may be disordered and invisible in the cryo-EM map.

The proteolytic cleavage of gp93 and gp162[28], separates the helix-rich domains from their N-terminal regions (Supplementary Fig. 3). The N-terminal regions of these proteins, as visualized in the cryo-EM, bind to the inner surface of the viral capsid, stabilizing the capsid structure and possibly driving its assembly (see above). Conversely, the helix-rich domains of these proteins are likely involved in forming the inner body of the virus, playing other roles in processes such as DNA packaging and ejection. Therefore, the proteolytic cleavage process appears to be a critical step for the multifunctional roles of the inner body proteins.

Additional cryo-EM densities, similar to the minor capsid protein network identified in the present ΦKZ structure, have been observed in the low-resolution capsid structures of various jumbo phages, such as phages ΦRSL1[13], ΦRSL2[14], ΦXacN1[14], PBS1[15], N3[15], 121Q[15] and G[12]. Furthermore, BLAST search using the UniProt database[44] identified homologs of the minor capsid proteins gp184, gp93, gp244, gp162, gp86, gp91, gp119, gp85 and gp28 in many other jumbo phages (Supplementary Table 2). This suggests that similar capsid assembly and stabilization mechanisms are probably employed by many other jumbo viruses.

Methods

ΦKZ virion preparation

The ΦKZ virions (ATCC BAA-28-B2) were propagated and purified following previously described methods[25,28]. Initially, an overnight culture of *P. aeruginosa* PAO1 (BNCC 360090) was diluted 1:1000 into 500 mL of LB media and incubated with shaking at 37 °C until the optical density at 600 nm (OD600) reached approximately 0.3. Once the culture reached the desired OD600, it was infected with ΦKZ at a multiplicity of infection (MOI) of 5. The infected culture was then incubated for about 3 hours until lysis of the host cells was observed. To remove bacterial debris, a low-speed spin was employed ($10,400 \times g$, 10 min), and the resulting supernatant was treated with DNAase I (Dingguo: DH113-6) and RNAaseI (Biosharp: BS109-25mg) at 37 °C for 30 minutes. The virions were further precipitated using 1 M NaCl and 10% PEG. The resulting pellets were then resuspended in SM buffer (50 mM Tris-HCl (pH 7.5), 100 mM NaCl, 8 mM MgSO4, 0.002% gelatin) and subjected to purification using cesium chloride gradient centrifugation ($110,000 \times g$) for 3 hours using a Beckman Coulter SW41 rotor. After purification, the virions were buffer exchanged to SM buffer and concentrated to a final concentration of about 10^{10}–10^{11} PFU/ml.

Cryo-EM sample preparation and data collection

A volume of 3 µl of purified ΦKZ virions was applied to a glow-discharged lacey carbon grid (Electron Microscopy Sciences) and was fast frozen using a Vitrobot Mark IV equipment. The grid was then loaded onto a Titan Krios microscope (ThermoFisher Scientific, operated at 300 kV) equipped with a K3 Summit direct electron detection device (Gatan) and a BioQuantum energy filter (Gatan). A cryo-EM dataset consisting of 10,916 movies was collected in super-resolution mode under a nominal magnification of 64,000 × (corresponding to a physical pixel size of 1.334 Å) with a total dose of 30 e−/Å² and a defocus range of −1.0 to −2.0 µm.

Cryo-EM data processing

Motion correction was applied to each movie using MotionCorr2[45]. The contrast transfer function (CTF) parameters were estimated using CTFFIND4[46]. The following processing was conducted using Relion 4.0[47] (Supplementary Fig. 1). For ΦKZ full capsid, a total of 21,004 particles were picked manually. Among them, 15,065 "good" particles were finally selected using 2D classification and 3D classification. The previously reported reconstruction (EMD-1392) was taken as the initial

reference map[25]. The good particles were then used for calculation of a two-fold binned consensus reconstruction of the ΦKZ genome-full capsid at a resolution of 5.4 Å.

Then, a block-based reconstruction was applied to the full capsid[10]. Briefly, the particles were symmetry-expanded to icosahedral, which generated 903,900 sub-particles. Each sub-particle was then re-centered on the capsid block using the re-extraction function in Relion 4.0[47]. To cover the full area of an icosahedral asymmetric unit of the capsid, two capsid blocks were reconstructed. The first block was centered on the five-fold vertex of the capsid. To save computing resources, the redundant C5-related sub-particles were removed. The resulting 180,780 non-redundant particles were used to perform the final reconstruction of the first block with C5 symmetry imposed. After per-particle CTF refinement, the first block was finally refined to a resolution of 3.5 Å according to the Fourier shell correlation (FSC) 0.143 cut-off (Supplementary Fig. 2a). The second block was centered close to the three-fold axis. All 903,900 sub-particles were used for the reconstruction without imposing any symmetry. After per-particle CTF refinement, the second block was finally refined to 3.6 Å according to FSC 0.143 cut-off (Supplementary Fig. 2a). Local resolutions of the two blocks were estimated using Relion 4.0[47] (Supplementary Fig. 2b). A composite map of the whole capsid was generated by combining the two blocks and applying the icosahedral symmetry as previously described[33].

Model building

The in-silico structures of all the ΦKZ encoded proteins were predicted using AlphaFold2[34]. The in-silico structure of the MCP was fitted into the cryo-EM map and manually adjusted using Coot[48]. Subtracting the densities of all MCP subunits from the cryo-EM map revealed the unmodeled components in the capsid, corresponding to the minor capsid proteins. The interpretation of the cryo-EM densities of the minor capsid proteins was started by manually building the Cα-trace models in Coot[48].

To assign sequences to these Cα-trace models, we compared each of them with all the above in-silico structures by manual inspection in UCSF Chimera[49]. For each Cα model of the minor capsid proteins gp28, gp35, gp85, gp91, gp119 and gp244, an in-silico structure was identified as a good match (Supplementary Fig. 12). Therefore, the sequence of the matched in-silico structure was assigned to its associated Cα-trace model. These sequence assignments were further validated using the findMySequence program[50], which confirmed the assignments with very high confidence (Supplementary Table 3).

The sequences of the remaining Cα-trace models for the minor capsid proteins gp86, gp93, gp162I, gp162II and gp184 were established using findMySequence[50]. The E-values associated with these sequence assignments ranged from 5.6×10^{-12} to 3.4×10^{-231} (Supplementary Table 3), exhibiting a high degree of certainty. A previous biochemical study has identified specific cleavage sites in the minor capsid proteins gp162, gp86 and gp93[28] (Supplementary Fig. 3). The correlation between the documented cleavage sites and the ends of the resolved segments of the polypeptide chains further supports the sequence assignments for these Cα-trace models. For example, the previous biochemical study has indicated that gp162 is cleaved between residues 17 and 18, and between residues 137 and 138. Notably, the atomic model for gp162II reveals a resolved segment extending from residue 19 to 134. This positioning places both ends of the resolved peptide chain in close vicinity to the reported cleavage sites.

The sequence assignments of all the polypeptide chains were supported by the fitting between the sidechains and cryo-EM densities (Supplementary Fig. 13) and the local inter-molecular chemical environments, and further validated by checkMySequence[51] (Supplementary Table 3). Finally, all atomic models were manually adjusted in Coot[48], refined using Phenix[52] and Rosetta[53] and then validated using

Phenix[52] (Supplementary Tables 1 and 3). Molecular graphics were generated using UCSF Chimera[49] and UCSF ChimeraX[54].

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The cryo-EM maps of ΦKZ's capsid generated in this study have been deposited in the Electron Microscopy Data Bank (EMDB) with accession codes EMD-39002 (composite cryo-EM map of ΦKZ capsid after applying the "block-based" reconstruction method), EMD-38845 (icosahedrally averaged cryo-EM reconstruction of ΦKZ capsid before applying the "block-based" reconstruction method), EMD-38846 (block one) and EMD-38848 (block two). Atomic model of ΦKZ's capsid have been deposited in the Protein Data Bank (PDB) with accession code PDB 8Y6V. All previously reported atomic models and cryo-EM map that were used in this study are PDB 1OHG, PDB 8BFL, PDB 8BFP and EMD-1392.

References

1. Mushegian, A. R. Are there 10(31) virus particles on earth, or more, or fewer? *J. Bacteriol.* **202**, e00052–20 (2020).
2. Hendrix, R. W. Jumbo bacteriophages. *Curr. Top. Microbiol Immunol.* **328**, 229–240 (2009).
3. Harding, K. R., Kyte, N. & Fineran, P. C. Jumbo phages. *Curr. Biol.* **33**, R750–R751 (2023).
4. Al-Shayeb, B. et al. Clades of huge phages from across Earth's ecosystems. *Nature* **578**, 425–431 (2020).
5. Malone, L. M. et al. A jumbo phage that forms a nucleus-like structure evades CRISPR-Cas DNA targeting but is vulnerable to type III RNA-based immunity. *Nat. Microbiol.* **5**, 48–55 (2020).
6. Laughlin, T. G. et al. Architecture and self-assembly of the jumbo bacteriophage nuclear shell. *Nature* **608**, 429–435 (2022).
7. Chaikeeratisak, V. et al. Assembly of a nucleus-like structure during viral replication in bacteria. *Science* **355**, 194–197 (2017).
8. Pires, D. P., Vilas Boas, D., Sillankorva, S. & Azeredo, J. Phage therapy: a step forward in the treatment of Pseudomonas aeruginosa infections. *J. Virol.* **89**, 7449–7456 (2015).
9. Krylov, V. et al. Phage phiKZ-the first of giants. *Viruses* **13**, 149 (2021).
10. Zhu, D. et al. Pushing the resolution limit by correcting the Ewald sphere effect in single-particle Cryo-EM reconstructions. *Nat. Commun.* **9**, 1552 (2018).
11. Zhang, X. & Zhou, Z. H. Limiting factors in atomic resolution cryo electron microscopy: no simple tricks. *J. Struct. Biol.* **175**, 253–263 (2011).
12. Gonzalez, B. et al. Phage G structure at 6.1 Å resolution, condensed DNA, and host identity revision to a lysinibacillus. *J. Mol. Biol.* **432**, 4139–4153 (2020).
13. Effantin, G. et al. Cryo-electron microscopy three-dimensional structure of the jumbo phage PhiRSL1 infecting the phytopathogen Ralstonia solanacearum. *Structure* **21**, 298–305 (2013).
14. Neumann, E. et al. 3D structure of three jumbo phage heads. *J. Gen. Virol.* **101**, 1219–1226 (2020).
15. Hua, J. et al. Capsids and genomes of jumbo-sized bacteriophages reveal the evolutionary reach of the HK97 fold. *mBio* **8**, e01579–17 (2017).
16. Ouyang, R. et al. High-resolution reconstruction of a Jumbo-bacteriophage infecting capsulated bacteria using hyperbranched tail fibers. *Nat. Commun.* **13**, 7241 (2022).
17. Yuan, Y. & Gao, M. Jumbo bacteriophages: an overview. *Front Microbiol.* **8**, 403 (2017).
18. Mesyanzhinov, V. V. et al. The genome of bacteriophage phiKZ of Pseudomonas aeruginosa. *J. Mol. Biol.* **317**, 1–19 (2002).
19. Sathe, N. et al. Pseudomonas aeruginosa: infections and novel approaches to treatment "Knowing the enemy" the threat of Pseudomonas aeruginosa and exploring novel approaches to treatment. *Infect. Med (Beijing)* **2**, 178–194 (2023).
20. Poole, K. Pseudomonas aeruginosa: resistance to the max. *Front Microbiol.* **2**, 65 (2011).
21. Lister, P. D., Wolter, D. J. & Hanson, N. D. Antibacterial-resistant Pseudomonas aeruginosa: clinical impact and complex regulation of chromosomally encoded resistance mechanisms. *Clin. Microbiol Rev.* **22**, 582–610 (2009).
22. Chegini, Z. et al. Bacteriophage therapy against Pseudomonas aeruginosa biofilms: a review. *Ann. Clin. Microbiol. Antimicrob.* **19**, 45 (2020).
23. Rai, P. et al. Characterisation of broad-spectrum phiKZ like jumbo phage and its utilisation in controlling multidrug-resistant Pseudomonas aeruginosa isolates. *Micro. Pathog.* **172**, 105767 (2022).
24. Krylov, V. et al. Selection of phages and conditions for the safe phage therapy against Pseudomonas aeruginosa infections. *Virol. Sin.* **30**, 33–44 (2015).
25. Fokine, A. et al. A three-dimensional cryo-electron microscopy structure of the bacteriophage phiKZ head. *J. Mol. Biol.* **352**, 117–124 (2005).
26. Fokine, A. et al. Cryo-EM study of the Pseudomonas bacteriophage phiKZ. *Structure* **15**, 1099–1104 (2007).
27. Wu, W., Thomas, J. A., Cheng, N., Black, L. W. & Steven, A. C. Bubblegrams reveal the inner body of bacteriophage phiKZ. *Science* **335**, 182 (2012).
28. Thomas, J. A. et al. Extensive proteolysis of head and inner body proteins by a morphogenetic protease in the giant Pseudomonas aeruginosa phage phiKZ. *Mol. Microbiol.* **84**, 324–339 (2012).
29. Lecoutere, E. et al. Identification and comparative analysis of the structural proteomes of phiKZ and EL, two giant Pseudomonas aeruginosa bacteriophages. *Proteomics* **9**, 3215–3219 (2009).
30. Wikoff, W. R. et al. Topologically linked protein rings in the bacteriophage HK97 capsid. *Science* **289**, 2129–2133 (2000).
31. Helgstrand, C. et al. The refined structure of a protein catenane: the HK97 bacteriophage capsid at 3.44 Å resolution. *J. Mol. Biol.* **334**, 885–899 (2003).
32. Chen, Z. 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. USA* **114**, E8184–E8193 (2017).
33. Fang, Q. et al. Structures of a large prolate virus capsid in unexpanded and expanded states generate insights into the icosahedral virus assembly. *Proc. Natl. Acad. Sci. USA* **119**, e2203272119 (2022).
34. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
35. Stone, N. P., Demo, G., Agnello, E. & Kelch, B. A. Principles for enhancing virus capsid capacity and stability from a thermophilic virus capsid structure. *Nat. Commun.* **10**, 4471 (2019).
36. Wang, Z. et al. Structure of the marine siphovirus TW1: evolution of capsid-stabilizing proteins and tail spikes. *Structure* **26**, 238–248 e233 (2018).
37. Fang, Q. et al. Near-atomic structure of a giant virus. *Nat. Commun.* **10**, 388 (2019).
38. Shao, Q. et al. Near-atomic, non-icosahedrally averaged structure of giant virus Paramecium bursaria chlorella virus 1. *Nat. Commun.* **13**, 6476 (2022).
39. Wang, C., Zeng, J. & Wang, J. Structural basis of bacteriophage lambda capsid maturation. *Structure* **30**, 637–645 e633 (2022).
40. Guo, F. et al. Capsid expansion mechanism of bacteriophage T7 revealed by multistate atomic models derived from cryo-EM reconstructions. *Proc. Natl. Acad. Sci. USA* **111**, E4606–E4614 (2014).
41. Ignatiou, A. et al. Structural transitions during the scaffolding-driven assembly of a viral capsid. *Nat. Commun.* **10**, 4840 (2019).

42. Fossati, A. et al. Next-generation proteomics for quantitative Jumbophage-bacteria interaction mapping. *Nat. Commun.* **14**, 5156 (2023).

43. Krylov, V. N. et al. Pseudomonas bacteriophage phi KZ contains an inner body in its capsid. *Can. J. Microbiol.* **30**, 758–762 (1984).

44. Zaru, R., Orchard, S. & UniProt, C. UniProt tools: BLAST, align, peptide search, and ID mapping. *Curr. Protoc.* **3**, e697 (2023).

45. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).

46. Rohou, A. & Grigorieff, N. CTFFIND4: fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).

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

48. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).

49. Pettersen, E. F. et al. UCSF Chimera-a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

50. Chojnowski, G. et al. findMySequence: a neural-network-based approach for identification of unknown proteins in X-ray crystallography and cryo-EM. *IUCrJ* **9**, 86–97 (2022).

51. Chojnowski, G. Sequence-assignment validation in cryo-EM models with checkMySequence. *Acta Crystallogr. D Struct. Biol.* **78**, 806–816 (2022).

52. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).

53. DiMaio, F., Tyka, M. D., Baker, M. L., Chiu, W. & Baker, D. Refinement of protein structures into low-resolution density maps using rosetta. *J. Mol. Biol.* **392**, 181–190 (2009).

54. Pettersen, E. F. et al. UCSF ChimeraX: structure visualization for researchers, educators, and developers. *Protein Sci.* **30**, 70–82 (2021).

Author contributions

Y.Y., Q.S., M.G., L.H., Z.C., L.S. and Q.F. designed the research. Y.Y., M.G., B.W. and J.P. purified the ΦKZ viral particles. L.H. prepared the cryo-EM sample and collected the cryo-EM data. Q.S. and L.H. processed the cryo-EM data. Y.Y., Q.S., M.G., X.Z., A.W. and X.L. interpreted the cryo-EM map. Y.Y., Q.S., M.G., L.H., A.F., L.S. and Q.F. wrote the manuscript, with further edits from all authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-024-50811-1.

Correspondence and requests for materials should be addressed to Lei Sun or Qianglin Fang.

Peer review information *Nature Communications* thanks Edward Egelman and the other, anonymous, reviewers for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.