

# Structural and biochemical insights into the mechanism of the Gabija bacterial immunity system

Received: 6 April 2023

Accepted: 15 January 2024

Published online: 29 January 2024

 Check for updatesYanwu Huo<sup>1,4</sup>✉, Lingfei Kong<sup>1,2,4</sup>, Ye Zhang<sup>1,2,4</sup>, Min Xiao<sup>1</sup>, Kang Du<sup>1</sup>, Sunyuntao Xu<sup>1</sup>, Xiaoxue Yan<sup>1</sup>, Jun Ma<sup>3</sup>✉ & Taotao Wei<sup>1,2</sup>✉

The Gabija system is a newly discovered bacterial immune system that consists of GajA and GajB. Here we report the cryo-EM structure of the Gabija complex from *Bacillus cereus* VDO45 at 3.6 Å, which provides the direct evidence of interactions between GajA and GajB. The Gabija complex is an octameric ring structure with four GajA and four GajB. GajA is an OLD nucleases family protein, while GajB belongs to the SF1 helicases. The Gabija complex has sequence-specific DNA nuclease activity and prefers circular rather than linear DNA as substrate, its activity is more sensitive to concentrations change of nucleotides compared to GajA alone. Our data suggest a mechanism of Gabija immunity: the nuclease activity of Gabija complex is inhibited under physiological conditions, while it is activated by depletion of NTP and dNTP upon the replication and transcription of invading phages and cleave the circular DNA to prevent phage DNA replication.

Bacteria have developed sophisticated defense systems to protect themselves from the invasion of bacteriophages. Different defense strategies have been employed to prevent the proliferation of phages. Examples include preventing phage DNA from entering cells, damaging the invading DNA and promoting host cell death. Numerous bacterial defense systems employ the strategy of degrading the invading DNA. This is achieved through sequence-specific DNA cleavage, such as CRISPR/Cas and R-M systems<sup>1,2</sup>. By understanding these mechanisms of immune defense, powerful molecular biology tools or gene editing tools have been developed, including the well-known restriction endonuclease, CRISPR/Cas9.

An increasing number of defense systems have been discovered recently<sup>3–8</sup>. Doron et al. predicted computationally and validated experimentally 10 previously unknown bacterial immune defense systems<sup>3</sup>. Among these newly discovered systems, the Gabija system is the most widespread, accounting for 8.5% of the sequenced bacterial and archaeal genomes (4360 genomes). The Gabija system consists of two genes, GajA and GajB. A previous study showed that the Gabija

system from *Bacillus cereus* VDO45 exhibits strong immunity against bacteriophages phi29, rho14, phi105 and SpBeta, indicating that the Gabija system is sufficient for bacterial immunity against certain bacteriophages<sup>3</sup>. GajA is a sequence-specific DNA-cleaving enzyme in the presence of Mg<sup>2+</sup>, its activity is completely inhibited by the high concentration of NTP and dNTP<sup>9</sup>. Prior research indicates that GajA plays an essential role in Gabija immunity system by sequence-specific DNA cleavage regulated by nucleotide concentration. However, the immunity of *E. coli* to bacteriophage T7 requires the whole Gabija gene cassette, indicating that the other gene of the Gabija system, GajB, is indispensable.

In this work, we purify and characterize the Gabija complex (GajA and GajB) from *B. cereus* VDO45 and present the structure of the Gabija complex at 3.6 Å resolution. The resolved structure shows that the Gabija complex adopts an octameric ring structure with four GajA and four GajB. This observation provides direct evidence of the interaction between GajA and GajB in the Gabija defense system. GajA is an OLD (overcoming lysogenization defect) nucleases family protein

<sup>1</sup>National Laboratory of Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, 15 Datun Road, Chaoyang District, Beijing 100101, China.

<sup>2</sup>School of Biological Sciences, University of Chinese Academy of Sciences, 19 Yuquan Road, Shijingshan District, Beijing 100049, China. <sup>3</sup>Institute of Infectious Diseases, Shenzhen Bay Laboratory, Gaoke Innovation Center, Guangqiao Road, Guangming District, Shenzhen, Guangdong 518132, China.

<sup>4</sup>These authors contributed equally: Yanwu Huo, Lingfei Kong, Ye Zhang ✉e-mail: [huoyw@ibp.ac.cn](mailto:huoyw@ibp.ac.cn); [majun@szbl.ac.cn](mailto:majun@szbl.ac.cn); [weitt@ibp.ac.cn](mailto:weitt@ibp.ac.cn)

containing an ATPase domain and a Toprim (topoisomerase/primase) domain. The ATPase domain of GajA encodes a conserved active site and adopts a productive conformation with the ATP-binding cassette (ABC) signature sequence orienting toward the active site of the opposing subunit in close approach to the P loop and bound nucleotide, while the Toprim domain has a conserved nuclease active site. On the other hand, GajB belongs to superfamily I (SF1) helicases and shares the essential features of this family conserved in spatial distribution. Our data further suggest that the Gabija complex has the sequence-specific DNA endonuclease activity and prefers supercoiled circular DNA as its substrate. The cleavage activity of Gabija complex is completely inhibited with 0.5 mM ATP towards the linear DNA substrate, while the cleavage activity on supercoiled circular DNA is still preserved at 0.5–2 mM ATP and inhibited by much higher concentrations of nucleotides. Together our data provide important insights into the structure and mechanisms of the Gabija bacterial immunity system.

## Results

### The overall structure and structural organization of the Gabija complex

The Gabija complex from *B. cereus VDO45* was heterogeneously expressed in *E. coli* and purified by chromatography (Supplementary Fig. 1a). Size exclusion chromatography coupled with multi-angle light scattering (SEC-MALS) revealed that the Gabija complex forms a stable hetero-octamer in solution (Supplementary Fig. 1b). In contrast, individual GajA and GajB exist in the form of tetramer and monomer, respectively (Supplementary Fig. 1c, d). Therefore, it is inferred that the octamer form of the Gabija complex is the minimum work unit of the Gabija system. Both GajA and GajB are essential for anti-phage activity of Gabija system; deletion of either gene leads to the loss of Gabija immunity revealed by phage resistance assay in vivo (Supplementary Fig. 2). Subsequently, the structure of Gabija complex is solved by the cryo-EM single particle analysis at a resolution of 3.6 Å (Supplementary Fig. 3 and Supplementary Table 1). The whole Gabija complex resembles an octameric ring with three 2-fold symmetry axes. The three dimensions of the octameric ring are 171 Å × 143 Å × 91 Å; The inner diameter of the ring is 35 Å at the narrowest region. Each octamer containing four GajA and four GajB (Fig. 1a, b and Supplementary Fig. 4a, b).

SEC-MALS results show that individual GajA and GajB exist in the form of tetramer and monomer respectively (Supplementary Fig. 1c, d), supporting the hypothesis that the octameric ring of Gabija complex is formed through two steps: the tetramer formation of GajA and subsequent octamer formation of Gabija complex. The tetramer formation of GajA leads to the interfaces between GajA1-GajA2 and GajA1-GajA3 (Fig. 1c, d). The A1-A2 interface involves symmetric interactions, the contact sites are mainly related to ATPase domain of GajA. They are  $\alpha 2$ ,  $\beta 6$ , the loop between  $\alpha 5$  and  $\beta 8$ , the loops on both sides of the  $\beta 6$  (Fig. 1c). The A1-A3 interface involves symmetric interactions. These contacting sites spread both the ATPase and Toprim domains of GajA. They are  $\alpha 5$ ,  $\alpha 6$ ,  $\alpha 7$ ,  $\alpha 8$ ,  $\alpha 10$ ,  $\alpha 15$ ,  $\alpha 16$ ,  $\beta 13$ , the loop between  $\alpha 14$  and  $\alpha 15$ , the loop between  $\alpha 10$  and  $\beta 14$ , the loop between  $\alpha 8$  and  $\beta 12$ , the loops on both sides of the  $\beta 13$ , the loop between  $\alpha 7$  and  $\beta 11$ , the loop between  $\beta 9$  and  $\beta 10$ , the loop between  $\alpha 1$  and  $\beta 3$ , the loop between  $\alpha 6$  and  $\beta 8$  (Fig. 1d). The octamer formation of Gabija complex leads to the interfaces between GajA1-GajB1, GajA1-GajB2 and GajB1-GajB2 (Fig. 1e, f). The interactions within GajA1-GajB1 are mediated by the ATPase domain of GajA and the 1A, 1B domains of GajB. The  $\alpha 3$  helix, the loops on both sides of the  $\alpha 3$  helix, the loop between  $\alpha 5$  and  $\beta 8$  and the  $\alpha 4$  helix of GajA are involved in this contact interface. In GajB, the interface contact sites are  $\alpha 4$ ,  $\alpha 7$ , the loop between  $\alpha 2$  and  $\beta 2$  and the loop between  $\alpha 3$  and  $\beta 3$  (Fig. 1e). The flexible CTR region (C-terminal region, residues 224–416) of GajB could not be traced in the electron density map. To determine the interactions between GajA and CTR

region of GajB, two constructs of GajB have been made, the NTR region (N-terminal region, residues 1–223) and the CTR region. Size exclusion chromatography demonstrates that the NTR region of GajB indeed binds to GajA, while the CTR region of GajB doesn't bind to GajA (Supplementary Fig. 5a, b). Interface A1-B2 is mediated by the  $\beta 5$ ,  $\beta 6$ , the loop between  $\beta 5$  and  $\beta 6$  of GajA ATPase domain and by the loop between  $\alpha 3$  and  $\beta 3$  of GajB 1A domain. Interface B1-B2 is mediated by the  $\alpha 5$  and the loops on both sides of the  $\alpha 5$  (Fig. 1f).

### Structural organization of GajA and GajB

GajA is an OLD nucleases family protein based on sequence homology, and contains two domains: an N-terminal ATPase domain (1–351 amino acid) and a C-terminal Toprim domain (352–578 amino acid) (Fig. 2a). The ATPase domain contains an eleven-stranded  $\beta$ -sheet with the order  $\beta 6$ - $\beta 5$ - $\beta 4$ - $\beta 1$ - $\beta 2$ - $\beta 11$ - $\beta 10$ - $\beta 3$ - $\beta 9$ - $\beta 8$ - $\beta 7$ . This  $\beta$ -sheet folds into two layers at the  $\beta 2$  and  $\beta 11$  strands and a central  $\alpha$  helix  $\alpha 1$  inserts into it, forming a sandwich-like structure. Helices  $\alpha 2$ ,  $\alpha 3$ , and  $\alpha 4$  flank the sandwich, capping the N terminal of  $\alpha 1$ . Helices  $\alpha 5$  and  $\alpha 6$  lie at the bottom of the sandwich, making a four-layer structure.

The Toprim domain contains a four-stranded  $\beta$ -sheet with the order  $\beta 13$ - $\beta 12$ - $\beta 14$ - $\beta 15$ . The helices  $\alpha 11$  and  $\alpha 12$  rest on the top of  $\beta 15$  within the same plane of the four-stranded  $\beta$ -sheet. The helices  $\alpha 7$ ,  $\alpha 9$ , and  $\alpha 10$  lie on the left side of the  $\beta$ -sheet plane, interacting with the  $\beta 11$ ,  $\beta 10$  and the loop within the  $\beta 10$ - $\beta 9$  of the ATPase domain. The helices  $\alpha 8$ ,  $\alpha 13$ ,  $\alpha 14$ ,  $\alpha 15$ ,  $\alpha 16$ ,  $\alpha 17$ , and  $\alpha 18$  lie on the right side of the  $\beta$ -sheet plane.

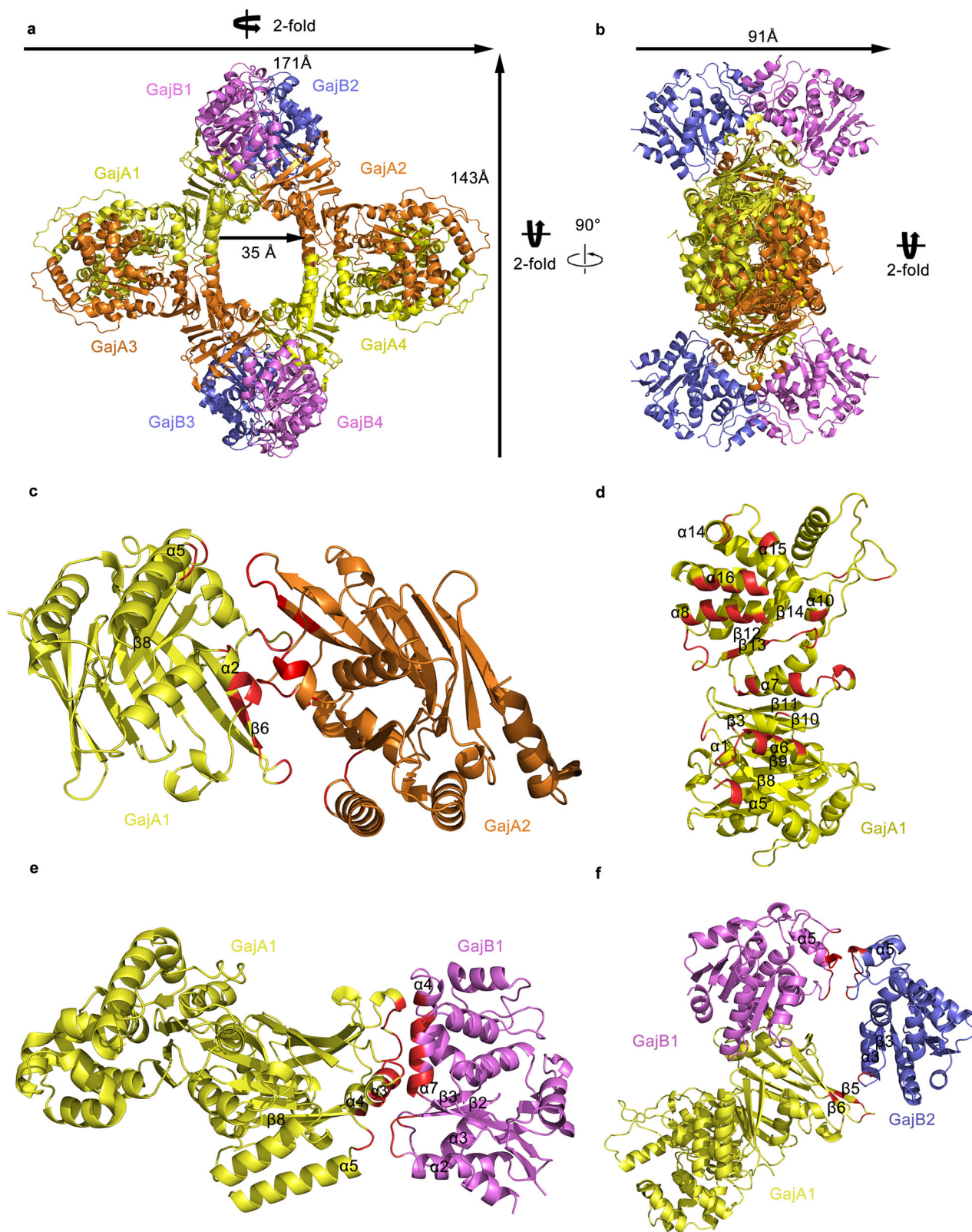
GajB subunit belongs to SF1 helicase based on sequence alignment, it contains four structural domains 1A (1–72, 157–223 AA), 1B (73–156 AA), 2A (224–305 AA, 417–494 AA) and 2B (306–416 AA) (Fig. 2b). Despite being present in the cryo-EM protein sample, domain 2A and 2B could not be traced in the electron density map, demonstrating that this region is quite flexible in the Gabija complex.

### Essential features of SF1 helicase of GajB

Sequence alignment reveals that 11 typical sequences commonly found in SF1 helicases are well conserved in GajB motifs (I-VI, Ia, Ib, Id, IVa and IVc) (Fig. 2c and Supplementary Fig. 6). The seven sequence motifs (I, Ia, II-VI) and Q motif conserved in SF1 helicase are involved in ATP binding as reported. In our structure, the ATP binding motifs (I, Ia, II, III) gather together and form the most conserved region (Fig. 2d). The Q motif is supposed to be situated in the  $\alpha 1$  helix of the GajB structure, while the Q motif is missing in  $\alpha 1$  helix. The adenine base is selected explicitly by Q motif, the missing of the Q motif results in the loss of selectivity of ATP. The GTPase activity of GajB is observed indeed in the following biochemical assay, but the GTPase activity of GajB appears much weaker than the ATPase activity, which indicating that the ATPase domain of GajB still has a higher affinity for adenine than guanosine. (Supplementary Fig. 7b, e). The good conservation of sequence alignment and spatial distribution of typical motifs imply that GajB may employ a similar mechanism for ATP hydrolysis as the SF1 helicases family.

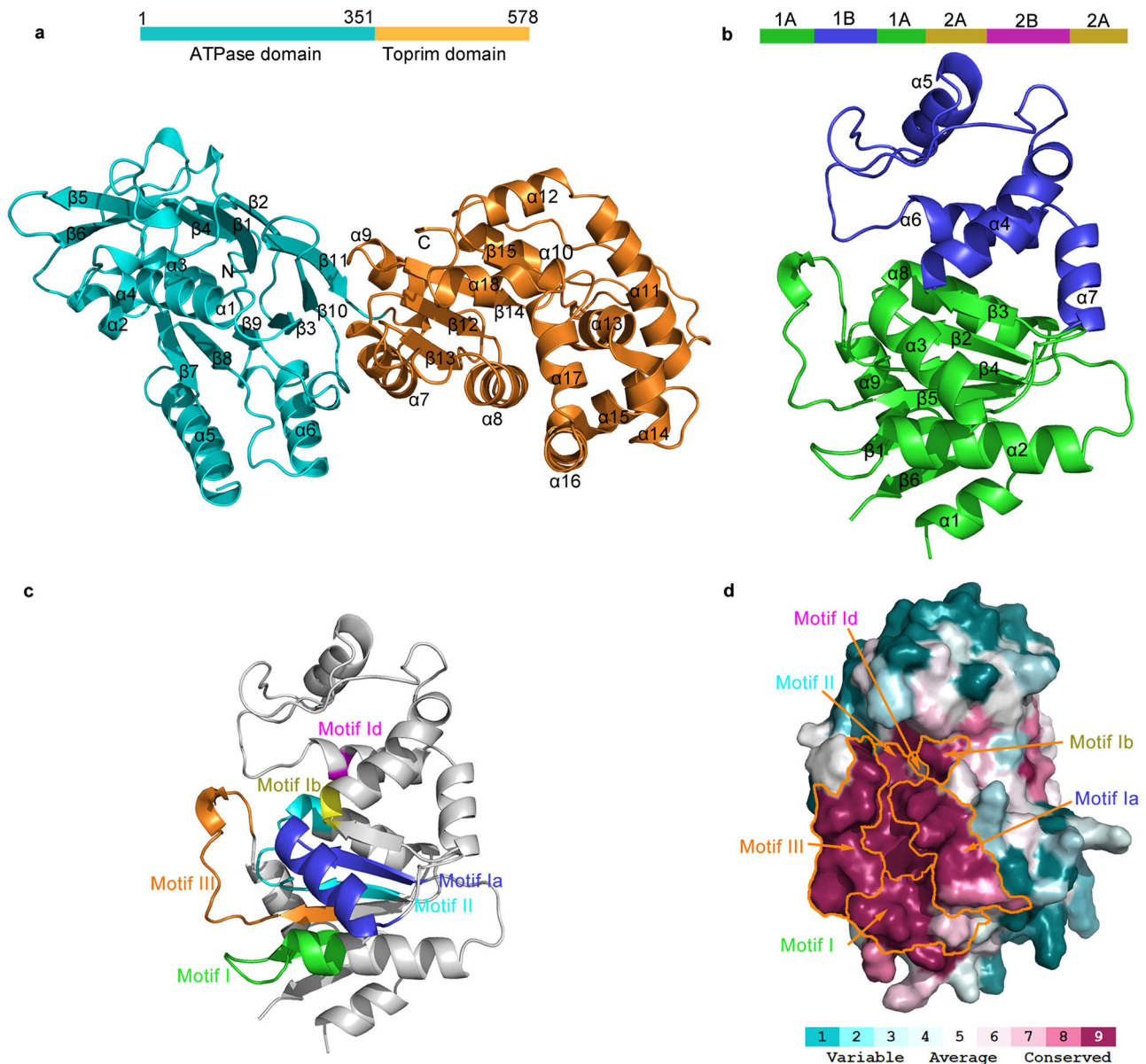
### GajA Toprim domain have a conserved nuclease active site

There is a highly conserved region (E379, E383, D432, D434, D511, D513, K541) in the Toprim domain (Fig. 3c). By structural alignment with OLD CTR structure from *Burkholderia pseudomallei* (OLD\_Bp), this region is determined to be the site of nuclease activity (Fig. 3a). GajA residues E379, E383, D432, D434, D511, D513, align well with the OLD\_Bp metal binding residues E400, E404, D455, D457, T506 and E508. GajA residue K541 roughly aligns with the OLD\_Bp K562 (Fig. 3b). K562 is the catalytic amino acid that plays a role in stabilizing the negative charge of transition state of the phosphoryl transfer reaction and protonate the leaving group. The spatial conservation of the catalytic site and good alignment with OLD\_Bp may imply that GajA employs conserved



**Fig. 1 | Overall structure of the Gabija complex and interactions between the subunits of the Gabija complex.** **a** The top views of the Gabija complex octameric ring structure. **b** The side views of the Gabija complex octameric ring structure. The three 2-fold symmetry axes and the measurement of the octameric ring dimensions are marked. **c, d** The tetramer formation of GajA leads to the interactions between GajA1-GajA2 and GajA1-GajA3 (interface A1-A2 in panel **c** and interface A1-A3 in

panel **d**), the contacting residues in the GajA1-GajA3 interface are symmetric related, for clarity, only subunit D is displayed. The octamer formation of Gabija complex leads to the interactions between GajA1-GajB1, GajA1-GajB2 and GajB1-GajB2; (interface A1-B1 in panel **e**, interface A1-B2 and B1-B2 in panel **f**). The four GajA are colored yellow and orange. The four GajB are colored blue and purple. The contacting residues in the interface are colored red.



**Fig. 2 | Structural analysis of GajA and GajB.** **a** Structural organization of full-length GajA. GajA contains two domains: an N-terminal ATPase domain (1-351 amino acid) and a C-terminal Toprim domain (352-578 amino acid), as illustrated by a schematic drawing of the structural domain. The number of boundaries is labeled. ATPase domain is colored cyan; Toprim domain is colored orange. **b** Structural organization of GajB. GajB contains four structural domains 1A (1-72, 157-223 AA), 1B (73-156 AA), 2A (224-305 AA, 417-494 AA) and 2B (306-416 AA) as illustrated by

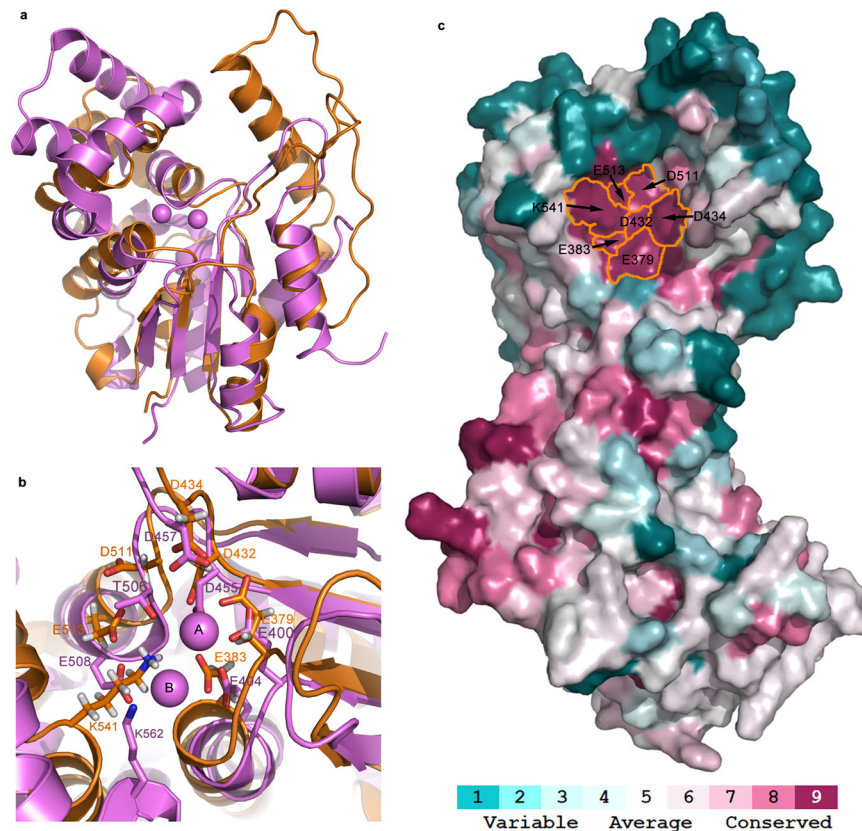
schematic drawing of the structural domain. 1A domain is colored green, 1B domain is colored blue, and the electron density of 2A and 2B domains is missing due to structural flexibility. **c, d** Dissection of the SF1 helicase of GajB. **c** Mapping of the SF1 helicase motifs onto the GajB structure. **d** Surface conservation in the GajB structure by ConSurf server<sup>21</sup>. Residues are colored from magenta to cyan with descending order of conservation.

catalytic machinery and a similar two-metal mechanism for nuclease cleavage<sup>10</sup>.

### The molecular mechanism of substrate selectivity of GajA

Based on sequence alignment, the GajA is homologous to the OLD family nucleases (OLD\_Bp and OLD\_Ts)<sup>10,11</sup>. The DNA cleavage activity of the homologous OLD nucleases is non-specific, while GajA exhibits sequence-specific DNA endonuclease activity. To explore the mechanism of substrate selectivity of GajA, we analyze the structure of GajA and carry out biochemical verification. The structure analysis reveals two unique lysine-rich structure elements in the GajA: an  $\alpha$ -helix (residue 474-493) and a flexible loop (residue 436-442). Two unique lysine-rich structure elements are divided into three patches according to the position relative to the active center: patch 1 (K436,

K438, K439, K441, K442), patch 2 (R481, R485, K487, K491) and patch 3 (K474, K476, K477, K478, K479). Patch 1 and 3 are just around the active center. In the contrary, patch 2 is on the opposite side of the active center (Fig. 4a). To examine whether the basic residues on three patches contribute to the substrate selectivity of GajA, we made three mutants, GajA<sub>mut1</sub> (K436A, K438A, K439A, K441A, K442A), GajA<sub>mut2</sub> (R481A, R485A, K487A, K491A) and GajA<sub>mut3</sub> (K474A, K476A, K477A, K478A, K479A). Non-specific DNA cleavage activity is only detected in the GajA<sub>mut1</sub> (Fig. 4b, c). These data suggest that the lysine-rich basic patch 1 of loop (residue 436-442) is responsible for the selection of sequence specific DNA digestion of GajA. Due to the lack of structural information of GajAB complex with DNA substrates, it could only be inferred from biological mutation experiments, the selection of sequence specific DNA digestion of GajA may be related to interactions



**Fig. 3 | GajA Toprim domains have a conserved nuclease active site.** **a** Structural alignment of GajA Toprim domains from *Bacillus cereus* VDO45 (orange) and *Burkholderia pseudomallei* (OLD\_Bp, PDB: 6NK8; purple) with the RMSD of 3.2 Å. **b** Close-up view of active sites of nuclease of GajA Toprim domains. Coloring as in **a**.

Bound magnesium ions in 6NK8 are shown as purple spheres marked with A and B. The key residues in the active sites are labeled. **c** Surface conservation of active site residues in the GajA structure by ConSurf server<sup>21</sup>. Residues are colored from magenta to cyan in descending order of conservation.

between the phosphate backbone of the DNA substrate palindrome structure and GajA<sub>muti</sub>.

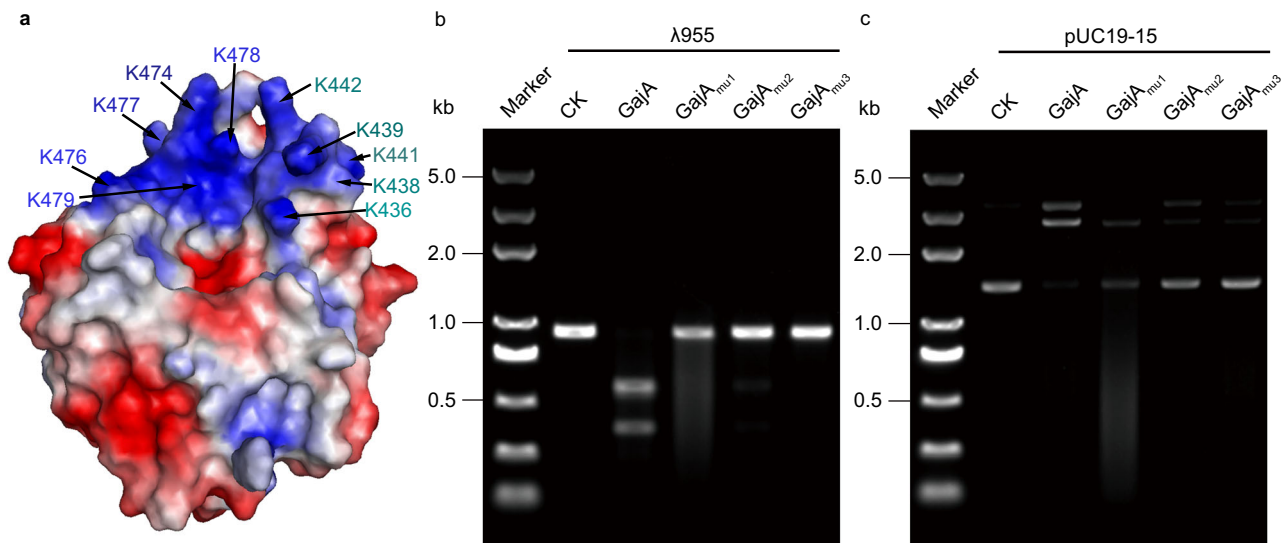
### GajA ATPase domains have a conserved active site and adopt a productive conformation

The ATPase domain of GajA is related to ABC proteins family based on the sequence alignment and structure comparison. Structure alignment by Dali<sup>12</sup> with OLD nuclease from *Thermus Scotoductus* (OLD\_TS, PDB: 6p74, Z-score = 20.5, RMSD = 3.4 Å) reveals that they have similar overall structure and identifies six conserved sequence motifs (P loop, Q loop, the ABC signature sequence, Walker B, D loop and H loop) in GajA that contribute to ATP binding and hydrolysis (Supplementary Fig. 8a). These six sequence motifs cluster together to form the most conserved region in the ATPase domain of GajA (Supplementary Fig. 8b). The conservation of sequence and spatial distribution in the active site of ATPase imply that the ATPase domain of GajA utilizes a conserved catalytic machinery for ATP binding and hydrolysis. The dimerization of nucleotide binding domain (NBD) of ABC ATPases is necessary for ATP binding and hydrolysis, which is mediated by sandwiching two ATP molecules at the interface between the associating subunits. After dimerization of NBD, the catalytic machinery for productive ATP hydrolysis is formed, which is not only composed of the P loop, Q loop, Walker B, and H loop from one subunit, but also composed of the ABC signature sequence and D loop from the other opposing subunit (Supplementary Fig. 8c). Structure analysis of the dimer of ATPase domain of GajA showed that GajA adopt a productive ATP hydrolysis state like chromosome partition protein SMC (PDB:5xg3). In this state, the ABC signature sequence is oriented toward the active site of the opposing subunit in close approach to the P loop and bound nucleotide (Supplementary Fig. 8c).

This orientation of the ATPase domains within GajA dimer is significant different from the OLD nuclease (PDB:6p74), which is the only full-length structure of OLD nuclease resolved so far. Superposition of ATPase domains of GajA dimer with OLD nuclease shows that one subunit from each ATPase domain dimer aligns well (Supplementary Fig. 8e). However, the second GajA subunit is rotated 180 degrees relative to OLD nuclease around the dimer axis (Supplementary Fig. 8f). The D loop in the second GajA subunit is far away nearly 48 Å from its counterpart in OLD nuclease (Supplementary Fig. 8f). This suggests that the conformations of the ATPase domains within the GajA dimer and OLD nuclease dimer are two different function state, these two conformations would inter-change during the ATP binding and hydrolysis cycle.

### Gabija complex prefers to cleave supercoiled circular DNA substrate to linear DNA

In concert with a previous study showing that GajA is a sequence-specific and metal-dependent DNA nuclease<sup>9</sup>, the GajA protein we purified showed the same cleavage activity on 955bp linear DNA substrate and supercoiled plasmids containing recognition sequence (5'-AATAACCCGATATT-3'). Intriguingly, the Gabija complex we purified showed substrate preference on supercoiled circular DNA rather than linear DNA substrate (Fig. 5a, b). The cleavage activity of the Gabija complex on linear DNA substrate decreased drastically as only about 30% of the linear DNA substrate is digested by Gabija complex (Fig. 5c). The cleavage activity of the Gabija complex on supercoiled plasmid is not affected at all compared to GajA protein alone, almost all the supercoiled plasmid was completely digested by Gabija complex (Fig. 5d). Together, these data suggest that the Gabija complex has



**Fig. 4 | The molecular mechanism of substrate selectivity of GajA.** **a** Electrostatic potential of GajA Toprim domains. The residues in patch 1 (K436, K438, K439, K441, K442) and patch 3 (K474, K476, K477, K478, K479) just around the active center is labeled. **b** Nuclease activity of wild-type GajA and mutant protein on  $\lambda$ 955 linear DNA substrate; **c** Nuclease activity of wild-type GajA and mutant protein on supercoiled plasmids substrate; CK, control reaction without GajA.  $\lambda$ 955 is a 955 bp

fragment from lambda DNA genome with the recognition sequence (5'-AATAACCCGGATATT-3'). The pUC19-15 plasmid contains the cleavage recognition sequence (5'-AATAACCCGGATATT-3'). GajA<sub>mut1</sub> (K436A, K438A, K439A, K441A, K442A), GajA<sub>mut2</sub> (R481A, R485A, K487A, K491A) and GajA<sub>mut3</sub> (K474A, K476A, K477A, K478A, K479A). All the assays are repeated at least 3 times with similar results. Source data are provided as a Source Data file.

stronger substrate selectivity with the presence of GajB compared to GajA alone.

### The cleavage activity of Gabija complex is regulated by nucleotides

The structural analysis demonstrate that Gabija complex has two ATPase active site, one is in the ATPase domain of GajA, and the other one is in the interface of A1 and A2 domain of GajB. We measured the ATPase activity of purified GajA, GajB, and the Gabija complex by monitoring the amount of free phosphate released. We observed that the  $V_{max}$  of GajA, GajB and Gabija complex activity of ATP hydrolysis is 4.088 mM/min, 0.953 mM/min and 7.801 mM/min, respectively (Supplementary Fig. 7a–c). When GajA and GajB form a complex, the positive synergistic effect of ATP hydrolysis capacity of the complex is observed, which suggests that GajB may be the regulator of GajA, and the interaction between them may affect their ATPase activity. We also observed that the  $V_{max}$  of GajA, GajB and Gabija complex activity of GTP hydrolysis is 1.364 mM/min, 0.218 mM/min and 1.829 mM/min, respectively (Supplementary Fig. 7d–f). In the case of GTP substrate, there is also a positive synergistic effect, although it is weaker than ATP substrate, which is consistent with the affinity of the substrate.

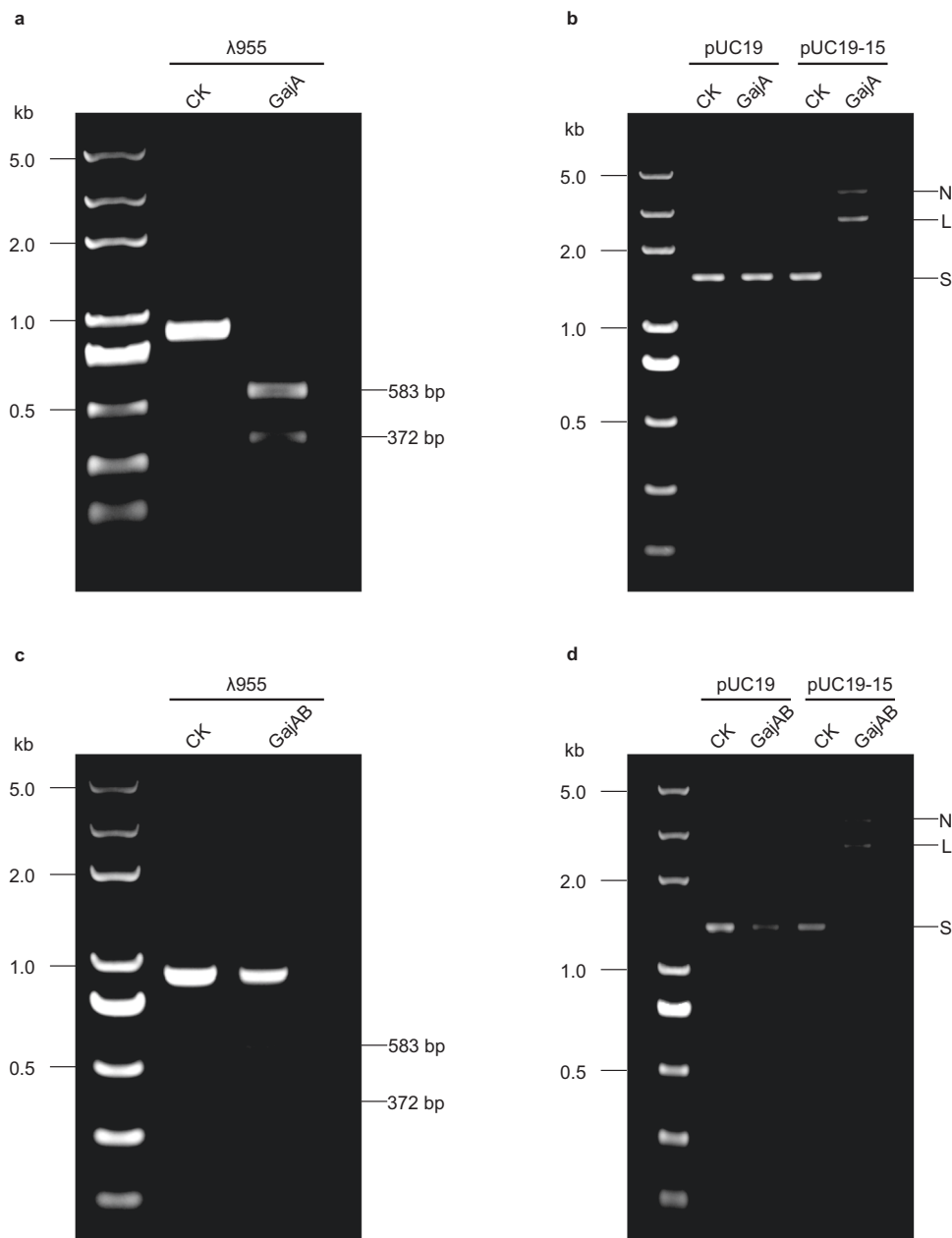
To further investigate the synergy of two the ATPase active sites in regulating Gabija nuclease activity, we conducted alanine screening of conserved residues in two ATPase active sites of the Gabija complex. Two mutants are constructed, one is H320A of GajA, and the other one is D162A and E163A of GajB. Both mutations in two ATPase active sites of Gabija complex abolish the anti-phage activity of the Gabija system (Supplementary Fig. 2). The most striking phenotype observed is that the AB<sub>mut</sub> Gabija complex is much more active in degrading linear DNA than the wild-type protein (Supplementary Fig. 7g). This result further revealed that the ATPase activity of GajB is related to the substrate selectivity of the Gabija complex, making Gabija complex prefer supercoiled circular over linear DNA. The loss of circular DNA substrate preference in AB<sub>mut</sub> Gabija complex may result in abolishing the phage resistance of Gabija system in vivo (Supplementary Fig. 2). D162A and E163A mutations in GajB partially relieve the inhibition of ATP and GTP on Gabija complex endonuclease activity only in the presence of optimum supercoiled plasmid substrates (Supplementary

Fig. 7g, h), H320A mutation in GajA show no effect on relief of inhibition of ATP and GTP on Gabija complex endonuclease activity both in linear and supercoiled DNA substrate, suggesting that the ATPase domain in the GajB may regulate the nuclease activity by sensing nucleotide change.

Next, we sought to test the effect of ATP or GTP on the nuclease activity of the Gabija complex. In the presence of linear DNA substrate, the low concentration (0.05 mM) of ATP/GTP both promote the nuclease activity of the Gabija complex; The high concentration (0.1 mM and above) of ATP/GTP both inhibit the nuclease activity of Gabija complex (Fig. 6a, b). In the presence of supercoiled plasmids substrate, the nuclease activity of the Gabija complex on the plasmid is significantly inhibited with the increased concentration of ATP and GTP. When the concentration of ATP reached 2 mM, the nuclease activity of the Gabija complex was entirely suppressed (Fig. 6c, d). We further measure the effect of different nucleotides on the cleavage activity of the Gabija complex. The results reveal that the cleavage activity of the Gabija complex is significantly inhibited by different NTP, dNTP, ADP, but not AMP. An interesting observation is that dNTP is more effective than NTP in the presence of pUC19-15 substrate, this implies that the depletion of dNTP caused by DNA replication may play a more important role in regulating cleavage activity of the Gabija complex (Fig. 6e). Apparently, the nuclease activity of the Gabija complex is negatively regulated by nucleotides with the exception that the low concentration of ATP/GTP promote the linear DNA substrate nuclease activity of the Gabija complex. We also measure the effect of non-hydrolysable analog of ATP on the cleavage activity of the Gabija complex (Supplementary Fig. 9). The results reveal that the cleavage activity of the Gabija complex is also inhibited by AMP-PNP, this indicates the binding of ATP is sufficient for inhibition of cleavage activity of the Gabija complex.

### Discussion

The Gabija bacterial defense system is a newly discovered bacterial immune defense system contains two components, GajA and GajB. The previous study shows that GajA is essential for bacterial immunity against certain bacteriophages and GajB is also indispensable for Gabija system. GajA is a sequence specific DNA cleaving enzyme and is



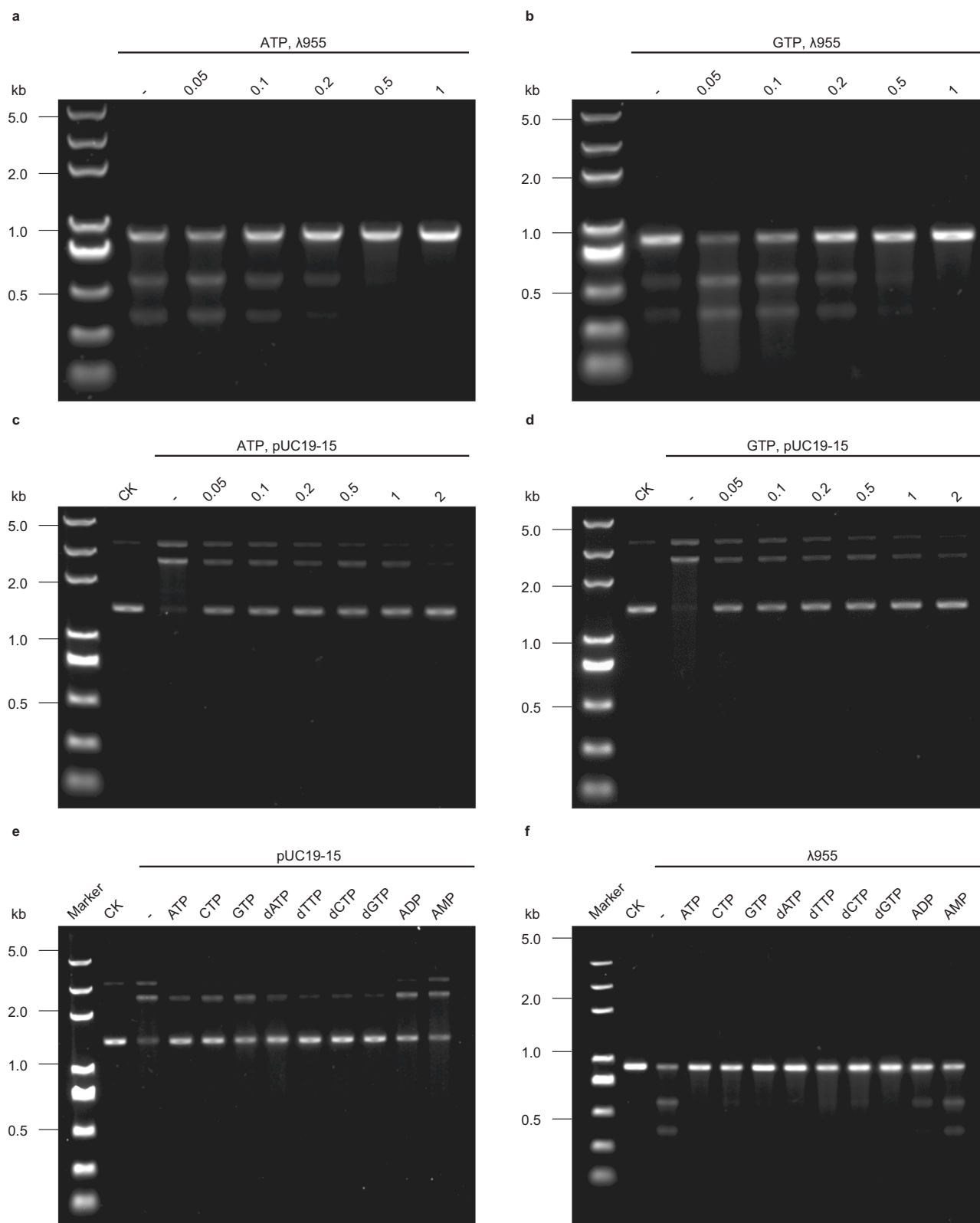
**Fig. 5 | The Gabija complex show substrate preference on supercoiled circular DNA to linear DNA substrate compared with GajA alone.** **a** Nuclease activity of GajA on λ955 linear DNA substrate; **b** Nuclease activity of GajA on supercoiled plasmids substrate; **c**, Nuclease activity of Gabija complex on λ955 linear DNA substrate; **d** Nuclease activity of Gabija complex on supercoiled plasmids substrate. λ955 is a 955 bp fragment from lambda DNA genome with the recognition sequence

(5'-AATAACCCGGATATT-3'). CK, control reaction without GajA or Gabija complex. The pUC19-15 plasmid contains the cleavage recognition sequence (5'-AATAACCCGGATATT-3') and the control is the pUC19 plasmid without the recognition sequence. 'N', 'L', and 'S' represent 'nicked', 'linear', and 'supercoiled' DNA, respectively. All the assays are repeated at least 3 times with similar results. Source data are provided as a Source Data file.

completely inhibited by 0.5 mM ATP. GajB was predicted as a UvrD-like SF1 helicase<sup>3</sup>. Our findings suggest GajB acts as a nucleotide sensor, at least when it is a part of the Gabija complex. Furthermore, our data revealed that with the aid of GajB, the substrate selectivity of the Gabija complex is more stringent, as it prefers supercoiled circular over linear DNA. In addition, compared to GajA, the Gabija complex is more sensitive to the concentration change of ATP. The Gabija complex began to cleave DNA below 2 mM ATP, while GajA began to cleave DNA below 0.5 mM ATP. Moreover, the low concentration (0.05 mM) of ATP promotes the digestion of the linear DNA substrate by the Gabija complex. The ATP concentration is about 3 mM, and the total nucleotide concentration is above 8.7 mM in *E. coli* at physiological conditions<sup>13</sup>. The Gabija complex nuclease activity is inhibited by 2 mM

ATP in vitro. The nuclease activity of the Gabija complex should be strictly inhibited at such high physiological concentrations of NTP. Indeed, co-expressing Gabija complex was found not toxic to *E. coli* during the process of purification of the Gabija complex.

Our findings here hence provide insights into the role of the Gabija complex in bacterial anti-phage immune surveillance (Supplementary Fig. 10). The Gabija complex nuclease activity is fully inhibited by nucleotides in the physiological state when bacteria grow normally. During phage invasion, the high-intensity DNA replication and transcription markedly reduced the concentration of nucleotides in the cell. In stage I, the ATP concentration is decreased to (0.5 mM-2 mM ATP) from 3 mM physiological concentration. Therefore, only the cleavage activity of the Gabija complex on supercoiled circular DNA is



**Fig. 6 | The nuclease activity of Gabija complex is regulated by nucleotides.** **a** Native agarose gel analysis of nuclease activity of Gabija complex on  $\lambda$ 955 linear DNA substrate with the increasing concentration of ATP; **b** Native agarose gel analysis of nuclease activity of Gabija complex on  $\lambda$ 955 linear DNA substrate with the increasing concentration of GTP; **c** Native agarose gel analysis of nuclease activity of Gabija complex on supercoiled plasmids substrate pUC19-15 with the increasing concentration of ATP; **d** Native agarose gel analysis of nuclease activity

of Gabija complex on supercoiled plasmids substrate pUC19-15 with the increasing concentration of GTP. **e** Effect of NTP, dNTP, ADP, and AMP on nuclease activity of Gabija complex in the presence of pUC19-15 substrate; **f** Effect of NTP, dNTP, ADP, and AMP on nuclease activity of Gabija complex in the presence of  $\lambda$ 955 substrate. All the assays are repeated at least 3 times with similar results. Source data are provided as a Source Data file.



activated. In stage II, the ATP concentration is decreased to below 0.5 mM, and the cleavage activity of the Gabija complex on linear DNA is activated. Thus, GajA of the Gabija complex cuts the phage DNA, stopping the infection due to phage replication failure. However, the specific sequence recognition is a double-edged sword. On the one side, it can selectively degrade invading bacteriophages and perform precise DNA substrate selection. Release of occupation of Gabija complex by less effective substrate could improve the efficiency of bacterial immunity. On the other side, bacteriophages would easily mutate away from Gabija defence. It is possible that the palindromic T/A sequences allow formation of a small hairpin in DNA. This would be preferentially formed in supercoiled DNA that is negatively supercoiled, consistent with supercoiled DNA substrate preference of Gabija complex. The recognition of palindromic structure may reduce the possibility of bacteriophage escape due to sequence mutations.

During the revision process, we became aware of three highly relevant preprints and papers that have been posted. The first paper, authored by Antine et al., describes the structural characterization of the Gabija complex<sup>14</sup>. They revealed the structural basis of the Gabija anti-phage defense complex and define a unique mechanism of viral immune evasion by GadI. The second paper, authored by Cheng et al., uncovers the mechanisms underlying nucleotide sensing by the Gabija complex<sup>15</sup>. GajB senses DNA termini nicked by GajA to hydrolyze (d)A/(d)GTP and results in an efficient abortive infection defense. The third preprint, authored by Fu et al., reports the structural assembly of Gabija complex<sup>16</sup>. These papers, in consistent with our findings reported here, define a detailed framework for understanding the anti-phage immune response by the Gabija complex.

## Methods

### Cloning, expression and purification of Gabija complex

Full-length coding sequence of GajA (residues 1-578) was cloned into PET-52b vectors between XmaI and BamHI sites harboring a Strep-tag II tag. Full-length coding sequence of GajB (residues 1-493) was cloned into PET-28a vectors between NdeI and XhoI restriction sites with a 6xHis tag. Constructs of full-length GajA and GajB were co-transformed into BL21 (DE3) cells, which were grown at 37 °C in LB medium containing 50 µg/ml kanamycin and 100 µg/ml ampicillin to an OD<sub>600</sub> of 0.6-0.7, and then protein expression was induced with 0.5 mM IPTG for 18 h at 16 °C. The cells were collected by centrifugation and resuspended in Buffer A (20 mM Tris-HCl, pH 7.5 at 25 °C, 300 mM NaCl and 2 mM β-mercaptoethanol), and then cells were lysed by sonication. The supernatant after centrifugation was applied onto a Strep-tactin Sepharose column pre-equilibrated in buffer A. After washing with 10-15 column volumes in buffer A, Gabija complex was eluted with buffer A plus 2.5 mM D-desthiobiotin. The elution was concentrated by Millipore Amicon Ultra-15 (30KD MWCO) and further purified by Superdex 200 column in Buffer B (10 mM Tris-HCl pH 7.5, 300 mM NaCl, 2 mM DTT). The peak fractions were collected and kept in Buffer C (50 mM Tris-HCl pH 7.5, 100 mM NaCl, 1 mM DTT, 0.1 mM EDTA, 50% glycerol and 0.1% Triton X-100). GajA and GajB mutants were constructed with site-directed mutagenesis. All the mutants were checked by SDS-PAGE (Supplementary Fig. 11). Sequences of primers are provided in the Supplementary Table 2.

### Size exclusion chromatography coupled to multi-angle light scattering

The purified GajA, GajB and Gabija complex were injected into an Agilent HPLC system with a Superdex 200 10/300 Increase column (GE) at 1 mg/ml in Buffer B at a flow rate of 0.5 ml/min. Eluent from the size exclusion column flowed directly to a static 18-angle light scattering detector (DAWN HELEOS-II) and a refractive index detector (Optilan T-REX) (Wyatt Technology) with data collected every second. Molar mass was determined using the ASTRA V software. Monomeric

BSA (1 mg/ml) (Sigma) was used for normalization of light scattering detectors and data quality control.

### ATPase/GTPase activity assay

The ATPase/GTPase activity was measured using QuantiChrom ATPase/GTPase assay kit (BioAssay Systems). All reactions were performed in the reaction buffer from the assay kit with various concentrations of ATP/GTP indicated and 0.2 µM protein at 37 °C for 30 min. Then follow the steps of the kit manual and the light absorbance of sample was measured at 620 nm. Nonlinear regression to the Michaelis-Menten equation and statistical analysis were performed using GraphPad Prism 8.

### DNA cleavage assay

The 20 nM various DNA substrate was digested by 0.2 µM protein in a final volume of 10 µl in DNA cleavage buffer (20 mM Tris-HCl, pH 9, 1 mM MgCl<sub>2</sub> and 0.1 mg/ml BSA). Reactions were incubated at 37 °C for 5 min and then terminated by the addition of 2 µl of 6×loading buffer containing 60 mM EDTA. Samples were analyzed via 1% native agarose gel electrophoresis and visualized using Bio-Rad ChemiDoc XRS+. After Gelred DNA dye (Sangon biotech) staining, the band of initial DNA substrate was measured and quantified using ImageJ software<sup>17</sup>. The DNA band in protein-free condition lane is taken as intact DNA control to assess the ratio of degraded DNA to intact DNA. The quantification bar graphs represent the average of three independent assays with error bars representing the standard error of the mean.

### Phage resistance assays

Phage cultivation and plaque assays were performed as below<sup>15,18</sup>. The *E. coli* BL21 strain harboring pQE82L vector inserted with various Gabija genes were grown to an OD<sub>600</sub> of 0.3 at 37 °C and then protein expression was induced with 0.2 mM IPTG. After cultured for 1 h, 500 µL bacterial cultures were added into 14.5 ml of 0.7% LB top agar and 15 mL mixture were poured onto LB plates containing 100 µg/ml ampicillin and 0.1 mM IPTG. A series of 10-fold dilutions of T7 phage solution was were spotted on double-agar plates. The plates were incubated at 37 °C overnight.

### Cryo-EM grid preparation and data collection

Freshly purified Gabija complex was diluted to 0.5 mg ml<sup>-1</sup> in 100 mM NaCl using low-salt buffer. The cryo-EM grids were prepared using an automatic plunge freezer Vitrobot Mark IV (ThermoFisher Scientific) under 100% humidity at 4 °C. Each aliquot of 3 µl Gabija complex was loaded onto the freshly glow-discharged homemade Graphene coated Au grid (RL2/L3, 300 mesh). The excess sample was blotted with filter paper for 3 s with a blot force of 3, and then the grid was vitrified by flash plunging into liquid ethane.

All datasets were automatically collected using SerialEM-3.7<sup>19</sup> through the beam-image shift data collection method<sup>20</sup> on a 200 kV Talos Arctica microscope (ThermoFisher Scientific) equipped with an energy filter and a K2 Summit detector (Gatan). Images were recorded under the super-resolution mode at a nominal magnification of 130,000× and a defocus value ranged between -1.0 to -3.0 µm, resulting in a pixel size of 1.0 Å per pixel. Each movie stack was exposed for 6.4 s with a total exposure dose of ~60 electrons per Å<sup>2</sup> over 32 frames.

### Data processing and reconstruction

Each movie stack was first binned and subjected to beam-induced motion correction and anisotropic magnification correction using MotionCor2-1.1.0<sup>21</sup>. The contrast transfer function (CTF) estimation was performed by CTFFIND4.1<sup>22</sup> using non-dose-weighted images. Further processing steps, including particle picking, 2D and 3D classification and auto-refinement were performed in Relion-3.08<sup>23</sup> using images with dose weighting. First, particles were automatic picked

using the Laplacian-of-Gaussian filter and processed with several rounds of reference-free 2D classification. Then, the best five images with clear structural features and different orientations were selected as templates for reference-based particle auto-picking. For Gabija complex, a total of 990,385 particles were picked from 3331 images and subjected to reference-free 2D classification. 615,824 particles with clear structural features were kept for further 3D classification with D2 symmetry, using the initial model generated by Relion-3.08. One among the five classes including 279,580 particles were selected for final 3D auto-refinement, resulting in a 3.6 Å density map. The local resolutions of final maps were calculated by ResMap-1.1.4<sup>24</sup>. Statistics for data collection and processing were summarized in Supplementary Table 1.

### Model building and refinement

For model building, the structure models of GajA and GajB were first predicted by AlphaFold 2<sup>25</sup>, respectively. Then the complete initial model of Gabija complex was generated by docking the predicted subunit models with a stoichiometric ratio of 4:4 into the reconstituted density map in UCSF Chimera-1.12<sup>26</sup>. After deleting the flexible areas without density map, a few iterative rounds of real-space refinement using PHENIX-1.13<sup>27</sup> and manual modification in Coot-0.8.8<sup>28</sup> were performed until no further improvement could be obtained. Statistics of the final models were summarized in Supplementary Table 1. Cryo-EM single-particle analysis of Gabija complex were summarized in Supplementary Fig. 3. GajA and GajB map to model fit for designated regions is shown in Supplementary Fig. 12. Comparison of AlphaFold 2 predicted model and final built model are shown in Supplementary Fig. 13. Figures were prepared in PyMOL and UCSF Chimera-1.12.

### Data availability

The cryo-EM map of Gabija complex has been deposited in the Electron Microscopy Data Bank (EMDB) under the accession code [EMD-35977](https://doi.org/10.2554/EMDB-35977). The structural coordinate of Gabija complex has been deposited in the RCSB Protein Data Bank (PDB) under the accession code [8J4T](https://doi.org/10.22541/au.171111111). Source data are provided with this paper.

### References

- Barrangou, R. et al. CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709–1712 (2007).
- Toock, M. R. & Dryden, D. T. The biology of restriction and anti-restriction. *Curr. Opin. Microbiol.* **8**, 466–472 (2005).
- Doron, S. et al. Systematic discovery of antiphage defense systems in the microbial pangenome. *Science* **359**, eaar4120 (2018).
- Gao, L. et al. Diverse enzymatic activities mediate antiviral immunity in prokaryotes. *Science* **369**, 1077–1084 (2020).
- Millman, A. et al. An expanded arsenal of immune systems that protect bacteria from phages. *Cell Host Microbe* **30**, 1556–1569 e1555 (2022).
- Rousset, F. et al. Phages and their satellites encode hotspots of antiviral systems. *Cell Host Microbe* **30**, 740–753.e745 (2022).
- Vassallo, C. N., Doering, C. R., Littlehale, M. L., Teodoro, G. I. C. & Laub, M. T. A functional selection reveals previously undetected anti-phage defence systems in the *E. coli* pangenome. *Nat. Microbiol.* **7**, 1568–1579 (2022).
- Tal, N. & Sorek, R. SnapShot: Bacterial immunity. *Cell* **185**, 578–578 e571 (2022).
- Cheng, R. et al. A nucleotide-sensing endonuclease from the Gabija bacterial defense system. *Nucleic Acids Res.* **49**, 5216–5229 (2021).
- Schiltz, C. J., Lee, A., Partlow, E. A., Hosford, C. J. & Chappie, J. S. Structural characterization of Class 2 OLD family nucleases supports a two-metal catalysis mechanism for cleavage. *Nucleic Acids Res.* **47**, 9448–9463 (2019).
- Schiltz, C. J., Adams, M. C. & Chappie, J. S. The full-length structure of Thermus scotoductus OLD defines the ATP hydrolysis properties

- and catalytic mechanism of Class 1 OLD family nucleases. *Nucleic Acids Res.* **48**, 2762–2776 (2020).
- Holm, L. & Rosenstrom, P. Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **38**, W545–549 (2010).
  - Buckstein, M. H., He, J. & Rubin, H. Characterization of nucleotide pools as a function of physiological state in *Escherichia coli*. *J. Bacteriol.* **190**, 718–726 (2008).
  - Antine, S. P. et al. Structural basis of Gabija anti-phage defense and viral immune evasion. *Nature*, **625**, 360–365 (2024).
  - Cheng, R. et al. Prokaryotic Gabija complex senses and executes nucleotide depletion and DNA cleavage for antiviral defense. *Cell Host Microbe* **31**, 1331–1334 (2023).
  - Fu, T., Yang, X., Shen, Z., & Lin, Q. Molecular basis of Gabija anti-phage supramolecular assemblies. *Res. Square* <https://doi.org/10.21203/rs.3.rs-3240148/v1> (2023).
  - Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–675 (2012).
  - Mazzocco, A., Waddell, T. E., Lingohr, E. & Johnson, R. P. Enumeration of bacteriophages using the small drop plaque assay system. *Methods Mol. Biol.* **501**, 81–85 (2009).
  - Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005).
  - Wu, C., Huang, X., Cheng, J., Zhu, D. & Zhang, X. High-quality, high-throughput cryo-electron microscopy data collection via beam tilt and astigmatism-free beam-image shift. *J. Struct. Biol.* **208**, 107396 (2019).
  - Ashkenazy, H. et al. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* **44**, W344–350 (2016).
  - Rohou, A. & Grigorieff, N. CTFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).
  - Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* **7**, e42166 (2018).
  - Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nat. Methods* **11**, 63–65 (2014).
  - Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
  - Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
  - Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 213–221 (2010).
  - Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 486–501 (2010).

### Acknowledgements

This research was funded by the National Key R&D Program of China (2022YFA1603702 and 2019YFA0508602 to T.W.) and the National Natural Science Foundation of China (22027810 to Y.H.; 92254304 and 92054108 to T.W.). The authors are indebted to Prof. Xinzhen Zhang (Institute of Biophysics, Chinese Academy of Sciences), Prof. Lixin Wan (Moffitt Cancer Center, USA) and Prof. Yu Cao (Shanghai Jiao Tong University, China) for valuable discussions, Dr. Rui Cheng (Huazhong University of Science and Technology, China) for providing phages and for helpful suggestions on phage assays, Fengrui Ren, Jiayun Li and Yining Zhu (Institute of Biophysics, Chinese Academy of Sciences) for technical assistance, and Xiaoxia Yu for support with SEC-MALS. Cryo-EM data collection was carried out at the Center for Biological Imaging (CBI), Core Facilities for Protein Science at the Institute of Biophysics, Chinese Academy of Sciences. We thank Boling Zhu, Lihong Chen, Xujing Li, Xiaojun Huang and other staff members at the CBI for their support with data collection.

## Author contributions

Y.H., Y.Z., M.X., K.D. and S.X. performed protein purification and biochemical experiments. J.M. and L.K. collected cryo-EM data and determined structure. X.Y. contributed to discussions. T.W. supervised experiments. Y.H. drafted the manuscript with input from all the others. All authors contributed to data analysis and manuscript preparation.

## 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/s41467-024-45173-7>.

**Correspondence** and requests for materials should be addressed to Yanwu Huo, Jun Ma or Taotao Wei.

**Peer review information** *Nature Communications* thanks Malcolm White and the other, anonymous, reviewer(s) 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.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024