A *Vibrio cholerae* viral satellite maximizes its spread and inhibits phage by remodeling hijacked phage coat proteins into small capsids

Caroline M. Boyd[1], Sundharraman Subramanian[2], Drew T. Dunham[1], Kristin N. Parent[2], Kimberley D. Seed[1]*

[1]Department of Plant and Microbial Biology, Seed Lab, University of California – Berkeley, CA 94720
[2]Department of Biochemistry and Molecular Biology, Parent Lab, Michigan State University, East Lansing, MI, 48824

*Kimberley Seed, 271 Koshland Hall, University of California – Berkeley, Berkeley, CA 94720, 510-664-7711, kseed@berkeley.edu

Email: kseed@berkeley.edu

Author Contributions:
Conceptualization: CMB, KDS; Investigation: CMB, SS, DTD; Formal analysis: CMB, SS, KNP; Data Curation: CMB, SS; Visualization: CMB, SS; Writing – Original Draft: CMB, KDS; Writing – Review & Editing: CMB, SS, DTD, KNP, KDS; Supervision: KDS; Project administration: KDS; Funding acquisition: CMB, DTD, KNP, KDS.

Competing Interest Statement: The authors declare no competing interests.

Keywords: phage, satellite, capsid, scaffold, transduction

Abstract

Phage satellites commonly remodel capsids they hijack from the phages they parasitize, but only a few mechanisms regulating the change in capsid size have been reported. Here, we investigated how a satellite from *Vibrio cholerae*, PLE, remodels the capsid it has been predicted to steal from the phage ICP1 (1). We identified that a PLE-encoded protein, TcaP, is both necessary and sufficient to form small capsids during ICP1 infection. Interestingly, we found that PLE is dependent on small capsids for efficient transduction of its genome, making it the first satellite to have this requirement. ICP1 isolates that escaped TcaP-mediated remodeling acquired substitutions in the coat protein, suggesting an interaction between these two proteins. With a procapsid-like-particle (PLP) assembly platform in *Escherichia coli*, we demonstrated that TcaP is a *bona fide* scaffold that regulates the assembly of small capsids. Further, we studied the structure of PLE PLPs using cryogenic electron microscopy and found that TcaP is an external scaffold, that is functionally and largely structurally similar to the external scaffold, Sid, encoded by the divergent satellite P4 (2). Finally, we showed that TcaP is largely conserved across PLEs. Together, these data support a model in which TcaP directs the assembly of small capsids comprised of ICP1 coat proteins, which inhibits the complete packaging of the ICP1 genome and permits more efficient packaging of replicated PLE genomes.

Introduction

Like the parasitism of bacteria by the phages that infect them, phages are parasitized by mobile genetic elements termed phage satellites. Satellites continue to be discovered across bacterial species and, despite the independent evolution of these elements, common themes have emerged (reviewed in (3). Universally, satellites are integrated genetic elements that are dependent on infection by a helper phage for their lifecycle. Helper phage infection triggers satellites' genome excision and replication. Following genome replication, satellites parasitize the helper phages' structural components and selectively package satellite genomes into proteinaceous shells comprised of coat proteins, called capsids (4). The piracy of essential phage structural proteins affords the phage satellites the luxury of reducing their coding capacity and, therefore, their genome size. As such, satellites' genomes can be packaged into smaller capsids compared to their helper phages. In all documented cases, the satellite encodes a strategy to direct the assembly of small capsids, a mechanism that excludes the complete packaging of the larger helper phage genome but permits the complete packaging of the smaller phage satellite genome (5–8). After the attachment of tails, also pirated from the helper phage, the mature virions harboring the satellite genome are released from the cell. Here, they transduce and integrate the satellite genome into neighboring susceptible bacteria.

PLEs (Phage inducible chromosomal island-Like Elements) are satellites of the lytic *V. cholerae* phage ICP1 (9). PLEs are one of four families of phage satellites that have been identified to date (3). In addition to their specificity to *V. cholerae* and unique complete inhibition of their helper phage, PLEs stand apart from other satellites in their genetic composition. They encode genes with similar functions, but without sequence identity, to those in other satellites. To date, 10 genetically distinct but related PLEs have been identified (10), with PLE1 being the most well-studied. PLEs, like other satellites, have a smaller genome (~18 kb) than the phage they parasitize ICP1 (~125 kb), and are dependent on their helper for excision (11, 12), replication (13), and virion production (1). PLE1 has been shown to produce virions with small ~50 nm wide capsids (Figure 1A), while ICP1 produces virions with large ~80 nm wide capsids (Figure 1B). Based on similarities to other capsid-remodeling satellites and the evidence that depletion of ICP1's coat protein by CRISPRi during infection results in reduced PLE transduction (1), it is hypothesized that PLE capsids are comprised of ICP1 coat proteins. However, the mechanism that PLE1 uses to achieve this capsid remodeling has yet to be explored.

As capsids for large double-stranded DNA viruses assemble through a stepwise pathway, remodeling must occur at the first stage, nucleation of the procapsid. Procapsids are the empty shell capsid precursors into which DNA is packaged (14). Procapsid size, referred to by a T number (which is the number of coat subunits divided by 60), and assembly are regulated by scaffolding proteins that guide coat proteins into their correct orientation around the pre-formed portal complex (15, 16). These capsid scaffolds can either be separately encoded proteins or contained within a domain of the coat protein, as is

seen in the phage HK97 (17, 18). So, to alter the size of the capsids, satellites must regulate the size of the procapsid.

Highlighting the convergent evolution across unique satellite families, four distinct strategies to produce small capsids have been characterized. Two representatives from different families of satellites exploit the role of scaffolds in the capsid assembly process to promote the formation of smaller capsids. First, some of the *Staphylococcus aureus* Pathogenicity Islands (SaPIs) encode an alternative internal scaffold, CpmB, that binds to the helper phage's coat proteins inside of the assembling procapsid (19). CpmB increases procapsid curvature, resulting in the smaller procapsid (19). The second example of capsid remodeling was described in the *Escherichia coli* phage satellite, P4 (5). Analogous to the satellite-encoded scaffold from SaPIs, P4 encodes an alternative scaffold, Sid, that functions to regulate assembly of the smaller procapsid, however, Sid binds to the outside of the procapsid (2, 20). Sid proteins form an external cage around the helper phage's coat proteins which promotes the assembly of the small procapsids (2). An alternative strategy to redirect capsid size is found in a subfamily of SaPIs, including SaPIbov5, which encodes a homolog of their helper phage's coat protein that exclusively forms the pentamers while the helper phage's coat proteins form the hexamers of the small capsid (7). These coat proteins encode their scaffolds within a domain of the coat, like HK97 (7). Exactly how the satellite-derived coat pentamers promote the assembly of small icosahedral capsids instead of the larger prolate capsid of the helper is not fully understood. The last characterized strategy used by satellites to make small capsids was recently discovered in the aptly named cfPICIs (capsid forming phage inducible chromosomal islands), which avoid altering the assembly of their helper phage capsids by instead encoding their own structural components of the capsid and stealing only the phage tails to construct satellite virions (21). Interestingly, despite the divergence in capsid remodeling strategies, there is convergence on capsid size where each satellite assembles icosahedral capsids with 240 subunits, or T=4 capsids. PLE's small capsids are similar in size to these T=4 capsids but PLE does not encode obvious homologs of CpmB or Sid nor capsid homologs, making it unclear how capsid remodeling is achieved in this divergent satellite. Notably, other helper phages have similar genome sizes, ~30-45 kb, while ICP1 has a significantly larger genome, ~125 kb and is the only known helper packaged into T=13 capsids (22, 23). This may suggest that PLE uses a unique mechanism to remodel the ICP1 coat proteins into the smaller ~50 nm wide PLE-sized capsids made from a considerably smaller number of subunits.

Here, we studied how PLE remodels the ICP1 capsid. We found a PLE-encoded protein that has an essential role in making small capsids, named TcaP for its tiny capsid phenotype. TcaP's activity was necessary and sufficient to make small capsids, which inhibited the production of infectious ICP1 virions and increased the efficiency of PLE transduction. We studied TcaP's mechanism of capsid remodeling using a heterologous procapsid-like-particle (PLP) assembly platform in *E. coli* and cryogenic electron microscopy (cryo-EM). These data revealed TcaP as an external scaffolding protein that directs the assembly of ICP1 coat proteins into small, PLE-sized capsids. Finally, we analyzed the known PLE variants and found TcaP is largely conserved in PLEs. Our work uncovered the mechanism of capsid remodeling in the phage satellite PLE and provides the first example of small capsid production benefitting satellite transduction.

Results

PLE encodes a single gene product, TcaP, that is necessary and sufficient to direct the assembly of small capsids

Given the observation that PLE virions have smaller capsids than ICP1 virions (Figure 1A&B), we hypothesized that PLE encodes a single protein responsible for assembling these smaller capsids. We reasoned that the generation of small capsids would limit the amount of DNA that could be accommodated within the capsids and block the complete packaging of the larger ICP1 genome, reducing ICP1 plaquing. In line with this prediction, other satellites' proteins that redirect capsid assembly are inhibitory to their helper phages for this reason (5, 6, 24). Using this logic and PLE1 as the representative PLE, we identified one candidate PLE protein, TcaP (previously Orf17 (AGG09411.1)), as a putative capsid remodeling protein.

First, to assess the capsid remodeling capacity of TcaP, we used transmission electron microscopy (TEM) to evaluate the morphology of virions produced after a single round of ICP1 infection in a PLE(-) *V. cholerae* strain expressing *tcaP* from a plasmid (p*tcaP*). Lysates collected from the p*tcaP* host showed an abundance of small, ~50 nm round particles reminiscent of PLE-sized procapsids (Figure 1C). Rarer particles in these samples resembled virions produced from a PLE(+) infection with small capsids and attached tails, though apparently lacking DNA (Figure S1A). In comparison, many virions produced from the empty vector control were ICP1-sized virions with large capsids, as expected (Figure 1D). Other particles observed in the control samples were full ICP1 capsids without tails, empty ICP1 procapsids or empty expanded capsids (Figure S1B). Importantly, these data demonstrate that *tcaP* expression during ICP1 infection, in the absence of other PLE-encoded products, results in the production of small capsids.

In agreement with the expectation that small capsid formation would be inhibitory to ICP1, p*tcaP* inhibited ICP1 plaque formation by 100-fold (Figure 1E). We then asked if ICP1 could escape TcaP-mediated inhibition through genetic mutation. Rare plaques isolated on the p*tcaP* strain were picked and serially propagated on the restrictive strain. Two escape phages were whole genome sequenced, and each harbored a single nonsynonymous mutation in the gene encoding the coat protein, resulting in substitutions R223H or E234K (Table S1). Substitutions at unique but proximal positions in coat protein from independent escape phages suggest that TcaP interacts with the coat protein, an expected feature of TcaP's predicted role as a capsid remodeling protein. Collectively, these genetic escapes coupled with TcaP's inhibition of ICP1 plaquing, which is accompanied by the production of small capsids, strongly support that TcaP is a capsid remodeling protein.

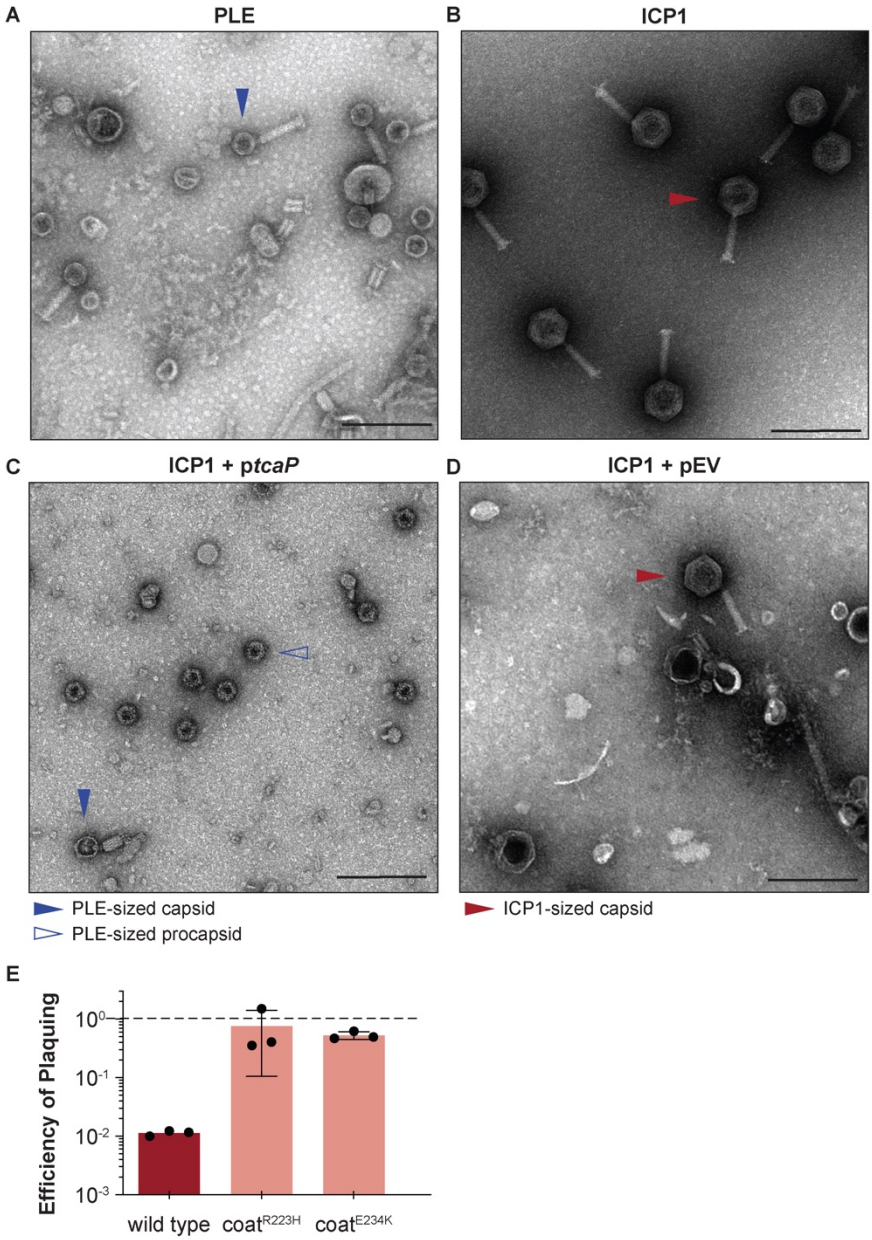


Figure 1. PLE encodes an inhibitory protein, TcaP, which modifies ICP1's capsid assembly process to produce small capsids

A-B) Representative transmission electron micrographs (TEMs) from 3 independent biological replicates show A) PLE virions have small, ~50 nm wide capsids and long contractile tails while B) ICP1 virions have large, ~80 nm wide capsids and long contractile tails. Scale bars are 200 nm.

C-D) Representative TEMs of lysates produced from ICP1 infection of *V. cholerae* expressing C) TcaP (p*tcaP*) or D) an empty vector (pEV). Arrowheads show exemplary capsids and their sizes according to the legend. Scale bars are 200 nm.

E) Efficiency of plaquing of wild type ICP1 or escape phages harboring the substitution indicated in the coat protein on *V. cholerae* expressing TcaP relative to an empty vector control. Each dot represents a biological replicate, bars represent the mean, and error bars show the standard deviation. The dotted line indicates an efficiency of plaquing of 1 where the expression of TcaP is not inhibitory to plaque formation.

While expression of TcaP during ICP1 infection was sufficient to redirect capsid size, we wanted to address if TcaP was necessary for PLE to generate particles with small capsids. Other capsid remodeling

satellites have been shown to use either a single protein (5, 24) or two proteins that act in concert to direct small capsid assembly (6). To assess if TcaP is required to assemble ICP1's coat protein into small capsids, virions resulting from ICP1 infection of wild type PLE and PLE∆*tcaP* were concentrated by centrifugation and their morphology was assessed by TEM. The wild type PLE control produced PLE-sized virions with small capsids as expected and no ICP1-sized virions were observed (Figure 2A). The PLE∆*tcaP* strain produced only virions with large, ICP1-sized capsids and no PLE-sized particles (Figure 2B), a phenotype that was rescued following the expression of *tcaP in trans* (Figure 2C). Importantly, *tcaP* is dispensable for PLE-mediated inhibition of ICP1 (25), indicating that the increase in capsid size in the absence of TcaP is specific to its role in capsid-size redirection and not a result of ICP1 escaping PLE. These data demonstrate that TcaP is necessary for PLE's redirection of ICP1 coat protein into small capsids and that PLE does not encode a redundant mechanism for capsid remodeling.

To further address the effect of deleting *tcaP* from PLE, we measured the transduction of PLE's genome following ICP1 infection in PLE∆*tcaP* hosts using a previously described assay using an antibiotic resistance marker in PLE (9). Interestingly, in the absence of *tcaP*, PLE transduction was 10-fold less efficient than wild type (Figure 2D). The defect in the transduction of the PLE∆*tcaP* strain was largely restored with p*tcaP* (Figure 2D). Together with the morphological data of transducing particles, these data suggest that TcaP-mediated capsid remodeling facilitates more efficient horizontal spread of the PLE genome to recipient *V. cholerae*. PLE is the first satellite to show dependency on small capsids for efficient transduction.

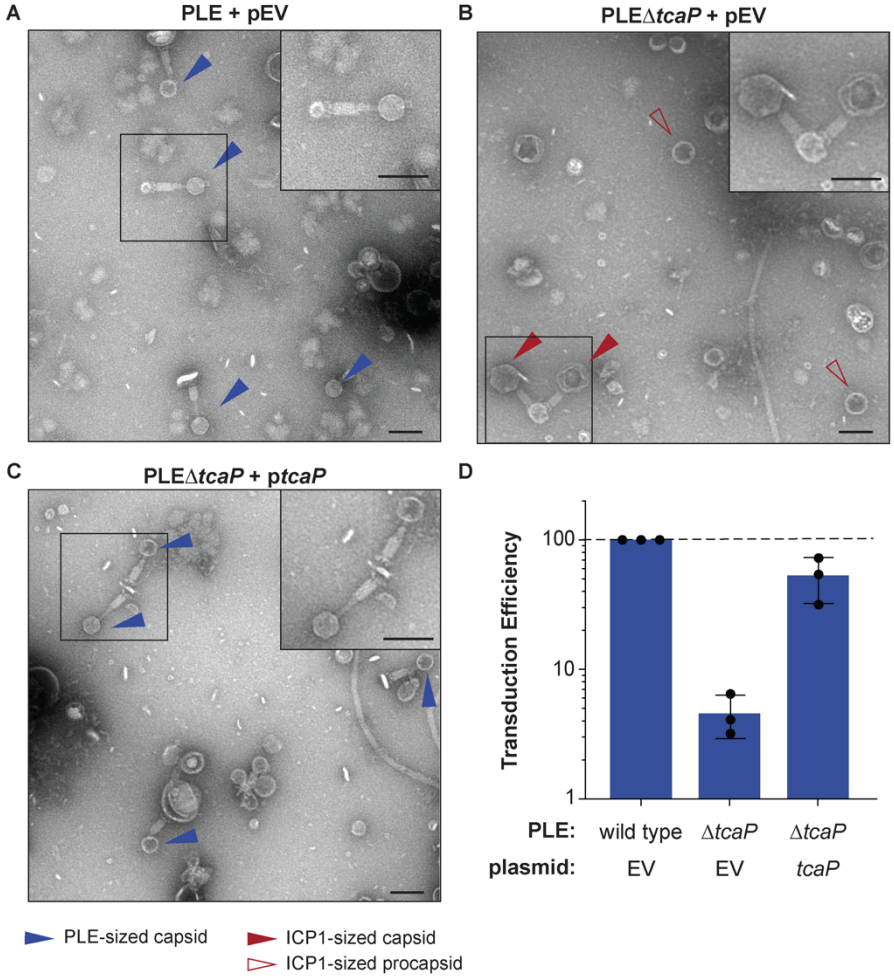


Figure 2. TcaP is the only PLE-encoded factor necessary for directing small capsid assembly, which is required for efficient PLE transduction.

A-C) Representative TEMs from 3 independent biological replicates of lysate from ICP1-infected strains of *V. cholerae* with wild type PLE or PLE∆*tcaP,* as indicated, carrying either the empty vector (EV) or a

vector expressing TcaP (*tcaP*). Insets are enlarged regions of the images highlighting representative particles, and arrowheads indicate capsid types and sizes, as described in the legend. Scale bars are 100 nm.
D) Quantification of PLE genome transduction for the strain indicated represented as the transduction efficiency relative to wild type PLE with an empty vector (pEV). Each dot represents a biological replicate, bars represent the mean, and error bars show standard deviation. The dotted line indicates an efficiency of 100%.

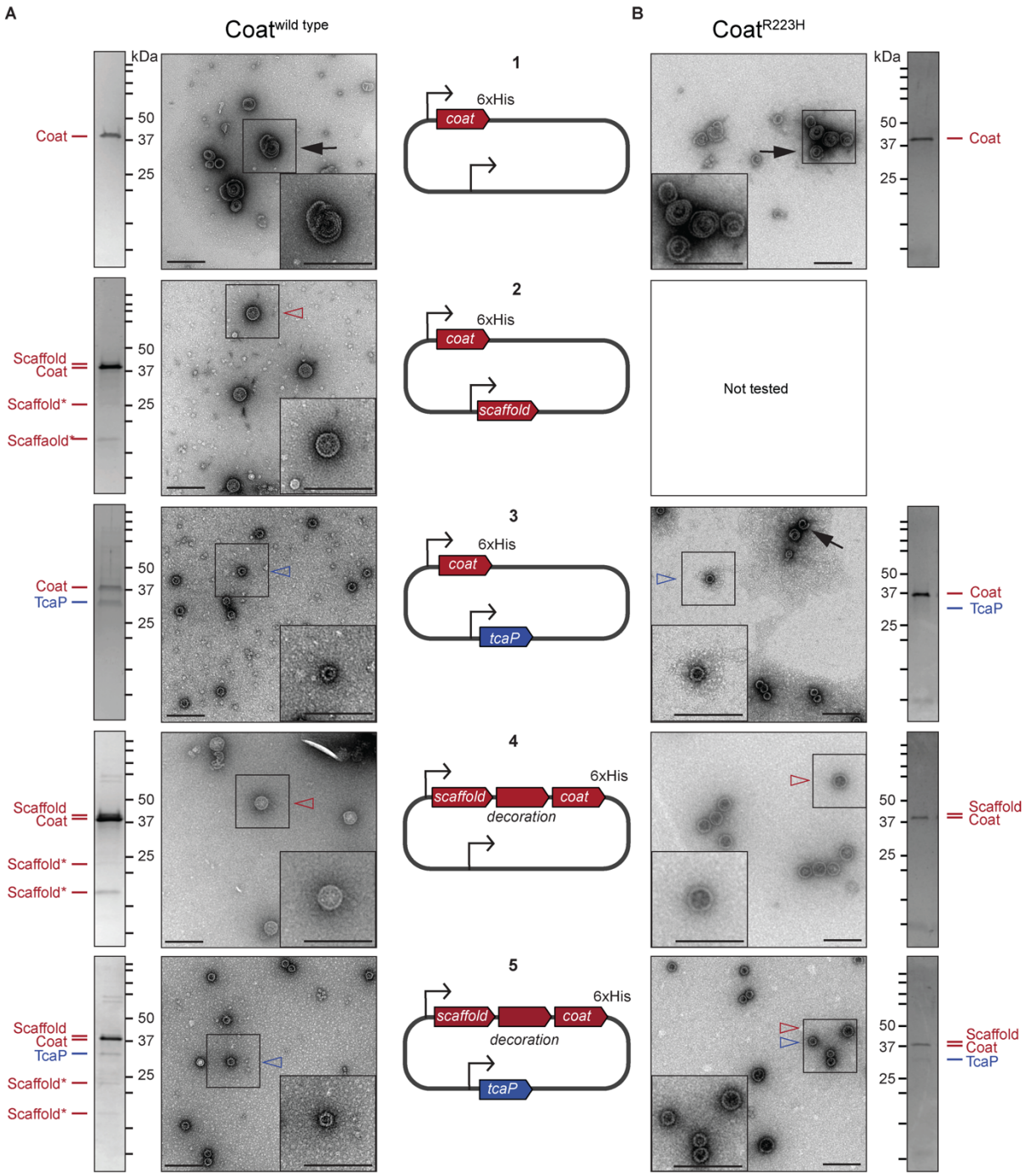
PLE-encoded TcaP is a *bona fide* capsid scaffold

Capsid scaffolding proteins are characteristically responsible for directly promoting the assembly of coat proteins and controlling capsid size (16). TcaP expression results in the formation of small capsids during ICP1 infection (Figure 1C), which suggests that it has scaffolding activity. However, it is possible that TcaP does not act directly on the coat proteins as a scaffold, but rather interferes with ICP1's capsid morphogenesis pathway in some way that decreases capsid size. To directly address if TcaP is a scaffold, we set up a heterologous procapsid-like-particle (PLP) assembly platform in *E. coli* similar to those previously described (26–28). Briefly, we co-expressed ICP1's coat and putative scaffolds from either ICP1 or PLE and monitored the production of PLPs. By adding a C-terminal six-histidine (6xHis) tag to the coat protein, we were able to purify coat-containing complexes by affinity chromatography, examine their protein content by SDS-PAGE/Coomassie staining, and their morphology by TEM. First, as a control, coat::6xHis (referred to as "coat" for simplicity) was expressed and purified. The coat proteins eluted as complexes, but they were irregular in size and shape, often forming spirals (Figure 3A.1), as is expected for coat proteins in the absence of scaffolds (16). As a second control, we confirmed that ICP1's putative scaffold could assemble coat proteins into uniform, ICP1-sized PLPs (Figure 3A.2). The presence of scaffold in these PLPs was supported by SDS-PAGE analysis in which we observed a band corresponding to the predicted size of the scaffold (39.3 kDa), indicating, as expected, that the scaffold co-eluted with coat-containing complexes (Figure 3A.2). Interestingly, two smaller bands of approximately 23 and 17 kDa also appeared in these samples. The size of these bands is consistent with the cleavage of the scaffold resulting in two fragments. As scaffolds are generally not a part of the mature capsid, they are removed from the procapsid by self-cleavage (29, 30) or cleavage by a protease (31, 32). Indeed, in the presence of protease inhibitors during purification, these bands were not observed, but the same-sized particles were produced, suggesting scaffold cleavage is not a prerequisite for assembly (Figure S2A). Analysis of ICP1 PLPs by mass spectrometry (LC-MS/MS) further supported the presence of ICP1's putative scaffold (Table S2). Together, these data confirm the predicted role of ICP1's scaffold and demonstrate that PLPs can be assembled and purified using this platform.

As other known satellite systems have varying dependencies on the phage-encoded scaffold and/or additional satellite-encoded factors for the assembly of small procapsids (6, 30, 33, 34), we wanted to address what proteins were required to make small capsids in the ICP1-PLE system. To test whether TcaP requires ICP1's scaffold to produce small PLPs, we co-expressed coat with either just TcaP, or with both TcaP and ICP1's scaffold simultaneously. First, analysis of the stained SDS-PAGE gel following co-expression of coat and TcaP showed that TcaP bound to coat and was co-eluted (Figure 3A.3). TEMs showed that TcaP was sufficient to assemble ICP1's coat into PLPs that were homogenously PLE-sized. These data demonstrate that TcaP has scaffolding activity for ICP1's coat protein and that it does not require additional factors to assemble small PLPs. Next, we tested if the scaffolding activity of TcaP could compete with the activity of ICP1's scaffold. To achieve the triple protein expression construct, the scaffold through coat was cloned downstream of one of the promoters in the plasmid as it occurs natively in ICP1's genome. Here, the putative decoration protein, Gp123, is encoded by the gene between the scaffold and coat. The decoration protein was not expected to incorporate into the PLPs as decoration proteins are typically added after genome packaging and capsid expansion exposes their binding site, as has been shown for phage Lambda (35). In line with this, the stained gel of the resulting purified ICP1-sized PLPs indicated that while the ICP1 scaffold co-eluted with coat-containing complexes as expected, the decoration protein was produced but not incorporated into the assembled particles (Figure 3A.4 and Figure S3). Interestingly, when we co-expressed ICP1 coat, decoration, and scaffold along with TcaP, we did not observe any large particles, suggesting TcaP is dominant over ICP1's scaffold (Figure 3A.5). SDS-PAGE analyses of particles produced in the presence of ICP1's scaffold and TcaP confirmed the

presence of TcaP as well as the phage-encoded scaffold and its cleavage products, suggesting TcaP does not block the incorporation of ICP1's scaffold into procapsids (Figure 3A.5). This experiment was repeated in the presence of protease inhibitors which eliminated the ICP1 scaffold cleavage products but reproduced the small particle morphology (Figure S2B). Collectively, these data demonstrate scaffolding activity for both ICP1's scaffold and PLE's TcaP and show that TcaP can assemble small PLPs in the absence or presence of ICP1's scaffold.

Having observed that TcaP is sufficient to make small PLPs in the heterologous assembly platform as well as when expressed from a plasmid during ICP1 infection, we next sought to determine if coat^{R223H} and coat^{E234K}, the previously identified *in vivo* genetic escapes of TcaP activity (Figure 1E), demonstrate escape from TcaP-mediated remodeling in the assembly platform. The coat^{E234K} variant was not assembly competent in our heterologous assembly platform (Figure S4), so we only continued with the coat^{R223H} variant. As expected, in the absence of any scaffolding proteins, coat^{R223H} formed similar spiral complexes as those observed with the wild type coat (Figure 3B.1). These data demonstrate that the R223H substitution does not compromise the protein's ability to bind to itself, nor does the substitution provide a means for the coat to form PLPs independent of a scaffold. Next, we expressed coat^{R223H} with TcaP, anticipating we would observe the formation of large ICP1-sized PLPs because this substitution was sufficient for ICP1 to avoid TcaP's inhibitory activity (Figure 1E). Unexpectedly, TcaP still directed the assembly of small PLPs comprised of coat^{R223H} (Figure 3B.3). However, TcaP was not robustly evident in the purified complexes as assessed by SDS-PAGE, and some particles were spirals, similar to those seen when the coat is expressed without a scaffold (Figure 3B.3). These data suggest that TcaP's scaffolding activity was partially compromised when the coat protein carried the R223H substitution. We hypothesized that with ICP1's scaffold, the coat^{R223H} proteins could be assembled into ICP1-sized PLPs. Leveraging the dual expression from a single promoter (Figure 3A.4), we confirmed that the ICP1 scaffold alone directed the assembly of ICP1-sized PLPs with coat^{R223H} (Figure 3B.4) but, we observed a mix of PLE and ICP1-sized PLPs following addition of TcaP (Figure 3B.5). We can attribute this phenotype to the substitution in the coat and not to the presence of protease inhibitors used during the purification as the addition of protease inhibitors during purification of the coat$^{wild\ type}$ co-expressed with the ICP1 scaffold and TcaP resulted in only small PLPs (Figure S2B). These data suggest that the R223H substitution in ICP1's coat can only partially escape TcaP-mediated small capsid formation, but *in vivo* this level of escape is sufficient to allow for approximately equal levels of viable progeny production as is seen in the absence of TcaP.



Name	Gene product (MW)	Accession
Coat	Gp122::6xHis (38.4 kDa)	AXY82215.1
Scaffold	Gp124 (39.3 kDa)	AXY82217.1
Decoration	Gp123 (13.2 kDa)	AXY82216.1
TcaP	Orf17 (33.2 kDa)	AGG09411.1

Figure 3. The size of procapsid-like-particles is determined by ICP1 and PLE scaffolds. Representative TEMs and Coomassie stained SDS-PAGE analyses of resulting affinity-purified procapsid-like-particles (PLPs) produced in the heterologous assembly platform in *E. coli* from 2-3 independent biological replicates following expression of A) coat$^{wild\ type}$ or B) coat^{R223H} proteins encoded on plasmids as shown in the diagrams in the central panel (numbered 1-5). ICP1-encoded genes are shown in red and PLE-encoded genes are shown in blue. Bent arrow icons indicate P_{tac} promoters. 6xHis

represents the tag fused to the C-terminus of the coat. Protein standards are indicated by black tick marks and a subset are marked by their sizes in kDa as indicated (standard range: 250, 150, 100, 75, 50, 37, 35, 20, 15, 10 kDa). Protein bands of interest are indicated by colored tick marks and labels (see legend for calculated molecular weights and accession numbers of these proteins, complete gene and protein information is provided in Table S4). In the absence of protease inhibitors, ICP1's scaffold appears to be cleaved and the resulting cleavage products are indicated by scaffold*. Protease inhibitors were included in all coat^{R223H} purifications. TEM insets are enlarged sections of the images highlighting representative particles and arrowheads indicate capsid types and sizes, as described in the legend. Scale bars are 200 nm.

Cryo-EM reveals TcaP is an external scaffold

The data obtained from the assembly of PLPs in *E. coli* demonstrate that TcaP has scaffolding activity; however, these data do not distinguish between TcaP acting as an internal or an external scaffold. To identify the localization of TcaP in PLE PLPs and address how TcaP interacts with the coat protein to assemble small capsids, we validated the protein content of PLE PLPs produced from co-expression of TcaP and coat by LC-MS/MS and subjected them to cryogenic electron microscopy (cryo-EM). As anticipated, LC-MS/MS confirmed PLE PLPs were comprised of coat and TcaP (Table S3). Representative micrographs from negatively stained samples (Figure 4A) and vitrified particles (Figure 4B) show a distinctly bumpy surface on the PLE PLPs, suggestive of external proteins. 379,643 particles were used for icosahedral reconstruction of the PLE PLP particle, with a final resolution of 3.4 Å (Figure 4C). The external density of the PLPs corresponded to TcaP, demonstrating that it functions as an external scaffold. Internal density was assigned to coat proteins with a canonical HK97-like fold, which exist as pentamers and hexamers within the 48 nm, T=4 icosahedral PLP. We used AlphaFold2 (36) to predict the structures of the coat protein and TcaP, which were fitted into density from the Cryo-EM and aided in modeling nearly all of the coat protein and approximately half of TcaP (residues 34–172) for which only a few side chains could be modeled. As expected, the A-domains of the coat proteins were centered in both pentamers and hexamers. TcaP dimers meet and form trimeric interactions at the three-fold axes between hexamers. Broadly, the reconstruction of PLE PLPs shows TcaP functions as an external scaffold and provides context for how TcaP's higher order structure regulates the number of coat proteins that can be accommodated within the assembling PLP.

Further, the reconstruction reveals details of the residues that were substituted in phages that escape TcaP-mediated capsid remodeling (R223H and E234K). Both residues are found in the A-domain which is oriented near the center of the hexamer where TcaP binds (Figure 4D). The arginines at position 223 were clearly visible in the map. Due to the two-fold symmetry in the hexamer, there are identical interactions between each half of the hexamer and the TcaP dimer. Importantly, there are three distinct interactions between TcaP and R223 (Figure 4E). First, R223 from coat$_D$ is in proximity to TcaP's residues 87-93, while coat$_B$ has R223 positioned near TcaP's residues 63-70, and R223 from coat$_C$ lies close to the second TcaP subunit. In this third interaction, the negatively charged aspartic acid at position 218 from the coat coordinates with the positive charges from the arginine residues from coat (R223) and TcaP (R133) and creates a salt bridge with residues measuring less than 4 Å apart. The electrostatics of the complex show TcaP's R133 fitting into a negatively charged pocket on the coat (Figure 4F). The side chain for the other residue substituted in the escape phage, E234, could not be confidently modeled, but its localization within the capsid is outside of the TcaP binding region. The exact nature of E234 is not yet clear, but perhaps it has a role in stabilizing the A-domain. Together, these data support the conclusion that ICP1 escapes TcaP's scaffolding activity by altering the coat to affect TcaP's binding site.

We next asked if these coat substitutions were observed in natural isolates of ICP1, as may be expected if ICP1 is under selective pressure to escape PLE-mediated capsid remodeling in nature. When we compared the coat alleles from 67 genetically distinct ICP1 isolates we found no differences within the A-domain of the alleles. These data show that ICP1 does not naturally encode mutations in the A-domain that would escape TcaP-mediated capsid remodeling.

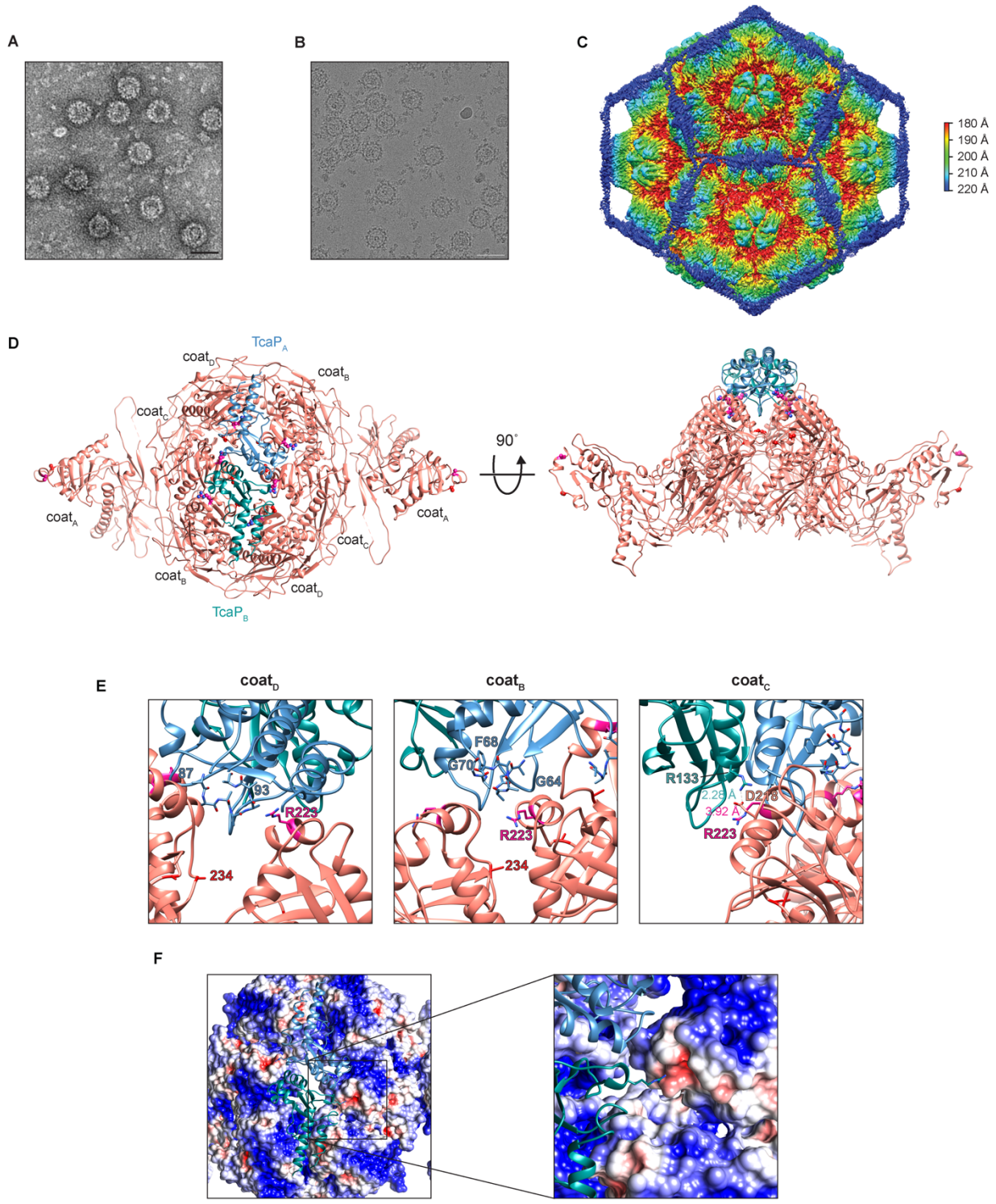


Figure 4. Cryo-EM reveals TcaP is an external scaffold.
A-B) Representative micrographs from 3 independent biological replicates from A) transmission electron and B) cryo-electron microscopy of PLE PLPs produced from co-expression of ICP1's cot and PLE's TcaP. Scale bars are 50 nm.
C) Isosurface reconstruction of PLE PLPs, resolved to 3.4 Å, colored radially (Å) as indicated on the legend.

D) Ribbon model of the solved structure for eight coat proteins (salmon), two from the adjacent pentamers (coatA) and six in a hexamer ($coat_{B,C,D}$) , and two partial TcaP proteins (blue and teal). Substituted residues in coat that escape TcaP-mediated remodeling are shown as spheres (pink for R223 and red for E234) and lie along the region of the hexamer where TcaP binds. Side chains that could not be fully resolved are modeled as alanines.
E) Details of the interactions between TcaP and the coat subunits. Residues in TcaP that contact coat are shown as sticks. For TcaP, the numbers, colored according to their chain, indicate the first and last residue within the region of contact. Distance measurements between residues are shown as dashed lines and measured in Å. Side chains that could not be fully resolved are modeled as alanines.
F) Electrostatic potential surface representation of the coat with a ribbon diagram of TcaP show the negative pocket on $coat_C$ that is filled by the positively charged arginine from TcaP. Electrostatic potential is calculated using APBS and the coloring is from –5 kT/e (red) to +5 kT/e (blue).

TcaP is conserved in PLEs

Having established that PLE1's TcaP acts as an alternative external scaffold that directs the assembly of ICP1's coat protein into small capsids, we wanted to assess the conservation of *tcaP* across genetically distinct PLEs. While PLE1 is the most well-studied PLE, there are nine additional PLEs that have been discovered to date (10). Therefore, we sought to evaluate if TcaP was conserved in all PLEs or if alternative strategies of capsid modification may exist amongst this family of satellites. Indeed, every PLE encodes a *tcaP* allele in the same locus, which encodes for proteins sharing 62-100% amino acid identity with $TcaP^{PLE1}$ (Figure 5A, Figure S5A). Largely, the regions that contact coat proteins in the PLP are conserved (Figure S5C). The conservation of this protein in the 10 PLEs suggests that all PLEs share the TcaP-mediated capsid remodeling strategy.

Curiously, the most divergent *tcaP* allele, $tcaP^{PLE5}$, is 237 nucleotides (~26%) shorter than $tcaP^{PLE1}$ due to an apparent truncation of the N-terminal coding region. However, a nucleotide alignment of the $tcaP^{PLE1}$ sequence with the $tcaP^{PLE5}$ 5'-UTR through the end of the coding sequence revealed two small deletions, one of which results in a frameshift and premature stop codon in the $tcaP^{PLE5}$ allele (Figures 5A and 5B). The alternative start site downstream of the deletions would restore the original reading frame for $tcaP^{PLE5}$ (Figure 5B). Given the conservation of full-length *tcaP* alleles in other PLEs, and the structural data highlighting interactions between the $TcaP^{PLE1}$ with ICP1's coat proteins, we wanted to assess if the truncated $TcaP^{PLE5}$ was functional to redirect ICP1 virion assembly.

To start, we generated PLE particles from the wild type PLE5 strain carrying an empty vector and assessed their morphology by TEM. We found the PLE5 virions did not have small capsids, but rather they had large, ICP1-size capsids (Figure 5C), similar to virions from the PLE1Δ*tcaP* strain (Figure 2B). Consistent with this, when we directly addressed the scaffolding activity of the truncated allele by expressing it *in trans* during ICP1 infection and assessing plaque formation, we found that unlike p$tcaP^{PLE1}$, which inhibited ICP1 plaquing (Figure 1E), p$tcaP^{PLE5}$ did not inhibit ICP1 plaque formation (Figure 5E), supporting the conclusion that $TcaP^{PLE5}$ is not a functional scaffold. The reconstruction of the PLE1 PLPs (Figure 4C) revealed residues 63-70 in $TcaP^{PLE1}$ that specifically contact $coat_B$. That region is notably lacking from the truncated TcaP (Figure S5C). Further, we predict that a TcaP protein missing 79 amino acids would not be long enough to form the scaffolding cage around an assembling T=4 procapsid explaining the lack of scaffolding activity for this TcaP variant.

Next, we wanted to address the transduction efficiency of PLE5. The efficiency of PLE1's transduction decreased in the absence of TcaP (Figure 2D), demonstrating an advantage to packaging the PLE genome into smaller capsids. To test if PLE5 had a transduction defect due to its inability to assemble small capsids, we measured the transduction efficiency of PLE5 in the presence of an empty vector or p$tcaP^{PLE1}$. Previous work reported similar transduction efficiency for PLEs1-5 (9), and we recapitulated those results here, showing PLE5 has only a subtle, ~2-fold decrease in transduction efficiency relative to PLE1 (Figure 5F). As this seemed in conflict with the PLE1Δ*tcaP* data, where there was over a 10-fold decrease in PLE1 transduction without TcaP (Figure 2D), we wanted to directly examine the effect of small capsid production on PLE5 transduction. First, we confirmed that p$tcaP^{PLE1}$ in PLE5(+) *V. cholerae* during ICP1 infection directed the assembly of small capsids (Figure 5D), which supported our hypothesis

that the unmodified capsids produced by PLE5 are the result of this PLE encoding a nonfunctional TcaP allele. Next, we found that PLE5 transduction increased with p*tcaP*^{PLE1}, exceeding the transduction efficiency of PLE5 and PLE1 (Figure 5F). This result suggests that the capacity for transduction is different for PLE1 and PLE5, but still nonetheless demonstrates that small capsids promote PLE transduction.

Since both PLE1 and PLE5 transduction efficiency was higher when small capsids were made, this raised the question of PLE5's maintenance in the *V. cholerae* population while lacking a functional scaffold. A recent analysis of sequenced *V. cholerae* isolates outlined a pattern where a PLE variant emerges, often rises to dominance, then is replaced by a different variant (10). This pattern revealed that after its dominance from before 1960 until its disappearance in 1991 (for which there are 21 *V. cholerae* isolates harboring identical PLE5 sequences), PLE5 re-emerged in 2016 and 2017. Genetic drift within a given PLE variant appears to be very rare: a particular PLE variant is typically 100% conserved at the nucleotide level across the entire mobile genetic element despite residing in different isolates of *V. cholerae* (9, 37). However, on rare occasions, single nucleotide polymorphisms between the same PLE variant are found in different *V. cholerae* isolates (38). Strikingly, the *tcaP*^{PLE5} sequence from 2017 showed a 14-nucleotide insertion which restored a full-length *tcaP* allele (Figure 5G and Figure S5B). The reestablishment of the full length *tcaP* allele in the contemporary PLE5 along with the preservation of full-length alleles in other PLEs support conservation of TcaP for its role in capsid remodeling to promote PLE transduction and inhibit ICP1.

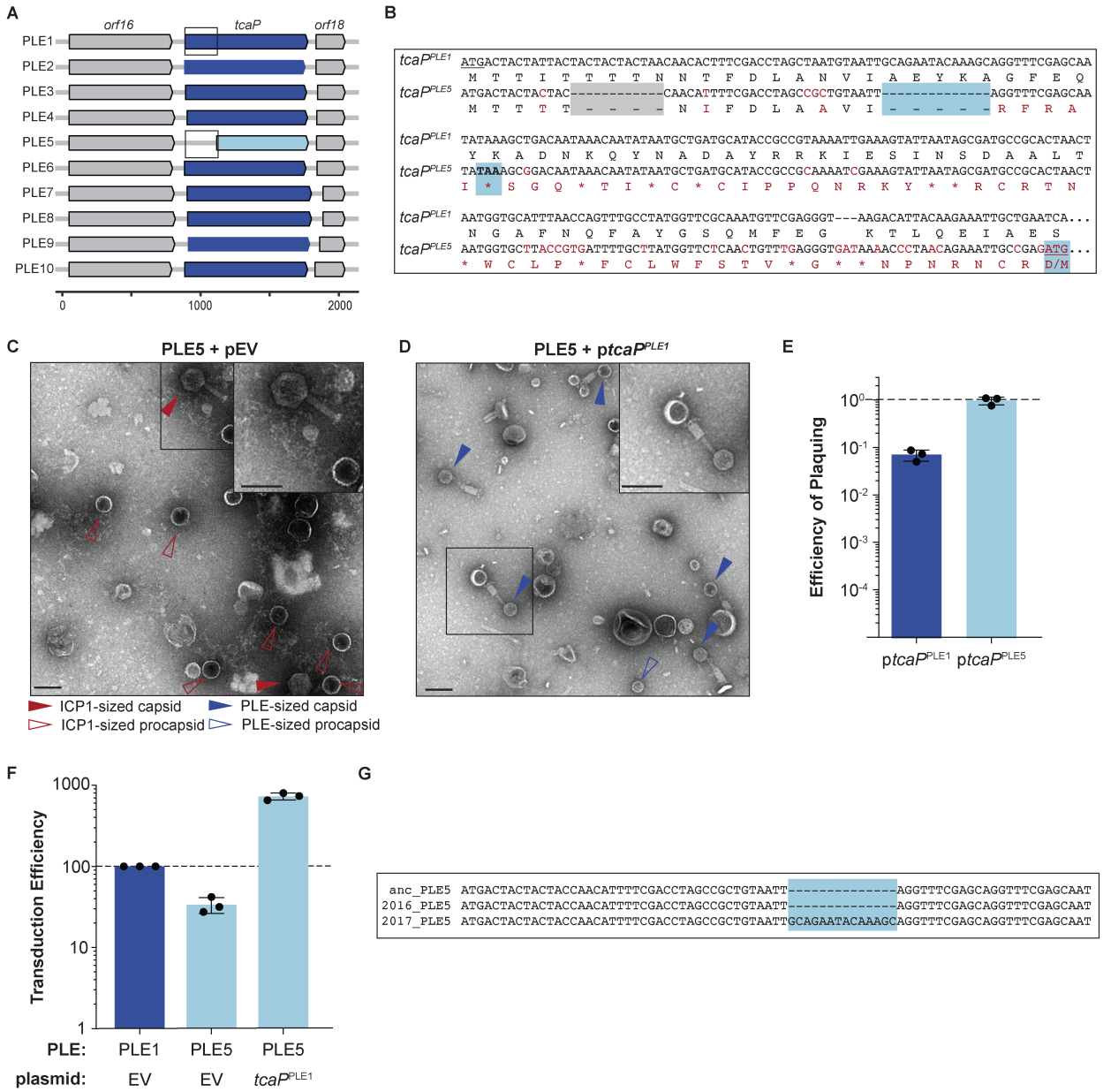


Figure 5. Genetic analysis of *tcaP* alleles from 10 PLEs reveals PLE5's nonfunctional TcaP variant.
A) Gene graphs of *tcaP* and its two neighboring genes, from the 10 known unique PLEs. PLE5's *tcaP*, the shortest allele, is shown in light blue, while the other, full-length *tcaP* alleles are shown in dark blue, and the neighboring genes are shown in gray. Lower scale bar is in nucleotides. Boxes outline regions aligned in (B).
B) Alignment of the region encoding *tcaP* and their translated products from PLE1 and PLE5 from the boxed regions in Panel A. Nonidentical nucleotides and amino acids are shown in red on the PLE5 sequence, gaps are shown as dashes on either sequence, and stop codons are shown as asterisks. The gray box indicates an in-frame deletion. The blue boxes indicate the notable features of the *tcaP*^{PLE5} sequence: the 14-nucleotide deletion that results in the frameshift, the resulting early stop codon, and the ATG and M of the alternative, originally annotated start site, which restores the original reading frame.
C-D) Representative TEMs from 2-3 independent biological replicates of lysate from ICP1-infected strains of PLE5(+) *V. cholerae* C) with an empty vector (pEV), or D) expressing *tcaP* from PLE1 (p*tcaP*^{PLE1}). Scale bars are 100 nm. Arrowheads show capsids and their sizes according to the legend.

E) Efficiency of ICP1 plaquing on *V. cholerae* expressing *tcaP* from PLE1 (p*tcaP*^{PLE1}) or *tcaP* from PLE5 (p*tcaP*^{PLE5}) (from the originally annotated start site producing the truncated allele) compared to an empty vector (pEV). Each dot represents a biological replicate, bars represent the mean, and error bars show the standard deviation. The dotted line indicates an efficiency of plaquing of 1, where the expression of TcaP is not inhibitory to plaque formation.
F) Transduction efficiency of the strain indicated relative to PLE1 with an empty vector. Each dot represents a biological replicate, bars represent the mean, and error bars show standard deviation. The dotted line indicates an efficiency of 100%.
G) Alignment of the first nucleotides of the *tcaP* alleles from the PLE5 variants encoding the "ancestral" (anc) sequences from before 1991, from 2016, or from 2017. The light blue box highlights the 14-nucleotide insertion in the PLE5 sequence from 2017.

Discussion

In this work, we characterized the mechanism of capsid remodeling in the phage-satellite system ICP1-PLE. We identified a PLE-encoded scaffold, TcaP, and demonstrated its necessity and sufficiency to redirect the assembly of ICP1's coat proteins into small capsids. Using a heterologous assembly system and cryo-EM, we discovered that TcaP functions as an external scaffold that assembles into a cage-like structure around the procapsid, favoring the smaller T=4 morphology (Figure 6A). We found that TcaP is conserved in all PLEs. Further, TcaP-mediated small capsid assembly is advantageous to PLE as the production of small capsids leads to more efficient transduction of PLE's genome, a unique feature not seen in other capsid-remodeling satellites. The data presented here further highlight similarities between PLE and other satellites while underscoring features that make PLE unique.

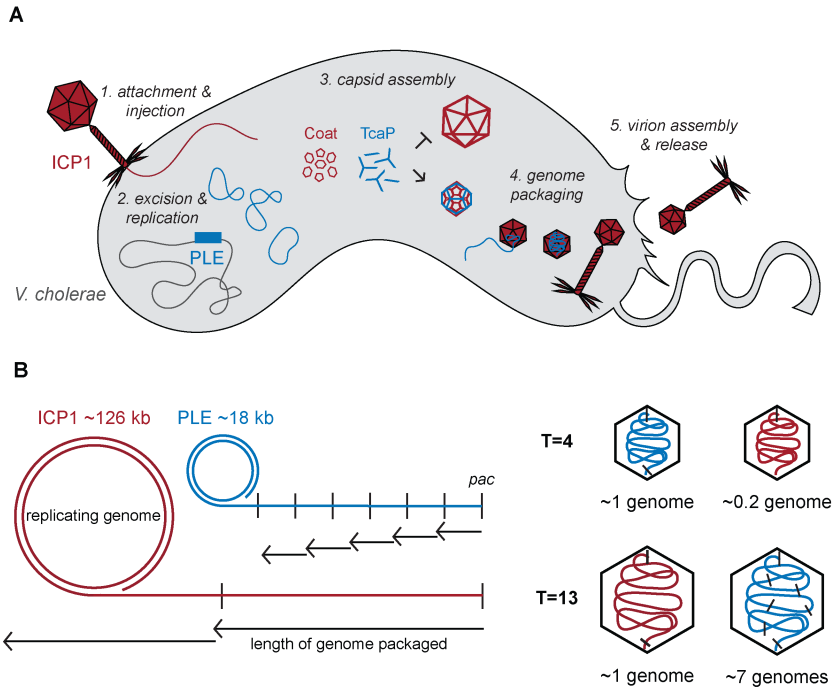


Figure 6. Model of TcaP-mediated small capsid assembly
A) A cartoon model of ICP1 infecting a PLE(+) *V. cholerae* cell. Injection of ICP1's genome triggers activation of PLE, then TcaP directs the assembly of coat proteins into small capsids, inhibiting the formation of large capsids. PLEs genome is then packaged into the small capsids. PLE virions are released from the cell. ICP1 components are red, PLE components are blue, and *V. cholerae* components are grey.
B) A model showing the impact of capsid size on genome packaging. The replicated ICP1 and PLE genomes form concatemers, from one *pac* site (indicated as a vertical line) to the next. Headful packaging results in ~105-110% of the genome within the capsid. The length of the genome packaged is

indicated by the small arrows for T=4 capsids and longer arrows for T=13 capsids. ICP1's genome is partially packaged into a T=4 capsid and several copies of PLE's genome are packaged into a T=13 capsid. ICP1's genome is red and PLE's is blue.

Initially, TcaP's function was unknown since homology-based and structural prediction searches were of low confidence and failed to identify known domains. However, the structure of PLE PLPs showed molecular details of the TcaP-coat interactions and revealed a striking resemblance to that of PLPs formed by the *E. coli* phage P2 coat protein (GpN), scaffold (GpO), and satellite P4 external scaffold (Sid) (2). Despite the low sequence conservation between the ICP1 and P2 coat proteins (20% identity), both have an HK97-like fold and are assembled into T=4 icosahedral procapsids measuring ~45 nm in diameter in the presence of their satellite-encoded external scaffolds. TcaP and Sid also share ~20% amino acid identity, however, TcaP is 53 amino acids longer than Sid. TcaP, like Sid, assembles into long filaments that form dimers that span the two-fold axes and trimers of dimers meeting at the three-fold axes of the procapsid. Given how TcaP and Sid interact with themselves to form dimers and trimers wherein hexamers are linked to each other by the length of TcaP or Sid dimers, it is not surprising how the resulting scaffold cage accommodates only procapsids with T=4 symmetry. Additional hexamers would not be accommodated within the cage formed by the external scaffolds, and thus the capsid is blocked from forming higher T number structures. Despite their similarities, TcaP and Sid differ in their specific interactions with the coat proteins. Unlike Sid, which only contacts two proteins in the hexamer (both GpN_B subunits), both TcaP monomers form interactions with three coat subunits, thus contacting all six coat subunits. TcaP and Sid also differ in their tertiary structures, further demonstrating the convergent evolution of the external scaffold capsid-remodeling strategy in divergent satellites.

Few external scaffolds have been described in viruses, and thus there are open questions about their role in capsid assembly, especially in relation to internal scaffolds. Outside of P4 and PLE, the only other identified external scaffolds are those in the ssDNA phages of the Microviridae family. A representative phage, ΦX174, requires an external scaffold (protein D) to form procapsids (39, 40), however, the structure of this protein is different from TcaP and Sid (41). Unlike the trimers formed by the satellite scaffolds, protein D seems to form tetramers in solution and directs the assembly of the coat proteins into a ~27 nm wide procapsid, despite the lack of direct coat-scaffold interactions. Notably, the small ΦX174 capsids still require an internal scaffold for assembly. Similarly, P4 is dependent on P2's internal scaffold (GpO) to form viable progeny virions, likely because the internal scaffold is important for the incorporation of the portal (30). Curiously, SaPIbov1, a satellite which uses an alternative internal scaffold to direct assembly of small capsids, also requires 80α's internal scaffold (Gp46) for viable satellite virions (42), likely for a similar reason. PLEs do not appear to encode portal proteins and thus we anticipate that PLE would similarly be dependent on ICP1 portal incorporation for progeny production. In the heterologous assembly assay, we show that TcaP's scaffolding activity is dominant over ICP1's scaffold, such that the resulting particles are satellite-sized in the presence of both scaffolds. These data suggest that if PLE depended on ICP1's scaffold for portal incorporation, TcaP's scaffolding activity would still be sufficient to direct the assembly of small capsids. It will be useful for future studies to directly address this requirement and to assess the complete protein makeup of PLE procapsids produced during ICP1 infection of PLE(+) *V. cholerae* to validate what PLE and ICP1 components comprise the procapsids.

Like ICP1 escaping TcaP activity, P2 can escape Sid-mediated capsid remodeling through point substitutions in the coat (2). Both of the two escape substitutions we identified in ICP1 lies in the A-domain, in the center of the hexamer across the 2-fold axis of symmetry where TcaP binds and one residue directly contacts TcaP. The A-domain has been implicated in regulating capsid size and assembly in many viral systems (reviewed in (43). Similarly, the five suppressor substitutions that have been identified in P2's coat protein lie in the hexamer along Sid's binding site, but they lie slightly outside of the A-domain and are instead at the end of the P-loop. Three substitutions are in residues that directly contact Sid, like R223 in TcaP. The other two residues in P2's coat likely function through other means to disrupt the interaction (2), as is probably the case for the second substitution seen in ICP1, E234K. Curiously, the PLP assembly data showed that the $coat^{R223H}$ only partially escaped TcaP (Figure 3) while this substitution was sufficient *in vivo* to escape TcaP's inhibition of plaque formation (Figure 1E). These inconsistencies can likely be explained by the lack of the additional components that comprise PLE native procapsids in our assembly platform, as well as the relative protein concentrations and dynamic

regulation that occurs during infection. However, despite the selection for phages that escape TcaP-mediated remodeling in the laboratory, no mutations in the A-domain exist in the current collection of 67 sequenced isolates of ICP1. Perhaps the mutations are not selected for in nature because the biophysical properties of these coat variants are incompatible with assembly in the natural conditions of the human gut or in estuaries where *V. cholerae* and ICP1 reside. In line with this hypothesis, one of the coat mutants we tested, coat^{E234K}, was not assembly competent in the heterologous assembly system (Figure S4), likely due to its predicted role in stabilizing the coat protein. Moreover, PLE is severely inhibitory to ICP1, employing several redundant strategies that independently restrict phage production, and thus *tcaP* is dispensable for phage inhibition (44). Instead of selecting for mutations that individually escape PLE's inhibitory proteins, ICP1 encodes broader strategies that degrade the PLE genome, such as the phage-encoded CRISPR-Cas system (45, 46), Odn (38), or Adi (12).

The inhibitory effects on the helper phage resulting from the production of small capsids from hijacked coat proteins are comparable between PLEs and other satellites (6, 24, 47). PLE, like other satellites, has a genome size of ~15-20 kb, which is compatible with the satellite-modified T=4 capsid size. However, the ICP1 genome does not follow the pattern seen in other helper phages, which typically are ~30-40 kb and packaged into similarly sized capsids (icosahedral T=7 or prolate T_{end}=4, T_{mid}=14). ICP1 has a large ~125 kb genome packaged into a T=13 capsid. Therefore, PLE and ICP1 have the most dramatic size difference between satellite and helper phage. The consistent level of inhibition is likely explained by the fact that regardless of how much smaller the capsid is, packaging anything less than the full length of the genome results in non-infectious virions. However, the effect of capsid size on the satellite packaging can be different. Phages and satellites that use *cos* site packaging strategies specifically cleave the genome only at those sites and thus package unit lengths of their genomes into a capsid. For satellites, if their genome is proportional to two or three lengths of their helper phages, they can package their genome into helper-sized capsids with no or reportedly little reduction in efficiency (5, 24). On the other hand, with headful packaging systems, which have been predicted for ICP1 and PLE (13), replicated genome concatemers will be threaded into the capsid until the capsid is full, signaling the terminase to stop packaging and cut the genome. Most phage's capsids accommodate 103-110% of their genome length (48). While the SaPIs that use headful packaging are unaffected by packaging their genomes into large capsids, PLE's transduction is severely reduced when packaged into ICP1-size capsids. The difference in capsid size between the SaPI helper, 80α (T=7), and PLE's helper, ICP1 (T=13), explains this result. Approximately two SaPI genomes would need to be packaged to fill the larger, T=7 capsid produced in the absence of its remodeling proteins CpmB and CpmA. Meanwhile, 6-7 PLE genomes would need to be packaged into ICP1's large, T=13 capsids in the absence of TcaP (Figure 6B). This larger difference in genome copies per capsid supports why PLE is the only satellite to suffer a transduction defect in the absence of capsid remodeling. The negative effect on PLE transduction in the absence of capsid remodeling suggests that PLEs benefit from TcaP's activity, explaining why it is largely conserved in all PLEs discovered to date.

By packaging each of its replicated genomes into capsids proportional to its size, PLE makes more transducing particles and increases the odds of horizontal transfer of its genome. Theoretically, high transduction efficiency would be advantageous and selected for in PLEs. If this is the case, the expectation would be that all PLEs encode a functional TcaP that directs the assembly of small capsids. However, older isolates of PLE5 lack such an allele and consequently package their genomes into large capsids (Figure 5C). If the loss of TcaP decreased PLE5's fitness by decreasing its ability to spread horizontally, it could be expected that this PLE would quickly be replaced by more fit PLEs in the population. Surveillance data show that PLE5 was detected in *V. cholerae* genomes from as early as 1931 until 1991 and again in 2016 and 2017 (10), suggesting PLE5 was maintained in the population for relatively long periods of time. Curiously, in an isogeneic background and under lab conditions, PLE5 transduces nearly as efficiently as PLE1 (9), though transduction efficiency is increased by the formation of small capsids (Figures 5D and 5F). These data suggest that PLE5 encodes other factors that contribute to its transduction and promote its maintenance in the population. It is unclear at what frequency and/or efficiency PLE transduction occurs in native conditions in estuaries or in the human intestinal tract where ICP1 preys on *V. cholerae*. It may be that low levels of transduction are sufficient to maintain a PLE in the population. Alternatively, PLEs could be maintained primarily through vertical transmission (10). The arsenal of anti-ICP1 factors encoded by PLE, even in the absence of TcaP

activity, seem beneficial for *V. cholerae*, supporting vertical transmission of PLE5 as a means for its maintenance in the population. However, the emergence of a restored full-length allele of *tcaP* in PLE5 in 2017, paired with the conservation of TcaP in other PLEs and evidence of horizontal PLE transfer in some clinical isolates (10), suggests capsid remodeling is advantageous due to its role in transduction and is selected for in PLEs in nature.

PLE is a unique satellite and its helper phage ICP1 is also dissimilar from other helper phages (3). In addition to its large capsid size, ICP1 is different from other documented helper phages in that it is an obligate lytic phage. This raises questions about whether the characteristics of the ICP1-PLE parasitism are specifically due to the lytic nature of ICP1. For example, PLE's anti-phage activity is the most potent of any satellite where ICP1 progeny production is completely blocked by PLE. Perhaps this more severe phage restriction is important for the protection of neighboring *V. cholerae* cells from lethal infection. In the case of temperate phages, the presence of many phage kin may promote lysogeny, and the infected cell would survive. As this is not a possibility for *V. cholerae* infected by ICP1, PLE's reduction of ICP1 in the environment may be a necessary means of protection for the bacterial population. In support of the selection for phage inhibition, PLEs encode multiple redundant mechanisms to block ICP1. It will be interesting to see if other lytic phages are parasitized by satellites and if there is conservation of distinct inhibitory mechanisms. Recent work has highlighted distantly related elements in other *Vibrio* species that may be elements with similarities to PLEs (49), but putative helper phages have not been identified. A recent study described a density separation and sequencing technique to identify novel satellites (50) that may help illuminate the diversity in phage-satellite pairings. As this work has highlighted, the convergent evolution of satellite strategies to hijack aspects of a helper phage's lifecycle may not be obvious at the sequence level, and the characterization of novel satellites can reveal features that unite or separate different families of satellites.

Materials and methods

Bacterial Growth Conditions

V. cholerae and *E. coli* were propagated at 37˚C on LB agar or in LB broth with aeration. Where needed, antibiotics were used at the following concentrations: streptomycin 100 µg/mL, kanamycin 75 µg/mL, spectinomycin 100 µg/mL, chloramphenicol 2.5 µg/mL (solid media) or 1.25 µg/mL (liquid media) for *V. cholerae* and 25 µg/mL for *E. coli*, and carbenicillin 50 µg/mL.

Strain Construction

V. cholerae carrying PLE1 marked with a kanamycin resistance cassette downstream of the last ORF (9) was made naturally competent and transformed with DNA fragments containing a spectinomycin resistance marker flanked by frt recombinase sites assembled to up and downstream regions of homology by splicing by overlap extension PCR as previously described (51). Spectinomycin resistant transformants were transformed with a plasmid carrying a FLP recombinase which was induced by IPTG (1 mM), allowing for the removal of the spectinomycin resistance cassette via recombination, resulting in an in-frame deletion. The strains were cured of the plasmid, and deletions were confirmed by PCR and Sanger sequencing.

Plasmid constructs were assembled using Gibson Assembly and transformed into *E. coli* BL21 or mated into *V. cholerae* via *E. coli* S17. A list of strains used in this study can be found in Table S5. Plasmids in *V. cholerae* have a P_{tac} promoter downstream of a theophylline sensitive riboswitch. A list of oligos used in this study can be found in Table S6.

ICP1 plaque assays

Overnight cultures of *V. cholerae* were diluted to OD_{600}=0.05 and grown in LB (supplemented with antibiotics where appropriate) with aeration at 37˚C either directly to OD_{600}=0.3 or to OD_{600}=0.2 then induced with 1 mM IPTG and 1.5 mM theophylline for 20 minutes to reach OD_{600}=0.3, then mixed with pre-diluted phage samples. Phage attachment was allowed for 7-10 minutes prior to plating in 0.5% molten top agar (supplemented with antibiotics and inducer where appropriate) followed by overnight incubation at 37˚C. Resulting individual plaques were counted.

ICP1 mutant purification and whole genome sequencing
V. cholerae carrying a plasmid encoding *tcaP* was used in a plaque assay as described above. As a control, ICP1_2011_Dha_A was used to infect *V. cholerae* expressing an empty vector. Plaques that formed on the *tcaP*-expressing strain were picked and purified on that strain two more times by plaque assay. High titer phage stocks were prepared by sodium chloride (1 mM) polyethylene glycol 8000 (10%) precipitation or by centrifugation (26,000 x *g* for 90 minutes) and stored in STE (5 mM Tris pH 8.0, 100 mM NaCl, 1 mM EDTA). Prior to collecting genomic DNA (gDNA), phage stocks were treated with DNase for 30 minutes at 37˚C to remove non-encapsidated DNA, then the enzyme was heat inactivated. gDNA was collected using a Qiagen DNeasy blood and tissue DNA purification kit according to the manufacturer's protocols and genomic libraries were prepared for Illumina sequencing using the NEBNext Ultra II DNA Library preparation kit as described in the manufacturer's protocols. Using an Illumina HiSeq4000 (University of California, Berkeley QB3 Core Facility), samples were sequenced by paired-end (2 x 150 bp). The genomes were assembled using SPAdes and analyzed by BreSeq (v0.33).

Virion production for TEM and/or transduction
50 mL cultures of *V. cholerae* strains carrying plasmids were grown and induced as described above. ICP1 was added at a multiplicity of infection (MOI) of 2.5 and cultures were incubated until lysis (30-90 minutes). 1 mL of lysate was used for transduction assays (see below). Remaining lysates were concentrated by centrifugation at 26,000 x *g* for 90 minutes, resuspended in Phage Buffer 2.0 (50 mM Tris-HCl, 100 mM NaCl, 10 mM $MgSO_4$, 1 mM $CaCl_2$) overnight, treated 1:1 with chloroform for 15 minutes, and centrifuged at 5,000 x *g* for 15 minutes. The aqueous layer was collected and 5 µL was applied to a grid for TEM.

PLE transduction assays were carried out as previously described (1, 9). Briefly, 1 mL of lysates from strains carrying PLE marked with a kanamycin resistance cassette downstream of the last ORF were treated with 10 µL chloroform, which was removed, along with bacterial debris, by centrifugation at 5,000 x *g* for 15 minutes. The supernatant was collected and mixed 10:100 with a saturated overnight culture of spectinomycin resistant recipient *V. cholerae* cells (Δ*lacZ::spec*^{R}) supplemented with 10 mM $MgSO_4$ immediately prior to transduction. Recipients were incubated with lysates for 20 minutes at 37˚C with shaking (220 rpm) and then serially 10-fold diluted. The resulting dilutions were plated on LB agar plates supplemented with spectinomycin and kanamycin. A colony represents one PLE virion/transducing unit.

TEM
For the preparation of grids, 5 µL samples were incubated on a copper mesh grid (Formvar/Carbon 300, Electron Microscopy Sciences) for 60 seconds, wicked, immediately washed with sterile ddH_2O for 15 seconds, wicked, immediately stained with 1% uranyl acetate for 30 seconds, wicked and allowed to dry completely. Micrographs were collected with a FEI Tecnai-12 electron microscope operating at 120 kV.

Production and Purification of Procapsid-like-Particles
Using Gibson Cloning, the coat gene fused to 6xHistidine was inserted into the pETDUET vector or the untagged coat was cloned into pCDFDuet. Either the ICP1 or PLE scaffold proteins were similarly inserted via Gibson Assembly downstream of a second T7 promoter. In other variations of this plasmid, the genes were cloned with the intergenic sequences found in ICP1_2011_Dha_A. The pETDUET constructs were transformed into *E. coli* BL21 and grown on LB agar or in LB broth supplemented with carbenicillin (50 µg/mL). To reduce aggregation of PLE PLPs for cryo-EM, a pETDUET construct encoding *gp122*::6xhis and *tcaP* was co-transformed with a pCDFDuet construct expressing *gp122* (untagged), grown on LB agar or in LB broth supplemented with carbenicillin (50 µg/mL) and streptomycin (100 µg/mL). For protein production, overnight cultures were diluted 1:100 in 0.5-1 L, grown to OD_{600}=0.2-0.4, induced with 1 mM IPTG for 3-5 hours, and collected by centrifugation at 4,000 x *g* for 20 minutes at 4˚C. Pellets were then resuspended in 10% of the volume of Phage Purification Buffer (Phage Buffer 2.0 supplemented with 20 mM imidazole, 1 mM BME, and protease and phosphatase inhibitors (cat#A32961/A32965, Pierce)), and frozen at -80˚C. For protein purification, frozen samples were thawed, sonicated, and centrifuged at 12,000 x *g* for 60-90 minutes to remove membranes and debris. The particles were then pelleted at 26,000 x *g* for 90 minutes. The pellet was nutated overnight in Phage Purification Buffer, and the nutate was loaded onto a column packed with HisPur nickel-nitrilotriacetic acid

(Ni-NTA) resin (cat# PI88222, Thermo Scientific) pre-equilibrated with ice-cold Phage Purification Buffer. After two passes over the column, the flow-through was collected and the column was washed with wash buffer (Phage Buffer 2.0 supplemented with 50 mM imidazole, 1 mM BME, and 10% glycerol). Proteins were then eluted with one column volume of Elution Buffers 1-3 (Phage Buffer 2.0 supplemented with 1 mM BME and 10% glycerol and 150, 250, or 350 mM imidazole) and a final six column volume elution with Elution Buffer 4 (Phage Buffer 2.0 supplemented with 1 mM BME and 10% glycerol and 500 mM imidazole). For all experiments with coat^{R223H} and the noted replicates of wild type coat, decoration, scaffold and TcaP, protease and phosphatase inhibitors were added to all purification and elution buffers. Aliquots of the fractions were either boiled in Laemlli buffer and assessed by SDS-PAGE stained with Coomassie or applied to a grid and imaged by TEM.

Cesium chloride-based purification following affinity purification were carried out as follows. PLE PLP (coat::6xHis + TcaP) samples from the final elutions from affinity chromatography were concentrated on a 100K MWCO Amicon ultra filter (cat#UFC810024, Millipore) to a final volume of ~1 mL. 1 mL steps of 1.6 g/cm^3, 1.5 g/cm^3, 1.4 g/cm^3, 1.3 g/cm^3 CsCl were made in Phage Buffer 2.0 (50 mM Tris-HCl, 100 mM NaCl, 10 mM $MgSO_4$, 1 mM $CaCl_2$ pH 8.0), filter sterilized, and layered in ultracentrifuge tubes (thinwalled WX 5 mL Cat#1131 Thermo Scientific) then topped with 500 µL of PLE PLP sample. Samples were spun for 2 hours at 36,000 rpm (~110,000 x *g*) at 18˚C (AH-650 swinging bucket rotor Thermo Scientific), then 250 µL fractions were manually collected. The protein content of fractions was assessed by SDS-PAGE/Coomassie. Fractions containing the most pure coat::6xHis and TcaP complexes were pooled, dialyzed against Phage Buffer 2.0 in 20k MWCO Slide-A-Lyzer dialysis cassettes (cat#66003, Thermo Scientific), concentrated on Amicon ultra filters (cat#UFC810024, Millipore) and prepared for cyro-EM.

Size exclusion purifications were carried out as follows. Lysates were prepared as described above and particles were pelleted and nutated in Phage Buffer 2.1 (same as Phage Buffer but with pH 7.4). The nutate was then spun at 12,000 x *g* for 10 minutes to remove aggregates then applied to an AKTA HiPrep 16/60 Sephacryl S-500 HR (cat#28935606, Cytiva) column and passed through at a flow rate of 0.5 mL/min. Protein content from elution peaks was assessed by SDS-PAGE/Coomassie and samples containing PLE PLPs (elutions at 26-32 mL) were concentrated by centrifugation at 26,000 x *g* for 90 minutes, resuspended in ~100-500 µL Phage Buffer 2.0 and prepared for cryo-EM.

CryoEM Sample preparation and data acquisition

CryoEM samples were prepared by applying 5 µL aliquots of purified PLE PLPs to R2/2 Quantifoil grids and R2/2 Quantifoil grids coated with 2nm ultrathin Carbon (QUANTIFOIL®) that had been glow discharged for 45 seconds in a Pelco Easiglow glow discharging unit. The samples were plunge frozen in liquid ethane using a Vitrobot Mark IV operated at 4˚C and 100% humidity, with a blot force of 1 and 5 seconds of blotting time per grid. The grids were screened at RTSF Cryo-EM facility using a Talos Arctica equipped with a Falcon 3EC direct electron detector and Cryo-EM data were collected at Purdue Cryo-EM facility using a Titan Krios equipped with a K3 direct electron detector, and operating at 300 keV with a post-column GIF (20 eV slit width) under low dose conditions. Micrographs were collected at 64,000X nominal magnification (0.664 Å/pixel) by recording 40 frames over 3.1 sec for a total dose of 36.41 e^-/Å^2.

Data processing for icosahedral reconstruction of PLE procapsids was carried out using Cryosparc v4.0.3. Briefly, the dose-fractionated movies were subjected to motion correction using patch motion correction with 2X binning and CTF estimation of the resulting images were estimated using the patch CTF estimation jobs. The particles were picked using the Template picker job with templates generated from blob picker. Particles were then extracted and subjected to 2D classification. CryoEM image processing statistics along with data collection parameters are listed in Table S7. 790,843 particles were used for 3D refinement with C1 symmetry, with a model generated from previous low-resolution data serving as the initial model. Refined particles were then subjected to a round of 3D classification in Relion 4.0 (52–54). 379,643 particles were then imported into Cryosparc and used for 3D refinement with Icosahedral symmetry. The overall resolution was estimated using Postprocess job with a spherical mask in Relion 4.0 based on the gold-standard Fourier shell correlation (FSC) = 0.143 criterion (52, 55). The final map was sharpened with a B-factor of -130. The final maps were deposited into EMDB (accession number EMD-29675).

AlphaFold2 (36) was used to generate homology models for all modeled protein chains. Initial models were then docked into EM maps for further refinement. For TcaP only residues 34 – 172 were modeled. Refinement was carried out using *Phenix* (56) and model adjustments were carried out in COOT (57). Model parameters were monitored using Molprobity in Phenix and the values are listed in Supplementary Table S8.

Mass Spectrometry

PLE and ICP1 PLPs subjected to mass spectrometry were further concentrated by trichloroacetic acid (TCA) precipitation, washed (10 mM HCl 90% acetone) three times and air dried. The pellets were then resuspended in 100 mM Tris pH 8.5, 8 M urea, digested by trypsin, and analyzed by LC-MS/MS (Thermo LTQ XL linear ion trap mass spectrometer at the Vincent J. Coates Proteomics/Mass Spectrometry Laboratory at UC Berkeley). A sequence database containing *E. coli* proteins as well as ICP1 coat, scaffold and decoration, as well as PLE TcaP were used to compare the masses recorded and identify proteins in the sample.

Acknowledgements

We would like to thank members of the Seed Lab for their useful feedback on this work. We would also like to thank the staff in the electron microscopy facility at the University of California Berkeley, RTSF Cryo-EM facility at Michigan State University, and the Purdue CryoEM facility for their assistance with the microscopy, and the staff in the Vincent J. Coates Proteomics/Mass Spectrometry Laboratory for their assistance with sample processing.

This work was supported by the National Institutes of Health [R01AI127652 to K.D.S; GM110185, GM140803, GM116789 to K.N.P; S10 Instrumentation Grant S10RR025622 in part for use of the Vincent J. Coates Proteomics/Mass Spectrometry Laboratory at University of California Berkeley], the National Institutes of Health NRSA Trainee Fellowship [5 T32 GM132022 to C.M.B. in part], a National Science Foundation Graduate Research Fellowship [2018257700 to D.T.D], and a National Science Foundation CAREER Award [1750125 to K.N.P.]. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the National Institute of Allergy and Infectious Diseases or NIH. K.D.S. holds an Investigators in the Pathogenesis of Infectious Disease Award from the Burroughs Wellcome Fund.

References

1. Z. Netter, C. M. Boyd, T. V. Silvas, K. D. Seed, A phage satellite tunes inducing phage gene expression using a domesticated endonuclease to balance inhibition and virion hijacking. *Nucleic Acids Res.*, 1–16 (2021).

2. J. L. Kizziah, C. M. Rodenburg, T. Dokland, Structure of the Capsid Size-Determining Scaffold of "Satellite" Bacteriophage P4. *Viruses* **12** (2020).

3. J. A. M. de Sousa, A. Fillol-Salom, J. R. Penadés, E. P. C. Rocha, "Identification and characterization of thousands of bacteriophage satellites across bacteria" (Microbiology, 2022) https:/doi.org/10.1101/2022.09.14.508007 (November 8, 2022).

4. G. E. Christie, T. Dokland, Pirates of the Caudovirales. *Virology* **434**, 210–221 (2012).

5. D. Shore, G. Deho, J. Tsipis, R. Goldstein, Determination of capsid size by satellite bacteriophage P4. *Proc. Natl. Acad. Sci. U. S. A.* **75**, 400–404 (1978).

6. P. K. Damle, *et al.*, The roles of SaPI1 proteins gp7 (CpmA) and gp6 (CpmB) in capsid size determination and helper phage interference. *Virology* **432**, 277–282 (2012).

7. N. C. Hawkins, J. L. Kizziah, J. R. Penadés, T. Dokland, Shape shifter: redirection of prolate phage capsid assembly by staphylococcal pathogenicity islands. *Nat. Commun.* **12**, 1–11 (2021).

8. N. Alqurainy, *et al.*, A widespread family of phage-inducible chromosomal islands only steals bacteriophage tails to spread in nature. *Cell Host Microbe*, S193131282200573X (2023).

9. B. J. O'Hara, Z. K. Barth, A. C. McKitterick, K. D. Seed, A highly specific phage defense system is a conserved feature of the Vibrio cholerae mobilome. *PLoS Genet.* **13**, 1–17 (2017).

10. A. Angermeyer, *et al.*, Evolutionary Sweeps of Subviral Parasites and Their Phage Host Bring Unique Parasite Variants and Disappearance of a Phage CRISPR-Cas System. *mBio* **13** (2022).

11. A. C. McKitterick, K. D. Seed, Anti-phage islands force their target phage to directly mediate island excision and spread. *Nat. Commun.* **9**, 1–8 (2018).

12. M. H. T. Nguyen, Z. Netter, A. Angermeyer, K. D. Seed, A phage weaponizes a satellite recombinase to subvert viral restriction. *Nucleic Acids Res.* **50**, 11138–11153 (2022).

13. Z. K. Barth, T. V. Silvas, A. Angermeyer, K. D. Seed, Genome replication dynamics of a bacteriophage and its satellite reveal strategies for parasitism and viral restriction. *Nucleic Acids Res.* **48**, 249–263 (2019).

14. A. C. Steven, J. B. Heymann, N. Cheng, B. L. Trus, J. F. Conway, Virus maturation: dynamics and mechanism of a stabilizing structural transition that leads to infectivity. *Curr. Opin. Struct. Biol.* **15**, 227–236 (2005).

15. J. King, S. Casjens, Catalytic head assembling protein in virus morphogenesis. *Nature* **251**, 112–119 (1974).

16. T. Dokland, Scaffolding proteins and their role in viral assembly. *Cell. Mol. Life Sci.* **56**, 580–603 (1999).

17. P. E. Prevelige, D. Thomas, J. King, Scaffolding protein regulates the polymerization of P22 coat subunits into icosahedral shells in vitro. *J. Mol. Biol.* (1988) https:/doi.org/10.1016/0022-2836(88)90555-4.

18. R. L. Duda, K. Martincic, R. W. Hendrix, Genetic basis of bacteriophage HK97 prohead assembly. *J. Mol. Biol.* **247**, 636–647 (1995).

19. A. D. Dearborn, *et al.*, Competing scaffolding proteins determine capsid size during mobilization of Staphylococcus aureus pathogenicity islands. *eLife* **6**, 1–26 (2017).

20. O. J. Marvik, *et al.*, The capsid size-determining protein Sid forms an external scaffold on phage P4 procapsids. *J. Mol. Biol.* **251**, 59–75 (1995).

22. C. M. Boyd, *et al.*, Bacteriophage ICP1: A Persistent Predator of. *Annu. Rev. Virol.* **12**, 1–20 (2021).

23. J. S. Depelteau, *et al.*, UVC inactivation of pathogenic samples suitable for cryo-EM analysis. *Commun. Biol.* **5**, 29 (2022).

24. N. Carpena, K. A. Manning, T. Dokland, A. Marina, J. R. Penadés, Convergent evolution of pathogenicity islands in helper cos phage interference. *Philos. Trans. R. Soc. B Biol. Sci.* **371**, 20150505 (2016).

25. S. G. Hays, K. D. Seed, Dominant Vibrio cholerae phage exhibits lysis inhibition sensitive to disruption by a defensive phage satellite. *eLife* **9**, 1–24 (2020).

26. M. S. Spilman, *et al.*, Assembly of bacteriophage 80α capsids in a Staphylococcus aureus expression system. *Virology* **434**, 242–250 (2012).

27. S. Wang, J. R. Chang, T. Dokland, Assembly of bacteriophage P2 and P4 procapsids with internal scaffolding protein. *Virology* **348**, 133–140 (2006).

28. M. E. Cerritelli, F. W. Studier, Assembly of T7 capsids from independently expressed and purified head protein and scaffolding protein. *J. Mol. Biol.* **258**, 286–298 (1996).

29. B. J. Robertson, P. J. McCann, L. Matusick-Kumar, V. G. Preston, M. Gao, Na, an autoproteolytic product of the herpes simplex virus type 1 protease, can functionally substitute for the assembly protein ICP35. *J. Virol.* **71**, 1683–1687 (1997).

30. J. R. Chang, A. Poliakov, P. E. Prevelige, J. A. Mobley, T. Dokland, Incorporation of scaffolding protein gpO in bacteriophages P2 and P4. *Virology* **370**, 352–361 (2008).

31. R. L. Duda, B. Oh, R. W. Hendrix, Functional domains of the HK97 capsid maturation protease and the mechanisms of protein encapsidation. *J Mol Biol* **23**, 1–7 (2013).

32. A. Huet, R. L. Duda, R. W. Hendrix, P. Boulanger, J. F. Conway, Correct Assembly of the Bacteriophage T5 Procapsid Requires Both the Maturation Protease and the Portal Complex. *J. Mol. Biol.* **428**, 165–181 (2016).

33. O. J. Marvik, P. Sharma, T. Dokland, B. H. Lindqvist, Bacteriophage P2 and P4 Assembly: Alternative Scaffolding Proteins Regulate Capsid Size. *Virology* **200**, 702–714 (1994).

34. E. W. Six, The helper dependence of satellite bacteriophage P4: Which gene functions of bacteriophage P2 are needed by P4? *Virology* **67**, 249–263 (1975).

35. G. C. Lander, *et al.*, Capsomer dynamics and stabilization in the T = 12 marine bacteriophage SIO-2 and its procapsid studied by CryoEM. *Structure* **20**, 498–503 (2012).

36. J. Jumper, *et al.*, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

37. A. C. McKitterick, K. N. LeGault, A. Angermeyer, M. Alam, K. D. Seed, Competition between mobile genetic elements drives optimization of a phage-encoded CRISPR-Cas system: Insights from a natural arms race. *Philos. Trans. R. Soc. B Biol. Sci.* **374**, 1–12 (2019).

38. Z. K. Barth, M. H. T. Nguyen, K. D. Seed, A chimeric nuclease substitutes a phage CRISPR-Cas system to provide sequence- specific immunity against subviral parasites. *eLife*, 1–21 (2021).

39. T. Dokland, *et al.*, Structure of a viral procapsid with molecular scaffolding. *Nature* **389**, 308–313 (1997).

40. S. Hafenstein, B. A. Fane, X174 Genome-Capsid Interactions Influence the Biophysical Properties of the Virion: Evidence for a Scaffolding-Like Function for the Genome during the Final Stages of Morphogenesis. *J. Virol.* **76**, 5350–5356 (2002).

41. J. E. Cherwa, L. J. Organtini, R. E. Ashley, S. L. Hafenstein, B. A. Fane, In vitro assembly of the øx174 procapsid from external scaffolding protein oligomers and early pentameric assembly intermediates. *J. Mol. Biol.* **412**, 387–396 (2011).

42. M. S. Spilman, *et al.*, Assembly of bacteriophage 80α capsids in a Staphylococcus aureus expression system. *Virology* **434**, 242–250 (2012).

43. M. M. Suhanovsky, C. M. Teschke, Nature's favorite building block: Deciphering folding and capsid assembly of proteins with the HK97-fold. *Virology* **479–480**, 487–497 (2015).

44. S. G. Hays, K. D. Seed, Dominant Vibrio cholerae phage exhibits lysis inhibition sensitive to disruption by a defensive phage satellite. *eLife* **9**, 1–24 (2020).

45. K. D. Seed, D. W. Lazinski, S. B. Calderwood, A. Camilli, A bacteriophage encodes its own CRISPR/Cas adaptive response to evade host innate immunity. *Nature* **494**, 489–491 (2013).

46. A. C. McKitterick, K. N. LeGault, A. Angermeyer, M. Alam, K. D. Seed, Competition between mobile genetic elements drives optimization of a phage-encoded CRISPR-Cas system: Insights from a natural arms race. *Philos. Trans. R. Soc. B Biol. Sci.* **374**, 1–12 (2019).

47. K. J. Kim, M. G. Sunshine, B. H. Lindqvist, E. W. Six, Capsid size determination in the P2-P4 bacteriophage system: Suppression of sir mutations in P2's capsid gene N by supersid mutations in P4's external scaffold gene sid. *Virology* **283**, 49–58 (2001).

48. V. B. Rao, M. Feiss, Mechanisms of DNA Packaging by Large Double-Stranded DNA Viruses. *Annu. Rev. Virol.* **2** (2015).

49. K. N. LeGault, Z. K. Barth, P. DePaola, K. D. Seed, A phage parasite deploys a nicking nuclease effector to inhibit viral host replication. *Nucleic Acids Res.*, 1–17 (2022).

50. J. M. Eppley, S. J. Biller, E. Luo, A. Burger, E. F. DeLong, Marine viral particles reveal an expansive repertoire of phage-parasitizing mobile elements. *Proc. Natl. Acad. Sci.* **119**, e2212722119 (2022).

51. A. B. Dalia, D. W. Lazinski, A. Camilli, Identification of a Membrane-Bound Transcriptional Regulator That Links Chitin and Natural Competence in Vibrio cholerae. *mBio* **5**, e01028-13 (2014).

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

53. D. Asarnow, E. Palovcak, Y. Cheng, pyem. Zenodo. https://doi.org/10.5281/zenodo.3576630. Deposited 16 December 2019.

54. A. Punjani, J. L. Rubinstein, D. J. Fleet, M. A. Brubaker, cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).

55. J. M. de la Rosa-Trevín, *et al.*, Scipion: A software framework toward integration, reproducibility and validation in 3D electron microscopy. *J. Struct. Biol.* **195**, 93–99 (2016).

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

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