



Published in final edited form as:

*Mol Microbiol.* 2012 April ; 84(2): 324–339. doi:10.1111/j.1365-2958.2012.08025.x.

## Extensive proteolysis of head and inner body proteins by a morphogenetic protease in the giant *Pseudomonas aeruginosa* phage $\phi$ KZ

Julie A. Thomas<sup>1</sup>, Susan T. Weintraub<sup>2</sup>, Weimin Wu<sup>3</sup>, Dennis C. Winkler<sup>3</sup>, Naiqian Cheng<sup>3</sup>, Alasdair C. Steven<sup>3</sup>, and Lindsay W. Black<sup>1</sup>

<sup>1</sup>Biochemistry and Molecular Biology, University of Maryland Baltimore, MD 21201 USA

<sup>2</sup>Biochemistry, University of Texas Health Science Center, San Antonio, TX 78229 USA

<sup>3</sup>Laboratory of Structural Biology Research, National Institute for Arthritis, Musculoskeletal and Skin Disease, National Institutes of Health, Bethesda, MD 20892

### Summary

Encased within the 280 kb genome in the capsid of the giant myovirus  $\phi$ KZ is an unusual cylindrical proteinaceous “inner body” of highly ordered structure. We present here mass spectrometry, bioinformatic, and biochemical studies that reveal novel information about the  $\phi$ KZ head and the complex inner body. The identification of 39 cleavage sites in 19  $\phi$ KZ head proteins indicates cleavage of many prohead proteins forms a major morphogenetic step in  $\phi$ KZ head maturation. The  $\phi$ KZ head protease, gp175, is newly identified here by a bioinformatics approach, as confirmed by a protein expression assay. Gp175 is distantly related to T4 gp21 and recognizes and cleaves head precursors at related but distinct S/A/G-X-E recognition sites. Within the  $\phi$ KZ head there are six high copy number proteins that are probable major components of the inner body. The molecular weights of five of these proteins are reduced 35–65% by cleavages making their mature form similar (26–31 kDa), while their precursors are dissimilar (36–88 kDa). Together the six abundant proteins sum to the estimated mass of the inner body (15–20 MDa). The identification of these proteins is important for future studies on the composition and function of the inner body.

### Keywords

myovirus; serine protease; cleavage motif; phiKZ; mass spectrometry; abundant capsid proteins; PF12699 paralog family; Hidden Markov Modelling

### Introduction

The *Pseudomonas aeruginosa* phage  $\phi$ KZ is the type virus of a genus of “giant” myoviruses, the  $\phi$ KZ-like phages, whose members have been isolated from diverse geographic locations and are infective for a variety of *Pseudomonas* species (Krylov *et al.*, 2007). The  $\phi$ KZ-like phages and their gene products are important due to their potential to control pathogenic pseudomonads (Briers *et al.*, 2008, Miroshnikov *et al.*, 2006, Paradis-Bleau *et al.*, 2007, Mesyanzhinov *et al.*, 2002) and genetically manipulate their hosts (Monson *et al.*, 2011). Some, such as  $\phi$ KZ, have already been incorporated into traditional phage therapy cocktails and continue to be examined for novel therapeutic applications [e.g., (Golshahi *et al.*, 2011, Matinkhoo *et al.*, 2011)].  $\phi$ KZ has a long, 280 kb linear dsDNA genome that encodes numerous proteins that are remarkably diverged from those in other

phages, thus making straightforward functional assignments very difficult (Mesyanzhinov *et al.*, 2002).

The  $\phi$ KZ virion also presents a singular research focus as its large T=27 head (Fokine *et al.*, 2005a) contains an unusual immense cylindrical protein “inner body” (IB) encased within the genomic DNA (Krylov *et al.*, 1984). The highly ordered structure of this IB has recently been solved (Wu *et al.*, 2012). The IB is a conserved feature in related phages [(e.g., 201  $\phi$ 2-1,  $\phi$ PA3, EL, OBP (Krylov *et al.*, 2007)] suggesting it is an essential component of the  $\phi$ KZ-like phages. However, the function, composition, and role in infection of the IB has remained elusive since it was first identified (Krylov *et al.*, 1984). The IB may well prove to be multi-functional and could have roles related to DNA packaging, packaged genome structure, DNA ejection, or it may play a role in all of these aspects of phage development. More specifically, the structurally defined IB may be important for size determination of the T=27 capsid and capsid stability; spooling of DNA tightly around the IB could reduce some of the DNA pressure that might otherwise be exerted on the capsid shell. Spooling of the DNA around the IB may also be necessary to organize such a length of dsDNA at ~500 mg/ml in a sufficiently ordered structure that allows efficient and rapid ejection (Black and Thomas, 2011). The IB may also represent a reservoir of proteins injected into the host cell together with the DNA that are necessary for host takeover. Supporting this hypothesis is that the IB is not detected after infection and injection of the packaged DNA (Krylov *et al.*, 1984). Moreover, in addition to the IB, a number of enzymes [e.g., multi-subunit- $\beta$ , $\beta'$ -RNA polymerase (RNAP)] are also found within the  $\phi$ KZ head with an unknown structural relationship to it.

Previous biochemical studies have illustrated the complexity of the  $\phi$ KZ head, containing at least 30 (Lecoutere *et al.*, 2009), but possibly more than 40 (Mesyanzhinov *et al.*, 2002) different proteins. This is in contrast to the head of T4 which contains 13 different proteins (Black *et al.*, 1993). The location of two head proteins (gp120 and gp89) in the complex  $\phi$ KZ SDS-PAGE profile were identified (Mesyanzhinov *et al.*, 2002), however ambiguity regarding the exact locations of other head proteins presented a major hurdle for even simple studies of the head. Also, other than gp89 and gp120, the identity of any other major or high copy number proteins in the  $\phi$ KZ head was unknown. However, the identity of high copy number head proteins is critical for further studies of the IB as any such proteins would become candidates for components of this large structure. Similarly, while gp120 and gp89 were identified as processed (Mesyanzhinov *et al.*, 2002) it was undefined whether other  $\phi$ KZ head proteins were cleaved or not, or which protein represented the prohead protease in any of the  $\phi$ KZ-like phages. Comprehensive mass spectral analyses of *P. chlororaphis* phage 201  $\phi$ 2-1 demonstrated that thirteen of its proteins were cleaved at a motif reminiscent of the T4 processing motif (Thomas *et al.*, 2010). However, the exact cleavage sites in all  $\phi$ KZ head proteins could not be inferred from the 201  $\phi$ 2-1 analyses since these phages are not closely related as illustrated by their homologous virion proteins having ~20–70 percent identity as determined by Blastp (Thomas *et al.*, 2008).

Here we report mass spectrometry (MS) and biochemical analyses of the  $\phi$ KZ tailless full head mutant and a prohead precursor and show that cleavage of numerous head proteins occurs in  $\phi$ KZ. This processing must play a major role in the formation of the IB, in addition to the capsid. Furthermore, using advanced bioinformatic approaches we have identified the  $\phi$ KZ prohead protease which cleaves multiple head proteins at a consensus amino acid recognition motif. This protease is strongly conserved among all known members of this group of phages, reflecting its important role in morphogenesis. We also identify six high copy number proteins (~100–200 copies) within the  $\phi$ KZ head as being IB proteins. This determination was made based on their release from full heads and by their presence in terminase-inhibited protein precursors-processed DNA-free proheads. These characteristics

lay the groundwork necessary for a genetic determination of the function(s) of the extraordinary IB structure.

## Results and Discussion

### Mass spectrometry of $\phi$ KZ tailless full heads

HPLC-electrospray ionization tandem mass spectrometry (HPLC-ESI-MS/MS) analyses of CsCl purified  $\phi$ KZ tailless mutant (Lavigne *et al.*, 2009, Fokine *et al.*, 2005a, Plotnikova *et al.*, 1982) tailless full heads permitted assignment of peptides to 50  $\phi$ KZ gene products (Figure 1). Twenty-nine of these proteins have previously been identified in tailless heads (Lecoutere *et al.*, 2009). However, one protein identified by Lecoutere *et al.*, gp128, was not found in our analyses. Conversely, 19 proteins were identified in the current study that were not reported to be present in tailless heads and/or wild type  $\phi$ KZ particles (Lecoutere *et al.*, 2009). Detection of these additional proteins most likely resulted from the approach used here which permitted sampling of low molecular weight proteins and low abundance proteins. Eleven proteins identified in this study (gps 27, 85, 91, 92, 119, 144, 153, 157, 184, 244 and 298) were not identified previously in the tailless heads, although they are  $\phi$ KZ virion proteins (Lecoutere *et al.*, 2009). Eight proteins (gps 42, 81, 96, 129, 177\_2, 199, 203 and 228) were identified in the tailless heads in this study that were not previously found in the  $\phi$ KZ virion. These proteins are unlikely to be contaminants as the heads were judged to be highly purified by electron microscopy and the reversion rate of the mutant is about  $10^{-5}$ – $10^{-6}$ . Many of these newly identified proteins are expected to be additional head proteins based on their possessing characteristics indicative of head proteins. For instance, the predicted function of gp129 is that it forms the portal (Cornelissen *et al.*, 2011). Similarly, other proteins have qualities that support them as being head proteins, such as being cleaved at a head prohead protease cleavage motif (e.g., gps 92, 96, 153, 129 and 203), and/or their gene being positioned within, or close to, a major head gene module (e.g., gps 91 and 177\_2)(Figure 2).

### Cleavage of $\phi$ KZ head proteins

Information gained by comparing observed and expected MS sequence coverage and 1-D gel migration provided insight into the catalog of  $\phi$ KZ head proteins that had undergone proteolytic cleavage (Table 1). However, the strongest evidence for cleavage was provided by the presence of a “semi-tryptic peptide”, that is, a peptide with a non-tryptic terminus (i.e., no arginine or lysine) in high confidence peptide assignments; these semi-tryptic peptides were detectable by including “semi-tryptic” cleavage as a parameter in the database searches of the tandem MS data. These efforts were facilitated by the detection of peptides from four different  $\phi$ KZ-derived preparations that underwent MS and are reported in this manuscript. High peptide coverage, particularly for the abundant proteins, was frequently observed (see Figure 2). Forty peptides were found that had been produced by cleavages immediately after a glutamic acid, confirming a consensus processing site(s) for 19 different proteins (Figure 2, Supplementary Table 1). Summation of the number of  $\phi$ KZ head proteins and their observed cleavage sites indicates the remarkable amount of processing that must be undertaken by the morphogenetic protease – the cleavage of at least 6200 peptide bonds per prohead, perhaps several fold more.

### Cleavage of $\phi$ KZ head proteins occurs at a modified T4-like protease motif

Inspection of the 39  $\phi$ KZ cleavage sites showed that in addition to cleavage occurring invariably immediately after a glutamic acid there is a strong preference for a serine in the third position upstream of the cleavage site (26 instances), although an alanine or a glycine is also accepted in this position (8 and 4 instances, respectively, Supplementary Table 1). In one instance (gp303) cleavage apparently occurred where phenylalanine is in the third

position upstream of the cleavage site. Notably, there is little requirement for any other specific residues, particularly downstream of the cleavage site (Supplementary Figure 2). Overall, the  $\phi$ KZ consensus cleavage sequence (S/A/G-X-E, where X is any amino acid) is similar to, but not the same as that of the T4 prohead protease which cleaves after the motif I/L-X-E (more than 13 cleavages in eight different precursor proteins)(Black *et al.*, 1993). In T4, there is also no requirement for specific residues downstream of the cleavage site. Also, in T4 there is evidence that a shift in the processing motif within a polypeptide results in a retargeting of the cleavage site by the protease (Black *et al.*, 1993, Mullaney and Black, 1996).

Identification of cleavage sites was critical for length determination of the mature polypeptide(s) for each  $\phi$ KZ protein (Table 1). Notably, for some proteins, processing was so extensive that only 30–50% of the precursor protein is retained in the mature head (e.g., gp97, see Table 2). The manner in which several proteins were processed is intriguing, with several proteins cleaved into two or three polypeptides that remain in the mature head (e.g., gps 89, 97, 177, 303). Some of the smaller fragments (e.g., gp89N and gp97N) may be non-functional debris and unable to be cleared from the capsid due to their locale or amino acid composition, similar to the case for the phage T4 “acid soluble peptides” that are found within the mature capsid; alternatively these fragments may have unknown functional significance.

Cleavage sites identified in both gps 89 and 120 were consistent with the *N*-termini identified by Edman degradation (Mesyanzhinov *et al.*, 2002). However, additional processing sites, *N*-terminal to those found by Edman degradation were also identified in both these proteins, as evidenced by the presence of low abundance semi-tryptic peptides (see Figure 3A for gp120). Similar semi-tryptic peptides that indicated that additional cleavages had occurred upstream of the maturation cleavage site were also identified in a number of other proteins [e.g., gp93 (Figure 4B) and gps 94, 95, 97, 162, 303 (Figure 2)]. The low abundance of these semi-tryptic peptides suggests that they are not present in every particle; the fact that they were identified at all is a consequence of the low detection limits of the MS analyses. Identification of these additional processing sites is important because they demonstrate that there is a multiplicity of processing in the precursor or immature regions of the  $\phi$ KZ head proteins. Examination of the regions in which multiple cleavages have been confirmed by semi-tryptic peptides typically also revealed other potential cleavage motifs in these regions (see Figure 3A and B), suggesting that the detected low abundance cleavages represent a fraction of what actually occurs in the  $\phi$ KZ head during maturation. It seems feasible that there is a redundancy of cleavage sites in these regions that may have evolved as a protective measure to ensure removal of regions of the protein not required in the mature head, raising the question as to whether there is a preferred length for cleaved fragments to ensure optimal removal from the prohead. Observation of sequences in the mature regions of proteins apparently adhering to the cleavage motif but which are apparently never cleaved (e.g., GFE-457 or ALE-723 in gp120, Figure 3A) indicates there are likely to be additional specificity or structural determinants required for cleavage in addition to the S/A/GXE motif.

### **Cleavage of numerous prohead proteins likely represents an essential head morphogenesis step in all $\phi$ KZ-related phages**

The identification of so many cleavage sites in  $\phi$ KZ head proteins enables us to draw conclusions as to the cleavage of head proteins in other phages, particularly 201  $\phi$ 2-1 and  $\phi$ PA3, phages that are members of the  $\phi$ KZ-like phage genus (Cornelissen *et al.*, 2011). Comparison of the identified cleavage sites in this study in  $\phi$ KZ and those identified in 201  $\phi$ 2-1 indicates that their respective proteases have very similar cleavage site residue requirements (S/A/G-X-E) and this requirement is likely to be a feature of phages in the

$\phi$ KZ genus. The cleavage sites in 12  $\phi$ KZ proteins (e.g., gps 86, 89, 96, 153, and 162) are conserved with identified cleavage sites in their respective homolog in 201  $\phi$ 2-1 (Thomas *et al.*, 2010) and from this work we can infer about another 15 cleavage sites in 201  $\phi$ 2-1 head proteins (Table S2). From the identified cleavage sites in  $\phi$ KZ and 201  $\phi$ 2-1 proteins we can also infer cleavages occur at about 27 sites in homologous proteins in  $\phi$ PA3 (Table S2).

We observed that frequently the degree in which a  $\phi$ KZ cleavage site was conserved in its homologs correlated with the overall degree of conservation of the homologs themselves (see Table S4). For instance, the cleavage sites in two highly conserved proteins (major capsid and protease, see below) are the only sites perfectly conserved in their 201  $\phi$ 2-1 and  $\phi$ PA3 homologs with regard to the three residues upstream of the processing site (Figure 3Ci and ii). However, despite more variations in cleavage site residues with more diverse homologs, we typically observed a conservation of putative cleavage site residues that mirrored the pattern seen in  $\phi$ KZ cleavages. A number of the  $\phi$ KZ cleavage sites are conserved in homologous proteins only with regard to the first and third residues upstream of the processing site, as seen in the homologs to gps 84 and 92 (SXE, Figure 3Ciii and iv). Similarly, in some homologous proteins the third residue upstream of a proposed cleavage site is replaced with an A or G as observed in some  $\phi$ KZ cleavages (e.g., Figure 3Cv and vi). In homologs of other  $\phi$ KZ processed proteins, sequences that adhere to the cleavage motif were frequently found in a position not aligning with, but within a few residues of the actual point of cleavage in the  $\phi$ KZ protein. We believe these motif-like sequences also represent cleavage sites (e.g., Figure 3Cvii).

Notably, we were unable to infer corresponding cleavage sites in the homologs to several  $\phi$ KZ proteins (egs. gps 93, 95, 97 and 303, Table S2). Typically, these are not strongly conserved proteins, however we suspect that numbers of these proteins in  $\phi$ KZ, and most likely their homologs in related phages, have an important role in the mature particle based on their abundance in the  $\phi$ KZ head (see below). Overall, the identification of numerous cleavage sites in  $\phi$ KZ greatly advances our knowledge and expectations of cleavage events in the  $\phi$ KZ-like phages. It also emphasizes that prohead protease cleavages form a major step in capsid maturation for these phages.

### Identification of $\phi$ KZ gp175 as the morphogenetic protease

Sequence-sequence based searches were unable to identify the enzyme responsible for the extensive number of cleavages in  $\phi$ KZ head; hence we sought to identify the  $\phi$ KZ prohead protease using Hidden Markov Model (HMM) based searches. To do this we used local implementations of the Sequence Analysis and Modeling System [SAM (Hughey and Krogh, 1996, Karplus *et al.*, 1998)] and HHpred (Söding, 2005). Initially, Hmmscore (a profile to sequence search) of a custom T4 gp21 model against a library of proteins from four  $\phi$ KZ-like phages ( $\phi$ KZ, EL,  $\phi$ PA3 and 201  $\phi$ 2-1) identified homologous proteins in the four phages: EL gp192 (E-8),  $\phi$ KZ gp175 (E-6), 201  $\phi$ 2-1 gp168 (E-5) and  $\phi$ PA3 gp205 (E-3).

To challenge these matches, an alignment of these four protease candidates was constructed using SAM and then converted into a HHM and used for HHpred searches (profile-profile comparisons). An HHsearch of the  $\phi$ KZ gp175 family HHM against all HHpred libraries (from <ftp://toolkit.lmb.uni-muenchen.de/HHsearch/databases/>) found the best match to be to PHA00911 gp21 prohead core scaffolding protein and protease. The E-value of this match was only 3, however the HHpred P-value, which takes secondary structure into account, was 2.5E-05. This match was solidified using a custom T4 gp21 HHM powerful enough to detect an HHM based on the U35 peptidase family at E-values of 3.5E-05. Members of the U35 peptidase family contain diverged, but known, homologs to the T4 protease, such as the HK97 protease (Liu and Mushegian, 2004, Cheng *et al.*, 2004). When searched against the

custom gp21 HHM the  $\phi$ KZ gp175 HHM scored slightly better ( $2.3E-06$ , Figure 4) than the U35 family HHM, indicating that  $\phi$ KZ gp175 and its homologs in the  $\phi$ KZ-related phages are homologous, albeit extremely diverged, from T4 gp21.

To confirm the identity of gp175 as the  $\phi$ KZ prohead protease we expressed its predicted mature form in *Escherichia coli* DH10B. An extract incubation assay of gp175 using expressed full length gp93 as a substrate showed little non-specific proteolysis of other proteins, whereas the majority of gp93 observable by SDS-PAGE was cleaved in a good yield to a length consistent with that found in the mature virion (Figure 5). Protease active site knockout control incubations (data not shown) will be published together with a more detailed enzymatic characterization.

### **$\phi$ KZ gp175 conserves catalytic residues and structural elements of serine proteases**

$\phi$ KZ gp175 conserves the proposed serine protease active site residues H-85 and S-140 in T4 gp21 (Figure 4); these residues have been shown to be conserved across the phage U9 and U35 families to the S21 family that contains herpesvirus protease (Liu and Mushegian, 2004, Cheng *et al.*, 2004). The final residue of the serine protease catalytic triad (an aspartate or glutamate in classical serine proteases) is less well conserved throughout the phage/virus phylogenetic spectrum (Cheng *et al.*, 2004). In the U9 family this amino acid position is most often an alanine or glycine (Cheng *et al.*, 2004), while in the  $\phi$ KZ-like phages it is a threonine (Figure 4B). This third residue of the catalytic triad is not considered to play a major role in catalysis and it has been proposed that for the U9 family catalysis might only require a catalytic dyad (Cheng *et al.*, 2004). Alternately, the third residue of the catalytic triad might have migrated—possibly to the well-conserved aspartate two residues downstream of the alanine/glycine residue (Cheng *et al.*, 2004), which is also conserved in the  $\phi$ KZ-like protease candidates (Figure 4B).

The structural elements from a major portion of the herpesvirus protease ( $\beta 1 - \beta 6$ ), and other members of the peptidase S21 superfamily, also have predicted secondary structure counterparts in phage proteases, with the exception of the herpesvirus  $\alpha B$ , which is involved in homodimer formation (Liu and Mushegian, 2004, Cheng *et al.*, 2004). Counterparts to each of the herpesvirus protease elements from  $\beta 1 - \beta 5$ , with the exception of  $\beta 6$ , are also predicted for  $\phi$ KZ gp175 (Figure 4).

### **C-terminal cleavage of gp175 may indicate activation by autocleavage**

$\phi$ KZ gp175 is a low copy head protein in the mature head based on an assignment of a relatively low number of spectra to the ~24 kDa protein in our MS experiments (29, compared to 1405 spectra for gp120, the ~65 kDa major capsid protein). This is similar to T4 where only three copies of the T4 protease remain in the mature head, although there are 72 copies of the zymogen present in the unprocessed prohead (Black *et al.*, 1993). Based on the T4 precedent, it is likely to be biological important that  $\phi$ KZ gp175 is C-terminally cleaved at E-210. In T4 the protease is the only head protein known to be C-terminally cleaved, probably at LAE-206 (Keller and Bickle, 1986). Notably, this cleavage of the protease zymogen is an autocleavage event that activates the protease (Showe *et al.*, 1976a, Showe *et al.*, 1976b). The protease auto-activation step is closely controlled by prohead assembly, likely regulated by addition of the vertex proteins (Black *et al.*, 1993). We expect that the processing of  $\phi$ KZ gp175 also activates its zymogen form.

### **Processed $\phi$ KZ head proteins of known or inferred function**

Only about one quarter of the processed  $\phi$ KZ proteins have either a known function or one that can be inferred by bioinformatics. These are the major capsid protein (gp120), the portal protein (gp129), the protease (gp175), the virion RNAP beta' subunit (gp180), and a helicase

(gp203). Several of these proteins have novel or unusual properties. For instance, the fact that the  $\phi$ KZ portal is processed is unusual as processing of the portal protein is an event rarely reported in tailed phages, although it is known that 21 and 26 residues are removed from the *N*-termini of the  $\lambda$  and P2 portal proteins, respectively (Walker *et al.*, 1982, Rishovd *et al.*, 1994, Rishovd *et al.*, 1998). In fact in T4 the portal is the only early prohead protein that is not processed despite this phage having a relatively high number (13) of processed proteins in its capsid (Black *et al.*, 1993). That the  $\phi$ KZ portal may have such an unusually large propeptide (262 residues) removed suggests it may have more than just a core-targeting function and may contribute significantly to the core itself, possibly connecting the portal to the precursor form of the IB in the prohead.

Helicases and multi-subunit RNAPs have not commonly been found in the virions of characterised phages. In  $\phi$ KZ, presumably the RNAP subunits are injected into the host for transcription of early genes (Thomas *et al.*, 2008). Possibly, they could have an additional role in genome translocation out of the capsid as in T7 (Grayson and Molineux, 2007). Possible functions for gp203 (one of three helicases encoded by  $\phi$ KZ) are more speculative. Gp203, homologous to the SNF-2 helicases, might be injected into the host where it may have a role in DNA transcription or replication, or it might be involved in organization of the DNA within the head. The identification of *C*-terminal cleavages at a protease motif in the RNAP beta' subunit and gp203 helicase is a novel finding and supports these enzymes being incorporated into the head based on our knowledge of T4 head morphogenesis (Black *et al.*, 1993).

### Abundant $\phi$ KZ head proteins that are candidates for major IB components

The major capsid protein for  $\phi$ KZ has been reported to be present in 1560 copies per virion (Fokine *et al.*, 2005a). In this study, the relative abundances of the proteins in the  $\phi$ KZ head were estimated from the number of peptide mass spectra assigned with high confidence, after normalization on the basis of calculated molecular weight (SC/Mw), and densitometry. SC/Mw for gps 89, 90, 93, 95, 97 and 162 were substantially higher than those for most of the other head proteins, ranging from 4.5–5.7 (Table 1). The SC/Mw for the major major capsid protein, gp120, was 25.8. However, as a consequence of saturation effects on relative quantification of the major capsid protein using this approach (Thomas *et al.*, 2010), the copy numbers derived from SC/Mw are expected to be over-estimated for gps 89, 90, 93, 95, 97 and 162. In order to obtain an alternative assessment of relative quantities, image analysis was conducted of the Coomassie-stained SDS-PAGE bands containing these six proteins in both purified phage and tailless heads. The densitometry results were in agreement with the MS data, supporting the conclusion that these six proteins are present in high abundance (>100 copies per head, Table 2). This indicates that each  $\phi$ KZ head is composed of more than 2000 protein subunits.

A number of lines of evidence identify these six abundant head proteins as probable major components of the IB. First, the genes encoding five of the abundant head proteins (gps 89, 90, 93, 95 and 97) are located within a single gene module (Figure 2). Although genes with related morphogenetic functions are often observed to be clustered in other phage genomes, the clustered location of these five genes is in  $\phi$ KZ more significant because of the otherwise overall absence of the normal head gene synteny in  $\phi$ KZ (Figure 2). Second, the sum of the observed masses of the IB candidate proteins is consistent with the total estimated mass of the IB (15 – 20 MDa)(Wu *et al.*, 2012). It is important to note that as a consequence of processing, the molecular weights of the mature forms of the six abundant  $\phi$ KZ IB proteins are remarkably similar (Table 2), despite major differences in the sizes of their precursors. Notably, three of the abundant proteins, gps 93, 95 and 162, contain a paralogous domain (PF12699) that is also present in two of the non-abundant head proteins, gp94 and gp163. By analogy to gene duplication events in other phages [e.g., T4 gps 23 and

24 (Fokine *et al.*, 2005b)], the  $\phi$ KZ proteins containing the PF12699 domain are likely to have evolved specialized, although related, functions. Each of the  $\phi$ KZ-related phages has a set of paralogs belonging to this family which we predict to have a major influence on the IB structure.

### $\phi$ KZ internal head proteins

The similarity in the proteolysis of the  $\phi$ KZ head proteins to those in T4 suggests similar overall head morphogenesis. In T4 an immature, uncleaved form of the prohead is first assembled around a structure-determining scaffolding core. The immature prohead is then converted by the gp21 protease to a more mature prohead whose outer shell is composed of the cleaved capsid proteins gp23 and gp24. Additionally, in the mature prohead most of the internal form-determining scaffolding components (gps 22, 21, 67 and 68) have been proteolytically removed. Other scaffolding components nonessential for prohead assembly undergo only limited *N*-terminal processing and are retained by the mature prohead into which DNA is then packaged by the terminase. Upon phage infection these proteins [the internal proteins (IPs), IPI, IPII, IPIII and Alt] are injected into the host together with the DNA (Black *et al.*, 1993). Addition of the phage T4 terminase inhibitor 9-aminoacridine (9AA) to infected bacteria blocks DNA packaging and leads to the accumulation of fully processed mature T4 proheads that contain the cleaved IPs but lack DNA (Schaerli and Kellenberger, 1980).

Host uptake of 9-aminoacridine has a strikingly similar effect on  $\phi$ KZ head morphogenesis enabling us to isolate a comparable packaging-inhibited  $\phi$ KZ prohead in which the abundant, candidate IB processed head proteins are components (Figure 6). The 9AA particles isolated by glycerol gradient centrifugation (Figure 6A) contained the mature capsid protein and a number of other abundant inner body proteins in proportions comparable to those found in the mature capsid (Figure 6B). The particles contained little, if any DNA (Figure 6C). Significantly, electron microscopy of these particles revealed a kernel-like object within the DNA-free prohead which apparently represents a precursor or degraded form of the IB that evidently can assemble in the absence of DNA packaging (Figure 7, Supplemental Figures 3 and 4). Isolation of these processed mature proheads shows that the  $\phi$ KZ protease operates like the T4 protease, prior to and independent of DNA packaging, and unlike the adenovirus protease that translocates along the packaged DNA to access its cleavage sites in the capsid (Mangel *et al.*, 2003).

Proteins released from the T4 capsid shell by osmotic shock were observed to be either scaffolding core-related or IP-related (Black *et al.*, 1993). Osmotic shock treatment of  $\phi$ KZ purified tailless heads followed by separation of the empty shells (Supplemental Figure 5) from the released proteins, enabled the identification of proteins released from the capsid, and therefore, most likely packaged within the capsid itself and not tightly associated with the capsid shell (Figure 8). The released proteins included the abundant IB proteins listed in Table 2 (Figure 8) and based on the T4 precedent are likely to have scaffolding core- and/or IP-related functions.

An additional experiment also indicates that most of the abundant  $\phi$ KZ IB proteins do not have one T4 IP-like signature characteristic. The T4 IPs and Alt precipitate with the still-condensed genomic DNA released from full heads by disruption of the capsid by treatment with 70% acetic acid (generally a protein solubilizing solvent) whereas these proteins are soluble as isolated from heads or proheads lacking DNA (Zachary and Black, 1991). The other T4 head proteins, with the exception of the portal protein, are found in the supernatant (Zachary and Black, 1991). The same treatment of  $\phi$ KZ tailless heads also resulted in most proteins being enriched in the supernatant as assessed by SDS-PAGE and the SC/Mw values for each protein in both fractions (Figure 9). As expected, most of the DNA was found in the



pellet. Only two processed, encapsidated proteins – the abundant protein (gp93) and the low abundance helicase (gp203) – were enriched in the pellet fraction. By this criterion therefore the soluble gp93 IB protein has a unique T4 IP-like association with the DNA unlike the remainder of the IB proteins. This suggests that most of the IB proteins are segregated away from the DNA in the protein rich cylinder (Wu *et al.*, 2012), whereas the T4 IPs are dispersed within the DNA (Black and Thomas, 2011).

### **Conservation of many $\phi$ KZ internal head proteins is analogous to the degree of conservation of T4 proteins with scaffolding/core functions**

Comparison of the conservation of the homologs to the processed proteins in T4 and  $\phi$ KZ among the members of their respective families supports that the candidate  $\phi$ KZ IB proteins, and in fact the majority of  $\phi$ KZ processed proteins, do not behave like T4 IP proteins, but more like the T4 scaffolding core proteins. Based on the degree by which the T4 processed proteins are conserved through the T4-like phages (from closely related phages, to more diverged members, Table S3) these proteins can be roughly be divided into three groups: (1) proteins that are highly conserved to approximately the same degree as known essential proteins (e.g., terminase, portal and tail sheath proteins), and include the major capsid protein, gp21 and the core-related protein, gp68; (2) proteins that are more diverged than the highly conserved proteins, but for most of which homologs could still be identified in the more distantly related T4-related phages (e.g., gp67 and gp22); and (3) proteins that are poorly conserved, and for which counterparts were not identifiable in many phages (e.g., IP and Alt proteins). Notably, the genes encoding the proteins in categories (1) and (2) are all located in one major region, however the genes encoding the IPs and Alt are in three different locales in the T4 genome.

A comparison between the processed proteins in  $\phi$ KZ and their homologs in the related phages 201  $\phi$ 2-1,  $\phi$ PA3 and EL (Table S4) revealed that all processed  $\phi$ KZ proteins have identifiable homologs in 201  $\phi$ 2-1 and  $\phi$ PA3, and the more highly conserved of these proteins have very diverged counterparts in EL [recently proposed to belong to a separate genus, the EL-like phages (Cornelissen *et al.*, 2011)]. Generally, the proteins of the phages 201  $\phi$ 2-1 (and  $\phi$ PA3) have similar degrees of divergence from their  $\phi$ KZ homologs as KVP40 proteins have from their T4 homologs (Thomas *et al.*, 2008). Therefore the very existence of identifiable homologs to the  $\phi$ KZ processed proteins in 201  $\phi$ 2-1 and  $\phi$ PA3 illustrates that their conservation is very dissimilar to the poor conservation of the IP proteins in the T4-like phages.

Several of the  $\phi$ KZ processed proteins are conserved to a comparable degree as the major morphogenetic proteins terminase and sheath, including the major capsid protein, portal, protease, virion RNAP subunits, helicase-gp203, gp89 (an abundant protein) and gp184 (Table S4).  $\phi$ KZ processed proteins more moderately conserved (divergence of 50–70%) include gps 84, 92, 163, 177 and 90, the latter an abundant head protein for which processing has not been observed. Lesser conserved processed  $\phi$ KZ proteins (70–80% divergence) include gps 86, 93, 94, 95, 96, 97 and 162. The degree of conservation of these more moderately and lesser conserved  $\phi$ KZ processed proteins is reminiscent of the conservation of the various T4 proteins with scaffolding core-related functions to their homologs in KVP40.

### **Concluding Remarks**

Proteolysis of head proteins is a critical step in the morphogenesis of many tailed phages, facilitating the conversion of the prohead to the mature capsid. In this study we show sequences of the cleavage sites in nineteen head proteins in  $\phi$ KZ. We identify informatically and directly by enzyme activity the protease likely responsible for all of these cleavages as

judged by action at consensus S/A/G-X-E recognition sequences at 39 cleavage sites. Processing of  $\phi$ KZ head proteins thus occurs at a sequence similar but different from the I/L-X-E recognition sequence of the distantly related T4 phage protease. We found evidence of multiple processing sites in nine  $\phi$ KZ proteins. This appears to represent a built-in mechanism by which the phage ensures that the majority of the propeptide regions are removed and emphasizes the essential nature of processing in  $\phi$ KZ head morphogenesis.

The other main conclusion of our study is the identification of six similar size, similar abundance internal head proteins that provide the biochemical basis for the regular multi-tiered IB cylinder structure (Wu *et al.*, 2012). Potentially, these six proteins could be segregated within individual tiers, or mixed in few or many combinations in several or more of these tiers. Further work will be required to determine their structural organization. Further study is also required on the IB precursor; the extensive cleavages in the IB proteins suggest that this structure may be quite different and not readily inferred from its mature form. Studies of related phages with differences in IB protein composition will enhance our understanding of the  $\phi$ KZ IB, and differences in IB structure among these phages may also help explain the major differences in genome length (280 kb,  $\phi$ KZ; 316 kb, 201  $\phi$ 2-1; 308 kb,  $\phi$ PA3; 211 kb, EL) found among these phages whose capsids are of similar size (Krylov *et al.*, 2007). The determination of whether or not the (thus far) unique IB of the PhiKZ-like phages is essential and, if so, the nature of its function(s) may be answered by a combined genetic, biochemical and structural approach.

## Experimental procedures

### Bacteria and phages

The Lory strain of *Pseudomonas aeruginosa* PAO1 (BEO3) was kindly provided by Dr Robert Ernst. Phage  $\phi$ KZ (Mesyanzhinov *et al.*, 2002, Fokine *et al.*, 2007) and its tailless mutant (Lavigne *et al.*, 2009, Fokine *et al.*, 2005a, Plotnikova, *et al.*, 1982) were generous gifts of Drs Andrei Fokine and Konstantin Miroshnikov, respectively. We also thank Dr Miroshnikov for advice on propagation of the tailless mutant.

### Purification of phages

$\phi$ KZ was propagated as described previously (Wu *et al.*, 2012). Purification of the tailless mutant was by step CsCl gradient as described previously (Fokine *et al.*, 2005a), with the exception that particles were grown in liquid culture and concentrated by PEG precipitation. An overnight culture of *P. aeruginosa* was diluted 1:1000 into 1L of LB and grown with shaking at 37°C until OD<sub>600</sub> was 0.33. The culture was then transferred to 40°C and infected with the tailless mutant phage at a MOI of 5 and grown for 2 hr, 15 min until lysis was observed. Bacterial debris was removed by a low speed spin (10,400 x g, 5 min) and the sample treated with Pancreatic DNAase I (Roche)(1500 U, 37°C, 30 min). Heads were concentrated by slow stirring overnight at 4°C in 1M NaCl and 10% PEG (final concentration) and then pelleted (11,300 x g, 4°C, 30 min). Pellets were resuspended in 20 ml SM buffer [50 mM Tris-HCl (pH 7.5), 100 mM NaCl, 8 mM MgSO<sub>4</sub>, 0.002% gelatin] containing Complete Protease Inhibitor (Roche) and then layered (5.8 ml/tube) onto CsCl step gradients composed of the following concentrations of CsCl: 1.59 g ml<sup>-1</sup> (1.0 ml), 1.52 g ml<sup>-1</sup> (0.75 ml), 1.5 g ml<sup>-1</sup> (1.5 ml), 1.30 g ml<sup>-1</sup> (1.5 ml) and 1.21 g ml<sup>-1</sup> (1.5 ml). The buffer used throughout the gradient was 10 mM Tris-HCl (pH 7.5) and 1 mM MgCl<sub>2</sub>. Tubes were spun (108,000 x g, 10°C, 3 h) in a Beckman Coulter SW41 rotor. The band judged to contain only heads by SDS-PAGE and transmission electron microscopy had a buoyant density of 1.46 g/ml. This is higher than the buoyant density of purified intact phage (1.36 g ml<sup>-1</sup>). Heads were found to be unstable upon dialysis unless the CsCl was removed by dialyzing initially against a buffer composed of 50 mM Tris-HCl (pH 7.5), 5M NaCl and

100 mM MgCl<sub>2</sub> (4°C), followed by a series of buffers composed of reducing concentrations of salts, concluding at 50 mM Tris-HCl (pH 7.5) and 10 mM MgCl<sub>2</sub>.

### Production of DNA-deficient prohead particles

Treatment of  $\phi$ KZ with the T4 terminase inhibitor 9-aminoacridine (9AA) was generally as described for T4 (Schaerli and Kellenberger, 1980), with the exception that preliminary experiments were undertaken to determine the concentration of 9AA required to inhibit phage particle production (15  $\mu$ g ml<sup>-1</sup>); the concentration that was subsequently used was 10-fold higher than required for the same effect on T4. This can be attributed, in part, to the higher tolerance of *P. aeruginosa* to 9AA than *E. coli* (Wainwright *et al.*, 1997). *P. aeruginosa* was propagated in 100 ml of M9S media at 37°C with shaking until OD<sub>600</sub> 0.45. Purified  $\phi$ KZ was used to infect at a MOI of 5; 9AA was added 30 seconds after infection. At OD<sub>600</sub> 0.68 the cells were harvested by centrifugation at 10,400 x g for 5 min. The pellet was resuspended in 800  $\mu$ l of SM buffer, DNAase (100 units) and CHCl<sub>3</sub> (40  $\mu$ l) were added and the mixture was incubated for 10 min (37°C) and then centrifuged for 5 min (4,300 x g, 4°C). The supernatant (200  $\mu$ l) was immediately loaded onto a pre-formed 15–45% glycerol gradient (50 mM Tris-HCl, pH7.5; 5 mM MgCl<sub>2</sub>) and centrifuged in a Beckman Coulter SW50.1 rotor (147,000 x g, 1hr, 4°C). Fractions (~350  $\mu$ l) were collected from the bottom of the tube and run on SDS-PAGE (10% polyacrylamide) and DNA agarose gels (0.7 – 1%). Non-dialysed fractions of interest were concentrated approximately 10-fold, by initially diluting the sample with 1280  $\mu$ l of SM and then centrifugation (38,700 x g, 30 min, 4°C) in a SS34 rotor. The pellet was resuspended in SM buffer.

### Preparation of disrupted $\phi$ KZ tailless particles

Purified heads in CsCl were osmotically shocked by dialysis from CsCl to a buffer containing 50 mM Tris-Cl (pH 7.5); 10 mM Mg Cl<sub>2</sub>. Particles were determined to be disrupted by a large increase in the viscosity of the sample such that it prevented it from being pipetted. This viscosity was reduced by treatment with DNAase (10U/200  $\mu$ l of sample) for 10 min (37°C) and the sample was then loaded onto a pre-poured, chilled glycerol gradient and treated as described for the 9AA samples above. SDS-PAGE bands containing proteins released from the capsid underwent MS analyses as described above.

### Acetic acid treatment of $\phi$ KZ tailless heads

Acetic acid treatment of  $\phi$ KZ tailless heads was as previously described (Zachary and Black, 1991). Heads were propagated as described above but instead of purification by CsCl gradient underwent differential centrifugation (2x 7,600 x g, 10 min; 38,700 x g, 30 min, 4°C). Pellets were resuspended in SM buffer and judged to contain only a small percentage of intact phage as judged by comparison of its SDS-PAGE profile with those of intact phage and purified heads. Acetic acid (70% final v/v) was added to the sample, vortexed, left at room temperature for 20 min, and then centrifuged in a benchtop centrifuge (16,000 x g, 4°C, 10 min). Pellets were dried by speedivac centrifugation (20 min, 60°C) then resuspended in 10 mM Tris-HCl (pH 7.5). The supernatant was dialyzed against water (1hr, 4°C), underwent Speedivac centrifugation, and was resuspended in an equal volume of 10 mM Tris-HCl (pH 7.5) as to the pellet fraction. Samples underwent SDS-PAGE, DNA gels and MS analyses as described above. For the MS analyses double the volume of pellet fraction was loaded on the SDS-PAGE gel as for the supernatant fraction. As a consequence, the number of spectra identified for each protein in the pellet fraction was halved for calculations of SC/Mw to enable a more realistic comparison of the abundance of each protein species in pellet versus supernatant fractions.

## Mass spectrometry

Samples were boiled for 10 min in SDS sample buffer (Bio-Rad) prior to electrophoresis on Criterion XT 12% SDS-PAGE reducing gels (Bio-Rad) in the presence of MOPS buffer (Bio-Rad). Proteins were visualized by staining with Coomassie blue. Gel lanes were divided into the following number of slices: acetic acid treatment of  $\phi$ KZ tailless heads, 7 slices; purified heads, 17 slices; disrupted  $\phi$ KZ tailless particles, 7 slices. Efforts were made to avoid transecting visible stained bands. Proteins in the gel slices were digested in situ with trypsin (Promega) and analyzed by HPLC-ESI-tandem mass spectrometry on either a Thermo Fisher LTQ mass spectrometer (acetic acid treatment of  $\phi$ KZ tailless heads) or a Thermo Fisher LTQ Orbitrap Velos mass spectrometer (purified heads and disrupted  $\phi$ KZ tailless particles). For the LTQ analyses, on-line HPLC separation of the digests was accomplished with an Eksigent NanoLC fitted with a PicoFrit™ (New Objective; 75  $\mu$ m i.d.) column packed to 11 cm with C18 adsorbent (Vydac; 218MS 5  $\mu$ m, 300 Å). A strategy was employed in which a survey scan was acquired followed by data-dependent collision-induced dissociation (CID) spectra of the seven most intense ions in the survey scan above a set threshold. For analyses on the Orbitrap, an Eksigent NanoLC-Ultra 2-D HPLC system was used with separation accomplished by a PicoFrit™ (New Objective; 75  $\mu$ m i.d.) column packed to 15 cm with C18 adsorbent (Vydac; 218MS 5  $\mu$ m, 300 Å). Precursor ions were acquired on the Orbitrap at a resolution of 60,000 ( $m/z$  400). Data-dependent CID spectra of the six most intense ions in the survey scan were acquired from the linear trap while the precursor ion spectra were being collected. Mascot (Matrix Science; London, UK) was used to search the uninterpreted CID spectra searched against a locally-generated  $\phi$ KZ protein database that had been concatenated with the SwissProt (version 51.6) database. Methionine was considered as a variable modification and semi-trypsin was specified as the proteolytic agent. Determination of probabilities of protein identifications and cross correlation of the Mascot results with X! Tandem were accomplished by Scaffold (Proteome Software). The tandem MS results obtained from the digest of each gel slice were searched separately using Mascot, and the data files were either evaluated individually or combined into datasets for processing by Scaffold (using the “MudPIT” option).

## Identification of abundant proteins in the $\phi$ KZ head

An estimate of abundance for each protein in the purified head sample was made by dividing its spectral counts by its molecular weight (SC/Mw). To obtain the number of spectra assigned to the mature region(s) of the processed proteins, a “Spectrum report” was exported from Scaffold into Microsoft Excel, which was then saved as tab-delineated text. Spectra over defined regions of each protein were then summed using the perl script MSsort (Thomas *et al.*, 2010). The program may be obtained from S.C. Hardies through the website at [http://biochem.uthscsa.edu/~hs\\_lab/scripting/MSprogs.html](http://biochem.uthscsa.edu/~hs_lab/scripting/MSprogs.html). Estimates of the copy number of each protein identified as abundant by SC/Mw were made by densitometry of the SDS-PAGE bands containing these proteins. Gels prepared with serially-diluted samples of purified  $\phi$ KZ and the tailless mutant were analyzed using ImageJ software (Rasband, 1997–2011). Controls used for copy number estimation of the abundant proteins were the bands containing the major capsid (gp120) and the major sheath (gp29).

## Cryo-electron microscopy and tomography of 9-aminoacridine treated $\phi$ KZ proheads

Cryo-EM was performed using a Philips CM20-FEG electron microscope as described (Cheng *et al.*, 2002).

Tomographic data collection and reconstruction were carried out as described (Harris *et al.*, 2006). Briefly, a sample of 9AA treated particles was mixed 50:50 with a BSA gold suspension (Aurion, Electron Microscopy Sciences, Hatfield, PA), applied to a Quantifoil R2/2 holey carbon grid (Quantifoil, SPI, West Chester, PA) and plunge-frozen using a

Vitrobot (FEI, Hillsboro, OR). Tilt series covering the range from  $-60^{\circ}$  to  $+60^{\circ}$  in  $2^{\circ}$  increments were collected at  $-4 \mu\text{m}$  nominal defocus, 38,500x magnification ( $7.8 \text{ \AA}$  pixel size), and a total electron dose per series of  $\sim 70 \text{ e}^{-} \text{ \AA}^{-2}$  using SerialEM (David N, 2005) on a Tecnai T12 electron microscope (FEI) equipped with a GIF 2002 energy filter (Gatan, Pleasanton, CA) operating in zero-loss mode. Reconstructions were calculated with IMOD (Kremer *et al.*, 1996). Capsids were extracted from the tomogram as subvolumes and denoised by 30 iterations of anisotropic non-linear diffusion (Frangakis and Hegerl, 2001) as implemented in Bsoft (Heymann *et al.*, 2008).

### Expression vector assay of the $\phi\text{KZ}$ protease

The gene encoding the full length form of gp93 was amplified by PCR using the primers 5'-GCGCCATGGCTCTACTTAAAATGCTAGACA-3' and 5'-GCGGGTACCTTACTTCTCATCATTCTCAGG-3'. The gene encoding the mature form of gp175 was amplified with the primers 5'-GCGCCATGGAACCATATCTAATTTTTCTGGT-3' and 5'-GCGTCTAGATTACTCTAATGCAGGGGAATTCCATTTA-3'. Both genes were cloned using the *NcoI* and *XbaI* sites of the vector pHERD20T (Qiu *et al.*, 2008), kindly provided by Dr Hongwei Yu, and transformed into *Escherichia coli* DH10B. Full length gp93 and mature gp175 were induced individually and the cells disrupted by freeze-thaw and lysozyme treatment. Cell lysates containing either gp93, gp175, or the two mixed together, were incubated for 30 minutes at  $37^{\circ}\text{C}$ . Samples were then spun in a benchtop centrifuge to separate soluble and insoluble fractions and the fractions assayed by SDS-PAGE. Only the soluble fractions are shown in Figure 5.

### Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

### Acknowledgments

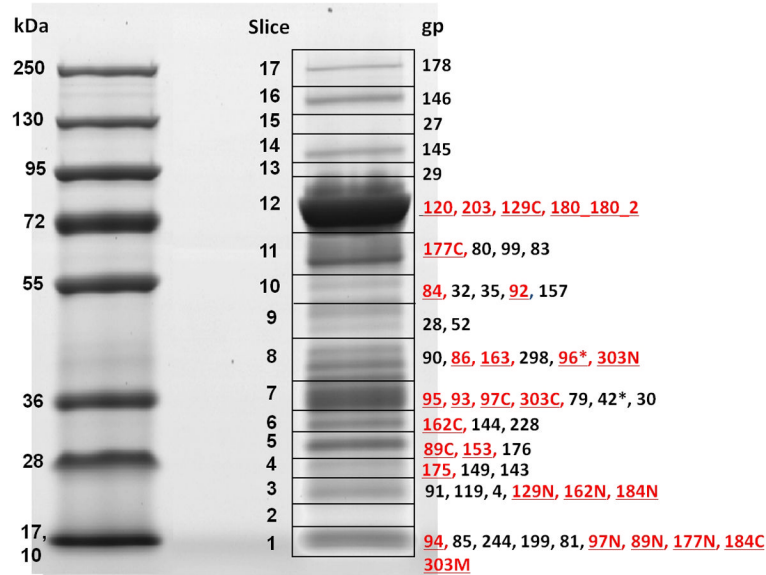
$\phi\text{KZ}$  and the  $\phi\text{KZ}$  tailless mutant were generously provided by Dr. A. Fokine (Purdue University) and Dr. K. Miroshnikov (Russian Academy of Sciences), respectively. *Pseudomonas aeruginosa* PAO1 was kindly provided by Dr R. Ernst (Univ. of Maryland Baltimore). The pHERD20T vector was provided by Dr Yu (Marshall University). We thank Kevin Hakala and Sam Pardo for the excellent MS analyses that were conducted in the UTHSCSA Institutional Mass Spectrometry Laboratory. We thank Ru-ching Hsia (Univ. of Maryland Baltimore) for her help with TEM and Qin Dan for her technical support. We are extremely grateful to Dr. S.C. Hardies for allowing access to his collection of bioinformatics software and the UTHSCSA Bioinformatics Center for assistance with computational aspects of the project. This work was supported by NIH grant AI11676 to L.W.B. and by the Intramural Research Program of NIAMS.

### References

- Black, LW.; Showe, MK.; Steven, AC. Morphogenesis of the T4 head. In: Karam, JD., editor. Molecular Biology of Bacteriophage T4. Washington, DC: ASM Press; 1993. p. 218-258.
- Black, LW.; Thomas, JA. Condensed genome structure. In: Rossman, M.; Rao, VB., editors. Viral Molecular Machines. New York: Springer; 2012. p. 469-488.
- Briers Y, Miroshnikov K, Chertkov O, Nekrasov A, Mesyanzhinov V, Volckaert G, Lavigne R. The structural peptidoglycan hydrolase gp181 of bacteriophage  $\phi\text{KZ}$ . Biochem Biophys Res Commun. 2008; 374:747–751. [PubMed: 18671939]
- Cheng H, Shen N, Pei J, Grishin NV. Double-stranded DNA bacteriophage prohead protease is homologous to herpesvirus protease. Protein Science. 2004; 13:2260–2269. [PubMed: 15273316]
- Cheng N, Trus BL, Belnap DM, Newcomb WW, Brown JC, Steven AC. Handedness of the Herpes Simplex Virus capsid and procapsid. J Virol. 2002; 76:7855–7859. [PubMed: 12097597]

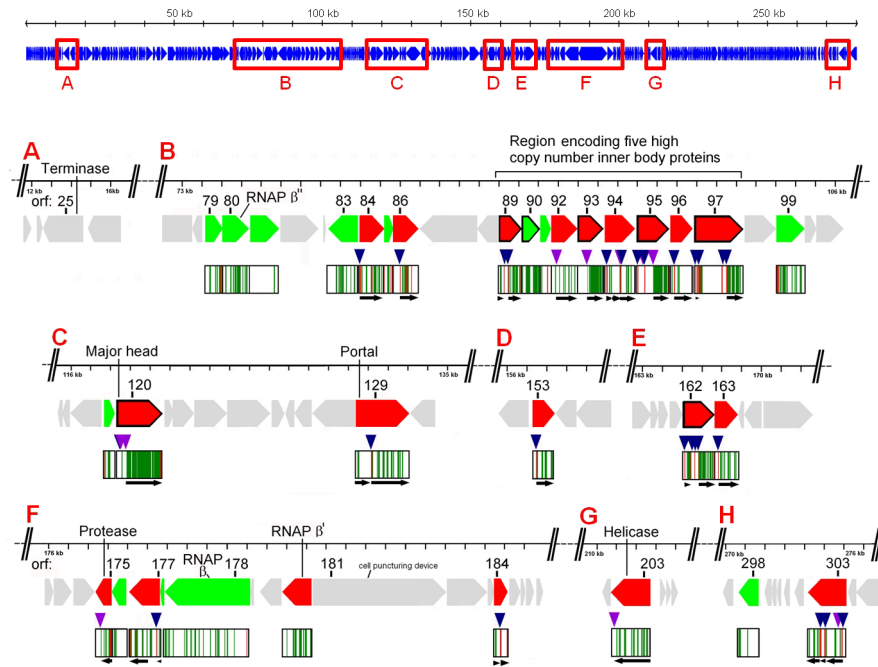
- Cornelissen A, Hardies SC, Shaburova OV, Krylov VN, Mattheus W, Kropinski AM, Lavigne R. Complete genome sequence of the giant virus OBP and comparative genome analysis of the diverse  $\phi$ KZ-related phages. *J Virol*. 2011; 86:1844–52. [PubMed: 22130535]
- David NM. Automated electron microscope tomography using robust prediction of specimen movements. *J Struct Biol*. 2005; 152:36–51. [PubMed: 16182563]
- Fokine A, Battisti A, Bowman V, Efimov A, Kurochkina L, Chipman P, et al. CryoEM Study of the *Pseudomonas* bacteriophage  $\phi$ KZ. *Structure*. 2007; 15:1099–1104. [PubMed: 17850749]
- Fokine A, Kostyuchenko VA, Efimov AV, Kurochkina LP, Sykilinda NN, Robben J, et al. A three-dimensional cryo-electron microscopy structure of the bacteriophage  $\phi$ KZ head. *J Mol Biol*. 2005a; 352:117–124. [PubMed: 16081102]
- Fokine A, Leiman PG, Shneider MM, Ahvazi B, Boeshans KM, Steven AC, et al. Structural and functional similarities between the capsid proteins of bacteriophages T4 and HK97 point to a common ancestry. *Proc Natl Acad Sci*. 2005b; 102:7163–7168. [PubMed: 15878991]
- Frangakis AS, Hegerl R. Noise reduction in electron tomographic reconstructions using nonlinear anisotropic diffusion. *J Struct Biol*. 2001; 135:239–250. [PubMed: 11722164]
- Golshahi L, Lynch KH, Dennis JJ, Finlay WH. *In vitro* lung delivery of bacteriophages KS4-M and  $\phi$ KZ using dry powder inhalers for treatment of *Burkholderia cepacia* complex and *Pseudomonas aeruginosa* infections in cystic fibrosis. *Journal of Applied Microbiology*. 2011; 110:106–117. [PubMed: 20875034]
- Grayson P, Molineux IJ. Is phage DNA ‘injected’ into cells—biologists and physicists can agree. *Curr Opin Microbiol*. 2007; 10:401–409. [PubMed: 17714979]
- Harris A, Cardone G, Winkler DC, Heymann JB, Brecher M, White JM, Steven AC. Influenza virus pleiomorphy characterized by cryoelectron tomography. *Proc Natl Acad Sci*. 2006; 103:19123–19127. [PubMed: 17146053]
- Heymann BJ, Cardone G, Winkler DC, Steven AC. Computational resources for cryo-electron tomography in Bsoft. *J Struct Biol*. 2008; 161:232–242. [PubMed: 17869539]
- Hughey R, Krogh A. Hidden Markov models for sequence analysis: Extension and analysis of the basic method. *Comput Appl Biosci*. 1996; 12:95–107. [PubMed: 8744772]
- Karplus K, Barrett C, Hughey R. Hidden markov models for detecting remote protein homologies. *Bioinformatics*. 1998; 14:846–856. [PubMed: 9927713]
- Keller B, Bickle TA. The nucleotide sequence of gene 21 of bacteriophage T4 coding for the prohead protease. *Gene*. 1986; 49:245–251. [PubMed: 3552886]
- Kremer JR, Mastronarde DN, McIntosh JR. Computer visualization of three-dimensional image data using IMOD. *J Struct Biol*. 1996; 116:71–76. [PubMed: 8742726]
- Krylov VN, Dela Cruz DM, Hertveldt K, Ackermann HW. “ $\phi$ KZ-like viruses”, a proposed new genus of myovirus bacteriophages. *Arch Virol*. 2007; 152:1955–1959. [PubMed: 17680323]
- Krylov VN, Smirnova TA, Minenkova IB, Plotnilova TG, Zhazikov IZ, Khrenova EA. *Pseudomonas* bacteriophage  $\phi$ KZ contains an inner body in its capsid. *Can J Microbiol*. 1984; 30:758–762. [PubMed: 6435847]
- Lavigne R, Darius P, Summer EJ, Seto D, Mahadevan P, Nilsson AS, et al. Classification of *Myoviridae* bacteriophages using protein sequence similarity. *Bmc Microbiol*. 2009; 9:224. [PubMed: 19857251]
- Lecoutere E, Ceysens PJ, Miroshnikov K, Mesyanzhinov V, Krylov VN, Noben JP, et al. Identification and comparative analysis of the structural proteomes of  $\phi$ KZ and EL, two giant *Pseudomonas aeruginosa* bacteriophages. *Proteomics*. 2009; 9:3215–3219. [PubMed: 19526553]
- Liu J, Mushegian A. Displacements of prohead protease genes in the late operons of double-stranded-DNA bacteriophages. *J Bacteriol*. 2004; 186:4369–4375. [PubMed: 15205439]
- Mangel WF, Baniecki ML, McGrath WJ. Specific interactions of the adenovirus proteinase with the viral DNA, an 11-amino-acid viral peptide, and the cellular protein actin. *Cellular and Molecular Life Sciences*. 2003; 60:2347–2355. [PubMed: 14625681]
- Matinkhoo S, Lynch KH, Dennis JJ, Finlay WH, Vehring R. Spray-dried respirable powders containing bacteriophages for the treatment of pulmonary infections. *Journal of Pharmaceutical Sciences*. 2011; 100:5197–5205. [PubMed: 22020816]

- Mesyanzhinov VV, Robben J, Grymonprez B, Kostyuchenko VA, Bourkaltseva MV, Sykilinda NN, et al. The genome of bacteriophage  $\phi$ KZ of *Pseudomonas aeruginosa*. *J Mol Biol.* 2002; 317:1–19. [PubMed: 11916376]
- Miroshnikov KA, Faizullina NM, Sykilinda NN, Mesyanzhinov VV. Properties of the endolytic transglycosylase encoded by gene 144 of *Pseudomonas aeruginosa* bacteriophage  $\phi$ KZ. *Biochemistry (Moscow).* 2006; 71:300–305. [PubMed: 16545067]
- Monson R, Foulds I, Foweraker J, Welch M, Salmond GPC. The *Pseudomonas aeruginosa* generalized transducing phage  $\phi$ PA3 is a new member of the  $\phi$ KZ-like group of ‘jumbo’ phages, and infects model laboratory strains and clinical isolates from cystic fibrosis patients. *Microbiology.* 2011; 157:859–867. [PubMed: 21163841]
- Mullaney JM, Black LW. Capsid targeting sequence targets foreign proteins into bacteriophage T4 and permits proteolytic processing. *J Mol Biol.* 1996; 261:372–385. [PubMed: 8780780]
- Paradis-Bleau C, Cloutier I, Lemieux L, Sanschagrin F, Laroche J, Auger M, et al. Peptidoglycan lytic activity of the *Pseudomonas aeruginosa* phage  $\phi$ KZ gp144 lytic transglycosylase. *FEMS Microbiol Lett.* 2007; 266:201–209. [PubMed: 17233731]
- Plotnikova TG, Dzhusupova AB, Khrenova EA, Krylov VN. Genetic and phenogenetic study of the ts-mutant group of *Pseudomonas aeruginosa* PA01 phage  $\phi$ KZ. *Genetika.* 1292; 18:1793–1798. [Russian]. [PubMed: 6818108]
- Qiu D, Damron FH, Mima T, Schweizer HP, Yu HD. PBAD-based shuttle vectors for functional analysis of toxic and highly regulated genes in *Pseudomonas* and *Burkholderia* spp. and other bacteria. *Appl Environ Micro.* 2008; 74:7422–7426.
- Rasband, WS. ImageJ. U. S. National Institutes of Health; Bethesda, Maryland, USA: 1997–2011. <http://imagej.nih.gov/ij/>
- Rishovd S, Holzenburg A, Johansen BV, Lindqvist BH. Bacteriophage P2 and P4 morphogenesis: Structure and function of the connector. *Virology.* 1998; 245:11–17. [PubMed: 9614863]
- Rishovd S, Marvik OJ, Jacobsen E, Lindqvist BH. Bacteriophage P2 and P4 morphogenesis: Identification and characterization of the portal protein. *Virology.* 1994; 200:744–751. [PubMed: 8178458]
- Schaerli C, Kellenberger. Head maturation pathway of bacteriophages T4 and T2. V Maturable epsilon-particle accumulating an acridine-treated bacteriophage T4-infected cells. *J Virol.* 1980; 33:830–844. [PubMed: 6997509]
- Showe MK, Isobe E, Onorato L. Bacteriophage T4 prehead proteinase: I. Purification and properties of a bacteriophage enzyme which cleaves the capsid precursor proteins. *J Mol Biol.* 1976a; 107:35–54. [PubMed: 12371]
- Showe MK, Isobe E, Onorato L. Bacteriophage T4 prehead proteinase: II. Its cleavage from the product of gene 21 and regulation in phage-infected cells. *J Mol Biol.* 1976b; 107:55–69. [PubMed: 1003460]
- Söding J. Protein homology detection by HMM-HMM comparison. *Bioinformatics.* 2005; 21:951–960. [PubMed: 15531603]
- Thomas JA, Rolando MR, Carroll CA, Shen PS, Belnap DM, Weintraub ST, et al. Characterization of *Pseudomonas chlororaphis* myovirus 201  $\phi$ 2-1 via genomic sequencing, mass spectrometry, and electron microscopy. *Virology.* 2008; 376:330–338. [PubMed: 18474389]
- Thomas JA, Weintraub ST, Hakala K, Serwer P, Hardies SC. Proteome of the large *Pseudomonas* myovirus 201  $\phi$ 2-1: Delineation of proteolytically processed virion proteins. *Molecular & Cellular Proteomics.* 2010; 9:940–951. [PubMed: 20233846]
- Wainwright M, Phoenix DA, Marland J, Wareing DR, Bolton FJ. *In-vitro* photobactericidal activity of aminoacridines. *J Antimicrob Chemother.* 1997; 40:587–589. [PubMed: 9372431]
- Walker JE, Auffret AD, Carne A, Gurnett A, Hanisch P, Hill D, Saraste M. Solid-phase sequence analysis of polypeptides eluted from polyacrylamide gels. *Eur J Biochem.* 1982; 123:253–260. [PubMed: 6210528]
- Wu W, Thomas JA, Cheng N, Black LW, Steven AC. Bubblegrams reveal the inner body structure of  $\phi$ KZ. *Science.* 2012; 335:182. [PubMed: 22246767]
- Zachary A, Black LW. Isolation and characterization of a portal protein-DNA complex from dsDNA bacteriophage. *Intervirology.* 1991; 33:6–16.



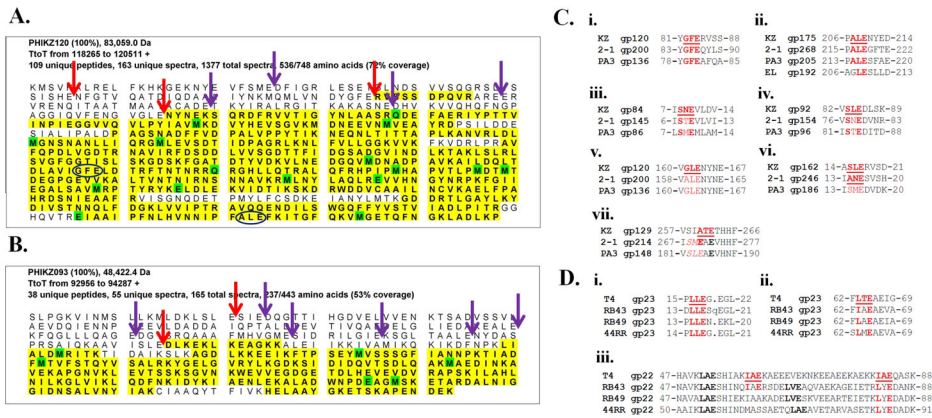
**Figure 1.** SDS-PAGE of proteins in purified  $\phi$ KZ tailless heads. The boundaries of slices that were digested and analyzed by MS are shown to the left of the processed gel lane. Each virion protein is listed next to the slice containing its maximum spectral count. gp refers to  $\phi$ KZ gene product. Virion proteins listed in red/underlined are proteolytically processed by the prohead protease, as confirmed by the detection of a semi-tryptic peptide (see Table 1). In cases where more than one polypeptide segment remained in the virion, the segments have been amended with N, M, or C to indicate *N*-terminal, middle- (if present) and *C*-terminal segments, respectively. Newly identified gene products in the tailless mutant are marked with an asterisk. Molecular weight standard was the Fermentas PageRuler Plus Prestained protein ladder.



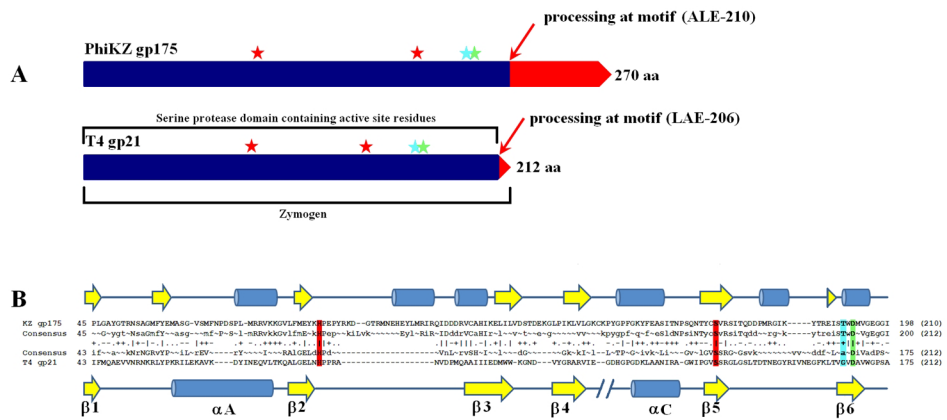


**Figure 2.**

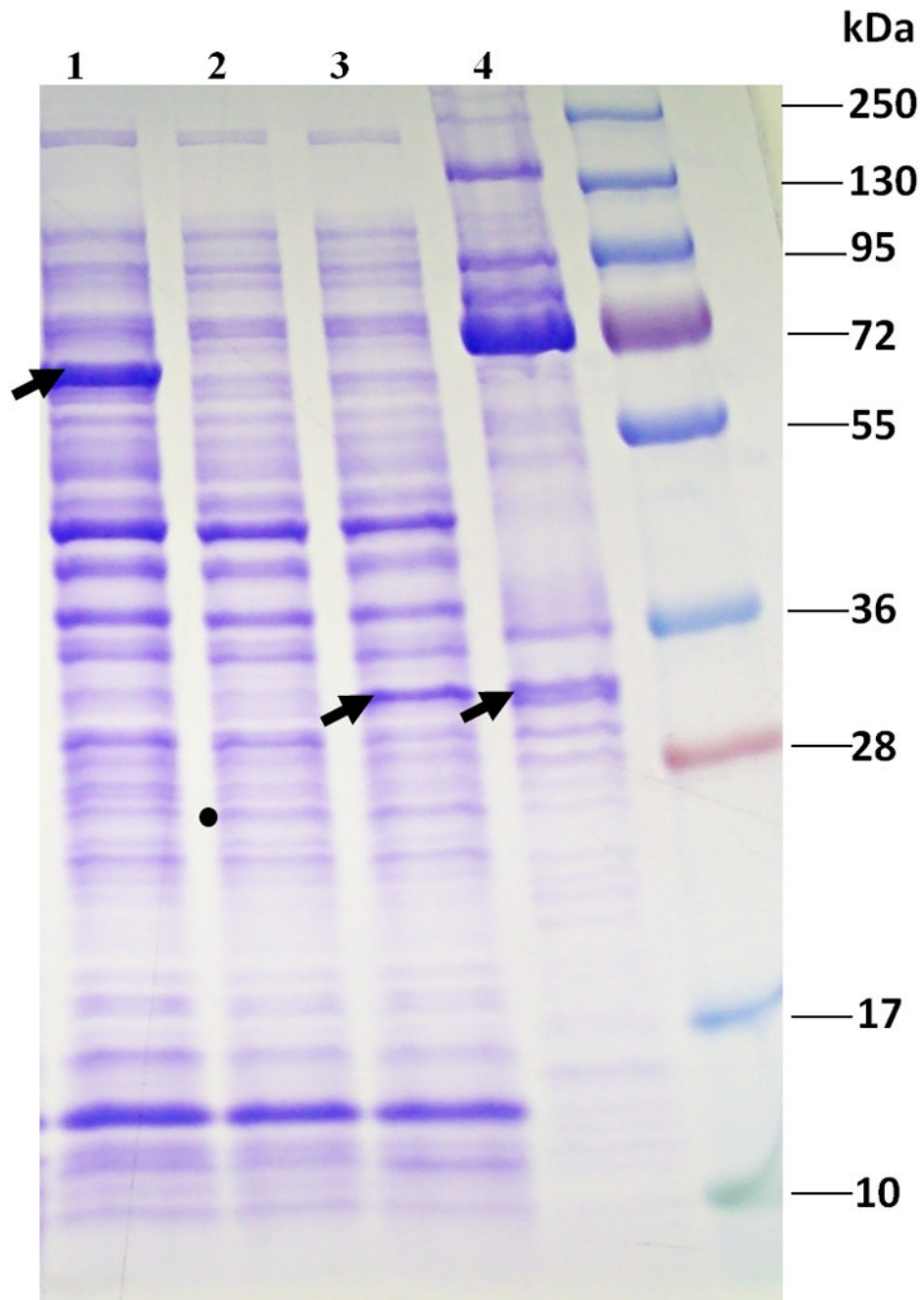
Regions of the  $\phi$ KZ genome encoding processed head proteins identified in full heads. Genomic segments from the full GBrowse display of  $\phi$ KZ with peptide coverage from MS of the tailless mutant mapped onto the corresponding genes. Legend: red arrows, proteins detected by MS and determined to be processed; light green arrows, proteins detected by MS, but not known to be processed; black outlined arrows, proteins determined to be abundant (see below); grey arrows, proteins not detected by MS. Boxed regions indicate peptide coverage in purified tailless mutant: lime green, tryptic peptide coverage; red, semi-tryptic peptide ending in glutamate after a processing motif (see Figure 4). Triangles above the peptide coverage boxes indicate prohead protease cleavage sites: dark blue, established by a semi-tryptic peptide in the purified head GeLCMS; purple, established by a semi-tryptic peptide detected in the osmotic shock or acetic acid treated samples that underwent MS analyses.



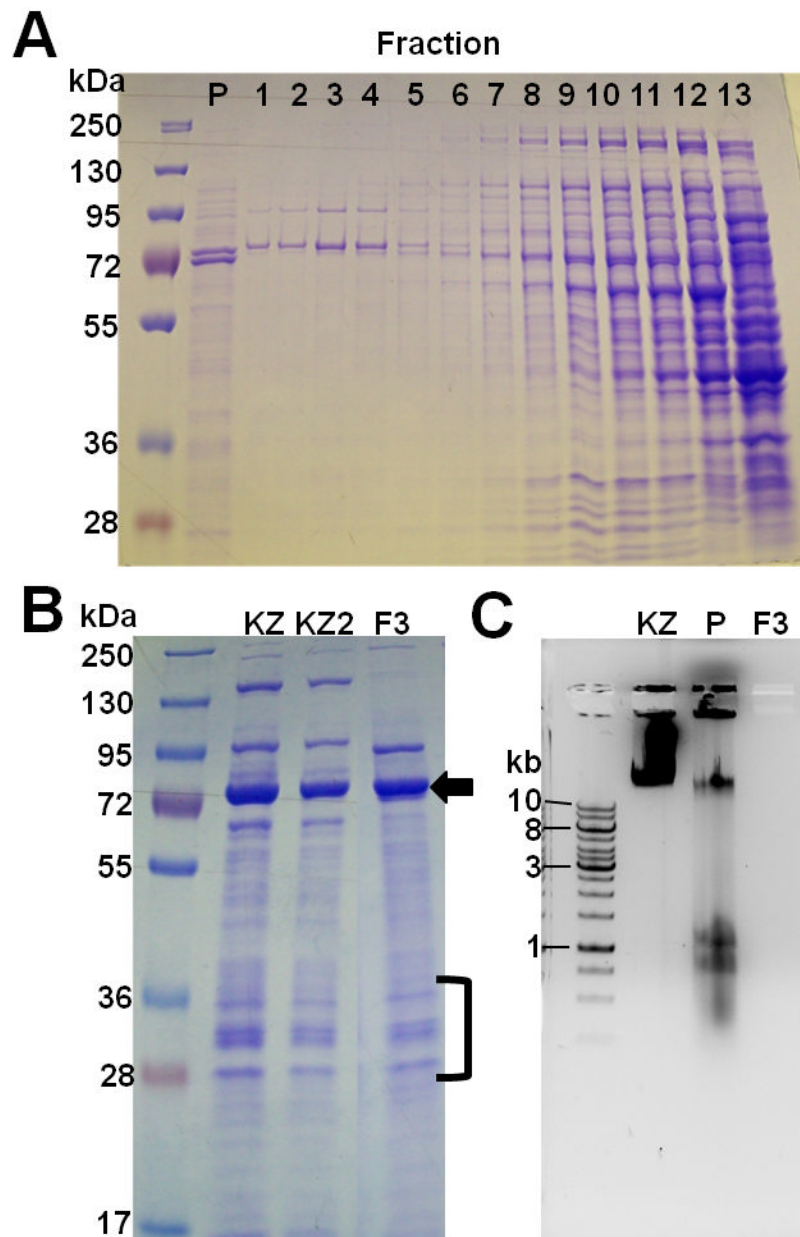
**Figure 3.** Cleavage sites in head proteins of  $\phi$ KZ and their conservation in related phages. Cleavage motifs in (A) the  $\phi$ KZ major head protein, gp120 and (B)  $\phi$ KZ abundant head protein, gp93. Red arrows indicate cleavage sites identified by MS; purple arrows indicate potential cleavage sites that adhere to the cleavage motif, see Figure 4. Blue circles indicate sequences that adhere to the processing motif but are not cleaved. (C) Conservation of cleavage motifs in proteins homologous to processed  $\phi$ KZ proteins. KZ, refers to  $\phi$ KZ, 2-1 refers to 201  $\phi$ 2-1, and PA3 refers to  $\phi$ PA3. Red underlined residues are identified cleavage sites in  $\phi$ KZ proteins and in (vi) 201  $\phi$ 2-1 (Thomas *et al.*, 2010). Conserved residues in other phage proteins are also red. Comparison of processing sites in homologs to the  $\phi$ KZ (i) major capsid protein, gp120 (ii) protease, gp175, (iii) gp84, (iv) gp92, (v) gp120 (maturation cleavage), (vi) gp162, and (vii) portal protein, gp129. (D) Conservation of processing motifs in proteins homologous to processed T4 proteins. 44RR refers to 44RRt.8. Red underlined residues are known processing sites in T4 proteins (Black *et al.*, 1993). Conserved residues, representing putative processing sites in homologous proteins are also red. Putative processing motifs in homologs to the T4 processed protein that are not strictly conserved are in bold. Comparison of processing sites in T4 (i) and (ii) major capsid protein, gp23, and (iii) major scaffold protein, gp22 are to their homologs.



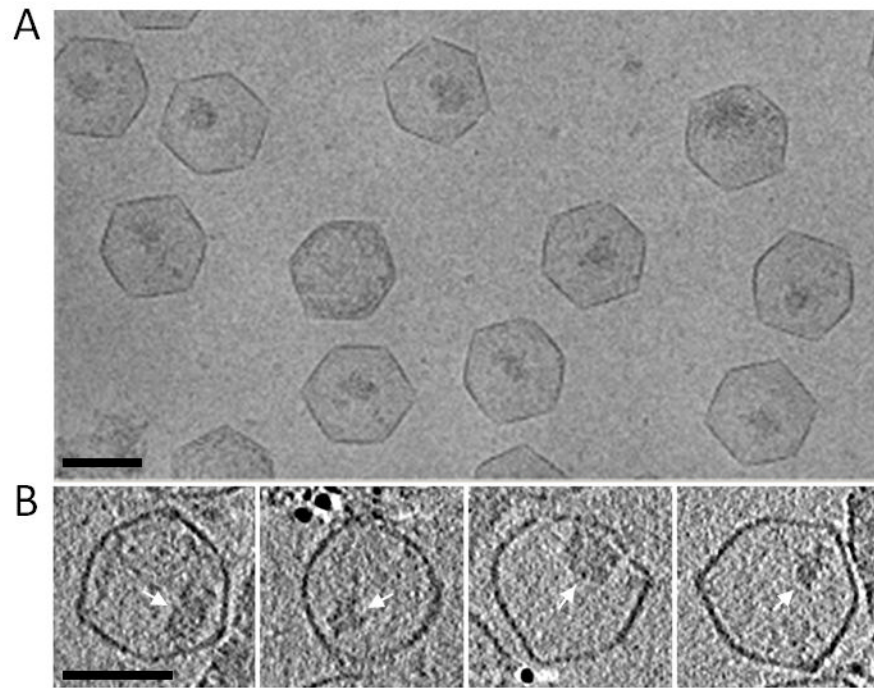
**Figure 4.** Features of the  $\phi$ KZ morphogenetic protease, gp175. (A) Comparison of the processing of  $\phi$ KZ gp175 compared to that of T4 gp21. Processing of gp21 converts the inactive form into an active serine protease. (B) HHpred alignment of  $\phi$ KZ gp175 HMM with that of T4 gp21. Above the alignment is the predicted secondary structure for gp175 obtained using Psi-Pred. Below the alignment are the structural elements of the herpesvirus fold positioned relative to the T4 sequence and labeled according to (Cheng *et al.*, 2004). Conserved catalytic residues (histidine and serine) are highlighted in red [red stars in (A)]. Candidate third catalytic triad residues are highlighted in light blue and green [light blue and green stars in (A)].



**Figure 5.** SDS-PAGE of an expression vector assay of the  $\phi$ KZ protease showing cleavage of  $\phi$ KZ head protein gp93. Lane 1, Arrow indicates full length gp93 expressed in DH10B. Lane 2, Dot shows position of mature gp175 in DH10B, much additionally found in the pellet fraction (data not shown). Lane 3, Arrow indicates cleaved gp93 after incubation with mature gp175. Lane 4, Arrow indicates the position of cleaved gp93 in purified wild-type  $\phi$ KZ virions. Lane 5, Molecular weight standard was the Fermentas PageRuler Plus Prestained protein ladder.

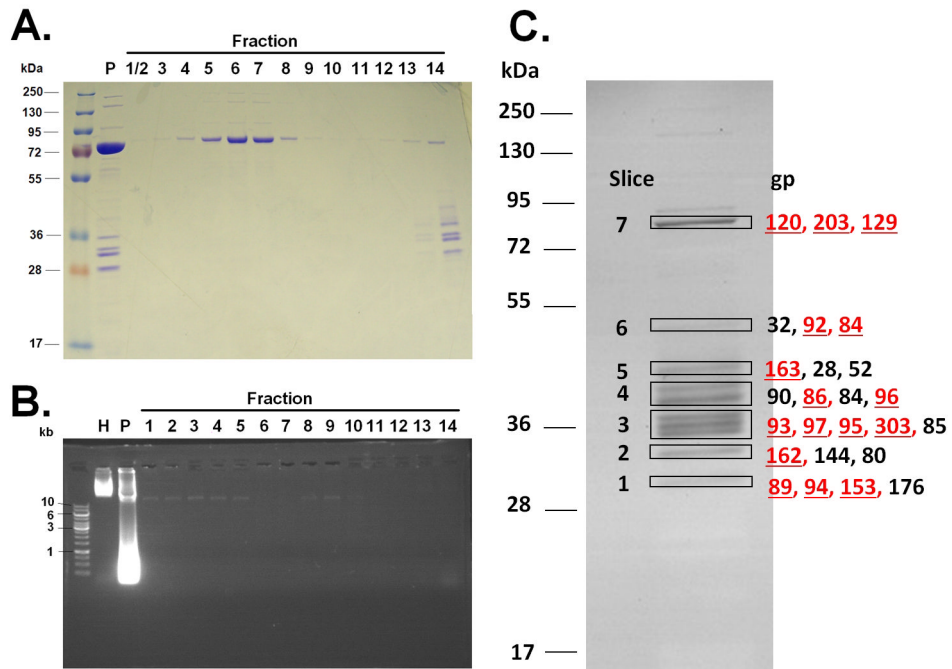


**Figure 6.** Purification of processed DNA-free IB containing proheads produced by terminase inhibitor 9AA. (A) SDS-PAGE of fractions from a glycerol gradient of 9AA-treated  $\phi$ KZ. Fraction numbers are given above each lane and are numbered from the bottom to the top of the gradient. P refers to the resuspended pellet fraction. (B) SDS-PAGE of purified whole  $\phi$ KZ (KZ; KZ2 refers to 2-fold dilution) and Fraction 3 (F3). Arrow indicates the band containing the mature form of the major capsid protein. Note the similar proportions of cleaved IB proteins in proheads and phages (bracketed). (C) Agarose gel stained with ethidium bromide showing DNA content of purified whole phage (KZ); pellet fraction (P); and Fraction 3 (F3).

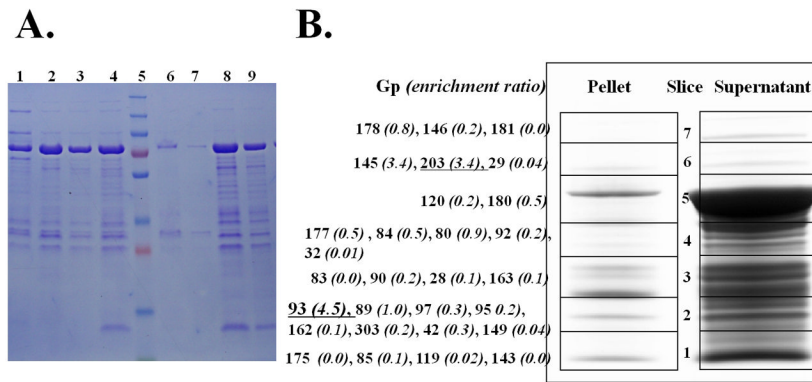


**Figure 7.**

(A) Cryo-electron micrograph of purified  $\phi$ KZ capsid-related particles produced in the presence of 9AA. The thin-walled polyhedral shape suggests that the capsids have the mature conformation. Most particles have an internal structure that appears to be related to the inner body but is shorter and more variable in appearance and may represent breakdown products. (B) Central 1.6 nm-thick slices through cryo-electron tomographic reconstruction of four 9AA-treated  $\phi$ KZ capsids. White arrows point to cylindrical internal densities that in some cases have about the same diameter as the inner body but are shorter and do not pass through the center of the capsid. Space bars represent 100 nm.



**Figure 8.** Osmotic shock treatment of  $\phi$ KZ tailless heads. (A) SDS-PAGE of glycerol gradient fractions of the tailless mutant after osmotic shock treatment. P, refers to resuspended pellet. Fractions are numbered from the bottom of the gradient (1) to the top of the gradient (14). Fractions 1 and 2 were combined. (B) DNA gel of the glycerol fractions, showing no or little DNA in fraction 6 that contained empty shells (see Supplemental Figure 3). H, refers to heads from purified  $\phi$ KZ tailless mutant. The standard used was the Fermentas GeneRuler 1 kb DNA ladder. (C) SDS-PAGE of fraction 14 showing the slices excised for MS identification of proteins released from within the capsid. Virion proteins in red/underlined are proteolytically processed as determined by the presence of a semi-tryptic peptide (see Table 1).



**Figure 9.**

SDS-PAGE of pellet and supernatant fractions after acetic acid treatment of  $\phi$ KZ tailless heads (A) 10% SDS-PAGE gel stained with Coomassie Blue stain. Lanes: (1) purified whole  $\phi$ KZ, (2) buoyant density purified tailless mutant, (3) step CsCl purified tailless mutant, (4) tailless mutant purified by differential centrifugation, (5) Fermentas Page Ruler, (6) pellet fraction from acetic acid treated sample from Lane 4, (7) 0.5 volume of sample in Lane 6, (8) supernatant fraction from acetic acid treated sample from Lane 4, and (9) 0.5 volume of sample loaded in Lane 8. (B) 12% SDS-PAGE gel stained with blue silver stain of pellet and supernatant fractions of acetic acid treated  $\phi$ KZ tailless heads. The boundaries of slices excised for GeLCMS are marked. Head proteins identified in the pellet fraction are provided and the ratio of their enrichment included in parentheses [(SC/Mw in pellet fraction)/(SC/Mw in supernatant fraction)]. Processed head proteins enriched in the pellet fraction are underlined.



Table 1

Prohead protease processed segments of head proteins from the  $\phi$ KZ tailless mutant determined by mass spectrometry. Proteins or segments predicted to be present in the mature virion in more than 100 copies are bolded.

gp/	Slice	Residue <sup>2</sup>		Mw <sup>3</sup>	SC <sup>4</sup>	SC/Mw <sup>5</sup>
		Start	Stop			
84	10	11*	413(C)	45.7	34	0.7
86	8	123*	428(C)	35.4	47	1.3
89N	1	1	122*	13.1	18	1.4
<b>89C</b>	<b>5</b>	<b>156*</b>	<b>387(C)</b>	<b>26.1</b>	<b>117</b>	<b>4.5</b>
92	10	86*	440(C)	41.1	11	0.3
<b>93</b>	<b>7</b>	<b>157*</b>	<b>435(C)</b>	<b>30.9</b>	<b>177</b>	<b>5.7</b>
94N	1	12*	95*	9.2	11	1.2
94M	1	96*	275*	20.1	15	0.7
94C	5	276*	505(C)	26.4	14	0.5
<b>95</b>	<b>7</b>	<b>261*</b>	<b>547(C)</b>	<b>31.6</b>	<b>177</b>	<b>5.6</b>
96	8	66*	371(C)	34.9	6	0.2
97N		14*	66*	5.8	24	4.1
<b>97C</b>	<b>7</b>	<b>533*</b>	<b>816(C)</b>	<b>30.7</b>	<b>152</b>	<b>5.0</b>
<b>120, capsid</b>	<b>12</b>	<b>164*</b>	<b>747(C)</b>	<b>64.9</b>	<b>1672</b>	<b>25.8</b>
129N	3	1	262*	29.3	5	0.2
129C, portal	12	263*	896(C)	70.1	89	1.3
153	5	53*	304(C)	28.2	42	1.5
162N	3/4	18*	137*	13.2	41	3.1
<b>162C</b>	<b>6</b>	<b>259*</b>	<b>522(C)</b>	<b>29.0</b>	<b>156</b>	<b>5.4</b>
163	8	55*	395(C)	38.1	45	1.2
175	4	1	210*	23.7	39	1.6
177N	1	1	60*	6.4	9	1.4
177C	11	61*	519(C)	52.4	126	2.4
180_180_2	12	1	505	57.8	36	0.6
184N	3	1	108*	12.1	6	0.5

gp <sup>1</sup>	Slice	Residue <sup>2</sup>		MW <sup>3</sup>	SC <sup>4</sup>	SC/MW <sup>5</sup>
		Start	Stop			
184C	1	109*	230	13.2	21	1.6
203	12	1	653*	75.3	71	0.9
303N	8	61	268	22.8	38	1.7
303M	1	269	366(C)	11.1	17	1.5
303C	7	367	646(C)	31.3	62	2.0

<sup>1</sup>The gene product name is appended with N (for N-terminal), M (for middle), or C (for C-terminal) when more than one segment of a gene product were detected in the virion.

<sup>2</sup>A coordinate is appended with an \* if confirmed by a semi-tryptic spectrum, or by (C) for an end defined by the stop codon. Semi-tryptic spectrum assignments were above 95% confidence except as noted in Supplementary Table 2.

<sup>3</sup>The molecular weight (MW) was calculated from the sequence.

<sup>4</sup>Peptide spectral counts (SC) were summed over all slices.

<sup>5</sup>SC/MW values were calculated as an approximate indicator of abundance of the polypeptides per virion.

**Table 2**

Abundant head proteins in  $\phi$ KZ, in addition to the major capsid protein (gp120, 1560 copies).

gp	Predicted Mw, kDa	Processed Mw, kDa (%reduction)	Copy number (densitometry)
89	43.2		
89C		26.1 (40)	60–100
90	35.9		190–230
93	47.6		
93C <sup>I</sup>		30.9 (35)	110–170
95	58.5		
95C <sup>I</sup>		31.6 (46)	110–170
97	87.8		
97C		30.7 (65)	110–170
162	57.6		
162C <sup>I</sup>		29.0 (50)	90–160

<sup>I</sup> proteins containing the paralogous domain (PF12699)