Published in final edited form as:
J Mol Biol. 2013 August 9; 425(15): 2765–2781. doi:10.1016/j.jmb.2013.05.002.

Functional domains of the HK97 capsid maturation protease and the mechanisms of protein encapsidation

Robert L. Duda[*], **Bonnie Oh**, and **Roger W. Hendrix**

Abstract

Tailed dsDNA bacteriophages and herpesviruses build capsids by co-assembling a major capsid protein with an internal scaffolding protein which then exits from the assembled structure either intact or after digestion *in situ* by a protease. In bacteriophage HK97, the 102 residue N-terminal *delta domain* of the major capsid protein is also removed by proteolysis after assembly and appears to perform the scaffolding function. We describe the HK97 protease that carries out these maturation cleavages. Insertion mutations at 7 sites in the protease gene produced mutant proteins that assemble into proheads, and those in the N-terminal two thirds were enzymatically inactive. Plasmid-expressed protease was rapidly cleaved *in vivo*, but was stabilized by co-expression with the delta domain. Purified protease was found to be active during the assembly of proheads *in vitro*. Heterologous fusions to the intact protease or to C-terminal fragments targeted fusion proteins into proheads. We confirm that the catalytic activity resides in the N-terminal 2/3 of the protease polypeptide and suggest that the C-terminal 1/5 of the protein contains a capsid targeting signal. The implications of this arrangement are compared to capsid targeting systems in other phages, herpesviruses, and encapsulins.

Introduction

Viral maturation proteases are widespread both in phages[1; 2] and viruses of higher organisms[3] where they are often absolutely required for the production of infectious particles and have served as the targets of anti-viral drug therapies[45]. Among these, the Herpesvirus proteases are the closest homologs to those of the tailed DNA bacteriophages[1]. The HK97 capsid maturation protease is also essential for production of viable phage[6789], and like the Herpesvirus proteases, plays its essential role in the processing of the scaffolding required for the assembly of a functional capsid[45], as described below.

Tailed bacteriophages and herpesviruses build icosahedral capsids by assembling the major capsid protein with a scaffolding protein which is not retained in the mature capsid[10]. These two proteins, together with a ring of portal protein (and additional components for more complex phages) form a prohead shell with the scaffolding protein on the inside[11]. A portal protein dodecamer sits at one of the 12 capsid vertices and forms the site where DNA goes in and out and where the tail is attached[1213]. Scaffolding proteins escape from the prohead interior, sometimes as full-size proteins (as in P22[14], T7[15]), or more often, after digestion *in situ* by a maturation protease. Bacteriophage HK97, like herpesviruses (HSV-1, CMV[16])

[*]Corresponding author, University of Pittsburgh, Department of Biological Sciences, Room 318 Langley Hall, 4249 Fifth Ave., Pittsburgh PA 15260, Phone: +1-412-624-4651, Fax: +1-412-624-4759.

and many phages (including T4[17], P2[18], and λ[19]), employs such a capsid maturation protease (gp4). However, HK97 does not have a separate scaffolding protein. Instead, the N-terminal 102 residues, or delta domain, of the HK97 major capsid protein (gp5) functions as the equivalent of a scaffolding protein that is removed by proteolysis[6]. Phage capsid genes are usually found in a cluster with the order (in the direction of transcription) of portal, protease, scaffolding protein, and major capsid protein[20]. In HK97 the order is portal, protease, (delta domain-) major capsid protein, showing that the delta domain occupies the genetically equivalent position of a scaffolding protein. The HK97 protease is apparently activated only after the capsid shell is formed with the protease inside, where it digests both the delta domain and itself, leaving a cleaved structure which has altered properties [7]. Capsid proteases are essential, since they play intrinsic roles in the function of scaffolding proteins or delta domains.

HK97 genes *3*, *4*, and *5* are required to build the T = 7 icosahedral HK97 procapsid[8] (Figure 1). These genes encode the portal or connector protein (46 kDa gp3, 12 copies), the protease, (25 kDa gp4, 50–60 copies) and the major capsid protein, (42 kDa gp5, 415 copies). In its natural host, *Escherichia coli*, the chaperonins GroEL and GroES are required for the correct folding of the major capsid protein subunit[21]. The HK97 assembly pathway can be recapitulated by expression of subsets of the capsid genes[6; 7]. When the major capsid protein alone is expressed from a plasmid, 460Å diameter Prohead I particles composed of uncleaved gp5 are made. When the protease is expressed along with the major capsid protein, the protease co-assembles and processes the capsid to produce Prohead II particles that have the cleaved form of the major capsid protein, 31 kDa gp5* (with the * indicating the cleaved form), and in which the gp5 delta domain and the protease are no longer present[6; 7; 22]. If the portal gene is added to our expression plasmid, then proheads with portals can be made. In the absence of the portal, normal-appearing particles without portals are made, and these are the simpler analogs of true proheads which we have studied extensively. DNA packaging causes Prohead II to expand to form 560Å diameter Head II[23; 24] which is stabilized by inter-subunit covalent bonds during the process[7; 21; 24; 25]. The later stages of HK97 capsid maturation have been studied extensively [6; 7; 22; 23; 24; 26; 27; 28; 29; 30; 31; 32; 33; 34; 35], while much less is known about the early steps [6; 36; 37; 38; 39] and very little about the capsid maturation protease. In the current work, we use genetic and biochemical approaches to reveal which parts of HK97 gp4 contain the proteolytic activity, how the activity may be regulated and which parts of the protein may be responsible for directing the assembly of the HK97 protease into proheads.

Results

Mutants in the HK97 protease gene 4

HK97 gene 4 was originally identified as the protease gene by a frameshift introduced by cutting and filling in a *Bst*BI restriction site[8] in our standard wild-type HK97 capsid expression plasmid[8], which normally makes HK97 Prohead II after induction. This mutation abolished proteolysis and can only make the first 26 residues of gp4 before shifting into a different reading frame. A comparison of the proteins and protein assemblies made by the wild-type plasmid and the protease knockout are shown in the first two lanes of Figure 2 (a) and (b) and form the basis of our assay for HK97 protease function. In the wild-type sample, nearly all of the 42 kDa major head protein is cleaved to its 31 kDa form, while the protease mutant only shows the uncleaved form. Agarose gels show which assembled forms of the proteins are present. The wild-type expression plasmid shows a sharp prohead band (Figure 2 (b) lane 1) that is identified as Prohead II because most of the major capsid protein was found in the cleaved form in the SDS gel lane above. A slower migrating sharp band of mature, expanded Head II is also found for wild-type, also indicating successful cleavage of the major capsid protein because only cleaved Prohead II can mature. In Figure 2 (b) lane 2,

the 4⁻ protease mutant also produces a sharp Prohead I band which migrates slightly slower than Prohead II, but is composed of uncleaved protein. A fuzzy, slower migrating band or spot is also seen in Figure 2 (b) lane 2 - this spot contains the hexamers and pentamers of the major capsid protein (capsomers) which were unassembled or had dissociated from Prohead I during sample preparation and thus indicates the presence of Prohead I. Note that Prohead II and Head II cannot dissociate into capsomers.

The first HK97 protease linker insertion mutant (LIM mutant) was made by inserting a 12 bp linker in the *Sma*I site in gene 4. The resulting LIM45 mutant has 4 amino acids inserted after residue 45. This mutant was found to block proteolysis, presumably by inactivating the protease, and about 50 copies of the inactive protein were found to co-purify with LIM 45 proheads (not shown, see [20]). Mutant LIM45 produced only uncleaved 42 kDa gp5 and uncleaved gp4 (Figure 2 (a) lane 3) and only Prohead I and capsomers were found (Figure 2 (b) lane 3). Subsequently, a set of linker insertion mutants in the protease gene were identified as part of a larger random mutant screen using enzymes that cut frequently in the expression plasmid (see Material and Methods). Mutants at seven sites in gene 4 region were identified by restriction mapping and further characterized by inducing plasmid expression under radiolabeling conditions and visualizing what was produced using gels. A 25 kDa protease band was detected in SDS gels for all of the LIM mutants except for mutant LIM-2 (compare Figure 2 (a) lane 6 with the other lanes). The linker in LIM-2 is inserted 5 bp upstream of the first gene 4 codon and abolished all synthesis of the protease, presumably by disrupting the gene's ribosome binding site (Figure 2 (c)). Insertions mutants at 5 sites in the N-terminal two-thirds of the gene produce proteins with no apparent proteolytic activity as demonstrated by presence of only uncleaved major capsid protein in SDS gels for mutants LIM45, LIM50, LIM78, LIM88 and LIM140 (lanes 3–7, and 10 in Figure 2). The agarose gels for these same five mutants show a Prohead I band and capsomers, but not a Head II band, confirming that no proteolytic processing occurred. Two mutants, LIM 130, and LIM179 retained measurable proteolytic activity (Figure 2 (a) lanes 8, 9). LIM179 showed the most cleavage, and LIM130, somewhat less, when compared to the wild-type example. LIM 130 and LIM179 both show a visible capsomer band in Figure 2 (b) (lanes 8 and 9) compared to wild type (lane 1), which has no capsomers. Complementation tests done using the plasmid-borne mutants and infecting amber mutant phage (Table 1) show that all of the LIM mutant proteases are defective for gene 4 complementation except for LIM179, which performed essentially like wild type in this assay. These results show that the HK97 protease catalytic activity requires contributions from residues spanning positions from ~40 to ~150 residues in the polypeptide, leaving the N-terminal ~40 residues and C-terminal 75 residues as regions that might be involved in targeting the protease for incorporation into proheads.

Homology to known protease structures and an active site mutant

At the onset of this work, no sequence similarity could be detected between the HK97 protease and any well characterized protease, and the only detectable similarity between a phage protease and known proteases were weak similarities between the λ capsid protease gpC and bacterial proteases[40]. The publication of crystal structures of Herpesviridae proteases[41,42,43] enabled detection of the link between the Herpesviridae protease and the phage capsid protease family that includes the HK97 protease.[12] Alignment of HK97 with a variety of putative phage proteases and the Herpesviridae proteases of known structure included the region of the HK97 gp4 covering residues 23 to 153 and predicted that the catalytic triad of the HK97 protease is composed of residues H65, S114, and E141[1] as noted in Figure 2 (c). We made a His 65 to Ala mutation in the protease gene and found that this change inactivated the protease, but it still allowed the protein to be incorporated into proheads (Figure 3). The mutant protein was found to make proheads that sediment faster than either Prohead II or Prohead I lacking the protease, and which contain 50–60 copies of

the inactive protease, as calculated from gel band intensities. This appears to confirm the assignment of one catalytic residue predicted by the alignment. Prohead I containing the inactivated protease was also used to grow the crystals used to determine a 5.2 Å x-ray structure of HK97 Prohead I[38], but the protease was not well ordered within these proheads and so no details of its structure were revealed.

Requirements for packaging the protease *in vivo*

None of the existing mutants seemed to affect the ability of the mutant proteins to be packaged into proheads, so a different approach was used to explore the requirements for capsid packaging in HK97. The collection of linker insertion mutants provided a convenient way to join other proteins to segments of the protease to see whether the fusion proteins produced would also be incorporated into HK97 proheads and reveal which parts of the protease are required for encapsidation. We fused the protease at most of the sites of the linker insertions to the C-terminus of a His-tagged variant of the bacteriophage λ decoration protein gpD, which is known to be efficiently made, highly soluble, and monomeric at very high concentrations[444546]. When we co-expressed these gpD-fusion variants with the HK97 major capsid protein, we found that a fusion protein was incorporated into proheads in every case (Figure 4). The control Prohead I contains only the prominent uncleaved 42 kDa Prohead I band, while proheads made with inactive protease (H65A mutant, both with or without a His-tag) produce proheads with an extra band corresponding to about 50–60 copies of the mutant gp4, as estimated from stained band intensities. Negatively stained electron micrographs of the same set of gpD-gp4 fusion-containing proheads reveal that these proheads do not appear remarkably different from empty wild-type Prohead I (Figure 5). The migration of these proheads in agarose gels is also identical and thus independent of their internal content (Figure 5(d).

The two mutants which incorporate the full length gpD and the full length protease show different types of evidence for incorporation. Where the inactive protease was fused with gpD (gpD-LIM-2-gp4 inactive), a band appears just below the 42 kDa major capsid protein band at a position consistent with the predicted molecular weight of 38 kDa for this fusion protein, showing that it is packaged. With the active protease (gpD-LIM-2-gp4 active), both Prohead I and Prohead II were made, as indicated by the accumulation of nearly equal amounts of cleaved and uncleaved major capsid protein (Figure 4). This shows that not only was the gpD-gp4 fusion incorporated, but it was also catalytically active.

Four fusion proteins that contain gpD and C-terminal fragments of the HK97 protease were incorporated into proheads, but showed no proteolytic activity as expected, because much of the putative catalytic domain has been removed. Fusion LIM50 has 174 residues, fusion LIM78 has 146 residues, fusion gpD-LIM140 has 85 residues, and fusion gpD-LIM179 has only 46 residues from gp4. This shows that 46 residues from the C-terminal tail of the HK97 protease are sufficient to direct incorporation of heterologous proteins into HK97 proheads (Figure 4), although the actual packaging signal may be smaller than the 46 residues defined here.

In vivo activities of His-tagged HK97 protease

We constructed plasmids that make six-histidine-tagged variants of the protease and tested whether the His-tagged protease retains its normal functions by expressing it along with the HK97 major capsid protein in *E. coli*, radio-labeling with ^{35}S-methionine and monitoring synthesis and cleavage of plasmid-encoded HK97 proteins using SDS polyacrylamide gels. The His-tagged protease was very active and able to carry out the maturation cleavage of the HK97 major capsid protein (see Figure 6 (a)), which indicates that it can successfully assemble into proheads. Note that the His-tagged protease is expressed at higher levels than

the wild-type protease due to the efficient translation initiation signal built into the His-tagged vector (Figure 6 (a)). We also tested the 6-His tagged gp4 producing plasmid for complementation activity (Table 1) and found that the plasmids expressing His-gp4 could complement phage mutants in gp4, but only when the cells were plated in the absence of inducer. In the presence of inducer, the unnaturally high levels of the protease produced by these plasmids actually inhibited phage growth, even for wild-type phage, showing that excess protease can somehow inhibit capsid formation. When the His-tagged protease was expressed on its own from plasmid pT7-6-6Hisgp4–38, the bulk of the protein synthesized was immediately reduced in molecular weight, suggesting that it may cleave itself under those conditions (Figure 6 (b) left). Fortunately, the His-tagged protease was protected from degradation when expressed in the presence of an N-terminal fragment of the HK97 major capsid protein containing the delta domain using the plasmid pTQ30-6Hisgp4-RV (Figure 6(b) right).

We investigated the nature of the rapid cleavage of the protease when it is expressed alone by repeating the expression experiment described in Figure 6(b) (left), except in LB medium and without radiolabelling. The cells were lysed, the pellet was dissolved in 6 M guanidine hydrochloride buffer (GuHCl), and protease fragments were purified using nickel affinity chromatography under denaturing conditions. The affinity purified fraction was reduced, alkylated with iodoacetamide and subjected to MALDI-TOF mass spectrometry. We found an ion peak with the mass of the full length His-tagged protease (27,086 Da), as well as four others that contained the following: (a) residues 1–183 (cleaved between K159 and S160 of the wild-type sequence, 19,985 Da), (b) residues 1–206 (cleaved between K182 and S183 of the wild-type sequence, 22,559 Da), (c) residues 3–206 (cleaved between K182 and S183 of the wild-type sequence and after R2 of the His-tag, 22,338 Da) and (d) a broad peak possibly corresponding to residues 1–215 or 3–217 (cleaved at or near R191 or K193 of the wild-type sequence and after R2 of the His-tag). The finding of these C-terminally cleaved fragments of the protease, in two (or possibly three) cases cleaved between lysine and serine residues (as found *in vivo* [7] and *in vitro* below) suggests that these cleavages are indeed catalyzed by the protease itself and that the protease has a trypsin-like specificity with a preference for cleavage between K and S residues.

In vitro activities of His-tagged protease

When we expressed proteins from the co-expression plasmid pTQ30-6Hisgp4-RV on a larger scale, we found that the His-tagged protease was well-expressed, but mostly partitioned into the cell debris during purification. Assuming the protease was in inclusion bodies, we washed the pellet, dissolved it in 6 M GuHCl buffer, and purified the protease using nickel-affinity chromatography under denaturing conditions. After removing the imidazole used for elution, the GuHCl concentration was lowered slowly to 0.5 M, which was found to keep the protease soluble, but inactive. If the GuHCl was completely removed, the protein was lost, presumably due to self digestion (data not shown).

We assembled the purified protease into HK97 proheads by diluting the protease into a mixture of HK97 capsomers in a Centricon centrifugal concentrator and concentrating the mixture 50-fold by centrifugation. The reactions were analyzed for prohead assembly and for polypeptide cleavage (Figure 7). In the absence of protease, the capsomers assembled into Prohead I containing uncleaved major capsid protein, but when His-tagged protease was added, we detected significant cleavage of the major capsid protein to the normal mature size. These results show that we could reconstitute the natural maturation cleavage activity of the HK97 protease and that only these two proteins are needed for this reaction to occur. The mass ratio of protease to major capsid protein in these reactions seems higher than would normally occur *in vivo*, but this was necessary to achieve the extent of cleavage we did obtain, which was plainly incomplete. We think at least thee factors contribute to the

incomplete cleavage: 1) an unknown fraction of the protease achieves folding and catalytic activity possibly due to purification under denaturing conditions followed by renaturation; 2) the reaction is not a simple mixing of enzyme with substrate, but requires complexes to form between the protease and capsid proteins and then for these complexes to assemble prior to cleavage; and 3) the substrate is experiencing assembly conditions and capable of assembling into capsid shells with no protease inside, which would render the substrate resistant to cleavage, which occurs only within the interior of the prohead. Under the best conditions, we found that about 50% of the major capsid protein was cleaved and that the extent of cleavage was greater when more protease was added. We confirmed that assembly occurred by showing that the capsomer band disappeared and a prohead band appeared in a native agarose gel (Figure 7 (a)). In addition to the mature 31 kDa cleavage product gp5*, we also observed additional bands of intermediate size between the full-size 42 kDa uncleaved major capsid protein and the mature form, as had been observed previously in radiolabeling experiments[8]. As a further control, we purified an inactive variant of the His-tagged protease with the LIM45 insertion and found that no proteolysis occurred when we added this variant to *in vitro* assembly reactions (data not shown). The products of an *in vitro* assembly reaction were separated on an SDS gel, electroblotted to a membrane, and individual bands were sequenced from their N-termini. Sequencing of the 31 kDa band revealed residues SLGSDAD, confirming that the band we see is indeed cleaved at the normal maturation cleavage site after Lys 103. Sequencing of the strongest intermediate band (marked 36 kDa in Figure 7) revealed residues SGTRLF, showing that cleavage also occurred after Lys 52. In both of these cases the protease cleaved between lysine and serine residues. Additional *in vitro* experiments showed that HK97 protease cleavage was blocked by serine protease inhibitor N-α-tosyl-L-lysyl-chloromethyl-ketone (TLCK, data not shown), an inhibitor of enzymes that cleave after lysine residues, further demonstrating that the HK97 protease has a specificity that favors cleavage after lysine.

Discussion

Interpretation of protease mutant phenotypes

There is no high resolution structure of the HK97 protease or any other phage capsid protease, so we can only interpret the mutant phenotypes in terms of our data and plausible models of how the protease structure might be arranged. Our mutant data seem to confirm the validity of primary sequence alignments of the HK97 protease with the Herpesvirus proteases[12]. The five insertion mutants which knock out all traces of HK97 protease activity, LIM45, LIM50, LIM78, LIM88 and LIM140, all lie within the aligned N-terminal region of the HK97 sequence (residues 23–153), which we believe is the major part of the proteolytically active domain. A sixth mutant in that region, LIM130, retained some proteolytic activity, but it and the other five mutants in the homologous region all fail to complement a gene 4 amber phage (Table 1). The gene 4 H65A mutant behaves exactly as expected for an active site mutant in its biochemical activities (Figure 3) and also fails to complement. The last of the insertion mutants, LIM179, retains proteolytic activity and is positive in gene 4 complementation, suggesting that it may be well outside of the active domain and in another region that may contain the targeting domain or may connect a targeting domain to the protease domain. Experiments with fusion proteins confirm that a targeting signal resides in the protease C-terminus (Figure 5).

How does the C-terminal domain of the HK97 protease function as a targeting signal for encapsidation? The HK97 protease is predicted (i.e., by PSIPRED[47]) to have a set of C-terminal α-helices in the last ~60 residues that may be configured to interact with the delta domain of the HK97 major capsid protein, and thereby direct the assembly of the protease into the HK97 prohead. The two partners of this interaction may both be alpha-helical. The delta domain of HK97 is known to be mostly α-helical from spectroscopic measurements[36])

with a strong prediction for coiled-coil interaction in the first ~60 residues[6], while the C-terminus of the HK97 protease is predicted to be alpha-helical with an additional prediction for coiled-coil interactions for the C-terminal ~25 residues (using the COILS Server[48]), leading us to speculate that they may interact via coiled-coil interactions. In this model, some of the α-helices of the HK97 delta domains, (which have been visualized by cryo-EM as interior-projecting domains of the hexons and pentons in Prohead I[6; 37; 49],) form coiled-coils with C-terminal α-helices of protease molecules before assembly, so that the tethered protease subunits can be enclosed in the assembling shell.

Capsid maturation protease function

A common feature of all tailed DNA phages and herpesviruses is that they have a prohead stage which is the mandatory precursor to the mature capsid. The herpesviruses and many phages utilize a maturation protease that acts in this transition from prohead to head, and the main function of the capsid protease is to degrade the parts of the prohead that are required for prohead assembly, but must be removed to complete maturation. The removed parts usually include the essential assembly core[50] or scaffolding proteins[10] and N-terminal delta domains of the major capsid protein (called delta domains because they are missing from mature capsids[51].) The protease may also remove pro-sequences from pro-enzymes or internal proteins that are packaged into the capsid for delivery into the host in more complex phages like T4[51; 52] and P1[53].

Other phages dispense with the need for a capsid maturation protease, but they still make use of a scaffolding protein for assembly. This was first described for P22[10], a tailed phage that has no protease and no need for one, perhaps because it has evolved a different solution to the removal of the scaffold: allowing the scaffolding protein to escape through portholes in the capsid that close during maturation after the scaffolding proteins have left[54]. The P22[14], ϕ29[55], T7[15], and SPP1[56] families of phages all can remove (and sometimes recycle and reuse) their scaffolding proteins without the aid of a protease. These phages may be the precursors to phages with proteases, or, they may have once utilized a capsid protease, but no longer need one after the scaffolding protein escape system had evolved.

Capsid targeting mechanisms and signals

For the rest of the phages and viruses that have a capsid maturation protease, there must be mechanisms for targeting the protease to the prohead interior and for regulating its activity. The work we present here shows that for HK97, a capsid targeting signal resides in the C-terminal region of the protease polypeptide. This is reminiscent of the packaging signal found for the *Thermotoga maritima* encapsulin and related bacterial particles[57]. In both cases the coding sequences for the parts of the proteins that bind to each other are right next to each other in their respective genetic maps (Figure 8 (a) and (b)). The encapsulins are capsid-like icosahedral particles found in bacteria that are made from proteins that have the HK97 fold and act as compartments that sequester stress response protein cargos (peroxidases or ferritin-like proteins) within their interiors. The type 1 encapsulin cargo proteins appear to bind via a specific C-terminal peptide sequence that was visualized *in situ* in the T=1 *Thermotoga encapsulin* x-ray structure described (PDB ID 3DKT). A different system for cargo incorporation was deduced for the larger (T=3) type 2 encapsulin from *Pyrococcus furiosus* (also called PfV or Pf-vlp, for *P. furiosus* virus-like particle)[58; 59]. In the *Pyrococcus* encapsulin, the cargo appears to be a ferritin-like protein domain that is co-translationally fused directly to the N-terminus of the major encapsulin subunit[57] (Figure 8 (c)). Encapsulin is the new and preferred name for the particles and genes often mis-labeled as "linocins" [57]. We refer to encapsulins using a C-terminal packaging signal as type 1 and those having a fused cargo as type 2.

Bacteriophage P2 has a slightly different scheme for incorporating its protease. The P2 protease is an N-terminal domain of gpO[18] which had previously been shown to be the scaffolding protein required for capsid assembly in P2[6061] and in its satellite phage, P4[62], as diagrammed in Figure 8 (d). So, for P2 and satellite phage P4, the protease and scaffolding protein are fused into a single protein that binds to the interior of the prohead by its C-terminal end. It has long been known that phage λ phage Mu, and HSV-1 use a curious variation on the scheme used by P2 and is illustrated Figure 8 (g) and (h). λ[19; 63; 64] Mu[65], and HSV-1[666768] each make a long protein that has a protease domain at the N-terminus and a scaffolding domain at the C-terminus, but they also make smaller scaffolding protein subunits (gpNu3 in λ, gpZ in Mu, the product of UL26.5 in Herpesvirus) in greater number by internal initiation within the coding sequence of the long protease open reading frame. In this scheme, the protease is incorporated via co-assembly with scaffolding protein, with each binding via their C-terminal regions to parts of the major capsid protein that lie on the interior of the capsid[16; 69].

The phage T4 protease has been studied extensively[17; 70; 71] and two polypeptides (full length and one internally initiated) appear to be required for protease activity when plasmids are used to complement a protease defective phage[72]. However, the T4 capsid is much more structurally complicated than the HK97 capsid[70; 73] and the mechanism of T4 protease incorporation into proheads remains a mystery. Better understood are a group of non-essential T4 proteins, which include gp*IPI*, gp*IPII*, gp*IPIII* and gp*alt*, that are incorporated into the T4 prohead, where they are proteolytically processed during maturation[51; 52; 74] and later injected into the host to perform a variety of functions [75; 76]. T4 internal protein gp*IPIII* has an N-terminal Capsid Targeting Sequence (CTS) that is required for packaging (Figure 8 (i)) [77]. The T4 CTS [78; 79; 80; 81; 82; 83] lies within the first ~10 residues of the T4 internal proteins, overlaps the cleavage site, and has a consensus sequence that suggests that this region may form an amphipathic α-helix.

Evolution and exploitation of capsid packaging strategies

From this brief survey of protein packaging strategies, it appears that there are only a few characterized examples where the targeting sequence employed for incorporating a protease or other cargo into a capsid is not also a scaffolding protein. These are HK97, *T. maritima* encapsulin, and the T4 internal proteins. This group also encompasses related capsids with the same gene arrangements as those mentioned: HK97 relatives D3, XP10, phi1026b, T5 and relatives, related encapsulins, and T4 relatives. Otherwise, it appears that the common strategy is to fuse the protease to either a scaffolding protein or the major capsid protein itself (Figure 8 (e) & (f), not discussed here, see original reports[8485; 86]). All of the capsid and capsid-like particles described here use a variant of the same basic subunit fold and therefore are likely to have evolved from a common ancestor, but that common ancestor is long absent, leaving us to speculate which variation may have emerged first. Some would argue that the encapsulins are devolved forms of phage capsids. If simpler forms are considered to have evolved first, then the strategy used by the *P. furiosus* T=3 encapsulin (Figure 8 (b), with a cargo fused to a protective shell[57], might be the simplest. A T=1 variant using this scheme could have evolved from a pentameric enzyme assembly (such as a protease) that developed the ability to form a closed icosahedral shell with the active site inside. It has not escaped our notice that the capsid packaging signal described here for HK97 could be used to make HK97-derived particles analogous to the encapsulins with a wide variety of contents for use as biochemical or biom edical nanotechnology, as well as to learn more about the mechanism of assembly. The creation of nanoparticles containing specific protein cargoes derived from phage or viral capsids and other natural systems have already been reported in several systems, including phage P22[87; 88], CCMV[89], Q-beta[90], and other natural systems[9192].

Specificity and regulation of protease activity

The mature cleavage site in HK97 site is between K103 and S104 and the sequence KS occurs 4 times in the delta domain of the major capsid protein and 4 times in the last 70 residues of the protease itself, and nowhere else in either protein. This led us to speculate that the specificity of the protease was for the sequence KS^6. We now believe that the HK97 protease may cleave after almost any lysine-containing peptide that can bind to it, with a preference for the sequence KS, based on analyses of the amino acid sequences of HK97 protease substrates (Figure 9). The gp5 delta domain has 13 lysines spaced about every 8 residues, while the rest of protein has only an additional 10 (and these 10 lysines would normally not be accessible to the protease). The protease has 16 lysines spread throughout. With a strict specificity for the sequence KS, it is unlikely that the delta domain and the protease would be completely digested to peptides small enough to exit from the prohead[6], which we believe is required for maturation. Other well-studied phage proteases are also serine proteases, but they have distinctly different and more restrictive recognition sequences: after the motif I/ L-X-E for T4[70] and the related motif S/ A/ G-X-E for phiKZ. A more narrow specificity may be required to correctly remove or retain parts of 10 distinct proteins for T4[70] and 19 for phiKZ[93].

The active form of the T4 protease is derived from a larger precursor[1771], posssibly by the removal of a C-terminal peptide[9394]. A similar situation appears to occur for the phiKZ protease[93], and both cases may represent autoactivation by C-terminal cleavage. Evidence for C-terminal self-cleavage of the HK97 protease (see Figure 6 and Results) suggests that HK97, may also use C-terminal cleavage autoactivation to control the HK97 protease activity. However, we argue that there are additional considerations for HK97, because the HK97 protease binds to the capsid protein via the C-terminus. The C-terminal cleavage may be required to untether the protease from its binding site and allow it to have unfettered access to its substrates throughout the interior of the HK97 prohead.

Methods

Prohead purification

We used the methods outlined previously[749] to produce HK97 Prohead I as a control and source of capsomers from T7 expression plasmid pVB in BL21(DE3) *E. coli* cells also containing plasmid pLysS[95]. In short, cells were grown at 28°C to ~ 4×10^8 cells per ml, induced with 90 mg/ ml IPTG (isopropyl-β-D-thiogalactopyranoside) for ~ 16 hrs, chilled, and harvested by centrifugation. All steps except chromatography were done in the cold. Cell pellets were resuspended in 50 mM Tris HCl (pH 8.0 @ 25°C, ~8.5 on ice) with 5 mM EDTA and lysed by addition of Triton X100 to 0.2 %, followed by 2-fold dilution in the same buffer and 3 cycles of heating to ~23°C and cooling on ice. Once cells were lysed, $MgSO_4$ was added to 7.5 mM, DNase I (Sigma D5025), added to 10 μg/ ml and the temperature was raised briefly to ~20°C until the viscosity was reduced significantly. Lysates were then clarified by centrifugation (Beckman JLA-16.250 rotor / 10 kRPM / 15 min) to remove cell debris. The clarified lysates were "PEG precipitated" by the addition of sodium chloride to 0.5 M and polyethylene glycol (8000 molecular weight) to 6% followed by a 1 hour incubation. The PEG precipitate was collected by centrifugation (Beckman JLA-16.250 rotor / 8 kRPM / 15 min) suspended in buffer G (20 mM Tris-HCl pH 7.5, 100 mM NaCl), and clarified by centrifugation (Beckman JLA-16.250 rotor / 12,000 RPM / 15 min). Proheads were concentrated by ultracentrifugation (Beckman Type 45 Ti for 2 hrs @ 35,000 RPM) and the pellets were left under buffer overnight to dissolve. The next day, the proheads were diluted to about 60 ml in Buffer G, clarified by centrifugation to remove insoluble material and run in ~5.7 ml aliquots on individual 10–30% glycerol gradients (in Buffer TKG50) in tubes appropriate for the Beckman SW28 rotor. Gradients were run at

27,0000 rpm for 2 hrs and the broad peak fractions were collected, dialyzed briefly to remove glycerol and chromatographed using a BioCad Sprint 60 chromatography system (Applied Biosystems) on a Perceptive Biosystems HQ20 column eluting with a 20 column volume gradient from 20 mM to 1 M NaCl in 20 mM of pH 7.5 Bis-Tris-Propane-Tris-HCl buffer. For large-scale preparations of proheads, we started with a 1 L culture, while for smaller preparations, such as those used for purifying the protease-fusion-containing proheads, we started with 50 ml cultures and scaled all steps accordingly. Fusion protein-containing prohead purifications were stopped after the first ultracentrifugation step because the purity was sufficient for analysis of incorporated proteins.

Mutagenesis

Linker insertion variants in gene 4 were made by inserting a 12-base pair dsDNA oligonucleotide (ggactcgagtcc) containing an XhoI restriction site into the plasmid pV0 (pT7-Hd2.9) after the plasmid was linearized using restriction enzymes *Hae*III, *Bst*UI, or *Eco*RV. Unphosphorylated linkers were added to linearized vector DNA using a linker-tailing protocol[96] that inserts only one linker. The 12 bp linker, ggactcgagtcc, has an XhoI site (underlined) that is unique in the plasmid to facilitate mapping. Individual clones were isolated after transformation of treated, ligated DNA into a host, and DNA from clones was tested for the presence of the linker. The linker positions were mapped by restriction endonuclease digestion and assigned to sites or site clusters designated alphabetically clockwise around the map of the plasmid. Mutant names used here are isolate numbers with a prefix indicating which enzyme was used for their creation: B for *Bst*UI, H for *Hae*III and M for *Eco*RV. The H65A mutant was made using the Stratagene QuickChange protocol that synthesizes the whole plasmid from two overlapping primers. The primers used were GCGATGTTTTTCAACGCCAAGACGTGGGAGCTG, and CAGCTCCCACGTCTTGGCGTTGAAAAACATCGC. Candidate clones were sequenced and a sequenced-verified segment from the mutant DNA was subcloned as a BanII fragment into a clean, non-mutagenized parent plasmid.

His-tagged protease vectors

Plasmids were derived from our standard HK97 expression plasmid (pT7-Hd2.9, also called pV0) or the variant pV0-LIM-2, described in Results, that has an XhoI linker inserted at a *Hae*III site 5 bp upstream of the ATG of gene 4. pTQ30-6-His-g4-g5 (expressing His-tagged gp4 and WT gp5) was made by inserting a 1994 bp XhoI-HindIII DNA insert from pV0-LIM-2 into the 2440 bp SalI-HindIII digested 6-His-tag vector pTQ30; the resulting vector has a T7 promoter, an efficient translation signal and a 25 residue 6-His-tag leader that is fused in-frame with gp4 at the SalI/ XhoI junction. pTQ30 was derived from Qiagen T3 promoter vector pQE30 by ligating the EcoRI-XbaI fragment containing the ribosome binding site, 6-Histidine leader and a drug resistance cassette into the T7 promoter vector pT7-5 (see GenBank AY230150) cut with EcoRI and XbaI.

A plasmid with the C-terminal part of gp5 deleted was made by digesting pTQ30-6-His-g4–g5 with EcoR*V* and Hind*III*, filling in the overhang with dNTPs and Klenow fragment of DNA polymerase, and recircularizing with T4 DNA ligase. This removes the gene 5 DNA beyond the EcoRV site to make p6xHis-gp4-EcoRV, which expresses tagged gp4 and a 20 kD gp5 N-terminal fragment with 177 of 182 residues from gp5. A plasmid expressing 6-His-gene 4 alone was made from plasmid p25kD#38[9]. p25kD#38 was made from an HK97 clone (#38[8]) containing a BamHI linker (CCGGATCCGG) inserted into the AluI site after codon 20 of gene 5: a 1060 bp BamHI-PstI fragment containing gene 4 was ligated into BamHI-PstI cut pT7-6 to make p25kD#38. A 626 bp AflIII-BstBI fragment containing the 6-His-tagged N-terminal region of gene 4 from pTQ30-6-His-g4–g5 was ligated into AflIII-

BstI cut p25kD#38 to yield p6His-gp4-ins38. This plasmid makes 6-His gp4 and potentially a 3.7 kDa peptide from the first 20 codons from gene 5.

Construction of gpD-gp4 fusions

was begun by making a set of vectors with a His-tagged gpD that had an XhoI site in all three reading frames. The phage λ gene D was amplified using primers TQ30gpDL1 with a 5' BglII site 5'ATCAGATCTGGTTCCATGGCGAGCAAAGAAACC, and TQ30gpDR1 with a KpnI site TCAGGTACCACCGGATCCAACGATGCTGATTGC. The product was ligated to the large fragment of BamHI/ KpnI cut pTQ30 (described above), producing a vector with 6-His-tagged gpD with adjacent BamHI and KpnI sites at the 3' end of the coding sequence. The XhoI site was added in 3 reading frames by opening the vector with BamHI and KpnI and ligating annealed oligonucleotide adaptors that bridge the gap and install an XhoI site at the end of the gene D open reading frame. The ad apters were annealed GATCTGGTTCTCGAGGTAC and CTCGAGAACCA for frame 3, GATCTGGTGGTCTCGAGGTAC and CTCGAGACCACCA for frame 2 and GATCTGGTGGCTCGAGGTAC and CTCGAGCCACCA for frame 1. After confirmation of each plasmid by sequencing, the fusion protein expression plasmids were constructed by cutting the gp4 LIM mutant plasmids and the gpD fusion donor plasmids each with XhoI and HindIII and ligating the pieces containing the C-terminus of gp4 through to the end of gene 5 to the large fragment of the gpD-containing plasmid. Frame 3 was used for LIM50, LIM78 and LIM-2; frame 1 was used for LIM130, LIM179 and LIM130.

Expression and purification of His-tagged protease

The HK97 protease (6-His-gp4 or an inactive variant with the LIM45 linker insertion described in Results) and a fragment of major capsid protein gp5 were co-expressed from the T7 promoter plasmid using *E. coli* strain BL21 (DE3) pLysS grown at 37 °C in LB. Cells were grown to ~ 4×10^8 per ml in LB, induced with 90 mg/ ml IPTG, harvested 3.5 hr after induction and frozen at −70° C. All steps except chromatography were done in the cold and intermediate fractions were stored at −20°C, if overnight and otherwise at −70°C. Cells were thawed and resuspended in 20 mM Tris HCl pH 7.9 (buffer A) and lysed by sonication. Insoluble material (i.e., inclusion bodies) was removed by centrifugation, suspended in the same buffer, and stored frozen. The bulk of the 6-His-gp4 and the gp5 fragment were insoluble, so guanidine hydrochloride (GuHCl) was used to dissolve the proteins for further purification. The suspension of insolubles was thawed, solid GuHCl was added to 6 M, imidazole was added to 5–10 mM, and the pH adjusted to ~8 by the addition of 1 M Tris base. The GuHCl extract was passed through a immobilized Ni^+ column (Chelating Sepharose Fast Flow) equilibrated with 5 mM imidazole in buffer A plus 6 M GuHCl. The bound material was washed with 20 mM imidazole and the protease was eluted with 0.8 M imidazole, all in buffer A plus 6 M GuHCl. The imidazole was removed by dialysis into 20 mM Tris HCl pH 7.5 with 1 mM DTT, 0.5 mM EDTA and 1% glycerol containing 5 M GuHCl. GuHCl was then reduced from 5 M to 0.5 M in ~15 hr by pumping 900 ml of 20 mM Tris HCl pH 7.5, 1% glycerol, 1 mM DTT into 100 ml of stirred buffer containing 2–4 ml gp4 in dialysis bags. Preparations were recovered and stored frozen.

Ultrafiltration-driven *in vitro* assembly of Prohead I

Re-assembly of capsomers derived from Prohead I was done by concentrating capsomers from ~2 ml to ~ 0.04 ml (about 50-fold) in a Centricon 50 centrifugal concentrator. Approximately 180–200 μg of capsomers were added to a total of 2 ml (mostly TKG50 buffer) in a Centricon 50 centrifugal concentrator and centrifuged a Beckman JA–20 rotor at 6,000 rpm until the volume was reduced to ~0.02 −0.04 ml (~1–2 hr). TKG50 buffer is 20 mM Tris-HCl pH 7.5, 50 mM potassium glutamate. The concentrator was sealed at both ends with Parafilm to prevent evaporation and left at room temperature overnight. The

concentrate was recovered by inverting the device and centrifuging again for 2 min at 2,000 rpm. For most experiments the volumes of the recovered samples were measured, equalized to ~0.04 ml and stored in the cold. The recovered material was assayed for assembly by agarose gel electrophoresis.

In vitro proteolysis

Tests for *in vitro* activity of the purified HK97 protease were done by diluting the protease (in 0.5 M GuHCl storage buffer) into in-Centricon assembly reactions (see above). The order of addition of components had a noticeable effect on the results (see under Figure 7). Normally the protease was added to ~1 mg/ ml capsomers and then diluted to 2 ml with buffer. After concentration and incubation, samples were TCA precipitated and run in SDS polyacrylamide gels to look for evidence of proteolysis, especially the conversion of HK97 gp5 (42 kDa) to gp5* (31 kDa). For tests for inhibition of proteolytic activity by N-α-tosyl-L-lysyl-chloromethyl-ketone HCl (TLCK, Sigma), a 50 mg/ ml (135 mM) stock of TLCK in methanol was added to assembly reactions to a final concentration of ~7 mM TLCK and 5% methanol in the reaction. A control reaction had 5% methanol without TLCK.

Gel Analysis

Native (non-denaturing) agarose gels were used to test for capsid assembly. Capsomers (hexamer and pentamer mixture) run as a diffuse spot, while Prohead I and Prohead II run as faster bands (both at about the same position, with Prohead I slightly ahead). Head II, which has transformed and expanded into the angular form observed on phage particles, also runs as a band that has a lower mobility than either Prohead I or II. SDS Polyacrylamide gels were run using the buffers described by Laemmli[74], but using an acrylamide stock containing 33.5% (w/ v) acrylamide and 0.3%(w/ v) bisacrylamide[97]. SDS gel samples were precipitated with ice cold 10% (w/ v) TCA, washed with acetone and dried under vacuum before being resuspended in SDS sample buffer (0.0625 M Tris-HCl pH 6.8, containing 2% (w/ v) SDS, 5% mercaptoethanol and 10% glycerol) and heated in a boiling water bath for 2.5 minutes. Estimates of stoichiometries of proteins present in gel lanes were done using photographs of Coomassie stained gels taken from images made using an Aristo V810 grid light box (Aristo Grid Lamp Products, Waterbury, CT) and the EDAS 290 camera system and 1D Image Analysis Software from Kodak (Rochester, NY).

Mass analysis

Purified protein previously eluted from a Ni^{+2} column in 6 M GuHCl plus imidazole was dialyzed into 6 M GuHCl/ 2 mM disodium EDTA in 0.2 M Tris HCL pH 8.3, reduced by adding DTT (dithiothreitol) to 15 mM and incubating at 42°C for 2.5 hr, and alkylated by adding freshly dissolved iodoacetamide to 50 mM and incubating at room temperature in the dark for 30 min. The reaction was quenched for 20 min with 40 mM additional DTT and precipitated with TCA (~10% w/ v). The pellets were washed 4 times with 90% (v/ v) acetone, dried under vacuum and stored frozen. Dried samples were dissolved in 0.1% trifluoroacetic acid in 50% (v/ v) acetonitrile:water, then diluted to 30% acetonitrile, mixed with sinapic acid matrix and d ried onto a MALDI plate. MALDI mass analysis was carried out using a Applied Biosystems Voyager DE-STR MALDI-TOF at the Carnegie-Mellon University Center for Mass Analysis (Pittsburgh, Pennsylvania) following calibration using apo-myoglobin.

Acknowledgments

We thank Kathleen Martincic for isolating the LIM mutants and doing much of their initial characterization. We also thank Lindsey Black, Julie Thomas, Alexis Huet and Craig Peebles for comments on the manuscript and our

many collaborators for their support and insightful discussions. This work was funded by NIH Grant R01 GM047795 to RWH and RLD.

References

1. Cheng H, Shen N, Pei J, Grishin NV. Double-stranded DNA bacteriophage prohead protease is homologous to herpesvirus protease. Protein Sci. 2004; 13:2260–2269. [PubMed: 15273316]
2. 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]
3. Tong L. Viral proteases. Chem Rev. 2002; 102:4609–4626. [PubMed: 12475203]
4. Hsu JT, Wang HC, Chen GW, Shih SR. Antiviral drug discovery targeting to viral proteases. Curr Pharm Des. 2006; 12:1301–1314. [PubMed: 16611117]
5. Steuber H, Hilgenfeld R. Recent advances in targeting viral proteases for the discovery of novel antivirals. Curr Top Med Chem. 2010; 10:323–345. [PubMed: 20166951]
6. Conway JF, Duda RL, Cheng N, Hendrix RW, Steven AC. Proteolytic and conformational control of virus capsid maturation: the bacteriophage HK97 system. J Mol Biol. 1995; 253:86–99. [PubMed: 7473720]
7. Duda RL, Hempel J, Michel H, Shabanowitz J, Hunt D, Hendrix RW. Structural transitions during bacteriophage HK97 head assembly. J Mol Biol. 1995; 247:618–635. [PubMed: 7723019]
8. Duda RL, Martincic K, Hendrix RW. Genetic basis of bacteriophage HK97 prohead assembly. J Mol Biol. 1995; 247:636–647. [PubMed: 7723020]
9. Duda RL, Martincic K, Xie Z, Hendrix RW. Bacteriophage HK97 head assembly. FEMS Microbiol Rev. 1995; 17:41–46. [PubMed: 7669350]
10. King J, Casjens S. Catalytic head assembling protein in virus morphogenesis. Nature. 1974; 251:112–119. [PubMed: 4421992]
11. Casjens, S.; Hendrix, RW. Control mechanisms in dsDNA bacteriophage assembly. In: Calendar, R., editor. The Bacteriophages. Vol. Vol. 1. New York: Plenum; 1988. p. 15-91.
12. Driedonks RA, Caldentey J. Gene 20 product of bacteriophage T4. II. Its structural organization in prehead and bacteriophage. J Mol Biol. 1983; 166:341–360. [PubMed: 6406677]
13. Driedonks RA, Engel A, tenHeggeler B, van D. Gene 20 product of bacteriophage T4 its purification and structure. J Mol Biol. 1981; 152:641–662. [PubMed: 7334518]
14. Casjens S, King J. P22 morphogenesis. I: Catalytic scaffolding protein in capsid assembly. J Supramol Struct. 1974; 2:202–224. [PubMed: 4612247]
15. Roeder GS, Sadowski PD. Bacteriophage T7 morphogenesis: phage-related particles in cells infected with wild-type and mutant T7 phage. Virology. 1977; 76:263–285. [PubMed: 319595]
16. Homa FL, Brown JC. Capsid assembly and DNA packaging in herpes simplex virus. Rev Med Virol. 1997; 7:107–122. [PubMed: 10398476]
17. 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. 1976; 107:35–54. [PubMed: 12371]
18. Chang JR, Spilman MS, Rodenburg CM, Dokland T. Functional domains of the bacteriophage P2 scaffolding protein: identification of residues involved in assembly and protease activity. Virology. 2009; 384:144–150. [PubMed: 19064277]
19. Medina E, Wieczorek D, Medina EM, Yang Q, Feiss M, Catalano CE. Assembly and maturation of the bacteriophage lambda procapsid: gpC is the viral protease. J Mol Biol. 2010; 401:813–830. [PubMed: 20620152]
20. Hendrix RW, Duda RL. Bacteriophage HK97 head assembly: a protein ballet. Adv Virus Res. 1998; 50:235–288. [PubMed: 9521001]
21. Popa MP, McKelvey TA, Hempel J, Hendrix RW. Bacteriophage HK97 structure: wholesale covalent cross-linking between the major head shell subunits. J Virol. 1991; 65:3227–3237. [PubMed: 1709700]
22. Gertsman I, Gan L, Guttman M, Lee K, Speir JA, Duda RL, Hendrix RW, Komives EA, Johnson JE. An unexpected twist in viral capsid maturation. Nature. 2009; 458:646–650. [PubMed: 19204733]

23. Helgstrand C, Wikoff WR, Duda RL, Hendrix RW, Johnson JE, Liljas L. The refined structure of a protein catenane: the HK97 bacteriophage capsid at 3.44 A resolution. J Mol Biol. 2003; 334:885–899. [PubMed: 14643655]
24. Wikoff WR, Liljas L, Duda RL, Tsuruta H, Hendrix RW, Johnson JE. Topologically linked protein rings in the bacteriophage HK97 capsid. Science. 2000; 289:2129–2133. [PubMed: 11000116]
25. Duda RL. Protein chainmail: catenated protein in viral capsids. Cell. 1998; 94:55–60. [PubMed: 9674427]
26. Conway JF, Wikoff WR, Cheng N, Duda RL, Hendrix RW, Johnson JE, Steven AC. Virus maturation involving large subunit rotations and local refolding. Science. 2001; 292:744–748. [PubMed: 11326105]
27. Duda RL, Ross PD, Cheng N, Firek BA, Hendrix RW, Conway JF, Steven AC. Structure and energetics of encapsidated DNA in bacteriophage HK97 studied by scanning calorimetry and cryo-electron microscopy. J Mol Biol. 2009; 391:471–483. [PubMed: 19540242]
28. Gan L, Conway JF, Firek BA, Cheng N, Hendrix RW, Steven AC, Johnson JE, Duda RL. Control of crosslinking by quaternary structure changes during bacteriophage HK97 maturation. Mol Cell. 2004; 14:559–569. [PubMed: 15175152]
29. Gan L, Speir JA, Conway JF, Lander G, Cheng N, Firek BA, Hendrix RW, Duda RL, Liljas L, Johnson JE. Capsid conformational sampling in HK97 maturation visualized by X-ray crystallography and cryo-EM. Structure. 2006; 14:1655–1665. [PubMed: 17098191]
30. Lata R, Conway JF, Cheng N, Duda RL, Hendrix RW, Wikoff WR, Johnson JE, Tsuruta H, Steven AC. Maturation dynamics of a viral capsid: visualization of transitional intermediate states. Cell. 2000; 100:253–263. [PubMed: 10660048]
31. Lee KK, Gan L, Tsuruta H, Hendrix RW, Duda RL, Johnson JE. Evidence that a local refolding event triggers maturation of HK97 bacteriophage capsid. J Mol Biol. 2004; 340:419–433. [PubMed: 15210344]
32. Lee KK, Gan L, Tsuruta H, Moyer C, Conway JF, Duda RL, Hendrix RW, Steven AC, Johnson JE. Virus capsid expansion driven by the capture of mobile surface loops. Structure. 2008; 16:1491–502. [PubMed: 18940605]
33. Ross PD, Cheng N, Conway JF, Firek BA, Hendrix RW, Duda RL, Steven AC. Crosslinking renders bacteriophage HK97 capsid maturation irreversible and effects an essential stabilization. Embo J. 2005
34. Ross PD, Conway JF, Cheng N, Dierkes L, Firek BA, Hendrix RW, Steven AC, Duda RL. A free energy cascade with locks drives assembly and maturation of bacteriophage HK97 capsid. J Mol Biol. 2006; 364:512–525. [PubMed: 17007875]
35. Wikoff WR, Conway JF, Tang J, Lee KK, Gan L, Cheng N, Duda RL, Hendrix RW, Steven AC, Johnson JE. Time-resolved molecular dynamics of bacteriophage HK97 capsid maturation interpreted by electron cryo-microscopy and X-ray crystallography. J Struct Biol. 2006; 153:300–306. [PubMed: 16427314]
36. Benevides JM, Bondre P, Duda RL, Hendrix RW, Thomas GJ Jr. Domain structures and roles in bacteriophage HK97 capsid assembly and maturation. Biochemistry. 2004; 43:5428–5436. [PubMed: 15122908]
37. Conway JF, Cheng N, Ross PD, Hendrix RW, Duda RL, Steven AC. A thermally induced phase transition in a viral capsid transforms the hexamers, leaving the pentamers unchanged. J Struct Biol. 2007; 158:224–232. [PubMed: 17188892]
38. Huang RK, Khayat R, Lee KK, Gertsman I, Duda RL, Hendrix RW, Johnson JE. The Prohead-I Structure of Bacteriophage HK97: Implications for Scaffold-Mediated Control of Particle Assembly and Maturation. J Mol Biol. 2011; 408:541–554. [PubMed: 21276801]
39. Xie Z, Hendrix RW. Assembly in vitro of bacteriophage HK97 proheads. J Mol Biol. 1995; 253:74–85. [PubMed: 7473718]
40. Baird L, Lipinska B, Raina S, Georgopoulos C. Identification of the Escherichia coli sohB gene, a multicopy suppressor of the HtrA (DegP) null phenotype. J Bacteriol. 1991; 173:5763–5770. [PubMed: 1885549]

41. Qiu X, Janson CA, Culp JS, Richardson SB, Debouck C, Smith WW, Abdel-Meguid SS. Crystal structure of varicella-zoster virus protease. Proc Natl Acad Sci U S A. 1997; 94:2874–2879. [PubMed: 9096314]
42. Shieh HS, Kurumbail RG, Stevens AM, Stegeman RA, Sturman EJ, Pak JY, Wittwer AJ, Palmier MO, Wiegand RC, Holwerda BC, Stallings WC. Three-dimensional structure of human cytomegalovirus protease. Nature. 1996; 383:279–282. [PubMed: 8805708]
43. Tong L, Qian C, Massariol MJ, Bonneau PR, Cordingley MG, Lagace L. A new serine-protease fold revealed by the crystal structure of human cytomegalovirus protease. Nature. 1996; 383:272–275. [PubMed: 8805706]
44. Forrer P, Jaussi R. High-level expression of soluble heterologous proteins in the cytoplasm of Escherichia coli by fusion to the bacteriophage lambda head protein D. Gene. 1998; 224:45–52. [PubMed: 9931426]
45. Iwai H, Forrer P, Pluckthun A, Guntert P. NMR solution structure of the monomeric form of the bacteriophage lambda capsid stabilizing protein gpD. J Biomol NMR. 2005; 31:351–356. [PubMed: 15929002]
46. Yang F, Forrer P, Dauter Z, Conway JF, Cheng N, Cerritelli ME, Steven AC, Pluckthun A, Wlodawer A. Novel fold and capsid-binding properties of the lambda-phage display platform protein gpD. Nat Struct Biol. 2000; 7:230–237. [PubMed: 10700283]
47. McGuffin LJ, Bryson K, Jones DT. The PSIPRED protein structure prediction server. Bioinformatics. 2000; 16:404–405. [PubMed: 10869041]
48. Lupas A, Van Dyke M, Stock J. Predicting coiled coils from protein sequences. Science. 1991; 252:1162–1164. [PubMed: 2031185]
49. Li Y, Conway JF, Cheng N, Steven AC, Hendrix RW, Duda RL. Control of Virus Assembly: HK97 "Whiffleball" Mutant Capsids Without Pentons. J Mol Biol. 2005; 348:167–182. [PubMed: 15808861]
50. Showe MK, Black LW. Assembly core of bacteriophage T4: an intermediate in head formation. Nat New Biol. 1973; 242:70–75. [PubMed: 4512235]
51. Tsugita A, Black LW, Showe MK. Protein cleavage during virus assembly: characterization of cleavage in T4 phage. J Mol Biol. 1975; 98:217–215. [PubMed: 1195380]
52. Black LW, Ahmad-Zadeh C. Internal proteins of bacteriophage T4D: their characterization and relation to head structure and assembly. J Mol Biol. 1971; 57:71–92. [PubMed: 4930576]
53. Streiff MB, Iida S, Bickle TA. Expression and proteolytic processing of the darA antirestriction gene product of bacteriophage P1. Virology. 1987; 157:167–171. [PubMed: 3029955]
54. Prasad BV, Prevelige PE, Marietta E, Chen RO, Thomas D, King J, Chiu W. Three-dimensional transformation of capsids associated with genome packaging in a bacterial virus. J Mol Biol. 1993; 231:65–74. [PubMed: 8496966]
55. Nelson RA, Reilly BE, Anderson DL. Morphogenesis of bacteriophage phi 29 of Bacillus subtilis: preliminary isolation and characterization of intermediate particles of the assembly pathway. J Virol. 1976; 19:518–532. [PubMed: 822176]
56. Droge A, Santos MA, Stiege AC, Alonso JC, Lurz R, Trautner TA, Tavares P. Shape and DNA packaging activity of bacteriophage SPP1 procapsid: protein components and interactions during assembly. J Mol Biol. 2000; 296:117–132. [PubMed: 10656821]
57. Sutter M, Boehringer D, Gutmann S, Gunther S, Prangishvili D, Loessner MJ, Stetter KO, Weber-Ban E, Ban N. Structural basis of enzyme encapsulation into a bacterial nanocompartment. Nat Struct Mol Biol. 2008; 15:939–947. [PubMed: 19172747]
58. Akita F, Chong KT, Tanaka H, Yamashita E, Miyazaki N, Nakaishi Y, Suzuki M, Namba K, Ono Y, Tsukihara T, Nakagawa A. The crystal structure of a virus-like particle from the hyperthermophilic archaeon Pyrococcus furiosus provides insight into the evolution of viruses. J. Mol. Biol. 2007; 368:1469–1483. [PubMed: 17397865]
59. Namba K, Hagiwara K, Tanaka H, Nakaishi Y, Chong KT, Yamashita E, Armah GE, Ono Y, Ishino Y, Omura T, Tsukihara T, Nakagawa A. Expression and molecular characterization of spherical particles derived from the genome of the hyperthermophilic euryarchaeote Pyrococcus furiosus. J Biochem. 2005; 138:193–199. [PubMed: 16091594]

60. Dokland T. Scaffolding proteins and their role in viral assembly. Cell Mol Life Sci. 1999; 56:580–603. [PubMed: 11212308]
61. Lengyel JA, Goldstein RN, Marsh M, Sunshine MG, Calendar R. Bacteriophage P2 head morphogenesis: cleavage of the major capsid protein. Virology. 1973; 53:1–23. [PubMed: 4574872]
62. Wang S, Palasingam P, Nokling RH, Lindqvist BH, Dokland T. In vitro assembly of bacteriophage P4 procapsids from purified capsid and scaffolding proteins. Virology. 2000; 275:133–144. [PubMed: 11017795]
63. Shaw JE, Murialdo H. Morphogenetic genes C and Nu3 overlap in bacteriophage lambda. Nature. 1980; 283:30–35. [PubMed: 6444246]
64. Ziegelhoffer T, Yau P, Chandrasekhar GN, Kochan J, Georgopoulos C, Murialdo H. The purification and properties of the scaffolding protein of bacteriophage lambda. J Biol Chem. 1992; 267:455–461. [PubMed: 1530932]
65. Morgan GJ, Hatfull GF, Casjens S, Hendrix RW. Bacteriophage Mu genome sequence: analysis and comparison with Mu-like prophages in Haemophilus, Neisseria and Deinococcus. J Mol Biol. 2002; 317:337–359. [PubMed: 11922669]
66. Liu FY, Roizman B. The herpes simplex virus 1 gene encoding a protease also contains within its coding domain the gene encoding the more abundant substrate. J Virol. 1991; 65:5149–5156. [PubMed: 1654435]
67. Newcomb WW, Homa FL, Thomsen DR, Ye Z, Brown JC. Cell-free assembly of the herpes simplex virus capsid. J Virol. 1994; 68:6059–6063. [PubMed: 8057482]
68. Preston VG, al-Kobaisi MF, McDougall IM, Rixon FJ. The herpes simplex virus gene UL26 proteinase in the presence of the UL26.5 gene product promotes the formation of scaffold-like structures. J Gen Virol. 1994; 75(Pt 9):2355–2366. [PubMed: 8077934]
69. Georgopoulos, C.; Tilly, K.; Casjens, S. Lambdoid phage head assembly. In: Lambda, II.; hendrix, RW.; Roberts, JW.; FW, S.; Weisberg, RA., editors. Cold Spring Harbor Laboratory. new york: Cold Spring Harbor; 1983. p. 279-303.
70. Black, LW.; Showe, MK.; Steven, AC. Morphogenesis of the T4 head. In: Karam, J., editor. Molecular Biology of Bacteriophage T4. Washington, DC: ASM Press; 1994. p. 518-558.
71. 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. 1976; 107:55–69. [PubMed: 1003460]
72. Hintermann E, Kuhn A. Bacteriophage T4 gene 21 encodes two proteins essential for phage maturation. Virology. 1992; 189:474–482. [PubMed: 1641978]
73. Rao VB, Black LW. Structure and assembly of bacteriophage T4 head. Virol J. 2010; 7:356. [PubMed: 21129201]
74. Laemmli UK. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature. 1970; 227:680–685. [PubMed: 5432063]
75. Bair CL, Rifat D, Black LW. Exclusion of glucosyl-hydroxymethylcytosine DNA containing bacteriophages is overcome by the injected protein inhibitor IPI*. J Mol Biol. 2007; 366:779–789. [PubMed: 17188711]
76. Rohrer H, Zillig W, Mailhammer R. ADP-ribosylation of DNA-dependent RNA polymerase of Escherichia coli by an NAD+: protein ADP-ribosyltransferase from bacteriophage T4. Eur J Biochem. 1975; 60:227–238. [PubMed: 173540]
77. 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]
78. Hong YR, Black LW. Protein folding studies in vivo with a bacteriophage T4 expression-packaging-processing vector that delivers encapsidated fusion proteins into bacteria. Virology. 1993; 194:481–490. [PubMed: 8503169]
79. Hong YR, Black LW. An expression-packaging-processing vector which selects and maintains 7-kb DNA inserts in the blue T4 phage genome. Gene. 1993; 136:193–198. [PubMed: 8294002]
80. Hong YR, Mullaney JM, Black LW. Protection from proteolysis using a T4::T7-RNAP phage expression-packaging-processing system. Gene. 1995; 162:5–11. [PubMed: 7557416]

81. Mullaney JM, Black LW. GFP:HIV-1 protease production and packaging with a T4 phage expression-packaging processing system. Biotechniques. 1998; 25:1008–1012. [PubMed: 9863054]
82. Mullaney JM, Black LW. Activity of foreign proteins targeted within the bacteriophage T4 head and prohead: implications for packaged DNA structure. J Mol Biol. 1998; 283:913–929. [PubMed: 9799633]
83. Mullaney JM, Thompson RB, Gryczynski Z, Black LW. Green fluorescent protein as a probe of rotational mobility within bacteriophage T4. J Virol Methods. 2000; 88:35–40. [PubMed: 10921840]
84. Lubbers MW, Waterfield NR, Beresford TP, Le Page RW, Jarvis AW. Sequencing and analysis of the prolate-headed lactococcal bacteriophage c2 genome and identification of the structural genes. Appl Environ Microbiol. 1995; 61:4348–4356. [PubMed: 8534101]
85. Effantin G, Figueroa-Bossi N, Schoehn G, Bossi L, Conway JF. The tripartite capsid gene of Salmonella phage Gifsy-2 yields a capsid assembly pathway engaging features from HK97 and lambda. Virology. 2010; 402:355–365. [PubMed: 20427067]
86. McClelland M, Sanderson KE, Spieth J, Clifton SW, Latreille P, Courtney L, Porwollik S, Ali J, Dante M, Du F, Hou S, Layman D, Leonard S, Nguyen C, Scott K, Holmes A, Grewal N, Mulvaney E, Ryan E, Sun H, Florea L, Miller W, Stoneking T, Nhan M, Waterston R, Wilson RK. Complete genome sequence of Salmonella enterica serovar Typhimurium LT2. Nature. 2001; 413:852–856. [PubMed: 11677609]
87. Patterson DP, Prevelige PE, Douglas T. Nanoreactors by programmed enzyme encapsulation inside the capsid of the bacteriophage P22. ACS Nano. 2012; 6:5000–5009. [PubMed: 22624576]
88. Kang S, Uchida M, O'Neil A, Li R, Prevelige PE, Douglas T. Implementation of p22 viral capsids as nanoplatforms. Biomacromolecules. 2010; 11:2804–2809. [PubMed: 20839852]
89. Minten IJ, Hendriks LJ, Nolte RJ, Cornelissen JJ. Controlled encapsulation of multiple proteins in virus capsids. J Am Chem Soc. 2009; 131:17771–17773. [PubMed: 19995072]
90. Rhee JK, Hovlid M, Fiedler JD, Brown SD, Manzenrieder F, Kitagishi H, Nycholat C, Paulson JC, Finn MG. Colorful virus-like particles: fluorescent protein packaging by the Qbeta capsid. Biomacromolecules. 2011; 12:3977–3981. [PubMed: 21995513]
91. Seebeck FP, Woycechowsky KJ, Zhuang W, Rabe JP, Hilvert D. A simple tagging system for protein encapsulation. J Am Chem Soc. 2006; 128:4516–4517. [PubMed: 16594656]
92. Fan C, Cheng S, Liu Y, Escobar CM, Crowley CS, Jefferson RE, Yeates TO, Bobik TA. Short N-terminal sequences package proteins into bacterial microcompartments. Proc Natl Acad Sci U S A. 2010; 107:7509–7514. [PubMed: 20308536]
93. Thomas JA, Weintraub ST, Wu W, Winkler DC, Cheng N, Steven AC, Black LW. Extensive proteolysis of head and inner body proteins by a morphogenetic protease in the giant Pseudomonas aeruginosa phage phiKZ. Mol Microbiol. 2012; 84:324–339. [PubMed: 22429790]
94. 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]
95. Studier FW, Rosenberg AH, Dunn JJ, Dubendorff JW. Use of T7 RNA polymerase to direct expression of cloned genes. Methods Enzymol. 1990; 185:60–89. [PubMed: 2199796]
96. Lathe RF, Kieny MP, Schmitt D, Curtis P, Lecocq JP. M13 bacteriophage vectors for the expression of foreign proteins in Escherichia coli: the rabies glycoprotein. J Mol Appl Genet. 1984; 2:331–342. [PubMed: 6330261]
97. Dreyfuss G, Adam SA, Choi YD. Physical change in cytoplasmic messenger ribonucleoproteins in cells treated with inhibitors of mRNA transcription. Mol Cell Biol. 1984; 4:415–423. [PubMed: 6717428]
98. Marvik OJ, Sharma P, Dokland T, Lindqvist BH. Bacteriophage P2 and P4 assembly: alternative scaffolding proteins regulate capsid size. Virology. 1994; 200:702–714. [PubMed: 8178454]
99. McGeoch DJ, Dalrymple MA, Davison AJ, Dolan A, Frame MC, McNab D, Perry LJ, Scott JE, Taylor P. The complete DNA sequence of the long unique region in the genome of herpes simplex virus type 1. J Gen Virol. 1988; 69(Pt 7):1531–1574. [PubMed: 2839594]

- A capsid maturation protease is essential for producing functional HK97 capsids
- Proteolytic activity is in the N-terminal 2/3 of the HK97 protease
- HK97 protease packaging signal is in the C-terminal 1/5 of the protein
- Packaging signals or cotranslation are common mechanisms of encapsidation
- Phage, herpesviruses and encapsulins share protein packaging mechanisms

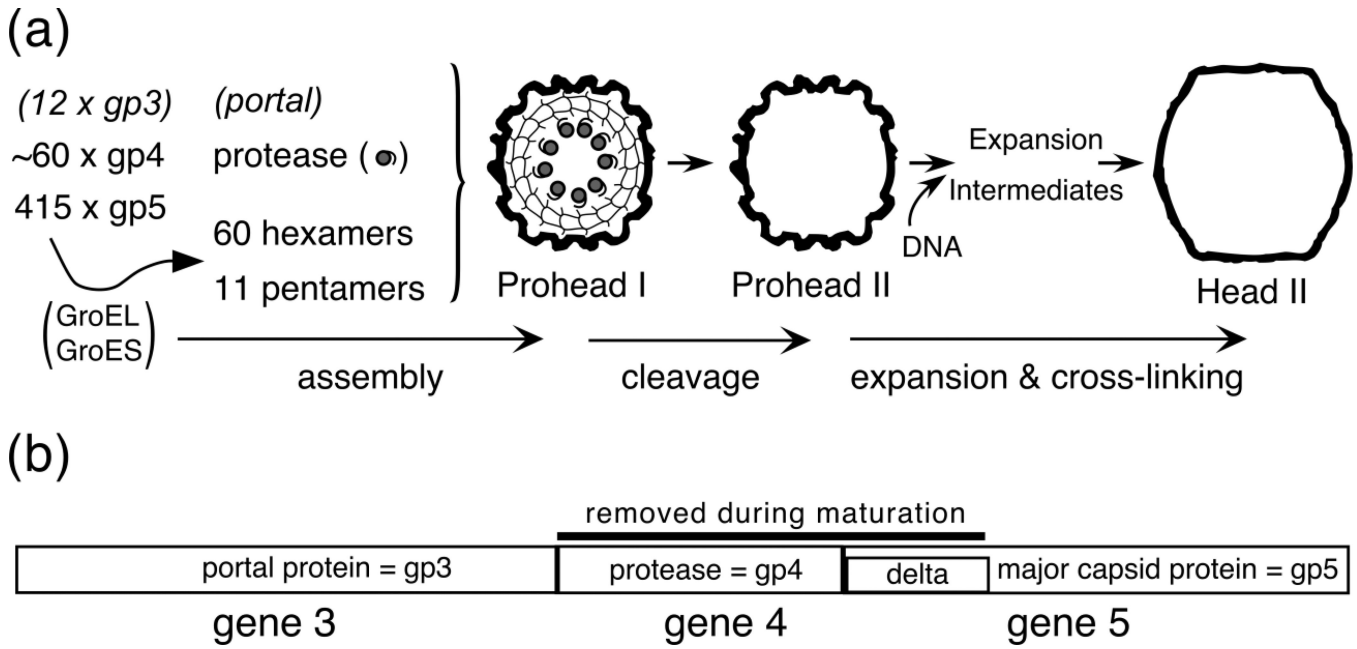

Figure 1. HK97 capsid assembly pathway and gene map

(a). A schematic view of HK97 capsid assembly showing how three HK97 proteins form a transient Prohead I, which is converted to Prohead II by the action of the co-assembled protease subunits. The proheads particles used in this study were made by expressing the proteins from plasmids and did not contain portals. Particles with no portal have 12 pentamers (b). Map of the 3 contiguous HK97 capsid genes.

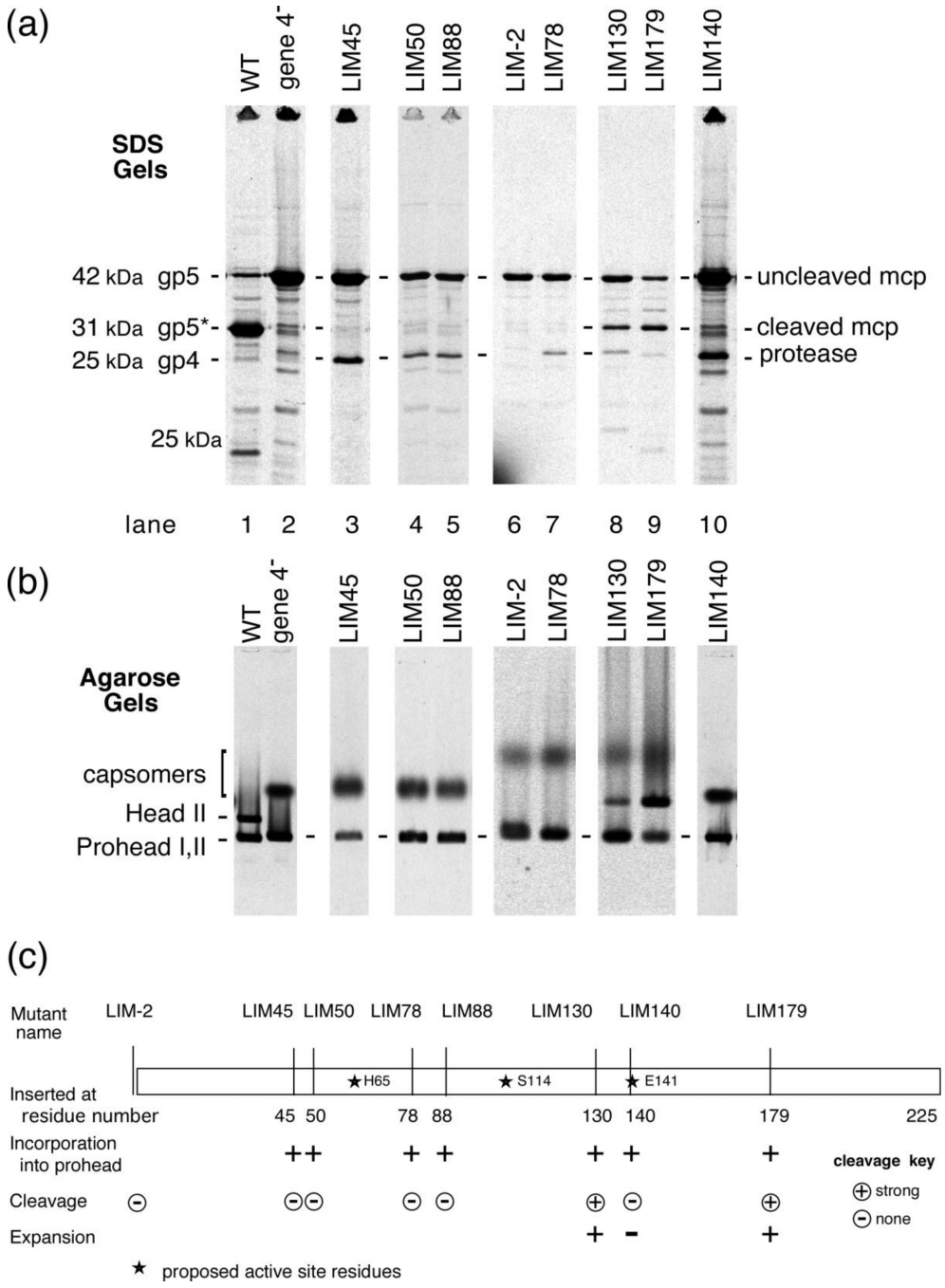

Figure 2. Biochemical phenotypes of protease mutants

analyzed by electrophoresis using (a) SDS gels (b) agarose gels. Cultures of cells bearing wild-type or mutant genes were grown in minimal medium, induced for protein expression and radio-labeled for 10 minutes after suppressing host protein synthesis using rifampicin. Cells were collected, lysed and samples were prepared for electrophoresis, which included TCA precipitation for the SDS gel samples. Gels were run, dried and autoradiographed. The figure shows images of the developed film. Positions of cleaved and uncleaved HK97 proteins are marked. Lysates were loaded directly onto the agarose gels without denaturation. In the agarose gel, note that Proheads I and II migrate similarly, but Heads migrate somewhat slower. Capsomers run as diffuse bands or spots because they are poorly

sieved by the gel. (c) Protease gene mutant map with phenotypes. The positions of the linker insertion mutants in the HK97 protease gene are shown to scale. Several properties of the mutants are shown using the symbols described in the figure itself. The location of the proposed active site residues are also marked.

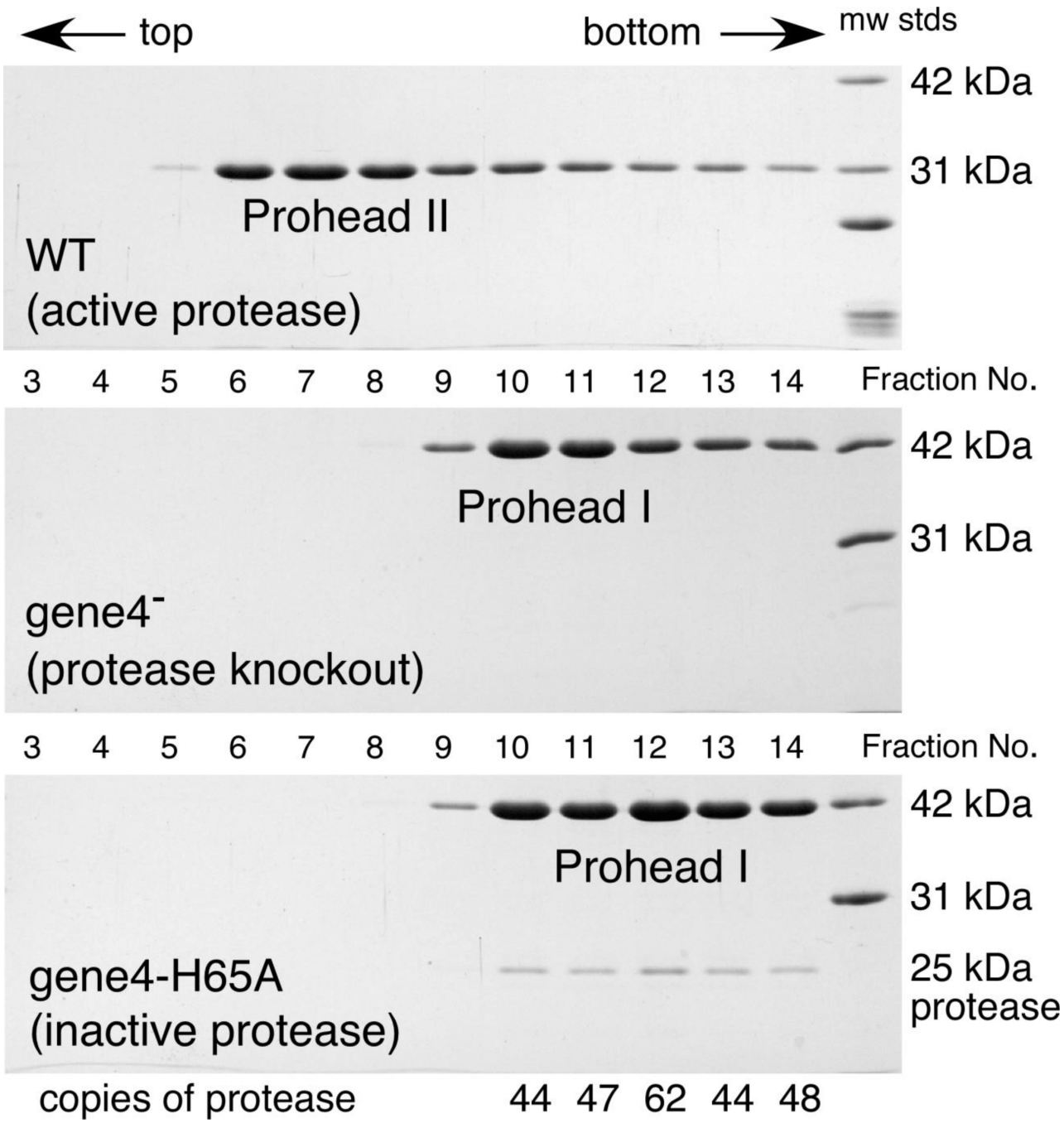

Figure 3. Analysis of HK97 proheads made with a catalytically inactive maturation protease

Extracts from 3 plasmid expression strains were run on 10–30% glycerol gradients at 40,000 rpm for 100 min in a Beckman SW41 rotor, and fractionated. Portions of fractions were TCA precipitated, heated with SDS, run on 12% SDS gels, and stained. The plasmids were as follows: (TOP) wild-type (pV0), producing wild-type gp4 & gp5 and thus producing Prohead II; (MIDDLE) (pVB) protease frameshift, producing Prohead I; and (BOTTOM) (pV0-gene4-H65A), producing Prohead I containing inactive protease molecules.

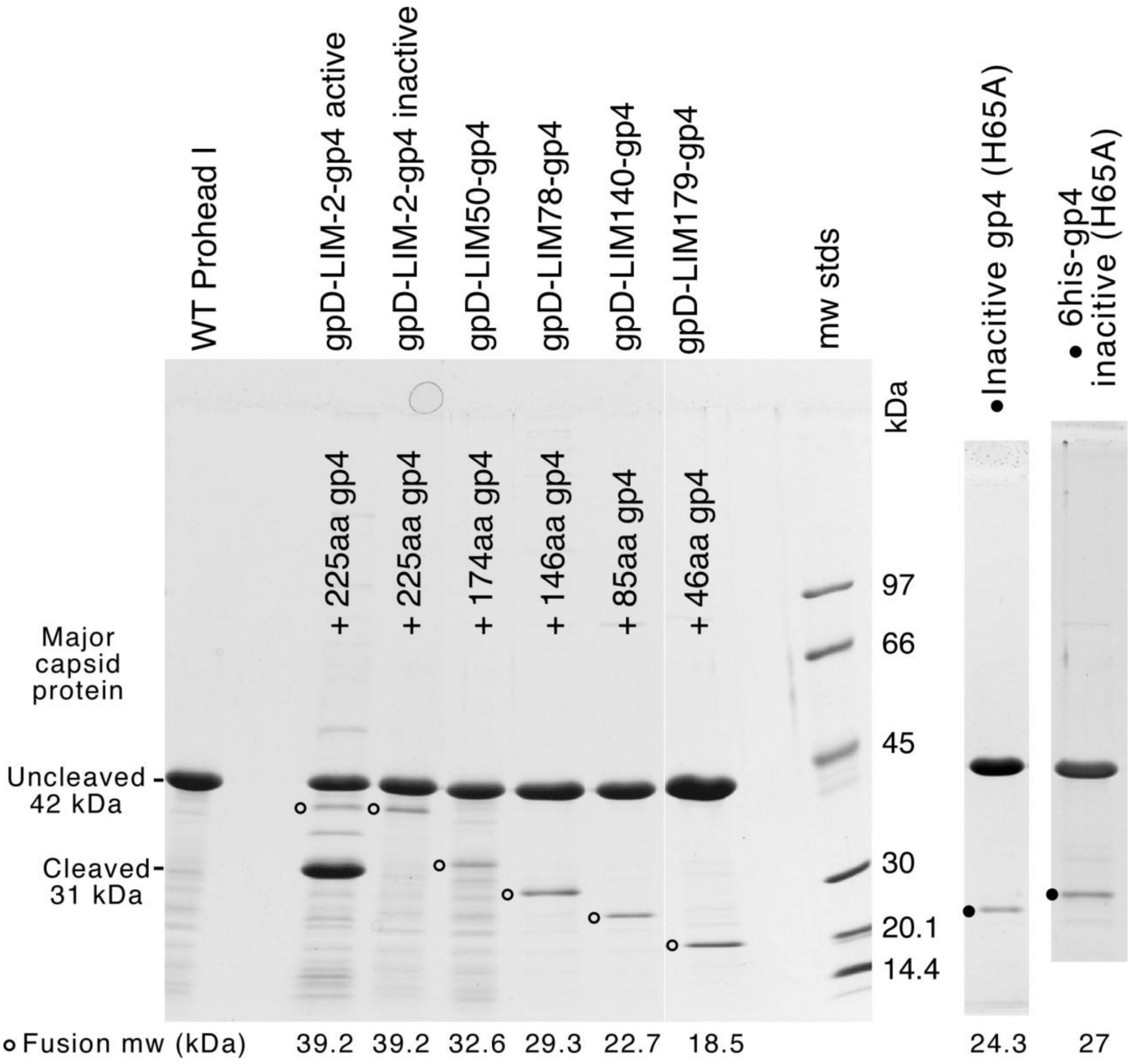

Figure 4. The C-terminus of HK97 protease is sufficient for packaging proteins into HK97 proheads

Proheads were made and purified using plasmids that express the HK97 major capsid protein gp5 and either wild-type protease, or gpD-protease fusion proteins made using several of the protease linker insertion mutants described in the text. The relative size and gp4 content of each is indicated in the figure. The lane on the right shows the protein composition of proheads made with the inactive protease mutant H65A for comparison. The predicted molecular weights for all protease variants are indicated at the bottom.

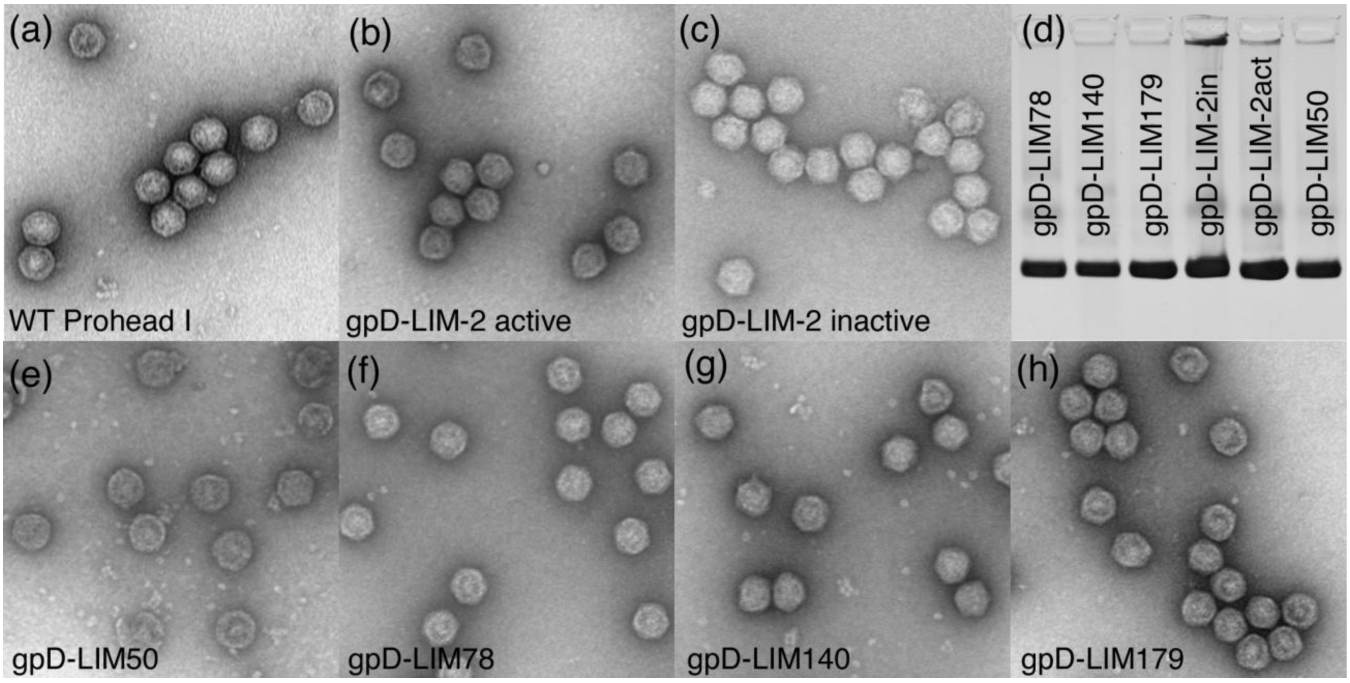

Figure 5. Electron micrographs and agarose gel analysis of Proheads containing gpD-gp4 fusion proteins

Panel (d) contains an agarose gel of the preparations used to make the micrographs in the other panels. Note that migration was essentially identical for all samples. Other panels show electron micrographs (stained with 1% uranyl acetate) of purified proheads containing six different gpD-gp4 fusion proteins, which are identified in the figure. The small dots in the background are capsomers - a result of the dissociation of some of the proheads in the samples. All samples were produced from small-scale cultures and purified through the ultracentrifugation step described in Materials and Methods. The wild-type Prohead I sample was produced separately and further purified on a glycerol gradient.

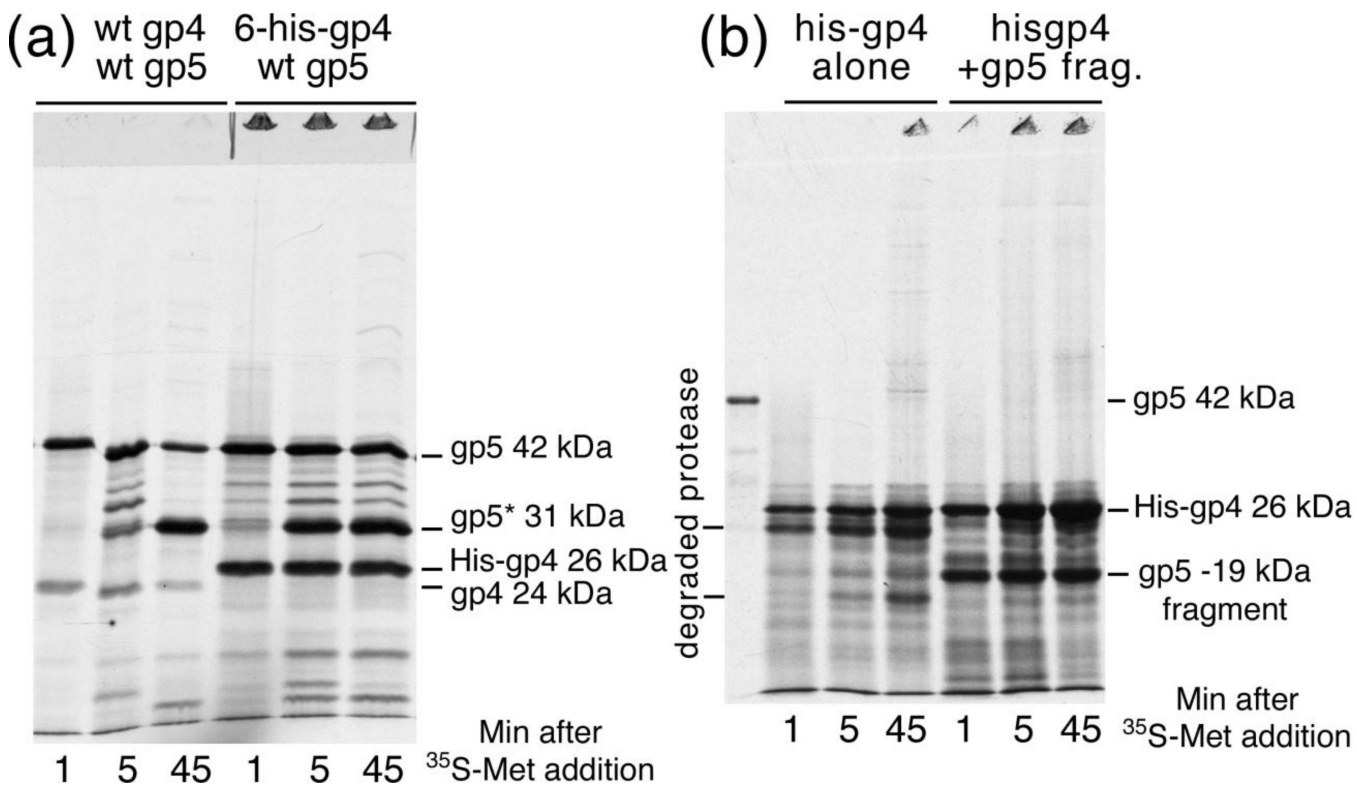

Figure 6. Expression of His-tagged gp4 in vivo

Cultures of cells bearing plasmids were treated as described in the legend to Figure 2 and under Materials and Methods, except samples were taken at 3 times as indicated in the figure. HK97 protein bands are identified in the figure. **(a). Cleavage of HK97 major capsid protein by 6-His protease**. The HK97 major capsid protein was expressed together with the wild-type protease from plasmid pV0 , and with the His-tagged protease from plasmid pTQ30-6Hisgp4–g5. **(b)**. Expression of the His-tagged protease by itself from plasmid pT7-6-6Hisgp4–38. The protease is rapidly cleaved to form a new band with a molecular weight a few kDa smaller; a band with an even lower molecular weight appears later. **(b). Stabilization of His-gp4 by co-expression with gp5 N-terminal fragment**. When the His-tagged protease is expressed from plasmid pTQ30-6Hisgp4-RV, the protease appears to stabilized by the N-terminal fragment of the major capsid protein that is co-expressed.

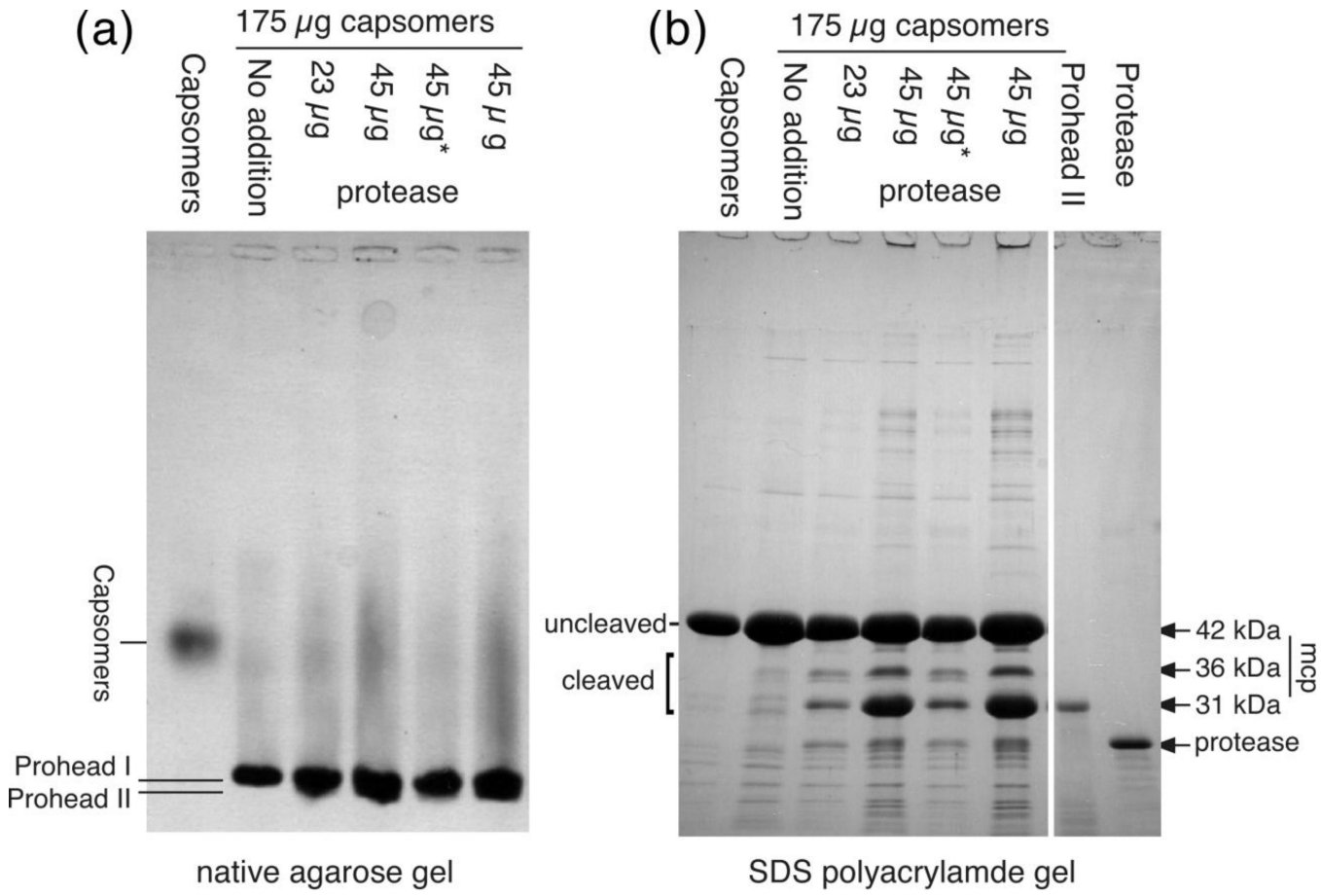

Figure 7. ***In vitro*** **assembly and proteolysis of HK97 proheads**

Capsomers (a mix of hexamers and pentamers) were mixed with protease in a Centricon 50 concentrator, diluted with Buffer TKG100 to a volume of 2 ml and centrifuged to concentrate the reactants and initiate assembly. The order of addition was protease, capsomers, buffer, except in the reactions labeled with "*", where the protease was added last. After overnight incubation at room temperature, the concentrates were collected, equalized in volume and run on gels for analysis. **(a) agarose gel (b) SDS gel**. HK97 capsomer, prohead and capsid protein bands are labeled in the figure.

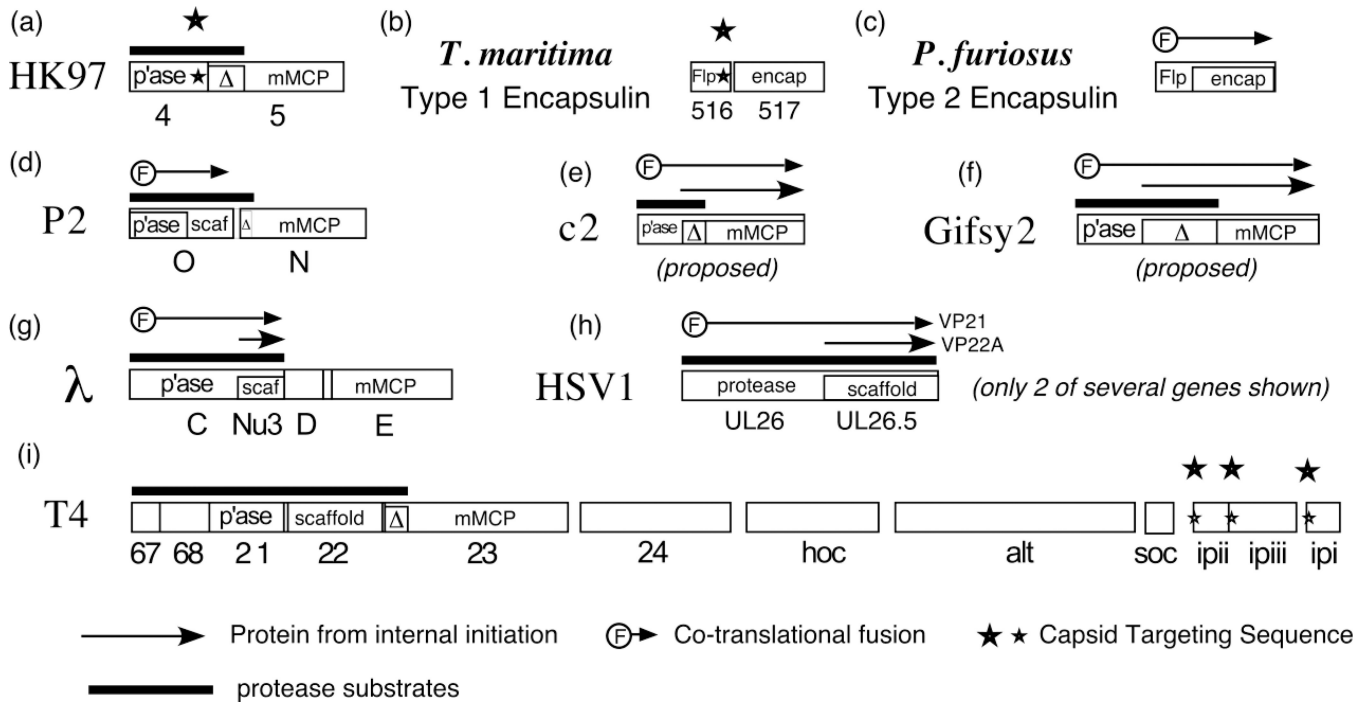

Figure 8. Arrangement of capsid targeting genes in phages

Sets of capsid genes relevant to capsid targeting are diagrammed here and further explained in the text. Genes are shown as rectangles, with overlapping genes shown as smaller rectangles within a larger one, and with the gene name or function within the rectangle. **Abbreviations**: mMCP, mature part of major capsid protein; Δ, delta domain of major capsid protein; p'ase, protease; scaf, scaffolding protein; Flp, ferritin-like protein; other symbols are defined in the figure. In (e) and (f), plausible ribosome binding sites are present upstream of internal Met codons that would allow the translation of a proposed shorter protein containing only a delta-domain like segment and the mature portion of major capsid protein, invoking a co-assembly mechanism like that found in λ, Mu, and HSV-1. References for these gene arrangements are as follows: (a) HK97[8;20](b) *Thermoga maritima*[57](c)*Pyrococccus furiosus*[57;58;59] (d) P2[18;61;98] (e) c2[84] (f) Gifsy2[85;86] (g) λ[63;64;69] (h) HSV-1[66; 67; 68; 99] (i) T4 [51; 52; 70; 73; 77]

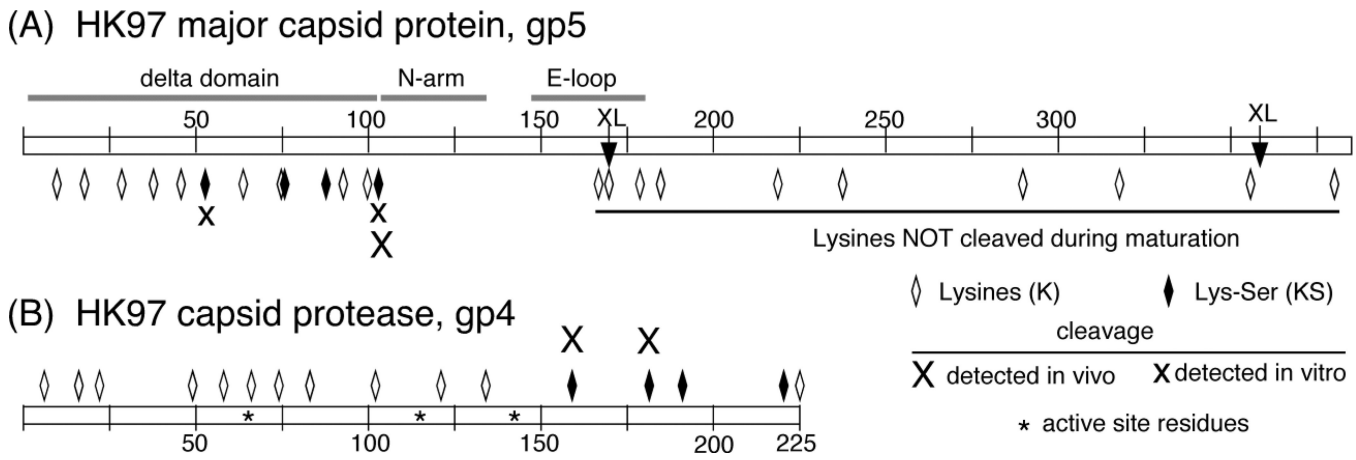

Figure 9. Locations of proposed and confirmed cleavage sites in the HK97 major capsid protein and maturation protease

Active site residues and cleavage sites are marked with symbols as indicated in the figure.

Table 1
Spot complementation tests with amber phage in genes 4 and 5

Plasmid containing strains were tested for their ability to support the growth of wild type phage or phage with amber mutations in the major capsid protein gene (*5*) or the protease gene (*4*), as described in Materials and Methods.

Plasmid	gene 5 am relative EOP	gene 4 am relative EOP	HK97 wild type relative EOP
wild type gp4-gp5	1*	1*	1*
gene 4 frameshift	1	0.001	1
gene 4 LIM45	1	0.001	1
gene 4 LIM50	1	0.001	1
gene 4 LIM78	1	0.001	1
gene 4 LIM88	1	0.001	1
gene 4 LIM130	1	0.001	1
gene 4 LIM140	0.3	0.001	1
gene 4 LIM179	1	0.5	1
gene 4 H65A	0.1	0.001	1
wild type gp4-gp5 induced	1*	1*	1*
wild type gp4-gp5 uninduced	0.1	0.01	1
gene 4 frameshift induced	1	0.001	1
gene 4 frameshift uninduced	0.01	0.01	0.1
6His-gp4–gp5 induced	0.1	0.1	0.1
6His-gp4–gp5 uninduced	1	1	1

The values are all relative to the complementation achieved by wild-type versions of genes - these are marked with a"*".