Published in final edited form as:

Mol Microbiol. 2010 July ; 77(2): 444–455. doi:10.1111/j.1365-2958.2010.07219.x.

Functional Analysis of the Highly Antigenic Outer Capsid Protein, Hoc, a Virus Decoration Protein from T4-like Bacteriophages

Taheri Sathaliyawala^{1,*}, Mohammad Z. Islam^{1,*}, Qin Li^{1,2}, Andrei Fokine³, Michael G. Rossmann³, and Venigalla B. Rao^{1,#}

¹Department of Biology, The Catholic University of America, Washington, DC, 20064, USA

²Laboratory Center, The Fourth Affiliated Hospital, China Medical University, Shenyang 110005, China

³Department of Biological Sciences, Purdue University, West Lafayette, IN 47907, USA.

Summary

Bacteriophage T4 is decorated with 155 copies of the highly antigenic outer capsid protein, Hoc. The Hoc molecule (40 kDa) is present at the center of each hexameric capsomer and provides a good platform for surface display of pathogen antigens. Biochemical and modeling studies show that Hoc consists of a string of four domains, three immunoglobulin (Ig)-like and one non-Ig domain at the C-terminus. Biochemical data suggest that the Hoc protein has two functional modules, a capsid binding module containing domains 1 and 4 and a solvent-exposed module containing domains 2 and 3. This model is consistent with the dumbbell-shaped cryo-EM density of Hoc observed in the reconstruction of the T4 capsid. Mutagenesis localized the capsid binding site to the C-terminal 25 amino acids, which are predicted to form two β -strands flanking a capsid binding loop. Mutations in the loop residues, ESRNG, abolished capsid binding, suggesting that these residues might interact with the major capsid protein, gp23*. With the conserved capsid binding module forming a foothold on the virus and the solvent-exposed module able to adapt to bind to a variety of surfaces, Hoc probably provides survival advantages to the phage, such as increasing the virus concentration near the host, efficient dispersion of the virus, and exposing the tail for more efficient contact with the host cell surface prior to infection.

Keywords

bacteriophage T4; virus assembly; Hoc; Soc; virus decoration protein; phage display

Introduction

Many viruses fortify their exterior surface by “decoration” proteins that provide survival advantages in their natural environments. First discovered in phage T4 and named Hoc and Soc (Ishii & Yanagida, 1975), such proteins have since been described for phages λ (gpD) (Sternberg & Weisberg, 1977), T5 (pb10) (Effantin *et al.*, 2006), L (Dec) (Tang *et al.*, 2006), ϕ 29 (gp 8.5) (Morais *et al.*, 2005), PRD1 (P30) (Abrescia *et al.*, 2004), adenoviruses (pIX) (Parks, 2005) and herpes viruses (VP23, VP19c) (Saad *et al.*, 1999). In addition to providing insights into the architectural principles of virus design (Qin *et al.*, 2010), these

[#]Corresponding author: Venigalla B. Rao, Department of Biology, 103 McCort Ward Hall, The Catholic University of America, 620 Michigan Ave., N.E., Washington, DC, 20064. Phone: (202) 319-5271; Fax: (202) 319-6161; rao@cua.edu.

*these authors contributed equally and share first authorship

proteins offer useful platforms to array antigens and proteins on virus surfaces for a variety of biotechnological applications (Ren *et al.*, 1996, Jiang *et al.*, 1997, Shivachandra *et al.*, 2006).

The T4 capsid is an elongated icosahedron (Figure 1), 120 nm long and 86 nm wide, built with three essential proteins (Fokine *et al.*, 2004, Iwasaki *et al.*, 2000): i) 930 copies of the major capsid protein, gene product (gp) 23* (49 kDa; * represents cleaved form), which form a hexagonal lattice characterized by triangulation numbers $T_{\text{end}}=13$ and $T_{\text{mid}}=20$; ii) 55 copies of the vertex protein gp24* (47 kDa), which form eleven of the twelve capsid vertices, one pentamer at each vertex (Black *et al.*, 1994; Fokine *et al.*, 2005); and iii) 12 copies of the portal protein gp20 (61 kDa), which form the unique twelfth vertex; the dodecameric ring through which DNA enters during packaging and exits during infection. In addition, two outer capsid proteins, Hoc, the highly antigenic outer capsid protein, and Soc, the small outer capsid protein, decorate the capsid surface (Ishii & Yanagida, 1975, Ishii & Yanagida, 1977, Ishii *et al.*, 1978). The Hoc molecule (40 kDa) is present at the center of most or all of 155 gp23 capsomers. The Soc molecules (10 kDa) bind to the capsid surface between the gp23* capsomers with up to 870 copies per capsid. The Hoc and Soc proteins are dispensable for capsid assembly. Deletion of either or both genes does not affect phage production, viability, or infectivity under standard laboratory conditions (Ishii & Yanagida, 1977). However it was shown that Soc helps the virus to survive under hostile environments (Ishii & Yanagida, 1977; Steven *et al.*, 1992; Qin *et al.*, 2010).

Molecules of the phage T4 major capsid protein gp23 assemble into a lattice around a scaffolding core consisting of the major core protein, gp22, the assembly maturation protease, gp21, and at least eight other proteins (Black *et al.*, 1994). The protease cleaves off the N-terminal 65 amino acids of gp23 molecules (producing gp23*) and degrades the scaffold into small peptides, which exit the capsid, creating an empty space in the interior of the capsid for genome packaging. The gp23* lattice expands resulting in an increase in the capsid volume by ~50% and the creation of binding sites for Hoc and Soc on the capsid surface (Carrascosa, 1978, Rao & Black, 1985). Hoc and Soc bind to the expanded capsid with high affinity and specificity, while the pentameric gp17 motor docked on the portal packages DNA into the capsid to near crystalline density (~500 mg/ml) (Sun *et al.*, 2008).

The structure of Soc was determined recently (Qin *et al.*, 2010). It is a tadpole-shaped molecule which binds to two gp23* molecules on the capsid surface. Interaction of Soc molecule with two gp23* molecules from adjacent capsomers glues the capsomers together. Furthermore, trimerization of Soc molecules through C-terminal interactions clamps together three gp23* hexameric capsomers. The assembly of 270 such triple clamps on the T4 capsid surface provides great stability to the phage allowing it to survive under hostile environments such as extreme pH (pH 11), high temperature (60°C), osmotic shock, and a host of denaturing agents.

Hoc binds at the center of the gp23* capsomer but it does not significantly enhance capsid stability (Ross *et al.*, 1985). Thus the function of Hoc and the selective advantage it provides to phage T4 remained unknown. The cryo-EM density observed for Hoc molecules in the reconstruction of T4 capsid has the shape of a dumbbell (Figure 1). Here, we propose that the two knobs of the Hoc dumbbell correspond to two functional modules, a capsid proximal module that interacts with gp23*, and a solvent-exposed module that might interact with host bacterial surface. The capsid binding site is localized to the C-terminal 25 amino acids, which contains a conserved predicted loop flanked by two β -strands that presumably orient the loop sequence for interaction with the major capsid protein, gp23*. This organization, a conserved capsid binding module forming a foothold on the virus and a variable surface module that can adapt to binding onto a variety of surfaces, likely provides survival

advantages to the virus such as enriching the virus near the host, efficient dispersion of the virus, or helping to orient the phage particles to expose the tail for efficient capture of the host for infection.

Results

Hoc encodes four domains three of which have an Ig-like fold

Sequence analyses show that Hoc sequence consists of four tandemly linked, ~10 kDa domains (Bateman *et al.*, 1997, Fraser *et al.*, 2006). The first three have an Ig-like fold, which typically consists of seven β -strands folded into two anti-parallel β -sheets packed into a β -sandwich (Halaby *et al.*, 1999, Bork *et al.*, 1994). The Ig-like domains are further classified into four superfamilies: the Ig domain (I-Set), the fibronectin 3 domain (FN3), the bacterial Ig-like domain (Big2), and the polycystic kidney disease domain (PKD) (Jing *et al.*, 2002). The first two Hoc domains show greater similarity to the PKD domain whereas the third domain is more similar to the I-set domain (Supplementary figure S1). Since such domains are present in numerous surface proteins and the exposed residues are involved in protein-carbohydrate and protein-protein interactions (Bateman *et al.*, 1996), the Hoc Ig domains may have a similar role. The fourth Hoc domain at the C-terminus however appears to be different as it does not show typical Ig-like features.

The C-terminal Hoc domain 4 is conserved among the T4-family phages

The C-terminal Hoc domain 4 is moderately well conserved among the T4-family phages. The N-terminal domain 1 is also conserved but less well than domain 4. In fact, the T4 Hoc domain 4 sequence is nearly identical to that in phages RB69 and JS98 (Figure 2). Domains 2 and 3 are not as well conserved or entirely missing in some phages. For instance, the phage JS98 does not have domain 2 and in phages 44RR2, 25, and 31, both domains 2 and 3 are missing (Figure 2; Supplementary figure S2). The RB69 Hoc has an additional Ig-like domain inserted into domain 2, and RB43 Hoc has only a short 61 amino acid peptide corresponding to a truncated domain 4. As Hoc binds to the major capsid protein gp23* that is well conserved among these phages (80–98% sequence similarity), the conservation of Hoc domain 4 indicates that it is at least one of the domains that interacts with gp23*.

The domain 4 contains a capsid binding site

Previous studies had shown that Hoc-fused with HIV capsid protein p24 at its N-terminus is displayed on phage T4 (Sathaliyawala *et al.*, 2006). This property was used to test the capsid binding properties of various Hoc domains. Fusion to p24 was also necessary because it increased the size of ~10 kDa size Hoc domains to a high molecular weight range (>35 kDa), allowing unambiguous analysis of capsid binding by SDS-PAGE.

Recombinant clones were constructed by fusing the N-terminus of each Hoc domain or a series of domains, to the C-terminus of p24 (Figure 3, panel A) and the fusion proteins were purified (panel B). The linker sequences between the Hoc domains were retained to minimize potential interference on folding. Native gel electrophoresis showed that the purified proteins migrated as compact bands suggesting that the proteins were folded correctly (panel C). To test for binding, the Hoc domain fusion proteins were incubated with *hoc-soc*- phage and the virus particles were separated from unbound proteins by high speed centrifugation. The bound and unbound fractions were analyzed by SDS-PAGE¹ (see Experimental Procedures).

¹The panels shown in Figure 3 depict the set of analyses done to analyze the capsid binding function of numerous Hoc proteins constructed in this study.

Only the fusion proteins containing domain 4, namely H4 and H234, bound to the capsid² (panel D, lanes 9 and 15). No binding was detected with any of the other domains even at 100-fold excess of the domain molecules to capsid binding sites (lanes 3, 5, 7, 11, and 13; data not shown). Furthermore, the H234 protein, but not the H123 protein, competitively inhibited the binding of another Hoc fusion protein, the 120 kDa anthrax PA-Hoc, to capsid (panel E) demonstrating the specificity of domain-4 binding to the capsid.

The conserved ESRNG residues in domain 4 are probably involved in capsid binding

A series of deletions and point mutations indicated that the C-terminal 25 amino acids of domain 4 are important for binding to capsid (Supplementary Table 1 and data not shown). In particular, the 5 amino acid ESRNG sequence is strictly conserved among all members of the T4 phage family (Figure 4, panel A). Secondary structure predictions show that these residues form a loop between two structurally conserved β -strands (panel A; Supplementary figure S3). This loop sequence is therefore a strong candidate for direct interaction with gp23* whereas the flanking β -strands might be required to precisely orient the loop residues.

To test this hypothesis, a recombinant RB49 Hoc was constructed and tested for its binding to T4 *hoc-soc*- phage. Unlike the T4 Hoc that shows three to four bands (one or two bands below the expected 40 kDa band and one band above; panel B, lane 2), the RB49 Hoc shows a single band at 55 kDa³ (panel B, lane 3) allowing more precise quantification of capsid binding. The binding data showed that the RB49 Hoc bound to T4 phage with nearly the same copy number as T4 Hoc (panel C). Four mutations were then introduced into the conserved ESRNG loop sequence. These mutants lost nearly all the capsid binding activity (panel D).

Domains 1 and 4 likely form the capsid binding module of Hoc

Biochemical analyses suggest that domains 1 and 4 interact to form one of the two knobs of Hoc observed in the cryoEM reconstruction (Fokine et al., 2004). The ectopically (*E. coli*) expressed Hoc is a monomer and highly soluble, and so are the individually expressed Hoc domains 1, 2, and 3, but domain 4 is insoluble. Domain 4 could be solubilized by urea denaturation, although only a small amount remained in solution after renaturation and purification by Histrap chromatography. The soluble protein was in an aggregated form as it eluted in the void volume of Superdex-200 gel filtration column that has an exclusion limit of ~600 kDa and apparently bound to the capsid as an oligomer as indicated by the high copy number of bound H4 (see Figure 3D, lane 9). Similar solubility behavior was observed with H234 in which domain 1 was deleted, but not with H123 or H23, which are soluble and monomeric. These results and consistent behavior by other Hoc mutants (Supplementary Table 1) demonstrated that any clone that contained domain 4 but not domain 1 resulted in aggregation or precipitation. But when domain 1 was present, including a direct fusion of domain 1 to domain 4 (see below), it resulted in the expression of monomeric and soluble protein. These results are consistent with the high incidence of hydrophobic amino acids in domain 4 (Figure 2). We hypothesize that certain hydrophobic residues of domain 4 may be solvent-exposed in the absence of domain 1 and cause aggregation unless they are stabilized by interaction with domain 1. Since mutational data as described above suggested that the capsid binding site is located in the C-terminus of domain 4, the knob formed by domain 1 and domain 4 interactions is likely to be the capsid binding module. Sequence alignments

²The copy number of bound H4 (Figure 3, panel D, lane 9) was higher than the expected 155 copies per capsid because the purified domain 4 aggregates and probably bound to the capsid as an oligomer.

³The T4 Hoc shows three to four bands by SDS-PAGE; one or two bands below the expected 40 kDa band and one band above (Figure 4, panel B, lane 2). On the other hand, the RB49 Hoc protein shows a single band at 55 kDa (Figure 4, panel B, lane 3) although its calculated molecular weight is 47 kDa. The reason(s) for these anomalous migrations of Hoc proteins are unknown. Unlike the T4 Hoc protein which is a monomer, the RB49 Hoc protein showed various oligomer forms (Supplementary figure S4).

showing the more conserved nature of domains 1 and 4 and more variable nature of domains 2 and 3 (Figure 2; Supplementary figure S3) further supported this interpretation.

A structural model for Hoc

In the cryo-EM reconstruction of T4 capsid, the density of Hoc has the shape of a dumbbell with two knobs (Fokine et al., 2004). One knob projects out on the surface, with its tip 60Å from the capsid, whereas the other knob, the capsid binding module, is buried in the central depression of the gp23* capsomer. Structural modeling using the program LOOPP (Learning, Observing and Outputting Protein Patterns), available on the bioinformatics server <http://cbsuapps.tc.cornell.edu/loopp.aspx> (Teodorescu *et al.*, 2004, Meller & Elber, 2001, Tobi & Elber, 2000), generated a horseshoe shaped model (Figure 5, panel A) based on the structure of Hemolin, an insect surface protein belonging to Ig superfamily (Su *et al.*, 1998). Hemolin, an immune protein, is induced by bacterial infection and binds to the surface lipopolysaccharide of the bacterium. Consistent with the sequence alignments (Figure 2 and Supplementary figure S3), the model has a string of four domains folded between domains 2 and 3, forming two modules that correspond to the two knobs of the dumbbell-shaped molecule seen in the cryo-EM reconstruction (Fokine et al., 2004).

The horseshoe Hoc model was manually fitted into the cryo-EM reconstruction of the T4-capsid (panel B) (Fokine et al., 2004). The volume of the cryo-EM density corresponding to Hoc molecule was insufficient to accommodate the entire atomic model. The weak density for the Hoc molecules in the cryo-EM reconstruction might be due to variability of the orientation and rotation of Hoc because each Hoc molecule can bind to the center of gp23 capsomers in six possible orientations related by the quasi 6-fold axis of the gp23 capsomer. Furthermore, different Hoc orientations in different T4 particles may result in a weak averaged density for the Hoc molecules observed in the three-dimensional reconstruction. The fitting suggested that regions of both domains 1 and domain 4, particularly some of the C-terminal amino acids that are shown to be important by biochemical studies, would be in contact with the major capsid protein (panel C).

Designing a “mini” T4 Hoc

Based on the above biochemical and modeling studies, a mini-T4 Hoc was engineered by deleting domains 2 and 3 and connecting domain 1 (amino acids 1–98) to domain 4 (amino acids 284–376) (H1-H4) through the intervening linkers. Consistent with the above data and interpretations, the mini-Hoc was found to be a soluble monomer (Supplementary Table 1). Binding experiments show that H1-H4 binds to *hoc*⁻*soc*⁻ phage with similar affinity and copy number as the full-length Hoc (Figure 6, panel A, and data not shown), which was further confirmed by Western blotting using anti-Hoc antibodies (panel B, lane 4).

Sequence alignments (Figure 2; Supplementary figures S2 and S3) suggest that Hoc from the T4- family phage 44RR2 contains only domains 1 and 4, which is equivalent to the mini-Hoc, the putative capsid binding module. This protein has significantly diverged from T4 Hoc, having only about 46% sequence similarity and 28% identity with T4 Hoc. Capsid binding experiments show that the recombinant 44RR2 Hoc binds to T4 phage (panel C), suggesting that the key amino acids involved in Hoc -gp23* interactions are conserved. Among these are the ESRNG loop residues that are required for capsid binding (Figure 4) and strictly conserved in 44RR2 Hoc.

Discussion

First discovered in 1975 as a nonessential structural protein (Ishii & Yanagida, 1975), Hoc and its homologues form a unique class of virus decoration proteins that attach to the outer

surface of the phage T4 capsid. However, their functional significance in virus life cycle remained unknown. Unlike the more commonly found “triplex” proteins such as Soc, which assemble as trimers at three fold vertices (Qin et al., 2010, Lander *et al.*, 2008, Yang *et al.*, 2000), Hoc is a monomer partly buried in the central depression of the hexameric capsomer (Fokine et al., 2004). Whereas the triplex proteins glue adjacent capsomers and reinforce the capsid, Hoc binding provides little, if any, stabilization. The surface exposed Hoc is highly immunogenic and has been extensively used to display pathogen antigens (Shivachandra et al., 2006, Shivachandra *et al.*, 2007, Li *et al.*, 2007).

Bioinformatics and biochemical data suggest that the Hoc sequence contains a string of four domains, three Ig-like domains and a fourth that appears to be a non-Ig domain at the C-terminus (Bateman et al., 1997, Fraser et al., 2006). The data also suggest that the Hoc molecule folds at its center forming two modules, a surface exposed module containing domains 2 and 3 and a capsid binding module containing domains 1 and 4. A structural model of Hoc resembles hemolin, an insect surface protein (Su et al., 1998), and can be satisfactorily fitted into the cryoEM density map of phage T4 capsid (Fokine et al., 2004). Biochemical and mutagenesis results suggest that domains 1 and 4 interact with each other and the capsid, and that the capsid binding site, at least in part, is present in domain 4. The similar solubility and binding properties an engineered mini-Hoc consisting only of domains 1 and 4 is consistent with this model. However, alternative models such as a linear string of domains, with the domain 4 at the capsid binding end, cannot be completely excluded. Since the T4 capsid is built with 155 gp23* hexameric capsomers, attachment of one Hoc to each capsomer generates a surface “armor” of symmetrically arrayed spikes that the virus can use for a variety of surface interactions.

Mutational studies mapped the capsid binding region to the C-terminal 25 amino acids of Hoc domain 4. Of these the ERSNG sequence is strictly conserved among all the T4-family phages. Mutation of any of these residues abolished capsid binding, suggesting that this motif is directly involved in Hoc binding to capsid. Secondary structure predictions show that this sequence forms a loop flanked by two structurally conserved β -strands. It therefore appears that the contact region between Hoc and the capsid, at least in part, consists on Hoc of the ESRNG loop that interacts with gp23*, whereas the flanking β -strands orient the loop residues appropriately for binding to capsid.

Hoc provides a useful platform for the display of pathogen antigens, which elicit strong immune responses (Shivachandra et al., 2006, Sathaliyawala et al., 2006, Wu *et al.*, 2007). Full-length antigens as large as the 90 kDa anthrax toxins have been successfully displayed up to the maximum copy number, 155 per capsid (Shivachandra et al., 2007). Consistent with the present results were the previous observations that when antigens were fused to the N-terminus of Hoc, there was no change in the binding affinity, whereas fusion to the C-terminus caused a 13–400 fold decrease in affinity (Sathaliyawala et al., 2006, Shivachandra et al., 2007). The latter was likely due to distortion of the binding site by the large antigen mass attached to the C-terminus.

The results presented here suggest new avenues for antigen display. The linker regions between domains 2 and 3 or between domains 1 and 4 of min-Hoc might be even better sites for display of pathogen antigens. Antigens at these sites are likely to be more exposed for presentation of epitopes to the immune system as they are at the distal end of the outward projecting Hoc molecule. Indeed, plague antigens inserted between domains 2 and 3 were successfully displayed on phage T4 (Mahalingam and Rao, unpublished data).

There are six equivalent orientations with which Hoc might bind to the center of each hexameric gp23* capsomer. However the slight curvature of each capsomer imposed by the

shape of the head might give some preference to some orientations with respect to others. In particular the capsomers near the pentameric vertices might not be conformationally equivalent to those at the center of the icosahedral face due to greater local curvature of the capsid at the vertices (Morais et al 2005). In addition, occupation of one of the sites sterically blocks access to the other available sites in the same hexamer, thus limiting the occupation of only one of the six sites by the Hoc monomer.

Previous studies showed that T4 *hoc*- mutant phage migrates as a faster species when compared to the wild-type phage. This behavior was shown to be related to the aggregation property of T4 *hoc*- mutant phage in low ionic strength buffers (Yamaguchi & Yanagida, 1980, Childs, 1980). The present results are consistent with this hypothesis in that binding of mini-Hoc to *hoc*- phage also reduced the electrophoretic mobility (data not shown). Thus, apparently Hoc helps T4 from aggregation either by charge repulsion (pI of T4 Hoc is 4.7) or by interfering with the interactions between phage particles. The symmetrically arrayed Hoc spikes that are raised by ~60Å from the capsid surface might prevent gp23*-gp23* interactions. Indeed, in cryoEM images we observed that the *hoc-soc*- phage particles frequently aggregate, apparently through interactions between capsids. Thus one of the functions of *hoc* might be to prevent the aggregation of phage particles in the infected cell where concentration of newly-assembled phage can be quite high.

Ig domains play essential roles in surface interactions between cells and organisms (Bork et al., 1994, Bateman et al., 1996, Jing et al., 2002). Recent bioinformatics evidence suggests that the Ig domains are widely distributed in phage genomes as insertions into the coding sequences of surface-exposed structural proteins (Fraser et al., 2006, 2007). For instance, the major tail protein of phage λ has a 246 amino acid Ig domain inserted into the C-terminus, and phages T3 and φ29 have an Ig domain inserted into the C-terminus of the major capsid protein. But Hoc is the only reported example where a separate protein enriched with Ig domains decorates the virus capsid. From the evolutionary standpoint, this would give maximum flexibility to adapt the phage surface to the virus' ecological niche. Indeed, the surface exposed domains 2 and 3 show the greatest variation and the phages RB69 and RB30 even contain a fourth Ig domain inserted into domain 2. Phage T5 has a Hoc-like protein on its surface with different Ig domains (Effantin et al., 2006). These observations suggest that the Hoc Ig domains are not essential for phage infection but might provide some advantages for survival in their natural environments. For instance, the surface exposed Ig domains might form weak nonspecific interactions with the surface carbohydrates and other molecules of the bacterial cell wall (Fraser et al., 2007). Such interactions would enrich phage particles on the bacterial surface where specific receptors for phage infection are localized, but not strong enough to immobilize the phage nonproductively on the bacterial surface. Alternatively, the phage can “piggy-back” on a bacterium which is not necessarily its host and “hitch-hike” to distant sites in search of *E. coli* host, or be protected in the mild biological environment of the bacterial surface. Hoc might also allow the phage to adhere to bacteria or biofilms exposing the tail injection machine and the associated receptor molecules to efficiently “capture” the host bacterium for infection. Surrounded by as many as 465 Ig domains, the Hoc-containing T4 phage could thus have a significant survival advantage over one that does not.

Experimental Procedures

Bacteria, phage, and DNAs

E. coli P301 (sup⁻) was used to prepare *hoc-soc*-(*hoc.Q21am-soc.del*) phage stocks and *E. coli* XL-10 cells Gold (Stratagene, La Jolla, CA) were used for initial transformation and maintenance of recombinant constructs. After confirming the sequences the clones were transferred into the expression strain, *E. coli* BL21 (DE3)/pLys-S (Stratagene, La Jolla, CA),

to allow IPTG (isopropyl- β -D-thiogalactopyranoside) induced over-expression of recombinant proteins (Studier *et al.*, 1990). The T7 expression plasmids pET15b (Amp^r) and pET28b (Kan^r) (EMD Biosciences, Inc) were used as the cloning vectors. Purified phage T4, 44RR2 and RB49 genomic DNAs were used as templates to amplify respectively the T4 Hoc and its mutants, 44RR2 Hoc, and RB49 Hoc and its mutants. The HIV DNA corresponding to the full length p24-gag sequence of HIV-1 su10 isolate was amplified from the pDAB72 plasmid.

Gene fusions and plasmid constructs

Gene fusions were constructed using the splicing-by-overlap extension (SOE) strategy (Shivachandra et al., 2006, Horton *et al.*, 1989, Kuebler & Rao, 1998). The end primers contained sites for appropriate restriction enzymes for directional cloning of the DNA into the vectors. The 5' end primers were designed so that the insertion resulted in in-frame fusion of the hexa-histidine-containing 25 amino acid sequence from the vector with the recombinant sequence. The amplified DNAs were purified by agarose gel electrophoresis, digested with NheI and BamHI for T4 Hoc constructs, and NdeI and BamHI for RB49 and 44RR2 Hoc constructs, and ligated with the gel-purified vector that was previously digested with the same restriction enzymes. The ligated DNA was then transformed into *E. coli* XL10 Gold cells (Stratagene) and plasmid DNAs were prepared from individual colonies by Miniprep kit (Fermentas). The entire insert was sequenced to confirm that there were no errors in the cloned DNA (Davis Sequencing). The clones were then transformed into *E. coli* BL21 (DE3) pLys-S or RIPL for over-expression of the recombinant Hoc proteins.

Purification of recombinant Hoc proteins

The BL21 (DE3) pLys-S cells harboring Hoc clones were induced with 1 mM IPTG for 3 h at 30°C. The cells were harvested by centrifugation at $4,000 \times g$ for 12 min at 4°C and pellets were stored at -70°C. The pellets from 1 liter culture were resuspended in 40 ml of Histrap binding buffer (50 mM Tris-HCl, pH 8.0, 20 mM imidazole, and 300 mM NaCl) and lysed by French-press (Aminco). The soluble fraction containing his-tagged fusion protein was separated from cell debris and any insoluble inclusion bodies by centrifugation at $34,000 \times g$ for 20 min at 4°C.

Constructs which expressed Hoc in soluble fraction were purified from the supernatant. Those clones that expressed Hoc in the insoluble fraction were centrifuged at $4,000 \times g$ for 5 min and the pellet was dissolved in 50 ml of urea buffer (8 M Urea, 50 mM Tris-HCl, pH 8.0, 20 mM imidazole and 300 mM NaCl). The sample was then centrifuged at $34,000 \times g$ for 20 min to remove cell debris and the supernatant was loaded onto a Histrap column (AKTA-prime, GE healthcare) pre-equilibrated with the same buffer. The protein was renatured by washing the column with a decreasing urea gradient (8 - 0 M) in binding buffer. The protein was eluted with 20 - 500 mM linear imidazole gradient in the same buffer. Most of the insoluble mutant proteins precipitated again immediately after elution. However, some of the proteins, e.g., H4 and H234, retained a small fraction of the protein in solution. This protein apparently was in an aggregated form as indicated by the elution of H4 in the void volume of Superdex-200 gel filtration column that has an exclusion limit of ~600 kDa. The soluble protein was used in the *in vitro* binding experiments.

The soluble Hoc proteins were purified from the supernatant using the same procedure except that the loading and wash buffers contained no urea. The peak fractions containing the Hoc protein were pooled and concentrated by Amicon Ultra-4 centrifugal filtration (3 kDa cut-off). The protein was then purified by size exclusion chromatography using Hi-Load 16/60 Superdex-200 (prep-grade) gel filtration column (GE Healthcare) in a buffer

containing 20 mM Tris-HCl (pH 8.0) and 100 mM NaCl. The protein was concentrated by Amicon Ultra-4 centrifugal filtration and stored at -70°C .

In vitro binding of Hoc on the T4 capsid

In vitro assembly of Hoc on T4 capsid was performed as described previously (Shivachandra et al., 2006, Shivachandra et al., 2007, Li et al., 2007). Briefly, about 2×10^{10} (or in some cases 3×10^{10}) T4 *hoc-soc*- phage were centrifuged at $34,000 \times g$ for 45 min in a low-bind Eppendorf tube. The sedimented phage particles were resuspended in binding buffer (50 mM sodium phosphate pH 7.0, 75 mM NaCl, and 1 mM MgSO_4). The purified recombinant Hoc proteins were also centrifuged at $34,000 \times g$ for 45 min to remove any precipitated protein. The supernatant was added to phage suspension at various ratios of Hoc molecules to capsid binding site in a total reaction volume of 150 μl . After incubating 15–45 min at room temperature, the phage particles were sedimented by centrifugation at $34,000 \times g$ for 45 min. The unbound supernatant was separated from the bound phage pellet and the pellet was washed twice with 1 ml each of binding buffer. When necessary, a control reaction was included in which the T4 phage was omitted. This accounted for any nonspecific precipitation of Hoc protein during incubation. In most experiments there was no significant background because, as described above, each protein was pre-sedimented at high speed to remove any aggregates. Furthermore, the reactions were performed in low-bind Eppendorf tubes to minimize non-specific binding. The final pellets were resuspended in 10 μl of assembly buffer and analyzed by SDS-PAGE. The gels were stained with Coomassie blue (BioRad) or Imperial stain (Thermo Scientific) and the protein bands were quantified by laser densitometry (PDSI, GE Healthcare). The density volumes of Hoc, gp23* and gp18 bands were determined for each lane separately and the number of Hoc molecules per capsid was calculated using the known copy numbers of gp23* or gp18. Where necessary, the apparent K_d (association constant) and B_{max} (maximum copies of Hoc bound per capsid) were determined by non-linear regression analysis using the equation $Y = B_{\text{max}}X / (K_d + X)$ (SigmaPlot 8.0 software).

Western blotting in Figure 6B was done by transferring the proteins separated by SDS-PAGE to PVDF membrane followed by treatments with rabbit polyclonal anti-Hoc antibodies and peroxidase conjugated secondary antibodies. The membrane was developed using ECL-plus reagent and exposed to film for 5 minutes.

Structural modeling and cryo-EM fitting

Three-dimensional models of T4 Hoc protein were generated using LOOPP (Learning, Observing and Outputting Protein Patterns) bioinformatics server <http://cbsuapps.tc.cornell.edu/loopp.aspx> (Teodorescu *et al.*, 2004, Meller & Elber, 2001, Tobi & Elber, 2000). One of the models based on the structure of Hemolin, an insect surface protein belonging to Ig superfamily (Su *et al.*, 1998) had a horseshoe-like shape, consistent with biochemical data. The model was visually fitted into cryo-EM density map using the program Chimera ([UCSF Chimera-a visualization system for exploratory research and analysis](#) (Pettersen et al., 2004)

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by a National Science Foundation grant (MCB-0443899) to M.G.R., a National Institutes of Health grant (NIAID, AI056443) to V.B.R., and a National Institutes of Health grant (NIAID, R56AI081726) to M.G.R. and V.B.R.

References

- Abrescia NG, Cockburn JJ, Grimes JM, Sutton GC, Diprose JM, Butcher SJ, Fuller SD, San Martin C, Burnett RM, Stuart DI, Bamford DH, Bamford JK. Insights into assembly from structural analysis of bacteriophage PRD1. *Nature*. 2004; 432:68–74. [PubMed: 15525981]
- Bateman A, Eddy SR, Chothia C. Members of the immunoglobulin superfamily in bacteria. *Protein Sci*. 1996; 5:1939–1941. [PubMed: 8880921]
- Bateman A, Eddy SR, Mesyanzhinov VV. A member of the immunoglobulin superfamily in bacteriophage T4. *Virus Genes*. 1997; 14:163–165. [PubMed: 9237357]
- Black LW, Showe MK, Steven AC. Morphogenesis of the T4 head. 1994:218–258.
- Bork P, Holm L, Sander C. The immunoglobulin fold. Structural classification, sequence patterns and common core. *J Mol Biol*. 1994; 242:309–320. [PubMed: 7932691]
- Carrascosa JL. Head maturation pathway of bacteriophages T4 and T2. IV. In vitro transformation of T4 head-related particles produced by mutants in gene 17 to capsid-like structures. *J Virol*. 1978; 26:420–428. [PubMed: 660719]
- Childs JD. Effect of hoc protein on the electrophoretic mobility of intact bacteriophage T4D particles in polyacrylamide gel electrophoresis. *J Mol Biol*. 1980; 141:163–173. [PubMed: 7441749]
- Effantin G, Boulanger P, Neumann E, Letellier L, Conway JF. Bacteriophage T5 structure reveals similarities with HK97 and T4 suggesting evolutionary relationships. *J Mol Biol*. 2006; 361:993–1002. [PubMed: 16876823]
- Fokine A, Chipman PR, Leiman PG, Mesyanzhinov VV, Rao VB, Rossmann MG. Molecular architecture of the prolate head of bacteriophage T4. *Proc Natl Acad Sci U S A*. 2004; 101:6003–6008. [PubMed: 15071181]
- Fokine A, Leiman PG, Shneider MM, Ahvazi B, Boeshans KM, Steven AC, Black LW, Mesyanzhinov VV, Rossmann MG. Structural and functional similarities between the capsid proteins of bacteriophages T4 and HK97 point to a common ancestry. *Proc Natl Acad Sci U S A*. 2005; 102:7163–7168. [PubMed: 15878991]
- Fraser JS, Maxwell KL, Davidson AR. Immunoglobulin-like domains on bacteriophage: weapons of modest damage? *Curr Opin Microbiol*. 2007; 10:382–387. [PubMed: 17765600]
- Fraser JS, Yu Z, Maxwell KL, Davidson AR. Ig-like domains on bacteriophages: a tale of promiscuity and deceit. *J Mol Biol*. 2006; 359:496–507. [PubMed: 16631788]
- Halaby DM, Poupon A, Mornon J. The immunoglobulin fold family: sequence analysis and 3D structure comparisons. *Protein Eng*. 1999; 12:563–571. [PubMed: 10436082]
- Horton RM, Hunt HD, Ho SN, Pullen JK, Pease LR. Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension. *Gene*. 1989; 77:61–68. [PubMed: 2744488]
- Ishii T, Yamaguchi Y, Yanagida M. Binding of the structural protein soc to the head shell of bacteriophage T4. *J Mol Biol*. 1978; 120:533–544. [PubMed: 650689]
- Ishii T, Yanagida M. Molecular organization of the shell of the T-even bacteriophage head. *J Mol Biol*. 1975; 97:655–660. [PubMed: 1185785]
- Ishii T, Yanagida M. The two dispensable structural proteins (soc and hoc) of the T4 phage capsid; their purification and properties, isolation and characterization of the defective mutants, and their binding with the defective heads in vitro. *J Mol Biol*. 1977; 109:487–514. [PubMed: 15127]
- Iwasaki K, Trus BL, Wingfield PT, Cheng N, Campusano G, Rao VB, Steven AC. Molecular architecture of bacteriophage T4 capsid: vertex structure and bimodal binding of the stabilizing accessory protein, Soc. *Virology*. 2000; 271:321–333. [PubMed: 10860886]
- Jiang J, Abu-Shilbayeh L, Rao VB. Display of a PorA peptide from *Neisseria meningitidis* on the bacteriophage T4 capsid surface. *Infect Immun*. 1997; 65:4770–4777. [PubMed: 9353063]
- Jing H, Takagi J, Liu JH, Lindgren S, Zhang RG, Joachimiak A, Wang JH, Springer TA. Archaeal surface layer proteins contain beta propeller, PKD, and beta helix domains and are related to metazoan cell surface proteins. *Structure*. 2002; 10:1453–1464. [PubMed: 12377130]
- Kuebler D, Rao VB. Functional analysis of the DNA-packaging/terminase protein gp17 from bacteriophage T4. *J Mol Biol*. 1998; 281:803–814. [PubMed: 9719636]

- Lander GC, Evilevitch A, Jeembaeva M, Potter CS, Carragher B, Johnson JE. Bacteriophage lambda stabilization by auxiliary protein gpD: timing, location, and mechanism of attachment determined by cryo-EM. *Structure*. 2008; 16:1399–1406. [PubMed: 18786402]
- Li Q, Shivachandra SB, Zhang Z, Rao VB. Assembly of the small outer capsid protein, Soc, on bacteriophage T4: a novel system for high density display of multiple large anthrax toxins and foreign proteins on phage capsid. *J Mol Biol*. 2007; 370:1006–1019. [PubMed: 17544446]
- Meller J, Elber R. Linear programming optimization and a double statistical filter for protein threading protocols. *Proteins*. 2001; 45:241–261. [PubMed: 11599028]
- Morais MC, Choi KH, Koti JS, Chipman PR, Anderson DL, Rossmann MG. Conservation of the capsid structure in tailed dsDNA bacteriophages: the pseudoatomic structure of phi29. *Mol Cell*. 2005; 18:149–159. [PubMed: 15837419]
- Parks RJ. Adenovirus protein IX: a new look at an old protein. *Mol Ther*. 2005; 11:19–25. [PubMed: 15585402]
- Qin L, Fokine A, O'Donnell E, Rao VB, Rossmann MG. Structure of the small outer capsid protein, Soc: a clamp for stabilizing capsids of T4-like phages. *J Mol Biol*. 2010; 395:728–741. [PubMed: 19835886]
- Rao VB, Black LW. DNA packaging of bacteriophage T4 proheads in vitro. Evidence that prohead expansion is not coupled to DNA packaging. *J Mol Biol*. 1985; 185:565–578. [PubMed: 4057255]
- Ren ZJ, Lewis GK, Wingfield PT, Locke EG, Steven AC, Black LW. Phage display of intact domains at high copy number: a system based on SOC, the small outer capsid protein of bacteriophage T4. *Protein Sci*. 1996; 5:1833–1843. [PubMed: 8880907]
- Ross PD, Black LW, Bisher ME, Steven AC. Assembly-dependent conformational changes in a viral capsid protein. Calorimetric comparison of successive conformational states of the gp23 surface lattice of bacteriophage T4. *J Mol Biol*. 1985; 183:353–364. [PubMed: 4020864]
- Saad A, Zhou ZH, Jakana J, Chiu W, Rixon FJ. Roles of triplex and scaffolding proteins in herpes simplex virus type 1 capsid formation suggested by structures of recombinant particles. *J Virol*. 1999; 73:6821–6830. [PubMed: 10400780]
- Sathaliyawala T, Rao M, Maclean DM, Bix DL, Alving CR, Rao VB. Assembly of human immunodeficiency virus (HIV) antigens on bacteriophage T4: a novel in vitro approach to construct multicomponent HIV vaccines. *J Virol*. 2006; 80:7688–7698. [PubMed: 16840347]
- Shivachandra SB, Li Q, Peachman KK, Matyas GR, Leppla SH, Alving CR, Rao M, Rao VB. Multicomponent anthrax toxin display and delivery using bacteriophage T4. *Vaccine*. 2007; 25:1225–1235. [PubMed: 17069938]
- Shivachandra SB, Rao M, Janosi L, Sathaliyawala T, Matyas GR, Alving CR, Leppla SH, Rao VB. In vitro binding of anthrax protective antigen on bacteriophage T4 capsid surface through Hoc-capsid interactions: a strategy for efficient display of large full-length proteins. *Virology*. 2006; 345:190–198. [PubMed: 16316672]
- Sternberg N, Weisberg R. Packaging of coliphage lambda DNA. II. The role of the gene D protein. *J Mol Biol*. 1977; 117:733–759. [PubMed: 609100]
- 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]
- Su XD, Gastinel LN, Vaughn DE, Faye I, Poon P, Bjorkman PJ. Crystal structure of hemolin: a horseshoe shape with implications for homophilic adhesion. *Science*. 1998; 281:991–995. [PubMed: 9703515]
- Sun S, Kondabagil K, Draper B, Alam TI, Bowman VD, Zhang Z, Hegde S, Fokine A, Rossmann MG, Rao VB. The structure of the phage T4 DNA packaging motor suggests a mechanism dependent on electrostatic forces. *Cell*. 2008; 135:1251–1262. [PubMed: 19109896]
- Tang L, Gilcrease EB, Casjens SR, Johnson JE. Highly discriminatory binding of capsid-cementing proteins in bacteriophage L. *Structure*. 2006; 14:837–845. [PubMed: 16698545]
- Teodorescu O, Galor T, Pillardy J, Elber R. Enriching the sequence substitution matrix by structural information. *Proteins*. 2004; 54:41–48. [PubMed: 14705022]
- Tobi D, Elber R. Distance-dependent, pair potential for protein folding: results from linear optimization. *Proteins*. 2000; 41:40–46. [PubMed: 10944392]

- Wu J, Tu C, Yu X, Zhang M, Zhang N, Zhao M, Nie W, Ren Z. Bacteriophage T4 nanoparticle capsid surface SOC and HOC bipartite display with enhanced classical swine fever virus immunogenicity: a powerful immunological approach. *J Virol Methods*. 2007; 139:50–60. [PubMed: 17081627]
- Yamaguchi Y, Yanagida M. Head shell protein hoc alters the surface charge of bacteriophage T4. Composite slab gel electrophoresis of phage T4 and related particles. *J Mol Biol*. 1980; 141:175–193. [PubMed: 7441750]
- 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]

Figure 1. Cryo-EM structure of phage T4 head

The major capsid protein gp23* (“*” represent the cleaved form), which forms a hexagonal lattice, is shown in cyan, the vertex protein gp24* in green, Hoc in the center of each gp23* hexameric capsomer in red, and Soc in white subunits (Fokine et al., 2004).

Figure 2. Sequence comparisons of Hoc proteins from T4-family bacteriophages
Sequence alignments of Hoc proteins from T4-family bacteriophages. Hoc protein sequences from T4-family bacteriophages are aligned by ClustalW (Accession numbers: T4_Hoc, NP_049793; RB69_Hoc, NP_861884; JS98_Hoc, YP_001595307; RB49_Hoc, NP_891739). The Hoc domains are highlighted in different colors: domain 1, blue; domain 2, green; domain 3, orange; domain 4, yellow. The fifth Ig-like insertion domain in RB69 is shown in red. Completely conserved residues are marked with an “*”, nearly completely conserved residues are marked with a “:” and partially conserved residues are marked with a “.”.

Figure 3. The Hoc domain 4 contains the capsid binding site

A. Schematics of recombinant Hoc domain constructs. Each recombinant has an N-terminal 25 amino acid peptide containing hexa-histidine tag (light green) and HIV-p24 (dark brown). The Hoc domains are shown in various colors (domain 1, dark blue; domain 2, light blue; domain 3, dark green; domain 4, red). The linkers between the domains are shown in light brown). The amino acids near the end points and junctions are shown above each construct. The numbers correspond to the number of amino acids between the sequences. **B.** SDS-polyacrylamide gel electrophoresis of purified p24-Hoc domain proteins. **C.** Native (nondenaturing) polyacrylamide gel electrophoresis of purified p24-Hoc domain proteins. The data from B and C show that the purified proteins are >90% pure and appear to be correctly folded. Had they been misfolded, the proteins would have migrated as a smear under non-denaturing conditions. **D.** Binding of p24-Hoc domain proteins to *hoc*⁻*soc*⁻ phage. Binding assays were performed according to the procedure described in Experimental Procedures. Lane 1, control *hoc*⁻*soc*⁻ phage; lanes 2, 4, 6, 8, 10, 12, and 14 show unbound protein present in the supernatant; lanes 3, 5, 7, 9, 11, 13, and 15 show protein bound to the phage. **E.** Hoc domain 4 inhibits binding of PA-Hoc to *hoc*⁻*soc*⁻ phage. Lane 1, control *hoc*⁻*soc*⁻ phage; lane 2, unbound PA-Hoc; lane 3, bound PA-Hoc; lane 4, unbound PA-Hoc and p24-H234; lane 5, bound PA-Hoc and H234; lane 6, unbound PA-Hoc and H123; lane 7, bound PA-Hoc and H123. The arrows show the positions of the bound proteins.

Figure 4. The conserved ESRNG loop residues are required for binding to capsid

A. Sequence alignment showing the C-terminal region of Hoc among the T4-family of phages. “E” represents β -strand, “H”, α -helix, and “-”, loop. The numbers at the end of each sequence correspond to the last amino acid of the Hoc coding sequence. **B.** SDS-PAGE of purified recombinant T4 Hoc and RB49 Hoc proteins. Marker lane shows the molecular weight standards with the sizes in kDa shown on the side of the panel. **C** *In vitro* binding of RB49 Hoc protein to *hoc-soc-* phage (2×10^{10} particles) at increasing ratio of Hoc molecules to capsid binding sites (0:1 to 50:1), as described in Experimental Procedures. The red arrow indicates the position of RB49 Hoc and the black arrow indicates the position of the major capsid protein, gp23*. **D** *In vitro* binding of mutant RB49 Hoc proteins to *hoc-soc-* phage (2×10^{10} particles) at a ratio of 100:1 Hoc molecules to capsid binding sites, as described in Experimental Procedures. Wild-type RB49 Hoc bound to T4 *hoc-soc-* phage (red arrow) but none of the mutants showed binding.

Figure 5. A structural model of Hoc

A. Homology model of Hoc based on the structure of Hemolin (Su et al., 1998). The polypeptide chain is colored with rainbow colors starting with blue at the amino end. **B.** The Hoc molecule was fitted into the cryo-EM reconstruction of phage T4 capsid. The figure shows the region corresponding to one gp23 capsomer. The dumbbell-shaped density in the center of the gp23 hexamer corresponds to the Hoc molecule. **C.** Interaction of Hoc molecule with the capsomer of the major capsid protein, gp23. The homology model of the gp23 hexamer (Fokine et al., 2005) is shown in magenta.

Figure 6. Mini-Hoc proteins bind to capsid

A. Binding of purified T4 mini-Hoc to *hoc-soc-* T4 phage. About 3×10^{10} *hoc-soc-* phage particles were incubated with 50:1 ratio of mini-Hoc protein to capsid binding sites and the samples were analyzed by SDS-PAGE. See Experimental Procedures for details. Lanes: 1, control *hoc-soc-* phage; 2, unbound mini-Hoc in the supernatant; 3, phage bound mini-Hoc in the pellet. Red arrows indicate the positions of mini-Hoc bands. **B.** Western blotting demonstrates the binding of mini-Hoc to *hoc-soc-* T4 phage. Polyclonal anti-Hoc antibodies were used to detect Hoc protein. Lanes: 1, *hoc-soc-* phage; 2, wild-type phage; 3, wild-type Hoc bound to *hoc-soc-* phage; 4, mini-Hoc bound to *hoc-soc-* phage. **C.** Binding of recombinant 44RR2 Hoc protein to the *hoc-soc-* phage. About 2×10^{10} phage particles were incubated with 44RR2 Hoc protein at an increasing ratio of Hoc molecules to capsid binding sites (0:1 to 20:1). See Experimental Procedures for details. The arrow indicates the position of bound 44RR2 Hoc that migrated very close to a phage T4 band.