How phage L capsid decoration protein distinguishes between nearly identical binding sites on an icosahedral virion

Rebecca Newcomer 1, Jason R. Schrad 2, Eddie B. Gilcrease 3, Sherwood R. Casjens 3, Michael Feig 2, Carolyn M. Teschke 1, Andrei Alexandrescu 1, and Kristin N. Parent 2

Affiliations
1 University of Connecticut, Department of Molecular and Cell Biology, Storrs, CT 06269; USA
2 Michigan State University, Department of Biochemistry and Molecular Biology, East Lansing MI, 48824; USA
3 University of Utah School of Medicine, Division of Microbiology and Immunology, Department of Pathology, Salt Lake City, 84112; USA

Running title: Structural biology of capsid binding by Dec

* To whom correspondence should be addressed
Kristin N. Parent
Dept. of Biochemistry & Molecular Biology
Michigan State University
603 Wilson Rd
519 Biochemistry Building
East Lansing MI, 48824

Key Words: bacteriophage, cryo-electron microscopy, cementing protein, capsid assembly, computational modeling, NMR
Abbreviations: cryo-EM; cryo-electron microscopy, 3DR; three-dimensional reconstruction, MD; molecular dynamics, NMR; nuclear magnetic resonance.

Abstract

The major coat proteins of most dsDNA tailed phages and viruses form capsids by a mechanism that includes active packaging of the dsDNA genome into a precursor procapsid, and subsequent expansion/stabilization of the capsid. Packaging of dsDNA genomes accrues ~10 - 60 atm of internal pressure. Phages and viruses have evolved diverse strategies to strengthen their capsids, such as non-covalent binding of auxiliary “decoration” proteins. The P22-like phages are paradigms for understanding the biophysics of capsid assembly and maturation. For example, phage L, Dec protein that has a highly unusual binding strategy where precisely distinguishes between nearly identical regions of the capsid. We combine cryo-electron microscopy and three-dimensional image reconstruction at near atomic resolution, with molecular dynamics techniques to discern the structure of native phage L particles. We used NMR to determine the structure/dynamics of Dec in solution. Key regions of coat and Dec that modulate the selectivity of Dec binding are elucidated.

Introduction

Viral icosahedral capsids are formed from multiple copies of a single or a few types of coat proteins that encapsidate the genome. Because of the symmetry of capsids that have more than 60 coat subunits, genetically identical proteins must assemble into capsid sites with different quasi-equivalent conformations, such that some coat proteins are in hexameric conformations and some are in pentameric conformations [1]. The resulting capsid shells often have additional decoration proteins bound to their exteriors that stabilize the virion against environmental insults. How decoration proteins bind to specific sites on quasi-equivalent capsids is poorly understood.
At least $10^{31}$ viruses are found in the biosphere [2], with double stranded DNA (dsDNA) bacteriophages being the most abundant. The majority of dsDNA containing phage virions have genomes enclosed in a pseudo-icosahedral capsid in which a tail that specifies host binding breaks the icosahedral symmetry. Members of the short-tailed phages (family *Podoviridae*) in the P22-like phage group, are a particularly well-studied model for understanding the biophysical mechanisms of capsid assembly [3] and capsid protein structures at near-atomic level detail [4,5]. The coat proteins of the P22-like phages, like all currently known tailed phages and herpesviruses, share a canonical HK97-like coat protein fold, one of the most ubiquitous protein folds in nature [6,7].

Tight packing of the dsDNA genome into the virion creates enormous internal pressure within the capsid, results in a high-energy state that primes the particles for infection, and facilitates delivery of the majority of the viral genome into the host. Phages have evolved diverse strategies to strengthen their capsids against internal pressures resulting from packaging of their genomes. A well-studied example is the phage HK97, where the coat protein forms covalent crosslinks with neighboring subunits, thereby creating a unique “chainmail” lattice [8]. In some other phages, a separately expressed auxilliary “decoration” protein (sometimes called “cementing” protein) functions as a molecular staple to stabilize mature capsids. The essential phage λ decoration protein gpD stabilizes its capsid by binding at all three-fold symmetry axes [9,10]. Conversely, many related phages do not utilize covalent crosslinks for stability, or auxillary decoration proteins. Instead, these capsids rely entirely on stabilizing interactions between protein subunits. For example, some coat protein P-loops create stabilizing interactions with neighboring capsid subunits around three-fold symmetry axes in the absense of auxiliary proteins [11-15]. The reasons why some phages require stabilization mechanisms such as auxillary proteins, and others do not, despite using highly similar functional capsid building blocks remain poorly understood.
The coat proteins of phage L and phage P22 share 99.6% identity, but only phage L encodes the small auxillary decoration protein, Dec [16]. Phage L and P22 encapsidate similar length genomes and have similar capsid diameters. Therefore, the internal pressures within both capsids are likely to be similar. Dec binds to P22 capsids \textit{in vitro} with nanomolar affinity at the same sites as on phage L virions, and increases the stability of P22 capsids at elevated temperatures when treated with a Mg$^{++}$ chelator \textit{in vitro} [16,17]. Given this, the role(s) of Dec in the life cycle of phage L compared to P22 is not immediately obvious, and, unlike gpD in phage λ, Dec is not essential to withstand genomic packaging forces in phage L virions (E. Gilcrease and S. Casjens, unpublished).

Dec binds only to mature particles, and not precursor procapsids [16] indicating that it recognizes the specific surface topology of mature capsids. P22 and phage L capsids are organized with a T=7 capsid geometry and have 11 vertices formed by coat protein pentons and an additional 60 hexons to create the capsid. The twelfth penton vertex is occupied by the portal channel that connects the phage virion to the tail apparatus. A T=7 organization results in different protein landscapes between neighboring capsomers (Figure 1A). For example, at the center of each of the 20 capsid facets there is a highly symmetric trimeric interaction formed by the three surrounding hexons (called the “true three-fold” axis; black triangle in Figure 1A). Between other hexons, there are “quasi-three-fold” symmetry axes, which have a subtly different arrangement than the true fold symmetry axes (magenta circle in Figure 1A). This quasi-three-fold site is less symmetric, and has bent contacts between capsomers when compared with the true-three-fold symmetry site. Lastly, there is a second type of quasi-three-fold symmetry site that occurs between hexons and the pentons (green triangle in Figure 1A). The tight binding (~nM affinity) of Dec occurs at only quasi-three-fold axes between hexons. Binding to the 20 true three-fold symmetry axes only occurs weakly, with at least an order of magnitude lower affinity [18]. The latter is sufficiently low that these sites are very sparsely occupied by Dec in the native phage L particles. No binding has
been detected at the quasi-three-fold sites surrounding the pentons \cite{17,18}. Dec’s ability to bind discriminately is in sharp contrast with the decoration proteins of other phages that fully saturate all symmetry-related sites in mature virions. Understanding the different mechanisms by which decoration proteins, such as gpD and Dec, can bind to surfaces of viral particles is not only important for understanding the principles of the underlying biology, but is also critical for structure-guided design of virus-inspired nanomaterials \cite{18-23}. Here we investigate how Dec is able to distinguish these subtly different binding sites with high specificity.

We report the capsid structure of the phage L virion, and elucidate the mechanism of the selective binding by Dec. We use a combination of structural approaches including: 1) near atomic resolution cryo-electron microscopy (cryo-EM) and three-dimensional image reconstructions (3DR) of native phage L particles, 2) computational modeling of the Dec trimer when bound to the capsid refined using fully atomistic molecular dynamics (MD) simulations, and 3) nuclear magnetic resonance (NMR) structure and dynamics of Dec in solution. Dec displays significant asymmetry and flexibility when bound to capsids. The N-terminal globular domain of the protein that is in contact with the capsid has an Oligonucleotide/oligosaccharide-Binding (OB-fold; \cite{24}) structure, and the C-terminal portion of Dec is comprised of a putative three-stranded $\beta$-helix domain. Several key residues within both the phage L coat protein lattice and the Dec trimer are predicted to modulate the selectivity of Dec binding. The contributions of these residues towards Dec binding were probed by site-directed mutagenesis. Our findings reveal that Dec has a novel protein fold for decoration proteins, and a binding mechanism that is not shared by other known phages and viruses.

**Results**

**Cryo-EM structure of phage L at near-atomic resolution:**
As a first step towards understanding the mechanism by which Dec binds to phage L, we determined a near-atomic level resolution structure of native phage L capsids using cryo-EM (Figure 1A). The resolution of the entire density map was determined to be 4.2 Å (see Supplementary Materials, Figure S1). Both the overall protein structure, as well as the specific contact points for attachment of Dec trimers to the coat protein subunits are clearly discernable near quasi-three-fold symmetry axes between hexons. As anticipated from previous studies [17,18], Dec very sparsely occupied true three-fold symmetry sites in phage L virions, as there was only a hint of very weak density observed at these positions, and no Dec occupancy was observed at quasi-three-fold sites between hexons and pentons. Segmentation of the Dec density from the virion density map for subsequent analysis was performed using UCSF Chimera’s Segger [25] (Figure 1B, C). The native phage L capsid map has been deposited in the EMDB database (accession number EMD-7787) along with the capsid protein asymmetric unit (PDB ID 6D2F) and the Dec model (trimer from molecular dynamic (MD)-based refinement against the cryo-EM map = PDB ID 6D2D).

**Structures of the phage L coat lattice and capsid-bound Dec trimers:**

As an initial guide for accurately fitting the phage L coat protein into the cryo-EM density, we used the most recently deposited structure of a P22 coat protein asymmetric unit (PDB ID: 5UU5; [26]), since phage L and phage P22 coat proteins are nearly identical, with only four amino acid differences [17]. Upon initial docking, we found deviations between the phage L capsid density and the P22 coat protein asymmetric unit. Therefore, to optimally fit the cryo-EM density we built a model of the phage L coat protein lattice where each capsid protein was allowed to move independently in the cryo-EM density envelope during MD-based refinement instead of assuming perfect icosahedral symmetry (Figure S2 panels A-C). The phage L capsid protein domains are named as defined in [26,27] for P22 coat protein. The amino acid sequence differences between phage L and P22 coat proteins were
accounted for during modeling; the R101H side chain is located in the spine helix pointing towards the capsid interior, I154L is located in the A-domain towards the hexamer center, M267L is in the I-domain, adjacent to but not interacting with Dec, and A276T in also located in the I-domain but on the distal end pointing towards center of the hexamer (Figure S2D). Overall, there were minor differences between the optimally fit P22 and phage L capsid lattices. The phage L coat protein chains deviate from those in the P22 structure (PDB 5UU5) by about 1 Å for the hexamer and 1.5 Å for the penton unit, which is in the range of what may be expected for structures of homologous proteins [28] (see Supplementary Materials, Table S1, for the complete measurements).

At lower resolution, the capsid-bound Dec trimer was previously found to have a tripod shape with three N-terminal legs interacting with the capsid surface and a protuding central C-terminal head or stalk [17,18]. Our results agree with the previously published structure, but the higher resolution map attained in this work allowed us to accurately model the majority of the Dec protein and to identify the specific amino acids in both Dec and coat protein that create the binding interface. The electron densities within the entire virion map were assessed by local resolution analysis [29] (Figure S3). The segmented trimeric Dec densities are asymmetric but well defined at the quasi-three-fold symmetry axes between hexons [18,20] (Figure 1A). The resolution within the Dec density is on average lower than the capsid, likely due to protein flexibility, especially in the C-terminal stalk region (Figure S3).

The Dec protein fold was determined by first using homology modeling (see Materials and Methods for details) followed by a fully atomistic refinement fit to the segmented cryo-EM density using MD-based fitting (Figures 1C, and S4). In brief, the N-terminal domain was modeled using an OB fold (described in more detail below) and the C-terminal domain with a three-stranded β-helix domain. The presence of an OB fold was identified by homology and a template-based model was initially docked into the cryo-EM
density envelope. Studies using NMR also revealed an OB fold from Dec monomers, not bound to the capsid (see NMR section below). The final trimeric Dec structure was refined using MD-based flexible fitting within the cryo-EM density envelope, using the NMR constraints to generate the final Dec structure.

The N-terminal domain of Dec (residues 10-88) was well defined by the cryo-EM density, but residues 89-134 were predicted with a lower confidence owing to the flexibility and lower resolution within that portion of the structure (Figures S3, S4). As we previously observed [18], the N-terminal domain forms the legs of the Dec “tripod” that are bound to the capsid, and the C-terminal region forms its protruding central stalk. As discussed in more detail below, the C-terminal domain appears to be involved in trimerization, while the N-terminal domains have little or no inter-subunit contacts within the trimer.

NMR structure of the Dec monomer has an OB-fold domain and an unfolded C-terminus:

To further investigate the properties of Dec, we characterized the unassembled protein in the absence of capsids by solution NMR. Initial NMR spectra of Dec at neutral pH gave extremely broad lines, with poorer signals than expected for a 43 kDa trimer, which was likely due to aggregation. We were able to achieve reproducibly good samples that gave NMR spectra with sharper lines using a protocol [30] in which Dec is acid-unfolded by lowering the pH to 2, followed by a refolding step induced by adjusting the pH to 4. Subsequent characterization of the protein using size exclusion chromatography, native-gel electrophoresis, and $^{15}$N NMR relaxation measurements (Figure S5A) showed that the unfolding/refolding protocol converts purified Dec from a trimer to partially folded monomers [30]. The refolded Dec monomers were stable for about a week after refolding but exhibited NMR line broadening, indicative of slow folding/oligomerization and aggregation after longer times at pH 4 or higher.
The good quality of the NMR data obtained following the acid-unfolding/refolding protocol, enabled us to obtain nearly complete (98%) NMR assignments for the monomeric Dec protein [30]. The NMR structure was determined using the program ARIA 2.3 [31], based on the experimental NMR constraints summarized in Table 1. The coordinates of the Dec monomer NMR structure ensemble have been deposited to the PDB under accession number 6E3C.

The N-terminus of the Dec monomer has an OB-fold motif (Figure 2). The NMR structure thus determined is close to an initial NMR model calculated with the program CS-ROSETTA [32], based on the assigned NMR chemical shifts in combination with template-based modeling (Figure S5). The backbone (Cα, C', N, O) RMSD is 2.3 Å over the folded parts of the structure (residues 12-89) between the NMR structure and the CS-Rosetta model. Submission of the NMR Dec monomer structure to the DALI server [33] identified about 80 matches to proteins that contain similar OB-fold domain structures. The strongest of these is a portion (residues 117-174) of chain E of the 40S ribosomal protein SA (PDB code 4KZX-E). This superposed on the Dec NMR structure with an Cα RMSD of 3.5 Å over 61 residues, supporting our interpretation that the structure of the Dec N-terminus is in the OB-fold family.

The canonical OB-fold motif [24] is a five-stranded Greek Key β-barrel, comprised of a β1-β3 meander and a β4-β5 hairpin, with an α-helix ‘αOB’ intervening between strands β3 and β4 (Figure 2A). The five-stranded OB-fold β-barrel is closed by an antiparallel pairing between strand β1 and β4, and a short parallel pairing between stands β3 and β5 [24]. In the Dec structure the anti-parallel pairing of stands β1 and β4 is conserved, but strands β3 and β5 are too distant (~ 13 Å) for any H-bond contacts. Thus, the Dec β-sheet is distorted to a more open structure compared with the classical five-stranded OB-fold β-barrel (Figure 2B). In general, the cores of OB-fold β-barrels are consolidated by three layers of hydrophobic
residues [24,34]. This arrangement is also present in the Dec structure (not shown). The canonical role of the helix $\alpha$OB is to provide a ‘hydrophobic plug’ for the bottom hydrophobic layer of the $\beta$-barrel [24,34]. Residue V55 appears to serve this role in Dec. In OB-fold proteins the orientation of the helix $\alpha$OB is more variable than that of the $\beta$-barrel [35,36], and in Dec, the $\alpha$OB helix extends almost directly between strand $\beta$3 and $\beta$5, with the helix axis in the plane of the $\beta$1-$\beta$3 meander, rather than below this structure (Figure 3B).

As with many OB-fold proteins [36], Dec has additional non-conserved elements of secondary structure at the periphery of the conserved motif. The short strand $\beta$0 forms an anti-parallel pair with the N-terminus of the OB-fold strand $\beta$1, and the helix $\alpha$C extends away from the last OB-fold strand $\beta$5 (Figure 2C). Outside of the folded globular part of the Dec monomer structure, residues 1-11 and 90-134 are disordered (Figure 2B). Although these segments were included in the NMR structure calculations, they have no defined structure in the Dec monomers (Figure 2B). The NMR structure agrees well with the model of trimeric Dec derived from MD-based flexible fitting within the cryo-EM density, with a range of RMSDs of 2.0-2.6 Å for the folded parts of the protein (residues 12-86), and 1.4-1.7 Å for the OB-fold portion of the structure (residues 20-77) (Figure 3, movie S1).

To further characterize the dynamics of Dec, we collected NMR $^{15}$N-relaxation data (Figure S5) and analyzed the dynamics of the monomeric protein in terms of the “Model-Free” formalism. The local backbone mobility of the Dec monomers is summarized in Figure 4. The N-terminus from residues 1-11 and the C-terminus from residues 90-134 have small $S^2$ order parameters characteristic of unfolded protein segments (Figure 4A). The random coil chemical shifts of these regions [30] and the lack of long-range NOEs, are also consistent with these regions being unstructured in the Dec monomers. Small and intermediate $S^2$ values indicative of increased flexibility on the ps-ns timescales are also seen at sites within the folded globular portion of Dec (Figure 4A). Information on motions
within slower µs-ms timescales can be garnered from exchange contributions to R2 transverse relaxation (R2\textsubscript{EX} parameters). The sites with the largest R2\textsubscript{EX} contributions are similar to those with low S2 order parameters, in particular L2-3, L4-5, and the region surrounding the loop L3-\(\alpha\) (purple and red in Figure 4B). The sites within the folded portion of the Dec monomers that have the largest flexibility based on the S\(^2\) and R2\textsubscript{EX} are shown in Figure 4 C,D, respectively.

**Mutagenesis of P22 coat protein and Dec reveals key residues that modulate binding:**

The cryo-EM map of the native phage L virion allowed us to fit models of both Dec and coat protein to visualize the binding interface. The resolution of the map in this region was sufficient to fit some bulky side chains to the coat and Dec proteins. Analysis of the binding interface revealed several residues in close contact between the phage L coat protein and the Dec trimer that appear to be important for binding. Previous work suggested that Dec binding to P22 capsids is driven primarily through electrostatic interactions [20], and our data suggests a highly charged binding interface (Figure 5).

To assess the role that specific coat protein surface residues play in modulating Dec binding, the following five coat protein residues in potential close contact with Dec were chosen for single amino acid substitution by site directed mutagenesis: E81R, P82S, R299E, P322S, and E323R (Figure 5; Supplementary Materials, movie S3). Site directed mutagenesis was performed on a plasmid that expresses P22 coat protein (see Materials and Methods) as described previously [37-39]. Cells carrying the plasmid were infected with P22 phages having an *amber* mutation in gene 5 (encodes coat protein), so that any phages produced are the result of complementation from the coat gene expressed from the plasmid. None of the amino acid substitutions affected virion production; phages assembled with each coat protein variant were grown to high titer and and were phenotypically indistinguishable from the parental phage.
We noted that coat proteins occupying different local conformations (as a result of the quasi-equivalent capsid lattice) contribute different residues to the binding interface (Figure 5A, Movie S3). For example, residues E81, P82, and R299 from the coat protein subunits that form the quasi-three-fold symmetry axes (grey subunits in Figure 5) contact Dec, whereas residues P322 and E323 contact Dec from adjacent and overlapping coat protein subunits (black subunits in Figure 5). While their closeness to Dec clearly indicates that it interacts near these coat residues, there are no obvious salt bridges between the coat protein and Dec molecules based on the MD experiments. To assess the role that specific Dec residues play in capsid binding, we also made five single amino acid substitutions in recombinant phage L Dec protein: K30D, Y31A, Y49E, Y71A, and E73R (see Materials and Methods). Modified Dec proteins were added to WT P22 mature phage particles, complexes were purified using CsCl gradients, and relative amounts of Dec protein bound to virions were quantitated. The global secondary structure of all Dec variants was indistinguishable from the WT protein by circular dichroism. Furthermore, native gel experiments performed as described previously [30] show that all Dec variants assemble as trimmers in solution. Therefore any changes in Dec occupancy in the variants are likely due to a disruption of the binding interface, rather than protein folding artifacts.

Among the five coat protein changes, coat amino acid substitutions E81R and E323R attenuated but did not completely abolish Dec binding ability (Figure 5 B,C). Therefore, we created a variant containing both amino acid substitutions, E81R:E323R. This coat protein variant displayed less Dec binding than each of the individual substitutions; however, residual Dec binding was still detected. Our data suggest that since the capsid quasi-three-fold binding sites for Dec are formed by coat proteins in two distinct quasi-equivalent conformations, weaker Dec binding (<20%) can still occur when this interface is disrupted, even by two amino acid substitutions. In addition, amino acid substitutions in Dec K30D and
Y49E displayed lower binding than the WT protein, consistent with the prediction of an electrostatic binding interface (Figure 5D, Movie S3).

The close contact region between Dec and coat can be thought of as including two binding sites. One includes coat residues E81 and P82 along with their closest interacting Dec residues K30, Y31 and Y49 (Figure 5D, “site 1”). For example, Dec residue Y49 is pointed directly towards the side chain of coat residue E81. The other patch includes Dec residues Y71 and E73 and coat residues P322, E323 and R299 (Figure 5D, “site 2”). Overall, variations in site 1 had the largest effect on binding. In Dec, site 1 residues K30 and Y49 had the largest effect on binding saturation, and in coat protein site 1 residue E81R had the largest effect relative to the other substitutions. Although site 1 clearly plays a strong role in modulating Dec binding, coat residue E323R which resides in site 2 had a reproducible, albeit modest effect. We conclude that the first site plays a substantially larger role in mediating Dec binding to coat protein, and is largely due to electrostatic interactions, when compared with the second site.

Not all of the variations tested in coat protein reduced binding, namely P82S, R299E and P322S (we intentionally made conservative proline substitutions as coat protein is highly aggregation prone [4]). Additionally, not all of the variations tested in Dec reduced binding, including Y31A, Y71A, and E73R. These observations are consistent in light of the 3D capsid structure. Coat protein residue P82 is at a lower virion radius than E81 and likely too far from the Dec residue Y49 to affect binding when it is altered. Coat protein change R299E did not alter binding, and this is consistent with the fact that when the closest adjacent Dec residue Y31 is changed to an alanine there was also no effect on binding. Lastly, coat residue P322 forms an interface with two Dec residues, Y71 and E73, and single substitutions at any of these positions have no effect on binding. Even though the cryo-EM density is strong in this region indicating protein:protein contact, electrostatic interactions may not be the driving force at site 2, but instead weaker interactions such as
van der Waals could contribute to binding. Further experiments will be required to fully characterize any subtle changes in binding affinity of the variants that did not obviously decrease binding saturation here.

Given that the binding interface is comprised of coat protein subunits in two distinct quasi-equivalent conformations, other weak protein-protein interactions may help promote binding at site 2, which was not dramatically affected by our mutagenesis. Consistent with this result, mutational analysis of the gpD decoration protein in phage \( \lambda \) shows that “numerous and redundant non-covalent binding interactions” are important for the binding and capsid stabilization mechanism, and that changes in binding are not solely based on single, critical residues [40]. Though gpD and Dec have different folds (see Discussion), perhaps multiple, redundant binding interactions are a common theme among capsid decoration proteins.

A structural basis for the discriminatory binding of Dec to capsids

We compared the observed interaction of Dec trimers with the capsid at the quasi-three-fold and one computationally placed at the point of very weak Dec density at the true three-fold symmetry axis using MD-based simulations (Figure 6). Overall, the interactions with Dec at the different capsid sites are very similar, but Dec modeled at the true three-fold axis is tilted further away relative to the capsid. Moreover, the potential MD-modeled contacts between Dec and coat protein at the true three-fold axis are not as close as the observed contacts at the true three-fold axes (see close-ups in Figure 6C and 6D). This is consistent with the reduced binding affinity of Dec at true three-fold axes and the only very weak density observed in our cryo-EM reconstructions at that site.

Discussion
Decoration proteins employ a variety of binding schemes to adhere to phage capsids:

For dsDNA tailed phages and related viruses, decoration proteins can occupy the capsid lattice in several different positions. Some bind capsids at the center of hexamers including decoration proteins in phages T4, T5, RB49, and Sf13 through Sf19, [41-44]. Alternatively, some decoration proteins bind the edges of hexamers as exemplified by gp17 in phage N4 [45]. Others, bind the capsid in areas between capsomers such as gpD in λ, gp56 in phage TW1, Soc in phage T4, gp87 in phage P74-26, and Dec in phage L [9,17,46-48]. In all of the aforementioned cases, the decoration proteins are added after capsid assembly. Conversely, the herpes virus heterotrimeric triplex protein decorates the outside of the capsid and is also essential during capsid assembly. For the majority of these examples, all possible quasi-equivalent binding sites on the lattice are occupied by the decoration proteins. By contrast, in phage L, Dec only binds at one type of symmetry axis with high affinity: the quasi-three-fold sites between hexamers (Figure 1A).

To function in a stabilizing capacity, decoration proteins must bind capsids with high enough affinity to remain associated even under harsh environmental conditions. Interestingly, the requirement for binding affinity must be balanced with binding specificity to allow discrimination between binding sites. Indeed there is a wide range of reported K_Ds for various decoration proteins. Phage L’s Dec and T4’s Hoc bind their respective capsids with nM affinity, whereas pb10 binds phage T5 with pM affinity [18,41,49]. Taken together, variations in binding affinities and saturating versus discriminating binding behavior suggests that decoration proteins are capable of recognizing subtle differences in capsid lattices. The reason some decoration proteins discriminate between similar quasi-equivalent sites is unknown, but some positions on the capsid lattice may be more vulnerable to environmental assaults than others, therefore some positions may not require stabilization from a decoration protein.
Dec has a novel architecture:

In dsDNA tailed phages, previously studied decoration proteins for which atomic-resolution structures are known have two basic types of protein folds. For decoration proteins such as phage T5’s pb10 and T4’s hoc, the overall fold is that of an Ig domain [41,44]. A second, more common type of decoration protein, exemplified by phage lambda’s gpD, P74-26’s gp87, phage 21’s SHP, and TW1 gp56, binds capsids at all local three-fold axes between capsomers. These proteins have similar polypeptide folds and form a symmetric trimer with an N-terminal β-tulip domain and an α/β subdomain that binds and stabilizes capsids through hydrophobic interfaces [40,48]. Our work shows that Dec represents a third type of decoration protein fold (Figure 7), the OB-fold, as well as a different capsid-binding mechanism that discriminates quasi three-fold capsid binding sites in favor of true three-fold sites. To our knowledge, Dec represents the first occurrence of an OB-fold structure in a virus decoration/cementing protein.

Crystal structures for gpD [50], gp87 [48], and SHP [51] lack resolution for several N-terminal residues, indicating that these regions are flexible when these decoration proteins are not associated with capsids. Cryo-EM maps of gpD in native phage λ show that the N-terminus becomes highly ordered when gpD is in a bound conformation, and that the N-terminus is the major part of the capsid binding mechanism as the N-terminus forms a stabilizing β-sheet with strands supplied from the capsid protein [9]. Like gpD, the N-terminus of the Dec monomer is disordered in solution as shown here by NMR. However, unlike gpD, our data indicates that the Dec N-terminus remains rather flexible in the capsid-bound form since there is (1) no cryo-EM density that is attributable to the first 10 residues, (2) our previous cryo-EM data show the N-terminus of Dec can be labeled with large cargo such as nanogold beads that are rather flexibly bound to the capsid [18], and (3) deletion of
the first 11 residues of Dec does not affect capsid binding affinity [20]. When we compare
the phage L Dec binding motif to the capsid stabilization mechanism in HK97, we see that
the residues that control Dec binding occupy similar spatial positions as the HK97 catalytic
residues that crosslink to form the lattice “chainmail”. Furthermore, comparison of Dec to the
β-tulip family decoration proteins highlights some interesting implications about binding
affinity and selectivity of binding locations in terms of symmetry of the auxiliary proteins
involved. Perfect trimeric symmetry seems to allow for a broader binding specificity. For
example, highly symmetric trimers in gpD and gp87 allow these decoration proteins to
occupy all quasi-equivalent binding sites including true three fold symmetry axes and also
both types of quasi-three-fold sutes. By contrast, the inherent asymmetry in the Dec trimer is
an unusual property among decoration proteins and seems to modulate Dec binding
selectivity to only one type of symmetry axis. Dec appears to represent a new evolutionary
lineage of decoration proteins that is distinctly different than those in the β-tulip family and
whose capsid binding is similarly driven mainly by electrostatics in site 1, with some van der
Waals interactions contributing to binding in site 2.

Summary

Phage L Dec protein has an N-terminal OB domain, and a C-terminal β-helix domain.
Capsid stabilization occurs through electrostatic interactions between Dec and coat protein
subunits. Discriminatory binding occurs at the quasi-three-fold symmetry axes between
hexons and not the true-three symmetry sites largely due to surface area of contact with the
capsid protein. Dec has a novel fold and capsid stabilization mechanism compared to other
known decoration proteins.

Materials and Methods
Strains and media: Phage L, its bacterial host, and purification procedures to produce high-titer stocks were previously described [16,17]. Stocks were stored in a 10 mM Tris (pH 7.6) and 10 mM MgCl₂ buffer. LB Miller broth and LB agar (Invitrogen) were used for all experiments.

Purification of P22 phage with various coat proteins: Salmonella enterica serovar Typhimurium strain DB7136 (leuA414am, hisC525am, su₀), expressing P22 coat protein from mutant pMS11 plasmids, were infected with P22 phage carrying amber mutations in gene 5 to stop production of phage-encoded coat protein as described previously [39]. This P22 strain also carried the c1-7 allele to prevent lysogeny. The resulting phages containing amino acid substitutions in coat protein, were purified using standard protocols [16,18].

Site-directed mutagenesis of P22 coat protein and phage L Dec protein: The plasmid pMS11 was mutated using Quickchange protocols as described previously [52] to generate plasmids that express coat protein containing the following substitutions: E81R, P82S, R299E, P322S, E323R, E81R:E323R. All plasmids were AmpR (100 µg/mL) and IPTG inducible (1 mM). Sequences were confirmed at the RSTF Genomics Core at Michigan State University. Plasmid pDec was mutated using Quickchange protocols, and was used to express wild type and variant Dec, each with a C-terminal histidine tag. Phage P22 purification [16,18] and binding assays were performed as described [17,18]. To measure binding, purified Dec was added to P22 phages made with WT or variant coat proteins. Free Dec was separated from Dec bound to P22 particles by cesium chloride density gradient sedimentation as previously described [16,17]. Bands containing P22 were TCA-precipitated, and analyzed by SDS-PAGE. Coomassie-stained gel bands corresponding to P22 coat protein and Dec were quantitated and normalized to the ratio found in native P22 particles.
when bound with Dec. All binding experiments were repeated two to four times. Representative data are shown.

**Cryo-EM:** Aliquots (~5 µL) of phage L virions were vitrified and examined using published procedures [53]. Briefly, this involved applying samples to Quantifoil R2/2 holey grids that had been plasma cleaned for ~20 sec in a Fischione model 1020 plasma cleaner. Grids were then blotted with Whatman filter paper for ~5 sec, and plunged into liquid ethane. Samples were pre-screened for concentration and purification quality in a JEOL JEM-2200FS TEM using a Gatan 914 specimen holder, using standard low-dose conditions as controlled by SerialEM [54]. High-resolution imaging on an FEI Titan Krios was performed at Florida State University. Micrographs were recorded on a Direct Electron DE-20 camera with a capture rate of 25 frames per second using a total of 53 frames, and a final dose of ~27 e/Å² at a final pixel size of 1.26 Å. Movie correction was performed using the Direct Electron software package, v2.7.1 [55] on entire frames.

**3D image reconstructions of icosahedral particles:** Micrographs exhibiting minimal astigmatism were selected for further processing. The objective lens defocus settings used to record each data set ranged 0.35 to 2.49 µm. In total, the final reconstruction used 7,879 of the best particles from 494 images. The program RobEM [http://cryoEM.ucsd.edu/programs.shtml](http://cryoEM.ucsd.edu/programs.shtml) was used to estimate micrograph defocus and astigmatism, extract individual phage L particles, and to preprocess the images. 150 particle images were used as input to the random-model computation procedure to generate an initial 3D density map at ~25 Å resolution [56]. Each map was then used to initiate determination and refinement of particle orientations and origins for the complete set of images using the current version of AUTO3DEM (v4.01.07) [57]. Phases of the particle structure factor data were corrected to compensate for the effects caused by the microscope
contrast-transfer function, as described [58]. The Fourier Shell Correlation (FSC0.5) “gold standard” criterion was used to estimate the resolution of the final 3D reconstructions [59]. Graphical representations were generated using the UCSF Chimera visualization software package [60]. Local resolution was estimated as previously described [29]. Map segmentation was performed using the Segger tool in Chimera [60].

**Molecular dynamics-based flexible fitting of capsid-bound Dec:** Initial structures of the capsid proteins except Dec were built via homology modeling based on the available structure for P22 (PDB ID 5UU5 [26]). Using the symmetry operations in the PDB structure, 1/8th of the capsid was constructed to cover the entire cryo-EM density map for phage L. Even if only part of a capsid protein copy was needed to cover the map, the entire protein was retained so that some capsid proteins extended beyond the volume covered by the density map. To obtain an initial structure, the entire capsid complex was fitted via rigid-body docking to the density map using UCSF Chimera [60].

The initial model for Dec was assembled in multiple stages. Secondary structure prediction for the Dec sequence suggests a predominantly β-sheet structure with some α-helical content. A search with PSI-BLAST [61] for known structural homologs of the Dec sequence did not result in any hits with sufficiently reliable E-values (<0.1). NMR chemical shift data for Dec also indicate a fold consisting mostly of β-sheets and two α-helices in the N-terminal part of Dec [30]. Based on the NMR chemical shifts, eleven initial models were built via CS-ROSETTA [32]. All of the resulting models feature an OB-fold, consistent with the predicted secondary structure and resulted in three-dimensional structures that fit approximately into the cryo-EM densities for Dec with good matches to the helical and strand regions that are clearly identifiable in the EM density map. The CS-ROSETTA [32] model that initially fit the EM density best was selected for further modeling. Initial trimer
models guided by the EM densities were constructed covering residues 11 to 88 and subsequently fit to the segmented EM density for Dec only via energy minimization (over 1,000 steps) and short molecular dynamics (2,000 steps at 100K, 5,000 steps at 200K, and 5,000K steps at 300K) followed by minimization again (1,000 steps) using the EMAP facility [62] in CHARMM (version 42a2) [63]. A distance dependent dielectric was applied to approximate an aqueous solvent environment. During the minimization, distinct structural elements (28 - 62, 12 - 26 and 63 - 77, and 79 - 86) were restrained separately so that they could move relative to each other in order to improve the fit to the EM density. All of the restraints were applied just to Cα coordinates allowing the rest of the backbone and side chains to move freely during the optimization.

After initial optimization of the N-terminal part of Dec, the N-terminus was extended to residue 1 using an extended chain, a linker was added to cover residues 89-92, and a homology model for the C-terminus based on a fragment from the bacteriophage T4 proximal long tail fibre protein gp34 (PDB ID: 4UXE [64]) was added. The EM density lacks sufficient resolution for reliably modeling the C-terminal part of Dec in terms of side chain orientations but indicates a barrel-shaped structure consistent with a β-helix seen in many bacteriophage tail structures. The modeled β-helix is also consistent with the secondary structure predicted for the C-terminus of Dec from its sequence [17,18,30]. The selection of this structure as a template was based on the dimension of the β-helix that was narrower than most β-helices found in other tail structures, and in better agreement with the EM density. The complete model of the Dec trimer was then re-optimized against the EM density with the same protocol as before, allowing the C-terminal part of the trimer to move separately from the N-terminus to find the best fit to the EM density. We note, that while the model for the N-terminal part of Dec is supported by NMR data and a good fit to the relatively high resolution EM density, the model for the C-terminus is speculative due to a lack of high-resolution experimental data.
The initial Dec model was combined with six capsid proteins at the quasi-three fold symmetry axis, and re-optimized via energy minimization to improve the fit at the interface between Dec and the capsid proteins. The optimized Dec model was further then combined with a larger capsid complex covering 1/8th of the capsid. A copy of Dec was placed at each quasi-three fold symmetry axis and an additional copy was placed at the true three-fold axis as there was also faint EM density at the latter sites. The combined capsid-Dec complex was finally subjected to extensive optimization against the EM density via molecular dynamics based flexible fitting using NAMD (version 2.10b2) via the MDFF module [65]. NAMD was chosen at this stage for better parallel computing performance [66] since the model of 1/8th of the capsid in full atomistic detail is quite large (>500,000 atoms). In the refinement against the entire EM density, each capsid protein was restrained to its internal conformation but allowed to move relative to other proteins. In addition, backbone torsions and hydrogen bonds were restrained to maintain secondary structure elements. VMD [67] was used to automatically setup the MDFF input. A simple constant dielectric environment (ε=80) was used and after minimization over 1000 steps, molecular dynamics simulations were carried out at 300K over 50,000 steps with a 1 fs integration time step. After a first round of refinement, symmetry copies of Dec and the capsid proteins were averaged, and a second round of refinement was carried out for the 1/8th capsid that started from the averaged capsid and Dec units but allowing each unit to again move independently.

NMR characterization of the structure and dynamics of monomeric Dec: Samples of recombinant Dec enriched in $^{15}$N, $^{13}$C and $^2$H isotopes for NMR studies were expressed in E. coli and purified as described [30]. To obtain samples suitable for NMR, 0.3 to 0.5 mM Dec was unfolded to pH 2 for 20 min, followed by refolding to pH 4.0 in 20 mM sodium acetate buffer containing 50 mM NaCl and 1 mM EDTA. The acid-unfolding/refolding procedure converted Dec to a monomer as monitored by native gel electrophoresis and size exclusion.
chromatography. All NMR data were collected for samples held at a temperature of 33 °C. Virtually complete NMR assignments for monomeric Dec (> 98% of backbone resonances) were obtained using a suite of 3D NMR experiments [30] and have been deposited in the Biological Magnetic Resonance Bank (http://www.bmrb.wisc.edu/) with the accession number 27435.

NMR structures for Dec were calculated with the program ARIA v. 2.3.1 [31] using the experimental restraints summarized in Table 1. Backbone (ϕ, ψ) and sidechain (χ1) torsion angles were calculated from assigned HN, N, Hα, Cα, Cβ, and CO NMR chemical shifts using the program TALOS-N [32]. NOE-based distance restraints were obtained from 3D 15N- and 13C-NOESY experiments collected on 15N/13C-labeled Dec samples, with or without 50% fractional deuteration. The NOESY experiments were collected on an 800 MHz spectrometer equipped with a cryogenic probe, using a mixing time of 100 ms. Hydrogen bond restraints were included based on H-bond donors and acceptors identified in a long-range HNCO experiment [68] recorded on a 2H, 13C, 15N triple-labeled Dec sample in TROSY mode at 800 MHz, and H-bond donors inferred from amide proton protection in 1H to 2H hydrogen exchange experiments. Water refinement in the program ARIA [31] was used as a final optimization step for the NMR structures.

Backbone dynamics of Dec were characterized using 15N NMR R1, R2, and 1H-15N NOE relaxation data recorded at 800 MHz. R1 rates were obtained using interleaved relaxation delays of 0.05, 0.13, 0.21, 0.49, 0.57, 0.71 and 0.99 s. R2 rates were determined using interleaved relaxation delays of 0.01, 0.03, 0.05, 0.07, 0.09, 0.11, and 0.15 s. A 2 s pre-acquisition delay was used for recovery to thermal equilibrium. 1H-15N NOE values were determined from the ratio of crosspeak intensities in a spectrum for which the proton signals were saturated for 2.5 s and a control spectrum in which the saturation period was replaced by a pre-acquisition delay of equivalent length. The processing and analysis of relaxation parameters was done according to published protocols [35]. Model-free analyses [69] of the
15N relaxation data were performed with the program Tensor2 [70], yielding an optimal global isotropic rotational correlation time of 7.4 ns.

Acknowledgements: We thank Timothy S. Baker (University of California, San Diego), Gabriel C. Lander and John E. Johnson (The Scripps Research Institute) for advice and support during the preliminary phase of this project, Giovanni Cardone for support in the local resolution analysis, Prof. Angela Gronenborn (U. Pittsburgh School of Medicine) for useful discussion, and Anne R. Kaplan for help with NMR structure calculations of Dec. This material is based upon work supported by the AAAS Marion Milligan Mason Award for Women in the Chemical Sciences to KNP, by grants NIH GM084953 and GM126948 to MF, NIH grant R01 GM076661 and a grant from the UConn Research Excellence Program to CMT and ATA. High resolution cryo-electron microscopy data were collected at Florida State University, and the FSU facility is supported by the following grants: S10 OD018142 and S10 RR025080 under PI Ken Taylor.

Competing Interests

The authors have no competing interests to declare.

References:

1. Caspar DLD, Klug A (1962) Physical principles in the construction of regular viruses. Cold Spring Harbor Symp Quant Biol 27: 1-24.
2. Bamford DH, Grimes JM, Stuart DI (2005) What does structure tell us about virus evolution? Curr Opin Struct Biol 15: 655-663.
3. Parent KN, Zlotnick A, Teschke CM (2006) Quantitative analysis of multi-component spherical virus assembly: Scaffolding protein contributes to the global stability of phage P22 procapsids. J Mol Biol 359: 1097-1106.
4. Teschke CM, Parent KN (2010) 'Let the phage do the work': Using the phage P22 coat protein structure as a framework to understand its folding and assembly mutants. Virology 401: 119-130.
5. Parent KN, Schrad JR, Cingolani G (2018) Breaking Symmetry in Viral Icosahedral Capsids as Seen through the Lenses of X-ray Crystallography and Cryo-Electron Microscopy. Viruses 10.

6. Hendrix RW (2002) Bacteriophages: evolution of the majority. Theor Popul Biol 61: 471-480.

7. Baker ML, Jiang W, Rixon FJ, Chiu W (2005) Common ancestry of herpesviruses and tailed DNA bacteriophages. J Virol 79: 14967-14970.

8. Popa MP, McKelvey TA, Hempel J, Hendrix RW (1991) Bacteriophage HK97 structure: wholesale covalent cross-linking between the major head shell subunits. J Virol 65: 3227-3237.

9. Lander GC, Evilevitch A, Jeemaeva M, Potter CS, Carragher B, Johnson, JE (2008) Bacteriophage lambda stabilization by auxiliary protein gpD: timing location, and mechanism of attachment determined by cryo-EM. Structure 16: 1399-1406.

10. Yang Q, Maluf NK, Catalano CE (2008) Packaging of a unit-length viral genome: the role of nucleotides and the gpD decoration protein in stable nucleocapsid assembly in bacteriophage lambda. J Mol Biol 383: 1037-1048.

11. Parent KN, Khayat R, Tu LH, Suhanovsky MM, Cortines JR, Teschke, CM, Johnson, JE Baker TS. (2010) P22 coat protein structures reveal a novel mechanism for capsid maturation: Stability without auxiliary proteins or chemical crosslinks. Structure 18: 390-410.

12. Chen DH, Baker ML, Hryc CF, DiMaio F, Jakana J, Wu W, Dougherty, M, Haase-Pettingell, C, Schmid, MF, Jiang, W, Baker, D, King, JA, Chiu, W. (2011) Structural basis for scaffolding-mediated assembly and maturation of a dsDNA virus. Proc Natl Acad Sci U S A 108: 1355-1360.

13. Spilman MS, Dearborn AD, Chang JR, Damle PK, Christie GE, Dokland, T. (2011) A conformational switch involved in maturation of Staphylococcus aureus bacteriophage 80a capsids. Journal of Molecular Biology 405: 863-876.

14. Parent KN, Gilcrease EB, Casjens SR, Baker TS (2012) Structural evolution of the P22-like phages: Comparison of Sf6 and P22 procapsid and virion architectures. Virology 427: 177-188.

15. Parent KN, Gilcrease EB, Casjens SR, Baker TS (2012) Structural evolution of the P22-like phages: Comparison of Sf6 and P22 procapsid and virion architectures. Virology 427: 177-188.

16. Gilcrease EB, Winn-Stapley DA, Hewitt FC, Joss L, Casjens SR (2005) Nucleotide sequence of the head assembly gene cluster of bacteriophage L and decoration protein characterization. J Bacteriol 187: 2050-2057.

17. Tang L, Gilcrease EB, Casjens SR, Johnson JE (2006) Highly discriminatory binding of capsid cementing protiens in bacteriophage L. Structure 14: 837-845.

18. Parent KN, Deedas CT, Egelman EH, Casjens SR, Baker TS, Teschke, CM. (2012) Stepwise molecular display utilizing icosahedral and helical complexes of phage coat and decoration proteins in the development of robust nanoscale display vehicles. Biomaterials 33: 5628-5637.

19. Douglas T, Young M (2006) Viruses: making friends with old foes. Science 312: 873-875.

20. Schwarz B, Madden P, Avera J, Gordon B, Larson K, Miettinen, HM, Uchida, M, LaFrance, B, Basu, G, Rynda-Apple, A, Douglas, T. (2015) Symmetry Controlled, Genetic Presentation of Bioactive Proteins on the P22 Virus-like Particle Using an External Decoration Protein. ACS Nano 9: 9134-9147.

21. McCoy K, Uchida M, Lee B, Douglas T (2018) Templated Assembly of a Functional Ordered Protein Macromolecular Framework from P22 Virus-like Particles. ACS Nano 12: 3541-3550.
22. Sharma J, Uchida M, Miettinen HM, Douglas T (2017) Modular interior loading and exterior decoration of a virus-like particle. Nanoscale 9: 10420-10430.

23. Catalano CE (2018) Bacteriophage lambda: The path from biology to theranostic agent. Wiley Interdiscip Rev Nanomed Nanobiotechnol.

24. Murzin AG (1993) OB(oligonucleotide/oligosaccharide binding)-fold: common structural and functional solution for non-homologous sequences. EMBO J 12: 861-867.

25. Pintilie GD, Zhang J, Goddard TD, Chiu W, Gossard DC (2010) Quantitative analysis of cryo-EM density map segmentation by watershed and scale-space filtering, and fitting of structures by alignment to regions. Journal of structural biology 170: 427-438.

26. Hryc CF, Chen DH, Afonine PV, Jakana J, Wang Z, Haase-Pettingell, C, Jiang, W, Adams, PD, King, JA, Schmid, MF, Chiu, W. (2017) Accurate model annotation of a near-atomic resolution cryo-EM map. Proc Natl Acad Sci U S A 114: 3103-3108.

27. Rizzo AA, Suhanovsky MM, Baker ML, Fraser LC, Jones LM, Rempel, DL, Gross, ML, Chiu, W, Alexandrescu, A, Teschke, CM. (2014) Multiple functional roles of the accessory I-domain of bacteriophage P22 coat protein revealed by NMR structure and CryoEM modeling. Structure 22: 830-841.

28. Eyrich VA, Marti-Renom MA, Przybylski D, Madhusudhan MS, Fiser A, Pazos, F, Valencia, A, Sali, A, Rost, B. (2001) EVA: continuous automatic evaluation of protein structure prediction servers. Bioinformatics 17: 1242-1243.

29. Cardone G, Heymann JB, Steven AC (2013) One number does not fit all: mapping local variations in resolution in cryo-EM reconstructions. J Struct Biol 184: 226-236.

30. Newcomer RL, Belato HB, Teschke CM, Alexandrescu AT (2018) NMR assignments of the phage-cementing protein Decorator. Biomol NMR Assign.

31. Bardiaux B, Malliavin T, Nilges M (2012) ARIA for solution and solid-state NMR. Methods Mol Biol 831: 453-483.

32. Shen Y, Vernon R, Baker D, Bax A (2009) De novo protein structure generation from incomplete chemical shift assignments. J Biomol NMR 43: 63-78.

33. Holm L, Rosenstrom P (2010) Dali server: conservation mapping in 3D. Nucleic Acids Res 38: W545-W549.

34. Alexandrescu AT, Gittis AG, Abeygunawardana C, Shortle D (1995) NMR structure of a stable "OB-fold" sub-domain isolated from staphylococcal nuclease. J Mol Biol 250: 134-143.

35. Alexandrescu AT, Shortle D (1994) Backbone dynamics of a highly disordered residue fragment of staphylococcal nuclease. J Mol Biol 242: 527-546.

36. Guardino KM, Sheftic SR, Slattery RE, Alexandrescu AT (2009) Relative stabilities of conserved and non-conserved structures in the OB-fold superfamily. Int J Mol Sci 10: 2412-2430.

37. Parent KN, Suhanovsky MM, Teschke CM (2007) Phage P22 procapsids equilibrate with free protein subunits. J Mol Biol 365: 513-522.

38. Suhanovsky MM, Parent KN, Dunn SE, Baker TS, Teschke CM (2010) Determinants of bacteriophage P22 polyhead formation: the role of coat protein flexibility in conformational switching. Mol Micro 77: 1568-1582.

39. Suhanovsky MM, Teschke CM (2013) An intramolecular chaperone inserted in bacteriophage P22 coat protein mediates its chaperonin-independent folding. J Biol Chem 288: 33772-33783.

40. Lambert S, Yang Q, De Angeles R, Chang JR, Ortega M, Davis, C, Catalano, CE. (2017) Molecular Dissection of the Forces Responsible for Viral Capsid Assembly and Stabilization by Decoration Proteins. Biochemistry 56: 767-778.

41. Vernhes E, Renouard M, Gilquin B, Cuniasse P, Durand D, England, P, Hoos, S, Huet, A, Conway, JF, Glukhov, A, Ksenzenko, V, Jacquet, E, Nhiri, N, Zinn-Justin, S.
Boulanger, P. (2017) High affinity anchoring of the decoration protein pb10 onto the bacteriophage T5 capsid. Sci Rep 7: 41662.

Doore SM, Schrad JR, Dean WF, Dover JA, Parent KN (2018) Shigella Phages Isolated during a Dysentery Outbreak Reveal Uncommon Structures and Broad Species Diversity. J Virol 92.

Sathaliyawala T, Islam MZ, Li Q, Fokine A, Rossmann MG, Rao, VB. (2010) Functional analysis of the highly antigenic outer capsid protein, Hoc, a virus decoration protein from T4-like bacteriophages. Mol Microbiol 77: 444-455.

Fokine A, Islam MZ, Zhang Z, Bowman VD, Rao VB, Rossmann, MG. (2011) Structure of the three N-terminal immunoglobulin domains of the highly immunogenic outer capsid protein from a T4-like bacteriophage. J Virol 85: 8141-8148.

Choi KH, McPartland J, Kaganman I, Bowman VD, Rothman-Denes LB, Rossmann, MG. (2008) Insight into DNA and protein transport in double-stranded DNA viruses: the structure of bacteriophage N4. J Mol Biol 378: 726-736.

Wang Z, Hardies SC, Fokine A, Klose T, Jiang W, Cho, BC, Rossmann, MG. (2018) Structure of the Marine Siphovirus TW1: Evolution of Capsid-Stabilizing Proteins and Tail Spikes. Structure 26: 238-248 e233.

Qin L, Fokine A, O'Donnell E, Rao VB, Rossmann MG (2010) Structure of the small outer capsid protein, Soc: a clamp for stabilizing capsids of T4-like phages. J Mol Biol 395: 728-741.

Stone NP, Hilbert BJ, Hidalgo D, Halloran KT, Lee J, Sontheimer, EJ, Kelch, BA. (2018) A Hyperthermophilic Phage Decoration Protein Suggests Common Evolutionary Origin with Herpesvirus Triplex Proteins and an Anti-CRISPR Protein. Structure 26: 936-947 e933.

Shivachandra SB, Rao M, Janosi L, Sathaliyawala T, Matyas GR, Alving, CR, Leppla, SH, Rao, VB. (2006) 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 345: 190-198.

Yang F, Forrer P, Dauter Z, Conway JF, Cheng N, Cerritelli, ME, Steven, AC, Pluckthun, A, Wlodawer, A. (2000) Novel fold and capsid-binding properties of the lambda-phage display platform protein gpD. Nat Struct Biol 7: 230-237.

Forrer P, Chang C, Ott D, Wlodawer A, Pluckthun A (2004) Kinetic stability and crystal structure of the viral capsid protein SHP. J Mol Biol 344: 179-193.

D’Lima NG, Teschke CM (2015) A Molecular Staple: D-Loops in the I Domain of Bacteriophage P22 Coat Protein Make Important Intercapsomer Contacts Required for Procapsid Assembly. J Virol 89: 10569-10579.

Baker TS, Olson NH, Fuller SD (1999) Adding the third dimension to virus life cycles: three-dimensional reconstruction of icosahedral viruses from cryo-electron micrographs. [erratum appears in Microbiol Mol Biol Rev 2000 Mar;64(1):237.]. Microbiology & Molecular Biology Reviews 63: 862-922.

Mastronarde DN (2005) Automated electron microscope tomography using robust prediction of specimen movements. J Struct Biol 1: 36-51.

Wang Z, Hryc C, Bammes B, Afonine PV, Jakana J, Chen, DH, Liu, J, Baker, ML, Kao, C, Ludtke, SJ, Schmid, MF, Adams, PD, Chiu, W. (2014) An atomic model of brome mosaic virus using direct electron detection and real-space optimization. Nature Communications 5: 4808.

Yan X, Dryden KA, Tang J, Baker TS (2007) Ab initio random model method facilitates 3D reconstruction of icosahedral particles. J Struct Biol 157: 211-225.

Yan X, Sinkovits RS, Baker TS (2007) AUTO3DEM-an automated and high throughput program for image reconstruction of icosahedral particles. J Struct Biol 157: 73-82.
58. Bowman VD, Chase ES, Franz AW, Chipman PR, Zhang X, Perry, KL, Baker, TS, Smith, TJ. (2002) An antibody to the putative aphid recognition site on cucumber mosaic virus recognizes pentons but not hexons. J Virol 76: 12250-12258.

59. Pintilie G, Chen DH, Haase-Pettingell CA, King JA, Chiu W (2016) Resolution and Probabilistic Models of Components in CryoEM Maps of Mature P22 Bacteriophage. J Struct Biol 157: 281-287.

60. Goddard TD, Huang CC, Ferrin TE (2007) Visualizing density maps with UCSF Chimera. J Struct Biol 157: 281-287.

61. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller, E, Lipman, Dl. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25: 3389-3402.

62. Wu X, Brooks B (2007) Modeling Of Macromolecular Assemblies with Map Objects. BioComp 411-420.

63. Brooks BR, Brooks CL, 3rd, Mackerell AD, Jr., Nilsson L, Petrella RJ, Roux, B, Won, Y, Archontis, G, Bartels, C, Boresch, C, Caflish, A., Caeves, L, Cui, Q, Dinner, AR, Feig, M, Fischer, S, Gao, J, Hodoscek, M, Im, W, Kuczera, K, Lazaridis, T, Ma, J, Ovchinnikov, V, Paci, E, Pastor, RW, Post, CB, Pu, JZ, Schafer, M, Tidor, B, Venable, RM, Woodcock, HL, Wu, X, Yang, W, York, DM, Karpus, M (2009) CHARMM: the biomolecular simulation program. J Comput Chem 30: 1545-1614.

64. Granell M, Namura M, Alvira S, Kanamaru S, van Raaij MJ (2017) Crystal Structure of the Carboxy-Terminal Region of the Bacteriophage T4 Proximal Long Tail Fiber Protein Gp34. Viruses 9.

65. Trabuco LG, Villa E, Mitra K, Frank J, Schulten K (2008) Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. Structure 16: 673-683.

66. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa, E, Chipot, C, Skeel, RD, Kale, L, Schulten, K. (2005) Scalable molecular dynamics with NAMD. J Comput Chem 26: 1781-1802.

67. Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. J Mol Graph 14: 33-38, 27-38.

68. Cordier F, Grzesiek S (1999) Direct observation of hydrogen bonds in proteins by interresidue 3HJNC' scalar couplings. J Am Chem Soc 121: 1601-1602.

69. Lipari G, Szabo, A. (1982) Model-Free Approach to the Interpretation of Nuclear Magnetic Resonance Relaxation in Macromolecules. 1. Theory and Validity. J Am Chem Soc 104: 4546-4559.

70. Dosset P, Hus JC, Blackledge M, Marion D (2000) Efficient analysis of macromolecular rotational diffuson from heteronuclear relaxation data. J Biomol NMR 16: 23-28.

71. Henderson R, Sali A, Baker ML, Carragher B, Devkota B, Downing, KH, Egelman EH, Feng, Z, Frank, J, Grigorieff, N, Jiang, W, Ludtke, SJ, Medalia, O, Penczek, PA, Rosenthal, PB, Rossman, MG, Schmid, MF, Schroeder, GF, Steven, AC, Stokes, DL, Westbrook, JD, Wriggers, W, Yang, H, Young, J, Berman, HM, Chiu, W, Kleywegt, GJ, Lawson, CL. (2012) Outcome of the first electron microscopy validation task force meeting. Structure 20: 205-214.

72. Kumar GS, Clarkson MW, Kunze MBA, Granata D, Wand AJ, Lindorff-Larsen, K, Page, R, Peti, W. (2018) Dynamic activation and regulation of the mitogen-activated protein kinase p38. Proc Natl Acad Sci U S A 115: 4655-4660.
### TABLE 1. Statistics for the top 20 NMR structures of Dec.

| Experimental Restraints |          |
|-------------------------|----------|
| Total number of NMR restraints | 1,009    |
| Total number of NOE distance restraints | 767      |
| Ambiguous               | 73       |
| Unambiguous             | 694      |
| Long range (|i-j|>4)        | 124      |
| Medium range (|i-j|≤4)       | 93       |
| Sequential (|i-j|=1)     | 281      |
| Intra-residue NOEs      | 269      |
| Total number of dihedral restraints | 174      |
| φ/ψ                     | 127      |
| χ1                      | 47       |
| Hydrogen bond restraints (34*2) | 68       |

#### RMSD from experimental restraints

|            |          |
|------------|----------|
| NOE distance (Å) | 0.054 ± 0.003 |
| Dihedral (°)    | 0.42 ± 0.14  |

#### RMSD from ideal geometry

|            |          |
|------------|----------|
| Bonds (Å)  | 0.0042 ± 0.0001 |
| Angles (°) | 0.55 ± 0.02   |
| Improper angles (°) | 1.68 ± 0.12 |
| E_{L-J} (kcal/mol) | -316 ± 44 |

#### RMSD from mean NMR structure

|            | Backbone b | All Heavy Atoms |
|------------|------------|-----------------|
| Entire protein 1-134 (Å) | > 9.5      | > 9.5           |
| Folded regions 12-86 (Å)  | 1.33 ± 0.21| 1.95 ± 0.32     |
| OB-fold 18-77 (Å)          | 1.08 ± 0.18| 1.66 ± 0.25     |
| OB-fold 2° structure c (Å) | 0.84 ± 0.15| 1.57 ± 0.31     |

#### Procheck Ramachandran Plot Statistics d

|          |          |
|----------|----------|
| Most favored (%) | 89.1      |
| Additionally allowed (%) | 10.9     |
| Generously allowed (%)  | 0.0       |
| Disallowed (%)        | 0.0       |

#### Quality Z Scores from PSVS e

|            |          |
|------------|----------|
| Procheck   | -2.64    |
| Molprobity Clash | -1.19    |

---

a Structures had no NOE violations > 0.5 Å nor dihedral violations > 5 degrees.
b Atoms: Cα, N, C, O
c Calculated over residues in the OB-fold portion: 21-32 (β1), 35-40 (β2), 47-50 (β3), 52-59 (βα3), 63-68 (β4), 73-77 (β5), a total of 41 amino acids.
d Calculated with the PSVS server (http://psvs-1.5-dev.nesg.org) using only the folded parts of Dec (residues 12-89).
Figure 1: Cryo-EM imaging and icosahedral image reconstruction of the mature phage L capsid. A) Image reconstruction of phage L overlaid with a black icosahedral cage. The different symmetry axes marked. The capsid is shown in grey and Dec density is shown in cyan. Enlarged view of a quasi-3-fold symmetry axis between hexons showing the B) coat protein and C) Dec. Both panels B and C show the density fitted with homology models optimized by molecular dynamics-based flexible fitting.
Figure 2: NMR structure of the Dec monomer. A) Ensemble of the 20 lowest-energy NMR structures. For clarity, the disordered N- and C-termini are not shown. The folded globular part of Dec has an OB-fold consisting of a β1-β3 meander (sky blue) and a β4-β5 hairpin (yellow), with an α-helix (green) intervening between strands β3 and β4. Additional secondary structure outside of the OB-fold includes the short N-terminal strand β0 (dark blue), and a C-terminal α-helix (pink). B) Cartoon showing the structure closest to the NMR average. The first 10 N-terminal residues and the last ~45 C-terminal residues, which are unstructured in the Dec monomers are colored red and orange, respectively. The coloring scheme for the rest of the protein in this and subsequent panels is the same as in A. C) Diagram summarizing the topology and secondary structure limits of the Dec monomers.
Figure 3: Comparison of cryo-EM and NMR structures. Superposition of the NMR Dec monomer structures onto the cryo-EM Dec trimer structure (maroon). The Dec monomer structures are colored as in Fig. 3.
Figure 4: Dynamics of the Dec monomer. A) $S^2$ order parameters describing the amplitude of fast motions on the ps-ns timescale. Rigid ($S^2 > 0.65$), moderately flexible ($0.65 \geq S^2 > 0.55$), and highly flexible ($0.55 \geq S^2$) sites are indicated by black, purple, and red symbols respectively. The secondary structure of Dec is given at the top of panel A. B) Contributions to R2 relaxation from slow conformational exchange on the µs-ms timescale. Amide protons with moderate $R2_{ex}$ values smaller than 3 Hz are shown in purple, those with large contributions above 3 Hz in red. The $S^2$ and $R2_{ex}$ parameters were obtained from a Model-Free analysis [69] of $^{15}$N R1, R2, and $^1$H-$^{15}$N NOE relaxation data for Dec (Figure S4) using the program TENSOR 2.0 [70]. In C) and D) the $S^2$ and $R2_{ex}$ values are mapped onto the NMR structure ensembles for Dec. Residues 1-10 and 90-134 which are unfolded and thus have the lowest $S^2$ values, are not shown in the structures.
**Figure 5: Binding assays of variant coat proteins and Dec to probe the binding interface.** A) Coat protein subunits depicted as ribbon diagrams around a quasi-3-fold axis Dec binding site. Three coat subunits directly surrounding a quasi-3-fold axis are shown in light grey, and three coat proteins occupying a neighboring quasi-equivalent conformation in the lattice shown in black. Residues selected for mutagenesis are shown as spheres. B) Normalized binding data of the ratio of variant coat protein bound to WT Dec, color-coded to match the corresponding residues in panel A. C) Dec trimer shown as ribbons, with each monomer a different shade of blue. Residues selected for mutagenesis are shown as spheres. D) Enlarged view of the Dec cryo-EM density with a monomer shown as a ribbon. K30 and Y49 residues are shown as spheres and the two different capsid binding sites are circled. E) Normalized binding data of the ratio of WT coat protein bound to variant Dec, and color-coded to match the corresponding residues in panels C and D.
Figure 6: Dec demonstrates discriminatory binding at different types of three-fold symmetry axes. Dec trimer (red/pink) with surrounding capsid proteins (blue/ice blue). Darker colors show the Dec trimer at a quasi three-fold symmetry axis, lighter colors show the Dec trimer at the true three-fold axis. A and B show side and top views of a local region of each Dec-capsid complex. The close-ups C and D show the interface between Dec and one of the capsid proteins. Both complexes (Dec plus capsid) were superimposed using only the Dec coordinates for residues 10-88.
Figure 7: Comparison of decoration proteins. A) Side and B) top-down views of decoration proteins gpD (PDB ID 1C5E), gp87 (PDB ID 6BL5) and Dec (PDB ID 6D2D). Chains of each trimer are rainbow colored from N-terminus (blue) to C-terminus (red).

Supplementary Figures/Tables/Movies Legends:

Table S1: Comparison of individual asymmetric units of the coat lattice with average asymmetric unit and P22 structure

We generated models for nine complete asymmetric units in order to completely fill the cryo-EM density of one eighth of the total capsid (Figure S2 A-C). Figure S2 indicates where each of these asymmetric units are located and corresponds to the data in Table S1. We
carried out the same analysis with respect to the P22 structure based on the PDB structure (5UU5).

Table S2: Comparison of individual Dec trimers with average structure

Figure S1: Phage L cryo-EM and 3D reconstruction data. A) Representative micrograph of frozen hydrated phage L particles. B) FSC curve with FSC\(_{0.143}\) cutoff shown with a dashed line. The red arrow points out the estimated global resolution of the map at 4.2 Å according to the “gold standard” method [71].

Figure S2: Phage L coat protein subunit and asymmetric unit. We generated models for nine complete asymmetric units in order to completely fill the cryo-EM density of one eighth of the total capsid (Figure S2 A-C, Table S1). We carried out the same analysis with respect to the P22 structure based on the P22 PDB structure (5UU5). The average phage L structure deviates from the P22 structure (PDB 5UU5) by about 1 Å for the hexamer and 1.5 Å for the penton unit, which is in the range of what may be expected for structures of homologous proteins [28]. A) Full phage L capsid shown in blue, with the octant of the map used for modeling highlighted. Coat protein segmentation is colored in grey and the Dec segmentation is colored in cyan. Note the map is rotated to peer down a true 3-fold axis in the center of the area that was modeled. B) Protein subunits modeled using molecular dynamics flexible fitting are shown as black ribbon diagrams and extend beyond the cryo-EM density map edges. C) Coat protein asymmetric units (P0X) and Dec trimers (D0X) labeled as listed in Tables S1 & S2. D) Single coat protein monomer shown color-coded according to domain boundaries (N-arm in red, P-domain in green, E-loop in yellow, A-domain in cyan, and I-domain in magenta). E) An asymmetric unit of coat proteins (PDB ID
6D2F), with six subunits from a hexon (A-F; red, orange, yellow, green, blue, and purple) and one subunit from a penton (subunit G, black).

**Figure S3:** Resolution of the phage L map is not uniform. A) Phage L surface rendered view, colored accorded to local resolution with an octant of the virion removed to show the internal genome organization. The color bar indicates resolution in Å.

**Figure S4. Dec trimers structures have a flexible C terminus;** Dec trimer from a top-down view (left) and rotated to a side view (right). Individual strands are colored yellow, green, or rainbow colored to highlight the OB fold. In the N-terminal OB fold region, the resolution of the map is much higher (ranges 3.8 – 4.5 Å) than the β-helix. We were able to reliably fit side chains into the majority of the N-terminal density envelope (residues 10-88). The average R2 and R1 values for the folded portions of Dec are about 13 Hz (Figure S5A) and 1.4 Hz (Figure S5B), respectively. For comparison, a Dec trimer with a MW of 43 KDa is expected to have R2 and R1 values near 50 Hz and 0.4 Hz, respectively [72]. Thus the $^{15}$N relaxation data for Dec samples subjected to our unfolding/refolding protocol are more consistent with a 14.4 KDa Dec monomer than a trimer. Similarly, the Model Free analysis of the $^{15}$N relaxation data gives an optimum global correlation time for isotropic rotational diffusion of 7.4 ns, consistent with a Dec monomer. For a 43 KDa trimer the value of the correlation time should be about 21 ns.

**Figure S5. CS-ROSETTA modeling of the Dec monomer structure.** (A) The lowest energy model based on the assigned NMR chemical shifts of Dec [30] has an OB-fold topology. Only residues 10-88 were included in the CS-ROSETTA simulations [32], since the program predicted that residues 1-9 and 89-134 are disordered, based on the assigned NMR chemical shifts. The 5-stranded, Greek Key β-barrel OB-fold motif [24] is formed from
a β1-β3 meander (light blue) and a β4-β5 hairpin (yellow), with an α-helix (green) intervening between strands β3 and β4. 

(B) Graph showing convergence of the CS-ROSETTA simulations. The statistics indicated that the CS-ROSETTA simulations did not fully converge. In spite of this, nine of the top ten lowest energy structures had an OB-fold structure with an average RMSD of 2.5 Å to the structure shown in (A).

**Figure S6.** $^{15}$N relaxation values for monomeric Dec. (A) R1, (B) R2, (C) $^1$H-$^{15}$N NOE. All data were recorded at 800 MHz, with a sample temperature of 33 °C. The secondary structure of Dec, derived from its NMR structure, is indicated at the top of the first panel.

**Supplementary Movies**

**Movie S1**

Comparison of cryo-EM Dec trimer (grey) overlaid with the top ensemble structures from NMR.

**Movie S2**

View of the segmentation and fitting process for a quasi-three fold axis of the phage L map. Segmented Dec density is shown in cyan and the segmented coat protein density is shown in grey. The Dec trimer backbone is shown as a ribbon diagram with each strand a different shade of blue. The coat protein backbone lattice is shown as a ribbon diagram with the three subunits directly surrounding the quasi-three-fold symmetry axis in grey. The three subunits in black occupy a different position in the quasi-equivalent lattice, but overlap with the grey subunits and contribute to the Dec binding site.
Movie S3. Key residues that modulate Dec binding affinity. View of the binding interactions between Dec and coat protein at a quasi-three-fold symmetry axis. Coat protein and Dec monomers are color coded as in Movie S2. Amino acid sites that were chosen for mutagenesis are shown as spheres and are color coded to match the data in Figure 6 of the main text.

Movie S4. Movie. S4 Definition of different regions of the Dec monomer NMR structure. To illustrate the precision of the structure all 20 members of the NMR ensemble are shown. The movie than flips through the 20 individual NMR structures to show the differences in precision between the structured OB-fold component, and the unstructured N- (yellow) and C-termini (orange). In the second half of the movie, the N- and C-termini are not shown to better illustrate differences in structural definition of the OB-fold component.