Published in final edited form as:

J Struct Biol. 2012 June ; 178(3): 215–224. doi:10.1016/j.jsb.2012.04.002.

Structure and size determination of bacteriophage P2 and P4 procapsids: function of size responsiveness mutations

Altaira D. Dearborn¹, Pasi Laurinmaki², Preethi Chandramouli³, Cynthia M. Rodenburg¹, Sifang Wang^{1,4}, Sarah J. Butcher², and Terje Dokland^{1,†}

¹Department of Microbiology, University of Alabama at Birmingham, Birmingham, AL ²Institute of Biotechnology, University of Helsinki, Finland ³Department of Molecular and Cell Biology, Harvard University, Cambridge, MA ⁴Experimental Therapeutics Centre, Singapore

Abstract

Bacteriophage P4 is dependent on structural proteins supplied by a helper phage, P2, to assemble infectious virions. Bacteriophage P2 normally forms an icosahedral capsid with $T=7$ symmetry from the gpN capsid protein, the gpO scaffolding protein and the gpQ portal protein. In the presence of P4, however, the same structural proteins are assembled into a smaller capsid with $T=4$ symmetry. This size determination is effected by the P4-encoded protein Sid, which forms an external scaffold around the small P4 procapsids. Size responsiveness (*sir*) mutants in gpN fail to assemble small capsids even in the presence of Sid. We have produced large and small procapsids by co-expression of gpN with gpO and Sid, respectively, and applied cryo-electron microscopy and three-dimensional reconstruction methods to visualize these procapsids. gpN has an HK97-like fold and interacts with Sid in an exposed loop where the *sir* mutations are clustered. The $T=7$ lattice of P2 has *dextro* handedness, unlike the *laevo* lattices of other phages with this fold observed so far.

Keywords

molecular piracy; capsid assembly; morphogenesis; cryo-electron microscopy; three-dimensional reconstruction; HK97 fold

1. Introduction

Bacteriophage P2 is a temperate, double-stranded (ds) DNA phage of the *Myoviridae* family that infects *Escherichia coli* and other enterobacteria (Bertani and Bertani, 1970; Bertani and Six, 1988; Nilsson and Haggård-Ljungquist, 2006). P2-like prophages are common in the environment (Breitbart et al., 2002), are present in about 30% of strains in the *E. coli* reference collection (Nilsson et al., 2004), and likely play an important role in horizontal gene transfer in bacterial evolution.

© 2012 Elsevier Inc. All rights reserved.

[†]Corresponding author. Address correspondence to: Terje Dokland Department of Microbiology University of Alabama at Birmingham 845 19th St South, BBRB 311 Birmingham, AL 35294 Tel: (205) 996 4502 Fax: (205) 996 2667 dokland@uab.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

The P2 virion consists of a 60 nm diameter isometric capsid, or head, attached to a 135 nm long contractile tail tipped by a base plate with six tail fibers attached. The capsids consist of 415 copies of capsid protein, gene product (gp) of the *N* gene (gpN) arranged with $T=7$ icosahedral symmetry (Dokland et al., 1992). P2 capsids are assembled as roughly spherical precursors called procapsids from gpN (40.2 kDa), several copies of a gpO scaffolding protein and a dodecameric connector or portal (gpQ) that forms a unique vertex through which the DNA is subsequently packaged (Fig. 1A) (Lengyel et al., 1973; Chang et al., 2008). At some point during the assembly process, gpN is processed by the removal of 31 N-terminal residues, yielding a mature cleavage product called N* (36.7 kDa) (Lengyel et al., 1973; Rishovd and Lindqvist, 1992). Likewise, gpQ is processed to Q* by the removal of 26 N-terminal amino acids, while gpO is cleaved between residues 141 and 142, leaving an N-terminal fragment, O*, that remains inside the mature capsid (Rishovd et al., 1994; Chang et al., 2008). These cleavage events are thought to be carried out by gpO, which contains a serine protease domain and possesses autoproteolytic activity (Wang et al., 2006; Chang et al., 2009). The proteolytic activity of gpO resides in the N-terminal half of the protein, while the C-terminal 90 amino acids comprise a scaffolding domain, which is required and sufficient for promoting capsid assembly (Wang et al., 2006; Chang et al., 2008; Chang et al., 2009).

Circular P2 genomic DNA substrates (33.6 kb) are packaged into these empty procapsids through the action of the terminase proteins gpM and gpP (Bowden and Modrich, 1985), upon which the capsid undergoes major structural transitions that lead to a more angular mature capsid (Lengyel et al., 1973; Marvik et al., 1994a; Chang et al., 2008) (Fig. 1B). The tails, which are assembled via a separate pathway, are attached to the gpQ connector at the unique vertex (Lengyel et al., 1974). A head completion protein, gpL, is added to the final assembly product (Chang et al., 2008).

P4 is an 11.6 kb replicon that can exist either as a plasmid or integrated into the host genome like a prophage (Briani et al., 2001; Deho and Ghisotti, 2006). It is genetically unrelated to P2, except in the cohesive (*cos*) ends. P4 does not normally get mobilized at high frequency and lacks genes encoding major structural proteins. In the presence of P2, however, P4 gets packaged into phage particles using structural proteins supplied by the P2 helper (Six, 1975). The capsid produced in the presence of P4 is smaller (45 nm) than the normal P2 capsid and contains 235 copies of gpN arranged with $T=4$ symmetry (Dokland et al., 1992; Lindqvist et al., 1993). This size difference is effected by the P4-encoded Sid (Size determination) protein (Barrett et al., 1976), which forms an external scaffold surrounding the P4 procapsids (Marvik et al., 1995) (Fig. 1C). P4 *sid* mutants fail to form small capsids; however, P4 DNA can still get packaged as dimers or trimers into large capsids at low efficiency (Shore et al., 1978). Mutants in gpN, called *sir* (sid responsiveness) render the capsid protein unable to form small capsids even in the presence of functional Sid (Six et al., 1991). Furthermore, certain mutations in Sid, named super-*sid* or *nms* (N mutation sensitive), can act as second-site suppressors of the *sir* mutations and enable the formation of small capsids even on a P2 *Nsir* background (Kim et al., 2001). Capsid assembly is completed by removal of Sid, cleavage of gpN, gpO and gpQ, and addition of the Psu decoration protein (Dokland et al., 1992; Rishovd and Lindqvist, 1992; Dokland et al., 1993; Chang et al., 2008) (Fig. 1D–F).

We showed previously that P4-size procapsids can be produced by co-expression of gpN and Sid alone (Dokland et al., 2002). Likewise, P2 procapsids can be produced by co-expression of gpN with a protease-defective version of gpO generated either by truncating the protein at the N-terminus or by mutating one of the three active site residues (Wang et al., 2006; Chang et al., 2009). In this paper, we present three-dimensional reconstructions of P2 and P4 procapsids produced by co-expression of gpN with gpO and Sid, respectively.

The gpN capsid protein has the HK97 fold seen in all other tailed dsDNA phages studied to date (Bamford et al., 2005; Johnson and Chiu, 2007). Unusually, the P2 procapsid has a $T=7$ *dextro* lattice, unlike the $T=7$ *laevo* lattices of all other bacteriophage capsids with the HK97 fold observed to date. gpO is not ordered with icosahedral symmetry and does not appear in the reconstruction. In the P4 procapsids, Sid forms a dodecahedral cage that surrounds the procapsid and interacts with the gpN hexamers. Sid-gpN interactions occur at a loop in the apical domain of gpN where the *sir* mutations are located.

2. Results

2.1. Cryo-EM of P2 and P4 procapsids

P4 procapsids were produced by the co-expression of gpN and Sid, while P2 procapsids were produced by co-expression of gpN and protease-deficient forms of gpO, either with or without the gpQ portal protein (Dokland et al., 2002; Wang et al., 2006; Chang et al., 2008) (Table 1). The procapsids were purified on sucrose gradients and prepared for cryo-EM as previously described (Dokland et al., 2002; Huiskonen et al., 2004; Dokland and Ng, 2006). The gpQ portals were not incorporated into capsids, but remained on top of the sucrose gradients after purification. The resulting shells were indistinguishable from those formed in the absence of gpQ.

P4 procapsids are hollow, somewhat angular shells with an average diameter of 40 nm (Fig. 2A). These particles have a characteristic “hairy” appearance resulting from the external Sid scaffold (Wang et al., 2000; Dokland et al., 2002), and lack an inner core, since they were produced in the absence of internal scaffolding. Very few aberrant particles were seen. P2 procapsids are roughly spherical shells with an average diameter of 54 nm and a shell thickness of about 4–5 nm (Fig. 2B). This appearance is quite distinct from the angular, thin-walled shells typical of expanded capsids (Chang et al., 2008). The procapsids appear to be partially filled, presumably by the gpO internal scaffolding protein. A few aberrant, incomplete or broken shells were also present.

2.2. Reconstruction of the P4 procapsid

Reconstructions of P4 procapsids produced both by in vitro assembly from purified gpN and Sid protein and from the co-expression of gpN and Sid were previously described (Wang et al., 2000; Dokland et al., 2002). The best resolution attained in those studies was 21 Å. The P4 procapsid reconstruction presented here (Fig. 3A) was generated from 2286 particle images and reaches a resolution of 9.5 Å based on the 0.5 Fourier shell correlation (FSC) criterion. At this resolution, several features corresponding to ordered α -helices can be distinguished, forming the basis for protein modeling based on the HK97 gp5 structure, as described below.

In the P4 procapsid, gpN is arranged into twelve pentamers and 30 elongated hexamers that sit on the icosahedral two-fold symmetry axes (Fig. 3A). The $T=4$ icosahedral symmetry dictates that there are four non-equivalent gpN subunits in the shell, denoted A–D, which are organized into an A_5 pentamer and a $(BCD)_2$ twofold symmetric hexamer (Dokland et al., 1992). The capsomers consist of protruding domains clustered around the symmetry axes, and basal domains that interact via two types of similar, trivalent contacts (Fig. 3A): a strict threefold contact at the icosahedral threefold axis (type 1), and a quasi-threefold contact between two hexamers and a pentamer (type 2).

The external scaffold made of Sid forms a dodecahedral cage that makes trivalent connections at the icosahedral threefold axes and interacts with the gpN hexamers at the icosahedral twofolds (Fig. 3A). This interaction involves primarily the B subunits.

2.3. Reconstruction of the P2 procapsid

The P2 procapsid was reconstructed from 7582 particle images to a resolution of 9.9 Å based on the 0.5 FSC criterion. The P2 procapsid reconstruction (Fig. 3B) shows the characteristic $T=7$ structure of capsid protein clustered in twelve protruding pentamers and 60 elongated (“skewed”) hexamers, similar to those seen in other bacteriophage procapsids, including lambda, HK97 and P22 (Dokland and Murialdo, 1993; Conway et al., 1995; Jiang et al., 2003). Unusually, the $T=7$ lattice in P2 is right-handed (*dextro*), rather than the *laevo* lattice found in other bacteriophages. (Although handedness cannot be determined from projection images, the correct hand of the reconstruction can be ascertained by comparing the detailed structure of the subunits to that of bacteriophage HK97—see below.) There are thus seven non-equivalent gpN subunits in the P2 shell, organized into an A₅ pentamer and a BCDEFG hexamer (Dokland et al., 1992). Contacts between capsomers in the P2 lattice are similar to those in P4, except that in addition to the type 1 and type 2 contacts mentioned above, the $T=7$ P2 lattice has an additional trivalent contact (type 3) formed between three hexamers that connect near the twofold symmetry axis (Fig. 3B).

We showed previously that gpO binds to gpN and is present in procapsid shells in about 120–160 copies per procapsid (Wang et al., 2006). P2 virions contain about 70 copies of O* (Geisselsoder et al., 1982). In spite of this, there is no evidence of the gpO internal scaffold in the reconstruction, apart from a diffuse inner density only slightly higher than the background (not shown). Presumably, gpO is not ordered symmetrically in the capsid.

2.4. Similarity of gpN to HK97 gp5

All tailed dsDNA bacteriophage (order *Caudovirales*) structures observed to date share a common structural motif (Bamford et al., 2005; Johnson and Chiu, 2007), referred to as the “HK97 fold” after the first bacteriophage structure solved at high resolution, the gp5 capsid protein of HK97 (Wikoff et al., 2000; Helgstrand et al., 2003; Gertsman et al., 2009). We therefore expected the same motif to be present in gpN. The HK97 fold has a characteristic two-domain structure, with a P domain that includes a long α helix called the “spine helix” and a three-stranded β sheet, and an A domain of mixed α/β structure that represents an insertion into the P domain (Fig. 4A,B). The A and P domains are separated by a flexible “ β hinge” that is a source of conformational diversity in many systems (Gertsman et al., 2009; Teschke and Parent, 2010). Other features include the E loop, a long β hairpin motif that protrudes from the P domain, and an extended N-terminal arm that contains little secondary structure.

In spite of the absence of sequence similarity between gpN and HK97 gp5 (or any other bacteriophage capsid protein), secondary structure prediction revealed a similar distribution of structural elements in gpN and gp5 when the sequences were aligned using the program FUGUE (Shi et al., 2001) (Fig. 4A). (The first 104 amino acids of gp5 constitute a “delta” domain, which is cleaved during capsid assembly, and was absent from the gp5 crystal structure. It was omitted from the alignment.) The N-terminal 55 residues of gpN have two long predicted α -helices (residues 2–18 and 35–55) that do not correspond directly to the extended N-terminal conformation in the gp5 crystal structure. However, bacteriophages ϵ 15, P22 and 80 α also have α -helices in the N-arm (Jiang et al., 2008; Parent et al., 2010; Spilman et al., 2011). The gpN-gp5 alignment also predicts a number of insertions in the gpN sequence; the longest one (24 residues) was located between helices α 3 and α 4 (Fig. 4A).

2.5. Fitting gp5 into the P2 procapsid map

The HK97 gp5 procapsid structure (PDB ID: 3E8K) was placed into the electron density for the P2 procapsid. There was good correspondence between the A and P domains of gp5 and

the protruding and basal parts of the hexamer density (Fig. 4C), in an orientation similar to that of the HK97 procapsid (Gertsman et al., 2009). The spine helix ($\alpha 3$) of gp5 was immediately recognizable as a tubular density in the map. Although the A domain density was generally less well defined than the P domain, perhaps due to flexibility, tubular densities corresponding to helices $\alpha 5$ and $\alpha 6$ were easily recognizable (Figs. 4C). However, after fitting the P domain of gp5 using the spine helix as a landmark, the A domain fell outside the density. The angle between the two domains was therefore adjusted at the β hinge to allow both domains to fit into the density (Fig. 4C), similar to the adjustments used to allow for the conformational variation in HK97 and P22 (Gertsman et al., 2009; Teschke and Parent, 2010). Fitting gp5 into the P4 procapsid reconstruction gave the same result (not shown).

It was clear that the structure could not be made to fit reasonably into a map with the opposite hand (Fig. 4D), showing that the *dextro* hand of the $T=7$ lattice was correct. To further validate this result, we carried out simultaneous classification against two HK97-based models, one with a $T=7$ *laevo* and one with a $T=7$ *dextro* lattice. Only the $T=7$ *dextro* lattice resulted in a reconstruction.

2.6. Modeling of gpN

Based on the fitting and the secondary structure alignment (Fig. 4), we were confident that HK97 gp5 constituted a suitable model for gpN. Using the gpN-gp5 alignment as a guide, we generated a model for gpN with the I-TASSER threading server, specifying HK97 gp5 (PDB IDs 1OHG or 3E8K) as a template (Zhang, 2008). The top resulting I-TASSER models all had the gp5 fold and differed primarily in the N-arm, the E loop (residues 70–104) and the insertion loops (Fig. 5A).

In order to fit the gpN I-TASSER model into the P2 and P4 procapsid reconstructions, it was first split at the β hinge that divides the A and P domains (Fig. 5B). The A and P domains were fitted individually into the maps as rigid bodies, using clearly recognizable structural motifs as a guide, such as the $\alpha 3$ spine helix and the β sheet (βD , βH and βJ) in the P domain and helices $\alpha 5$ and $\alpha 6$ in the A domain (Fig. 5B–D). This was followed by minor additional relative adjustments of motifs within the domains. Some of the interspersing loops, which could not be modeled with confidence, were omitted from the model. Part of the E loop, for which there was no clear density, was omitted. The individual strands of the three-stranded β sheet could not be resolved in the map. However, they match a flat density that is connected to $\alpha 2$ at one end of the P domain (Fig. 5C, D). There were only small differences in gpN between P2 and P4, which could be accommodated by small shifts in the A domain helices (Fig. 5C,D).

Helix $\alpha 5$ in the A domain is part of a motif (residues 192–224) that includes four of the five known *sir* loci (Table 2; Figs 4A and 5B) (Six et al., 1991). Hence, we refer to this motif as the “*sir* loop”. Two of the *sir* loci reside in $\alpha 5$ (A217 and L221), while two are at the apex of the loop (D206, Y207). A fifth *sir* locus, M184, is located in the P domain, just before the *sir* loop. The *sir* loop matches an arch-like density in the A domain (Figs. 5C, D). This density makes contact with the Sid external scaffold in the P4 procapsid (Fig. 5D), consistent with a role of the *sir* mutations in affecting Sid–gpN interactions (see below).

The I-TASSER model predicts two α -helices ($\alpha 0$ and $\alpha 1$) in the N-terminal arm of gpN (Figs. 4A, 5A). In contrast, the gp5 N-arm in the HK97 virion contains little secondary structure and extends across to the E-loop of the adjacent subunit (Helgstrand et al., 2003) (Figs. 4A,B). In the HK97 procapsid, however, the N-arm points towards the inside of the shell (Gertsman et al., 2009) and in bacteriophages $\epsilon 15$, P22 and 80 α , there is a helix in this position (Jiang et al., 2008; Parent et al., 2010; Teschke and Parent, 2010; Spilman et al.,

2011). In the P2 and P4 maps, helix $\alpha 1$ most likely corresponds to a tubular density that extends from the A-P domain interface towards the interior of the shell, similar to $\alpha 1$ in P22 and 80 α (Fig. 5B,C,D). An additional tubular density underneath the P domain most likely corresponds to $\alpha 0$, although the connectivity to the rest of the subunit is unclear (Fig. 5B,C,D).

2.7. Intra- and intercapsomeric interactions

Intra-capsomeric interactions in both P2 and P4 involve contacts between the E loop and the P domain of the adjacent subunit, and between $\alpha 5$ (the *sir* loop) and $\alpha 6$ of the adjacent subunit (Figs. 6A-D), similar to interactions observed in other phages, including HK97 (Helgstrand et al., 2003). Conformational variation between subunits is evident in the $\alpha 6/\alpha 6'$ loop, which adjusts to fill in the central hole in the capsomers, and in the *sir* loop (Figs. 6A,B). The hexamer structures of P2 and P4 are remarkably similar, and almost perfectly superimposable (Fig. 6E), even though the P2 hexamers are not constrained by symmetry, while the P4 hexamers sit on icosahedral twofold axes (Fig. 3). The only obvious difference in a P4–P2 hexamer difference map is the external scaffold made of Sid (Fig. 6E).

Interactions between capsomers occur at trivalent contacts (Fig. 3), in which the P domains of adjacent subunits form a threefold annulus (Fig. 6F). These interactions are formed at the site of an insertion in gpN between βH and βJ (Fig. 4A), equivalent to the “P loop” of phages 80 α and P22 (Parent et al., 2010; Spilman et al., 2011).

2.8. The Sid scaffold

The external scaffold made of Sid consists of multiple segments of tubular density that connect in an annulus at the threefold axis and form an interlocking hook above the hexamers at the twofold axes (Figs. 3A and 6C–E). The size and shape of the tubular densities is consistent with α -helices. There is no homology model available for Sid; however, sequence analysis and biophysical characterization suggested that Sid is virtually entirely α -helical with a high propensity for coiled coil formation (Wang et al., 2000).

The scaffold spans a distance of about 9 nm from the twofold to threefold axes, constituting one icosahedral asymmetric unit. If Sid were 100% α -helical, it would have a length of 36.6 nm when fully extended. The tubular densities observed in the map account for a distance of about 25 nm—not taking into account connecting loops—which is thus roughly consistent with a single copy of Sid. The exact connectivity between the helices cannot be resolved in the map, and it is therefore impossible to say whether one Sid subunit stretches from twofold to threefold or if it has a more intertwined structure, e.g. stretching from threefold to threefold as a dimer. The super-sid (*nms*) mutations are clustered near the C-terminus of Sid (Kim et al., 2001). If we assume that these mutations are located at the gpN–Sid interface where the *sir* mutations are found, the C-terminus of Sid is most likely located in the finger-like protrusion that extends from the scaffold toward the *sir* loop (Figs. 5B,D and 6D).

3. Discussion

3.1. Structure and assembly of gpN

Over the past decade it has become clear that all tailed dsDNA bacteriophages (order *Caudovirales*) are structurally related and share a common capsid protein fold (Bamford et al., 2005; Johnson and Chiu, 2007), first described at high resolution in the *E. coli* bacteriophage HK97 (Wikoff et al., 2000). Apart from the HK97 gp5 capsid protein, examples of this fold are found in the structures of T4, T7, $\phi 29$, P22 and $\epsilon 15$ as well as herpesviruses and the archaeal virus-like particle PfV (Baker et al., 2005; Fokine et al., 2005; Morais et al., 2005; Agirrezabala et al., 2007; Akita et al., 2007; Jiang et al., 2008;

Parent et al., 2010). Thus, gpN was expected to conform to the same structure, in spite of a lack of detectable sequence homology. While the two folds are indeed similar, gpN differs from gp5 by having several insertions into its sequence, primarily in the P domain, conformational differences in the A domain, a partly disordered E loop and an ordered, α -helical N-arm (Figs. 4A and 5A,B). However, unlike e.g. P22 or T4 (Fokine et al., 2005; Parent et al., 2010), there is no additional, independent domain inserted into the gpN sequence.

One unexpected difference was the $T=7$ *dextro* lattice of the P2 capsid. Other $T=7$ phages observed so far, including HK97, P22, T7 and ϵ 15, all have $T=7$ *laevo* lattices. (The papovaviruses have pseudo- $T=7$ *dextro* lattices, but this is based on an all-pentamer architecture (Belnap et al., 1996).) It is conceivable that this is an artifact of the expression system and that native procapsids have $T=7$ *laevo* hand, but this seems unlikely. We also considered the possibility that the sample contained a mixture of both lattices, but refinement against two models of opposite hands yielded only a $T=7$ *dextro* structure. In spite of this difference in handedness, however, interactions between subunits and capsomers are retained: gpN forms the same kind of trivalent contacts between P domains as in many other bacteriophages (Fig. 3B, 6F). This anomaly also illustrates the point that protein chirality does not necessarily determine the handedness of the supramolecular assembly it forms (Chothia, 1991).

Non-equivalence in icosahedral capsids is reflected in conformational variability and differences in protein-protein interactions (Caspar, 1980; Dokland, 2000). In the P2 and P4 reconstructions, conformational variation is manifested through small shifts in the A domain helices, $\alpha 5$ and $\alpha 6/\alpha 6'$ (Figs. 6A–D). However, there is no distinctive conformational difference that distinguishes P2 and P4. Rather, small perturbations in the capsid protein structure and interactions propagated over the whole lattice is sufficient to accommodate the difference in curvature between the $T=4$ and $T=7$ shells.

3.2. The scaffolding proteins, gpO and Sid

The P2/P4 system is unique in that two distinct capsid sizes can be assembled from the same capsid protein under control of the two scaffolding proteins, gpO and Sid (Marvik et al., 1994b) (Fig. 1). gpO promotes the formation of $T=7$ shells primarily by curtailing the intrinsic flexibility of gpN, much like a molecular chaperone (Dokland et al., 1992; Dokland et al., 1999; Dokland, 2000; Chang et al., 2009). In the presence of both gpO and Sid, Sid takes precedence, leading to the formation of $T=4$ capsids (Wang et al., 2006). Consistent with this observation, gpO interacts only transiently with gpN, while Sid is stably associated with the procapsids until expansion occurs (Wang et al., 2003; Chang et al., 2008).

Sid is unique in that it forms an external, ordered scaffold, unlike the internal scaffolding cores found in most dsDNA bacteriophages (Dokland, 1999). The other example of an external scaffolding protein is that of the *Microviridae* family of ssDNA viruses (Dokland et al., 1997), which is a globular protein, unlike Sid. Sid has a propensity for coiled coil formation and spontaneously forms trimers and higher oligomers in solution (Wang et al., 2000). Sequence prediction suggests a high likelihood of disorder in solution. Many proteins are known to possess this kind of “intrinsic disorder” and become ordered only upon binding to a substrate (Uversky and Dunker, 2010). Once bound to gpN hexamers, the elongated Sid molecules constitute a tether that would constrain the possible shell assemblies to the smaller $T=4$ shell. This oligomerization with a stably bound Sid–gpN complex thereby overrides the transient, chaperone-like action of gpO.

We previously suggested that the Sid scaffold was made up of 120 copies of Sid protein, based on measurements on Coomassie-stained gels, and proposed an arrangement in which

two copies of elongated Sid subunits extend in tandem from the twofold to threefold axes of symmetry (Dokland et al., 2002). However, the P4 procapsid reconstruction shown here is consistent with the presence of only 60 copies of Sid, most likely in an arrangement where one copy of Sid stretches from the twofold to the threefold axes (Fig. 6C)

3.3. Maturation and capsid expansion

In P2 and P4, as in many dsDNA bacteriophages, capsid maturation involves cleavage of the capsid protein. Cleavage occurs after assembly (Marvik et al., 1994a), and pre-cleaved protein does not assemble correctly (Wang et al., 2000). Cleavage occurs prior to DNA packaging, however, and in procapsids isolated from a wildtype infection, gpN is already processed to its mature N* form (Rishovd and Lindqvist, 1992; Marvik et al., 1994a; Chang et al., 2008). In the P2 procapsid particles used here, cleavage of gpN and gpO does not occur because a protease-defective gpO was used (Wang et al., 2000; Chang et al., 2009). The 31-residue N-terminal peptide that is cleaved off from gpN includes the predicted helix $\alpha 0$ (Fig. 4A). By analogy with other phage capsid proteins, especially P22 (Parent et al., 2010), it is likely that a tubular density underneath the P domain corresponds to the $\alpha 0$ helix (Fig. 5C,D). This helix might contact the gpO scaffolding protein, in which case cleavage could lead to scaffold release. Alternatively, cleavage might destabilize intersubunit interactions in preparation for expansion. Indeed, experimental removal of the N-terminal 29 residues by trypsin treatment was previously shown to cause spontaneous capsid expansion of P4 procapsid shells (Wang et al., 2003). The presence of gpO might play a role in delaying the expansion until DNA packaging occurs [the P4 procapsids that were expanded *in vitro* (Wang et al., 2003) did not contain gpO].

3.4. Function of *sir* mutations

It was previously hypothesized that the *sir* mutations affected the flexibility in a presumed hinge region in gpN (Six et al., 1991; Kim et al., 2001). This hypothesis was based on the clustering of the *sir* mutations roughly in the middle of the gpN sequence and the observation of an apparent two-domain structure in the low resolution reconstructions of P2 and P4 mature capsids (Dokland et al., 1992). Variation in the observed angle between these two domains appeared to support this idea. This hypothesis was made before any high resolution structures of bacteriophage capsids had been solved. In light of the subsequent structures of HK97 gp5 and numerous other bacteriophage capsid proteins (Wikoff et al., 2000; Fokine et al., 2005; Jiang et al., 2008; Parent et al., 2010), it is now clear that this model is incorrect. Although the HK97 fold has two domains, their sequences are intertwined, with the A domain essentially forming an insertion in the P domain (Fig. 4A,B). Consequently, the *sir* mutations end up largely clustered in one loop (the *sir* loop) at the apex of the A domain (Fig. 5B). This loop forms the primary, if not the only, interaction point between Sid and gpN in the P4 procapsid (Fig. 5D). Thus, we conclude that the *sir* mutations most likely disrupt gpN–Sid interactions, either directly or indirectly, through changes in loop conformation. *sir3*, which affects M184, is located in the sequence preceding the *sir* loop, and might be more likely to affect loop conformation or flexibility.

The *nms* or “super-sid” mutations are clustered near the C-terminal end of Sid, and it was previously suggested that these mutations led to a strengthening of the Sid lattice at the threefold interaction point (Kim et al., 2001; Dokland et al., 2002). In light of the above results, it is more likely that the super-sid mutants suppress the *sir* phenotype by restoring interactions that were disrupted by the *sir* mutations. The C-terminus of Sid is therefore most likely located near the gpN–Sid interface, in the finger-like protrusions of the scaffold at these points (Fig. 5B,D).

3.5. Perspectives

The P2/P4 system represents a fascinating example of molecular piracy, in which a replicon usurps the structural gene products of an unrelated replicon for its own mobilization and spread. Satellite viruses are not uncommon in eukaryotes, but these viruses generally encode their own capsid proteins. Hepatitis delta virus, for example, uses envelope proteins supplied by hepatitis B virus, but encodes its own nucleocapsid protein (Sureau, 2006). P4 is different in that it is a complete and independent replicon with a highly specialized machinery for helper exploitation, depends on the helper for structural proteins, and has the ability to redirect the capsid assembly process. As it turns out, the P2/P4 system is not the only example of such a mechanism. In *Staphylococcus aureus*, genetic elements called *S. aureus* pathogenicity islands (SaPIs) are mobilized by specific helper phages (Lindsay et al., 1998; Novick et al., 2010) and are packaged into phage-like transducing particles using structural proteins supplied by the helper phage (Poliakov et al., 2008). In some cases, the SaPI capsids are smaller than those normally formed by the helper. In the *S. aureus* system, capsid size determination apparently depends on two internal scaffolding proteins (Dearborn et al., 2011; Spilman, 2011).

4. Materials and methods

4.1. Cloning and expression of co-expression constructs

The plasmids used are listed in Table 1. pLucky7, which expresses gpN and Sid under control of separate T7 promoters from a chimeric vector, was previously described (Dokland et al., 2002). pET16b-derived plasmids pSW101 and pJRC49 express protease-deficient truncated and mutated versions of gpO, respectively, in tandem with gpN under control of the same T7 promoter (Wang et al., 2006; Chang et al., 2009). The *Q* gene was inserted after *OΔ25* and *N* in pSW101 to generate plasmid pCMRP1 that co-expresses *OΔ25*, gpN and gpQ.

All cloning was performed in the *E. coli recA⁻* strain DH5α (Invitrogen). For expression, clones were transformed into *E. coli* strains BL21(DE3) or BLR(DE3) (Novagen). Expression was induced by IPTG and after 2 hr of growth, the cells were lysed in buffer containing 10 mM Tris-HCl pH 7.4, 200 mM NaCl, 1 mM PMSF, 1% Triton X-100, and 0.5% deoxycholate by freeze-thawing and high pressure disruption in a French Press (Thermo Fisher) or in an Emulsiflex EF-C3 high-pressure cell disruptor (Avestin Inc., Ottawa). Procapsids were harvested from the clarified lysate by pelleting at 110,000×g followed by one or two cycles of purification on 10–40% sucrose gradients in 50 mM Tris-HCl, pH 8.0, 100 mM or 1 M NaCl, 10 mM MgCl₂ for 2 h at 110,000×g, 4 °C. Gradient fractions (1 ml each) were collected manually from the top of the gradient. Fractions containing procapsids were pooled, pelleted by centrifugation for 40 min at 178,000×g and resuspended in a low-ionic strength buffer containing 10mM Tris-HCl, pH 8.0, 20 mM NaCl and 2 mM MgCl₂.

4.2. Electron microscopy and three-dimensional reconstruction

Cryo-EM was done by standard methods: the sample was applied to either non-glow discharged, chloroform-washed lacey grids (Electron Microscopy Sciences, Hartfield, PA) or glow discharged Quantifoil R2/2 grids (Quantifoil Micro Tools, Jena), blotted for 1–3 sec with filter paper, plunged into liquid ethane using a manual plunger, and transferred to a Gatan 626 cryo-sample holder (Baker et al., 1999; Huiskonen et al., 2004; Dokland and Ng, 2006). All samples were observed in an FEI Tecnai F20 field-emission gun (FEG) electron microscope operated at 200 kV. Images to be used for reconstruction were collected on Kodak SO-163 film under low dose conditions, at magnifications of 50,000x or 62,000x and typical defocus values of 1–3 μm.

The P4 procapsid images were digitized on a CreoScitex EverSmart scanner with a step size of 4.545 μm (220 dots/mm) and a 3x binning factor, leading to a pixel size of 2.73 $\text{\AA}$ /pixel with respect to the sample. P2 procapsid images were digitized with a Nikon CoolScan 9000ED film scanner with a raster step size of 12.7 μm (2000 dpi), corresponding to 2.05 or 2.54 $\text{\AA}$ /pixel.

Three-dimensional reconstruction was done with EMAN versions 1.8-1.9 (Ludtke et al., 1999; Ludtke et al., 2001; Ludtke et al., 2004) and with AUTO3DEM (Yan et al., 2007). The contrast transfer function (CTF) curves were determined interactively using the program CTFIT in the EMAN suite (Ludtke et al., 1999; Saad et al., 2001) and a phase correction applied to the images. For P4, the existing low resolution reconstruction was used as a starting model, while for P2, the STARTICOS routine was used to generate a model from scratch. To check for handedness, the images were simultaneously classified against two models created by placing the HK97-based model in $T=7$ lattices of opposite hands. The final maps from EMAN were used as starting models for AUTO3DEM refinement. The final maps were calculated to 8 $\text{\AA}$ resolution, and an inverse B factor of 200–250 $\text{\AA}^2$ was applied for map sharpening using bfilter from the Bsoft suite (Heymann and Belnap, 2007). Map visualization, segmentation and modeling were done using UCSF Chimera (Pettersen et al., 2004; Goddard et al., 2007). The P2 and P4 procapsid maps were deposited in EMDB with accession codes EMD-25276 and EMD-25273, respectively.

4.3. Modeling

Prediction of secondary structure was done with the PROF (Ouali and King, 2000) and GOR (Garnier et al., 1996) servers. The gpN sequence was aligned with the HK97 gp5 structure (PDB ID: 1OHG) using FUGUE (Shi et al., 2001), giving an initial alignment of the secondary structure elements of the two proteins. A gpN starting model was generated by I-TASSER (Zhang, 2008), specifying the HK97 gp5 structure (PDB IDs 1OHG or 3E8K; both resulted in essentially identical models) as a template. The model was separated into the A and P domains, which were fitted separately into the corresponding density in UCSF Chimera (Pettersen et al., 2004; Goddard et al., 2007). Minimal additional adjustments were made in the A and P domains, and some loops and other features that did not clearly correspond to density were omitted from the final model.

Acknowledgments

Funding for this work at UAB was partly provided by The National Institutes of Health grants R21 AI071982 and R01 AI083255 to T.D. This work was initiated while T.D. was at the Institute for Molecular Agrobiolgy, Singapore, and was funded by Singapore's Agency for Science, Technology and Research (A*STAR). We thank the Biocenter Finland National Cryo Electron Microscopy Unit, Institute of Biotechnology, Helsinki University, for providing services.

References

- Agirrezabala X, Velasquez-Muriel JA, Gomez-Puertas P, Scheres SHW, Carazo JM, Carrascosa JL. Quasi-atomic model of bacteriophage T7 procapsid shell: insights into the structure and evolution of a basic fold. *Structure*. 2007; 15:461–472. [PubMed: 17437718]
- Akita F, Chong KT, Tanaka H, Yamashita Y, Miyazaki N, Nakaishi Y, Suzuki M, Namba K, Ono Y, Tsukihara T, Nakagawa A. The crystal structure of a virus-like particle from the hyperthermophilic archaeon *Pyrococcus furiosus* provides insight into the evolution of viruses. *J. Mol. Biol.* 2007; 368:1469–1483. [PubMed: 17397865]
- Baker ML, Jiang W, Rixon FJ, Chiu W. Common ancestry of herpesviruses and tailed DNA bacteriophages. *J. Virol.* 2005; 79:14967–14970. [PubMed: 16282496]

- Baker TS, Olson NH, Fuller SD. Adding the third dimension to virus life cycles: three-dimensional reconstruction of icosahedral viruses from cryo-electron micrographs. *Microbiol. Molec. Biol. Rev.* 1999; 63:862–922. [PubMed: 10585969]
- Bamford DH, Grimes JM, Stuart DI. What does structure tell us about virus evolution? *Curr. Opin. Struct. Biol.* 2005; 15:655–663. [PubMed: 16271469]
- Barrett KJ, Marsh ML, Calendar R. Interactions between a satellite bacteriophage and its helper. *J. Mol. Biol.* 1976; 106:683–707. [PubMed: 789896]
- Belnap DM, Olson NH, Cladel NM, Newcomb WW, Brown JC, Kreider JW, Christensen ND, Baker TS. Conserved features in papillomavirus and polyomavirus capsids. *J. Mol. Biol.* 1996; 259:249–263. [PubMed: 8656427]
- Bertani, LE.; Six, E. The P2-like phages and their parasite, P4. Calendar, R., editor. Plenum Press; New York: 1988. p. 73-143.
- Bertani LE, Bertani G. Preparation and characterization of temperate, noninducible bacteriophage P2 (host: *Escherichia coli*). *J. Gen. Virol.* 1970; 6:201–212. [PubMed: 4908574]
- Bowden DW, Modrich P. In vitro maturation of circular bacteriophage P2 DNA. *J. Biol. Chem.* 1985; 260:6999–7007. [PubMed: 2987239]
- Breitbart M, Salamon P, Andresen B, Mahaffy JM, Segall AM, Mead D, Azam F, Rohwer F. Genomic analysis of uncultured marine viral communities. *Proc. Natl Acad. Sci. USA.* 2002; 99:14250–14255. [PubMed: 12384570]
- Briani F, Deho G, Forti F, Ghisotti D. The plasmid status of satellite bacteriophage P4. *Plasmid.* 2001; 45:1–17. [PubMed: 11319927]
- Caspar DL. Movement and self-control in protein assemblies. Quasi-equivalence revisited. *Biophys. J.* 1980; 32:103–138. [PubMed: 6894706]
- Chang JR, Poliakov A, Prevelige PE, Mobley JA, Dokland T. Incorporation of scaffolding protein gpO in bacteriophages P2 and P4. *Virology.* 2008; 370:352–361. [PubMed: 17931675]
- Chang JR, Spilman MS, Rodenburg CM, Dokland T. Functional domains of the bacteriophage P2 scaffolding protein: identification of residues involved in assembly and protease activity. *Virology.* 2009; 384:144–150. [PubMed: 19064277]
- Chothia, C. Asymmetry in protein structures.. In: Bock, GR.; Marsh, J., editors. *Biological asymmetry and handedness*. John Wiley & Sons; Chichester, England: 1991. p. 36-49.
- Conway JF, Duda RL, Cheng N, Hendrix RW, Steven AC. Proteolytic and conformational control of virus capsid maturation: the bacteriophage HK97 system. *J. Mol. Biol.* 1995; 253:86–99. [PubMed: 7473720]
- Dearborn A, Spilman M, Damle P, Chang J, Monroe E, Saad J, Christie G, Dokland T. The *Staphylococcus aureus* Pathogenicity Island 1 Protein gp6 Functions as an Internal Scaffold during Capsid Size Determination. *J. Mol. Biol.* 2011; 412:710–722. [PubMed: 21821042]
- Deho, G.; Ghisotti, D. The satellite phage P4.. In: Calendar, R., editor. *The bacteriophages*. Oxford University Press; New York: 2006. p. 391-408.
- Dokland T. Scaffolding proteins and their role in viral assembly. *Cell. Mol. Life Sci.* 1999; 56:580–603. [PubMed: 11212308]
- Dokland T. Freedom and restraint: themes in virus capsid assembly. *Structure.* 2000; 8:R157–R167. [PubMed: 10997898]
- Dokland T, Bernal RA, Burch A, Pletnev S, Fane BA, Rossmann MG. The role of scaffolding proteins in the assembly of the small, single-stranded DNA virus ϕ X174. *J. Mol. Biol.* 1999; 288:595–608. [PubMed: 10329166]
- Dokland T, Isaksen ML, Fuller SD, Lindqvist BH. Capsid localization of the bacteriophage P4 Psi protein. *Virology* 1993; 194:682–687.
- Dokland T, Lindqvist BH, Fuller SD. Image reconstruction from cryo-electron micrographs reveals the morphopoietic mechanism in the P2-P4 bacteriophage system. *EMBO J. The European Molecular Biology Organization Journal.* 1992; 11:839–846.
- Dokland T, McKenna R, Ilag LL, Bowman BR, Incardona NL, Fane BA, Rossmann MG. Structure of a viral procapsid with molecular scaffolding. *Nature (London).* 1997; 389:308–313. [PubMed: 9305849]

- Dokland T, Murialdo H. Structural Transitions during maturation of bacteriophage lambda capsids. *J. Mol. Biol.* 1993; 233:682–694. [PubMed: 8411174]
- Dokland, T.; Ng, ML. Transmission electron microscopy of biological specimens. Dokland, T.; Huttmacher, DW.; Ng, ML.; Schantz, JT., editors. World Scientific Press; Singapore: 2006. p. 153-208.
- Dokland T, Wang S, Lindqvist BH. The structure of P4 procapsids produced by coexpression of capsid and external scaffolding proteins. *Virology.* 2002; 298:224–231. [PubMed: 12127785]
- 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. USA.* 2005; 102:7163–7168. [PubMed: 15878991]
- Garnier J, Gibrat JF, Robson B. GOR secondary structure prediction method version IV. *Methods Enzymol.* 1996; 266:540–553. [PubMed: 8743705]
- Geisselsoder J, Sedivy JM, Walsh RB, Goldstein R. Capsid structure of satellite phage P4 and its helper. *J. Ultrastruct. Res.* 1982; 79:165–173. [PubMed: 7077742]
- Gertsman I, Gan L, Guttman M, Lee K, Speir JA, Duda RL, Hendrix RW, Komives EA, Johnson JE. An unexpected twist in viral capsid maturation. *Nature.* 2009; 458:646–651. [PubMed: 19204733]
- Goddard TD, Huang CC, Ferrin TE. Visualizing density maps with UCSF Chimera. *J. Struct. Biol.* 2007; 157:281–287. [PubMed: 16963278]
- Helgstrand C, Wikoff WR, Duda RL, Hendrix RW, Johnson JE, Liljas L. The refined structure of a protein catenane: the HK97 bacteriophage capsid at 3.44Å resolution. *J. Mol. Biol.* 2003; 334:865–899.
- Heymann J, Belnap D. Bsoft: image processing and molecular modeling for electron microscopy. *J. Struct. Biol.* 2007; 157:3–18. [PubMed: 17011211]
- Huiskonen JT, Kivela HM, Bamford DH, Butcher SJ. The PM2 virion has a novel organization with an internal membrane and pentameric receptor binding spikes. *Nat. Struct. Mol. Biol.* 2004; 11:850–856. [PubMed: 15286721]
- Jiang W, Baker ML, Jakana J, Weigele PR, King J, Chiu W. Backbone structure of the infectious e15 virus capsid revealed by electron cryomicroscopy. *Nature.* 2008; 451:1130–1135. [PubMed: 18305544]
- Jiang W, Li Z, Zhang Z, Baker ML, Prevelige PE, Chiu W. Coat protein fold and maturation transition of bacteriophage P22 seen at subnanometer resolutions. *Nat. Struct. Biol.* 2003; 10:131–135. [PubMed: 12536205]
- Johnson JE, Chiu W. DNA packaging and delivery machines in tailed bacteriophages. *Curr. Opin. Struct. Biol.* 2007; 17:237–243. [PubMed: 17395453]
- Kim K-J, Sunshine MG, Lindqvist BH, Six EW. Capsid size determination in the P2-P4 bacteriophage system: suppression of sir mutations in P2's capsid gene N by supersid mutations in P4's external scaffold gene sid. *Virology.* 2001; 283:49–58. [PubMed: 11312661]
- Lengyel JA, Goldstein RN, Marsh M, Calendar R. Structure of the bacteriophage P2 tail. *Virology.* 1974; 62:161–164. [PubMed: 4607804]
- Lengyel JA, Goldstein RN, Marsh M, Sunshine MG, Calendar R. Bacteriophage P2 head morphogenesis: cleavage of the major capsid protein. *Virology.* 1973; 53:1–23. [PubMed: 4574872]
- Lindqvist BH, Deho G, Calendar R. Mechanisms of genome propagation and helper exploitation by satellite phage P4. *Microbiol. Rev.* 1993; 57:683–702. [PubMed: 8246844]
- Lindsay JA, Ruzin A, Ross HF, Kurepina N, Novick RP. The gene for toxic shock toxin is carried by a family of mobile pathogenicity islands in *Staphylococcus aureus*. *Mol. Microbiol.* 1998; 29:527–543. [PubMed: 9720870]
- Ludtke SJ, Baldwin PR, Chiu W. EMAN: semiautomated software for high-resolution single-particle reconstructions. *J. Struct. Biol.* 1999; 128:82–97. [PubMed: 10600563]
- Ludtke SJ, Cheng DH, Song JL, Chuang DT, Chiu W. Seeing GroEL at 6 Å resolution by single particle electron cryomicroscopy. *Structure.* 2004; 12:1129–1136. [PubMed: 15242589]
- Ludtke SJ, Jakana J, Song JL, Chuang DT, Chiu W. A 11.5 Å single particle reconstruction of GroEL using EMAN. *J. Mol. Biol.* 2001; 314:253–262. [PubMed: 11718559]

- Marvik OJ, Dokland T, Nøkling RH, Jacobsen E, Larsen T, Lindqvist BH. The capsid size-determining protein Sid forms an external scaffold on phage P4 procapsids. *J. Mol. Biol.* 1995; 251:59–75. [PubMed: 7643390]
- Marvik OJ, Jacobsen E, Dokland T, Lindqvist BH. Bacteriophage P2 and P4 morphogenesis: assembly precedes proteolytic processing of the capsid proteins. *Virology.* 1994a; 205:51–65. [PubMed: 7975237]
- Marvik OJ, Sharma P, Dokland T, Lindqvist BH. Bacteriophage P2 and P4 assembly: alternative scaffolding proteins regulate capsid size. *Virology.* 1994b; 200:702–714. [PubMed: 8178454]
- 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]
- Nilsson, AS.; Haggård-Ljungquist, E. The P2-like bacteriophages.. In: Calendar, R., editor. *The bacteriophages*. Oxford University Press; New York: 2006. p. 365-390.
- Nilsson AS, Karlsson JL, Haggård-Ljungquist E. Site-specific recombination links the evolution of P2-like coliphages and pathogenic enterobacteria. *Mol. Biol. Evol.* 2004; 21:1–13. [PubMed: 12949155]
- Novick RP, Christie GE, Penades JR. The phage-related chromosomal islands of Gram-positive bacteria. *Nat. Rev. Microbiol.* 2010; 8:541–551. [PubMed: 20634809]
- Ouali M, King RD. Cascaded multiple classifiers for secondary structure prediction. *Protein Sci.* 2000; 9:1162–1176. [PubMed: 10892809]
- Parent KN, Khayat R, Tu LH, Suhanovsky MM, Cortines JR, Teschke CM, Johnson JE, Baker TS. P22 coat protein structures reveal a novel mechanism for capsid maturation: stability without auxiliary proteins or chemical crosslinks. *Structure.* 2010; 18:390–401. [PubMed: 20223221]
- Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera - a visualization system for exploratory research and analysis. *J. Comput. Chem.* 2004; 25:1605–1612. [PubMed: 15264254]
- Poliakov A, Chang JR, Spilman MS, Damle PK, Christie GE, Mobley JA, Dokland T. Capsid size determination by *Staphylococcus aureus* pathogenicity island SaPII involves specific incorporation of SaPII proteins into procapsids. *J. Mol. Biol.* 2008; 380:465–475. [PubMed: 18565341]
- Rishovd S, Lindqvist BH. Bacteriophage P2 and P4 morphogenesis: protein processing and capsid size determination. *Virology.* 1992; 187:548–554. [PubMed: 1546453]
- Rishovd S, Marvik OJ, Jacobsen E, Lindqvist BH. Bacteriophage P2 and P4 morphogenesis: identification and characterization of the portal protein. *Virology.* 1994; 200:744–751. [PubMed: 8178458]
- Saad A, Ludtke SJ, Jakana J, Rixon FJ, Tsuruta H, Chiu W. Fourier amplitude decay of electron cryomicroscopic images of single particles and effects on structure determination. *J. Struct. Biol.* 2001; 133:32–42. [PubMed: 11356062]
- Shi J, Blundell TL, Mizuguchi K. FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties. *J. Mol. Biol.* 2001; 310:243–257. [PubMed: 11419950]
- Shore D, Deho G, Tsipis J, Goldstein R. Determination of capsid size by satellite bacteriophage P4. *Proc. Natl. Acad. Sci. USA.* 1978; 75:400–404. [PubMed: 272656]
- Six EW, Sunshine MG, Williams J, Haggård-Ljungquist E, Lindqvist BH. Morphopoietic switch mutations of bacteriophage P2. *Virology.* 1991; 182:34–46. [PubMed: 1840708]
- Six EW. The helper dependence of satellite bacteriophage P4: which gene functions of bacteriophage P2 are needed by P4? *Virology.* 1975; 67:249–263. [PubMed: 1099784]
- Spilman, MS. Molecular piracy in the mobilization of *Staphylococcus aureus* pathogenicity island 1. Dept. of Microbiology Ph.D.; 2011.
- Spilman MS, Dearborn AD, Chang JR, Damle PK, Christie GE, Dokland T. A conformational switch involved in maturation of *Staphylococcus aureus* bacteriophage 80alpha capsids. *J. Mol. Biol.* 2011 in press.
- Sureau C. The role of the HBV envelope proteins in the HDV replication cycle. *Curr. Top. Microbiol. Immunol.* 2006; 307:113–131. [PubMed: 16903223]

- Teschke CM, Parent KN. 'Let the phage do the work': Using the phage P22 coat protein structures as a framework to understand its folding and assembly mutants. *Virology*. 2010; 401:119–130. [PubMed: 20236676]
- Uversky VN, Dunker AK. Understanding protein non-folding. *Biochim. Biophys. Acta*. 2010; 1804:1231–1264. [PubMed: 20117254]
- Wang S, Chandramouli P, Butcher S, Dokland T. Cleavage leads to expansion of bacteriophage P4 procapsids in vitro. *Virology*. 2003; 314:1–8. [PubMed: 14517054]
- Wang S, Chang JR, Dokland T. Assembly of bacteriophage P2 and P4 procapsids with internal scaffolding protein. *Virology*. 2006; 348:133–140. [PubMed: 16457867]
- Wang S, Palasingam P, Nøklung RH, Lindqvist BH, Dokland T. In vitro assembly of bacteriophage P4 procapsids from purified capsid and scaffolding proteins. *Virology*. 2000; 275:133–144. [PubMed: 11017795]
- Wikoff WR, Liljas L, Duda RL, Tsuruta H, Hendrix RW, Johnson JE. Topologically linked protein rings in the bacteriophage HK97 capsid. *Science*. 2000; 289:2129–2133. [PubMed: 11000116]
- Yan X, Sinkovits R, Baker T. AUTO3DEM--an automated and high throughput program for image reconstruction of icosahedral particles. *J. Struct. Biol*. 2007; 157:73–82. [PubMed: 17029842]
- Zhang Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*. 2008; 9:40. [PubMed: 18215316]

Figure 1.

Schematic diagram of the P2/P4 capsid assembly pathway. (A) Capsid protein gpN, scaffolding protein gpO and connector (portal) protein gpQ are assembled into the $T=7$ P2 procapsid. (B) GpN, gpO and gpQ are then processed to their mature forms, N*, O* and Q*, respectively. DNA is packaged into the procapsid by the P2 terminase complex (gpM and gpP), accompanied by expansion into a mature capsid (B). (C) In the presence of the P4-encoded protein Sid, P2 structural proteins are instead assembled into a small, $T=4$ procapsid with Sid forming an external scaffold. (D) Loss of Sid, (E) protein processing and DNA packaging leads to the mature, expanded P4 capsid. (F) The decoration protein Psu is added to the mature capsid.

Figure 2.
Cryo-electron micrographs of P4 (A) and P2 (B) procapsids. Scale bar, 100 nm.

Figure 3.

Isosurface representations of the P4 (A) and P2 (B) procapsid reconstructions, viewed down a twofold axis of symmetry. The large triangles correspond to one icosahedral face, delimited by three fivefold axes of symmetry (pentagons). Three twofold symmetry axes (ovals) are located along the edges of the triangle. The four non-equivalent gpN subunits in the asymmetric unit in P4 and the seven subunits in P2 are labeled from A through G. The same triangle is represented schematically below the corresponding reconstructions, showing the subunit arrangement in the $T=4$ and $T=7$ *dextro* lattices. The icosahedral (type 1) threefold axis is indicated by a filled triangle, while the local type 2 and type 3 threefold axes are indicated by open triangles.

Figure 4.

Structure of gpN and HK97 gp5. (A) Alignment of the secondary structure elements of gpN and gp5. α -helices and β -strands are shown as boxes and arrows, respectively, labeled according to (Helgstrand et al., 2003) and colored as follows: red, N-arm; pink, E loop; blue, P domain β sheet; gold, spine helix motif; green, A domain. The dashed lines in gp5 correspond to insertions in gpN. The purple triangles indicate the location of *sir* mutations in the gpN sequence. (B) Ribbon diagram of HK97 gp5, colored as in (A). (C) Ribbon diagram of HK97 gp5 fitted into the P2 procapsid (gpN subunit B) density, after adjusting the angle between the A and P domains (colored as in A and B). The original, unmoved gp5 is shown in gray. (D) Comparison of fitting the HK97 P domain into the P2 procapsid hexamer density in the *T=7 laevo* and *T=7 dextro* lattices. Three subunits are shown (gold, red, purple).

Figure 5.

Modeling of gpN. (A) Ribbon diagram showing the superposition of the top two ITASSER models (red and blue) on the HK97 gp5 template (gray). (B) The gpN model, based on the ITASSER models, modified to fit in the P2 procapsid map, colored the same way as gp5 in Fig. 4. Pertinent secondary structure elements are labeled. The amino acids affected by the *sir* mutations are shown as purple spheres. (C) The gpN model fitted into the P2 procapsid subunit B density. (D) gpN model fitted into the P4 procapsid subunit B density. The density corresponding to Sid is shown in red, with a model for the proposed C-terminal helix fitted in.

Figure 6.

GpN and Sid interactions. (A) Isosurface representation (mesh) of a hexamer from the P2 procapsid reconstruction, viewed down the quasi-sixfold axis. The six gpN subunits are shown in ribbon representation (B, yellow; C, red; D, blue; E, purple; F, green; G, orange). (B) Oblique view of the same hexamer (yellow isosurface), showing details of the interactions between two subunits, B (yellow) and C (red). (C) View of the P4 hexamer, viewed down the icosahedral twofold axis, with gpN subunits fitted (B, yellow; C, red; D, blue). The Sid scaffold is shown as a solid red isosurface. Icosahedral threefold axes are indicated by triangles. (D) Oblique view of the P4 procapsid hexamer, including the Sid scaffold. (E) Isosurfaces of the P2 (blue mesh) and P4 (yellow solid surface) hexamers

superimposed. The calculated difference map between the two is shown as a red isosurface, showing that the only significant difference between the P2 and P4 hexamers is the Sid scaffold. (F) Superposition of P2 (blue mesh) and P4 (red mesh) density at the type 2 (ABC) trivalent interaction. The three subunits involved in the interaction are shown in ribbon representation (A, pink; B, yellow; C, red).

Table 1

Expression plasmids used in this study.

Plasmid name	Vector	Proteins expressed	Reference
pLucky7	pET11a	gpN, Sid	(Dokland et al., 2002)
pSW101 [§]	pET16b	OΔ25, gpN	(Wang et al., 2006)
pJRC49 [¶]	pET16b	O(S107A), gpN	(Chang et al., 2009)
pCMRP1	pET16b	OΔ25, gpN, gpQ	This work

[§]Equivalent to pOΔ25N described in (Wang et al., 2006)

[¶]Referred to as O(S107A)+N in (Chang et al., 2009)

Table 2

N sir mutations and their location in the gpN model (Six et al., 1991).

Mutation name	Residue number	Mutation	Location
<i>sir3</i>	184	M→T	In P domain, before <i>sir</i> loop
<i>sir5</i>	206	D→Y	apex of <i>sir</i> loop
<i>sir9</i>	206	ΔD	apex of <i>sir</i> loop
<i>sir2</i> , <i>sir6</i> , <i>sir8</i>	207	Y→F	apex of <i>sir</i> loop
<i>sir22</i>	217	A→E	<i>sir</i> loop helix (α5)
<i>sir22q</i>	217	A→Q	<i>sir</i> loop helix (α5)
<i>sir1</i>	221	L→V	<i>sir</i> loop helix (α5)
<i>sir4</i>	221	L→Q	<i>sir</i> loop helix (α5)