VIROLOGY JOURNAL

REVIEW   **Open Access**

Structure and assembly of bacteriophage T4 head

Venigalla B Rao[1*], Lindsay W Black[2]

Abstract

The bacteriophage T4 capsid is an elongated icosahedron, 120 nm long and 86 nm wide, and is built with three essential proteins; gp23*, which forms the hexagonal capsid lattice, gp24*, which forms pentamers at eleven of the twelve vertices, and gp20, which forms the unique dodecameric portal vertex through which DNA enters during packaging and exits during infection. The past twenty years of research has greatly elevated the understanding of phage T4 head assembly and DNA packaging. The atomic structure of gp24 has been determined. A structural model built for gp23 using its similarity to gp24 showed that the phage T4 major capsid protein has the same fold as that found in phage HK97 and several other icosahedral bacteriophages. Folding of gp23 requires the assistance of two chaperones, the *E. coli* chaperone GroEL and the phage coded gp23-specific chaperone, gp31. The capsid also contains two non-essential outer capsid proteins, Hoc and Soc, which decorate the capsid surface. The structure of Soc shows two capsid binding sites which, through binding to adjacent gp23 subunits, reinforce the capsid structure. Hoc and Soc have been extensively used in bipartite peptide display libraries and to display pathogen antigens including those from HIV, *Neisseria meningitides*, *Bacillus anthracis*, and FMDV. The structure of Ip1*, one of the components of the core, has been determined, which provided insights on how IPs protect T4 genome against the *E. coli* nucleases that degrade hydroxymethylated and glycosylated T4 DNA. Extensive mutagenesis combined with the atomic structures of the DNA packaging/terminase proteins gp16 and gp17 elucidated the ATPase and nuclease functional motifs involved in DNA translocation and headful DNA cutting. Cryo-EM structure of the T4 packaging machine showed a pentameric motor assembled with gp17 subunits on the portal vertex. Single molecule optical tweezers and fluorescence studies showed that the T4 motor packages DNA at a rate of up to 2000 bp/sec, the fastest reported to date of any packaging motor. FRET-FCS studies indicate that the DNA gets compressed during the translocation process. The current evidence suggests a mechanism in which electrostatic forces generated by ATP hydrolysis drive the DNA translocation by alternating the motor between tensed and relaxed states.

Introduction

The T4-type bacteriophages are ubiquitously distributed in nature and occupy environmental niches ranging from mammalian gut to soil, sewage, and oceans. More than 130 such viruses that show similar morphological features as phage T4 have been described; from the T4 superfamily ~1400 major capsid protein sequences have been correlated to its 3D structure [1-3]. The features include large elongated (prolate) head, contractile tail, and a complex baseplate with six long, kinked tail fibers radially emanating from it. Phage T4 historically has served as an excellent model to elucidate the mechanisms of head assembly of not only T-even phages but of large icosahedral viruses in general, including the widely distributed eukaryotic viruses such as the herpes viruses. This review will focus on the advances in the past twenty years on the basic understanding of phage T4 head structure and assembly and the mechanism of DNA packaging. Application of some of this knowledge to develop phage T4 as a surface display and vaccine platform will also be discussed. The reader is referred to the comprehensive review by Black et al [4], for the early work on T4 head assembly.

* Correspondence: rao@cua.edu
[1]Department of Biology, The Catholic University of America, Washington, DC, USA
Full list of author information is available at the end of the article

Structure of phage T4 capsid

The overall architecture of the phage T4 head determined earlier by negative stain electron microscopy of the procapsid, capsid, and polyhead, including the positions of the dispensable Hoc and Soc proteins, has basically not changed as a result of cryo-electron microscopic structure determination of isometric capsids [5]. However, the dimensions of the phage T4 capsid and its inferred protein copy numbers have been slightly altered on the basis of the higher resolution cryo-electron microscopy structure. The width and length of the elongated prolate icosahedron [5] are T_{end} = 13 laevo and T_{mid} = 20 (86 nm wide and 120 nm long), and the copy numbers of gp23, Hoc and Soc are 960, 155, and 870, respectively (Figure 1).

The most significant advance was the crystal structure of the vertex protein, gp24, and by inference the structure of its close relative, the major capsid protein gp23 [6]. This ~0.3 nm resolution structure permits rationalization of head length mutations in the major capsid protein as well as of mutations allowing bypass of the vertex protein. The former map to the capsomer's periphery and the latter within the capsomer. It is likely that the special gp24 vertex protein of phage T4 is a relatively recent evolutionary addition as judged by the ease with which it can be bypassed. Cryo-electron microscopy showed that in the bypass mutants that substitute pentamers of the major capsid protein at the vertex, additional Soc decoration protein subunits surround these gp23* molecules, which does not occur in the gp23*-gp24* interfaces of the wild-type capsid [7].

Figure 1 Structure of the bacteriophage T4 head. **A)** Cryo-EM reconstruction of phage T4 capsid [5]; the square block shows enlarged view showing gp23 (yellow subunits), gp24 (purple subunits), Hoc (red subunits) and Soc (white subunits); **B)** Structure of RB49 Soc; **C)** Structural model showing one gp23 hexamer (blue) surrounded by six Soc trimers (red). Neighboring gp23 hexamers are shown in green, black and magenta [28]; **D)** Structure of gp24 [6]; **E)** Structural model of gp24 pentameric vertex.

Nevertheless, despite the rationalization of major capsid protein affecting head size mutations, it should be noted that these divert only a relatively small fraction of the capsids to altered and variable sizes. The primary determinant of the normally invariant prohead shape is thought to be its scaffolding core, which grows concurrently with the shell [4]. However, little progress has been made in establishing the basic mechanism of size determination or in determining the structure of the scaffolding core.

The gp24 and inferred gp23 structures are closely related to the structure of the major capsid protein of bacteriophage HK97, most probably also the same protein fold as the majority of tailed dsDNA bacteriophage major capsid proteins [8]. Interesting material bearing on the T-even head size determination mechanism is provided by "recent" T-even relatives of increased and apparently invariant capsid size, unlike the T4 capsid size mutations that do not precisely determine size (e.g. KVP40, 254 kb, apparently has a single T_{mid} greater than the 170 kb T4 T_{mid} = 20) [9]. However, few if any in depth studies have been carried out on these phages to determine whether the major capsid protein, the morphogenetic core, or other factors are responsible for the different and precisely determined volumes of their capsids.

Folding of the major capsid protein gp23

Folding and assembly of the phage T4 major capsid protein gp23 into the prohead requires a special utilization of the GroEL chaperonin system and an essential phage co-chaperonin gp31. gp31 replaces the GroES co-chaperonin that is utilized for folding the 10-15% of *E. coli* proteins that require folding by the GroEL folding chamber. Although T4 gp31 and the closely related RB49 co-chaperonin CocO have been demonstrated to replace the GroES function for all essential *E. coli* protein folding, the GroES-gp31 relationship is not reciprocal; i.e. GroES cannot replace gp31 to fold gp23 because of special folding requirements of the latter protein [10,11]. The N-terminus of gp23 appears to strongly target associated fusion proteins to the GroEL chaperonin [12-14]. Binding of gp23 to the GroEL folding cage shows features that are distinct from those of most bound *E. coli* proteins. Unlike substrates such as RUBISCO, gp23 occupies both chambers of the GroEL folding cage, and only gp31 is able to promote efficient capped single "cis" chamber folding, apparently by creating a larger folding chamber [15]. On the basis of the gp24 inferred structure of gp23, and the structures of the GroES and gp31 complexed GroEL folding chambers, support for a critical increased chamber size to accommodate gp23 has been advanced as the explanation for the gp31 specificity [14]. However, since
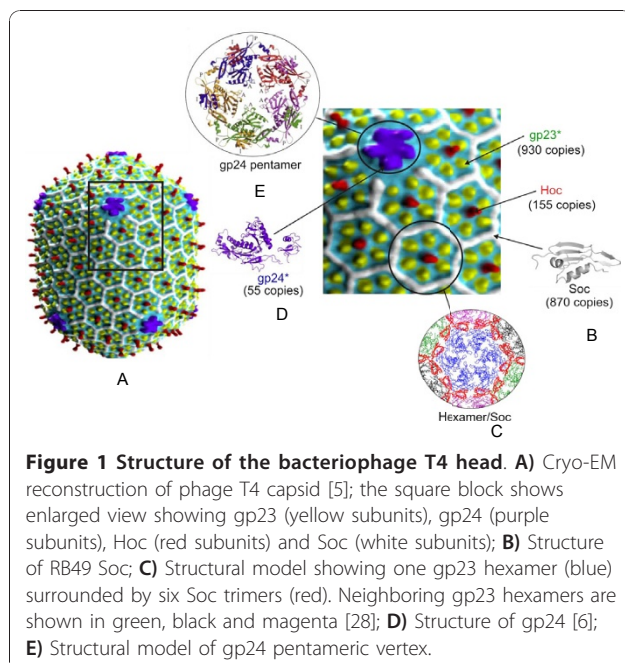
comparable size T-even phage gp31 homologs display preference for folding their own gp23s, more subtle features of the various T-even phage structured folding cages may also determine specificity.

Structure of the packaged components of the phage T4 head

Packaged phage T4 DNA shares a number of general features with other tailed dsDNA phages: 2.5 nm side to side packing of predominantly B-form duplex DNA condensed to ~500 mg/ml. However, other features differ among phages; e.g. T4 DNA is packed in an orientation that is parallel to the head tail axis together with ~1000 molecules of imbedded and mobile internal proteins, unlike the DNA arrangement that traverses head-tail axis and is arranged around an internal protein core as seen in phage T7 [16]. Use of the capsid targeting sequence of the internal proteins allows encapsidation of foreign proteins such as GFP and staphylococcal nuclease within the DNA of active virus [17,18]. Digestion by the latter nuclease upon addition of calcium yields a pattern of short DNA fragments, predominantly a 160 bp repeat [19]. This pattern supports a discontinuous pattern of DNA packing such as in the icosahedral-bend or spiral-fold models. A number of proposed models (Figure 2) and experimental evidence bearing on these are summarized in [17].

In addition to the uncertain arrangement at the nucleotide level of packaged phage DNA, the structure of other internal components is poorly understood in comparison to surface capsid proteins. The internal
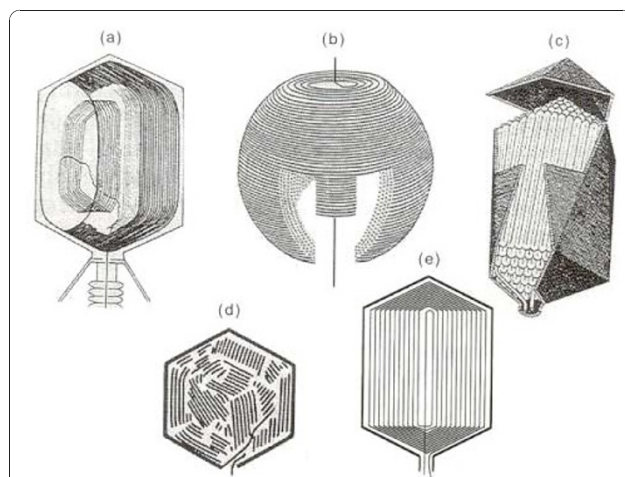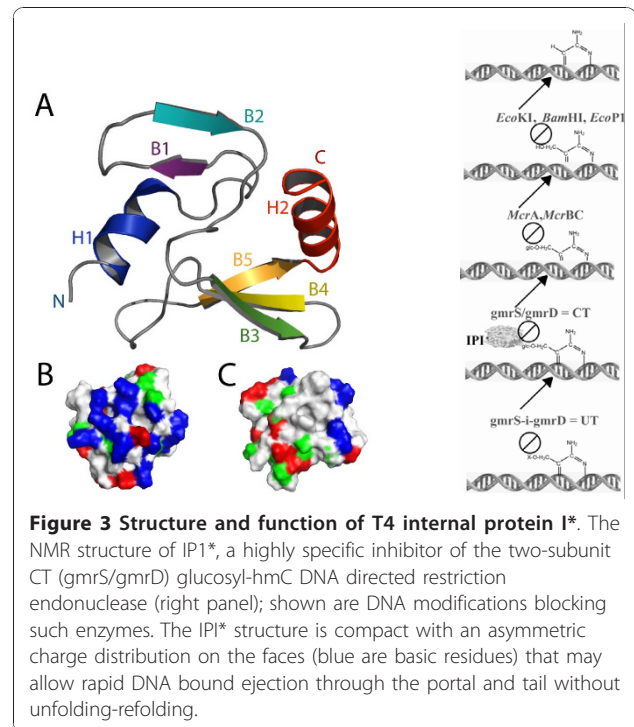

Figure 3 Structure and function of T4 internal protein I*. The NMR structure of IP1*, a highly specific inhibitor of the two-subunit CT (gmrS/gmrD) glucosyl-hmC DNA directed restriction endonuclease (right panel); shown are DNA modifications blocking such enzymes. The IPI* structure is compact with an asymmetric charge distribution on the faces (blue are basic residues) that may allow rapid DNA bound ejection through the portal and tail without unfolding-refolding.

protein I* (IPI*) of phage T4 is injected to protect the DNA from a two subunit gmrS + gmrD glucose modified restriction endonuclease of a pathogenic *E. coli* that digests glucosylated hydroxymethylcytosine DNA of T-even phages [20,21]. The 76-residue proteolyzed mature form of the protein has a novel compact protein fold consisting of two beta sheets flanked with N- and C-terminal alpha helices, a structure that is required for its inhibitor activity that is apparently due to binding the gmrS/gmrD proteins (Figure 3) [22]. A single chain gmrS/gmrD homolog enzyme with 90% identity in its sequence to the two subunit enzyme has evolved IPI* inhibitor immunity. It thus appears that the phage T-evens have co-evolved with their hosts, a diverse and highly specific set of internal proteins to counter the hmC modification dependent restriction endonucleases. Consequently the internal protein components of the T-even phages are a highly diverse set of defense proteins against diverse attack enzymes with only a conserved capsid targeting sequence (CTS) to encapsidate the proteins into the precursor scaffolding core [23].

Genes 2 and 4 of phage T4 likely are associated in function and gp2 was previously shown by Goldberg and co-workers to be able to protect the ends of mature T4 DNA from the recBCD exonuclease V, likely by binding to the DNA termini. The gp2 protein has not been identified within the phage head because of its low abundance but evidence for its presence in the head comes from the fact that gp2 can be added to gp2

Figure 2 Models of packaged DNA structure. **a)** T4 DNA is packed longitudinally to the head-tail axis [91], unlike the transverse packaging in T7 capsids [16]**(b)**. Other models shown include spiral fold **(c)**, liquid-crystal **(d)**, and icosahedral-bend **(e)**. Both packaged T4 DNA ends are located in the portal [79]. For references and evidence bearing on packaged models see [19].

deficient full heads to confer exonuclease V protection. Thus gp2 affects head-tail joining as well as protecting the DNA ends likely with as few as two copies per particle binding the two DNA ends [24].

Solid state NMR analysis of the phage T4 particle shows the DNA is largely B form and allows its electrostatic interactions to be tabulated [25]. This study reveals high resolution interactions bearing on the internal structure of the phage T4 head. The DNA phosphate negative charge is balanced among lysyl amines, polyamines, and mono and divalent cations. Interestingly, among positively charged amino acids, only lysine residues of the internal proteins were seen to be in contact with the DNA phosphates, arguing for specific internal protein DNA structures. Electrostatic contributions from internal proteins and polyamines' interactions with DNA entering the prohead to the packaging motor were proposed to account for the higher packaging rates achieved by the phage T4 packaging machine when compared to that of Phi29 and lambda phages.

Display on capsid

In addition to the essential capsid proteins, gp23, gp24, and gp20, the T4 capsid is decorated with two non-essential outer capsid proteins: Hoc (<u>h</u>ighly antigenic <u>o</u>uter <u>c</u>apsid protein), a dumbbell shaped monomer at the center of each gp23 hexon, up to 155 copies per capsid (39 kDa; red subunits); and Soc (<u>s</u>mall <u>o</u>uter <u>c</u>apsid protein), a rod-shaped molecule that binds between gp23 hexons, up to 870 copies per capsid (9 kDa; white subunits) (Figure 1). Both Hoc and Soc are dispensable, and bind to the capsid after the completion of capsid assembly [26,27]. Null (amber or deletion) mutations in either or both the genes do not affect phage production, viability, or infectivity.

The structure of Soc has recently been determined [28]. It is a tadpole shaped molecule with two binding sites for gp23*. Interaction of Soc to the two gp23 molecules glues adjacent hexons. Trimerization of the bound Soc molecules results in clamping of three hexons, and 270 such clamps form a cage reinforcing the capsid structure. Soc assembly thus provides great stability to phage T4 to survive under hostile environments such as extreme pH (pH 11), high temperature (60°C), osmotic shock, and a host of denaturing agents. Soc-minus phage lose viability at pH10.6 and addition of Soc enhances its survival by ~10^4-fold. On the other hand, Hoc does not provide significant additional stability. With its Ig-like domains exposed on the outer surface, Hoc may interact with certain components of the bacterial surface, providing additional survival advantage (Sathaliyawala and Rao, unpublished results).

The above properties of Hoc and Soc are uniquely suited to engineer the T4 capsid surface by arraying pathogen antigens. Ren et al and Jiang et al developed recombinant vectors that allowed fusion of pathogen antigens to the N- or C-termini of Hoc and Soc [29-32]. The fusion proteins were expressed in *E. coli* and upon infection with *hoc⁻soc⁻* phage, the fusion proteins assembled on the capsid. The phages purified from the infected extracts are decorated with the pathogen antigens. Alternatively, the fused gene can be transferred into T4 genome by recombinational marker rescue and infection with the recombinant phage expresses and assembles the fusion protein on the capsid as part of the infection process. Short peptides or protein domains from a variety of pathogens, *Neisseria meningitides* [32], polio virus [29], HIV [29,33], swine fever virus [34], and foot and mouth disease virus [35], have been displayed on T4 capsid using this approach.

The T4 system can be adapted to prepare bipartite libraries of randomized short peptides displayed on T4 capsid Hoc and Soc and use these libraries to "fish out" peptides that interact with the protein of interest [36]. Biopanning of libraries by the T4 large packaging protein gp17 selected peptides that matches with the sequences of proteins that are thought to interact with p17. Of particular interest was the selection of a peptide that matched with the T4 late sigma factor, gp55. The gp55 deficient extracts packaged concatemeric DNA about 100-fold less efficiently suggesting that the gp17
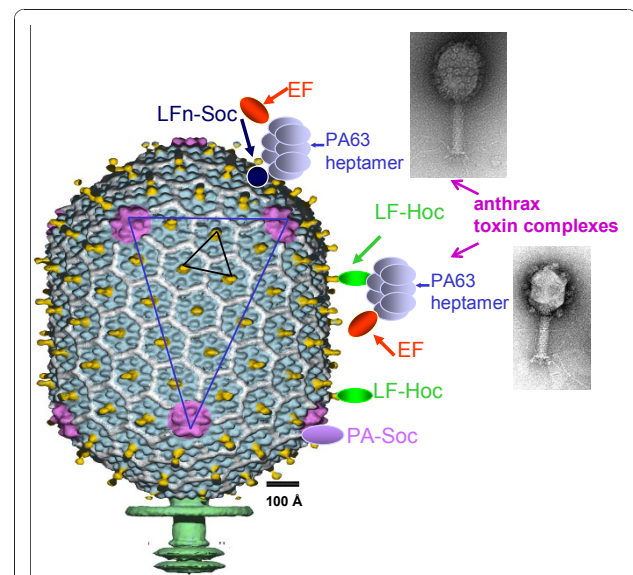

Figure 4 In vitro display of antigens on bacteriophage T4 capsid. Schematic representation of the T4 capsid decorated with large antigens, PA (83 kDa) and LF (89 kDa), or hetero-oligomeric anthrax toxin complexes through either Hoc or Soc binding [39,41]. See text for details. The insets show electron micrographs of T4 phage with the anthrax toxin complexes displayed through Soc (top) or Hoc (bottom). Note the copy number of the complexes is lower with the Hoc display than with the Soc display.

interaction with gp55 helps loading the packaging terminase onto the viral genome [36,37].

An in vitro display system has been developed taking advantage of the high affinity interactions between Hoc or Soc and the capsid (Figure 4) [38,39]. In this system, the pathogen antigen fused to Hoc or Soc with a hexa-histidine tag was overexpressed in *E. coli* and purified. The purified protein was assembled on *hoc⁻soc⁻* phage by simply mixing the purified components. This system has certain advantages over the in vivo display: i) a functionally well characterized and conformationally homogeneous antigen is displayed on the capsid; ii) the copy number of displayed antigen can be controlled by altering the ratio of antigen to capsid binding sites; and iii) multiple antigens can be displayed on the same capsid. This system was used to display full-length antigens from HIV [33] and anthrax [38,39] that are as large as 90 kDa.

All 155 Hoc binding sites can be filled with anthrax toxin antigens, protective antigen (PA, 83 kDa), lethal factor (LF, 89 kDa), or edema factor (EF, 90 kDa) [36,40]. Fusion to the N-terminus of Hoc did not affect the apparent binding constant (K_d) or the copy number per capsid (B_{max}), but fusion to the C-terminus reduced the K_d by 500-fold [32,40]. All 870 copies of Soc binding sites can be filled with Soc-fused antigens but the size of the fused antigen must be ~30 kDa or less; otherwise, the copy number is significantly reduced [39]. For example, the 20-kDa PA domain-4 and the 30 kDa LFn domain fused to Soc can be displayed to full capacity. An insoluble Soc-HIV gp120 V3 loop domain fusion protein with a 43 aa C-terminal addition could be refolded and bound with ~100% occupancy to mature phage head type-polyheads [29]. Large 90 kDa anthrax toxins can also be displayed but the B_{max} is reduced to about 300 presumably due to steric constraints. Antigens can be fused to either the N- or C-terminus, or both the termini of Soc simultaneously, without significantly affecting the K_d or B_{max}. Thus, as many as 1895 antigen molecules or domains can be attached to each capsid using both Hoc and Soc [39].

The in vitro system offers novel avenues to display macromolecular complexes through specific interactions with the already attached antigens [41]. Sequential assembly was performed by first attaching LF-Hoc and/or LFn-Soc to *hoc⁻soc⁻* phage and exposing the N-domain of LF on the surface. Heptamers of PA were then assembled through interactions between the LFn domain and the N-domain of cleaved PA (domain 1' of PA63). EF was then attached to the PA63 heptamers, completing the assembly of the ~700 kDa anthrax toxin complex on phage T4 capsid (Figure 4). CryoEM reconstruction shows that native $PA63_{(7)}$-$LFn_{(3)}$ complexes are assembled in which three adjacent capsid-bound

LFn "legs" support the PA63 heptamers [42]. Additional layers of proteins can be built on the capsid through interactions with the respective partners.

One of the main applications of the T4-antigen particles is their potential use in vaccine delivery. A number of independent studies showed that the T4-displayed particulate antigens without any added adjuvant elicit strong antibody responses, and to a lesser extent cellular responses [28,32]. The 43 aa V3 loop of HIV gp120 fused to Soc displayed on T4 phage was highly immunogenic in mice and induced anti-gp120 antibodies; so was the Soc-displayed IgG anti-EWL [29]. The Hoc fused 183 aa N-terminal portion of HIV CD4 receptor protein is displayed in active form. Strong anthrax lethal-toxin neutralization titers were elicited upon immunization of mice and rabbits with phage T4-displayed PA either through Hoc or Soc ([38,40], Rao, unpublished data). When multiple anthrax antigens were displayed, immune responses against all the displayed antigens were elicited [40]. The T4 particles displaying PA and LF, or those displaying the major antigenic determinant cluster mE2 (123 aa) and the primary antigen E2 (371 aa) of the classical swine fever virus elicited strong antibody titers [34]. Furthermore, mice immunized with the Soc displayed foot and mouth disease virus (FMDV) capsid precursor polyprotein (P1, 755 aa) and proteinase 3C (213 aa) were completely protected upon challenge with a lethal dose of FMDV [34,35]. Pigs immunized with a mixture of T4-P1 and T4-3C particles were also protected when these animals were co-housed with FMDV infected pigs. In another type of application, T4-displayed mouse Flt4 tumor antigen elicited anti-Flt4 antibodies and broke immune tolerance to self-antigens. These antibodies provided antitumor and anti-metastasis immunity in mice [43].

The above studies provide abundant evidence that the phage T4 nanoparticle platform has the potential to engineer human as well as veterinary vaccines.

DNA packaging

Two nonstructural terminase proteins, gp16 (18 kDa) and gp17 (70 kDa), link head assembly and genome processing [44-46]. These proteins are thought to form a hetero-oligomeric complex, which recognizes the concatemeric DNA and makes an endonucleolytic cut (hence the name "terminase"). The terminase-DNA complex docks on the prohead through gp17 interactions with the special portal vertex formed by the dodecameric gp20, thus assembling a DNA packaging machine. The gp49 EndoVII Holliday structure resolvase also specifically associates with the portal dodecamer thereby positioning this enzyme to repair packaging-arrested branched-structure-containing concatemers [47]. The ATP-fueled machine translocates DNA into the capsid

until the head is full, equivalent to about 1.02 times the genome length (171 kb). The terminase dissociates from the packaged head, makes a second cut to terminate DNA packaging and attaches the concatemeric DNA to another empty head to continue translocation in a processive fashion. Structural and functional analyses of the key parts of the machine - gp16, gp17, and gp20 - as described below, led to models for the packaging mechanism.
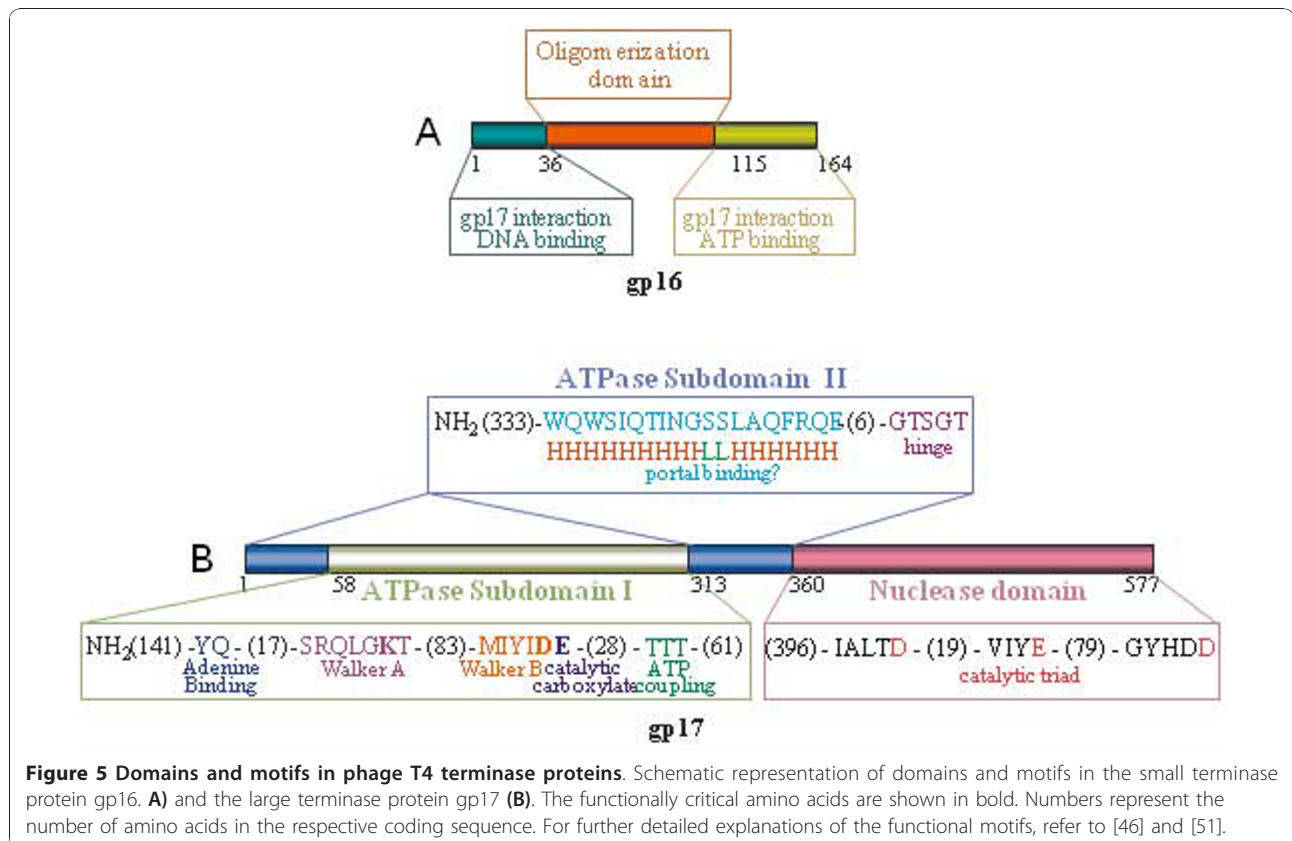
gp16

gp16, the 18 kDa small terminase subunit, is dispensable for packaging linear DNA in vitro but it is essential in vivo; amber mutations in gene 16 accumulate empty proheads resulting in null phenotype [37,48].

Mutational and biochemical analyses suggest that gp16 is involved in the recognition of viral DNA [49,50] and regulation of gp17 functions [51]. gp16 is predicted to contain three domains, a central domain that is important for oligomerization, and N- and C-terminal domains that are important for DNA binding, ATP binding, and/ or gp17-ATPase stimulation [51,52] (Figure 5). gp16 forms oligomeric single and side-by-side double rings, each ring having a diameter of ~8 nm with ~2 nm central channel [49,52]. Recent mass spectrometry determination shows that the single and double rings are 11-mers and 22-mers respectively [53]. A number of *pac* site phages produce comparable small terminase subunit multimeric ring structures. Sequence analyses predict 2-3 coiled coil motifs in gp16 [48]. All the T4 family gp16s as well as other phage small terminases consist of one or more coiled coil motifs, consistent with their propensity to form stable oligomers. Oligomerization presumably occurs through parallel coiled-coil interactions between neighboring subunits. Mutations in the long central α-helix of T4 gp16 that perturb coiled coil interactions lose the ability to oligomerize [48].

gp16 appears to oligomerize following interaction with viral DNA concatemer, forming a platform for the assembly of the large terminase gp17. A predicted helix-turn-helix in the N-terminal domain is thought to be involved in DNA-binding [49,52]. The corresponding motif in the phage lambda small terminase protein, gpNu1, has been well characterized and demonstrated to bind the DNA. In vivo genetic studies and in vitro DNA binding studies show that a 200 bp 3'-end sequence of gene 16 is a preferred "*pac*" site for gp16 interaction [49,50]. It was proposed that the stable gp16 double rings were two turn lock washers that constituted the structural basis for synapsis of two *pac* site DNAs. This could promote the gp16 dependent gene amplifications observed around the *pac* site that can be selected in *alt-* mutants that package more DNA; such

Figure 5 Domains and motifs in phage T4 terminase proteins. Schematic representation of domains and motifs in the small terminase protein gp16. **A)** and the large terminase protein gp17 **(B)**. The functionally critical amino acids are shown in bold. Numbers represent the number of amino acids in the respective coding sequence. For further detailed explanations of the functional motifs, refer to [46] and [51].

synapsis could function as a gauge of DNA concatemer maturation [54-56].

gp16 stimulates the gp17-ATPase activity by > 50-fold [57,58]. Stimulation is likely via oligomerization of gp17 which does not require gp16 association [58]. gp16 also stimulates in vitro DNA packaging activity in the crude system where phage infected extracts containing all the DNA replication/transcription/recombination proteins are present [57,59], but inhibits the packaging activity in the defined system where only two purified components, proheads and gp17, are present [37,60]. It stimulates gp17-nuclease activity when T4 transcription factors are also present but inhibits the nuclease in a pure system [51]. gp16 also inhibits gp17's binding to DNA [61]. Both the N- and C-domains are required for ATPase stimulation or nuclease inhibition [51]. Maximum effects were observed at a ratio of approximately 8 gp16 molecules to 1 gp17 molecule suggesting that in the holoterminase complex one gp16 oligomer interacts with one gp17 monomer [62].

gp16 contains an ATP binding site with broad nucleotide specificity [49,51], however it lacks the canonical nucleotide binding signatures such as Walker A and Walker B [52]. No correlation was evident between nucleotide binding and gp17-ATPase stimulation or gp17-nuclease inhibition. Thus it is unclear what the role of ATP binding plays in gp16 function.

The evidence thus far suggests that gp16 is a regulator of the DNA packaging machine, modulating the ATPase, translocase, and nuclease activities of gp17. Although the regulatory functions can be dispensable for in vitro DNA packaging, these are essential in vivo to coordinate the packaging process and produce an infectious virus particle [51].
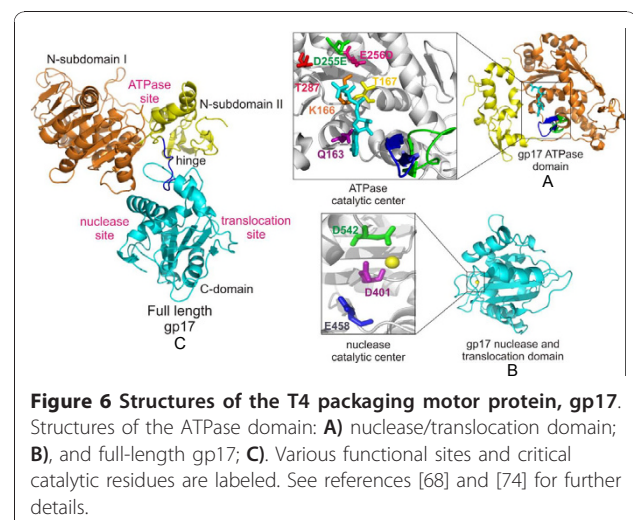
gp17

gp17 is the 70 kDa large subunit of the terminase holoenzyme and the motor protein of the DNA packaging machine. gp17 consists of two functional domains (Figure 5); an N-terminal ATPase domain having the classic ATPase signatures such as Walker A, Walker B, and catalytic carboxylate, and a C-terminal nuclease domain having a catalytic metal cluster with conserved aspartic and glutamic acid residues coordinating with Mg [62].

gp17 alone is sufficient to package DNA in vitro. gp17 exhibits a weak ATPase activity (K_{cat} = ~1-2 ATPs hydrolyzed per gp17 molecule/min), which is stimulated by > 50-fold by the small terminase protein gp16 [57,58]. Any mutation in the predicted catalytic residues of the N-terminal ATPase center results in a loss of stimulated ATPase and DNA packaging activities [63]. Even subtle conservative substitutions such as aspartic acid to glutamic acid and *vice versa* in the Walker B motif resulted in complete loss of DNA packaging

suggesting that this ATPase provides energy for DNA translocation [64,65].

The ATPase domain also exhibits DNA binding activity, which may be involved in the DNA cutting and translocation functions of the packaging motor. There is genetic evidence that gp17 may interact with gp32 [66,67], but highly purified preparations of gp17 do not show appreciable affinity for ss or ds DNA. There seem to be complex interactions between the terminase proteins, the concatemeric DNA, and the DNA replication/recombination/repair and transcription proteins that transition the DNA metabolism into the packaging phase [37].

One of the ATPase mutants, the DE-ED mutant in which the sequence of Walker B and catalytic carboxylate was reversed, showed tighter binding to ATP than the wild-type gp17 but failed to hydrolyze ATP [64]. Unlike the wild-type gp17 or the ATPase domain which failed to crystallize, the ATPase domain with the ED mutation crystallized readily, probably because it trapped the ATPase in an ATP-bound conformation. The X-ray structure of the ATPase domain was determined up to 1.8 Å resolution in different bound states; apo, ATP-bound, and ADP-bound [68]. It is a flat structure consisting of two subdomains; a large subdomain I (NsubI) and a smaller subdomain II (NsubII) forming a cleft in which ATP binds (Figure 6A). The NsubI consists of the classic nucleotide binding fold (Rossmann fold), a parallel β-sheet of six β-strands interspersed with helices. The structure showed that the predicted catalytic residues are oriented into the ATP pocket, forming a network of interactions with bound ATP. These also include an arginine finger that is proposed to trigger βγ-phosphoanhydride bond cleavage. In addition, the structure showed the movement of a loop near the adenine binding motif in response to ATP hydrolysis,

Figure 6 Structures of the T4 packaging motor protein, gp17. Structures of the ATPase domain: **A)** nuclease/translocation domain; **B)**, and full-length gp17; **C)**. Various functional sites and critical catalytic residues are labeled. See references [68] and [74] for further details.

which may be important for transduction of ATP energy into mechanical motion.

gp17 exhibits a sequence nonspecific endonuclease activity [69,70]. Random mutagenesis of gene 17 and selection of mutants that lost nuclease activity identified a histidine-rich site in the C-terminal domain being critical for DNA cleavage [71]. Extensive site-directed mutagenesis of this region combined with the sequence alignments identified a cluster of conserved aspartic acid and glutamic acid residues that are essential for DNA cleavage [72]. Unlike the ATPase mutants, these mutants retained the gp16-stimulated ATPase activity as well as the DNA packaging activity as long as the substrate is a linear molecule. However these mutants fail to package circular DNA as they are defective in cutting DNA that is required for packaging initiation.

The structure of the C-terminal nuclease domain from a T4-family phage, RB49, which has 72% sequence identity to the T4 C-domain, was determined to 1.16Å resolution [73] (Figure 6B). It has a globular structure consisting mostly of anti-parallel β-strands forming an RNase H fold that is found in resolvases, RNase Hs and integrases. As predicted from the mutagenesis studies, the structures showed that the residues D401, E458 and D542 form a catalytic triad coordinating with Mg ion. In addition the structure showed the presence of a DNA binding groove lined with a number of basic residues. The acidic catalytic metal center is buried at one end of this groove. Together, these form the nuclease cleavage site of gp17.

The crystal structure of the full-length T4 gp17 (ED mutant) was determined to 2.8Å resolution (Figure 6C) [74]. The N- and C-domain structures of the full-length gp17 superimpose with those solved using individually crystallized domains with only minor deviations. The full-length structure however has additional features that are relevant to the mechanism. A flexible "hinge" or "linker" connects the ATPase and nuclease domains. Previous biochemical studies showed that splitting gp17 into two domains at the linker retained the respective ATPase and nuclease functions but DNA translocation activity was completely lost [62]. Second, the N- and C-domains have a > 1000 square Å complementary surface area consisting of an array of five charged pairs and hydrophobic patches [74]. Third, the gp17 has a bound phosphate ion in the crystal structure. Docking of B-form DNA guided by shape and charge complementarity with one of the DNA phosphates superimposed on the bound phosphate aligns a number of basic residues, lining what appears to be a shallow translocation groove. Thus the C-domain appears to have two DNA grooves on different faces of the structure, one that aligns with the nuclease catalytic site and the second that aligns with the translocating DNA (Figure 6). Mutation of one

of the groove residues (R406) showed a novel phenotype; loss of DNA translocation activity but the ATPase and nuclease activities are retained.
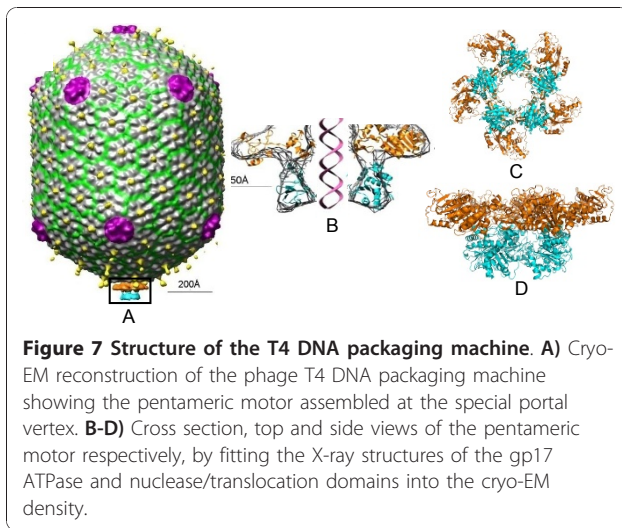
Motor

A functional DNA packaging machine could be assembled by mixing proheads and purified gp17. gp17 assembles into a packaging motor through specific interactions with the portal vertex [75] and such complexes can package the 171 kb phage T4 DNA, or any linear DNA [37,60]. If short DNA molecules are added as the DNA substrate, the motor keeps packaging DNA until the head is full [76].

Packaging can be studied in real time either by fluorescence correlation spectroscopy [77] or by optical tweezers [78]. The translocation kinetics of rhodamine (R6G) labeled 100 bp DNA was measured by determining the decrease in diffusion coefficient as the DNA gets confined inside the capsid. Fluorescence resonance energy transfer between the green fluorescent protein labeled proteins within the prohead interior and the translocated rhodamine-labeled DNA confirmed the ATP-powered movement of DNA into the capsid and the packaging of multiple segments per procapsid [77]. Analysis of FRET dye pair end labeled DNA substrates showed that upon packaging the two ends of the packaged DNA were held 8-9 nm apart in the procapsid, likely fixed in the portal channel and crown, and suggesting that a loop rather than an end of DNA is translocated following initiation at an end [79].

In the optical tweezers system, the prohead-gp17 complexes were tethered to a microsphere coated with capsid protein antibody, and the biotinylated DNA is tethered to another microsphere coated with streptavidine. The microspheres are brought together into near contact, allowing the motor to capture the DNA. Single packaging events were monitored and the dynamics of the T4 packaging process were quantified [78]. The T4 motor, like the Phi29 DNA packaging motor, generates forces as high as ~60 pN, which is ~20-25 times that of myosin ATPase and a rate as high as ~2000 bp/sec, the highest recorded to date. Slips and pauses occur but these are relatively short and rare and the motor recovers and recaptures DNA continuing translocation. The high rate of translocation is in keeping with the need to package the 171 kb size T4 genome in about 5 minutes. The T4 motor generates enormous power; when an external load of 40 pN was applied, the T4 motor translocates at a speed of ~380 bp/sec. When scaled up to a macromotor, the T4 motor is approximately twice as powerful as a typical automobile engine.

CryoEM reconstruction of the packaging machine showed two rings of density at the portal vertex [74] (Figure 7). The upper ring is flat, resembling the ATPase domain structure and the lower ring is spherical,

Figure 7 Structure of the T4 DNA packaging machine. **A)** Cryo-EM reconstruction of the phage T4 DNA packaging machine showing the pentameric motor assembled at the special portal vertex. **B-D)** Cross section, top and side views of the pentameric motor respectively, by fitting the X-ray structures of the gp17 ATPase and nuclease/translocation domains into the cryo-EM density.

resembling the C-domain structure. This was confirmed by docking of the X-ray structures of the domains into the cryoEM density. The motor has pentamer stoichiometry, with the ATP binding surface facing the portal and interacting with it. It has an open central channel that is in line with the portal channel and the translocation groove of the C-domain faces the channel. There are minimal contacts between the adjacent subunits suggesting that the ATPases may fire relatively independently during translocation.

Unlike the cryoEM structure where the two lobes (domains) of the motor are separated ("relaxed" state), the domains in the full-length gp17 are in close contact ("tensed" state) [74]. In the tensed state, the subdomain II of ATPase is rotated by 6° degrees and the C-domain is pulled upwards by 7Å, equivalent to 2 bp. The "arginine finger" located between subI and NsubII is positioned towards the βγ phosphates of ATP and the ion pairs are aligned.

Mechanism

Of many models proposed to explain the mechanism of viral DNA translocation, the portal rotation model attracted the most attention. According to the original and subsequent rotation models, the portal and DNA are locked like a nut and bolt [80,81]. The symmetry mismatch between the 5-fold capsid and 12-fold portal means that only one portal subunit aligns with one capsid subunit at any given time, causing the associated terminase-ATPase to fire causing the portal, the nut, to rotate, allowing the DNA, the bolt, to move into the capsid. Indeed, the overall structure of the dodecameric portal is well conserved in numerous bacteriophages and even in HSV, despite no significant sequence similarity. However, the X-ray structures of Phi29 and SPP1 portals did not show any rigid groove-like features that are complementary to the DNA structure [81-83]. The

structures are nevertheless consistent with the proposed portal rotation and newer, more specific, models such as the rotation-compression-relaxation [81], electrostatic gripping [82], and molecular lever [83], have been proposed.
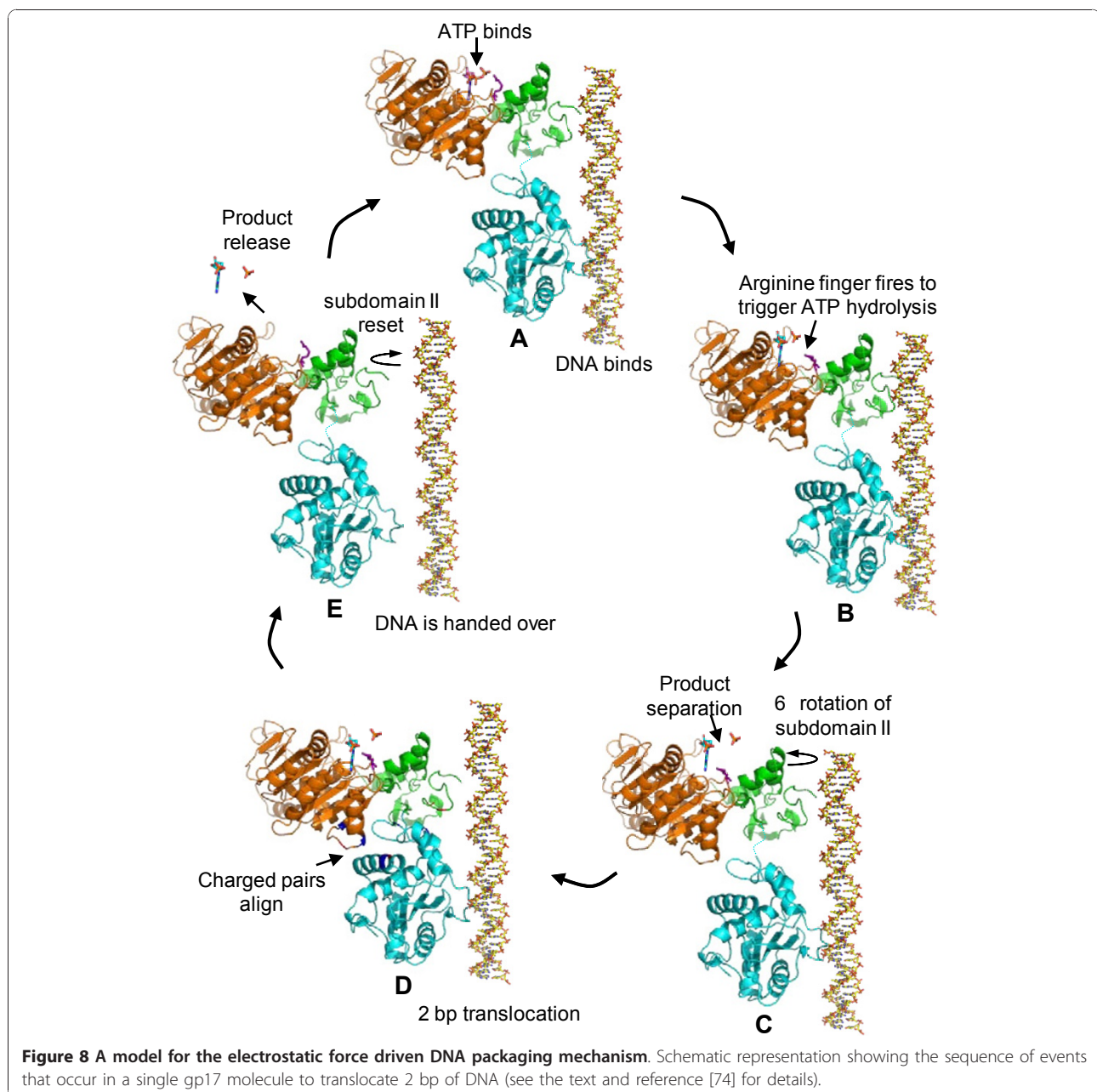
Protein fusions to either the N or C terminal end of the portal protein could be incorporated into up to ~one-half of the dodecamer positions without loss of prohead function. As compared to wild-type, portals containing C-terminal GFP fusions lock the proheads into the unexpanded conformation unless terminase packages DNA, suggesting that the portal plays a central role in controlling prohead expansion. Expansion is required to protect the packaged DNA from nuclease but not for packaging itself as measured by FCS [84]. Moreover retention of DNA packaging function of such portals argues against the portal rotation model, since rotation would require that the bulky C-terminal GFP fusion proteins within the capsid rotate through the densely packaged DNA. A more direct test tethered the portal to the capsid through Hoc interactions [85]. Hoc is a nonessential T4 outer capsid protein that binds as a monomer at the center of the major capsid protein hexon (see above; Figure 1). Hoc binding sites are not present in the unexpanded proheads but are exposed following capsid expansion. To tether the portal, unexpanded proheads were first prepared with 1 to 6 of the 12 portal subunits replaced by the N-terminal Hoc-portal fusion proteins. The proheads were then expanded in vitro to expose Hoc binding sites. The Hoc portion of the portal fusion would bind to the center of the nearest hexon, tethering 1 to 5 portal subunits to the capsid. The Hoc-capsid interaction is thought to be irreversible and thus should prevent the rotation of the portal. If portal rotation were to be central to DNA packaging, the tethered expanded proheads should show very little or no packaging activity. However, the efficiency and rate of packaging of tethered proheads were comparable to those of wild-type proheads, suggesting that portal rotation is not an obligatory requirement for packaging [85]. This was more recently confirmed by single molecule fluorescence spectroscopy of actively packaging Phi29 packaging complexes [86].

In the second class of models, the terminase not only provides the energy but also actively translocates DNA [87]. Conformational changes in the terminase domains cause changes in the DNA binding affinity resulting in binding and releasing DNA, reminiscent of the inchworm-type translocation by helicases. gp17 and numerous large terminases possess an ATPase coupling motif that is commonly present in helicases and translocases [87]. Mutations in the coupling motif present at the junction of NSubI and NSubII result in loss of ATPase and DNA packaging activities.

The cryoEM and X-ray structures (Figure 7) combined with the mutational analyses led to the postulation of a terminase-driven packaging mechanism [74]. The pentameric T4 packaging motor can be considered to be analogous to a five cylinder engine. It consists of an ATPase center in NsubI, which is the engine that provides energy. The C-domain has a translocation groove, which is the wheel that moves DNA. The smaller NsubII is the transmission domain, coupling the engine to the wheel via a flexible hinge. The arginine finger is a spark plug that fires ATPase when the motor is locked in the firing mode. Charged pairs generate electrostatic force by alternating between relaxed and tensed states (Figure 8). The nuclease groove faces away from translocating DNA and is activated when packaging is completed.

In the relaxed conformational state (cryoEM structure), the hinge is extended (Figure 8). Binding of DNA to the translocation groove and of ATP to NsubI locks the motor in translocation mode (A) and brings the arginine finger into position, firing ATP hydrolysis (B). The repulsion between the negatively charged ADP(3-) and Pi(3-) drive them apart, causing NsubII to rotate by 6° (C), aligning the charge pairs between the N- and C-

Figure 8 A model for the electrostatic force driven DNA packaging mechanism. Schematic representation showing the sequence of events that occur in a single gp17 molecule to translocate 2 bp of DNA (see the text and reference [74] for details).

domains. This generates electrostatic force, attracting the C-domain-DNA complex and causing 7Å upward movement, the tensed conformational state (X-ray structure) (D). Thus 2 bp of DNA is translocated into the capsid in one cycle. Product release and loss of 6 negative charges causes NsubII to rotate back to original position, misaligning the ion pairs and returning the C-domain to the relaxed state (E).

Translocation of 2 bp would bring the translocation groove of the adjacent subunit into alignment with the backbone phosphates. DNA is then handed over to the next subunit, by the matching motor and DNA symmetries. Thus, ATPase catalysis causes conformational changes which generate electrostatic force, which is then converted to mechanical force. The pentameric motor translocates 10 bp (one turn of the helix) when all five gp17 subunits fire in succession, bringing the first gp17 subunit once again in alignment with the DNA phosphates. Synchronized orchestration of the motor's movements translocates DNA up to ~2000 bp/sec.

Short (< 200 bp) DNA substrate translocation by gp17 is blocked by nicks, gaps, hairpin ends, RNA-containing duplexes, 20-base mismatches and D-loops, but not by 10-base internal mismatches [88]. Packaging of DNAs as short as 20 bp and initiation at almost any type DNA end suggests translocation rather than initiation deficiency of these short centrally nicked or gapped DNAs. Release from the motor of 100 bp nicked DNA segments supported a torsional compression portal-DNA-grip-and-release mechanism, where the portal grips the DNA while the gp17 imparts a linear force that may be stored in the DNA as compression or dissipated by a nick (Figure 9). Use of a DNA leader joined to a Y-DNA structure showed packaging of the leader segment; the Y-junction was arrested in proximity to a prohead portal containing GFP fusions, allowing FRET transfer between the Y-junction located dye molecule and the portal GFPs [89] (Figure 9D). Comparable stalled Y-DNA substrates containing FRET-pair dyes in the Y-stem showed that the motor compresses the stem held in the portal channel by 22-24% (Figure 9E. This finding supports the proposal that torsional compression of B DNA by the terminase motor by a portal-DNA-grip-and-release mechanism helps to drive translocation [88]. Attaching a longer DNA leader to the Y-DNA allows such abnormal structure substrates to be anchored in the procapsid for successful translocation, most likely by multiple motor cycles [89]. Differences in DNA substrate size may at least in part account for much less stringent DNA structural requirements measured in the Phi29 packaging system [90].
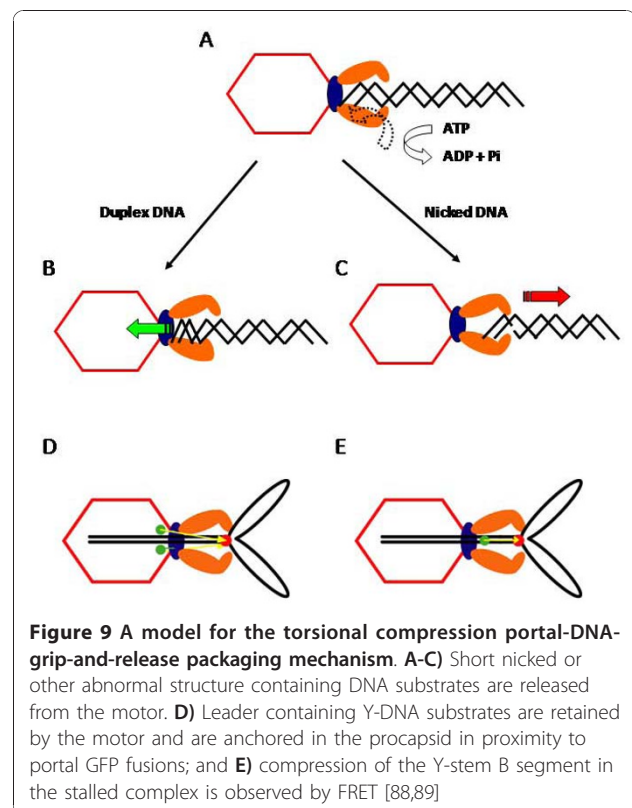

Figure 9 A model for the torsional compression portal-DNA-grip-and-release packaging mechanism. **A-C)** Short nicked or other abnormal structure containing DNA substrates are released from the motor. **D)** Leader containing Y-DNA substrates are retained by the motor and are anchored in the procapsid in proximity to portal GFP fusions; and **E)** compression of the Y-stem B segment in the stalled complex is observed by FRET [88,89]

Conclusions

It is clear from the above discussion that major advances have been made in recent years on the understanding of the phage T4 capsid structure and mechanism of DNA packaging. These advances, by combining genetics and biochemistry with structure and biophysics, set the stage to probe the packaging mechanism with even greater depth and precision. It is reasonable to hope that this would lead to the elucidation of catalytic cycle, mechanistic details, and motor dynamics to near atomic resolution. The accumulated and emerging basic knowledge should also lead to medical applications such as the development of vaccines and phage therapy.

Author details
[1]Department of Biology, The Catholic University of America, Washington, DC, USA. [2]Department of Biochemistry and Molecular Biology, University of Maryland Medical School, Baltimore, MD, USA.

Authors' contributions
VR and LWB made equal contributions to drafts of this review. Both authors revised all sections of the article and read and approved the final manuscript.

Competing interests
The authors declare that they have no competing interests.

References
1. Comeau AM, Krisch HM: **The capsid of the T4 phage superfamily: the evolution, diversity, and structure of some of the most prevalent proteins in the biosphere.** *Mol Biol Evol* 2008, **25(7)**:1321-32.
2. Krisch HM, Comeau AM: **The immense journey of bacteriophage T4–from d'Herelle to Delbruck and then to Darwin and beyond.** *Res Microbiol* 2008, **159**:314-324.
3. Tetart F, Desplats C, Krisch HM: **Genome plasticity in the distal tail fiber locus of the T-even bacteriophage: recombination between conserved motifs swaps adhesin specificity.** *J Mol Biol* 1998, **282**:543-556.
4. Black LW, Showe MK, Steven AC: **Morphogenesis of the T4 head.** 1994, 218-258.
5. Fokine A, Chipman PR, Leiman PG, Mesyanzhinov VV, Rao VB, Rossmann MG: **Molecular architecture of the prolate head of bacteriophage T4.** *Proc Natl Acad Sci USA* 2004, **101**:6003-6008.
6. 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.
7. Fokine A, Battisti AJ, Kostyuchenko VA, Black LW, Rossmann MG: **Cryo-EM structure of a bacteriophage T4 gp24 bypass mutant: the evolution of pentameric vertex proteins in icosahedral viruses.** *J Struct Biol* 2006, **154**:255-259.
8. 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.
9. Miller ES, Kutter E, Mosig G, Arisaka F, Kunisawa T, Ruger W: **Bacteriophage T4 genome.** *Microbiol Mol Biol Rev* 2003, **67**:86-156, table of contents.
10. Keppel F, Rychner M, Georgopoulos C: **Bacteriophage-encoded cochaperonins can substitute for Escherichia coli's essential GroES protein.** *EMBO Rep* 2002, **3**:893-898.
11. Andreadis JD, Black LW: **Substrate mutations that bypass a specific Cpn10 chaperonin requirement for protein folding.** *J Biol Chem* 1998, **273**:34075-34086.
12. Snyder L, Tarkowski HJ: **The N terminus of the head protein of T4 bacteriophage directs proteins to the GroEL chaperonin.** *J Mol Biol* 2005, **345**:375-386.
13. Bakkes PJ, Faber BW, van Heerikhuizen H, van der Vies SM: **The T4-encoded cochaperonin, gp31, has unique properties that explain its requirement for the folding of the T4 major capsid protein.** *Proc Natl Acad Sci USA* 2005, **102**:8144-8149.
14. Clare DK, Bakkes PJ, van Heerikhuizen H, van der Vies SM, Saibil HR: **An expanded protein folding cage in the GroEL-gp31 complex.** *J Mol Biol* 2006, **358**:905-911.
15. Clare DK, Bakkes PJ, van Heerikhuizen H, van der Vies SM, Saibil HR: **Chaperonin complex with a newly folded protein encapsulated in the folding chamber.** *Nature* 2009, **457**:107-110.
16. Cerritelli ME, Cheng N, Rosenberg AH, McPherson CE, Booy FP, Steven AC: **Encapsidated conformation of bacteriophage T7 DNA.** *Cell* 1997, **91**:271-280.
17. Mullaney JM, Thompson RB, Gryczynski Z, Black LW: **Green fluorescent protein as a probe of rotational mobility within bacteriophage T4.** *J Virol Methods* 2000, **88**:35-40.
18. Mullaney JM, Black LW: **Capsid targeting sequence targets foreign proteins into bacteriophage T4 and permits proteolytic processing.** *J Mol Biol* 1996, **261**:372-385.
19. Mullaney JM, Black LW: **Activity of foreign proteins targeted within the bacteriophage T4 head and prohead: implications for packaged DNA structure.** *J Mol Biol* 1998, **283**:913-929.
20. Bair CL, Rifat D, Black LW: **Exclusion of glucosyl-hydroxymethylcytosine DNA containing bacteriophages is overcome by the injected protein inhibitor IPI*.** *J Mol Biol* 2007, **366**:779-789.
21. Bair CL, Black LW: **A type IV modification dependent restriction nuclease that targets glucosylated hydroxymethyl cytosine modified DNAs.** *J Mol Biol* 2007, **366**:768-778.
22. Rifat D, Wright NT, Varney KM, Weber DJ, Black LW: **Restriction endonuclease inhibitor IPI* of bacteriophage T4: a novel structure for a dedicated target.** *J Mol Biol* 2008, **375**:720-734.
23. Repoila F, Tetart F, Bouet JY, Krisch HM: **Genomic polymorphism in the T-even bacteriophages.** *EMBO J* 1994, **13**:4181-4192.
24. Wang GR, Vianelli A, Goldberg EB: **Bacteriophage T4 self-assembly: in vitro reconstitution of recombinant gp2 into infectious phage.** *J Bacteriol* 2000, **182**:672-679.
25. Yu TY, Schaefer J: **REDOR NMR characterization of DNA packaging in bacteriophage T4.** *J Mol Biol* 2008, **382**:1031-1042.
26. Ishii T, Yanagida M: **The two dispensable structural proteins (soc and hoc) of the T4 phage capsid; their purification and properties, isolation and characterization of the defective mutants, and their binding with the defective heads in vitro.** *J Mol Biol* 1977, **109**:487-514.
27. Ishii T, Yamaguchi Y, Yanagida M: **Binding of the structural protein soc to the head shell of bacteriophage T4.** *J Mol Biol* 1978, **120**:533-544.
28. Qin L, Fokine A, O'Donnell E, Rao VB, Rossmann MG: **Structure of the Small Outer Capsid Protein, Soc: A Clamp for Stabilizing Capsids of T4-like Phages.** *J Mol Biol* 2010, **29;395(4)**:728-41.
29. Ren ZJ, Lewis GK, Wingfield PT, Locke EG, Steven AC, Black LW: **Phage display of intact domains at high copy number: a system based on SOC, the small outer capsid protein of bacteriophage T4.** *Protein Sci* 1996, **5**:1833-1843.
30. Ren ZJ, Baumann RG, Black LW: **Cloning of linear DNAs in vivo by overexpressed T4 DNA ligase: construction of a T4 phage hoc gene display vector.** *Gene* 1997, **195**:303-311.
31. Ren Z, Black LW: **Phage T4 SOC and HOC display of biologically active, full-length proteins on the viral capsid.** *Gene* 1998, **215**:439-444.
32. Jiang J, Abu-Shilbayeh L, Rao VB: **Display of a PorA peptide from Neisseria meningitidis on the bacteriophage T4 capsid surface.** *Infect Immun* 1997, **65**:4770-4777.
33. Sathaliyawala T, Rao M, Maclean DM, Birx DL, Alving CR, Rao VB: **Assembly of human immunodeficiency virus (HIV) antigens on bacteriophage T4: a novel in vitro approach to construct multicomponent HIV vaccines.** *J Virol* 2006, **80**:7688-7698.
34. Wu J, Tu C, Yu X, Zhang M, Zhang N, Zhao M, Nie W, Ren Z: **Bacteriophage T4 nanoparticle capsid surface SOC and HOC bipartite display with enhanced classical swine fever virus immunogenicity: a powerful immunological approach.** *J Virol Methods* 2007, **139**:50-60.
35. Ren ZJ, Tian CJ, Zhu QS, Zhao MY, Xin AG, Nie WX, Ling SR, Zhu MW, Wu JY, Lan HY, *et al*: **Orally delivered foot-and-mouth disease virus capsid protomer vaccine displayed on T4 bacteriophage surface: 100% protection from potency challenge in mice.** *Vaccine* 2008, **26**:1471-1481.
36. Malys N, Chang DY, Baumann RG, Xie D, Black LW: **A bipartite bacteriophage T4 SOC and HOC randomized peptide display library: detection and analysis of phage T4 terminase (gp17) and late sigma factor (gp55) interaction.** *J Mol Biol* 2002, **319**:289-304.
37. Black LW, Peng G: **Mechanistic coupling of bacteriophage T4 DNA packaging to components of the replication-dependent late transcription machinery.** *J Biol Chem* 2006, **281**:25635-25643.
38. Shivachandra SB, Rao M, Janosi L, Sathaliyawala T, Matyas GR, Alving CR, Leppla SH, Rao VB: **In vitro binding of anthrax protective antigen on bacteriophage T4 capsid surface through Hoc-capsid interactions: a strategy for efficient display of large full-length proteins.** *Virology* 2006, **345**:190-198.
39. Li Q, Shivachandra SB, Zhang Z, Rao VB: **Assembly of the small outer capsid protein, Soc, on bacteriophage T4: a novel system for high**

density display of multiple large anthrax toxins and foreign proteins on phage capsid. *J Mol Biol* 2007, **370**:1006-1019.

40. Shivachandra SB, Li Q, Peachman KK, Matyas GR, Leppla SH, Alving CR, Rao M, Rao VB: **Multicomponent anthrax toxin display and delivery using bacteriophage T4.** *Vaccine* 2007, **25**:1225-1235.

41. Li Q, Shivachandra SB, Leppla SH, Rao VB: **Bacteriophage T4 capsid: a unique platform for efficient surface assembly of macromolecular complexes.** *J Mol Biol* 2006, **363**:577-588.

42. Fokine A, Bowman VD, Battisti AJ, Li Q, Chipman PR, Rao VB, Rossmann MG: **Cryo-electron microscopy study of bacteriophage T4 displaying anthrax toxin proteins.** *Virology* 2007, **367**:422-427.

43. Ren SX, Ren ZJ, Zhao MY, Wang XB, Zuo SG, Yu F: **Antitumor activity of endogenous mFlt4 displayed on a T4 phage nanoparticle surface.** *Acta Pharmacol Sin* 2009, **30**:637-645.

44. Rao VB, Black LW: **DNA packaging in Bacteriophage T4.** In *"Viral Genome Packaging Machines: Genetics, Structure, and Mechanism".* Edited by: Catalano CE. Eurekah.com and Kluwer Academic/Plenum Publishers; 2006:40-58.

45. Black LW: **DNA packaging in dsDNA bacteriophages.** *Annu Rev Microbiol* 1989, **43**:267-292.

46. Rao VB, Feiss M: **The bacteriophage DNA packaging motor.** *Annu Rev Genet* 2008, **42**:647-681.

47. Golz S, Kemper B: **Association of holliday-structure resolving endonuclease VII with gp20 from the packaging machine of phage T4.** *J Mol Biol* 1999, **285**:1131-1144.

48. Kondabagil KR, Rao VB: **A critical coiled coil motif in the small terminase, gp16, from bacteriophage T4: insights into DNA packaging initiation and assembly of packaging motor.** *J Mol Biol* 2006, **358**:67-82.

49. Lin H, Simon MN, Black LW: **Purification and characterization of the small subunit of phage T4 terminase, gp16, required for DNA packaging.** *J Biol Chem* 1997, **272**:3495-3501.

50. Lin H, Black LW: **DNA requirements in vivo for phage T4 packaging.** *Virology* 1998, **242**:118-127.

51. Al-Zahrani AS, Kondabagil K, Gao S, Kelly N, Ghosh-Kumar M, Rao VB: **The small terminase, gp16, of bacteriophage T4 is a regulator of the DNA packaging motor.** *J Biol Chem* 2009, **284**:24490-24500.

52. Mitchell MS, Matsuzaki S, Imai S, Rao VB: **Sequence analysis of bacteriophage T4 DNA packaging/terminase genes 16 and 17 reveals a common ATPase center in the large subunit of viral terminases.** *Nucleic Acids Res* 2002, **30**:4009-4021.

53. Duijn EV: **Current Limitations in Native Mass Spectrometry Based Structural Biology.** *Journal of the American Chemical Society* 2010, **21(6)**:971-8.

54. Wu CH, Lin H, Black LW: **Bacteriophage T4 gene 17 amplification mutants: evidence for initiation by the T4 terminase subunit gp16.** *J Mol Biol* 1995, **247**:523-528.

55. Black LW: **DNA packaging and cutting by phage terminases: control in phage T4 by a synaptic mechanism.** *Bioessays* 1995, **17**:1025-1030.

56. Wu CH, Black LW: **Mutational analysis of the sequence-specific recombination box for amplification of gene 17 of bacteriophage T4.** *J Mol Biol* 1995, **247**:604-617.

57. Leffers G, Rao VB: **Biochemical characterization of an ATPase activity associated with the large packaging subunit gp17 from bacteriophage T4.** *J Biol Chem* 2000, **275**:37127-37136.

58. Baumann RG, Black LW: **Isolation and characterization of T4 bacteriophage gp17 terminase, a large subunit multimer with enhanced ATPase activity.** *J Biol Chem* 2003, **278**:4618-4627.

59. Rao VB, Black LW: **Cloning, overexpression and purification of the terminase proteins gp16 and gp17 of bacteriophage T4. Construction of a defined in-vitro DNA packaging system using purified terminase proteins.** *J Mol Biol* 1988, **200**:475-488.

60. Kondabagil KR, Zhang Z, Rao VB: **The DNA translocating ATPase of bacteriophage T4 packaging motor.** *J Mol Biol* 2006, **363**:786-799.

61. Alam TI, Rao VB: **The ATPase domain of the large terminase protein, gp17, from bacteriophage T4 binds DNA: implications to the DNA packaging mechanism.** *J Mol Biol* 2008, **376**:1272-1281.

62. Kanamaru S, Kondabagil K, Rossmann MG, Rao VB: **The functional domains of bacteriophage t4 terminase.** *J Biol Chem* 2004, **279**:40795-40801.

63. Rao VB, Mitchell MS: **The N-terminal ATPase site in the large terminase protein gp17 is critically required for DNA packaging in bacteriophage T4.** *J Mol Biol* 2001, **314**:401-411.

64. Mitchell MS, Rao VB: **Functional analysis of the bacteriophage T4 DNA-packaging ATPase motor.** *J Biol Chem* 2006, **281**:518-527.

65. Goetzinger KR, Rao VB: **Defining the ATPase center of bacteriophage T4 DNA packaging machine: requirement for a catalytic glutamate residue in the large terminase protein gp17.** *J Mol Biol* 2003, **331**:139-154.

66. Franklin JL, Haseltine D, Davenport L, Mosig G: **The largest (70 kDa) product of the bacteriophage T4 DNA terminase gene 17 binds to single-stranded DNA segments and digests them towards junctions with double-stranded DNA.** *J Mol Biol* 1998, **277**:541-557.

67. Mosig G: **Recombination and recombination-dependent DNA replication in bacteriophage T4.** *Annu Rev Genet* 1998, **32**:379-413.

68. Sun S, Kondabagil K, Gentz PM, Rossmann MG, Rao VB: **The structure of the ATPase that powers DNA packaging into bacteriophage T4 procapsids.** *Mol Cell* 2007, **25**:943-949.

69. Bhattacharyya SP, Rao VB: **A novel terminase activity associated with the DNA packaging protein gp17 of bacteriophage T4.** *Virology* 1993, **196**:34-44.

70. Bhattacharyya SP, Rao VB: **Structural analysis of DNA cleaved in vivo by bacteriophage T4 terminase.** *Gene* 1994, **146**:67-72.

71. Kuebler D, Rao VB: **Functional analysis of the DNA-packaging/terminase protein gp17 from bacteriophage T4.** *J Mol Biol* 1998, **281**:803-814.

72. Rentas FJ, Rao VB: **Defining the bacteriophage T4 DNA packaging machine: evidence for a C-terminal DNA cleavage domain in the large terminase/packaging protein gp17.** *J Mol Biol* 2003, **334**:37-52.

73. Alam TI, Draper B, Kondabagil K, Rentas FJ, Ghosh-Kumar M, Sun S, Rossmann MG, Rao VB: **The headful packaging nuclease of bacteriophage T4.** *Mol Microbiol* 2008, **69**:1180-1190.

74. Sun S, Kondabagil K, Draper B, Alam TI, Bowman VD, Zhang Z, Hegde S, Fokine A, Rossmann MG, Rao VB: **The structure of the phage T4 DNA packaging motor suggests a mechanism dependent on electrostatic forces.** *Cell* 2008, **135**:1251-1262.

75. Lin H, Rao VB, Black LW: **Analysis of capsid portal protein and terminase functional domains: interaction sites required for DNA packaging in bacteriophage T4.** *J Mol Biol* 1999, **289**:249-260.

76. Leffers G, Rao VB: **A discontinuous headful packaging model for packaging less than headful length DNA molecules by bacteriophage T4.** *J Mol Biol* 1996, **258**:839-850.

77. Sabanayagam CR, Oram M, Lakowicz JR, Black LW: **Viral DNA packaging studied by fluorescence correlation spectroscopy.** *Biophys J* 2007, **93**:L17-19.

78. Fuller DN, Raymer DM, Kottadiel VI, Rao VB, Smith DE: **Single phage T4 DNA packaging motors exhibit large force generation, high velocity, and dynamic variability.** *Proc Natl Acad Sci USA* 2007, **104**:16868-16873.

79. Ray K, Ma J, Oram M, Lakowicz JR, Black LW: **Single Molecule- and Fluorescence Correlation Spectroscopy-FRET Analysis of Phage DNA Packaging: Co-localization of the Packaged Phage T4 DNA Ends within the Capsid.** *J Mol Biol* 2010, **395**:1102-1113.

80. Hendrix RW: **Symmetry mismatch and DNA packaging in large bacteriophages.** *Proc Natl Acad Sci USA* 1978, **75**:4779-4783.

81. Simpson AA, Tao Y, Leiman PG, Badasso MO, He Y, Jardine PJ, Olson NH, Morais MC, Grimes S, Anderson DL, *et al*: **Structure of the bacteriophage phi29 DNA packaging motor.** *Nature* 2000, **408**:745-750.

82. Guasch A, Pous J, Ibarra B, Gomis-Ruth FX, Valpuesta JM, Sousa N, Carrascosa JL, Coll M: **Detailed architecture of a DNA translocating machine: the high-resolution structure of the bacteriophage phi29 connector particle.** *J Mol Biol* 2002, **315**:663-676.

83. Lebedev AA, Krause MH, Isidro AL, Vagin AA, Orlova EV, Turner J, Dodson EJ, Tavares P, Antson AA: **Structural framework for DNA translocation via the viral portal protein.** *EMBO J* 2007, **26**:1984-1994.

84. Ray K, Oram M, Ma J, Black LW: **Portal control of viral prohead expansion and DNA packaging.** *Virology* 2009, **391**:44-50.

85. Baumann RG, Mullaney J, Black LW: **Portal fusion protein constraints on function in DNA packaging of bacteriophage T4.** *Mol Microbiol* 2006, **61**:16-32.

86. Hugel T, Michaelis J, Hetherington CL, Jardine PJ, Grimes S, Walter JM, Falk W, Anderson DL, Bustamante C: **Experimental test of connector rotation during DNA packaging into bacteriophage phi29 capsids.** *PLoS Biol* 2007, **5**:e59.

87. Draper B, Rao VB: **An ATP hydrolysis sensor in the DNA packaging motor from bacteriophage T4 suggests an inchworm-type translocation mechanism.** *J Mol Biol* 2007, **369**:79-94.

88. Oram M, Sabanayagam C, Black LW: **Modulation of the packaging reaction of bacteriophage t4 terminase by DNA structure.** *J Mol Biol* 2008, **381**:61-72.

89. Ray K, Sabanayagam CR, Lakowicz JR, Black LW: **DNA crunching by a viral packaging motor: Compression of a procapsid-portal stalled Y-DNA substrate.** *Virology* 2010, **398**:224-232.

90. Aathavan K, Politzer AT, Kaplan A, Moffitt JR, Chemla YR, Grimes S, Jardine PJ, Anderson DL, Bustamante C: **Substrate interactions and promiscuity in a viral DNA packaging motor.** *Nature* 2009, **461**:669-673.

91. Earnshaw WC, King J, Harrison SC, Eiserling FA: **The structural organization of DNA packaged within the heads of T4 wild-type, isometric and giant bacteriophages.** *Cell* 1978, **14**:559-568.