THE CATHOLIC UNIVERSITY OF AMERICA

Mechanism of Bacteriophage T4 DNA Packaging ATPase

A DISSERTATION

Submitted to the Faculty of the

Department of Biology

School of Arts and Sciences

Of The Catholic University of America

In Partial Fulfillment of the Requirements

For the Degree

Doctor of Philosophy

By

Song Gao

Washington, D.C.

2010

Mechanism of Bacteriophage T4 DNA Packaging ATPase

Song Gao, Ph.D.

Director: Venigalla B. Rao, Ph.D.

The DNA packaging motors of bacteriophages are the most powerful force-generating biological machines reported to date. The T4 DNA packaging motor is composed of the dodecameric portal vortex protein gp20 (61 kDa) and a pentamer of the large terminase protein gp17 (70 kDa). gp17 is a weak ATPase whose activity is dramatically stimulated either by the small terminase gp16 (18 kDa) or by assembly into a packaging motor. This ATPase activity, which is mapped to the N-terminal domain of gp17 is central to the packaging mechanism and energetically coupled to DNA translocation. In this study, the small and large terminase proteins from bacteriophages RB49 and KVP40, and small terminases from other T4 family phages were studied for ATPase activity and stimulation. The nuclease and translocase activities, which are also essential for DNA packaging, were as well tested. The results show that the interactions between gp16 and gp17 are highly specific. gp17 from the T4 phage is stimulated by its own gp16 but not by a gp16 from another family member phage. Extensive domain swapping experiments to switch the specificity of one phage terminase to another showed, surprisingly, that the interaction sites are localized to both

the N- and C-terminal domains of gp16. But the most stringent specificity region was mapped to the C-terminal ~20 amino acids. For gp17, partial change in specificity was observed when the N-terminal ~80 amino acids were swapped. Moreover, M13 phage display library screening of peptides binding to T4 gp16 revealed several potential gp16 interacting sites on gp17, including a region close to the important ATPase coupling motif ($TTT_{285\text{-}287}$). Combining these results with the available structural and modeling data on gp17 and gp16, a novel model involving multiple interaction sites on both gp17 and gp16 is proposed for the mechanism of ATPase stimulation. Multiple weak interactions between binding partners, one a molecular motor (gp17) and the other a regulator (gp16), might provide the greatest flexibility to fine-tune the functions of a complex motor: assembly, ATPase, DNA translocation and headful nuclease.

This dissertation by Song Gao fulfills the dissertation requirement for the doctoral degree in biology approved by Venigalla B. Rao, Ph.D., as Director, and by James J. Greene, Ph.D., and Pamela L. Tuma, Ph.D., as Readers.

_____

Venigalla B. Rao, Ph.D., Director

_____

James J. Greene, Ph.D., Reader

_____

Pamela L. Tuma, Ph.D., Reader

Table of Contents

List of Abbreviations

aa: amino acid

Amp: ampicillin

ATP: adenosine triphosphate

BLAST: basic local alignment search tool

bp: base pair

Chl: chloramphenicol

C-terminal: carboxy terminal

Da: Dalton

DNA: deoxyribonucleic acid

dNTP: deoxynucleotide triphosphate

dsDNA: double stranded deoxyribonucleic acid

gp: gene product

IPTG: Isopropyl β-D-1-thiogalactopyranoside

Kan: kanamycin

kb: kilobase

kDa: kilodalton

L: liter

LB: Luria-Bertani

M: molar

ml: milliliter

MQ: Millicue

nM: nanomolar

N-terminal: amino terminal

PCR: polymerase chain reaction

RNA: ribonucleic acid

SDS-PAGE: sodium dodecyl sulfate polyacrylamide gel electrophoresis

SOE-PCR: splicing by overlap extension polymerase chain reaction

ssDNA: single stranded deoxyribonucleic acid

TLC: thin layer chromatography

μl: microliter

μM: micromolar

WT: wild-type

Acknowledgements

This work has been carried out at the Department of Biology, The Catholic University of America. It is a pleasure to thank the many people who made this work possible.

I am heartily thankful to my Ph.D. supervisor, Dr. Venigalla B. Rao, for his continued encouragement, guidance and support. I would also like to thank Dr. James J. Greene and Dr. Pamela L. Tuma for providing insightful comments that improved the quality of this work. Furthermore, I am deeply indebted to all student and postdoc colleagues in Dr. Rao's lab for providing a stimulating and happy environment in which to learn and grow.

Finally, I want to thank my family. I would like to express special thanks to my wife Zhihong Zhang. Without her help and encouragement, this work would not have been completed. Also, a special thought is devoted to my parents for a never-ending support.

Introduction

Bacteriophages, or phages, are viruses that infect bacteria. It is believed that they are the most abundant organisms in the biosphere, with the total population estimated to be ~10^{31} particles.[1, 2] They are thought to be very ancient, as their host organisms, bacteria, were the first, and for some time, the only inhabitants of Earth.[3] Because of their relatively small genome size, short life cycle, and simplicity of isolation, bacteriophages have served as models to elucidate basic mechanisms of molecular biology.

Many fundamental biological problems were solved with the help of phage studies. The research on bacteriophage T2 demonstrated that DNA, not protein, is the hereditary material.[4] Other basic biological findings, such as the triplet codon system, the nature of messenger RNA, recombination during DNA replication, and DNA repair mechanisms, were accomplished with the studies of phages, T2 and T4 in particular.[5] Bacteriophages were once considered to be good therapeutic agents to fight against bacterial infections, which were abandoned after the discovery of antibiotics. But as simple model systems, phage studies continued to be exploited to understand the basic mechanisms in the molecular biology field. In recent years, phage research is again emphasized, as this field is

shifting toward 'ecology-oriented'. It is becoming increasingly clear that phages, as organisms of extreme diversity, contributed to the evolution of bacteria, and more importantly, the virulence of pathogens.[3] Basic research of bacteriophages could eventually lead to novel therapeutic strategies against hard-to-treat bacterial infections.

The dsDNA tailed-phages, or *Caudovirales*, are one of the eight major phage phyla.[6] Although there's great diversity within this group, tailed-phages do have features in common, e. g. they package DNA with a common mechanism. In brief, DNA is translocated into a preformed protein procapsid with an ATP-powered molecular motor.[7] Bacteriophage T4, which infects *Escherichia coli*, is an excellent model organism of the dsDNA tailed-phages. Its entire genome is sequenced and mapped, and its detailed life cycle has been well established. And, its efficiency of infection reaches the theoretical limit, 1.0, that each phage is capable of productively infecting a bacterial cell.[5]

The bacteriophage T4 life cycle

The T4 mature virus consists of a 1150 Å-long and 850 Å-wide prolate head with hemi-icosahedral ends that is filled with the genomic DNA, a 1000 Å-long, 210

Å-diameter contractile tail, a 460 Å-diameter baseplate and six 1450 Å-long tail fibers (Fig. 1).[8-12] As previously described,[13, 14] T4 infection of its host *E. coli* starts with the attachment of the tail fibers to the lipopolysaccharide receptors in the outer membrane of the host cell. This sends a signal to the baseplate and causes conformational changes of the baseplate proteins, which opens up the end of the tail tube for the phage DNA to pass through. Further conformational changes of the protein gene product (gp) 18 on the tail sheath lead to tail contraction and penetration of the tail tube through the outer and inner membranes, and the phage genomic DNA is injected into the cytosol. Immediately after this, several early phage genes are transcribed by an *E. coli* RNA polymerase and a phage coded DNA polymerase initiates the replication of phage DNA.

The T4 DNA replication is primed with the RNA transcripts synthesized by the host RNA polymerase at specific promoter sites. Phage T4 utilizes a recombination-dependent mechanism for DNA replication, in which the 3′ end of the template strand is primed for replication by recombination with homologous sequences in adjacent DNA molecules. This origin-independent replication produces a highly branched DNA concatamer that is then resolved and packaged into the procapsids (proheads). After DNA replication, the late

genes are transcribed and the structural proteins are produced for virus assembly (see below). At the end of the 25 minutes life cycle, the host cell lyses and the mature virus particles are released into the medium (Fig. 2).

Morphogenesis of the bacteriophage T4 virion

The head, tail, and fibers of bacteriophage T4 assemble via independent sequentially ordered pathways and converge together to form a mature virion (Fig. 3).[11] Assembly of the T4 head initiates with the gp20-gp40 complex formed at the inner side of the cytoplasmic membrane.[15] Upon the formation of this initiation complex, a scaffold is built with proteins gp21, gp22, gp67, gp68, IPI, IPII, IPIII, and gpAlt. The scaffold is then coated with gp23 and gp24 to produce the prohead. The proteins gp68, IPI, IPII, IPIII, and gpAlt are not absolutely required, but rather help the correct assembly. After the prohead is fully assembled, gp21 is converted into an active protease called T4PPase by slow self-cleavage, and cleaves the amino termini of gp23, gp24, IPI, IPII, IPIII, and gpAlt and extensively digests gp22, gp21, gp67, and gp68.[15-19] The activation of gp21 only happens when it is incorporated into the assembled prohead, suggesting that an exact 3D arrangement of several gp21 subunits and, possibly, other prohead proteins is required for the activation. Post-assembly self-cleavage of the

structural proteins is a common feature that is found in many viruses, such as picornaviruses and retroviruses.

After the proteolytic step, the proheads dissociate from the cytoplasmic membrane for DNA packaging. The long, branched, 'endless' concatemeric T4 DNA is efficiently packaged into the proheads by a gp16-gp17 hetero-oligomeric complex called the terminase.[20] The T4 and other dsDNA tailed phages, such as λ, HK97, T3, T7, and P22, also undergo prohead expansion upon initiation of DNA packaging.[21, 22] T4 prohead expansion stabilizes the capsid structure and creates binding sites for gpHoc and gpSoc.[23]

Upon completion of DNA packaging (when the head is full), the terminase cuts the concatemeric DNA, separating the DNA-full head from the rest of the DNA substrate and dissociates from the head. The T4 head assembly is completed with the attachment of the neck proteins, gp13, gp14 and gp15 to the portal vertex. Subsequently, the independently assembled tail and long tail fibers are attached to produce the infectious virus particles.

DNA packaging in bacteriophage T4

DNA packaging during the assembly of tailed-phages is a highly efficient process. Bacteriophage T4 has to compact the ~171-kb 56-μm long DNA molecule within a 120 nm × 86 nm capsid shell to a near crystalline density in less than 5 minutes.[24] Studies using optical tweezers measurements of single DNA molecule packaging dynamics in phage T4 demonstrated that the T4 packaging motor generates forces >60 pN and produces a power density of ~5,000 kilowatts/m^3, implying high ATP turnover rates of >300 s^{-1}.[25] T4 DNA packaging motor is capable of translocating DNA up to ~2,000 bp/s,[25] making it the fastest and most powerful molecular motor reported to date.

The T4 DNA packaging machine is composed of the portal (gp20) and the terminase complex (gp16 and gp17).[26] The portal is assembled with twelve copies of gp20 (61 kDa) at a unique vertex of the capsid. The terminase complex, defined as the nuclease activity that generates the DNA termini for packaging, is formed with gp16 (18 kDa) and gp17 (70 kDa).[26, 27] gp17 is essentially a monomer in solution and assembles as a pentamer at the prohead portal, while gp16 forms stable ring-like oligomers.[28, 29] As shown in Fig. 4, at the beginning of the T4 DNA packaging pathway, the terminase recognizes the concatameric viral DNA, makes the first cleavage and brings the DNA end to the empty channel of the

prohead portal. DNA is then translocated into the capsid utilizing the energy of ATP hydrolysis. When one headful DNA, equivalent to a ~1.02-length of T4 genome, is packaged into the capsid, the terminase makes a second cut in the DNA and dissociates from the filled head. The terminase complex with unpackaged DNA is now ready for the next round of packaging.[26, 30, 31]

T4 DNA packaging, as the representation of tailed-phage packaging, shares mechanistic similarities to other nucleic acid processing enzymes. Study of the phage packaging system would help to elucidate the details of many important biological events, e. g. genome segregation, chromosome condensation, DNA replication and repair, and transcription.[32, 33] The implications could be very broad because these biological events are central to cellular functions in higher organisms.

The large and small terminase proteins of bacteriophage T4

The key player of the T4 DNA packaging machine is the large terminase protein, gp17. gp17 is the catalytic component of the machine, or the motor, which possesses the ATPase, nuclease and translocase activities that are essential for DNA packaging.[30, 34] The ATPase activity, which is weak by gp17 itself, is central

for DNA packaging, and is stimulated during active packaging as a portal-assembled motor, or by the small terminase, gp16.[34, 35] Structural work and biochemical studies have determined the domain organization of the protein as well as critical functional motifs (Fig. 5).[29, 30, 36-40] gp17 consists of an N-terminal ATPase domain (N360) and a C-terminal nuclease domain (C360). The ATPase domain has two subdomains, the larger subdomain I and the smaller subdomain II. The subdomain I encodes the classic ATPase signatures: the adenine binding signature YQ (aa 142-143), Walker A ($SRQLGKT_{161–167}$), Walker B ($MIYID_{251–255}$), catalytic carboxylate (E256), and the ATPase coupling motif (C-motif, $TTT_{285–287}$). Mutagenesis work of Walker A showed that no substitutions are tolerated at the highly conserved GKT signature. Conservative substitutions (e. g. G165A, K166R, and T167A) resulted in a null phenotype. Biochemical analyses showed that the Walker A mutants only retained the nuclease activity but the stimulated ATPase and DNA packaging activities were lost.[37] For Walker B, any mutations at the conserved Asp255 resulted in a null phenotype. The mutants could bind ATP but completely lost stimulated ATPase and DNA packaging activities.[41] Also, no substitutions are tolerated at the conserved catalytic carboxylate, E256.[42] The C-motif mutants (Thr287Ala/Asp) were deficient in turnover of ATP hydrolysis, thus losing stimulated steady-state ATPase and DNA translocation activities.[43]

The last 33 amino acids (578-610) of T4 gp17 are not required for its functions; deleting these residues renders it protease resistant during protein purification.[44] So the K577 truncated construct is used in this study as the T4 gp17 wild-type.

The N-terminal RecA-like nucleotide binding subdomain I and the subdomain II form an ATP binding cleft and hydrolyze ATP to provide energy required for DNA translocation.[40] An 'arginine finger positioning' mechanism was proposed for the stimulation of this ATPase activity. An arginine (Arg-162) was found to be appropriately located at the junction of ATPase subdomains I and II. On the other hand, the C-terminal nuclease domain has an RNaseH fold closely resembling resolvases and a putative DNA translocation groove.[29] The C-terminal domain is connected with the N-terminal domain by a flexible hinge.

The Cryo-EM reconstruction of the T4 DNA packaging motor indicates that the motor consists of five subunits of gp17 forming a continuous channel with the portal channel through which DNA is translocated into the capsid.[29] During packaging, the putative DNA translocation groove on the C-terminal domain grabs the DNA for translocation. The nuclease center facing the outside of the motor is responsible for the DNA cleavages at the time of packaging initiation

and termination.

Different conformational states of the T4 packaging motor were identified from the data obtained by X-ray structural determination and cryo-electron microscopy reconstruction. These discoveries led to the proposal of an electrostatic force-dependent DNA packaging mechanism.[29] The two different conformational states of the motor are the tensed state and the relaxed state. In the relaxed state, the flexible linker between the N- and C- terminal domains is in an extended conformation (Fig. 6). Upon ATP binding to the N-terminal domain and dsDNA binding to the C-terminal domain on the same gp17 molecule, conformational changes occur that are transmitted by the N-Subdomain II, resulting in the insertion of Arg-162 into the ATPase active center to trigger ATP hydrolysis. ATP hydrolysis causes the rotation of the N-Subdomain II by 6°, resulting in the alignment of an array of complementary charged amino acid pairs and hydrophobic surfaces at the N- and C-domain interface. The electrostatic force between opposing charges attracts the C-domain toward the N-domain/procapsid by 7 Å (and the motor is in the tensed state, see Fig. 6), which is equivalent to the translocation of two base pairs of DNA. When the ADP^{3-} and Pi^{3-} products are released, the net loss of six negative charges causes

the N-Subdomain II to rotate back to its original orientation, attenuating the electrostatic interactions between the N- and C-terminal domains. The C-terminal domain returns to its original relaxed position. The movement of DNA by two base pairs places the DNA phosphates in an optimal position to align with the DNA translocation groove of the adjacent gp17 C-domain, and the cycle is repeated.

During DNA translocation, the gp17-ATPase is stimulated and coupled with DNA movement, while the nuclease activity is inactive. After one headful DNA is packaged, gp17 receives the 'headful' signal, dissembles from the capsid, and makes the termination cut.[31] Thus, in various DNA packaging steps, the ATPase, translocase, and nuclease activities of the packaging motor are precisely regulated and coordinated.

The small terminase protein, gp16, is the other component of the T4 terminase complex. Sequence alignments, mutational and biochemical analyses show that it contains three functional domains, a central domain (amino acids 36–115) for oligomerization, and an N-terminal domain (amino acids 1–35) and a C-terminal domain (amino acids 116–164) for activity[38, 45] (Fig. 5). A DNA binding site is

predicted in the N-terminal domain, and the C-domain is shown to be required for ATP binding.[28, 38, 45, 46] ATP binding activity is also reported on other phage small terminase proteins, such as gpNu1 (λ) and gp1 (SPP1), and is hypothesized to serve as an ATP reservoir for the respective large terminases.[47, 48] The central domain is not directly involved in functional activities, but rather provides the structural platform for the interaction partners (gp17 and/or DNA).[45]

The T4 DNA packaging machine consists of a motor and a regulator

Although gp16 is dispensable for DNA packaging *in vitro*, it is essential for phage viability.[41] It has been established that in *cos* phages and *pac* phages, the small terminase recognizes a specific viral genome sequence and recruits the large terminase to form a holoenzyme complex. And the nuclease activity of the large terminase generates the DNA end at a nearby site to initiate packaging.[49] Though studies only showed weak nonspecific binding of T4 gp16 to DNA, it is still thought to participate in the genome recognition step.[26, 28]

Evidence further suggested that gp16 plays broader roles throughout the packaging process. It stimulates the gp17-ATPase activity to ~50 fold by triggering the hydrolysis of gp17-bound ATP in a manner similar to that of the

GTPase-activating proteins (GAPs).[38] gp17-nuclease activity is restricted by gp16 to the putative packaging initiation cut, and extensive random DNA cutting is inhibited.[38] gp16 inhibits gp17-dependent DNA packaging *in vitro* but enhances it *in vivo*.[34, 35] These data suggest that gp16 is a regulator of the T4 DNA packaging motor.

The regulation of gp16 on the packaging motor gp17 is important for successful DNA packaging and production of infectious phage particles. As described above, gp17 is a flexible molecule with a number of domains and binding sites, undergoing dynamic movements and conformational transitions during the packaging process.[41] Its activities have to be precisely coordinated to perform different functions at different stages of DNA packaging. How a small molecule, gp16, regulates the various functions of the packaging motor is a fascinating biological question. Here the large and small terminase proteins of T4 family phages were systematically studied to dissect the various functional properties of the small terminase. The specificity of the gp16-gp17 interactions is established and the potential regions involved in the interactions are identified. Models for the mechanism of regulation of the complex DNA packaging motor are discussed.

Materials and Methods

Bacteria, DNAs and phage

E. coli XL-10 gold strain (Stratagene, La Jolla, CA) was used for preservation of the T4 family g17 and g16 wild-type and recombinant plasmids. *E. coli* BL21 (DE3) pLysS strain (Stratagene) which produces chloramphenicol resistance was used for overexpression of the T4 family terminase proteins and recombinants. This strain was also used in the *in vivo* nuclease assay. The expression vector pET-15b (Novagen, San Diego, CA) was used for construction of T4 family large terminase proteins and recombinants. The expression vector pET-28b (Novagen) was used for construction of T4 family small terminase proteins and recombinants. These two expression vectors contain antibiotic resistance markers (ampicillin and kanamycin, respectively) for gene insertion selection, a phage T7 promoter for inducible expression of the inserted gene, and a hexahistidine sequence upstream of the multiple cloning sites. Phenol-chloroform extraction purified T4 family phage DNAs were available in the laboratory and were used as templates for amplification of g17 and g16 sequences. *E. coli* P301 (*sup*$^-$) strain was used for preparation of T4 phage proheads, with the infection of the *17am18am-rII* packaging defective mutant phage. *17am18am-rII* was constructed by standard phage crosses.[50] The 29 kb P1-pBR hybrid vector, pAD10, was used as the

substrate of the *in vitro* nuclease assay. Construction of pAD10 plasmid was described earlier,[51] and the plasmid DNA was linearized by *BamHI* digestion. Phage λ DNA and 500 bp linear DNA used in the DNA packaging assays were purchased from Fermentas, Glen Burnie, MD. The Ph.D.-12 phage display peptide library and *E. coli* ER2738 host strain used in biopanning were purchased from New England Biolabs, Ipswich, MA.

Sequences and alignments

The amino acid sequences of T4, RB49 and KVP40 terminase proteins were obtained from GenBank. Searches of other T4 family phage terminase sequences were performed using BLAST from the National Center for Biotechnology Information (NCBI) database. Sequences were analyzed by DNAMAN software (Lynnon Corporation, Quebec, Canada). DNAMAN sequence alignments were conducted with default parameter settings.

Cloning and overexpression of T4 family terminase proteins and mutants

PCR

The *g17* and *g16* coding sequences of T4 family phages were amplified by Polymerase Chain Reaction (PCR) with primers corresponding to the ends of

respective genes. Each reaction was performed according to the procedure provided with the Phusion PCR Master Mix (New England Biolabs). In brief, the reactions were set up in 50 μl volume containing 25 μl of the 2x Master Mix, forward and reverse primers at a final concentration of 0.5 μM and ~100 ng of purified mature phage DNA as templates. The cycling conditions were 98°C 5-10 s for denaturation, T_m + 3°C 10-30 s for annealing and 72°C 15-30 s/ 1 kb for extension for 25-35 cycles. PCR amplification was verified with agarose gel electrophoresis.

The terminase mutant clones constructed in this study include N- and C-terminal truncations, fusion proteins and point mutants. The truncations were constructed by PCR amplification of desired part of terminase DNA sequence with specific primers. Most fusions and point mutants were constructed using the PCR-directed splicing by-overlap-extension (SOE PCR) strategy.[52] In general, DNA fragments from the genes that are to be fused, or sequences before and after a mutation site, are generated in separate PCRs. The stitch primers are designed so that the ends of the PCR products for recombination contain complementary sequences. If constructing a mutant, mutations are included in the stitch primers as well. When these PCR products are mixed, denatured and reannealed in a stitch PCR reaction, the strands having the matching sequences at their 3' ends

(including mutations if introduced) overlap and act as primers for each other. Extension of this overlap produces a molecule in which the original sequences are 'spliced' together or point mutations are incorporated. When polypeptides are to be fused or the sites to be mutated were close to the termini of a gene, SOE PCR was not used, and altered sequences were directly introduced as part of the forward or reverse primers. The primers used to construct the T4 family terminase protein and mutant clones are listed in Appendix 1. Restriction enzyme sites were added to the 5' ends of the forward and reverse primers to facilitate efficient insertion into the vectors.

Purification of PCR products

The amplified DNA fragments were purified by ammonium acetate/isopropanol precipitation. An equal volume of 8 M ammonium acetate and two volumes of isopropanol were added to each PCR mixture for precipitation at room temperature for 20-30 min. DNA fragments were sedimented by high-speed centrifugation in a bench-top centrifuge. The DNA pellets were washed twice with 78% ethanol, air dried at 37°C for 2 min and resuspended in 30-50 μl of sterile Milli-Q water.

Restriction enzyme digestion to generate 'sticky ends' for ligation

Purified DNA was put in a restriction enzyme digestion system at 37°C for 16 hours. The system includes purified DNA, appropriate amount of 10X restriction digestion buffer, restriction enzyme(s) and sterile Milli-Q water. The restriction enzymes were heat-inactivated at 67°C for 20 min after digestion. gp17 DNA fragments were digested with *BamHI* (New England Biolabs) for insertion into pET-15b vector. gp16 DNA fragments were double-digested with *NdeI* (New England Biolabs) and *BamHI* for in-frame insertion into pET-28b vector.

Purification of DNA fragments from restriction digestion

The restriction digestion mixtures were mixed with appropriate amount of 6X agarose gel loading buffer and electrophoresed on a 0.8% agarose gel. The appropriate DNA bands were visualized with ultraviolet light and were sliced out. The DNA was extracted using the Qiagen gel extraction protocol. In brief, 3 volumes of Buffer QG were added to 1 volume of gel to dissolve at 50°C for 10 min. 1 gel volume of isopropanol was added and the dissolved DNA was applied to the QIAquick Spin Column and centrifuged for 30-60 sec. The bound DNA was washed with 0.75 ml Buffer PE and eluted with 30 μl Buffer EB or sterile Milli-Q water. To confirm the concentration, 2 μl of extracted DNA inserts were

analyzed by agarose gel electrophoresis.

Ligation of gp17 and gp16 fragments into phage T7 expression vectors

The phage T7 expression vectors (pET-15b and pET-28b) were linearized by single- or double-digestion with restriction enzymes and dephosphorylated with alkaline phosphatase (New England Biolabs). pET-15b was single-digested with *BamHI* for insertion of gp17 DNAs. pET-28b was double-digested with *NdeI* and *BamHI* for orientation-specific insertion of gp16 DNAs. The prepared vectors were incubated with respective DNA inserts at a molar ratio of 1:3-1:10 (vector: insert) and appropriate amount of 10X ligase buffer and T4 DNA ligase (Fermentas) for 2 hours at 22°C. Successful ligation was confirmed by agarose gel electrophoresis.

Transformation for plasmid maintenance

The ligated DNAs were transformed into *E. coli* XL10 Gold competent cells for plasmid maintenance. For each transformation, 100 μl of competent cells were thawed and aliquoted into a pre-chilled Falcon 2095 polypropylene tube. 4 μl of β- mercaptoethanol was added to the aliquot and the tube was incubated on ice for 10 min with gentle swirling every 2 min. 2 μl of a ligation mixture was added

to the aliquot and mixed by gentle swirling. After 30 min incubation on ice, the cells were heat-pulsed in a 42°C water bath for 30 sec and immediately cooled on ice for 2 min. 0.9 ml of preheated (42°C) SOC medium was added followed by an incubation period of 1 hour (for ampicillin resistant pET-15b) or 3 hours (for kanamycin resistant pET-28b) at 37°C shaking at 250 rpm. 50 μl or 200 μl of the transformation mixture was plated on LB agar plates containing appropriate antibiotic (50 μg/ml ampicillin or 35 μg/ml kanamycin). Plates were incubated at 37°C overnight. Individual colonies were picked and grown in small scale LB medium containing the appropriate antibiotic for transformant preservation (in 50% glycerol stock) and plasmid mini-preparation.

Plasmid mini-preparation

Mini-prep plasmid DNAs were prepared by the alkaline lysis procedure using Qiagen buffers. Approximately 1.5 ml cell culture was centrifuged in a bench-top centrifuge to remove the culture medium. The cell pellet was completely resuspended in 0.2 ml of the Cell Resuspension Buffer (P1) by vortex. 0.2 ml of the Cell Lysis Buffer (P2) was added and mixed by inverting the tube 4-6 times. After incubation for 2-5 min, 0.2 ml of the Neutralization Buffer (P3) was added and immediately mixed by inverting the tube 4-6 times. The tube was put for

centrifugation at ~12,000 x g in a bench-top centrifuge for 15 min to remove the white precipitate. The plasmid containing supernatant was retained and mixed with an equal amount of isopropanol and incubated at room temperature for 20 min to precipitate the DNA. After incubation, the DNA was pelleted by centrifugation at ~12,000 x g for 20 min and washed twice with 78% ethanol. The plasmid DNA pellet was air dried at 37°C for 2 min and resuspended in 30-50 μl of sterile Milli-Q water.

DNA insert and orientation test

PCR reactions were performed to check the insertion of *g17* and *g16* DNA sequences into vectors. Also, the orientation of *g17* insertions was tested in this step. The mini-prep plasmids were used as the PCR template. The forward primer was the T7 promoter primer (5'-TAATACGACTCACTATAGGG-3', upstream of the insertion site for both pET-15b and pET-28b) and the reverse primers were the specific reverse primers of the inserts. The PCR results were analyzed on an agarose gel and only the plasmids with inserts in the right orientation would give positive results.

DNA sequencing

Plasmid DNA samples with inserts in the right orientation were commercially sequenced (Macrogen, Inc., Rockville, MD or Davis Sequencing, Inc., Davis, CA) to confirm the in-frame insertion and absence of any mutation of the inserted DNA sequence. The T7 promoter primer was used for sequencing of gp16 inserts. Both the T7 promoter primer and the T7 terminator primer (5'-GCTAGTTATTGCTCAGCGG-3', downstream of the insertion site for both pET-15b and pET-28b) were used for sequencing of gp17 inserts.

Transformation for protein expression

Appropriate plasmids were transformed into *E. coli* BL21 (DE3) pLysS competent cells for protein expression. For each transformation, 20 μl of competent cells were thawed and aliquoted into a pre-chilled 1.5 ml polypropylene microcentrifuge tube. 1 μl of a plasmid DNA was added to the aliquot and mixed by gentle stirring. After 5 min incubation on ice, the cells were heat-pulsed in a 42°C water bath for 30 sec and immediately cooled on ice for 2 min. 80 μl of room temperature SOC medium was added followed by an incubation period of 1 hour at 37°C shaking at 250 rpm. 50 μl of the transformation mixture was plated on LB agar plates containing appropriate antibiotic (50 μg/ml ampicillin or 35 μg/ml kanamycin, plus 34 μg/ml chloramphenicol). Plates were incubated at

37°C overnight. Individual colonies were picked and grown in small scale LB medium containing appropriate antibiotic for transformant preservation (in 50% glycerol stock).

Small scale overexpression of terminase proteins and mutants

Small scale overexpression of the terminase proteins and mutants were done in 20-50 ml of Moore medium (2% tryptone, 1.5% yeast extract, 0.2% dextrose, 0.8% NaCl, 0.2% Na_2HPO_4 and 0.1% KH_2PO_4) with the appropriate antibiotic. For each protein, the small scale LB culture prepared above was inoculated freshly into Moore medium at a ratio of 1:50. The culture was allowed to grow at 30°C shaking at 250 rpm until the cells reached log phase (~4×10^8 cells/ml). Protein expression was induced by addition of isopropyl-1-thio-β-D-galactopyranoside (IPTG) to a final concentration of 1 mM. The induction period was 2 hours at 30°C shaking at 250 rpm. 0.5 ml aliquots were retained before and 2 hours after the addition of IPTG for sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) analysis. The SDS-PAGE samples were prepared by pelleting out the cells from the 0.5 ml aliquots with centrifuge and resuspension with sterile Milli-Q water plus SDS sample buffer. The samples were boiled for 5 min and cooled on ice before loading.

Solubility test

The solubility of terminase proteins and mutants were tested with B-PER Bacterial Protein Extraction Reagent (Pierce, Rockford, IL). For each protein, 1.5 ml of cell culture induced for 2 hours as described above was centrifuged to pellet the cells. 300 μl of the B-PER Reagent was used to completely resuspend the cell pellet by vigorously vortexing. The cells were vortexed for an additional minute to ensure homogeneous resuspension. The cell suspension was centrifuged at ~12,000 x g in a bench-top centrifuge for 5 min to separate the soluble proteins from the insoluble proteins. The soluble fraction (supernatant) was collected and the insoluble fraction (pellet) was resuspended in 300 μl of the B-PER Reagent. 10 μl each of the soluble and insoluble fraction was used for SDS-PAGE to determine the solubility of overexpressed protein.

Purification of T4 family terminase proteins and mutants

Large scale overexpression of terminase proteins and mutants

For each terminase protein or mutant, the expression *E. coli* strain BL21 (DE3) pLysS containing a given terminase clone from transformant preservations was streak-purified to inoculate 20-50 ml Moore medium with appropriate antibiotic. This culture was grown at 37°C shaking at 250 rpm for 6-8 hours and inoculated

freshly into 1-2 L of Moore medium at a ratio of 1:50. The large culture was grown at 30°C shaking at 175-225 rpm until the cells reached log phase (~4×10^8 cells/ml). IPTG was added to a final concentration of 1 mM for an induction period of 2.5-3.5 hours. 0.5 ml culture aliquots were retained before and after induction to verify overexpression of the terminase by SDS-PAGE. Cells were harvested by centrifugation at 8,200 x g for 15 min at 4°C. The cell pellet was either used fresh or frozen at -70°C.

French Press cell lysis

The cell pellet from large volume culture was resuspended to homogeneity in 40-80 ml lysis/binding buffer (20 mM Tris-HCl pH 8.0, 300 mM NaCl, 20 mM imidazole, 5 mM $MgCl_2$ and 0.5 mM ATP). 1 or 2 protease inhibitor cocktail tablets (Complete, Mini, EDTA-free; Roche Applied Science, Indianapolis, IN) were added to the resuspension to protect the expressed terminase from nonspecific proteolysis. The cells were lysed completely with a constant pressure of ~18,000 psi provided by a French Press motor and pressure cell (Aminco; Thermo Fisher Scientific Inc., Waltham, MA). The cell lysate was centrifuged at 34,000 x g for 20 min at 4°C to separate the soluble fraction (supernatant) from the insoluble fraction (pellet). Small amounts of supernatant and pellet

suspension were analyzed by SDS-PAGE to ensure the presence of expressed terminase in the soluble fraction. The supernatant was filtered with a 0.20 μm filter (Minisart; Sartorius Stedim Biotech, Aubagne Cedex, France) for subsequent purification steps.

Chromatography

The soluble terminase proteins and mutants were purified from the French Press supernatant by successive chromatography on Histrap HP (affinity for the hexahistidine tag) and Hiload Superdex 200 prep-grade (size exclusion) columns using AKTA-PRIME and AKTA-FPLC systems (GE Healthcare, Pittsburgh, PA). The supernatant was loaded onto a Histrap HP column pre-equilibrated with the lysis/binding buffer. After sample loading, the column was washed with wash buffer (20 mM Tris-HCl pH 8.0, 300 mM NaCl, 50 mM imidazole, 5 mM $MgCl_2$ and 0.5 mM ATP) to remove any unbound proteins. The His-tagged protein was eluted with a 50 mM to 400 mM imidazole gradient of elution buffer (20 mM Tris-HCl pH 8.0, 100 mM NaCl, 400 mM imidazole, 5 mM $MgCl_2$ and 0.5 mM ATP). Peak fractions were pooled and EDTA was added to a final concentration of 5 mM. Pooled fractions were concentrated by an Amicon Ultra-15 centrifugal filter unit (5 kDa cut off; Millipore, Temecula, CA) and purified on a HiLoad

16/60 Superdex 200 prep-grade gel filtration column pre-equilibrated with 20 mM Tris-HCl and 100 mM NaCl. Small amounts of protein samples were retained from each step for SDS-PAGE analyses. The purified proteins were further concentrated and stored at -70°C as small aliquots.

Concentration determination of purified proteins

Protein concentrations were determined by the Bradford assay (Bio-Rad, Hercules, CA). Briefly, the protein sample was mixed with the Bio-Rad Protein Assay Dye Reagent and the absorbance at 595 nm was compared with a standard curve from the optical density plotting against varying acetylated BSA (USB Corporation, Cleveland, Ohio) concentrations (SmartSpec Plus Spectrophotometer, Bio-Rad). Protein concentrations were cross-checked by laser densitometric quantification of the Coomassie blue staining of the protein bands from SDS-PAGE, using acetylated BSA as the protein standard (Personal Densitometer, GE Healthcare).

Oligomeric states of terminase proteins

Different oligomeric species of purified proteins were separated with native gradient (4-12% or 4-20%, w/v) polyacrylamide gel electrophoresis (PAGE) under

non-denaturing conditions.

Assaying gp16 bound DNA

Freshly purified gp16 (1-2 mg/ml) was treated with Benzonase nuclease (Novagen) at a concentration of 200 U/ml (excessive amount) for overnight at room temperature. The treated gp16 sample was concentrated to < 1 ml volume by an Amicon Ultra-15 centrifugal filter unit (Millipore) and was loaded on a HiLoad 16/60 Superdex 200 prep-grade gel filtration column pre-equilibrated with 20 mM Tris-HCl and 100 mM NaCl. Peak fractions were pooled and Proteinase K (Fermentas) (10 U per 1 mg of gp16, excessive amount) was added to digest the protein and release the nuclease protected DNA at 65°C for one hour. The released DNA was purified with the GeneJET™ PCR Purification Kit (Fermentas) as per the manufacture's protocol. The purified gp16 bound DNA was stored at -20°C. Samples were taken from each step for PAGE/agarose gel analysis.

***In vivo* nuclease assay**

The *E. coli* BL21 (DE3) pLysS cells containing cloned large terminase constructs were streak-purified to inoculate Moore medium with the appropriate antibiotic.

The cells were grown at 37°C shaking at 250 rpm to ~4×10^8 cells/ml, and IPTG was added to a final concentration of 1 mM for induction. 1.5 ml culture aliquots were collected immediately before induction and at 60 and 120 min after induction. Half of each aliquot was used to verify the overexpression of the terminases by SDS-PAGE. The remaining half of each aliquot was centrifuged at 3,000 x g for 5 min and the pellet was subjected to plasmid DNA preparation by the alkaline lysis procedure. The prepared plasmid DNAs were electrophoresed on a 0.8% (w/v) agarose gel. Uninduced cells or induced cells containing nuclease-deficient constructs show no cleavage of plasmid DNA. Induced nuclease-proficient constructs cleave resident plasmid and *E. coli* genomic DNAs and yield a DNA smear throughout the lane upon electrophoresis.[44]

ATPase assays

Procedures of ATPase assays were developed according to Leffers and Rao.[34] The purified gp17s (0.2-2 μM), either alone or with gp16 (at a gp16:gp17 molar ratio of 10:1), were incubated at 37°C for 20 min in a 20 μl reaction mixture. The reaction mixture contained 0.05-2 mM unlabeled (cold) ATP and 75 nM [γ-^{32}P]ATP (specific activity 3000 Ci/mmol; GE Healthcare) in ATPase buffer (50 mM Tris-HCl pH 7.5, 0.1 M NaCl, 5 mM $MgCl_2$). EDTA was added to a final

concentration of 50 mM to stop the reaction, and the ATP hydrolysis products (0.6-1.0 μl aliquot from each reaction) were separated by thin layer chromatography (TLC) on 10×10 cm polyethyleneimine-cellulose plates (Sigma-Aldrich, St. Louis, MO). The reaction mixtures were spotted at about 1.5 cm from the bottom of the plates and the solvent (0.5 M LiCl and 1 M formic acid) was allowed to migrate to the top in a TLC chamber. The plates were air-dried for autoradiography and phosphorimaging (Storm 820, Molecular Dynamics, GE Healthcare). The radioactive signals were quantified with ImageQuant (Molecular Dynamics) software.

In the quantitative experiments, kinetic data were obtained by varying unlabeled ATP concentrations (0.05-2 mM) while keeping concentrations of other components constant. Total [Pi] produced was calculated from the ^{32}Pi values, and the kinetic parameters were determined by using SigmaPlot 8.0 (Systat Software, Inc., Chicago, IL) software. Data shown were an average of duplicate or triplicate values.

Purification of proheads

T4 proheads were produced by infecting *E. coli* P301 (*sup*$^-$) strain with the packaging defective mutant phage, *17am18am-rII*, at 37°C. Infection at this

temperature yields a mixture of two different expanded statuses of proheads, ELPs (empty large particles) and ESPs (empty small particles), but largely enriches for ELPs, which are more active for *in vitro* DNA packaging.[50] The *E. coli* culture used was prepared from streak-purified strain preservation. A small culture was allowed to grow at 37°C shaker in 50:50 LB-M9A medium for 6-8 hours. This culture was inoculated into a large LB-M9A medium (500 ml) at a dilution of 1:50 and incubated at 37°C shaker until the density reached 4×10^8 cells/ml. Infection with the mutant phage was conducted at a multiplicity of 4 and the cells were super-infected at the same multiplicity 7 min later. The infected culture was incubated at 37°C shaker for another 22 min and the cells were harvested by centrifugation at 8,200 x g for 12 min. The pellet was resuspended in 50 ml of prohead buffer-I (50 mM Tris-HCl pH 7.5, 5 mM $MgCl_2$, and 3 mM β-mercaptoethanol) plus 20 μg/ml DNase I (Sigma-Aldrich) and 1 protease inhibitor cocktail tablet. Potassium Glutamate (Sigma-Aldrich) was added to a final concentration of 50 mM as well as 20-30 drops of chloroform (Mallinckrodt, Phillipsburg, NJ,). The resuspension was incubated at 37°C shaker for 30 min. To this, 50 ml of prohead buffer-II (buffer-I plus 600 mM NaCl) was added and the sample was left in 37°C shaker for another 5 min before centrifugation at 8,200 x g for 12 min. The supernatant containing proheads was

further centrifuged at 34,000 x g for 45 min, and the pellet was resuspended completely in 15 ml prohead buffer-I by pipetting. Cell debris was removed by a low speed centrifugation and the sample was filtered with a 0.20 μm filter before loaded onto a HiPrep DEAE column (HiPrep 16/10 DEAE FF, GE Healthcare) pre-equilibrated with prohead buffer-I. The proheads were eluted with a 0-300 mM linear NaCl gradient using an AKTA-PRIME system, and the fractions were concentrated to 5-10 ml and changed to prohead buffer-I (no NaCl) by an Amicon Ultra-15 centrifugal filter unit (5 kDa cut off). The sample was then further purified by a high pressure Mono-Q column (Mono Q 5/50 GL, GE Healthcare) using a 0-300 mM linear NaCl gradient with an AKTA-FPLC system. The proheads eluted as two major peaks, both of which contained >95% of ELPs as judged by SDS-PAGE (see below). The pooled peak fractions were concentrated by centrifugation at 34,000 x g for 1 hour and resuspended in prohead buffer-I and stored at -70°C.

Treatment of SDS to proheads at room temperature will dissociate ESPs but not ELPs. This property was used to distinguish the amount of ELPs *versus* ESPs in a prohead preparation.[35] The prohead samples were mixed with SDS-sample buffer at room temperature or were boiled. With SDS-PAGE analysis, the fraction of ESPs was determined by comparing the amount of major capsid protein (gp23)

of the room temperature (unboiled) sample to that of the boiled sample. The gp23 amount was quantified using densitometry (Personal Densitometer). The number of prohead particles in a given sample was also determined from the gp23 bands of the sample and a *hoc*$^{-}$*soc*$^{-}$ phage control. Typically 4-5×10^{12} ELPs were produced from 1 liter of infected cells.

DNA packaging assays

The purified gp17s (0.1-3.0 μM) were incubated with purified proheads (1-2×10^{9} particles) and 300 ng of 48-kb phage λ DNA in a 20 μl reaction mixture containing 50 mM Tris-HCl pH 7.5, 5 mM $MgCl_2$, 1 mM spmeridine, 1 mM putriscine, 5% (w/v) polyethylene glycol (Fluka, Sigma-Aldrich), 1 mM ATP and 100 mM NaCl for 45 min at room temperature. Unpackaged DNA was degraded by adding DNase I to a final concentration of 0.5 μg/μl and incubated for 30 min at 37°C. Proteinase K (Fermentas) was added to a final concentration of 0.5 μg/μl in 50 mM EDTA (pH 8.0), 0.2% SDS and incubated for 30 min at 65 °C to release the packaged (DNase-protected) DNA. The reaction mixtures were loaded on an 0.8% agarose gel for electrophoresis. The ethidium bromide stained DNA was quantified by Gel DOC XR imaging system (Bio-Rad) using a control (30 ng λ DNA) run on the same gel.

Packaging-ATPase assays

The purified gp17s (0.5-1.0 μM) were incubated with purified proheads (2×10^9 particles) and 300 ng of λ DNA in a 20 μl reaction mixture as described for the DNA packaging assays for 30 min at 37°C. The reactions also included 1 mM unlabelled ATP and 75 nM [γ-^{32}P]ATP (spec. act. 3000 Ci/mmol). After incubation, 0.6 μl aliquot was removed from each reaction for thin layer chromatography followed by autoradiography and phosphorimaging (Storm 820, Molecular Dynamics). The followed steps of the *in vitro* DNA packaging assay were conducted with the remaining samples.

***In vitro* nuclease assay**

The purified gp17 (1 μM) was allowed to degrade the linearized pAD10 plasmid DNA (29 kb; 80 ng) in a 20 μl reaction mixture containing 5 mM Tris-HCl pH 8.0, 6 mM NaCl and 5 mM $MgCl_2$ for 15 min at 37 °C. When testing the inhibition of DNA degradation by gp16, 5 μM of the T4 family gp16s was added. The reactions were terminated by addition of EDTA to a final concentration of 50 mM, and the samples were electrophoresed on an 0.8% (w/v) agarose gel followed by ethidium bromide staining.

Screening for the peptides that interact with T4 gp16 from a phage display library

Biopanning of the Ph.D.-12 phage display peptide library (New England Biolabs) with T4 gp16 was carried out in 96-well microtiter plates. This M13 display library has a random 12-mer peptide-gIII fusion. The plate was coated with 150 μl per well of 100 μg/ml purified T4 gp16 diluted with 0.1 M $NaHCO_3$ (pH 8.6) and incubated with gentle agitation in a humidified chamber overnight at 4°C. Coated wells were blocked with 10% milk (w/v) in blocking buffer (0.1 M $NaHCO_3$ pH 8.6, 0.02% NaN_3) for 1 hour at 4°C and washed 6 times with TBST buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.1-0.5% [v/v] Tween-20). To bind peptide-displayed phages with gp16, 10^{11} phages from the Ph.D.-12 phage display peptide library, diluted in 100 μl TBST, were loaded into each coated well and incubated for 15 min at room temperature with gentle shaking. The nonbinding phages were washed off with TBST 10 times and the bound phages were eluted by different means with three different selection methods (see below). The eluted phages were amplified by infection with the *E. coli* ER2738 host strain (New England Biolabs) and the expansions were titrated on LB/IPTG/Xgal plates. 10^{11} phages from the expansions were used for the consecutive round of panning.

Selection method a) involved four rounds of panning. From round one to round three, bound phages were eluted with 100 μl elution buffer (0.2 M Glycine-HCl pH 2.2, 1 mg/ml BSA) for 10 min. The eluate was neutralized with 15 μl 1 M Tris-HCl pH 9.1. In the fourth round, bound phages were eluted with 100 μl of 1 μM T4 gp17 in TBS buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl) for 30 min. Selection method b) involved two rounds of panning and for both rounds, bound phages were eluted with 100 μl of 1 μM T4 gp17 in TBS buffer for 30 min. Method c) had four rounds of panning, and the acidic elution buffer was used for phage elution for each round. After the final round of panning of each method, eluted phages were titrated and the plaques from titering were isolated for DNA preparation. The nucleotide sequences corresponding to the 12-mer peptides fused to gIII were commercially sequenced (Davis Sequencing, Inc.) using the -96 gIII sequencing primer (New England Biolabs). The panning selected peptide sequences were aligned with T4 gp17 by DNAMAN software. Only the sequences of the phages affinity re-isolated at least two times were considered significant.

Assay of phage-displayed peptides binding with T4 gp16

Assaying the binding of the biopanning selected peptides to T4 gp16 was done

using enzyme-linked immunosorbent assays (ELISAs). The plaques of phages displaying peptides of interest were amplified by infecting the *E. coli* ER2738 culture and concentrated by PEG/NaCl precipitation. For each clone/peptide to be characterized, one row (12 wells) of ELISA plate was coated with 100 μl of 100 μg/ml T4 gp16 in 0.1 M $NaHCO_3$ (pH 8.6) per well and incubated in a humidified container overnight at 4°C. Coated wells were blocked with 10% milk (w/v) in blocking buffer (0.1 M $NaHCO_3$ pH 8.6, 0.02% NaN_3) for 1 hour at 4°C and washed 6 times with TBST buffer. Each clone was carried out with four-fold serial dilutions of the phages in 200 μl of TBST, starting with 10^{12} virions and ending with 2×10^5 virions, and each dilution was loaded into each of the 12 blocked wells. The plate was incubated at room temperature for 1 hour with agitation. After washing the wells with TBST 6 times, bound phages were eluted with 100 μl of 1 μM T4 gp17 in TBS buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl) for 30 min and titrated on LB/IPTG/Xgal plates. The titers were compared with the controls (wells coated with BSA). $>10^2$ fold of difference on titers would indicate binding of the peptide to T4 gp16.

Results

Sequence identity among T4 family terminases

Terminase sequences from a number of phages related to T4 have been analyzed. These include terminase proteins from coliphages T4, RB32, RB14, RB51, RB69, JS10, JS98, RB49, JSE, Phi1 and RB43, *Aeromonas* phages 31, 44RR2.8t, 25 and Aeh1, and *Vibrio* phages KVP40 and KVP20. Among these, small and large terminase sequences from five coliphages (T4, RB69, JS98, RB49 and RB43), two *Aeromonas* phages (44RR2.8t and Aeh1) and two *Vibrio* phages (KVP40 and KVP20) were selected based on various degree of divergence, and BLAST searches of the NCBI database were performed.

The search results are summarized in Table 1. The sizes of the terminase proteins are similar, ranging between 154 a.a. and 182 a.a. for the nine small terminase subunits, and between 600 a.a. and 633 a.a. for the nine large terminases subunits. The overall sequence identity percentage for large terminase subunits is 53%, and for small terminase subunits, the percentage is 41%. Lower percentage of identity was observed when gp16 and gp17 of T4 were compared with *Vibrio* phages KVP40 or KVP20. In addition, the terminase proteins of *Aeromonas* phages show more similarity to coliphages than to *Vibrio* phages, suggesting a closer

relationship between coliphage and *Aeromonas* phage terminase proteins. This result is consistent and complementary with previous reported sequence analysis of T4, RB49, KVP40 and KVP20.[46]

The high level of similarity among T4 family terminases is a reflection of their common function in viral genome packaging.[41] Of these nine T4 family phages, terminase proteins from coliphage RB49 that is distantly related to T4, and from *Vibrio* phage KVP40 that is also distantly related, were selected to perform detailed functional analysis. The sequence identities of RB49 large and small terminases to T4 are 70% and 49%, respectively. KVP40 terminases are less similar to T4, with the identity percentages being 53% and 41%, respectively. Studying these terminases is expected to provide divergent functional features which are important for understanding the mechanisms of terminase functions.

Cloning and overexpression of T4 family terminase proteins

Small and large terminase proteins from bacteriophages RB49 and KVP40 were cloned into the T7 expression vectors (pET-28b or pET-15b) for overexpression in *E. coli*. These expressed proteins contained an N-terminal hexahistidine tag. T4 gp17 (with the C-terminal 33 amino acid truncation), T4-N360, the gp17 N-

terminal ATPase domain, and T4 gp16 were also overexpressed.[44] The polypeptides produced by 2-hour IPTG induction comprised 5-20% of the total cell protein, as shown on the SDS-polyacrylamide gel (Fig. 7). The new protein bands overexpressed correspond to the sizes calculated from the predicted amino acid sequences, as judged by comparison with the molecular weight standards (Fig. 7, *arrows*). All of these terminase proteins exhibited fairly good solubility as determined by B-PER Bacterial Protein Extraction Reagent (Pierce) (data not shown).

Purification and oligomeric analysis of T4 family terminase proteins

All the overexpressed terminase proteins and the T4-N360 domain were purified by Ni-agarose affinity column followed by size-exclusive gel filtration. The purified proteins were estimated to be >95% homogeneous as judged by SDS-PAGE (Fig. 8a). The protein yield for gp17s and T4-N360 was about 2-3 mg/liter of induced culture, and the yield for gp16s was about 20 mg/liter of induced culture. On the SDS gel, one or two shorter bands could be seen in addition to the predominant band of RB49 gp17, KVP40 gp17 or RB49 gp16 (Fig. 8a). In the case of KVP40 gp16, additional shorter bands were also seen upon overloading the gel (data not shown). Since these observations were consistent with T4 gp17 and

gp16 purified using a similar protocol, the shorter form(s) was likely to be the corresponding terminase proteins arising from proteolytic cleavages or alternative translational terminations.[34]

The oligomeric states of purified terminase proteins were analyzed by native-PAGE (Fig. 8b). RB49 gp17 formed a ladder of bands corresponding to monomers, dimers, trimers, etc., which was similar to T4 gp17.[34] KVP40 gp17, however, showed a smear pattern with most of the protein present in stacking gel. This may indicate inappropriate folding and aggregation of the purified protein.

Previous analyses have determined that T4 gp16 exists as oligomers in solution and forms single and double rings and clustered double rings.[28, 34] A critical coiled coil motif responsible for gp16 oligomerization is present in all T4 family phage small terminases including RB49 and KVP40.[45] In the native-PAGE analysis the majority of all three gp16s migrated as high molecular weight oligomeric species, suggesting their common structural and biochemical characteristics, although the distribution of the proteins to different oligomeric species varied (Fig. 8b).

Purified gp16 was DNA bound

There is evidence that gp16 shows weak binding to dsDNA, but this was only demonstrated by denaturation-renaturation (oligomer dissociation/re-association) process.[28] In this study, the *E. coli* overexpressed and purified gp16 proteins were shown to be DNA-bound. Most of the DNA molecules co-purified with RB49 gp16 were only loosely (and presumably nonspecifically) bound, as they were separated from the gp16 protein bands on native-PAGE (Fig. 9a and 9b, *lane 2*). These DNA molecules were degraded by Benzonase nuclease (Fig. 9a and 9b, *lane 3*). However, some of the DNA molecules were tightly bound and protected from the nuclease digestion, and were released only when gp16 protein was digested (Fig. 9a and 9b, *lanes 3* and *4*). The existence of tightly bound and protected DNA was further confirmed by purification of DNA from the proteinase digestion mixture and running on an agarose gel (Fig. 9c). Electrophoresis of the protected DNA showed some concentrated bands at the bottom of the gel (Fig. 9a, *lane 4*, and 9c), suggesting that the binding of gp16 to DNA has size preference, ~40-80 bp. Similar results were obtained with T4 gp16 (data not shown). These results suggest that gp16 binds to dsDNA and forms stable complexes.

Nuclease activity of RB49 and KVP40 large terminase proteins

The large terminase of T4 has nuclease activity that generates linear DNA substrates from the concatemeric T4 DNA for packaging.[53] Presumably, large terminases of other T4 family phages exhibit this important nuclease activity. An *in vivo* nuclease assay was used to analyze the nuclease activity of the large terminases of RB49 and KVP40. When gp17 with active nuclease activity has been overexpressed by induction, the plasmid and *E. coli* genomic DNA is cleaved to fragments, and the plasmid DNA prepared from the induced cells appears as an extended smear on an agarose gel. In contrast, a nuclease-deficient gp17 construct yields intact plasmid bands.[44] Similar to T4 gp17, RB49 gp17 exhibited nuclease activity (Fig. 10). KVP40 gp17, however, showed no detectable activity in this assay which is likely due to the improper folding of this overexpressed protein (Fig. 8b).

T4 family terminase proteins exhibit gp16-stimulated ATPase activity

The T4 large terminase gp17 possesses a weak ATPase activity that is dramatically stimulated by the small terminase gp16.[34] This ATPase activity, which is mapped to the N-terminal domain of gp17, is suggested to be critically required for DNA packaging.[37, 44] Conserved ATPase signatures were found in the

large terminases of all the T4 family phages, suggesting their common functional features.[46] gp17s of RB49 and KVP40 exhibited weak basal level of ATPase activity, hydrolyzing ~2-3% of the ATP present in the reaction mixture under the conditions used (Fig. 11, *lanes RB49 gp17* and *KVP40 gp17*, and Table 2). The RB49 gp17, like the T4 gp17, was greatly stimulated by gp16 and the fold of stimulation was comparable to that of T4 gp17 (Fig. 11, *lane RB49 gp17 + gp16*, and Table 2). However, the gp16-stimulation of ATPase on KVP40 gp17 was only marginal (Fig. 11, *lane KVP40 gp17 + gp16*, and Table 2). Again, this was consistent with its putative misfolding problem.

The large terminases of T4 family phages exhibit *in vitro* DNA packaging activity

In order to test the DNA packaging activity of the large terminases, a defined system was established using purified components.[35] Only two essential components, proheads and gp17, are required to achieve efficient packaging.[35] When these were mixed with DNA and ATP under optimal conditions, up to ~20% of the input DNA was packaged and remained DNase resistant. The purified RB49 and KVP40 gp17s were tested for DNA translocation activity in this defined system, using T4 proheads purified with ion-exchange

chromatography (Fig. 12). The proheads appeared as one major gp23 band upon SDS-PAGE and no gp18 band, which indicated the high purity.

When T4 gp17 was tested in the DNA packaging assay, a lag was observed at low gp17 concentration and the packaging increased after a threshold concentration (0.5-1 μM) (Fig. 13, *solid line*). Similar results were obtained for RB49 gp17, but the maximum packaging activity was 30% lower as compared to T4 gp17, and the packaging decreased after 1.5 μM (Fig. 13, *dotted line*). This may be due to the fact that T4 proheads were used in the reaction and the lower activity suggests decreased recognition between gp17 and the prohead portal. Interestingly, KVP40 gp17 showed significant DNA packaging activity at low concentrations (Fig. 13, *dashed line*), although the activity is about 25% that of T4 gp17 and about 10% that of RB49 gp17.

In the defined packaging system where gp16 is not present, T4 gp17-ATPase is stimulated up to 8-10 fold when DNA translocation occurs.[35] This ATPase stimulation was also demonstrated with RB49 and KVP40 gp17s (Fig. 14). For both RB49 and KVP40 gp17s, like T4 gp17, a weak basal ATPase activity was seen when any of the packaging components were missing and there was no

packaging. And when DNA packaging was in place, the ATPase activity was stimulated (Fig. 14, compare *columns* of *RB49 gp17* and *KVP40 gp17* with *T4 gp17 columns*). Thus, gp17s of RB49 and KVP40 possess the same functional features as T4 large terminase and are able to actively translocate DNA into T4 proheads.

gp16s inhibit DNA packaging

It has been shown previously that gp16 behaves differently in crude and defined DNA packaging systems. In the crude system where transcription, replication and recombination proteins are present, gp16 enhances packaging efficiency.[28, 34] In the defined system containing only proheads and gp17, it is inhibitory.[35, 54] Quantitative analysis of this effect in this study showed that complete inhibition required a very low gp16 concentration, e. g. at a ratio of 1 gp16 oligomer: 60 gp17 monomer[38] (Fig. 15). This is in contrast to the ratio required for optimal gp17-ATPase stimulation, which is 1 gp16 oligomer: 1 gp17 monomer.[34]

The mechanisms for inhibition of DNA packaging in the defined system by gp16 could possibly be: 1) titration of gp17 from binding to the prohead portal; 2) titration of DNA ends where packaging presumably initiates from accessed by gp17; 3) interaction with prohead-bound gp17 molecules, which resulted in

packaging abortion. In the reaction where the inhibition was nearly complete, there were 6×10^9 proheads, 6×10^9 λ DNA molecules (or ~10^{12} 500 bp DNA molecules), 3×10^{11} gp16 oligomers and 1.8×10^{13} gp17 monomers (Fig. 15, *lane 4*). As gp17 forms a pentamer at the prohead portal,[29] there would be ~3×10^{10} prohead-bound gp17 molecules. Since one gp16 oligomer is thought to interact with one gp17 monomer,[44] there would be a large excess of free unbound gp17. So the first possibility mentioned above is unlikely. The second possibility is also unlikely because when 500 bp DNA was used as the packaging substrate, inhibition was relieved with the same or even lower amount of gp16, whereas the number of DNA ends was increased by ~200 fold (Fig. 15, compare *lanes 5-8* of *λ DNA* vs. *500 bp DNA*). Moreover, in a packaging system using purified prohead-gp17-DNA complex and no free gp17, addition of gp16 resulted in inhibition of DNA translocation at a ratio of 1 gp16 oligomer : 1 gp17 monomer.[38] Taken together, these data suggest that in the defined system, gp16 stalls DNA packaging by interacting with the prohead-assembled gp17 molecules.

This phenomenon was also observed with T4 family terminases. When T4 gp17, RB49 gp17 or KVP40 gp17 was incubated with one of the T4 family gp16s in the defined DNA packaging reactions, strong inhibition occurred with the same low

concentrations of gp16s (Fig. 16). The extent of inhibition was similar to T4 gp17 by all three gp16s, but varied when RB49 or KVP40 gp17s were used for packaging (Fig. 16, compare *groups* of *RB49 gp17* and *KVP40 gp17* with *T4 gp17* in *histogram*). This might reflect that there is specificity for the interaction between gp16 and prohead-bound gp17.

gp17 nuclease is regulated by gp16s

The nonspecific nuclease activity associated with gp17 is required for the cutting of concatemeric DNA at the packaging initiation and termination.[30, 31, 39] This nuclease activity is strictly controlled so that cutting only occurs before and after packaging. It was thought that the terminating cut is related to the headful packaging, while the initiation cut is regulated by gp16.[41] If gp16 is present, gp17 can only convert circular DNA to linear form, but further degradation is inhibited.[38] When incubated with the linear pAD10 DNA in an *in vitro* system, T4 gp17 cleaved the DNA into a series of short fragments (Fig. 17, *lane 2*). The cleavage is largely inhibited by gp16 of T4, RB49 or KVP40 at a molar ratio (gp16: gp17) of 5:1 (*lanes 3*, *4*, *5*). Similar to the gp16 inhibition of DNA packaging in the defined system (Fig. 16), T4 family gp16s cross-regulate T4 gp17 nuclease.

Combinations of terminase proteins from different T4 family phages show stringent specificity of gp17-ATPase stimulation

The gp17-ATPase activity of T4 gp17 and RB49 gp17 was dramatically increased upon gp16 stimulation, while purified KVP40 gp17 that was presumably not folded properly showed little ATPase stimulation with gp16 (Fig. 11). When T4 gp17, T4-N360 or RB49 gp17 was combined with different T4 family gp16s, ATPase was stimulated differentially, indicating specific interactions between the large and small terminases (Fig. 18). For T4 gp17, RB49 gp16 could stimulate the ATPase activity to 14% as compared to T4 gp16, and KVP40 gp16 could only stimulate to 4% (Fig. 18a). In the case of T4-N360 (T4 gp17 ATPase domain), the stimulation by RB49 gp16 was 9% of the activity stimulated by T4 gp16, and the stimulation by KVP40 gp16 was not detectable (Fig. 18b). When using RB49 gp17, the activity stimulated by T4 gp16 was 12% to that stimulated by RB49 gp16, and as low as 2% was stimulated by KVP40 gp16 (Fig. 18c). These data suggest stringent specificity of gp17-ATPase stimulation by gp16, even when there is about 70% overall sequence identity between the T4 and RB49 gp17s (Table 1). Moreover, the specificity appears to be associated with the ATPase domain of the large terminase (Fig. 18b). The K_m values for the gp17-ATPase activity stimulated by different gp16s were similar, while the catalytic rate of ATP hydrolysis

(represented by V_{max}) varied (Fig. 18, see K_m and V_{max} values). This indicates that the specificity is at the level of catalysis, but not binding of ATP to the catalytic pocket. Since previous results showed that gp16 stimulates gp17-ATPase by increasing the rate of hydrolysis of already bound ATP,[34] these results suggest that the specificity of gp16-gp17 interactions affect this catalytic step.

The gp17-ATPase stimulation specificity is consistent with phage divergence

In addition to the gp16s from RB49 and KVP40, three other T4 family small terminase proteins, RB69 gp16, RB43 gp16 and 44RR gp16, were overexpressed and purified (Fig. 19a). Native-PAGE showed that, like T4 gp16, the purified T4 family gp16s migrated as high molecular weight complexes on the gel (Fig. 19b), suggesting that they too oligomerized. The specificity for ATPase stimulation was observed when these gp16s were combined with T4 gp17 (Fig. 20). gp16 from the closely related coliphage RB69 (90% sequence identity to T4 gp16) stimulated the ATPase activity of T4 gp17 to 46% of that simulated by T4 gp16, while gp16s from more distantly related phages RB49 and RB43 could only stimulate to 16% and 20%, respectively. The *Aeromonas* phage 44RR gp16 and the *Vibrio* phage KVP40 gp16 were able to stimulate to <5%. The results show that the gp17-ATPase stimulation specificity correlates with their divergence and heterologous

terminases that show closer sequence relationship work better for ATPase stimulation.

Alignment of T4 family gp16 sequences shows three divergent regions

T4 family gp16s with various sequence similarities to T4 were selected for multiple sequence alignments. The majority of the sequences, largely the central domain of the proteins, showed high level of homology, while three divergent regions with <50% homology were observed in the N- and C- domains[38] (Fig. 21). Divergent region I is located in the N-terminal domain between amino acids 1-25 (T4 gp16 sequence); region II is at the junction of the central domain and the C-domain comprising amino acids 112-126, and region III is at the C-terminus (amino acids 148-164) (Fig. 21). Other evidence has shown that the central domain (which is relatively conserved) is responsible for oligomerization and that the N- and C- domains interact with gp17, ATP and/or DNA.[38] The divergent regions in the N- and C- domains of gp16 are potentially essential for determining the gp17-ATPase stimulation specificity.

Divergent region III at the C-terminus of gp16 is essential for the gp17-ATPase stimulation specificity

In order to distinguish which region(s) was important for the gp17-ATPase stimulation specificity, a series of gp16 swap fusion clones were constructed where one or more of the divergent regions from one phage gp16 were swapped to another (Fig. 22). These gp16 recombinants include alternation of the amino acid sequences of region I, II or III from T4 to RB49 (Fusion 6, 11, 13 and 16), from RB49 to T4 (Fusion 17, 18 and 19) and from KVP40 to T4 (Fusion 20). All of the swap fusion constructs were purified to near homogeneity as soluble proteins (Fig. 23a, see 'Materials and Methods' for details of the purification procedure). Swapping of these gp16 regions did not affect oligomerization or folding, as they migrated as compact bands to the position similar to T4 gp16 oligomers on the Native-PAGE gel (Fig. 23b). Had they been misfolded, the protein would have migrated as a smear. Quantitative analysis of ATPase stimulation to T4 gp17 and RB49 gp17 demonstrated the specificity of the fusion proteins (Table 3). Fusions 6 and 13, with the divergent regions I and II swapped from T4 to RB49 respectively, showed no change in specificity. Significant changes in specificity were observed for fusions 11 and 16, where region III was altered. Moreover, for Fusion 17, change of the region III sequence from RB49 to T4 completely converted the specificity to T4. Even for KVP40 gp16 whose ATPase stimulation activity was <4% (Fig. 18 and 20), swapping the region III sequence to T4 increased the

activity to ~20%, and T4 gp17 was stimulated higher than RB49 gp17 (Table 3, *Fusion 20*). These data suggest that the divergent region III is essential for the gp17-ATPase stimulation specificity. Further studies showed that partial swap of the region III sequence did not change the specificity (Table 3, *fusions 18* and *19*), indicating that the entire region is participating in the specificity determination.

These results were verified with another set of experiments where the stimulation was tested as a function of the [gp16]: [gp17] ratio (Fig. 24 and 25). The ATPase activity of T4 gp17 was highly stimulated with T4 gp16, Fusion 6 and Fusion 17 (possessing T4 region III sequence) (Fig. 24), and that of RB49 gp17 was highly stimulated with RB49 gp16 and Fusion 11 (possessing RB49 region III sequence) (Fig. 25). Furthermore, the specificity is mostly related to different K_{cat}, but not the [gp16]: [gp17] ratio required for maximum stimulation, which remained the same, at about 8-10:1 (Fig. 24 and 25). The K_{cat} decrease at very high [gp16]: [gp17] ratios (especially at 120:1) was also observed and described previously.[45]

Multiple sites from both N- and C- domains of gp16 are responsible for the gp17-ATPase stimulation specificity

Progressive deletions at the C terminus of T4 gp16 up to V115 (amino acids 1-V115 remained) resulted in reduction of the gp17-ATPase stimulation, but significant level of the activity was still retained.[38] The recombinant T4 I128 (T4 gp16 amino acids 1-I128, Fig. 26 and 27), for which most of the C-domain including the divergent region III was truncated, was able to stimulate the ATPase of T4 gp17 to ~63% (Table 4). Interestingly, this protein retained the specificity, as RB49 gp17 was only stimulated to ~5%, although the important region III was not present (Table 4). A gp16 fusion with the same truncation and the N-domain sequence changed to RB49 (Fusion 9, Fig. 26 and 27) stimulated the ATPase activity of RB49 gp17 to ~11%, while the stimulation to T4 gp17 was not detectable (Table 4). This result indicates that the N-domain of gp16 shows gp17-ATPase stimulation specificity when the C-domain is truncated. Taken together, the data suggest that the gp17-ATPase stimulation specificity is a combined outcome from the gp17-gp16 interactions through several different regions and multiple sites from both the N- and C- domains of gp16.

Both the N- and C- domains of gp16 are important for gp17-ATPase stimulation

Only partial loss of the gp17-ATPase stimulation activity was observed when the

gp16 C-domain was truncated.[38] T4 gp16 mutants with deletions at the N terminus, L9 (amino acids L9-164), I12 (amino acids I12-164) and S36 (amino acids S36-164) also retained oligomerization and gp17-ATPase stimulation (>80% as compared with the WT, Fig. 28). However, deletions at both N and C termini of gp16 lost the gp17-ATPase stimulation activity, though their oligomerization was intact.[38] The gp17-ATPase activity can be stimulated partially with either the N- or the C-domain, but for maximum stimulation both domains are required. These data further confirm that multiple sites from both N- and C- domains of gp16 are involved in the interaction of the large and small terminase proteins.

Multiple sites on gp17 contribute to the specificity, including sites in the N-terminal 85 amino acids

The large terminase has a higher amino acid sequence similarity as compared to the small terminase among T4 family phages (Table 1). Sequence alignment of selected T4 family large terminases by BLAST showed that most divergent regions on gp17 were the N-terminal ~80 amino acids and the extreme C-terminus.[46] Since the ATPase stimulation specificity is associated with the ATPase domain (Fig. 18b), the divergent region at the N-terminus could contribute to it, and was considered to be a good candidate for sequence swap. A fusion gp17

was constructed with the N-terminal 85 amino acid sequence swapped from T4 to RB49 (Fig. 29). This fusion construct was successfully purified as a soluble protein (Fig. 30) and tested for ATPase stimulation with T4 and RB49 gp16s (Table 5). The fusion gp17 showed a significant change of the stimulation specificity although a complete switch of the specificity was not achieved. The stimulation by RB49 gp16 was 77% as compared to that stimulated by T4 gp16, indicating that the fusion gp17 possesses specificity for both gp16s (Table 5, compare with the K_m and V_{max} values in Fig. 18, a and c). The change of the specificity by swapping the N-terminal 85 amino acids suggest that sites in this region are involved in the specificity determination, but other sites outside this region may also contribute. Multiple sites on gp17 are involved in its interaction with gp16, and are responsible for the ATPase stimulation specificity.

Several regions on gp17 which are potentially interacting with gp16 were identified by biopanning with phage M13 displayed random peptides

In order to identify the regions/amino acid residues on gp17 with which gp16 may interact, biopanning experiments were performed using a random 12-mer phage display peptide library against purified gp16. gp16 was coated on plates and screened for the peptides that bind to it. Multiple rounds of selection

employing three different elution methods led to the determination of 14 phage display peptide sequences (Table 6). Each sequence was isolated more than once, and 9 out of the 14 sequences were selected with at least two methods. Especially, two peptide sequences were affinity re-isolated (recovered) 17 times (Table 6, *Clone No. 5-5*) and 9 times (*Clone No. 8-24*), respectively. The phage display peptide sequences were aligned with the T4 gp17 sequence and the matches were displayed in Table 7. Among these sequences, the most compelling matches on the basis of frequency were to the region of gp17 N-subdomain I residues 290-315. 4 sequences and 28 matches (counting the repeats of selection) were clustered to this region. Besides, another region with high frequency of matching was N-subdomain II residues 37-52, at which 2 sequences and 11 matches were found. All of the peptide sequences matching these two regions were confirmed to be interacting with gp16 (Table 7, column 4).

A number of features suggest that the regions matching the selected sequences may potentially interact with gp16, especially residues 37-52 and residues 290-315. 1) The region of residues 290-315 locates near to the gp17 C-motif (the ATPase coupling motif, $TTT_{285\text{-}287}$), an important motif that is functioning as an ATP hydrolysis sensor.[43] Although the mechanism of gp17-ATPase stimulation by

gp16 is not yet clear, it is quite possible that a region close to the C-motif is involved in the gp17-gp16 interaction, resulting in elevated ATPase catalysis. 2) It was shown that sites in the N-terminal 85 amino acids of gp17 were involved in the interaction with gp16 (Table 5). Two matches were found to be in this part of sequence, including residues 37-52. 3) Based on the X-ray structure and the Cryo-EM reconstruction information of gp17,[29] all the biopanning identified regions are accessible to gp16. Moreover, the two highly possible regions, residues 37-52 and residues 290-315, are located at a key N- C- domain interface, which makes the regulation of activities associated with the two distinct domains efficient (see discussion).

Discussion

Tailed bacteriophages and herpes viruses translocate their dsDNA genome into a preformed protein capsid shell by powerful molecular motors.[20, 41, 55] This is a complex process which involves a series of enzyme activities, protein-protein and protein-DNA interactions.[41] In bacteriophage T4, the DNA packaging machine is composed of the procapsid, the portal protein gp20, and the large and small terminases, gp17 and gp16.[26] Packaging starts with the initiation cleavage of the T4 DNA concatemer by gp17. When the pentameric DNA packaging motor is assembled at the portal, gp17 is converted from the cutting mode to the DNA translocation mode. After one headful DNA is packaged, gp17 dissociates from the capsid and makes the termination cut.[31]

A novel mechanism of DNA packaging was proposed based on the data from biochemical functional analyses of the terminases, X-ray structures of domains and full-length protein, and electron microscopic reconstruction of the packaging motor.[29] The large terminase, gp17, is acting as the actual packaging motor and is central to this electrostatic force-dependent mechanism. This mechanism does not involve the small terminase, gp16, which is dispensable for *in vitro* packaging, but is essential for phage viability.[20, 41] gp16 does not possess any enzymatic

activities; however, it seems to be actively involved in DNA packaging by interacting with the large terminase and DNA.[41] Missense mutations in gp16 results in loss of DNA packaging *in vivo* and accumulation of empty proheads.[56] Though the precise role(s) of the small terminase in DNA packaging is still a mystery, evidence suggests that it acts as a regulator for the DNA packaging motor.[38] gp17 is a flexible protein which possesses various functions and undergoes many transitions during the DNA packaging process. The regulatory role of gp16 is important as the various functions of the packaging motor have to be precisely coordinated at different stages of DNA packaging. Otherwise uncontrolled motor activities would be suicidal to the virus.

Biochemical analyses have shown that the various activities of gp17 were modulated by gp16, but the detailed mechanism(s) remained poorly understood. Here, the study of the terminase proteins from T4 family phages show that specific interactions between gp17 and gp16 through multiple sites on both proteins. Multiple weak interactions between gp17 and gp16 depending on the conformational state of the motor may provide the flexibility required for the regulation of the complex phage DNA packaging machine. This could also be a novel new mechanism broadly applicable to other molecular motors.

T4 family terminase proteins share common functional features

Sequence analyses showed close relationship of T4 family terminase proteins. BLAST searches of the NCBI database using T4 terminase returned several T4 family phage terminases with high scores. Large terminases share the same domain organization, with an N-domain containing the classic ATPase signatures, and a C-domain with nuclease and DNA translocation sites.[44] Despite the relatively lower similarity of the T4 family small terminases (Table 1), the predicted secondary structures and the domain organizations are essentially the same.[38] This indicates that the basic structure and functions of T4 family terminases are well conserved. On the other hand, differences are expected in their interacting specificities which are likely localized in the divergent regions.

The terminase proteins from RB49 and KVP40 were systematically studied for their basic activities, and the results are summarized in Table 8. Similar to T4 large terminase, the RB49 gp17 exhibited the gp16-stimulated ATPase activity, the nuclease activity and the DNA packaging activity with packaging stimulated ATPase. The purified protein also showed multiple bands corresponding to monomers and multimers on native-PAGE gel (Fig. 8b). The overexpressed KVP40 gp17 behaved differently. It showed a smear on the native-PAGE gel,

indicating improper folding and aggregation. The nuclease activity of this protein was not detectable (Fig. 10). Consistent with these data, purified KVP40 gp17 possesses very low gp16-stimulated ATPase activity (<2 fold, see Fig. 11 and Table 2). On the other hand, this purified protein exhibited significant DNA packaging activity and packaging stimulated ATPase, but only at low concentrations (Fig. 13 and 14). This suggests that although this protein had folding problems and formed largely aggregates, either a small portion of the purified protein was folded properly and functional, or the presence of proheads in the DNA packaging reaction mixture allowed proper folding as a result of its interaction with portal. In either case, the fact that KVP40 gp17 is functional in the DNA packaging system indicates that this protein has the same functional features as T4 gp17. These results verified that large terminase proteins from T4 family phages possess common basic functions.

The small terminase proteins of tailed phages exist as stable ring-like oligomers in solution.[28, 47, 57, 58] Sequence analyses of T4 family gp16 showed coiled-coil motifs responsible for oligomerization in the central region of the proteins.[45] With native-PAGE analysis, as has been observed on T4 gp16, the purified RB49 and KVP40 gp16s migrated as two major oligomeric species (Fig. 8b). However, the

distribution of the species varied among the three small terminases. It has been suggested that the two major oligomeric species of gp16 correspond to oligomers (single rings) and dimers of oligomers (double rings),[28, 45] but neither the mechanism of the formation of dimers (and multimers) of oligomers nor the meaning of it for gp16's function is known. Presumably, RB49 and KVP40 gp16s form similar ring-like structures, and the different distribution patterns to the two oligomeric species are the reflection of the proteins' intrinsic properties.

In contrast to the stimulation of DNA translocation in the crude packaging system that mimics *in vivo* DNA packaging,[34] gp16 is strongly inhibitory in the defined system containing only two purified components, proheads and gp17 (Fig. 15). It is possible that in the crude system, there are other phage factors involved in the formation of a 'native' packaging machine, whereas in the defined system containing only two components, gp16 interacts with the portal-bound gp17 and traps the packaging motor in an inactive conformation. It is also possible that gp16 interfered with the binding of gp17 to the portal to assemble the motor, or to DNA ends to make the initiation cut. However, the data in this study showed that both of the above are unlikely, as inhibition was observed even when gp17 and DNA were in excess. The hypothesis that gp16 interacts

with the prohead-assembled gp17 and stalls the motor was further supported by other evidence: 1) When purified prohead-gp17-DNA complex was used for packaging, gp16 inhibits DNA translocation as well as the packaging ATPase (V. I. Kottadiel and Z. Zhang, personal communication). 2) In the presence of gp16, reconstruction of the prohead-assembled gp17 showed substantial change in the gp17 cryoEM density.[38] It was thought that a major conformational change in the prohead-assembled gp17 motor would occur after gp16 interaction.

Strong inhibitory effects were also observed with RB49 and KVP40 gp16s, when the respective gp17s were used in the defined system. Interestingly, the inhibition of DNA packaging is cross-active. Significant reduction of DNA packaging occurred when any of the three gp16s were added (Fig. 16). This result confirms that the functions of T4 family terminases are conserved. On the other hand, the gp16-gp17 interaction possesses specificity, as different extent of inhibition was observed when different gp16s were used.

gp17-nuclease activity is strictly regulated such that the concatemeric DNA is cut once for the first packaging initiation and gets a series of termination cuts following each headful packaging. It was thought that gp16 is responsible for the

regulation of the first initiation cut, while the termination cuts are linked to headful packaging.[41] In the presence of gp16 (and no ATP), gp17 nuclease is restricted to the first cut to convert circular DNA to linear form but further cutting of linear DNA is inhibited.[38] This regulatory role is exhibited by RB49 and KVP40 gp16s as well (Fig. 17). The cross-regulation of the nuclease of T4 gp17 by T4 family gp16s also suggests that common functional features are shared within small terminase proteins. As shown in Table 8, the activities of small terminase required for the regulation of the packaging motor are exhibited by all three T4 family gp16s.

The gp16-gp17 interaction is highly specific

DNA packaging motors in tailed phages are extremely powerful, with the T4 motor translocating DNA at a rate of up to 2,000 bp/second. However, the packaging process must be well-controlled and various domain movements need to be synchronized. Not only are the various functions of the packaging motor precisely coordinated in different stages of DNA packaging, but also packaging of viral DNA and interactions of components have to be specific to the respective phage in mixed infections that frequently occur in nature. The packaging motor itself does not possess intrinsic specificity, as any dsDNA molecule could be

picked-up and packaged in the defined *in vitro* packaging system.[35] It is also shown in this study that the recognition of packaging motor to prohead portal had no stringent specificity, as RB49 or KVP40 gp17 formed competent motors at T4 prohead portal (Fig. 13). gp16, as the regulator of the packaging motor, could be responsible, at least partly, for the stringent specificity required during DNA packaging.

gp16 interacts with gp17 and regulates its functions, as were represented by ATPase stimulation, nuclease modulation and DNA translocation inhibition. These interactions showed specificity when heterologous terminases were put together, and the specificity was most significantly demonstrated with the ATPase stimulation (Fig. 18). Only low (<15%) ATPase cross-stimulation was observed when gp17 and gp16 proteins from different T4 family phages were combined. This indicates stringent specificity of the gp16-gp17 interactions. Not only did gp16 have no effect on the ATPase activities of other nonspecific ATPase such as Na^+/K^+ -ATPase and phage λ large terminase,[34] it also had a greatly reduced stimulation within the closely related T4 family terminase proteins. Consistent with the fact that the gp17-ATPase is stimulated by elevation of the catalytic rate of hydrolysis,[34] the stringent specificity also affects the catalytic

capacity of gp17. As a result, different gp16s having different levels of specificities stimulates gp17-ATPase to different extents, but the gp17's binding affinity toward the ATP substrate is not changed. The terminase proteins of T4 and RB49 are quite similar with respect to their amino acid sequences, domain organizations and functions, but are significantly different in their specificity, indicating two aspects of the ATPase stimulation mechanism: the catalytic function conducted by the conserved regions/motifs and the specificity function determined by the divergent regions. This is further supported by the data showing that the heterologous gp16s more diverged from T4 showed lower stimulation on T4 gp17 (Fig. 20).

The gp16-gp17 interaction involves multiple sites

Alignment of T4 family gp16s showed that the least conserved amino acids were clustered into three regions (Fig. 21). These divergent regions were carefully studied in domain swapping experiments as they possessed potential importance for specificity. Region III was identified to be the most stringent region, as swapping the amino acid sequence of this region resulted in almost entire specificity switch (Table 3, Fig. 24 and 25). Further studies showed that the whole region III participated in the specificity determination, indicating that specificity

is not determined by interactions between one or two key amino acid residues. Thus it appears that complementary surfaces that encompass a number of amino acid residues might be involved.

Another important finding is that region III is not the only region involved in specificity. When region III was not present (e. g. gp16 truncations with C-terminal deletions), the N-domain showed contribution to specificity (Table 4). Similarly, more than one site on gp17 appears to be involved in the specificity determination, as a swap of N-terminal 85 amino acids on gp17 only resulted in partial change in specificity (Table 5).

The interaction sites are clustered at a key position to facilitate efficient regulation

The domain organization of T4 gp16 has been proposed earlier.[38] Evidence have shown that the central domain is essential for oligomerization, whereas the N- and C- domains are interacting with various molecules (gp17, DNA and ATP) to perform function.[28, 38, 45, 46] Importantly, the gp17-interaction sites are located in both the N- and C- domains. These observations suggest that the central domain of gp16 forms the structural core whereas the N- and C-domains form the

functional core. This domain organization is further supported by structural modeling. Based on the recently resolved crystal structure of the phage Sf6 small terminase,[30] a T4 gp16 structure model was constructed (Fig. 31a). The Sf6 small terminase is shorter on both N- and C- domains as compared to T4 gp16, and forms octamers in solution. Thus the T4 gp16 structure was modeled as an octamer and was missing the N-terminal 11 amino acid residues and the C-terminal 24 amino acid residues. However, it still gives information about how T4 gp16 oligmerizes, and indicates the possibility of the C-domain contacting the N-domain of a neighboring molecule to form the functional core at the edge of the gp16 ring.

As shown in the sequence analysis, greater sequence divergence was found in the N- and C- domains of gp16, and the central domain is well conserved (Fig. 21). Since the formation of ring-like oligomers is a common feature of the small terminase proteins, the conserved structural core may provide a suitable platform to present the functional domains, whereas the divergence of the functional core is responsible for the specific recognition of the respective gp17 partner and/or viral DNA. Consistent with this hypothesis, this study has identified the specificity determinant region III that is located at the end of the C-

domain.

The gp16-gp17 interactions are weak and dynamic. No gp16-gp17 complex could be isolated so far under a variety of experimental conditions. However, two regions are considered to be involved based on the phage display library screening: amino acids residues 37-52 of the N-Subdomain II and amino acids residues 290-315 of the N-Subdomain I. Both of the regions are located at the interface of the N- and C- domains, suggesting the interface position is where gp16 interacts (Fig. 31b). The N- C-domain interface is a key position where communications with both the N-domain ATPase center and the C-domain nuclease center could be established. For motor regulation, the functional core of gp16 is presented at the edge of the ring-shaped oligomer, positioned to the N- and C- domains interface of gp17, and causes conformational changes through multiple site interactions, resulting in different outcomes depending on the functional state of gp17. When packaging is at initiation (or termination), interaction with gp16 regulates the nuclease activity. When the DNA is under translocation, the ATPase is stimulated and coupled with the DNA movement.

gp17-nuclease is regulated by gp16 through a β-hairpin at the interface of the

ATPase domain

Comparing the recently resolved structure of the nuclease domain of the large terminase of SPP1 with T4 gp17 showed a conserved feature of phage terminases that is not present in other RNase H family proteins: there is a β-hairpin located at the interface of the N- and C- domains.[39, 59] This β-hairpin is flexible because it is differently positioned in T4 and SPP1 large terminase, and is disordered in the RB49 gp17 C-domain structure.[29] In SPP1 large terminase nuclease domain, the β-hairpin partially blocks the active site and a linear DNA molecule modeled into the catalytic site of the structure clashed with the β-hairpin. Thus it has been proposed that the nuclease activity is regulated by the β-hairpin.

Conserved acidic residues are required for catalysis in RNase H family proteins. In T4 gp17, these catalytic residues are D401, E458 and D542 located in a DNA binding groove (Fig. 31b and 32). The β-hairpin in T4 gp17 flanked by two β-strands, amino acid residues 526-529 and amino acid residues 532-535, is located at the interface of the ATPase domain, and at one end of the DNA binding groove with the possibility to block DNA accessibility (Fig. 31b and 32, compare the two conformations of the β-hairpin). Conformational changes of the ATPase domain could affect the position of the β-hairpin. In particular, W533 in the β-sheet

formed by amino acid residues 532-535 is closely positioned with the hydrophobic residues Y295, W298 and V302 from the ATPase domain. Interestingly, these three residues are from the region of amino acid residues 290-315 that likely interacts with gp16, as shown by biopanning studies. Thus interaction of gp16 with the ATPase domain and the following conformational changes likely change the position of the β-hairpin and affect nuclease activity by controlling the accessibility of the DNA binding groove.

It was thought that the initiation cut before DNA packaging is regulated by the gp16-gp17 terminase complex.[41] Small terminases from many phages have been reported to be responsible for selection of viral DNA. Though there's currently no direct evidence showing T4 gp16 is responsible for this, it indeed binds dsDNA (Fig. 9). One hypothesis is that at the beginning of T4 DNA packaging, the concatemeric phage DNA bound to gp16 is recruited to the gp17 nuclease site for initiation cut, and the β-hairpin of gp17 is restricted by gp16 to prevent further cleavage. After one headful DNA is packaged, the conformational changes induced by the 'headful signal' would release the DNA packaging motor (and the β-hairpin), and the termination cut would occur on the DNA at the portal.

Though the nuclease activity of gp17 is an independent activity, it is not separated from the ATPase. *In vitro* results have shown that the nuclease activity of the C-domain is 13-fold lower than the full length gp17.[31] When ATP is present, the nuclease activity of gp17 is not inhibited by gp16 but stimulated (M. Ghosh-Kumar, personal communication). Mutagenesis data also showed the linkage of the nuclease activity with the ATPase domain. For instance, the gp17 D255E-E256D mutant in which the sequence of Walker B aspartate and catalytic glutamate is flipped, binds ATP but could not hydrolyze it. This mutant has a greatly reduced nuclease activity.[36] On the other hand, the E256Q mutant that also binds ATP and lost ATPase shows greater nuclease activity than the wild type.[42] In the ATPase coupling mutant, T287D, where ATP is hydrolyzed once and the products remain bound to the active site, nuclease activity is lost.[43] These results suggest that the nuclease activity is intimately linked with the conformation and the nucleotide state of the ATPase domain, to which the gp16 regulator interacts.

gp16 stimulation of gp17-ATPase mimics the conformational changes occurring during active packaging

The energy used in T4 DNA packaging is coming from the ATP hydrolysis by the

N-domain of gp17.[44] gp17's ATPase activity is essentially inactive in its monomeric state, but is dramatically stimulated during active packaging as a portal-assembled motor. The same ATPase is also stimulated by gp16, when gp17 is in solution as free molecules without packaging.[34] It has been discussed earlier that the ATPase activity of gp17 is stimulated with a mechanism called 'arginine finger positioning'.[29] A potential arginine finger (Arg-162) was found in gp17, which is appropriately located at the junction of ATPase subdomains I and II. It has been proposed that during active DNA packaging, the contact of the translocating DNA with one gp17 molecule of the assembled motor causes a conformational change through the N-Subdomain II, resulting in the precise positioning of Arg-162 into the catalytic center (Fig. 33). ATP hydrolysis leads to a 6° rotation of the N-Subdomain II, and the charged pairs between the N- and C-domains are aligned and the DNA is moved along with the C-domain by the electrostatic force. Hydrolysis product release resets this gp17 molecule and the DNA is handed over to the neighboring gp17 molecule for the next translocation cycle. In this scenario ATP is hydrolyzed in a controlled manner and coupled with DNA movement.

gp16 stimulation of gp17-ATPase utilizes the same mechanism, as the other well

characterized mechanism, nucleotide exchange, is unlikely.[38] gp16 enhances the catalytic rate of ATP hydrolysis, but not the affinity of gp17 to the ATP substrate.[34] In addition to that, ADP binding to gp17 was unaffected by gp16.[38] Thus gp16 works like a GAP for gp17. Limited proteolysis experiments showed that gp17's susceptibility to proteinase changed upon gp16 interaction, indicating that gp16 interaction caused conformational changes of gp17 (B. Draper, personal communication). Presumably, these conformation changes are similar to those that happen in active packaging. When gp17 is a free molecule in solution but not assembled as a motor, gp16 oligomer interacts with both N-subdomains I and II, leading to a conformational change of gp17 favoring the positioning of Arg-162 into the nucleotide-binding pocket. Because gp17 is a free and flexible molecule and there are no constrains as in the assembled motor, the arginine finger would be positioned simultaneously causing uncontrolled ATP hydrolysis. In a real phage infection, with the help of other phage factors, gp16 would be assembled onto the packaging motor in an appropriate way and would initiate/stabilize the motor conformation for highly efficient ATP hydrolysis and DNA packaging. Consistent with this hypothesis is the finding that in the crude DNA packaging system where other phage factors are present mimicking *in vivo* packaging, gp16 stimulates the packaging efficiency.[34]

Model for packaging implicating gp16 in assembly and regulation of phage T4 DNA packaging machine

We proposed that gp16 small terminase is a global regulator of the T4 DNA packaging motor, modulating the ATPase, translocase, and nuclease activities to fine-tune the motor functions throughout the packaging process. As postulated in the model (Fig. 34), gp16 binds to the T4 concatemeric DNA and forms the gp16-DNA complex. Interaction with the large terminase gp17 at the N- C-domain interface brings the DNA to the nuclease site and place gp16 oligomer at a key position to regulate motor activities (A). After the initiation cleavage of DNA, conformational changes on the complex lead to the assembly of the DNA packaging machine at the prohead portal (B). This sets the stage for optimal ATP hydrolysis and coordinated DNA translocation. Following completion of packaging, the DNA-full head is separated from the complex by the termination cut (C), making room for docking of another empty prohead on the now open and active holoterminase complex (D), while the packaged head goes on to completion stage (E).

It has been well established that the small terminase is important for the recognition of viral genome in *cos* and *pac* phages, and similarly thought of in

phage T4.[26, 28] Results in this study supported this inference by showing that *E. coli* overexpressed gp16 bound DNA (Fig. 9). The bound DNA was co-purified with the gp16 oligomer, and was protected from nuclease digestion, suggesting tight binding. Analyses of the bound DNA showed some sequence preference (K. Kondabagil, personal communication). Previous experiments suggested that the gp16 monomer, but not the oligomer, binds to dsDNA with some preference to a putative *pac* site within gene *16*.[28] The results here are not in conflict with them, as the gp16 oligomers used in the previous experiments could be DNA-bound already, and only when the oligomers dissociated and the bound DNA released, binding to dsDNA was demonstrated. The binding of gp16 to phage DNA, presumably to the putative *pac* sites, is critical. Whether the *pac* sites require sequence specific interactions is under investigation. The gp16 binding to DNA would serve as a nucleus for the subsequent packaging machine assembly events. Coupled with the limited expression of terminase genes, such assembly nuclei would be created at a small fraction of the available *pac* sites, thereby minimizing packaging initiation at adjacent genomes, or within the same genome, which could be suicidal.

Once the holoterminase complex is assembled at the portal and the DNA

packaging machine is formed, the motor is set to perform the highest packaging efficiency with the modulation of gp16. The N- C- domain interface where gp16 assembled is facing outside of the motor, accessible by gp16 oligomers, and is a critical position to communicate with both the ATPase and the nuclease domains (Fig. 35). The gp16 regulator appropriately assembled at this position would optimize the motor activities by restricting the nuclease activity and stimulating the ATPase activity to achieve maximal DNA packaging. One point should be noted that, it is likely that other phage factors such as late transcription factors, DNA recombination and repair enzymes may also be involved, as proper assembly of active holoterminase complex relied on these phage factors (stimulation of DNA packaging by gp16 only occurs when other phage factors are present; otherwise gp16 stalls the motor). These phage factors may also stabilize the complex and work in concert with the DNA translocation. For example, the phage T4 recombinational endonuclease VII, gp49, was shown to interact with the portal protein and was proposed to be an integral part of the packaging machine.[60]

In conclusion, this study revealed mechanisms of regulation of the T4 DNA packaging motor, gp17, by the regulator, gp16. As discussed above, the

regulation is achieved by weak interactions through multiple sites clustered at a key position of the ATPase domain. Distantly located activity sites (ATPase, nuclease) are well communicated with the interactions taken place at the key position, via conformational changes of critical structural elements (e. g. the position of the β-hairpin affecting DNA access; the conformations of N-Subdomains I and II orienting the arginine finger). The involvement of multiple interaction sites may provide the great flexibility to coordinate motor activities at different conformational or functional states of the packaging motor. This study has also established the stringent specificity between the motor and the regulator that allows their co-evolution for a given T4 family phage. This might allow selection of their own genome and packaging machine in mixed infections in nature.

Most phage genomes encode both small and large terminases, and all thus far analyzed possess similar functional features.[41, 61, 62] Modulation by a specific regulator through multiple weak interactions at a key position of the motor might be a common theme in phage/virus DNA packaging motors as well as other complex molecular motors.

Figures and Tables

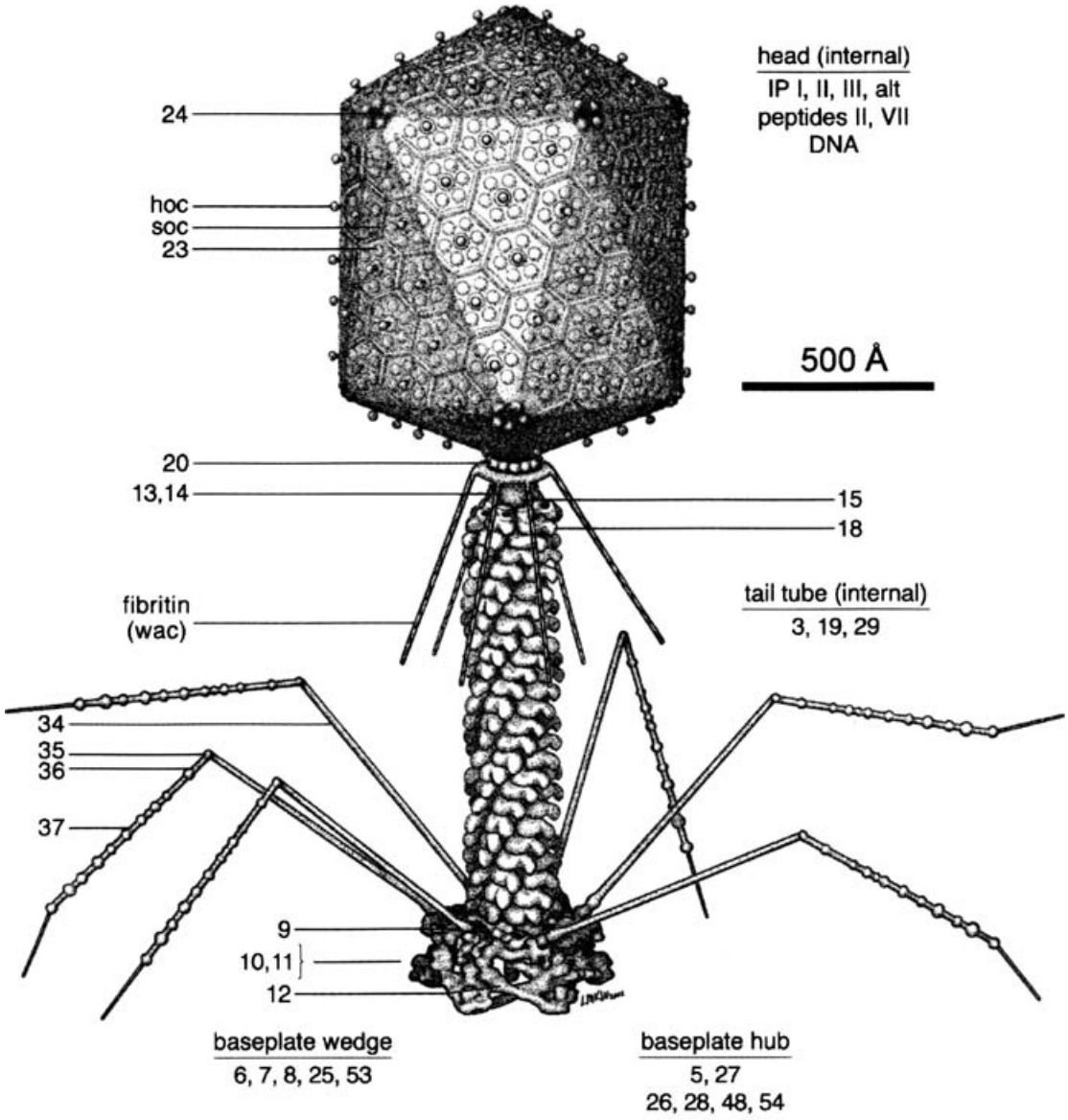


Figure 1. Bacteriophage T4 mature virus (modified from Eiserling and Black[12]).

Labels are gene numbers or names of the proteins on the virion.

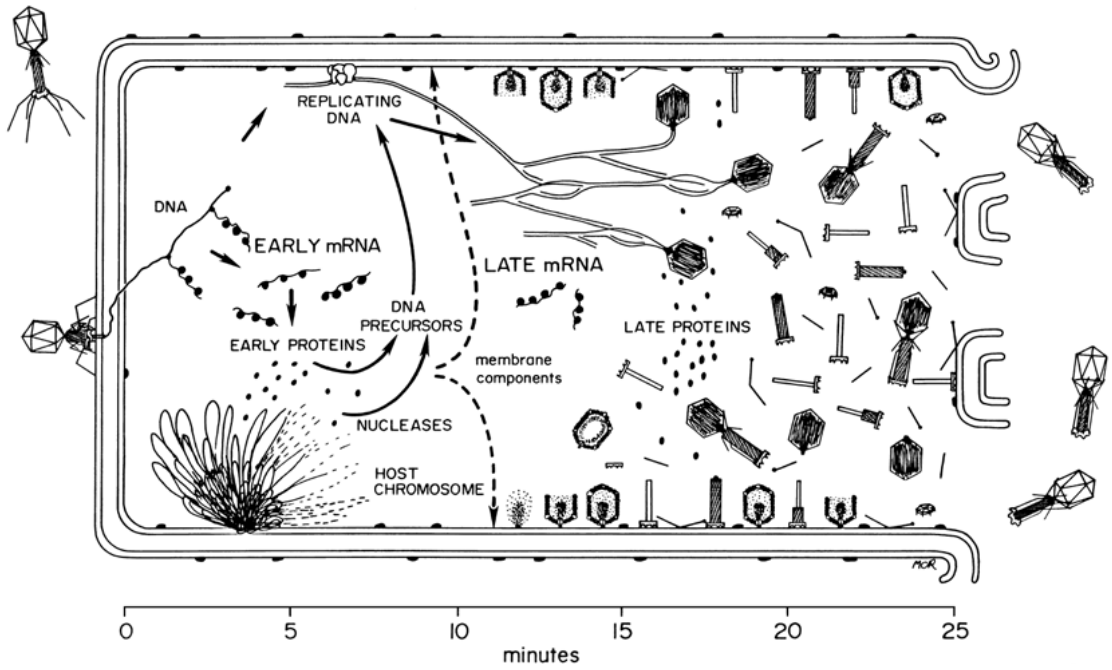


Figure 2. Bacteriophage T4 life cycle.[14]

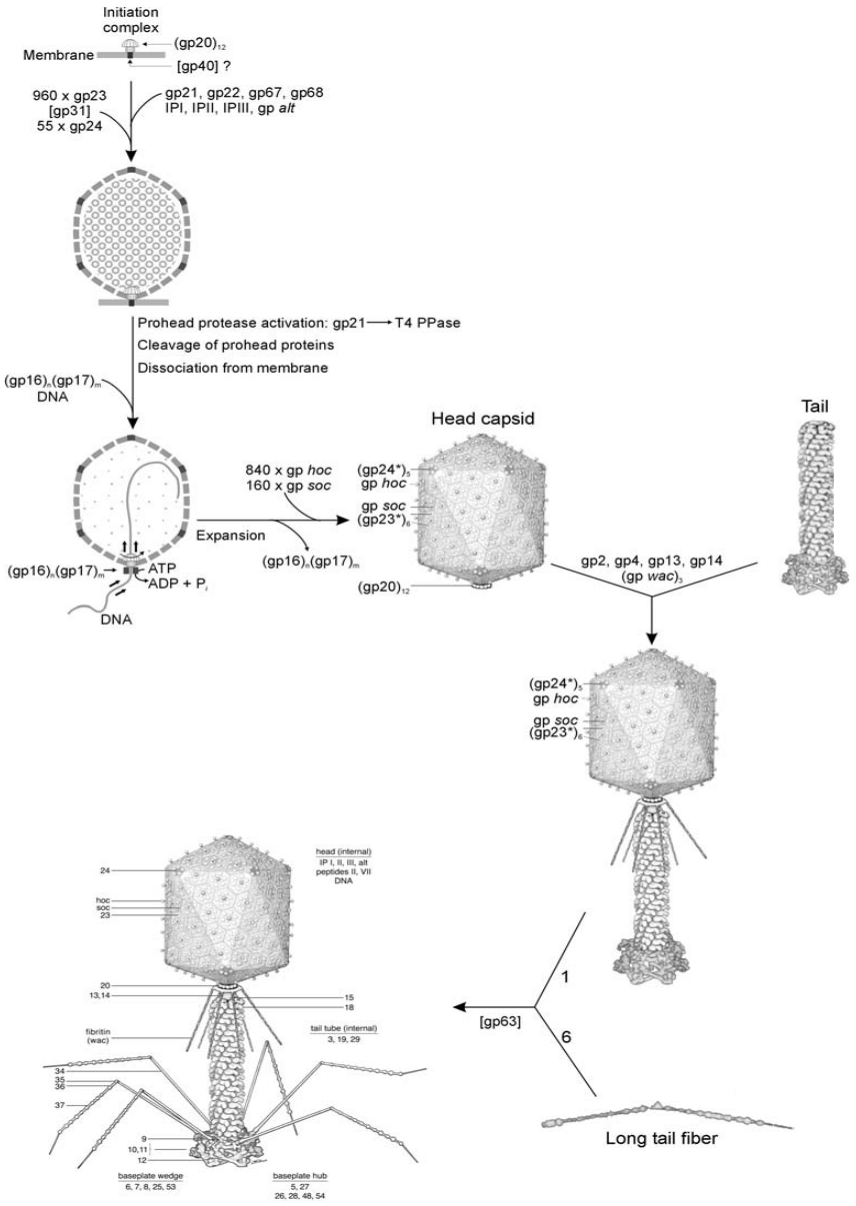


Figure 3. Morphogenesis of the bacteriophage T4 virion.[11] The overall assembly pathway can be divided into three independent stages: head, tail, and long tail fiber assembly. The chaperonines and catalytic proteins are indicated in brackets near the protein, or assembly step, that requires the chaperonine. Known protein stoichiometries are given as subscripts.

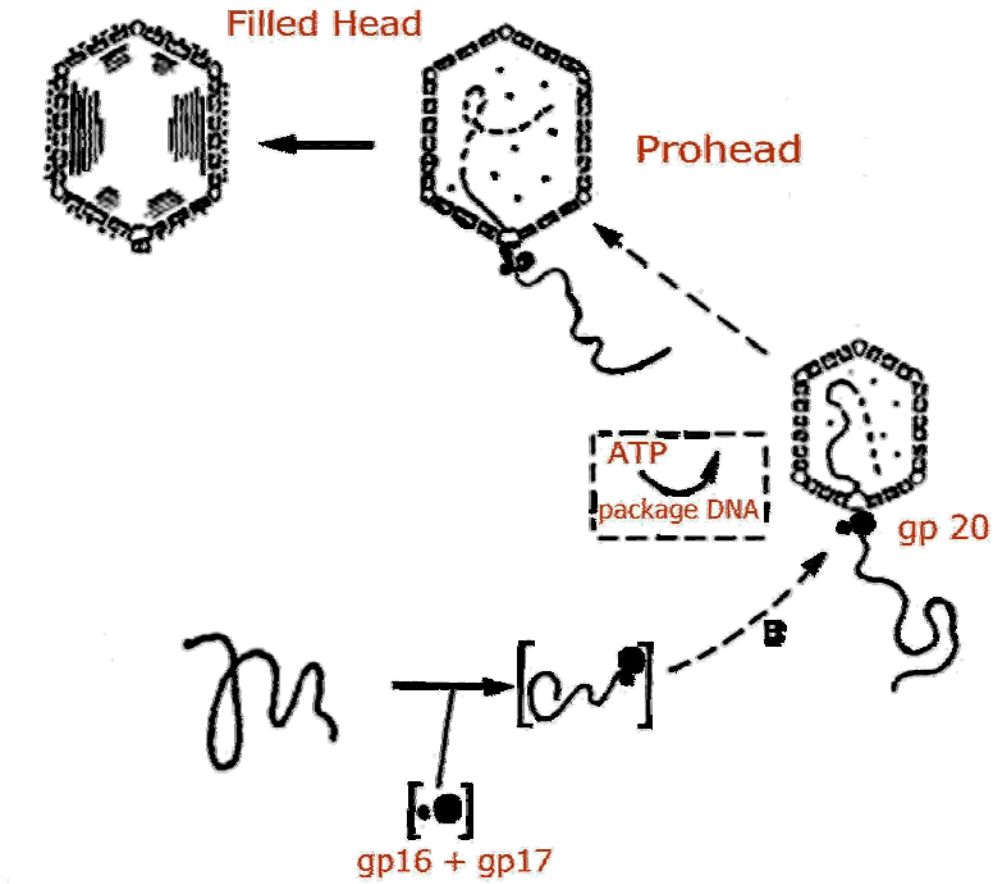


Figure 4. Bacteriophage T4 DNA packaging pathway (modified from Karam *et al.*[5]).

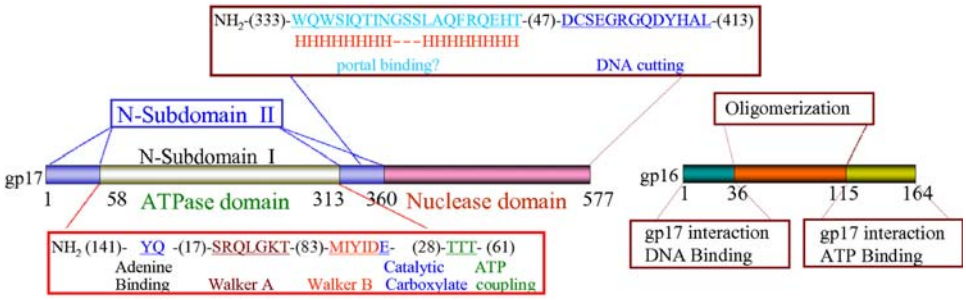


Figure 5. Domain organization of phage T4 terminase proteins.[38] Domain end points of the large terminase, gp17, were determined by the x-ray structures of ATPase and nuclease domains and the full-length gp17.[29, 40] The functional motifs were determined by mutational and biochemical studies.[30, 36, 37, 39] The domain organization of the small terminase, gp16, is predicted by this and previous studies.[38, 45] *Numbers* represent the number of amino acids in the respective coding sequence.

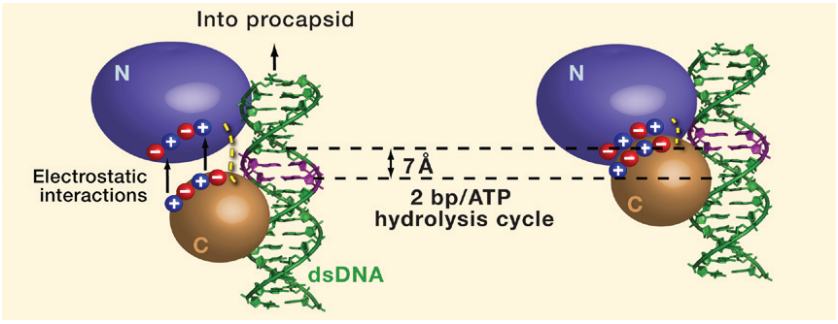


Figure 6. Schematic of two experimentally observed conformational states.[63] The tensed state (right) and the relaxed state (left) imply that a 7 Å translocation movement driven by electrostatic forces moves the DNA toward the procapsid by two base pairs per cycle of ATP hydrolysis.

No.	Phage	Protein	No. of amino acids	Identity percentage								
				1	2	3	4	5	6	7	8	9
(A) Small terminase subunit protein sequences												
1	T4	gp16	164	100	90	80	49	46	49	48	41	42
2	RB69	gp16	164		100	80	48	47	49	48	41	42
3	JS98	gp16	163			100	53	47	48	47	42	43
4	RB49	gp16	165				100	47	39	44	35	35
5	RB43	Truncated gp16	178					100	39	44	29	30
6	44RR2.8t	gp16	154						100	41	32	34
7	Aeh1	gp16	172							100	30	31
8	KVP40	gp16v	177								100	93
9	KVP20	gp16v	182									100
(B) Large terminase subunit protein sequences												
1	T4	gp17	610	100	91	83	70	69	66	60	53	53
2	RB69	gp17	611		100	84	70	69	66	58	53	53
3	JS98	gp17	611			100	68	69	66	60	51	51
4	RB49	gp17	607				100	66	67	58	50	50
5	RB43	gp17	609					100	70	59	54	53
6	44RR2.8t	gp17	613						100	60	51	51
7	Aeh1	gp17	633							100	54	54
8	KVP40	gp17v	600								100	97
9	KVP20	gp17v	600									100

Table 1. Small and large terminase subunit protein sequences.

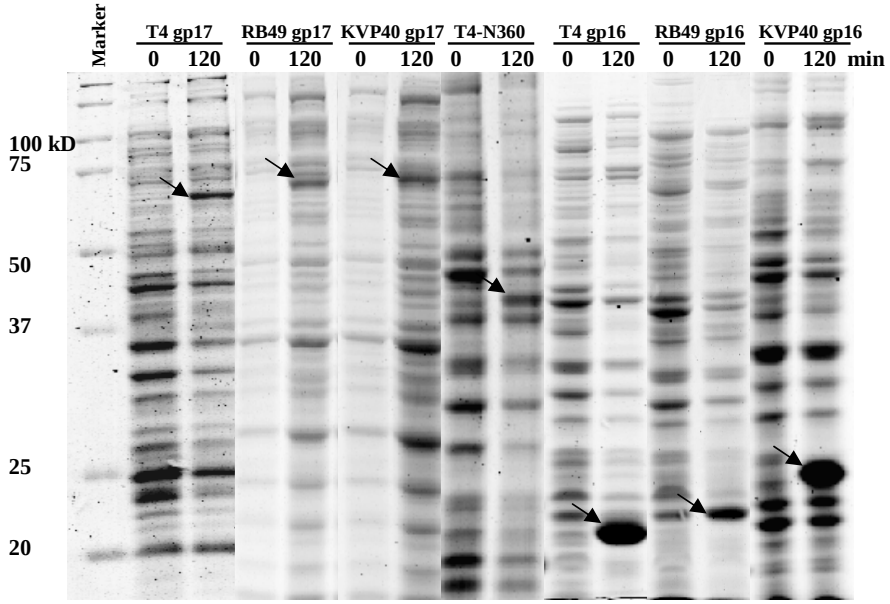


Figure 7. Overexpression of terminase proteins. The T4 family terminase proteins and T4 gp17-N domain (N360) were overexpressed in *E. coli* BL21 (DE3) pLysS cells. Cell cultures were grown to 4×10^8 cells/ml at 30°C and induced with 1 mM IPTG. 0 and 120 min culture aliquots (indicated by numbers at the top of the lanes) were obtained after induction and were analyzed with 10% SDS-PAGE. Arrows correspond to new protein bands overexpressed after IPTG induction. The molecular mass of each induced band is consistent with its predicted molecular weight.

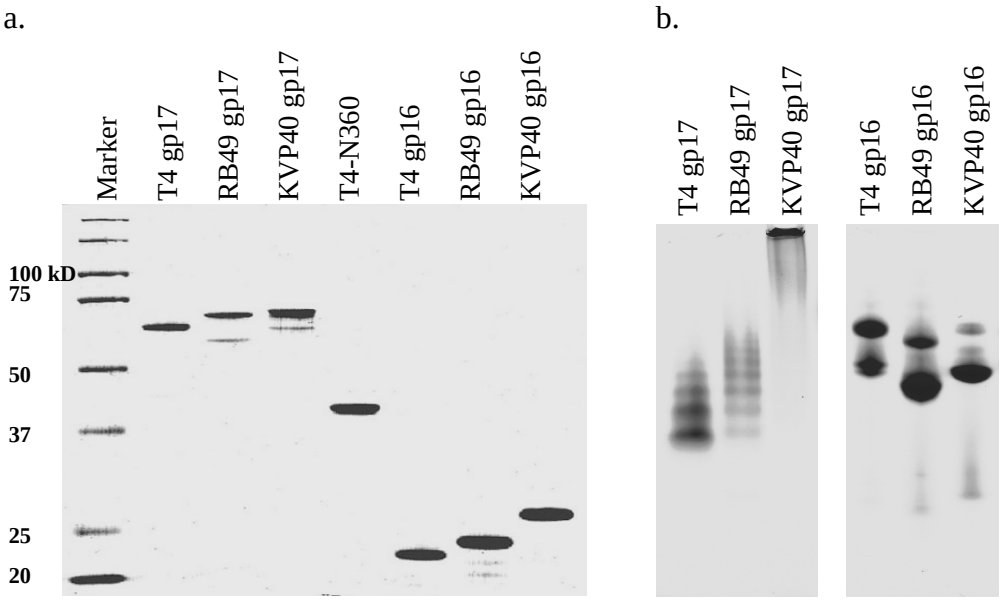


Figure 8. Purified terminase proteins. All the His-tagged terminase proteins were purified by Ni-agarose affinity chromatography followed by gel filtration as described in 'Materials and Methods'. Purified protein samples were electrophoresed on a 10% SDS-polyacrylamide gel (a) or a 4-12% non-denaturing native-polyacrylamide gel (b). The proteins were stained with Coomassie Blue R.

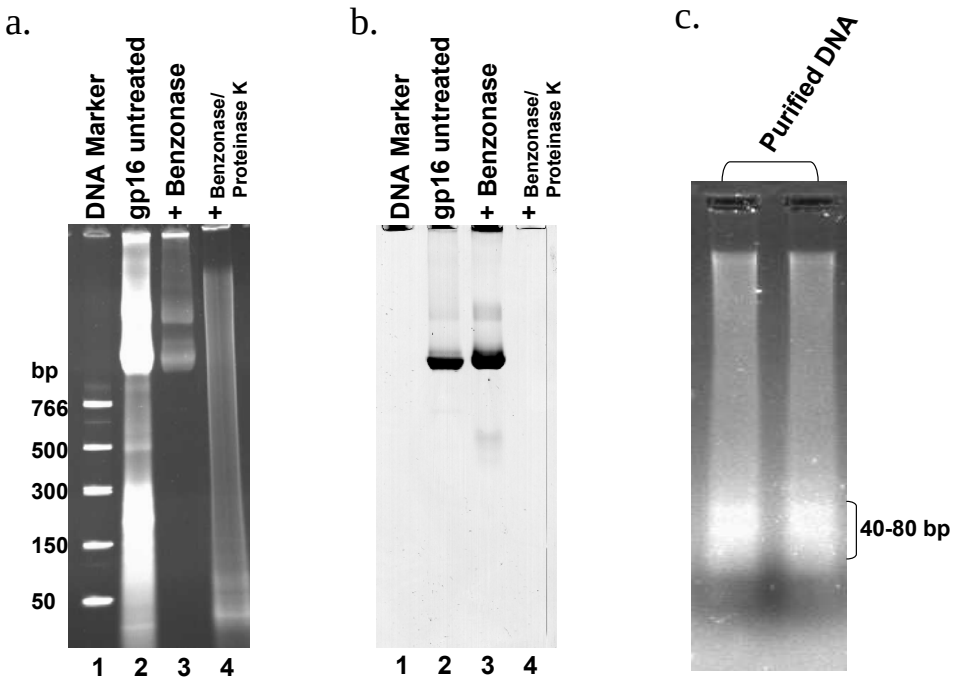


Figure 9. The purified RB49 gp16 was DNA bound. Purified RB49 gp16 was treated with Benzonase nuclease for overnight and passed through gel-filtration chromatography. The nuclease protected DNA was released by Proteinase K digestion and was purified with GeneJET™ DNA Purification Kit (Fermentas). See 'Materials and Methods' for details. The Benzonase and Proteinase K treated samples were run on a 4-20% PAGE gel (*lanes 3* and *4*) under non-denaturing condition along with DNA marker (*lane 1*) and untreated gp16 control (*lane 2*). The PAGE gel was stained with SYBR Green for DNA (*a*) followed by Coomassie Blue R staining for protein (*b*). The purified DNA was run on a 2% agarose gel in duplicates (*c*).

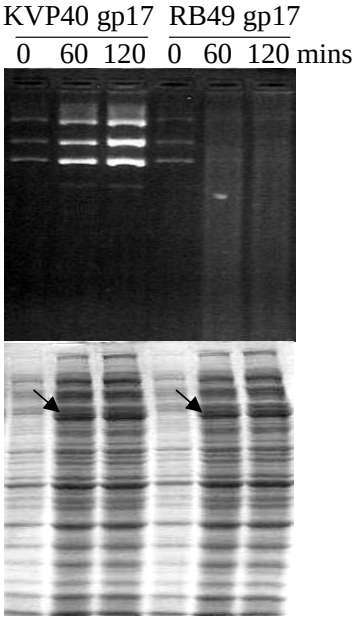


Figure 10. RB49 gp17 exhibits nuclease activity. The terminase proteins in the genetic background of *E. coli* BL21 (DE3) pLysS were grown to 4×10^8 cells/ml at 37°C and induced with 1 mM IPTG. Aliquots of cultures were collected at 0, 60, and 120 min following induction. Half of each aliquot was electrophoresed on a 12% SDS-PAGE gel to confirm overexpression (*lower panel*; arrows correspond to the new protein bands that are overexpressed following IPTG induction). From the rest of the aliquots miniprep plasmid DNAs were prepared by the alkaline lysis procedure and electrophoresed on a 0.8% (w/v) agarose gel (*upper panel*). Smeared appearance of DNA indicates nuclease activity.

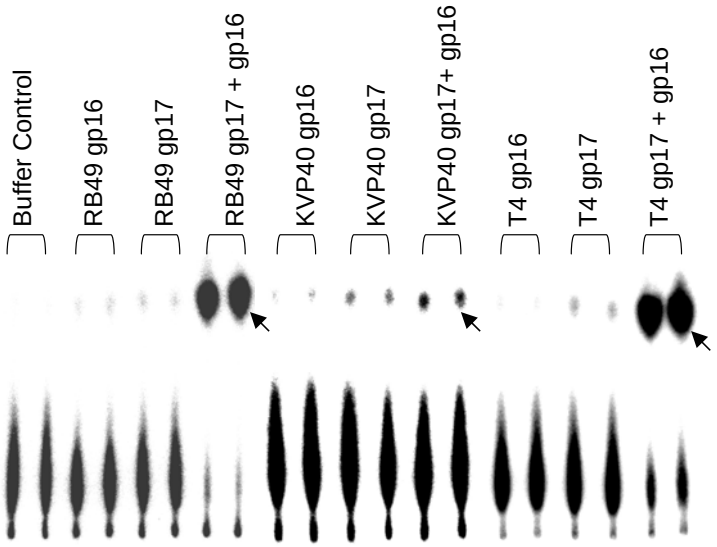


Figure 11. The terminases exhibit gp16-stimulated ATPase activity. Autoradiogram showing ATPase activity of gp17s. All ATPase assays were performed using the purified gp17s (1 μM) either alone or with homologous gp16s (10 μM) at a cold ATP concentration of 0.5 mM with trace amounts of [γ-^{32}P]ATP (75 nM). Each reaction was conducted in duplicate. Buffer control and gp16 controls were also included. *Arrows* correspond to the elevated ^{32}Pi produced by gp16-stimulated ATP hydrolysis. This result is a representation of four independent experiments.

Terminase proteins	% of ATP hydrolysis
RB49 gp17	1.7
RB49 gp17 + gp16	95
KVP40 gp17	3.2
KVP40 gp17 + gp16	6.1
T4 gp17	2.3
T4 gp17+ gp16	86

Table 2. Quantification of ATPase activity of gp17s. Quantification of the autoradiogram shown in Figure 4. The separated ^{32}Pi values were compared with the total ATP present for determination of the percent of ATP hydrolysis.

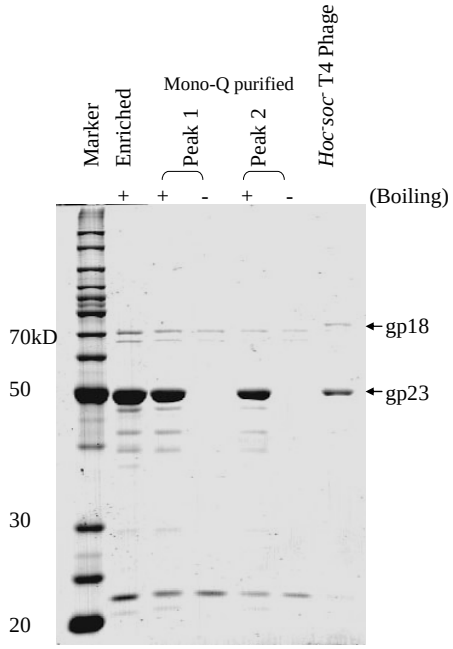


Figure 12. Purification of proheads. T4 proheads were produced by infection of *E. coli* P301 (*sup*$^-$) with *17am18am-rII* phage at 37°C to enrich for ELPs and were purified by ion-exchange chromatography (DEAE FF and Mono-Q columns; see 'Materials and Methods' for details). The proheads were eluted as two major peaks by a linear NaCl gradient from the Mono-Q column. The peak fractions were concentrated by high-speed centrifugation and were analyzed on a 10% SDS-PAGE gel. After addition of SDS-sample buffer, the samples were either kept at room temperature (unboiled; -) or in a boiling water bath (boiled; +) prior to electrophoresis. Enriched represents the pelleted proheads applied to chromatography. A known amount of *hoc*$^-$*soc*$^-$ T4 phage was used as control for determining the number of prohead particles yielded by quantifying gp23 bands.

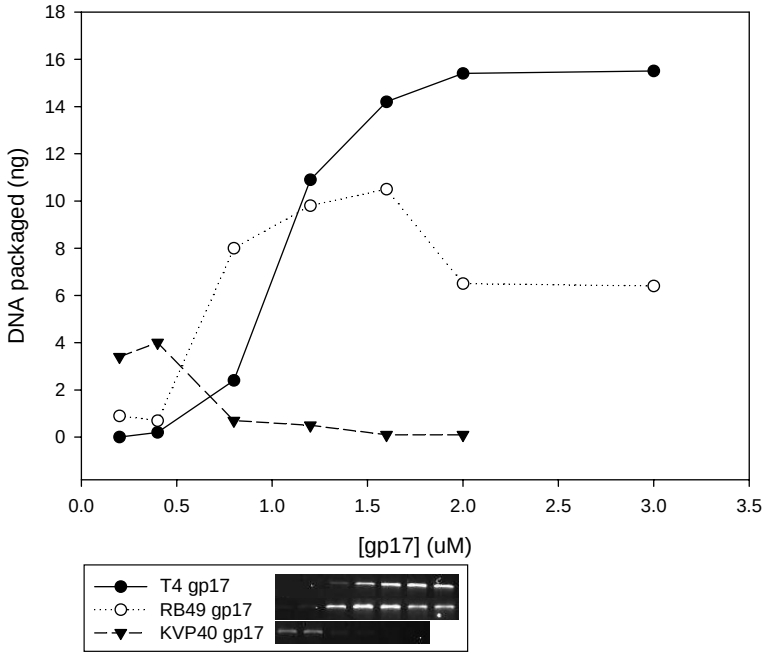


Figure 13. DNA packaging by the large terminases. In all the *in vitro* DNA packaging assays, the concentrations of purified gp17s were varied over a range of 0.1-3 μM in the presence of 10^9 prohead particles, along with 300 ng of linear λ DNA at a saturating concentration of ATP (1 mM). Packaged DNA was electrophoresed on an agarose gel, stained with ethidium bromide (*inset*) and quantified by Gel DOC XR imaging system. This result is a representation of three independent experiments.

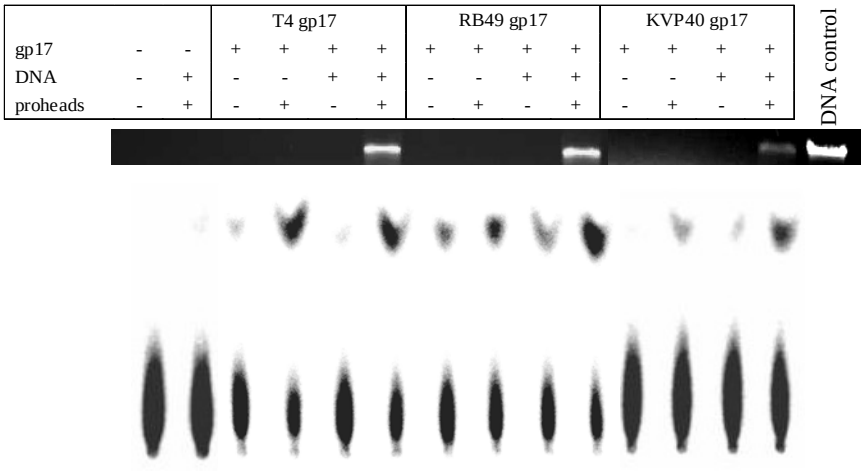


Figure 14. DNA packaging stimulates gp17-ATPase. The autoradiogram at the bottom shows the hydrolysis of ATP by T4, RB49 and KVP40 gp17s. The *inset* above the autoradiogram is the agarose gel image showing the DNA packaged under each condition. The + and - symbols in the box on top of the agarose gel image indicates the presence and absence of a component of the packaging system. The gp17 concentrations for each set of reactions are 1.0 μM, 1.0 μM and 0.5 μM for T4 gp17, RB49 gp17 and KVP40 gp17, respectively. This result is a representation of three independent experiments.

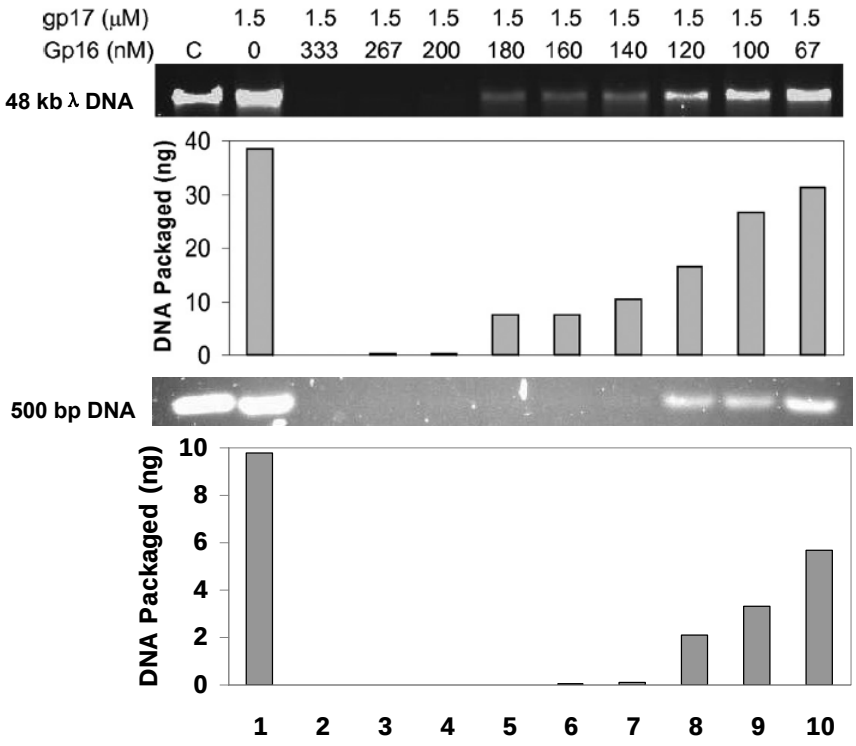


Figure 15. gp16 interaction to prohead-assembled gp17 inhibits DNA packaging. T4 gp17 was incubated alone (1.5 μM; 1.8×10^{13} molecules) or with increasing concentrations of T4 gp16 (67-333 nM; $1\text{-}5\times10^{11}$ gp16 oligomers, assuming gp16 an octamer[28]) in the presence of 6×10^{9} proheads and 300 ng DNA (phage λ DNA or 500 bp linear DNA) in the defined *in vitro* DNA packaging system. Control lanes C where 30 ng DNA (10% of the total DNA) was loaded were used for quantification of the packaged DNA, and the quantified data are shown in the histograms. This result is a representation of three independent experiments. (Published data.[38])

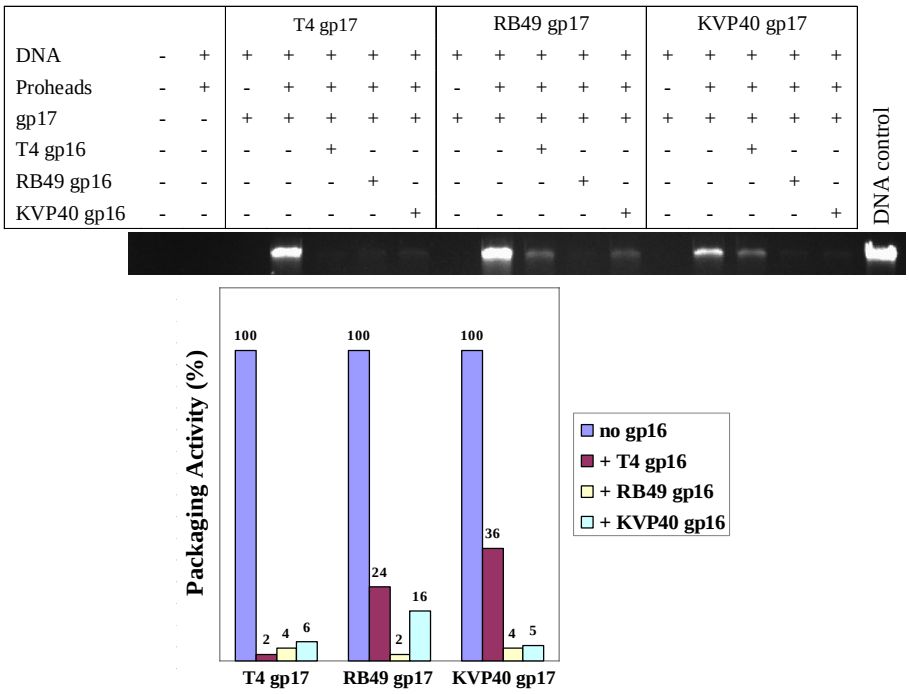


Figure 16. Inhibition of DNA packaging by T4 family gp16s. In reaction mixtures where 6×10^9 proheads and 300 ng λ DNA were present, gp17s (1.0 μM, 1.0 μM and 0.5 μM for T4 gp17, RB49 gp17 and KVP40 gp17, respectively) either alone or with gp16s (200 nM, ~3×10^{11} oligomers for each gp16) were incubated for DNA translocation. The + and - symbols on top of the agarose gel image inset indicates the presence and absence of a component. The histogram below shows the quantification of DNA packaging activity for each set of reactions according to the agarose gel band densities. The quantitative data were normalized to the activities without gp16 inhibition. This result is a representation of three independent experiments.

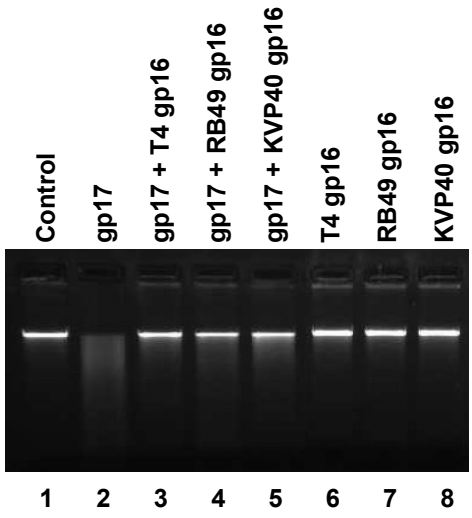
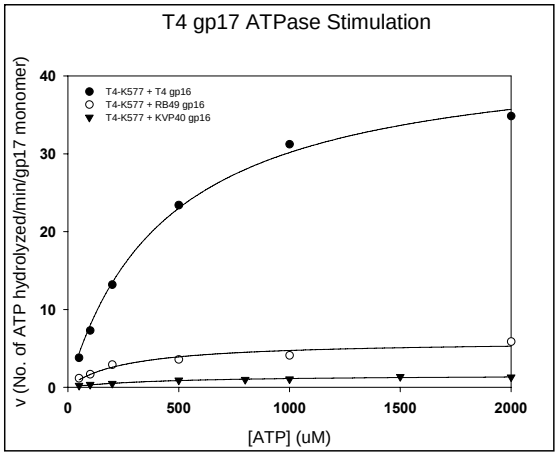


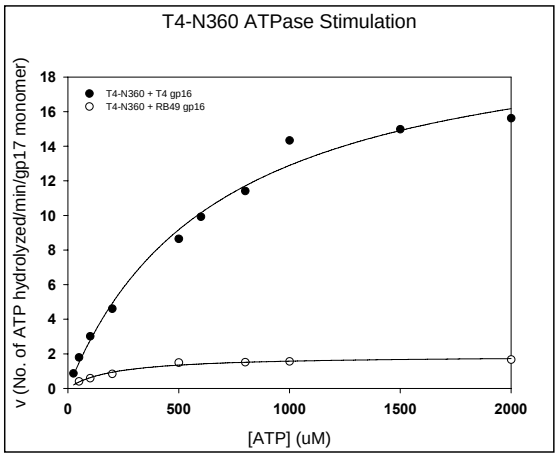
Figure 17. gp17 nuclease is regulated by gp16s. gp17 (T4 gp17, 1 μM) was incubated either alone or in the presence of different T4 family gp16s (5 μM for each gp16) in a reaction mixture containing 80 ng of pAD10 plasmid DNA (29 kb), 5 mM Tris-HCl pH 8.0, 6 mM NaCl and 5 mM $MgCl_2$. Controls include a sample with no gp16 or gp17 (*lane 1*) and samples without gp17 (*lanes 6, 7, 8*).

a.



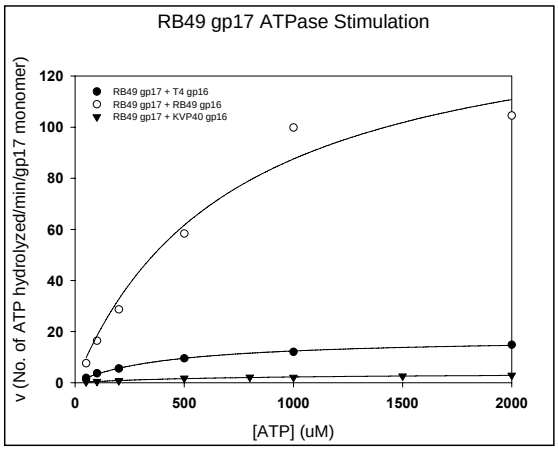
T4 gp17 +	K_m [1]	V_{max} [2]	%
T4 gp16	451	43	100%
RB49 gp16	264	6	14%
KVP40 gp16	434	1.6	4%

b.



N360 +	K_m	V_{max}	%
T4 gp16	614	21	100%
RB49 gp16	206	1.9	9%
KVP40 gp16	ND[3]	ND	ND

c.



RB49 gp17 +	K_m	V_{max}	%
T4 gp16	427	18	12%
RB49 gp16	725	152	100%
KVP40 gp16	713	3.8	2%

1) K_m (μM)
2) V_{max} (No. of ATP hydrolyzed/ min/ gp17 monomer)
3) Not detectable

Figure. 18. Combinations of T4 family terminases show specificity of ATPase stimulation. All ATPase assays were performed using purified gp17s (0.2-2 μM) with different gp16s at a gp16: gp17 molar ratio of 10:1. The concentration of cold ATP was varied (0.05-2 mM) in the presence of trace amounts of [γ-^{32}P]ATP (75 nM). The reaction rates were determined by phosphorimaging, and the kinetic parameters were determined by using SigmaPlot 8.0 software. See 'Materials and Methods' for details. *a*. ATPase activity of T4 gp17 stimulated by gp16s of T4, RB49 and KVP40. *b*. ATPase activity of T4-N360 stimulated by different gp16s. *c*. ATPase activity of RB49 gp17 stimulated by different gp16s. The curves in the charts represent the *v* for ATPase stimulation as a function of ATP concentration, and the kinetic parameters are listed to the right. The percentages (%) of ATPase stimulation were normalized to the combinations of homologous large and small terminases, based on the V_{max} values. Values represent average of duplicates from two independent experiments.

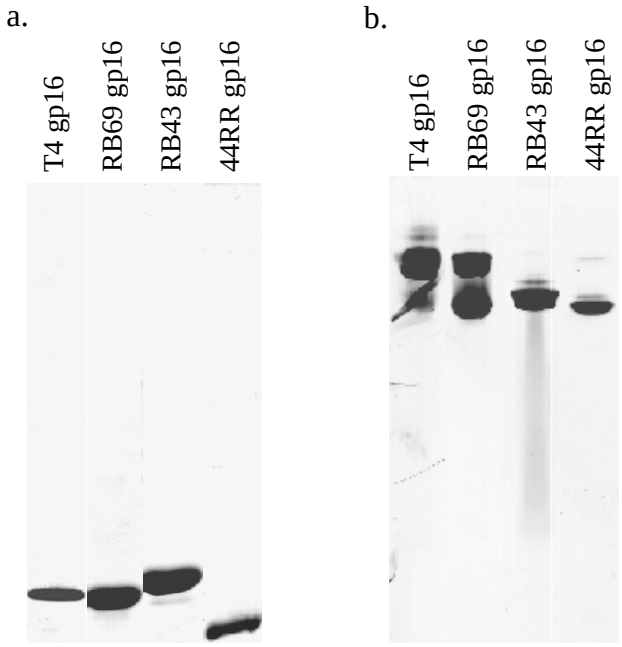


Figure 19. Purified T4-family small terminase proteins. *a*. Small terminase proteins were purified as described in the 'Materials and Methods' and electrophoresed on a 10% SDS-polyacrylamide gel. *b*. All gp16s formed oligomers, as shown on the 4-20% native non-denaturing gradient gel. Both gels were stained with Coomassie Blue R.

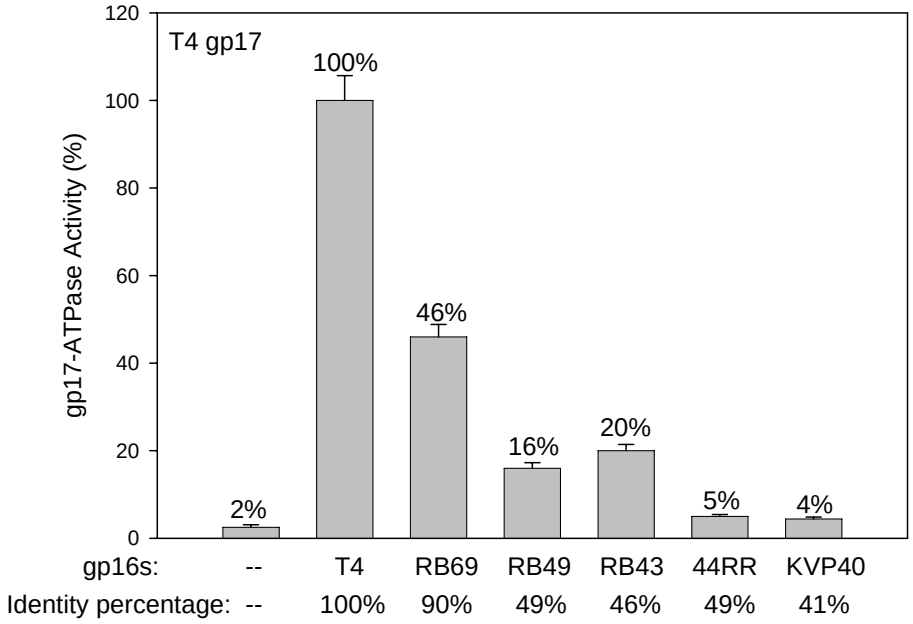


Figure 20. The specificity of gp17-ATPase stimulation is consistent with phage divergence. Histogram showing T4 gp17 ATPase activity stimulated by different gp16s as percentage of the T4-gp16 stimulated activity. Each T4 family gp16 used and its sequence identity percentage to T4 gp16 are shown at the *bottom* of columns. ATPase assays were performed using the purified T4 gp17 (0.6 μM) either alone or with gp16s (6 μM) at a cold ATP concentration of 0.6 mM with 75 nM [γ-^{32}P]ATP. Each reaction was conducted in duplicates. The amount of Pi produced was quantified and compared with the total ATP present for determination of the ATPase activity.

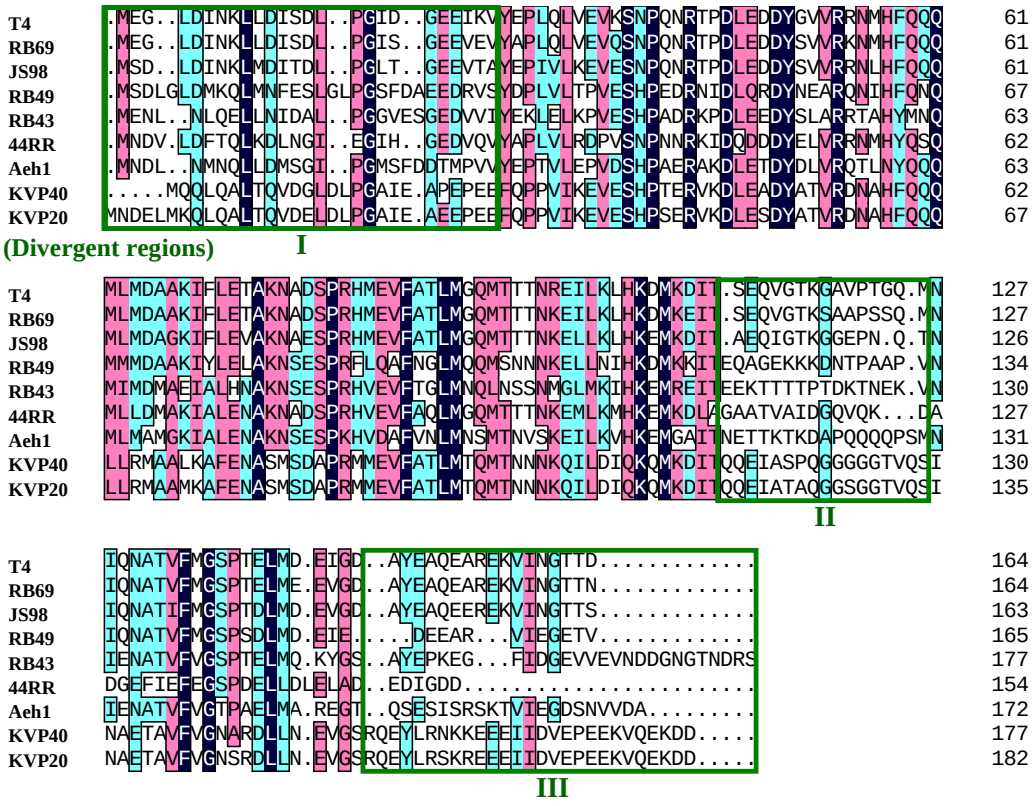


Figure 21. Alignment of T4 family gp16 sequences shows three divergent regions. T4 family gp16 sequences were selected for multiple sequence alignment by DNAMAN software. The three divergent regions suggested by the alignment are boxed in *green*. Sequences are shown after the name of each phage and the number of amino acid residues in each line is listed to the right. Amino acid residues with homology levels of 100%, 75% and 50% are highlighted in *black*, *magenta* and *cyan*, respectively.

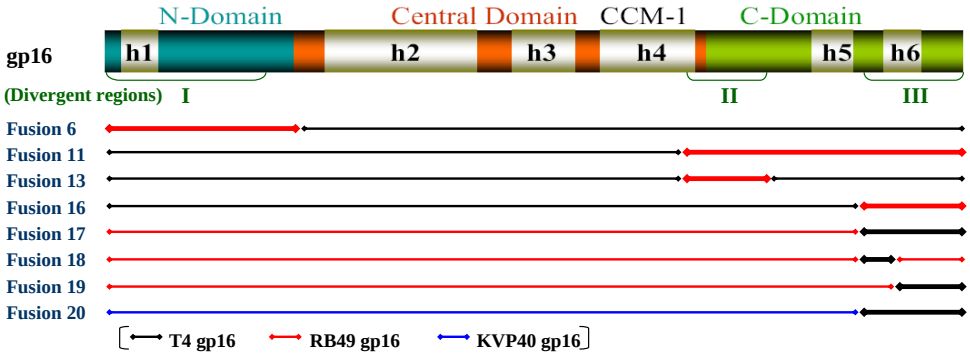


Figure 22. Configurations of gp16 fusion constructs. Schematic of gp16 polypeptide showing domain organization (N-domain in *blue*, central domain in *orange*, and C-domain in *green*), coiled-coil motif (CCM-1), predicted helices (h1–h6), and divergent regions (I, II and III).[38] The configurations of gp16 fusion polypeptides are represented by double arrow lines with different colors (*black*, T4 gp16 sequence; *red*, RB49 gp16 sequence; *blue*, KVP40 gp16 sequence. e.g. Fusion 6 was constructed with the N-domain sequence of RB49 gp16 followed by the T4 gp16 sequence from the start of the central domain).

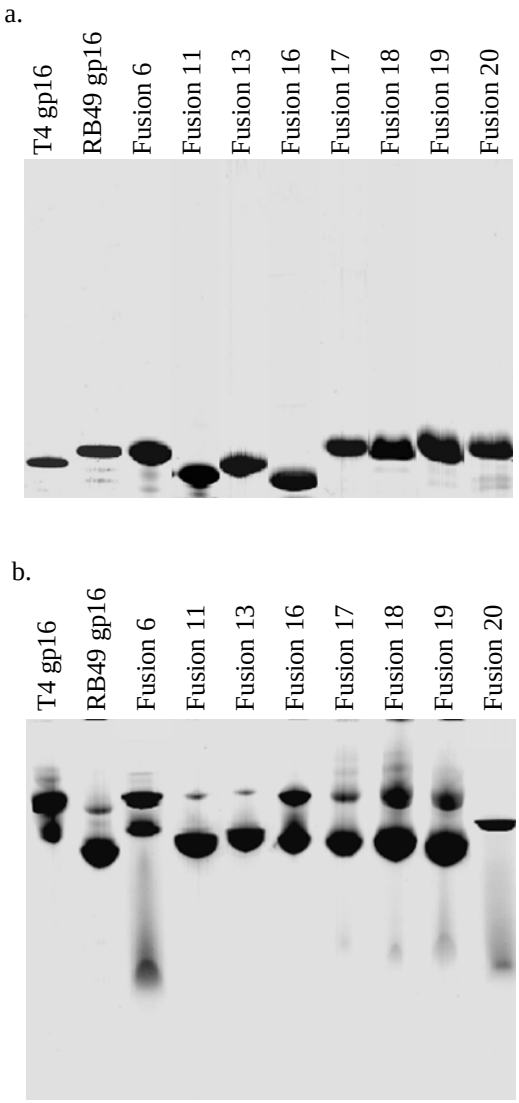


Figure 23. Purified gp16 fusion constructs. *a*. SDS-polyacrylamide gel (10%) showing the purified fusion proteins. Cloning, overexpression and purification were performed as described in 'Materials and Methods'. *b*. 4-12% native non-denaturing gradient gel showing the oligomeric state of the fusion proteins. Both the gels were stained with Coomassie Blue R. All of the fusion proteins produced oligomers.

gp16 fusion proteins	Description	T4 gp17 ATPase stimulation activity (% of WT)	RB49 gp17 ATPase stimulation activity (% of WT)
T4 gp16	Wild-type	100	20.1
RB49 gp16	Wild-type	23.8	100
Fusion 6	Region I: T4 -> RB49	106.6	23.8
Fusion 11	Regions II & III: T4 -> RB49	36.9	116.2
Fusion 13	Region II: T4 -> RB49	73.2	16.8
Fusion 16	Region III: T4 -> RB49	55.7	74.2
Fusion 17	Region III: RB49 -> T4	106	25.2
Fusion 18	Part of region III: RB49 -> T4	26	86
Fusion 19	Part of region III: RB49 -> T4	44.8	95.3
Fusion 20	Region III: KVP40 -> T4	22	19.2

Table 3. Divergent region III at the C-terminus of gp16 is essential for the gp17-ATPase stimulation specificity. T4 and RB49 gp17 ATPase stimulation activity of gp16 fusion proteins as percentage of WT. WT (wild-type) refers to the ATPase activity stimulated by homologous gp16s (e.g. the ATPase activity of T4 gp17 stimulated by T4 gp16). Values represent average of duplicates from two independent experiments. Detailed kinetic parameters are listed in Appendix 2.

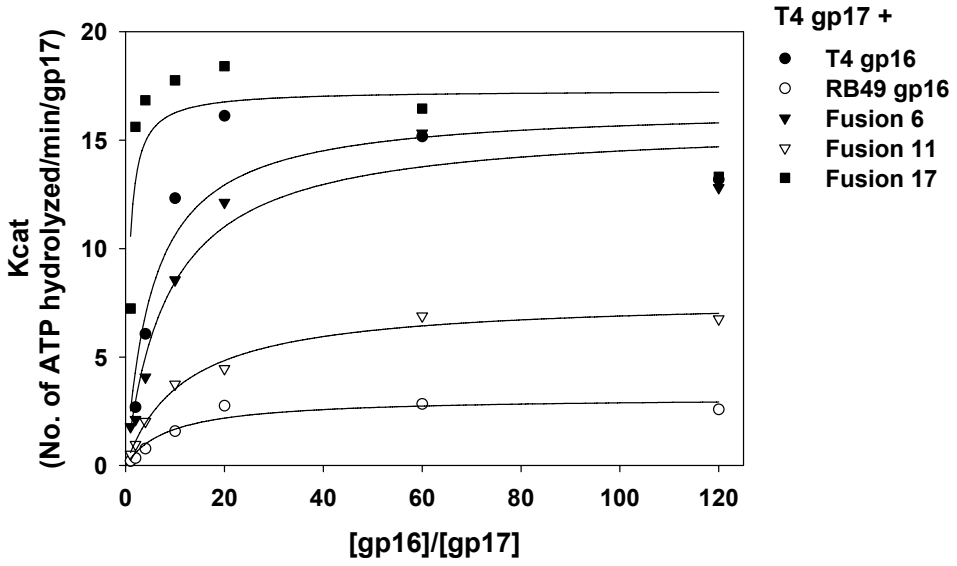


Figure 24. K_{cat} for stimulated T4 gp17 ATPase as a function of [gp16]: [gp17] ratio. T4, RB49 and three fusion gp16s were tested. gp16 concentration was varied at a constant gp17 concentration and a saturating ATP concentration. Values represent average of duplicates from two independent experiments. See 'Materials and Methods' for details on ATPase assays.

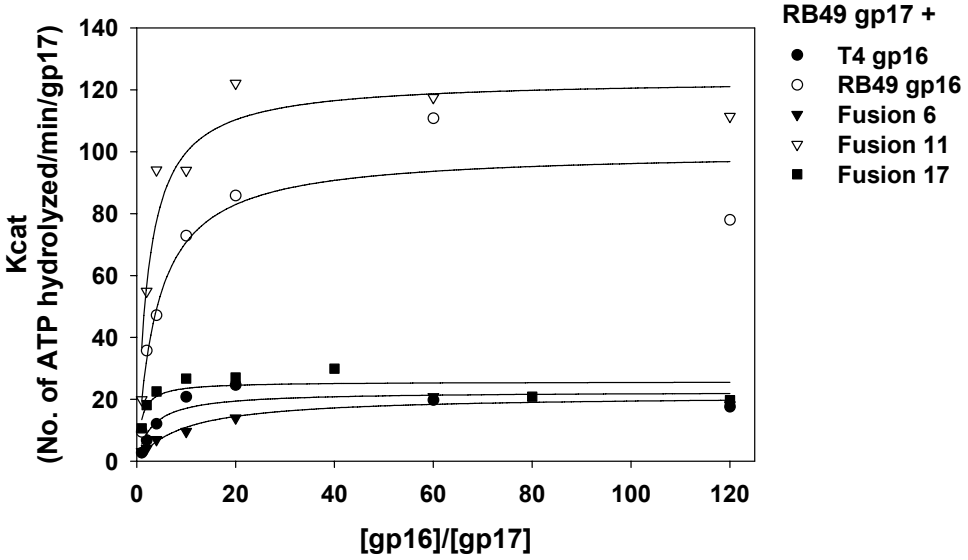


Figure 25. K_{cat} for stimulated RB49 gp17 ATPase as a function of [gp16]: [gp17] ratio. T4, RB49 and three fusion gp16s were tested. gp16 concentration was varied at a constant gp17 concentration and a saturating ATP concentration. Values represent average of duplicates from two independent experiments. See 'Materials and Methods' for details on ATPase assays.

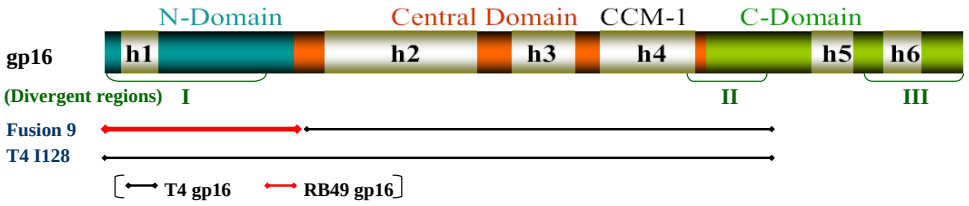


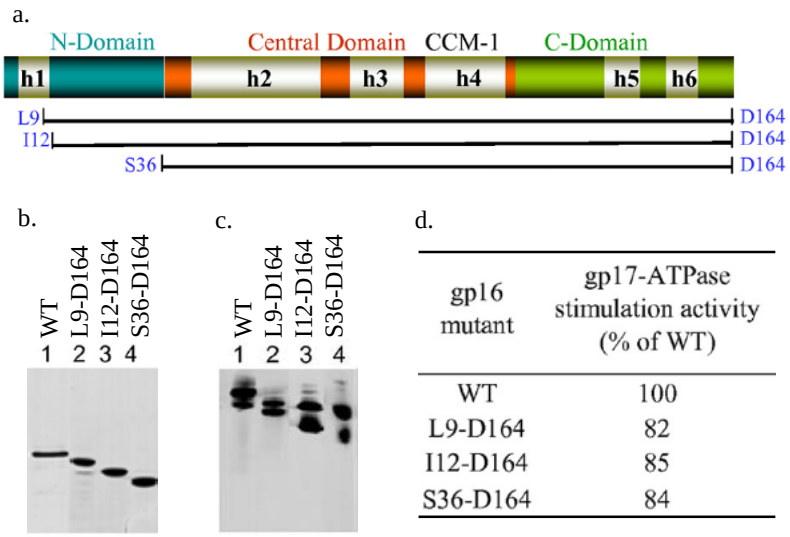
Figure 26. Configurations of gp16 truncated fusion recombinants. Schematic of gp16 polypeptide showing the configurations of gp16 recombinants (see Fig. 22 legend for description of the schematic). The C-terminal truncation endpoint is the T4 gp16 isoleucine 128 residue.

Figure 27. Purified gp16 truncated fusion recombinants. SDS-polyacrylamide gel (12%) showing the purified gp16 recombinants. The gel was stained with Coomassie Blue R. Cloning, overexpression and purification were performed as described in 'Materials and Methods'.

gp16 fusion proteins	T4 gp17 ATPase stimulation activity (% of WT)	RB49 gp17 ATPase stimulation activity (% of WT)
T4 gp16	**100**	**20.1**
T4 I128	**62.8**	**5.5**
Fusion 9	**ND***	**10.7**

*** Not detectable**

Table 4. N-domain of gp16 shows gp17-ATPase stimulation specificity when C-domain is truncated. T4 and RB49 gp17 ATPase stimulation activity of gp16 recombinants as percentage of WT. WT (wild-type) refers to the ATPase activity stimulated by homologous gp16s (e.g. the ATPase activity of T4 gp17 stimulated by T4 gp16). Values represent average of duplicates from two independent experiments. The detailed kinetic parameters are listed in Appendix 2.



d.

gp16 mutant	gp17-ATPase stimulation activity (% of WT)
WT	100
L9-D164	82
I12-D164	85
S36-D164	84

Figure 28. The gp16 N-domain truncation mutants show partial loss of gp17-ATPase stimulation. *a*. Schematic of the gp16 polypeptide showing the endpoints of the N-terminal truncation clones (T4 gp16 L9-D164, I12-D164, and S36-D164; see Fig. 22 legend for description of the schematic). *b*. SDS-polyacrylamide gel (12%) showing the purity of mutant proteins. Cloning, overexpression and purification procedures are described in 'Materials and Methods'. *c*. 4-20% native non-denaturing gel showing the oligomeric state of gp16 mutants. Both the gels were stained with Coomassie Blue R. All the mutants produced oligomers. *d*. T4 gp17-ATPase stimulation activity of gp16 mutants expressed as percentage of the WT T4 gp16. Values represent average of duplicates from two independent experiments. The detailed kinetic parameters are listed in Appendix 2. (Published data.[38])

Figure 29. Construction of a fusion gp17. Schematic of the domain organization of the T4 large terminase, gp17, and the construction of a fusion gp17 clone (represented by double arrow lines; *black*, T4 gp17 sequence; *red*, RB49 gp17 sequence; the N-terminal 85 amino acids were changed). gp17 consists of two domains,[44] the N-domain (ATPase domain, amino acids 1-360) and the C-domain (Nuclease domain, amino acids 361-577, in *magenta*). The N-domain is divided into N-Subdomain I (in *white*) and N-Subdomain II (in *blue*), based on the x-ray structures of the ATPase domain and the full-length gp17.[29, 40]

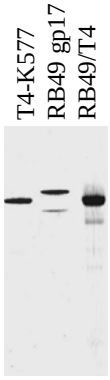


Figure 30. Purified gp17 fusion protein. SDS-polyacrylamide gel (10%) showing the purity of the fusion protein. The gel was stained with Coomassie Blue R. Cloning, overexpression and purification procedures are described in 'Materials and Methods'.

gp16	RB49/T4 fusion gp17 ATPase stimulation activity		
	K_m (mM)	K_{cat} (No. of ATP hydrolyzed/ min/ gp17 monomer)	K_{cat}/K_m (%)
T4 gp16	**0.56**	**89**	**159 (100%)**
RB49 gp16	**0.22**	**27**	**123 (77%)**

Table 5. Change of the N-terminal 85 amino acid sequence of gp17 affects the ATPase stimulation specificity. ATPase activity of the fusion gp17 stimulated by T4 and RB49 gp16s. ATPase assays were performed using the fusion gp17 (0.5 μM) with gp16 (5 μM) at increasing concentrations of cold ATP (0.05-2 mM) and trace amounts of [γ-^{32}P]ATP (75 nM). The reaction rates were determined by phosphorimaging, and the kinetic parameters were determined by SigmaPlot 8.0 software. Values represent average of duplicates from two independent experiments. The percentage (%) of stimulation was based on the K_{cat}/K_m values with the higher value arbitrarily set as 100%.

Peptide sequence*	Clone number	Selection method** (number of clones)
KVVQLSQMPQMM-	5-2	c(3)
KVFTLYPPNANV-	5-3	b (1), c(3)
NAKVWTVPSKPP-	5-5	a (6), c(11)
KLWTLPLSQTHP-	5-6	b (1), c(4)
KLWQIPDLSFLK-	5-8	c(2)
QPHKVFFPNLPR-	6-1	c(2)
KVWPWHSEQIIP-	6-3	b (1), c(1)
HLLSPELKLLAR-	6-4	a (1), c(2)
DFPARLMPYGFL-	7-21	c(2)
YSIRHDYPLRFW-	7-24	a (1), c(2)
APTKCCFAQSPG-	8-21	c(3)
HSLRPEWRMPGP-	8-24	a (5), b (1), c(3)
VNAKCCHAYLAR-	8-29	a (1), c(1)
GLKIWSLPPHHG-	F1-1	a(2), b(1)

* Sequences of the phages which were affinity re-isolated at least two times are listed.
** *a*, *b* and *c*, three different panning selection methods used in this study. Method a, T4 gp17 was used for competitive elution in the final (4th) panning round. Method b, gp17 competitive elution for both panning rounds. Method c, no competitive elution involved. See 'Materials and methods' for details. *Number of clones* in parentheses reflects the number of times the sequence selection occurred with a particular method.

Table 6. Amino acid sequences of 14 phage-displayed peptides isolated by biopanning against T4 gp16.

Domain and function	Region of interaction	Alignment of peptide sequences with respective T4 gp17 region* (clone number)		Binding to gp16 (by ELISA)
N-Subdomain II, ATPase	37-52	**HWIKSQW**DGK**WYPEKF**		
		---K-VWP--WHSEQIIP	(6-3)	+
		HSLRPEWRMPGP	(8-24)	+
N-Subdomain I, ATPase	60-77	**KIVKI**PNNS**DKPELFQTY**		
		KVVQLS---QMPQMM	(5-2)	-
		------------KVFTLYPPNANV	(5-3)	+
N-Subdomain I, ATPase	90-98	------**LPNLKRANI**		
		KVFTLYPPN---ANV	(5-3)	+
		QPHKVFFPNLPR	(6-1)	-
N-Subdomain I, ATPase	174-187	-**A**HFV**CF**NKDKAV**G**		
		APTKCCFAQ--SPG	(8-21)	+
N-Subdomain I, ATPase	290-315	**GL**NHF**YDIWTAAVEGKSGFEPYTAIW**		
		----NAKVWT--VPSKP---P	(5-5)	+
		------KLWTLPLSQT---HP	(5-6)	+
		-----YSI-------RHDY-PLR-FW	(7-24)	+
		GL----KIWS--------LPPHHG	(F1-1)	+
C-domain, Nuclease	436-453	-**HLILPDIVMRYLVE**YNE**C**		
		KLWQIPDLS--FLK	(5-8)	-
		-HLLSPELKL--LAR	(6-4)	+
		------------VN-AKCCHAYLAR	(8-29)	+
C-domain, Nuclease	542-553	**DL**VMS**LVIFGWL**		
		DFPARLMPYGFL	(7-21)	+

* *Grey* indicates identical amino acid residues; *yellow* indicates similar amino acid residues by computer match (DNAMAN software, Lynnon BioSoft). The gp17 amino acid residues matching the aligned sequences are shown in bold.

Table 7. T4 gp17 amino acid sequence regions that were identified as potential candidates to interact with gp16.

Large terminase	Oligomeric state*	gp16-stimulated ATPase	Nuclease	DNA packaging & packaging stimulated ATPase
T4 gp17	Monomers & multimers	Good	Yes	Yes
RB49 gp17	Monomers & multimers	Good	Yes	Yes
KVP40 gp17	Aggregation	Marginal	Not detectable	Yes (at low concentrations)
Small terminase		ATPase stimulation	Nuclease regulation	Inhibition of DNA packaging
T4 gp16	Oligomers	Yes	Yes	Yes
RB49 gp16	Oligomers	Yes	Yes	Yes
KVP40 gp16	Oligomers	--	Yes	Yes

* Based on the Native-PAGE analyses of the purified proteins.

Table 8. Functional properties of the purified T4 family terminase proteins.

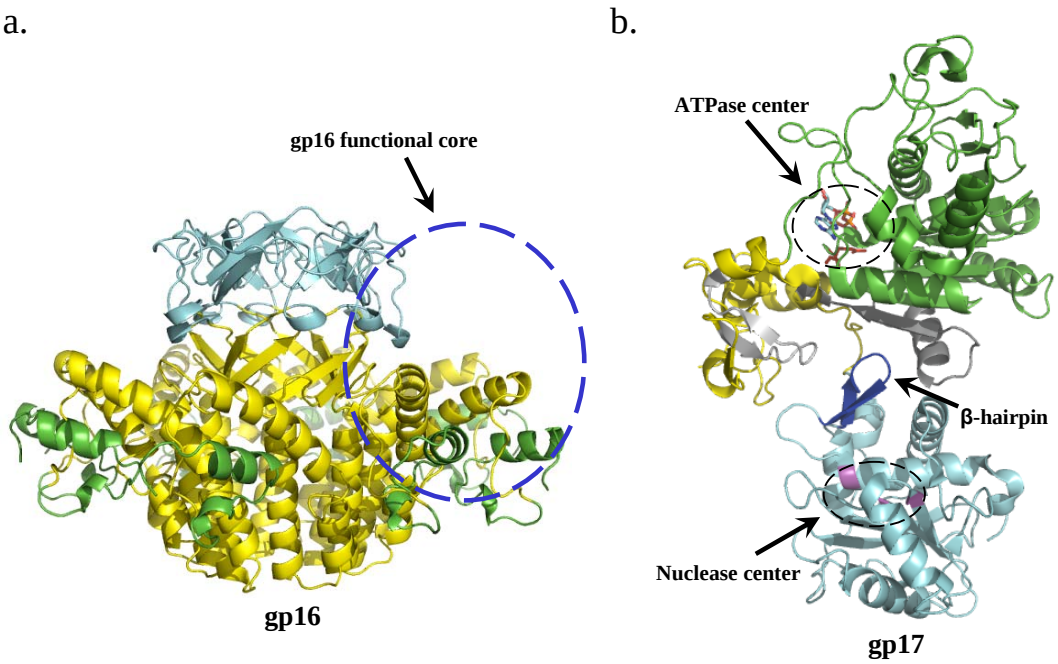


Figure 31. The possible gp16-gp17 interaction sites are clustered at a key position. *a.* Ribbon diagram of the T4 gp16 octamer model constructed based on x-ray structure of the phage Sf6 small terminase (RCSB PDB code 3hef) (model by courtesy of B. Draper). *Green*, the N-domain. *Yellow*, the central domain. *Cyan*, the C-domain. *Blue circle* indicates the possible position of gp16 functional core. *b.* Ribbon diagram of x-ray structure of T4 gp17 (RCSB PDB code 3cpe). N-Subdomain I is in *green*. N-Subdomain II is in *yellow*. C-domain is in *cyan*. The putative gp16 interaction regions are in *grey*. *Blue*, the β-hairpin that may control the nuclease activity. *Magenta*, the nuclease catalytic residues in the circled nuclease center. Arg-162 is shown as *orange sticks* in the circled ATPase center. The ribbon diagrams were generated with the PyMOL program.[64]

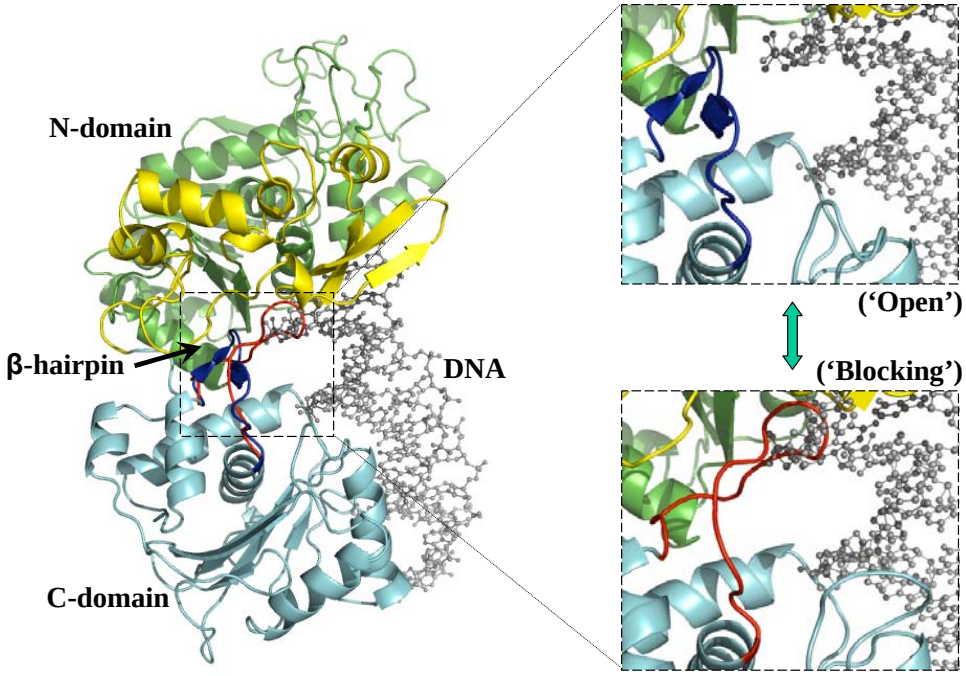


Figure 32. Ribbon diagram of x-ray structure of T4 gp17 showing the flexible β-hairpin blocking the nuclease DNA binding groove. *Green*, the N-Subdomain I. *Yellow*, the N-Subdomain II. *Cyan*, the C-domain. A short dsDNA molecule (shown as *grey balls and sticks*) is modeled into the nuclease DNA binding groove. The β-hairpin is magnified in the *insets* showing an 'open' conformation (in *blue*) as determined in the structure, or a possible 'blocking' conformation (in *red*) that was modeled (model by courtesy of B. Draper). The figure was generated with the PyMOL program.[64]

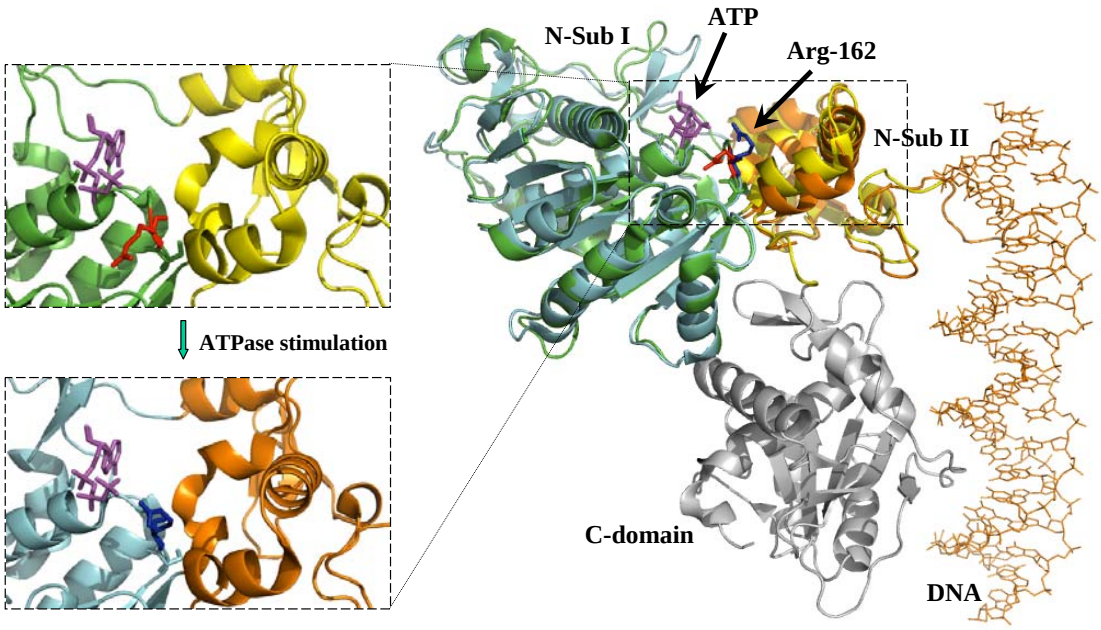


Figure 33. Ribbon diagram of x-ray structure of T4 gp17 showing the conformational change of N-Subdomain I and II affecting the position of Arg-162. An N-domain structure[40] (N-SubI, *cyan*; N-SubII, *orange*) with the Arg-162 (*blue*) at active position (RCSB PDB code 2o0h) is superpositioned into the T4 gp17 structure (N-SubI, *green*; N-SubII, *yellow*; C-domain, *grey*) where the Arg-162 (*red*) is at non-stimulation position. The Arg-162 in the two different orientations is magnified in the *insets*. ATP is shown as *magenta sticks*. A dsDNA molecule (shown as *orange sticks*) is modeled into the DNA translocation groove. The figure was generated with the PyMOL program.[64]

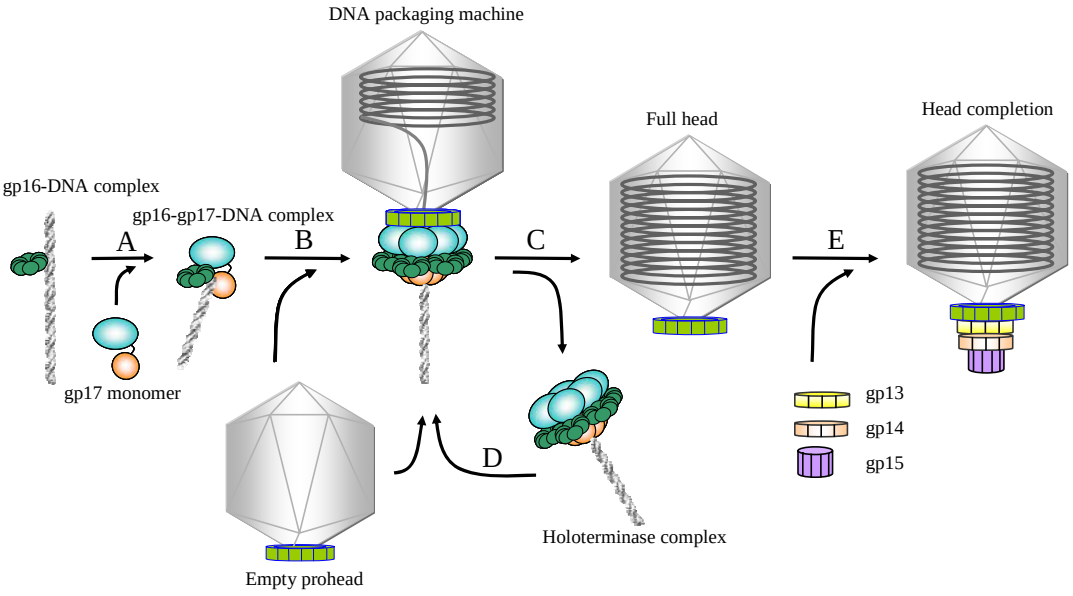


Figure 34. Model for packaging implicating gp16 in assembly and regulation of phage T4 DNA packaging machine. See Discussion for details. gp16 oligomers are depicted in *green*. The N- and C- domains of gp17 are shown in *cyan* and *orange*, respectively. Packaged DNA is represented as a *barrel* inside the prohead. The *green collar* at the unique vertex of prohead represents the dodecameric portal formed by 12 molecules of gp20. The head completion products, gp13, gp14, and gp15, are shown as *rings* added to the open surface of portal; the stoichiometry of these complexes is unknown.

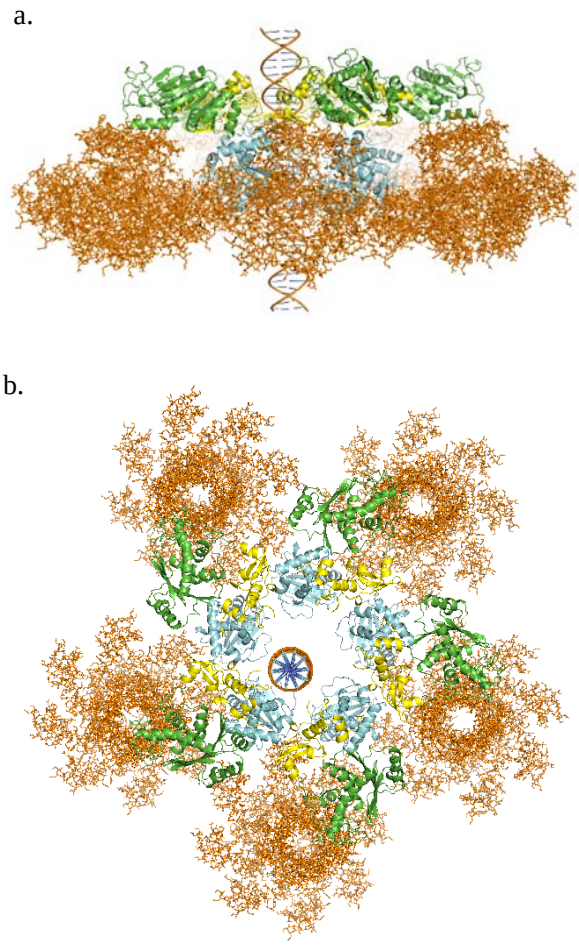


Figure 35. The speculated complete T4 DNA packaging machine. *a.* The organization of the complete packaging machine shown with the ribbon diagram of the pentameric T4 gp17 model from EMfit (RCSB PDB code 3ezk), a modeled dsDNA molecule in the central channel of the gp17 pentamer, and the modeled T4 gp16 octamers shown as *orange sticks*. The N-SubI, N-SubII and C-domain of gp17 are colored with *green*, *yellow* and *cyan*, respectively. This is a side view. *b.* Top view of the complete DNA packaging machine. The diagrams were generated with the PyMOL program.[64]

Appendix 1

RB49 gp17

Forward: 5′-*CGC***GGATCC**AATGAGCGTGATTGAAGGGATT-3′ (nt 1-21)

Reverse: 5′-*CGC***GGATCC**TTATATCATTGTTAGACCTTC-3′ (nt 1804-1824)

RB49 gp16

Forward: 5′-*CCATCA***CATATG**ATGTCAGATCTAGGCTTAGAT-3′ (nt 1-21)

Reverse: 5′-*CGC***GGATCC**TCATACTGTCTCACCTTCAAT-3′ (nt 478-498)

KVP40 gp17

Forward: 5′-*CGC***GGATCC**AATGTCGAGCCAGAAGAAAAAG-3′ (nt 1-21)

Reverse: 5′-*CGC***GGATCC**CTAAAACATACCGAAACTATT-3′ (nt 1783-1803)

KVP40 gp16

Forward: 5′-*CCATCA***CATATG**ATGCAACAACTTCAGGCATTG-3′ (nt 1-21)

Reverse: 5′-*CGC***GGATCC**TTAATCGTCTTTTTCTTGAAC-3′ (nt 514-534)

Fusion 6

Forward: 5′-*CCATCA***CATATG**ATGTCAGATCTAGGCTTAGAT-3′

Reverse: 5′-*CGC***GGATCC**TTAATCGGTTGTTCCATTTATCAC-3′

Stitch forward: 5′-CTCGTCTTGACTCCGGTCGAAAGCAATCCACAAAACCGT

ACT-3′

Stitch reverse: 5′-AGTACGGTTTTGTGGATTGCTTTCGACCGGAGTCAAGACG

AG-3′

Fusion 9

Forward: 5′-*CCATCA***CATATG**ATGTCAGATCTAGGCTTAGAT-3′

Reverse: 5′-*CGC***GGATCC**TTAAATATTCATTTGACCTGTAGG-3′

Stitch forward: 5′-CCGGTCGAAAGCCATCCAGAAAACCGTACTCCAGACTT

AGAA-3′

Stitch reverse: 5′-TTCTAAGTCTGGAGTACGGTTTTCTGGATGGCTTTCGACC

GG-3′

Fusion 11

Forward: 5′-*CCATCA***CATATG**ATGGAAGGTCTTGATATAAACAAA-3′

Reverse: 5′-*CGC***GGATCC**TCATACTGTCTCACCTTCAAT-3′

Stitch forward: 5′-CATAAAGATATGAAAGACATTACAGAGCAAGCTGGAGA

GAAAAAGAAA-3′

Stitch reverse: 5′-TTTCTTTTTCTCTCCAGCTTGCTCTGTAATGTCTTTCATATC

TTTATG-3′

Fusion 13

Forward: 5′-*CCATCA***CATATG**ATGGAAGGTCTTGATATAAACAAA-3′

Reverse: 5′-*CGC***GGATCC**TTAATCGGTTGTTCCATTTATCAC-3′

Stitch 1 forward: 5′-CATAAAGATATGAAAGACATTACAGAGCAAGCTGGAG

AGAAAAAGAAA-3′

Stitch 1 reverse: 5′-TTTCTTTTTCTCTCCAGCTTGCTCTGTAATGTCTTTCATAT

CTTTATG-3′

Stitch 2 forward: 5′-GATAATACCCCAGCCGCACCAGTTAATATTCAGAATGC

GACAGTATTC-3′

Stitch 2 reverse: 5′-GAATACTGTCGCATTCTGAATATTAACTGGTGCGGCTGG

GGTATTATC-3′

Fusion 16

Forward: 5′-*CCATCA***CATATG**ATGGAAGGTCTTGATATAAACAAA-3′

Reverse: 5′-*CGC***GGATCC**TCATACTGTCTCACCTTCAATAACTCGCGCTTCCT

CGTCTTCAATTTCGTCCATTAATTCTGTTGG-3′

Fusion 17

Forward: 5′-*CCATCA***CATATG**ATGTCAGATCTAGGCTTAGAT-3′

Reverse: 5′-*CGC***GGATCC**TTAATCGGTTGTTCCATTTATCACCTTCTCACGAG

CTTCTTGAGCCTCGTAAGCATCACCGATTTCATCCATTAAATCACT-3′

Fusion 18

Forward: 5′-*CCATCA***CATATG**ATGTCAGATCTAGGCTTAGAT-3′

Reverse: 5′-*CGC***GGATCC**TCATACTGTCTCACCTTCAATAACTCGCGCTTCTT

GAGCCTCGTAAGCGTCTTCGATTTCATCCATTAA-3′

Fusion 19

Forward: 5′-*CCATCA***CATATG**ATGTCAGATCTAGGCTTAGAT-3′

Reverse: 5′-*CGC***GGATCC**TTAATCGGTTGTTCCATTTATCACCTTCTCTCGCGC

TTCCTCGTCTTCGAT-3′

Fusion 20

Forward: 5′-*GGAATTC***CATATG**CAACAACTTCAGGCATTG-3′

Reverse: 5′-*CGC***GGATCC**TTAATCGGTTGTTCCATTTATCACCTTCTCACGAG

CTTCTTGAGCCTCGTAAGCATCACCAACTTCGTTCAGAAGATC-3′

T4 L9-D164

Forward: 5′-*CCATCA***CATATG**CTTTTAGATATTTCTGACCTCCCC-3′

Reverse: 5′-*CGC***GGATCC**TTAATCGGTTGTTCCATTTATCAC-3′

T4 I12-D164

Forward: 5′-*CCATCA***CATATG**ATTTCTGACCTCCCCGGAATTGAC-3′

Reverse: 5′-*CGC***GGATCC**TTAATCGGTTGTTCCATTTATCAC-3′

T4 S36-D164

Forward: 5′-*CCATCA***CATATG**AGCAATCCACAAAACCGTACTCCA-3′

Reverse: 5′-*CGC***GGATCC**TTAATCGGTTGTTCCATTTATCAC-3′

RB49/T4 gp17

Forward: 5′-*CCATCA***CATATG**ATGAGCGTGATTGAAGGGATT-3′

Reverse: 5′-*CGC***GGATCC**TTATTTTGAAAATACTTCAGATGC-3′

Stitch forward: 5′-AAAGATAGTGACAATATTCGCACGCGGTATATGGGTCTT CCTAACTTG-3′

Stitch reverse: 5′-CAAGTTAGGAAGACCCATATACCGCGTGCGAATATTGTCA CTATCTTT-3′

Appendix 1. Primers used to construct T4 family terminase protein and mutant clones. The nucleotide numbers in parentheses correspond to the coding sequence of gp17 or gp16. Restriction enzyme digesting sites (shown in bold) were added to the 5′ ends of the forward and reverse primers to facilitate efficient insertion into the vectors. Italicized nucleotides are protecting sequence added to the 5′ ends for efficient enzyme cutting.

Appendix 2

gp16 mutant	T4 gp17 ATPase stimulation activity			RB49 gp17 ATPase stimulation activity		
	K_m [a]	K_{cat} [b]	% of WT [c]	K_m	K_{cat}	% of WT
T4 gp16	451	43	100	427	18	20.1
RB49 gp16	264	6	23.8	725	152	100
KVP40 gp16	434	1.6	3.9	713	3.8	2.5
T4 I128	334	20	62.8	565	6.5	5.5
Fusion 6	610	62	106.6	902	45	23.8
Fusion 9	ND [d]	ND	ND	370	8.3	10.7
Fusion 11	369	13	36.9	624	152	116.2
Fusion 13	308	21.5	73.2	572	20.2	16.8
Fusion 16	240	12.8	55.7	579	90	74.2
Fusion 17	1218	123	106	756	40	25.2
Fusion 18	238	5.9	26	334	60	86
Fusion 19	213	9.1	44.8	494	99	95.3
Fusion 20	124	2.6	22	249	10	19.2
T4 L9-D164	395	31	82.4			
T4 I12-D164	384	31	84.7			
T4 S36-D164	727	58	83.7			

a. K_m (μM)

b. K_{cat} (No. of ATP hydrolyzed/ min/ gp17 monomer)

c. Based on the K_{cat}/K_m values; WT (wild-type) refers to the ATPase activity stimulated by homologous gp16s

d. Not detectable

Appendix 2. Kinetic parameters of gp17-ATPase stimulation activity. ATPase assays were performed using purified gp17s (0.2-2 μM) and gp16s at a gp16: gp17 molar ratio of 10:1. The concentration of cold ATP was varied (0.05-2 mM) in the presence of trace amounts of [γ-^{32}P]ATP (75 nM). The reaction rates were determined by phosphorimaging, and the kinetic parameters were determined by using SigmaPlot 8.0 software. Values represent average of duplicates from two independent experiments. See 'Materials and Methods' for details.

References

1. Suttle, C.A. Viruses in the sea. *Nature* **437**, 356-61 (2005).

2. Wommack, K.E., Colwell, R. R. Virioplankton: viruses in aquatic ecosystems. *Microbiol Mol Biol Rev* **64**, 69-114 (2000).

3. Brussow, H., Canchaya, C., Hardt, W. D. Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. *Microbiol Mol Biol Rev* **68**, 560-602 (2004).

4. Hershey, A.D. & Chase, M. Independent functions of viral protein and nucleic acid in growth of bacteriophage. *J Gen Physiol* **36**, 39-56 (1952).

5. Karam, J.D., Drake, J.W., Kreuzer, K.N., Mosig, G., Hall, D.H., Eiserling, F.A., et al. Molecular biology of bacteriophage T4. Miller, E.S. (ed.). *American Society for Microbiology, Washington, D.C.* (1994).

6. Fauquet, C., Mayo, M., Maniloff, J., Desselbergre, U. & Ball, A., editors. Virus Taxonomy. Vol. VIII. *Amsterdam: Elsevier Academic Press* (2005).

7. Catalano, C., editor. Viral genome packaging machines: genetics, structure, and mechanism. *Landes Biosceince, Georgetown, TX* (2005).

8. Baschong, W. et al. Head structure of bacteriophages T2 and T4. *J Ultrastruct Mol Struct Res* **99**, 189-202 (1988).

9. Crowther, R.A., Lenk, E.V., Kikuchi, Y. & King, J. Molecular reorganization

in the hexagon to star transition of the baseplate of bacteriophage T4. *J Mol Biol* **116**, 489-523 (1977).

10. Cerritelli, M.E., Wall, J.S., Simon, M.N., Conway, J.F. & Steven, A.C. Stoichiometry and domainal organization of the long tail-fiber of bacteriophage T4: a hinged viral adhesin. *J Mol Biol* **260**, 767-80 (1996).

11. Leiman, P.G., Kanamaru, S., Mesyanzhinov, V.V., Arisaka, F. & Rossmann, M.G. Structure and morphogenesis of bacteriophage T4. *Cell Mol Life Sci* **60**, 2356-70 (2003).

12. Eiserling, F.A. & Black, L.W. Pathways in T4 morphogenesis. In: Molecular Biology of Bacteriophage T4. Karam, J.D., ed. *ASM Press, Washington, D.C.* (1994).

13. Matthews, C.K. An overview of the T4 developmental program. In: Molecular Biology of Bacteriophage T4. Karam, J.D., ed. *ASM Press, Washington, D.C.* (1994).

14. Mosig, G. & Eiserling, F.A. T4 and Related Phages: Structure and Development. In: The Bacteriophages, 2nd edition. Calender, R., ed. *Oxford University Press, London* (2004).

15. Laemmli, U.K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680-5 (1970).

16. Laemmli, U.K., Amos, L.A. & Klug, A. Correlation between structural transformation and cleavage of the major head protein of T4 bacteriophage. *Cell* **7**, 191-203 (1976).

17. Steven, A.C., Couture, E., Aebi, U. & Showe, M.K. Structure of T4 polyheads. II. A pathway of polyhead transformation as a model for T4 capsid maturation. *J Mol Biol* **106**, 187-221 (1976).

18. Showe, M.K., Isobe, E. & Onorato, L. Bacteriophage T4 prehead proteinase. I. Purification and properties of a bacteriophage enzyme which cleaves the capsid precursor proteins. *J Mol Biol* **107**, 35-54 (1976).

19. Showe, M.K., Isobe, E. & Onorato, L. Bacteriophage T4 prehead proteinase. II. Its cleavage from the product of gene 21 and regulation in phage-infected cells. *J Mol Biol* **107**, 55-69 (1976).

20. Black, L.W. DNA packaging in dsDNA bacteriophages. *Annu Rev Microbiol* **43**, 267-92 (1989).

21. Jardine, P.J. & Coombs, D.H. Capsid expansion follows the initiation of DNA packaging in bacteriophage T4. *J Mol Biol* **284**, 661-72 (1998).

22. Jardine, P.J., McCormick, M.C., Lutze-Wallace, C. & Coombs, D.H. The bacteriophage T4 DNA packaging apparatus targets the unexpanded prohead. *J Mol Biol* **284**, 647-59 (1998).

23. Steven, A.C., Greenstone, H.L., Booy, F.P., Black, L.W. & Ross, P.D. Conformational changes of a viral capsid protein. Thermodynamic rationale for proteolytic regulation of bacteriophage T4 capsid expansion, co-operativity, and super-stabilization by soc binding. *J Mol Biol* **228**, 870-84 (1992).

24. Fokine, A. et al. Molecular architecture of the prolate head of bacteriophage T4. *Proc Natl Acad Sci U S A* **101**, 6003-8 (2004).

25. Fuller, D.N., Raymer, D.M., Kottadiel, V.I., Rao, V.B. & Smith, D.E. Single phage T4 DNA packaging motors exhibit large force generation, high velocity, and dynamic variability. *Proc Natl Acad Sci U S A* **104**, 16868-73 (2007).

26. Rao, V.B. & Black, L.W. DNA Packaging in Bacteriophage T4. In Viral Genome Packaging Machines: Genetics, Structure, and Mechanism. Catalano, C.E., ed. *Landes Biosciences, Georgetown, TX*, pp. 40-58 (2005).

27. Rao, V.B. & Black, L.W. 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* **200**, 475-88 (1988).

28. Lin, H., Simon, M.N. & Black, L.W. Purification and characterization of the

small subunit of phage T4 terminase, gp16, required for DNA packaging. *J Biol Chem* **272**, 3495-501 (1997).

29. Sun, S. et al. The structure of the phage T4 DNA packaging motor suggests a mechanism dependent on electrostatic forces. *Cell* **135**, 1251-62 (2008).

30. Kuebler, D. & Rao, V.B. Functional analysis of the DNA-packaging/terminase protein gp17 from bacteriophage T4. *J Mol Biol* **281**, 803-14 (1998).

31. Alam, T.I. et al. The headful packaging nuclease of bacteriophage T4. *Mol Microbiol* **69**, 1180-90 (2008).

32. Massey, T.H., Mercogliano, C.P., Yates, J., Sherratt, D.J. & Lowe, J. Double-stranded DNA translocation: structure and mechanism of hexameric FtsK. *Mol Cell* **23**, 457-69 (2006).

33. Iyer, L.M., Makarova, K.S., Koonin, E.V. & Aravind, L. Comparative genomics of the FtsK-HerA superfamily of pumping ATPases: implications for the origins of chromosome segregation, cell division and viral capsid packaging. *Nucleic Acids Res* **32**, 5260-79 (2004).

34. Leffers, G. & Rao, V.B. Biochemical characterization of an ATPase activity associated with the large packaging subunit gp17 from bacteriophage T4. *J Biol Chem* **275**, 37127-36 (2000).

35. Kondabagil, K.R., Zhang, Z. & Rao, V.B. The DNA translocating ATPase of bacteriophage T4 packaging motor. *J Mol Biol* **363**, 786-99 (2006).

36. Mitchell, M.S. & Rao, V.B. Functional analysis of the bacteriophage T4 DNA-packaging ATPase motor. *J Biol Chem* **281**, 518-27 (2006).

37. Rao, V.B. & Mitchell, M.S. The N-terminal ATPase site in the large terminase protein gp17 is critically required for DNA packaging in bacteriophage T4. *J Mol Biol* **314**, 401-11 (2001).

38. Al-Zahrani, A.S. et al. The small terminase, gp16, of bacteriophage T4 is a regulator of the DNA packaging motor. *J Biol Chem* **284**, 24490-500 (2009).

39. Rentas, F.J. & Rao, V.B. 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* **334**, 37-52 (2003).

40. Sun, S., Kondabagil, K., Gentz, P.M., Rossmann, M.G. & Rao, V.B. The structure of the ATPase that powers DNA packaging into bacteriophage T4 procapsids. *Mol Cell* **25**, 943-9 (2007).

41. Rao, V.B. & Feiss, M. The bacteriophage DNA packaging motor. *Annu Rev Genet* **42**, 647-81 (2008).

42. Goetzinger, K.R. & Rao, V.B. 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* **331**, 139-54 (2003).

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

44. Kanamaru, S., Kondabagil, K., Rossmann, M.G. & Rao, V.B. The functional domains of bacteriophage t4 terminase. *J Biol Chem* **279**, 40795-801 (2004).

45. Kondabagil, K.R. & Rao, V.B. 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* **358**, 67-82 (2006).

46. Mitchell, M.S., Matsuzaki, S., Imai, S. & Rao, V.B. 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* **30**, 4009-21 (2002).

47. Chai, S., Lurz, R. & Alonso, J.C. The small subunit of the terminase enzyme of Bacillus subtilis bacteriophage SPP1 forms a specialized nucleoprotein complex with the packaging initiation region. *J Mol Biol* **252**, 386-98 (1995).

48. Hwang, Y., Catalano, C.E. & Feiss, M. Kinetic and mutational dissection of the two ATPase activities of terminase, the DNA packaging enzyme of

bacteriophage Chi. *Biochemistry* **35**, 2796-803 (1996).

49. Catalano, C.E., Cue, D. & Feiss, M. Virus DNA packaging: the strategy used by phage lambda. *Mol Microbiol* **16**, 1075-86 (1995).

50. Rao, V.B. & Black, L.W. DNA packaging of bacteriophage T4 proheads in vitro. Evidence that prohead expansion is not coupled to DNA packaging. *J Mol Biol* **185**, 565-78 (1985).

51. Pierce, J.C. & Sternberg, N.L. Using bacteriophage P1 system to clone high molecular weight genomic DNA. *Methods Enzymol* **216**, 549-74 (1992).

52. Horton, R.M., Hunt, H.D., Ho, S.N., Pullen, J.K. & Pease, L.R. Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension. *Gene* **77**, 61-8 (1989).

53. Bhattacharyya, S.P. & Rao, V.B. A novel terminase activity associated with the DNA packaging protein gp17 of bacteriophage T4. *Virology* **196**, 34-44 (1993).

54. Black, L.W. & Peng, G. Mechanistic coupling of bacteriophage T4 DNA packaging to components of the replication-dependent late transcription machinery. *J Biol Chem* **281**, 25635-43 (2006).

55. Earnshaw, W.C. & Casjens, S.R. DNA packaging by the double-stranded DNA bacteriophages. *Cell* **21**, 319-31 (1980).

56. Luftig, R.B. & Ganz, C. Bacteriophage T4 head morphogenesis. IV. Comparison of gene 16-, 17-, and 49-defective head structures. *J Virol* **10**, 545-54 (1972).

57. Nemecek, D. et al. Subunit conformations and assembly states of a DNA-translocating motor: the terminase of bacteriophage P22. *J Mol Biol* **374**, 817-36 (2007).

58. White, J.H. & Richardson, C.C. Gene 18 protein of bacteriophage T7. Overproduction, purification, and characterization. *J Biol Chem* **262**, 8845-50 (1987).

59. Smits, C. et al. Structural basis for the nuclease activity of a bacteriophage large terminase. *EMBO Rep* **10**, 592-8 (2009).

60. Golz, S. & Kemper, B. Association of holliday-structure resolving endonuclease VII with gp20 from the packaging machine of phage T4. *J Mol Biol* **285**, 1131-44 (1999).

61. Burroughs, A.M., Iyer, L.M. & Aravind, L. Comparative genomics and evolutionary trajectories of viral ATP dependent DNA-packaging systems. *Genome Dyn* **3**, 48-65 (2007).

62. Casjens, S.R. Comparative genomics and evolution of the tailed-bacteriophages. *Curr Opin Microbiol* **8**, 451-8 (2005).

63. Williams, R.S., Williams, G.J. & Tainer, J.A. A charged performance by gp17 in viral packaging. *Cell* **135**, 1169-71 (2008).

64. DeLano, W.L. The PyMOL molecular graphics system (http://www.pymol.org). (2002).