THE CATHOLIC UNIVERSITY OF AMERICA

Assembly of Bacteriophage T4 DNA Packaging Motor: Analysis of Portal-Terminase Interactions

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

Shylaja M. Hegde

Washington, D.C.

2010

Assembly of Bacteriophage T4 DNA Packaging Motor: Analysis of Portal-Terminase Interactions

Shylaja M. Hegde, Ph.D.

Director: Venigalla B. Rao, Ph.D.

The icosahedral double-stranded DNA bacteriophages and herpes viruses package their genomes into preformed proheads by a powerful ATP driven motor. The packaging motor, an oligomer of gp17 (large terminase) in phage T4, is assembled at the special portal vertex of the empty prohead. The T4 motor is the fastest and most powerful motor reported to date. gp17 has an N-terminal ATPase that powers DNA translocation and a C-terminal translocase that causes DNA movement. The dynamic interactions between the motor (gp17) and the portal (gp20) are however poorly understood. Here, using biochemistry, bioinformatics, structure, and molecular genetics, the site in gp17 that interacts with the dodecameric portal protein is determined.

Biochemical and structural studies suggest that the N-terminal domain of gp17 interacts with gp20, and that the stoichiometry of prohead-gp17 complex is five subunits of gp17 to twelve subunits of gp20. Sequence alignments predict that there are two potential portal binding sites in gp17. Mutational studies show that the portal binding site I in the N-terminal domain is critical for function whereas the site II in the C-terminal domain is not critical. Second site suppressors of site I D331Q mutant (temperature sensitive) show a single intragenic mutation in the helix-loop-helix (HLH) of N-terminal sub-domain II, suggesting the importance of this motif in portal interaction. Fitting the X-ray structure of gp17 into the cryo-EM density of portal-motor complex showed the same HLH (amino acids 333-352) in contact with gp20. A peptide corresponding to the HLH motif specifically binds to proheads as well as

inhibits DNA packaging in vitro. Swapping of non-conserved residues of the helix, but not the conserved residues of the loop, from T4-family phages relieves the DNA packaging inhibition. Together these data for the first time identify a HLH motif in gp17 that interacts with gp20, leading to models for symmetry mismatch between the packaging motor and the portal as well as implications to the mechanism of viral DNA translocation.

This dissertation by Shylaja M. Hegde fulfills the dissertation requirement for the doctoral degree in biology approved by Venigalla B. Rao, Ph.D., as Director, and by Johan E. Golin, Ph.D., and Michael Mullins, Ph.D., as Readers.

_____

Venigalla B.Rao, Ph.D., Director

_____

Johan E. Golin, Ph.D, Reader

_____

Michael Mullins, Ph.D, Reader

TABLE OF CONTENTS

	Page
List of Figures	iv
List of Tables	vi
Acknowledgements	vii
Introduction	1
Materials and Methods	20
Experimental Results	47
Discussion	82
References	94

LIST OF FIGURES

Figure	Page
1. The schematic structure of T4 bacteriophage	5
2. Overview of the developmental cycle of T4 bacteriophage	6
3. T4 DNA packaging pathway	8
4. Schematic of the T4 bacteriophage DNA packaging machine	11
5. The functional domain of gp17 protein	12
6. A ribbon diagram of the T4 gp17 crystal structure	13
7. Ribbon diagrams of crystal structure of phi-29 portal protein	14
8. Model for DNA translocation by bacteriophage	16
9. Map of pET 15b vector	22
10. Map of pET 28 vector	23
11. Splicing overlap extension	25
12. Schematic of prohead binding assay	48
13. Specific binding of gp17 to proheads	49
14. Construction and functional characterization of gfp-full length gp17	50
15. Binding of gfp-gp17 to the proheads	52
16. gfp-gp17 specifically interacts with gp20 on the prohead	54
17. Prohead binding assay with N360 and C360 domain of gp17	56
18. Stochiometry of terminase in DNA packaging complex	58
19. Cryo-EM reconstruction of gp17 bound to the prohead	59
20. Cryo-EM reconstructions prohead assembled packaging motor	60

21. Predicted gp20 interacting region on gp20 61

22. Sequence alignment of eight T4 family phage terminases 62

23. Mutagenesis of portal binding site I and II residues 64

24. X-ray of structure of gp17 with portal binding sites and HLH highlighted 68

25. Construction and purification of helix-loop-helix peptide 71

26. Schematic of *in vitro* DNA packaging inhibition assay 72

27. Inhibition of *in vitro* DNA packaging by helix-loop-helix peptide 73

28. Effect of HLH peptides on gp16 stimulated ATPase activity of gp17 74

29. Specific binding of gfp-wt peptides to the prohead 75

30. T4 family phage terminases were aligned with CLUSTAL W 78

31. Construction and purification of HLH swap peptide mutants 79

32. Effects of swap peptide mutants on DNA packaging activities 80

33. Characterization of gfp-gp17 GSS-AAA loop mutant 81

34. Symmetry mismatch model for DNA packaging 92

LIST OF TABLES

Table		Page
1.	Characterization of portal binding site I functional clones	66
2.	Second site mutation and phenotype of (D331Q) *ts* suppressor phage	67

ACKNOWLEDGEMENTS

First of all, I would like to convey my gratitude to my research advisor Dr. Venigalla B. Rao, for his advice, guidance and support throughout my work. His critical comments and encouragement enabled me to develop thorough understanding of the subject. I am proud to be a part of his lab and the project. I also would like to thank my committee for taking the time to review my work, offer suggestions, and serve at my defense.

It is my pleasure to thank all members of my lab who offered me their help during my work. Many thanks go in particular to Dr. Bonnie Drapper and Zhihong Zhang.

I owe my deepest gratitude to my Family and Friends for their inseparable support and encouragement all these years. A special thanks to my mom and sisters for their constant support, love and care; this dissertation is simply impossible without them.

Above all I thank god for all his blessings in my life.

Introduction

History of Bacteriophages

Bacteriophages are viruses that infect bacteria and kill them by intracellular replication and cell lysis. According to most estimates, bacteriophages first inhabited earth about 3-5 billion years ago, and since that time they have controlled the levels of bacteria in the environment through a classical "predator-prey" relationship. However, the existence of bacteriophages was acknowledged only a century ago. In 1896, Earnest Hankin reported that something which could pass through a very fine porcelain filter, in the waters of the Ganges and Jumna rivers in India, had antibacterial action. Twenty years later a British bacteriologist, Frederick Twort, actually isolated these filterable microbes, which are capable of infecting a bacterium, inside which they can multiply to produce progeny viruses and ultimately lyse the cell to release the daughter virus particles. However, Twort did not further explore his findings. Two years later, Felix d'Herelle, a French Canadian microbiologist from the Pasteur Institute in Paris, reported the same phenomenon. He then named these sub-microscopic microbes "bacetriophages" or "bacteria-eaters" (In greek *phago* meaning *to eat* or *devour*) (Dr. Felix d'Herelle, Science News, 1949).

Immediately following the discovery of bacteriophage, part of the scientific community believed it would help doctors to successfully treat bacterial infections. Richard Bruynoghes and Joesph Maisin (1921), from France, were the first to use phages for treating human skin disease caused by Staphylococcus bacteria. Later on, phages were also used for treating skin infections caused by Klebsiella, Proteus, *E.coli* and for

Staphylococcal lung and pleural infections. However the development and wide spread use of antibiotics made phage research less interesting and was later completely abandoned by the biomedical community. Nonetheless, slowly bacteriophages became the most used model organisms for basic research to understand many concepts of molecular biology. The most famous and important one is the demonstration of "DNA as a hereditary material." It was also used to understand the central dogma of molecular biology of DNA-RNA-protein, discovery of RNA, genetic code etc.

In 1952, Alfred Hershey and Martha Chase used phage T2 to demonstrate that the DNA is the material that is passed on to progeny, not the protein. Two sets of experiments were done. In one set phages were grown in the presence of radioactively labeled phosphorous which was incorporated into the DNA phosphate backbone. In the second set, phages were grown in the presence of radioactively labeled sulfur which is incorporated into the cysteine and methoinine residues of proteins. These labeled phages were then used to infect the *E.coli* cells. After the infection, only radioactively labeled DNA entered into the host *E.coli* cells, which was then passed on to the virus progeny (Hershey and Chase, 1952).

T4 as a model organism

After Hershey and Chase's famous experiment, the T even bacteriophages continue to be an ideal model system to elucidate the basic principle of molecular biology. By using T4 bacteriophage, Francis Crick elucidated the degenerate triplet codon system (Crick *et al*., 1961) and Sydney Brenner proved the existence of mRNA in the cell (Brenner *et al*., 1961). Since then T4 bacteriophage has become the phage of choice to

study the mechanisms involved in the assembly and morphogenesis of complex biological structures (Kellenberger *et al*., 1990).

The entire genome of T4 has been sequenced. The genome encodes 289 open reading frames, of which 150 genes have already been characterized. T4 has a short life cycle of about 25-30 minutes and labeled amino acids and deoxynucleotides can be incorporated with high efficiency to follow assembly pathways. The developmental cycle of T4 is well understood. A large collection of conditional lethal mutants are available. These well established features of T4 make it a compelling system to elucidate the mechanisms of fundamental biological processes.

Life cycle of T4 bacteriophage

T4 bacteriophage is a linear dsDNA bacteriophage. It consists of 120 nm long and 80 nm wide icosahedral head and a 10 nm long and 2 nm wide tail. The head encapsulates the dsDNA genome. The contractile tail has a base plate and attached to it are long and short tail fibers (Figure 1; Eiserling, 1992). It infects *E.coli* by binding of the tips of tail fibers to the lippopolysaccharide receptor on the surface of the bacterial cell. This causes conformational changes in the base plate and contraction of tail sheath, which in turn causes the tail lysozyme (gp5) to puncture the outer membrane of bacteria. The lysozyme domain of gp5 then degrades the periplasmic peptidyoglycan and inner membrane. The genome is then ejected from the head into the host cell (Matthews *et al*., 1994).

As soon as the DNA gains entry into the host cell, the host RNA polymerase transcribes the early genes required for DNA replication and inhibition of the host cell

metabolism. The phage enzymes take over the host cell machinery for replication of its own DNA. The replication process produces concatemeric DNA that is a head to tail polymer of unit length genome as well as highly branched. The late genes which code for structural proteins are then transcribed. Structural proteins are assembled into head, tail, and tail fibers by independent pathways. Once the DNA is packaged into the capsid, the tail and tail fibers attach to the capsid. The entire life cycle takes approximately 25 minutes and mature phages are released into the medium by lysis of the host cell (Figure 2; Mathews *et al*., 1994).

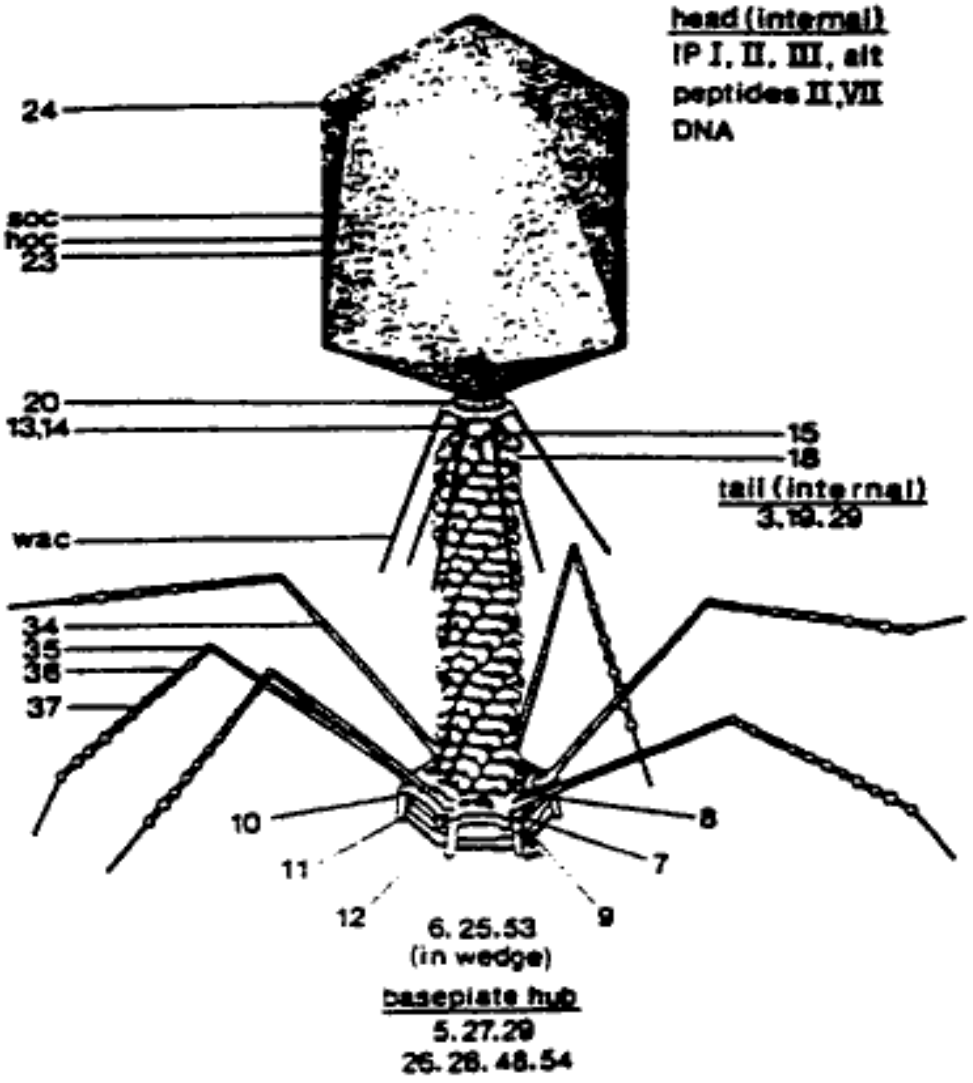


Figure 1: The schematic structure of T4 bacteriophage (Eiserling, 1992)

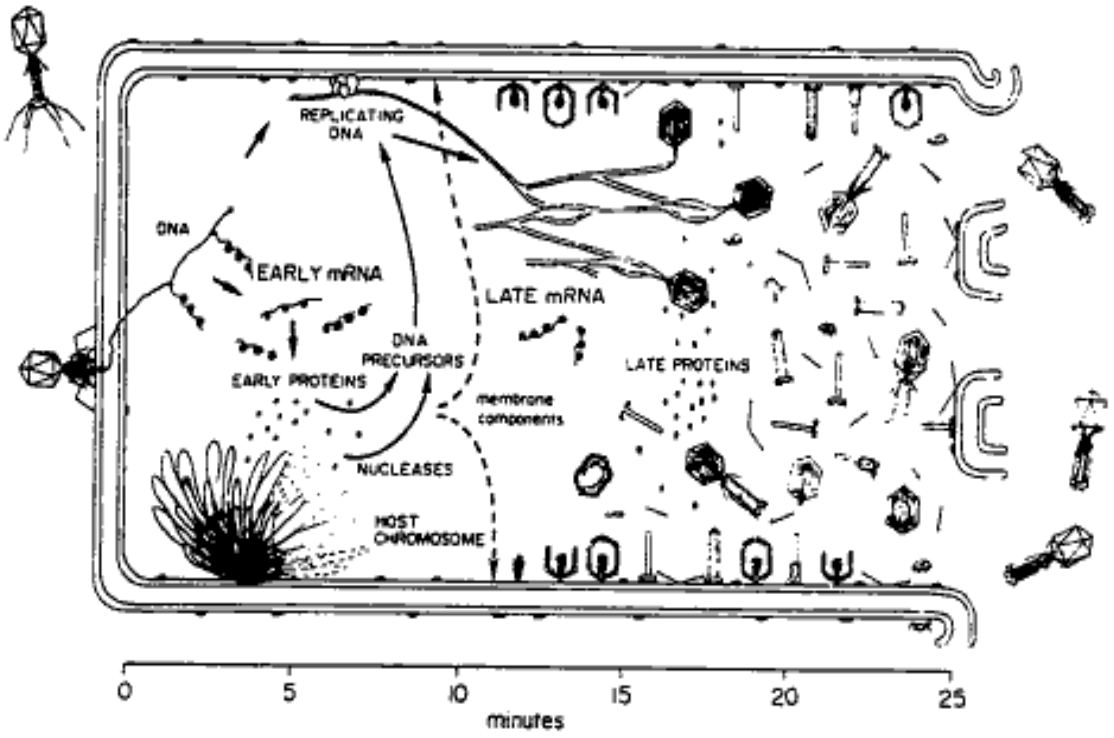


Figure 2: Overview of the developmental cycle of T4 bacteriophage (Mathews et al., 1994).

DNA Packaging

DNA packaging is a fundamental biological process, in T4, herpes viruses, and other icoshedral phages, it involves: i) recognition of concatemeric viral DNA, ii) cutting of DNA to produce an end, iii) translocation of DNA into the preformed capsid, iv) condensation into a crystalline density inside the capsid, and v) cutting of DNA to terminate DNA packaging. The process of DNA packaging by dsDNA viruses share the mechanistic similarities with other important biological process such as genome segregation in prokaryotes (Massey *et al*., 2006) and chromosome condensation in eukaryotes (Iyer *et al*., 2004). Therefore the simpler bacteriophages are good model systems to study this complex biological problem.

Packaging initiation

In phages like lambda, DNA packaging is initiated when DNA is cut at the specific site called "*cos*". In some other phages like P1, P22 *or* T1, Spp1, the initial cut is made in the vicinity of a specific site called "*Pac*". The termination cut in lambda is made at the next *cos* site, but it is nonspecifically made after one "headful" genome length is packaged in case of P1, Spp1 and P22 (Feiss *et al*., 2001; Casjens *et al*., 2005). In the case of T4, both the initial and final cuts are not made at a specific sequence. T4 packages DNA strictly by a headful mechanism (Stringer *et al*., 1976). The heedful signal from the packaged capsid triggers the termination cut. Thus there is a strict linkage between capsid size and genome length.

In T4, DNA packaging is initiated by binding of a small terminase subunit gp16 to the newly replicated concatemeric DNA. The large terminase protein gp17 makes the

first cut to generate the free end. The terminase-DNA complex then docks the free end of DNA to the empty capsid by interacting with the portal protein gp20, which is situated at the special five fold vertex of the capsid (Figure 3; Black *et al*., 1994). Using ATP hydrolysis as the driving force, the terminase protein translocates DNA through the portal channel into the capsid (Kondabagil and Rao 2006). Once a headful volume of DNA is translocated, gp17 makes the second cut to terminate packaging and dissociate from the packaged prohead. It then docks the DNA to another empty capsid and repeats the process until all the genomes are packaged inside the capsid (Baumann and Black, 2003; Bhattacharya and Rao, 1993).

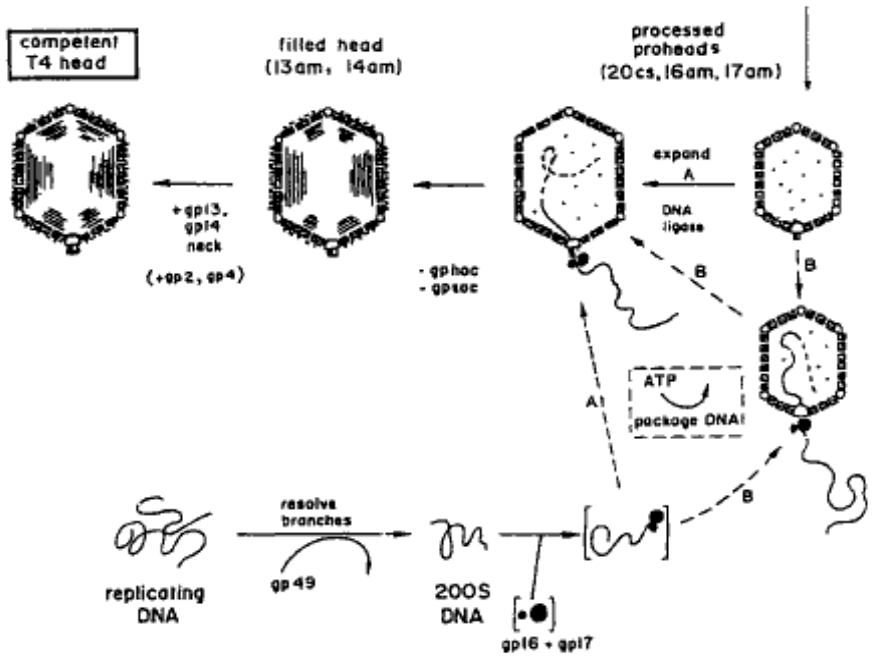


Figure 3: T4 DNA packaging pathway showing the processing and translocation of DNA into an empty capsid by the terminase complex formed by gp16 and gp17 (Black *et al*., 1994).

The bacteriophage T4 packaging machine

The minimal packaging motor of T4 consists of the terminase protein gp17 and the portal protein gp20 (Figure 4; Sun *et al*., 2008).

Large terminase: The large terminase protein gp17 is a 70 kDa protein. It has an N terminal ATPase domain (N360), which fuels the DNA translocation and a C terminal nuclease domain (C360), which cleaves the DNA before and after packaging (Kanamaru *et al*., 2004). The N-domain (N360) has the classic Walker A (Mitchel *et al*., 2002), Walker B (Mitchel and Rao 2006), catalytic carboxylate (Goetzinger and Rao 2003), and ATPase coupling motifs (Draper and Rao 2007). The C-domain (C360) has a triad of acidic residues which coordinate with Mg to form a catalytic center for DNA cleavage (Rentas and Rao, 2003; Sun *et al*., 2008). The N-domain is sufficient for the ATPase activity whereas the C-domain is sufficient for the nuclease activity. However, when the protein is separated into two domains it loses DNA packaging activity despite retaining ATPase and nuclease functions (Figure 5). Thus, communication between the ATPase and the nuclease domains is required for successful translocation of DNA into the capsid (Kanamaru *et al*., 2004).

Structure: The crystal structure of gp17 has been solved recently. T4 bacteriophage is the only system in which the structure of the terminase protein is known. The ATPase domain is a flat structure, consisting of six parallel beta sheets, a classic Rossmann nucleotide binding fold (NBD), a structure similar to other ATPases such as RecA (Sun *et al*., 2007). On the other hand, the nuclease domain is a globular structure consisting of anti-parallel beta strands. The structure resembles the RNase H fold found in RNase H

and Ruv C resolvase (Sun *et al*., 2008). The N-terminal domain is further divided into larger ATP binding subdomain I and smaller subdomain II. The subdomain II links the ATP binding domain (subdomain I) to the nuclease domain, connecting the ATP hydrolysis to DNA binding and translocation (Figure 6; Sun *et al*., 2008).

Portal protein: The portal protein is a 61 kDa dodecameric protein which assembles to form a channel at a special vertex of the capsid (Figure 7). The channel is cone shaped, with the broader end interacting with capsid and the narrower end protruding out from the capsid. The overall structure and symmetry of the portal protein are strictly conserved among the bacteriophages and herpes viruses (Simpson *et al*., 2000). It is critical for assembly as well as expansion and stabilization of the capsid (Hsiao and Black 1978). It is proposed to provide the docking site for terminase during DNA translocation (Hsiao and Black, 1977; Black, 1989). It also acts as a headful packaging gauge by controlling the length of the DNA packaged, (Casjens *et al*., 1992; Tavares *et al*., 1992) and a connector between the head and the tail (Driedonks and Caldentey, 1983; Rishovd *et al*., 1998).

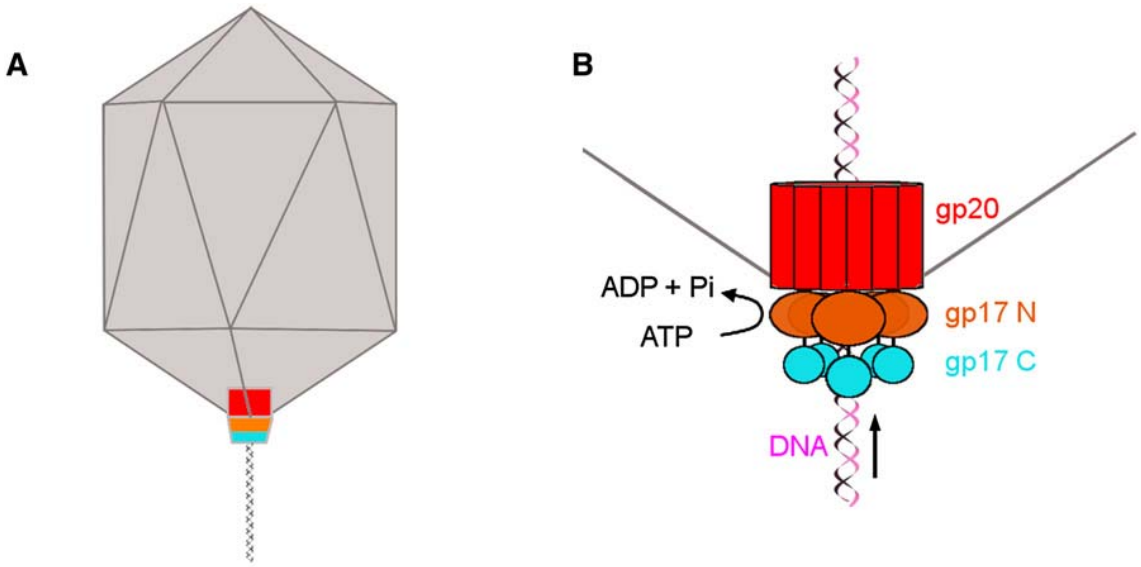


Figure 4: Schematic of the T4 bacteriophage DNA packaging machine. (Sun *et al*., 2008) A: Packaging motor assembles on the prohead through interaction with the gp20 and translocates the DNA into the capsid. B: Closer view of interaction of gp17 motor with gp20. gp20: red cylinders, gp17 N domain: orange big spheres, and gp17 C domain: blue small spheres.

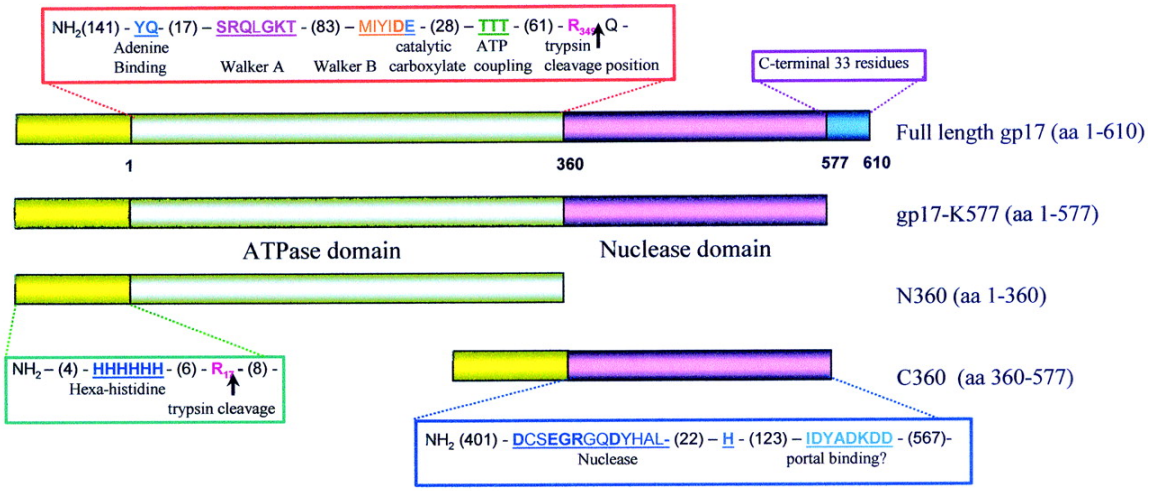


Figure 5: The functional domain of gp17 protein. The full length gp17 (610 aa) has two domains N terminal ATPase domain (1-360 aa, indicated as a white box) and C terminal nuclease domain (360-610, indicated as purple box). The well conserved motifs of ATPase are shown in the box above it. The gp17 K577 construct is same as full-length protein, with the last 33 amino acids (blue box) truncated. gp17-K577 retains all the functions of full length protein. The last 33 amino acids are susceptible to protease degradation during the protein purification process. All the constructs are tagged with hexa-histidine tags (yellow box) for Ni-agarose chromatography purification (Kanamaru *et al*., 2004).

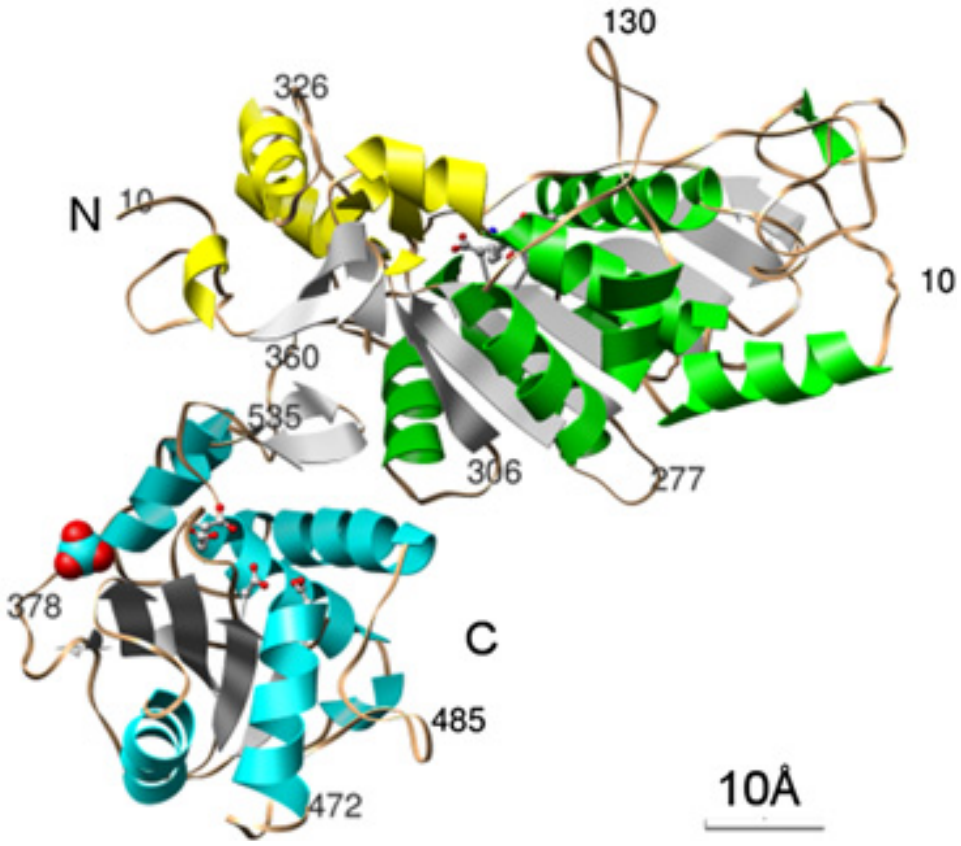


Figure 6: A ribbon diagram of the T4 gp17 crystal structure. The N-terminal subdomain I helices are colored green, N-terminal subdomain II helices are in yellow, C-terminal domain helices are in Cyan (Sun *et al*., 2008)

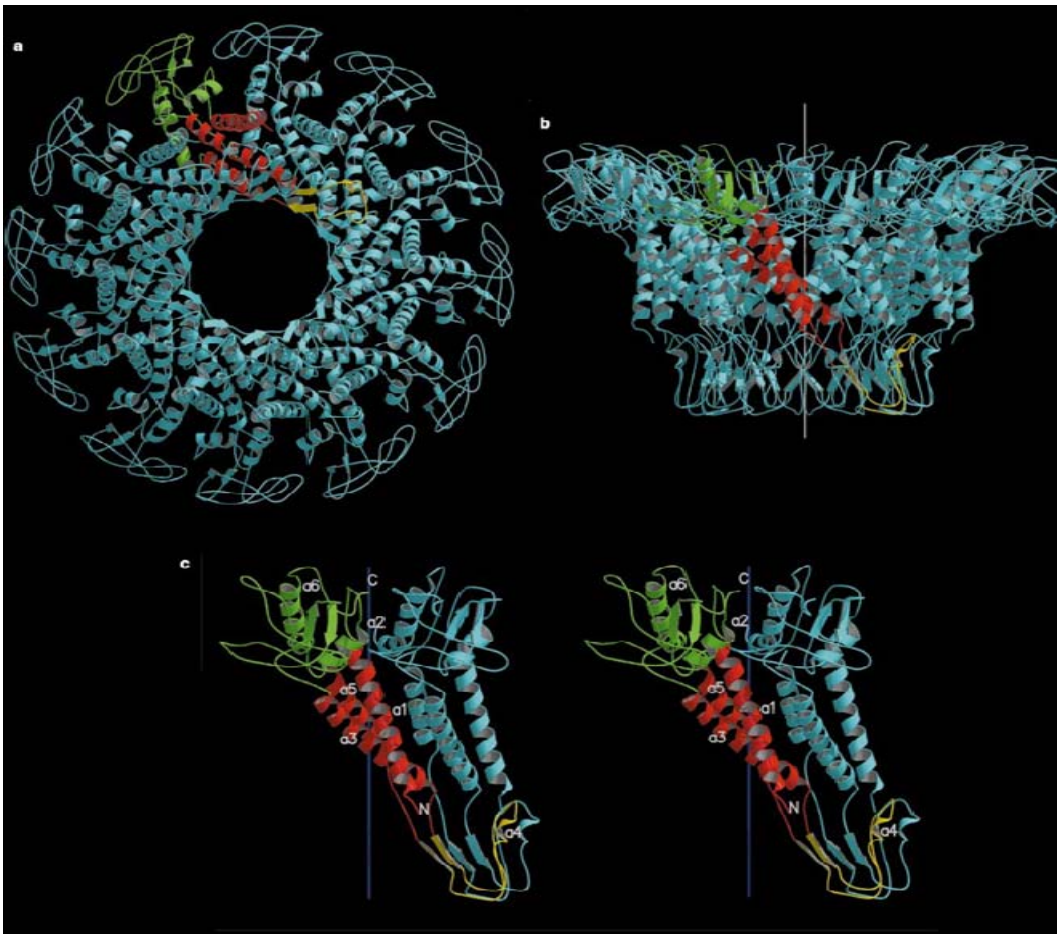


Figure 7: Ribbon diagrams of crystal structure of phi-29 portal protein. A: Top view of the portal channel. B: Side view of the channel. C: Diagram of two monomers. Single subunit is colored green in the wide end domain that resides in the capsid, red in the central domain, and yellow in the narrow end (Rossmann *et al*., 2000).

DNA packaging models

Several packaging models have been proposed to define the basic mechanism by which the DNA is translocated into the capsid. One of the early and widely accepted models was the symmetry mismatch model. According to this model, the symmetry mismatch between the portal (12- fold) and the capsid (5- fold) results in ATP hydrolysis by the terminase protein that causes rotation of the portal, which is coupled to the translocation of DNA into the capsid (Hendrix *et al*., 1978). A number of related models considering the portal as the rotary pump have also been proposed. Experimentally these models have not been proven in any phage system, and recent evidence on T4 and phi29 phage motors argue against the portal rotation model (Black *et al*., 2006; Hugel *et al*., 2007). A different model in which the terminase acts as a linear motor has also been proposed. According to this model, the terminase protein is the translocating pump, whereas the portal protein is a passive conduit, providing a channel for passage of DNA during translocation (Rao and Mitchell, 2001; Black *et al*., 1978; Morita *et al*., 1994). In depth experimental analysis of terminase protein supports its direct role in DNA translocation (Kubler and Rao 1998; Lin and Black 1998; Leffers and Rao, 2000; Mitchell and Rao, 2006; Draper and Rao, 2007). More recently based on biochemical, mutational and structural studies an electrostatic dependent DNA translocation model has been proposed for T4 bacteriophage. According to this model, an electrostatic force between two domains of gp17, generated by alternating between tensed and relaxed conformational states drives DNA translocation (Figure 8).

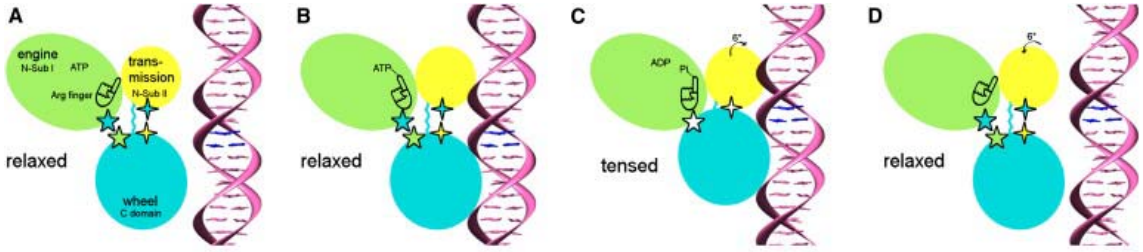


Figure 8: Model for DNA translocation by bacteriophage T4 packaging motor. gp17 C-terminal domain, N-terminal subdomain I and subdomain II are represented as cyan, green and yellow ovals respectively. The charged interactions between subdomain I subdomain II and C-domain are represented by stars. The flexible linker between N- and C-domain is shown in zigzag cyan line A: In relaxed state the flexible linker between N and C terminal is in extended form. B: DNA binds to the C domain and ATP binds to the N domain. The conformational changes in N domain cause insertion of arginine finger into the ATPase active centre and trigger ATP hydrolysis. C: The repulsion between negatively charged ADP^{3-} and Pi^{3-} causes 6^{o} rotation of subdomain II. The rotation of subdomain II aligns charged pairs between N and C domains. The electrostatic force between opposite charges pulls the C domain-DNA complex towards the N domain, translocating DNA into the capsid. D: ADP^{3-} and Pi^{3-} are released and subdomain II rotates back; the C-domain returns to original relaxed position. DNA is released and aligned to bind the C-domain of the adjacent gp17 subunit and continue the next translocation cycle (Sun *et al*., 2008).

Purpose

It is now well established that gp17 is the packaging motor of T4 bacteriophage. The assembly of this motor onto the portal of the prohead is the initial and key, step for DNA translocation. Although extensive work has been done to characterize the gp17 molecular motor, not much is known about the nature of interaction and communication between gp17and gp20 during DNA translocation. In fact, direct evidence for gp17-gp20 interaction has not been established in any phage system. Also, very little is known about the stoichiometry of gp17 in the packaging complex.

The main purpose of this study is to rigorously analyze the gp17-gp20 interaction by using molecular genetic, biochemical, bioinformatics and structural approaches. This study will aim to provide direct evidence for gp17-gp20 interaction, determine the stoichiometry of the gp17 packaging motor and map the gp20 interacting region on gp17. The results would provide insight on the dynamic interactions between the portal and the packaging motor that are central to the DNA packaging mechanism.

Overview of experimental approaches

Biochemical, genetic, bioinformatic, and structural approaches were used to analyze the interaction between the terminase and portal proteins. A direct binding assay, using purified prohead and terminase, was developed to show a specific interaction between gp17 and gp20 on the prohead.

The importance of two predicted portal binding site I and II was tested by bioinformatics and genetic studies. Multiple sequence alignments of T4 family phages were used to select the most conserved residues (D330, D331) from site I and (D560,

Y561) from site II for combinatorial mutagenesis. Second site suppressor were isolated for D331Q temperature sensitive mutant and mapped to the near by sequence (I337M and I337K).

The solved structures of T4 gp17 and Spp1 portal were used to model the gp20 interacting region on gp17. These analyses predicted that a helix-loop helix region from residues 333-354 in gp17 contacts the gp20 loop that protrudes out of the capsid. A helix-loop-helix peptide fused to the soc molecule was used to test whether or not peptide binds to gp20 and inhibit DNA packaging. The interacting regions are mapped to the helix residues of the helix-loop-helix motif.

Contribution and originality

DNA packaging in bacteriophages is an important model to understand the fundamental mechanisms of DNA translocation, transduction of biochemical energy into mechanical work, and chromosome condensation and de-condensation. Central to DNA packaging is the dynamic of protein-protein interactions between the portal and the packaging motor. This is the first study which focused to elucidate these interactions using combinations of precise approaches. The information generated on the sites of interactions, stoichiometry, and ATPase motor function will shed light on the inner workings of the packaging mechanism.

Many double stranded bacteriophages and pathogenic viruses like herpes viruses share similar assembly pathways. So the findings of this study will have broad implications to the understanding of DNA translocation in viruses, a key step in the virus

life cycle. The results would offer potential avenues to block DNA packaging by small compound therapeutics.

Material and methods

Bacterial strain

E.coli XL-10 gold cells (Stratagene, La Jolla, CA) were used for transformation of recombinant clones. Fusion genes of phage T4 sequence were cloned under the control of T7 promoter in the pET-28b vector. XL-10 gold cells lack T7 RNA polymerase gene required for the expression of recombinant gene, which can be toxic to the host strain. Therefore, XL-10 gold cells were used for stable maintenance of fusion gene.

For the expression of gene products, plasmids were transformed into the expression strain of *E.coli*, BL-21(DE3) pLys-S (Novagen) or BL-21 codon plus (DE3)-RPIL (Stratagene). The expression strain of *E.coli* has T7 RNA polymerase gene under the control of Lac operon: over expression of protein was achieved by addition of IPTG.

E.coli P301 (sup-) and E.coli B40 (sup+) were used to prepare wt and 17am18amrII phages. *E.coli* P301 (sup-) was also used in preparing empty proheads.

Bacteriophage strain

17am18amrII constructed earlier in our lab was used to prepare phage stocks and empty proheads. gp17 F329am and gp17 F558am were constructed in this study and were used in the marker rescue assays of gp17 DD330-331 and gp17 DY560-561 libraries respectively. Mutants gp17 K166am and H436am that were constructed earlier were used to construct F329am and F558am phages respectively.

Bacterial cultures were grown either in Luria-Bertani (LB) medium or Moore medium (2 % bacto-tryptone, 0.2% Na2HP04, 0.1% KH2PO4, 0.8% NaCl, 15% yeast

extract, and 0.2% dextrose) containing one or combinations of antibiotics, ampicilin (50 ug/ml), kanamycin (36 ug/ml) chloromophenical (36 ug/ml). For preparing T4 phage stocks and proheads, LB medium and M9CA (50:50) were used.

Plasmid vectors and other DNA templates

pET-15b ampicilin resistance (ampR) (Figure 9) and and pET-28b kanamycin resistance (kanR) (Figure 10) vector that carries T7 expression system were used for cloning all mutant constructs (Studier *et al*., 1990). pET-15b has a hexahistidine tag upstream of the multiple cloning site (MCS). Ligation of insert DNA into the vector in correct orientation will fuse hexahistidine tag to the N-terminus of expressed protein. In pET-28b, the hexahistidine tag is both on the upstream as well as the downstream of MCS, which allows the fusion of histag to either N- or C- terminus of recombinant protein. His-tagged proteins were purified by Ni-agarose affinity chromatography and gel filtration purification chromatography.

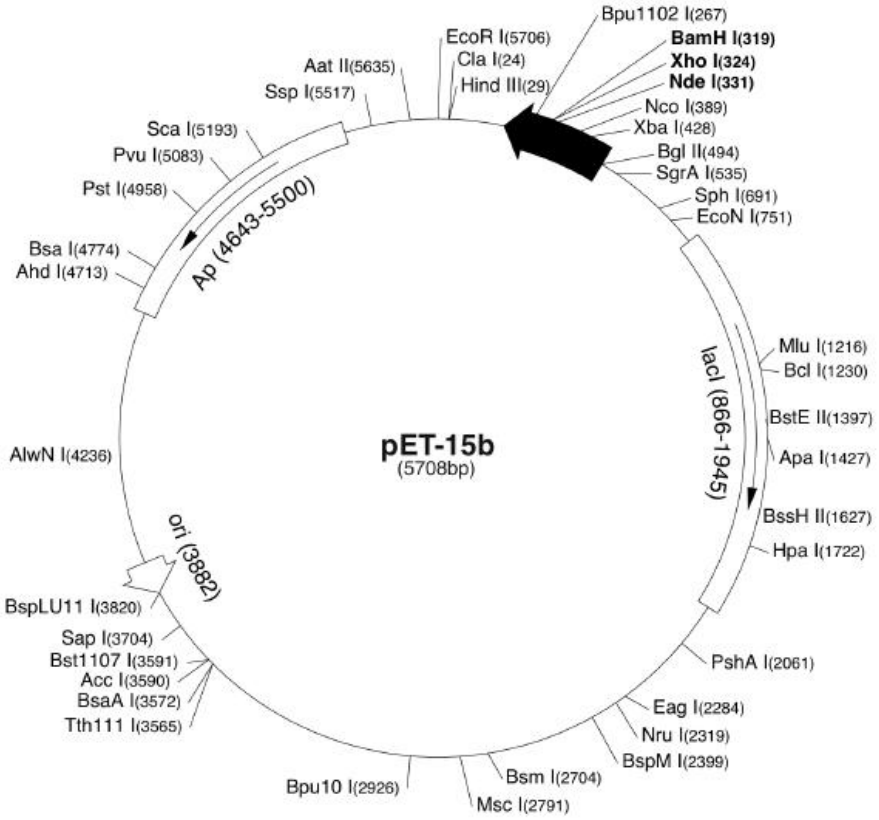


Figure 9: Map of pET-15b vector (www.novagen.com)

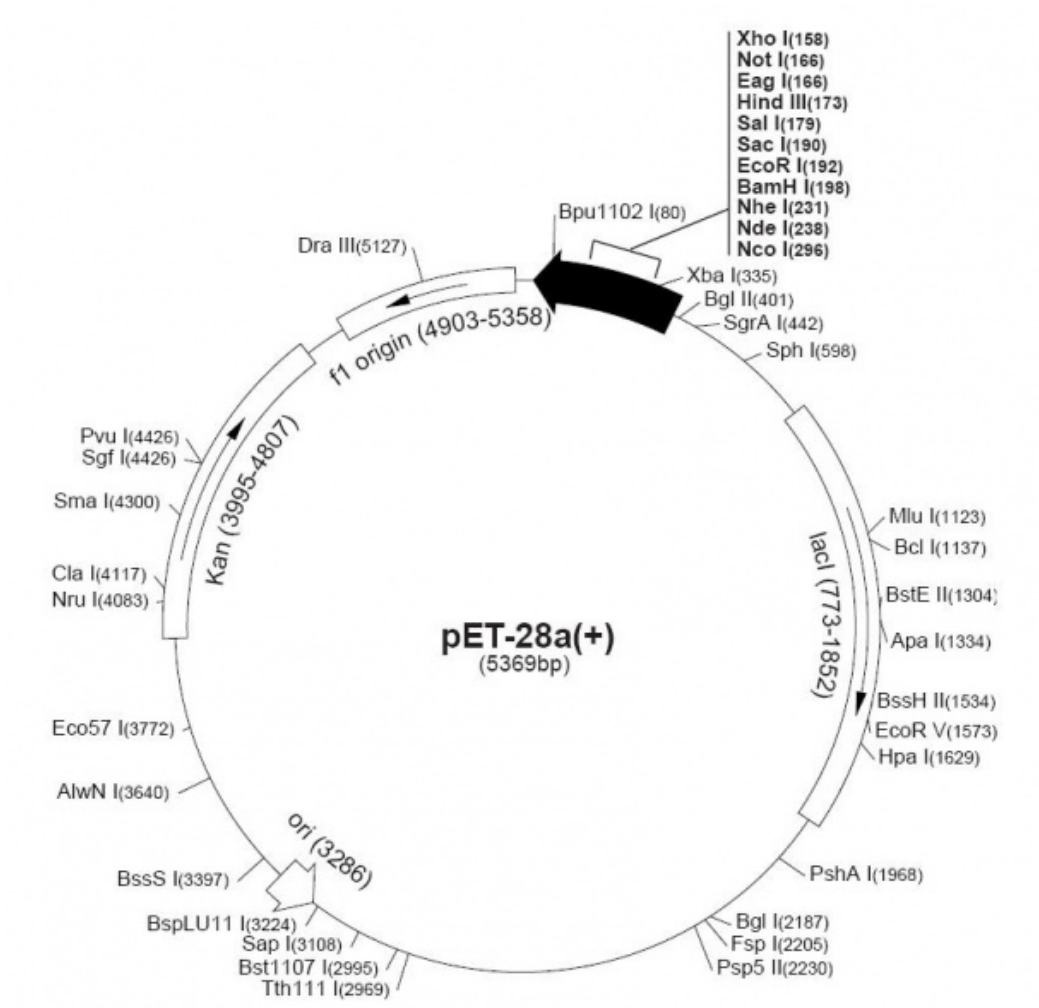


Figure 10: Map of pET-28 vector (www.novagen.com)

Construction of recombinant DNA using splicing by overlap extension (SOE)

All the recombinants used in this study were constructed either by one-step PCR or by PCR based splicing by overlap extension (SOE) strategy (Horton *et al.,* 1990). In one- step PCR, two sets of primers were used. Primer 1 is the forward primer and primer 2 is the reverse primer with the interested mutation inserted. In case of SOE method, as shown in Figure 11, four oligonucleotide primers and two successive PCR were used to construct the stitched DNA. In the first round of PCR, two halves of the gene were amplified separately by using primer 1 (forward primer) and primer 3 (stitch reverse primer) in the first PCR and primer 2 (stitch forward primer) and primer 4 (reverse primer) in second PCR tube. Both the PCR tubes contained 2X Master Mix (contains dNTPs, Taq polymerase and buffer), template DNA, and water to make up the volume to 25 µl. In the 2nd round of PCR, the two halves are stitched together to reconstitute a full-length gene. This PCR reaction mixture contained: 1ul of PCR reaction from each half, primer 1 (forward primer), primer 4 (reverse primer), 2X Master Mix and water to make up the reaction volume (50 µl). Formation of full-length gene product was confirmed by agarose gel electrophoresis (0.8% w/v) and EtBr staining (1 µg/ml).

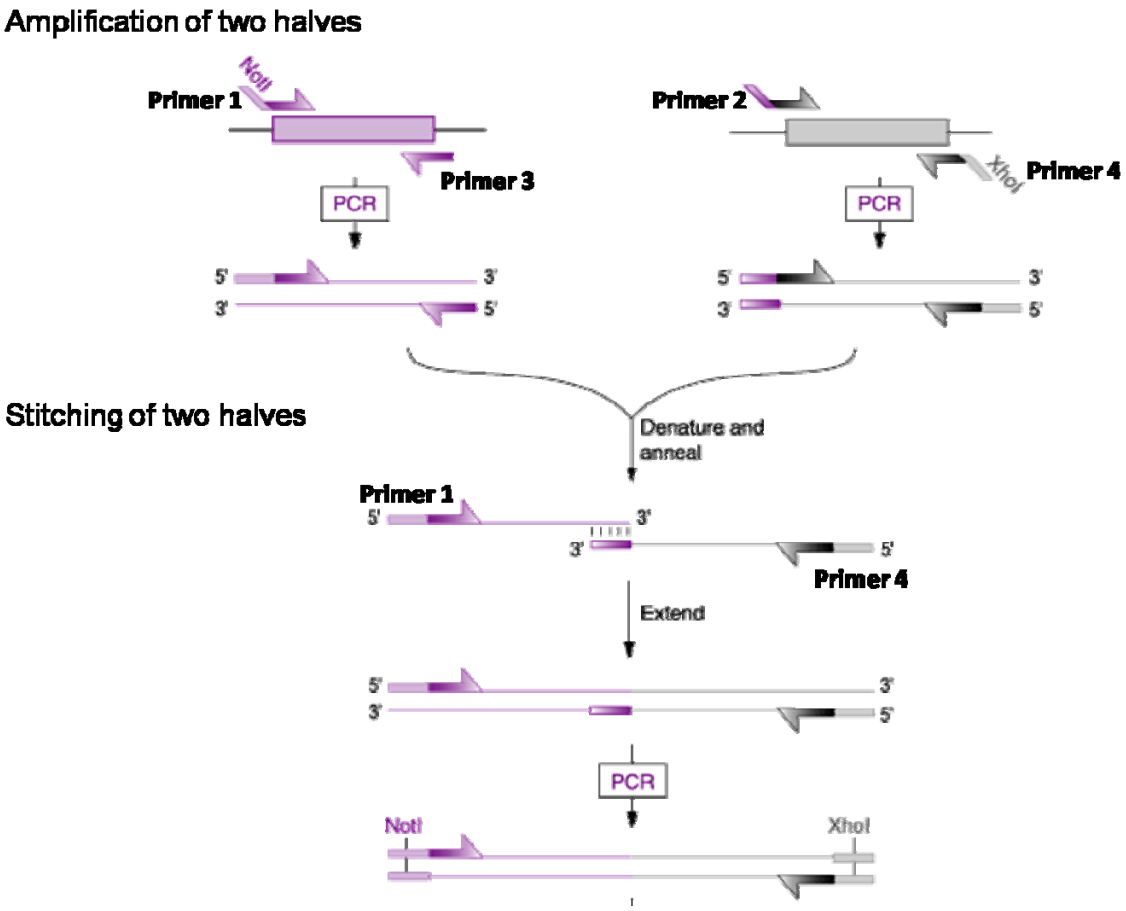


Figure 11: Splicing Overlap Extension. Primers 1 and primers 4 are end primers, while primer 2 and 3 are mutant primers (http://www.biochem.arizona.edu/classes/bioc47).

Primers used in this study

- **gp17 F329amber mutant:**

1. Gp17 forward

   5’ ***CGC*GGATCC** GATG GAA CAA CCG ATT AAT GTA TTA

2. Stitch primer forward

   5’ TTG TAT AAC GAT GAA ATT TAG GAC GAT GGA TGG CAA TGG TCG 3’

3. Stitch primer reverse 5’

   CGA CCA TTG CCA TCC ATC GTC CTA AAT TTC ATC GTT ATA CAA 3’

4. Gp17 reverse

   5’ ***CGC*GGATCC** T TAT ACC ATT GAC ATA CCA TGA GAT AC

- **gp17 DD330-331 double mutant library:**

1. Gp17 forward

   5’ ***CGC*GGATCC** GATG GAA CAA CCG ATT AAT GTA TTA 3’

2. Stitch primer forward

   5’CTT TAT AAC GAT GAA GAT ATT TTT NNN NNN GGA TGG CAA TGG AGC ATA CAA ACC 3’

3. Stitch primer reverse

   5’ GGT TTG TAT GCT CCA TTG CCA TCC NNN NNN AAA AAT ATC TTC ATC GTT ATA AAG

4. Gp17 reverse

5’ ***CGC*GGATCC** T TAT ACC ATT GAC ATA CCA TGA GAT AC 3’

- **gp17 F559amber mutant**

1. Gp17 forward

   5’ ***CGC*GGATCC** GATG GAA CAA CCG ATT AAT GTA TTA 3’

2. Stitch primer forward

   5’ TGG TTA TCA ACG CAG TCA AAA TAG ATT GAT GCG GAT AAA GAT GAC ATG C 3’

3. Stitch primer reverse

   5’G CAT GTC ATC TTT ATC CGC ATA ATC AAT CTA TTT TGA CTG CGT TGA TAA CCA 3’

4. Gp17 reverse

   5’ ***CGC*GGATCC** T TAT ACC ATT GAC ATA CCA TGA GAT AC 3’

- **gp17 DY 560-561 double mutant library**

1. Gp17 forward

   5’ ***CGC*GGATCC** GATG GAA CAA CCG ATT AAT GTA TTA 3’

2. Stitch primer forward

   5’ TGG TTA TCA ACG CAG TCA AAA TTT ATT NNN NNN GCG GAT AAA GAT GAC ATG C 3’

3. Stitch primer reverse

   5’G CAT GTC ATC TTT ATC CGC NNN NNN AAT AAA TTTT TGA CTG CGT TGA TAA CCA 3’

4. Gp17 reverse

5'' ***CGC*GGATCC** T TAT ACC ATT GAC ATA CCA TGA GAT AC 3'

- **gfp**

1. gfp forward

   5' ***CTA* GCT AGC** ATG AGT AAA GGA GAA CTT TTC ACT 3'

2. gfp reverse

   5' ***CTA* GCT AGC** TTT GTA TAG TTC ATC CAT GCC ATG TGT

- **gp17 full length**

1. GP17 forward

   5' ***CGC*GGATCC** GATG GAA CAA CCG ATT AAT GTA TTA 3'

2. Gp17 reverse

   5' ATG CAT GCC TCG AGT TAT ACC ATT GAC ATA CCA TGA GAT AC

- **gfp- wt peptide (helix loop-helix 333-352)**

1. gfp forward

   5' ***CTA* GCT AGC** ATG AGT AAA GGA GAA CTT TTC ACT 3'

2. Stitch primer forward

   5' CTG GTT CCG CGC GGC AGC TGG CAA TGG AGC ATA CAA ACC ATT AAT GGT TCT 3'

3. Stitch primer reverse

   5' GCT GCC GCG CGG AAC CAG TTT GTA TAG TTC 3'

4. Gp17 reverse

5’ CTC GAG GCA TGC ATT TAA TGT TCT TGA CGG AAT TGA GCT AAT GAA GAA CCA TTA ATG GT

- **gfp-scrambeled peptide**

1. gfp forward

   5’ ***CTA*** **GCT AGC** ATG AGT AAA GGA GAA CTT TTC ACT 3’

2. Stitch primer forward

   5’ TGG ACC TCT CAA ATT GGT CAT TTA CAA TGG TTC GTC AAT CAT AAA TGC ATG CCT CGA G 3’

3. Stitch primer reverse

   5' GCT GCC GCG CGG AAC CAG TTT GTA TAG TTC 3’

4. Scr-peptide reverse

   5’ CTC GAG GCA TGC ATT TAT GAT TGA CGAG AAC CAT TGT AAA TGA CCA ATT TGA GAG GTC CA 3’

- **gfp- loop-helix-loop-helix peptide**

1. gfp forward

   5’ ***CTA*** **GCT AGC** ATG AGT AAA GGA GAA CTT TTC ACT 3’

2. Stitch primer forward

   5’ATTACACATGGCATGGATGAACTATACAAAGAAGATATTTTT GACGATGGATGGCAATGG 3’

3. Stitch primer reverse

   5’CCATTGCCATCCATCGTCAAAAATATCTTCTTTGTATAGTTCA TCCATGCCATGTGTAAT 3’

4. Peptide reverse

   5'***ATG CAT GCC*** **TCG AG** TTA ACC ACT TAC TGG TGT AGG 3'

- **Rb69 soc- wt peptide**

1. Rb69-Soc forward

   5' ***CTA*** **GCTAGC**ATGGGTGGTTATGTAAACATCAAAACC 3'

2. Stitch primer forward

   5'CCT ACA CCA GTA AGT GGT TGG CAA TGG AGC ATA CAA AC 3'

3. Stitch primer reverse

   5'GT TTG TAT GCT CCA TTG CCA ACC ACT TAC TGG TGT AGG 3'

4. Peptide reverse

   5'***ATG CAT GCC*** **TCG AG** TTA ACC ACT TAC TGG TGT AGG 3'

- **Rb69-soc-scrambeled peptide**

1. Rb69-soc forward

   5' ***CTA*** **GCTAGC**ATGGGTGGTTATGTAAACATCAAAACC 3'

2. Stitch primer forward

   5'CCTACACCAGTAAGTGGTAGCGCTGAACAAAATATATGGACC TCTCAA 3'

3. Stitch primer reverse

   5'TTGAGAGGTCCATATATTTTGTTCACGCTACCACTTACTGGTG TAGG

4. Scr-peptide reverse

   5'***ATGCATGC*****CTCGAG**TTATGAACGGAACCATTGTAAATGACCA

   ATTTGAGAGGTCCATAT 3'

- **Rb69-soc F348A mutant peptide**

1. Rb69-soc forward

   5' ***CTA*** **GCTAGC**ATGGGTGGTTATGTAAACATCAAAACC 3'

2. Peptide 348A mutant reverse

5'***ATGCATGC*****CTCGAG**TTAATGTTCTTGACGAGCTTGAGCTAATGAAGA

ACCATTAATGGTTTGTATGCTCCATTGCCA 3'

- **Rb69-soc swap 1 mutant peptide**

1. Rb69-soc forward

   5' ***CTA*** **GCTAGC**ATGGGTGGTTATGTAAACATCAAAACC 3'

2. Swap 1 reverse

   5'***ATGCATGC*****CTCGAG**TTAATGTTCTTGTAAAGCAGCTTCTTTTG

   AAGAACCAGCAATCATTTTAGCGCTCCATTGCCA 3'

- **Rb69-soc swap 2 mutant peptide**

1. Rb69-soc forward

   5' ***CTA*** **GCTAGC**ATGGGTGGTTATGTAAACATCAAAACC 3'

2. Swap 2 reverse

   5'***ATGCATGC*****CTCGAG**TTAATGTTCTTGTAAAGCTTGTTCTAATG

   AAGAACCAGCAATAGCTTGAGAGCTCCATTGCCA3'

- **Rb69-soc swap 3 mutant peptide**

1. Rb69-soc forward

   5’ ***CTA*** **GCTAGC**ATGGGTGGTTATGTAAACATCAAAACC 3’

2. Swap 3 reverse

   5’***ATGCATGC*****CTCGAG**TTAATGTTCTTGACGAGCTTGAGCTAAA GAGGTATTTTTAATGGTTTGTATGCTCCATTGCCA 3’

- **Rb69-soc loop-helix-loop-helix peptide**

1. Rb69-soc forward

   5’ ***CTA*** **GCTAGC**ATGGGTGGTTATGTAAACATCAAAACC 3’

2. Stitch primer forward

   5’GCTATGTTTACACCTACACCAGTAAGTGGTGAAGATATTTTTG ACGATGGATGGCAATGG 3’

3. Stitch primer reverse

   5’CCATTGCCATCCATCGTCAAAAATATCTTCACCACTTACTGGT GTAGGTGTAAACATAGC 3’

4. Peptide reverse

   5’***ATG CAT GC*****C TCG AG** TTA ACC ACT TAC TGG TGT AGG 3’

Sequence in bold letters indicates the restriction enzyme cutting site and italicized sequence indicates the tag sequence inserted for efficient cutting by the restriction enzyme.

gfp-gp17 fusion protein

For construction of gfp-gp17 gene, the gfp gene was amplified by using gfp forward and gfp reverse primers with Nhe1 restriction site on both the ends. gfp was

cloned into pET- 28b vector using the Nhe1 restriction cutting site. The gp17 gene was amplified with BamH1 site at the N-terminus and Xho1 site at the C-terminus. The amplified gene was cloned into gfp vector by cutting the vector with BamH1 and Xho1 restriction enzymes, which is downstream of gfp gene with a seven amino acids T7 tag as linker between gfp and gp17.

Ammonium acetate/isopropanol precipitation

Equal volume of 8 M ammonium acetate was added to the PCR mixture. The total volume was doubled by adding isopropanol and then incubated at room temperature for 20 minutes. DNA was sedimented by centrifugation at 14,000 rpm for 30 min. The supernatant was discarded and the pellet was washed twice with ice cold 80% ethanol. The pellet was dried for a minute at 37°C to evaporate the left-over ethanol. Finally the DNA pellet was dissolved in 30-50 µl of sterile Milli-Q H2O.

Restriction enzyme digestion

Restriction enzyme digestions was done in a final reaction volume of 30 µl, containing 2 µl BamH1 (in case of single digestion) and 1 µl of Nhe1 plus 1 µl of Xho1 (in case of double digestion) (using restriction enzymes from New England Biolabs or Fast digest from Fermentas). Digestions were performed in 30 µl of reaction volume containing, 20 µl of substrate DNA (2.0 ug), 3 µl of 10X buffer, and sterile Milli-Q water. The reaction mixture was incubated for 1 h at 37°C, and were stopped either by heat inactivation of the restriction enzyme i.e. transferring the reaction mixture tube to 67°C for 20 min or by adding 10X agarose loading dye containing EDTA.

Preparation of vector

pET-28b plasmids were purified using plasmid isolation kit from Promega. The purified plasmids were digested with NheI and Xho1 restriction enzymes (Fast Digest from Fermentas). To prevent self-ligation of the vector, the vector was dephosphorylated by alkaline phosphatase (Promega).

Agarose gel purification of restriction fragments

Agarose gel electrophoresis was used to purify the restriction digested inserts and vector DNA from heterogeneous mixture of DNA. The DNA loaded gel was run at 100V. After 1 h, the DNA was visualized on ultraviolet tansiluminator and the desired fragments were excised from the gel using a clean razor blade. Qiaquick PCR purification kit (Qiagen Valencia, CA) was used to extract the DNA from the gel. The DNA was eluted using 30 µl of elution buffer containing (10 mM Tris HCl pH 8.5)

Ligation reaction

To ligate the insert and vector DNAs, a 15 µl of reaction mixture of 5:1 insert DNA to vector ratio was used. The reaction mixture was prepared by mixing 11.5 µl of insert DNA with 1 µl of either pET-15b (BamH1 digested) or pET-28b (Nhe1 and Xho1 digested) plus, 1.5 µl of 10X ligase reaction buffer. The reaction mixture was incubated at 68°C for 8 min and cooled on ice for 2 min and 1 µl of T4 DNA ligase (New England biolab) was added and incubated in 16°C water bath for overnight. The product was analyzed on 0.8% agarose gel. The formation of ladder of insert DNAs is indicative of successful ligation. The ligated DNA was transformed into the XL-10-Gold *E.coli* for long-term maintenance of the plasmid.

Transformation of plasmids into XL-10-Gold ultracompetent cells

XL-10-Gold ultracompetent cells were gently thawed on ice and a 25 ul aliquot was transferred to a pre-chilled 2059 polypropylene falcon tube. 1ul of β-mercaptoethanol was added and incubated for 10 min on ice, occasionally swirling the tube. 1 ul of ligation mixture was added to the tube and incubated for another 30 min on ice. The cells were heat-shocked exactly for 30 sec at 42°C. The tube was then immediately cooled on ice for 2 min. 500 ul of SOC medium was added to the tube and the culture was incubated at in 37°C shaker for 1-3 h. 50 and 100 ul of the cells were separately plated on LB agar plate containing either ampicilin (50 ug/ml) (for pET-15b vector) or kanamycin (36 ug/ml) (for pET-28b vector). The plates were incubated overnight at 37°C. Next day the isolated colonies were picked and grown in 2 ml of LB+amp or LB+Kan media for further analysis. The presence of insert was confirmed either by colony PCR or by restriction digestion of purified plasmids.

Plasmid preparation

Colonies were picked and grown overnight in LB media containing the appropriate antibiotics. Plasmids were purified using Qiagen manual purification method or Promega/ Fermentas plasmid purification kit.

Transformation of plasmid into expression into BL-21(DE3)pLys-S or BL-21(DE3)-RPIL cells

Either BL -21(DE3)pLys-S or BL-21(DE3)-RPIL cells from Novagen were gently thawed on ice and 20 ul of cells were transferred into pre-chilled 2059 polypropylene falcon tube. 1 ul of β-mercaptoethanol was added and incubated for 10 min on ice by

occasionally swirling the tube. One ul of purified plasmid DNA was added and incubated for another 30 min on ice. The cells were heat-shocked for 30 sec and immediately cooled on ice for 2 min. 200 ul of SOC medium was added to the tube and the culture was incubated at 37°C for 1 h. 50 and 100 ul of the cells were separately plated on LB agar containing appropriate antibiotics. The plates were incubated for overnight at 37°C incubator. Well isolated colonies were picked and grown for 7 h in liquid media containing appropriate antibiotics.

Expression and solubility tests

To test the expression of the protein, 200 µl of culture was added to 10 ml of media containing appropriate antibiotics and was grown at 30°C. When the cell density was $4x10^8$, IPTG to a final concentration of 1 mM was added. 500 µl of culture was taken out before and after 1 h and 2 h, IPTG induction. The culture was centrifuged at 6000 rpm for 5 minutes. Supernatant was discarded and the pellets were dissolved in 100 µl of Milli-Q water. 10 ul of sample plus 10 ul of 2X SDS PAGE buffer was used to load the gel. For testing the solubility of the protein, 1.5 ml culture at 2 h time point was taken and centrifuged at 6000 rpm for 10 min. The supernatant was discarded and the pellet was resuspended in 300 ul of B-PER reagent from Pierce. The sample was subjected to vortex for 5 min and centrifuged at high speed for 20 min. The supernatant was transferred to a new eppendorf tube and pellet was resuspended in 300 ul of B-PER reagent. 10 ul of supernatant and 10 ul of pellet subjected to SDS-PAGE analysis.

Once the mutant was confirmed for protein expression and solubility, the plasmids were sent for sequencing. When the sequencing results confirmed of no second site mutation in the gene, the large scale protein purification was carried out.

Large scale expression and purification of recombinant proteins

Frozen cultures of streak purified BL-21(DE3) pLys-S cells or BL-21(DE3)-RPIL cells expressing desired protein were used to inoculate 20 ml of Moore medium containing kanmycin + chloromophenicol antibiotics and grown at 37°C shaker for 7-8 h. Next day the culture was inoculated into 1 liter of Moore medium containing kanamycin and chloromophenicol and grown at 30°C. When the cell count was $4x10^8$cells/ml, IPTG was added. 500 ul of cell culture was taken out before and after 2 h of IPTG addition for SDS-PAGE analysis. After 2 h of protein induction, cells were centrifuged at 7000 rpm for 12 min at 4°C. The supernatant was discarded and the pellets were stored at -70^{o}C.

Following day, the cells were thawed and completely resuspended in 40 ml of buffer 1 (20 mM Tris HCl pH 8.0, 100 mM of NaCl, 5 mM MgCl2, and 1 mM of ATP) containing 10 mM of immidazole (binding buffer) and one EDTA-free protease tablet from ROCHE. The cells were lysed by passing through the French Press cell (Aminco) at a pressure of 1000-1400 psi. The cell lysate was centrifuged at 16000 rpm for 25 min. Supernatant was collected in 50 ml tube and pellet was resuspended in equal volume of buffer 1 (20 ul of which was used in parallel to run the SDS-PAGE). The supernatant was filtered using 0.2 uM filter (Milipore) and loaded onto 1 ml Histrap column (GE healthcare) pre-equilibrated with buffer1 (binding buffer). The column was washed with buffer 2 (buffer 1 + 40 mM immidazole) (wash buffer). The flow- through and wash from

the column were collected for SDS-PAGE analysis. The bound protein was eluted with 50 ml of 40-500 mM immidazole gradient using buffer 2 and buffer 3 (buffer 1 + 500 mM immidazole). Histagged protein elutes at 80% of gradient (400 mM immidazole concentration). The elution was monitored by UV absorbance at 280 nm. The peak fractions were collected pooled together and 5 ml of which was injected onto 10/60 Superdex 200 (prep-grade) gel filtration column (GE healthcare), pre-equilibrated with gel filtration buffer (20 mM Tris HCl pH 8.0, 100 mM of NaCl, 5 mM Mgcl2). The peak fractions from Histrap and gel filtration column were collected separately and analyzed for the presence and purity of protein. The major peak from gel filtration column corresponding to the molecular weight of interested protein was collected, pooled together and concentrated using 15 ml Amicon centrifugal filter with a molecular weight cut-off of 3 kDa. The concentrated fractions were stored in aliquots at -70°C until further use. The concentration of protein was determined by SDS-PAGE and absorbance at 280 nm.

In some cases where the recombinant protein was insoluble it was purified from the pellet fraction. The pellets were solubilized in 10 ml of 8 M urea. Solubilized sample was centrifuged at 15000 rpm for 30 min at room temperature. Supernatant was collected and loaded onto 1ml Histrap column pre equilibrated with 8 M urea. The protein was renatured on the column with 80 ml of (20 mM Tris pH8.0, 100 mM NaCl) buffer containing 8-0 M urea gradient. All the procedures were carried out at the room temperature. After renaturation, the protein was eluted and as processed as described earlier.

Preparation of 17am18amrII phage stocks

Single colony of *E.coli* B40 was inoculated in 20 ml of LB+M9CA media (50:50). The cells were grown for 5 h in 37°C shaker. Six plates of 17am18amrII phages were prepared by incubating 300 ul of B40 with 300 ul plaque purified 17am18amrII (~$2x10^6$/ml) at 37°C for 7 min then adding 2.5 ml of top agar and pouring on LB plates. The plates were incubated at 37°C for overnight. The following day, 20 ml of B40 were inoculated into 1 liter of LB+M9CA media and cells were grown at 37°C for 2-2.5 h. When the cell count was $3x10^8$cells/ml, the cells were infected with mutant phage by scraping off the top agar from six plates into the culture. The culture was allowed to shake for 2-3 h and then centrifuged at 12500 rpm for 60 min using GSA rotor. The supernatant was discarded and the pellet was completely resuspended in 60 ml of Pi-MgSO4 buffer (50 mM phosphate buffer pH 7.0, 70 mM NaCl, 1 mM MgSO4) containing 60 ul of 7.5 ug/ul DNase (Sigma) and several drops of chloroform. The sample was incubated in 37°C shaker for 60 min and centrifuged at 7000 rpm for 12 min. Supernatant containing the mutant phage was collected in a fresh tube and phage titer was determined by plating serially diluted phages on B40. The purified mutant phages were also plated on P301 plates to check the revertant rate and wt phage contamination. If the concentration of the mutant phage was above $7x10^{10}$ it was then used for prohead preparation.

Preparation of proheads

Single colony of *E.coli* P301 was inoculated in 10 ml of LB+M9CA medium (50:50) and grown for 5 h in 37°C shaker. Next day, 10 ml of P301 was inoculated into 50:50 LB+M9CA medium (500 ml) and grown at 37°C for 2-2.5 h. When the cell count was $3x10^8$cells/ml, it was infected with 17am18amrII mutant phage at M.O.I of 4 (multiplicity of infection). After 7 min it was superinfected with the same M.O.I of mutant phage. The culture was kept in 37°C shaker for another 28-30 min and then centrifuged at 7000 rpm for 12 min. The pellet was resuspended in 50 ml of prohead buffer (50 mM Tris HCl pH 7.5, 5 mM MgCl2, 3 mM β-mercaptoethanol) containing 50 mM of potassium glutamate, 10 ug/ml DNAse, half EDTA-free protease tablet from Roche and few drops of chloroform. The pellet was incubated at 37°C shaker for 25 min and after adding 50 ml of prohead buffer containing 600 mM of NaCl it was centrifuged at 7000 rpm for 12 min. The supernatant was then transferred to 6 Corex glass tubes and centrifuged at 18000 rpm for 45 min. The supernatants were discarded and the pellets were resuspended in 10 ml of prohead binding buffer (50 mM Tris HCl pH 7.5, 5 mM MgCl2, 3 mM β-mercaptoethanol) and centrifuged at 7000 rpm for 12 min. The supernatant were collected and filtered using 0.2 um filter. The filtered sample was loaded onto DEAE Sephacel (GE health care) column pre-equilibrated with prohead binding buffer. The DEAE column separates the particles based on charge. It separates the ELPs (Empty Large Particles) from ESPs (Empty Small Particles) and the` DNA contaminant. 80 ml gradient of 0-300 mM NaCl, was used to elute the proheads. Three major peaks eluted from the DEAE column, the first large sharp peak corresponds to

ELPS, the second smaller peak corresponds to ESPs and the third large peak at the end of the gradient corresponds to DNA (Rao *et al*., 2006). Only the first major peak, which elutes at 70% of salt gradient (~210 mM NaCl) corresponding to the ELPs were collected. The fractions were concentrated to 1 ml by using 15 ml Amicon filters and then diluted with 4 ml of binding buffer to reduce the salt concentration to 20 mM. The concentrated sample was injected into a 1 ml of MonQ column 5/50 GL (GE healthcare) pre-equilibrated with 10 ml of prohead binding buffer. The column was washed with 10 ml of binding buffer and the proheads were eluted with 40 ml of 0-300 mM of salt gradient. Depending on the UV absorbance value proheads were further concentrated or directly stored in aliquots at -70°C.

Prohead binding assay

The proheads ($2x10^{11}$) were incubated with gp17 at 1:50 ratio of gp20:gp17 or a control protein in 500ul of binding buffer (50 mM Tris HCl pH 7.5, 100 mM NaCl, 5 mM MgCl2) for 30 min at room temperature. Proheads were sedimented by centrifugation at 18000 rpm for 45 min at 4°C. The supernatant was collected in new tube and 20 ul of the sample was subjected to SDS-PAGE analysis. The pellets were washed with 1ml of binding buffer and centrifuged at 18000 rpm for 45 min. The supernatant was discarded and to the pellet 20 ul of binding buffer was added, resuspended and transferred to a new tube containing 20 ul of 2X SDS-PAGE buffer. The proheads without any protein added was used as a control.

Quantification of gp17 copy number

Destained SDS-PAGE gels were washed thoroughly with Milli-Q water and scanned using laser densitometry. Image-quant software was used to quantify the density of gp17 and gp20. The copy number of gp17 was calculated by normalizing the density of the gp17 band to that of the gp20 band, whose copy number was known to be 12-copies/prohead particle.

***In vitro* DNA packaging inhibition assay**

The peptides used in DNA packaging inhibition assay were fused to the soc protein of RB69 phage. To minimize any binding of soc-peptides to the soc binding sites on the capsid, proheads were first saturated with RB69 wt-soc. $4x10^{9}$ proheads were incubated with 1:10 proheads: soc ratio at 37°C for 15 min. soc-saturated proheads were then incubated with soc-peptide constructs at 1:6 gp17: peptide ratio in 20 ul buffer (65

mM of Tris HCl pH 7.5, 5 mM MgCl2, 80 mM NaCl). After 15 min incubation at 37°C, other packaging components; 600 ng of lambda DNA, gp17 (1 uM final concentration) and final concentration of 1 mM ATP was added and incubated at 37°C for 45 min. To degrade the unpackaged DNA DNase I was added to a final concentration of 0.5 ug/ul and incubated at 37°C for 30 min. The reaction mixture was then treated with proteinase K cocktail (50 mM EDTA, 0.5 ug/ul proteinase K, 0.2% SDS) and incubated at 67°C for 30 min. The reaction mixture was run on 0.8% agarose gel for further analysis. The packaged DNA in the absence of any soc peptide was considered as 100% and the packaging activity in the presence of various peptides were normalized accordingly.

Construction of F329am and F559am mutant phages

The gp17 F329am gene in the plasmid maintained in *E.coli* BL-21 cells was transferred into T4 genome by recombinational marker rescue. *E.coli* P301 or *E.coli* phenylalanine containing phenylalanine amber suppressor was used for marker rescue. Amber suppressor of *E.coli* has the modified UAG suppressor t-RNA which is charged with one of the following amino acids. alanine, cystenine, glutamic acid, lysine, tyrosine, phenylaalanine, glutamine, arginine, aspargine. *E.coli* suppressor plates were prepared by mixing 300 ul of $2x10^9$ cells/ml with 3 ml of top agar and pouring on LB agar plate. 2ul of transformants were spotted on the plates containing P301 and phenylalanine amber suppressor lawn. The K166am phage (2 ul) was spotted on top of the BL21 (F329am) cells. The plates were incubated at 37°C for overnight. The transformants containing F329am inserts after recombinational exchange rescue the defective phage from K166am mutation. Along with it, the F329am mutation is transferred into the T4 genome. The

resultant phages were screened from the heterogeneous mixture of progeny phages by comparing the lysis in P301 and phenylalanine suppressor plate. wt phages will form plaques on both the plates, K166am cannot form plaques on any plates and F329am will form plaques only on phenylalanine suppressor plates. Those plaques were picked dispensed into 500 ul of Pi-Mg phage dilution buffer containing few drops of chloroform. Serial dilutions of phages were made and the infection process was repeated on the phenylalanine suppressor cells and P301 to confirm the purity of the isolated F329am phage. The purified F329am phage was spot tested on 11 available *E.coli* amber suppressors. 2 ul of F329am phage zero stock was spotted on the suppressor lawn and allowed to air dry. The plates were incubated overnight at 37°C. If the lysis was observed in any of the plates, that particular amino acid is considered as tolerated at F329 position of gp17. Following the same procedure, F558am phage was constructed.

Construction of double mutant library (DD330-331, DY560-561)

The XL-10 Gold cells carrying plasmids with gp17 mutant libraries either DD330-331 library or DY560-561 library were transferred into the T4 genome by marker rescue using the F329am and F558am mutant phages respectively. To do marker rescue assay, *E.coli* P301 plates were prepared as described earlier. 2 ul of transformants were first spotted followed by 2 ul of either F329am (concentration ~ $2x10^6$) in case of DD330-331 library or F558am (concentration ~ $2x10^6$) in case of DY560-561 library. Once the spots were completely dried the plates were incubated at 37°C for overnight. Next day, the plates were observed for lysis. Based on the plaque morphology, the clones were grouped into wt, null mutants (no lysis) or small plaque mutants. A wt gp17

construct and a plasmid with no insert were used as positive and negative control respectively. Large numbers of clones were screened to ensure that all possible combinations were represented in the library. Functional clones were further tested for their plaque forming ability at 3 different temperatures, 25°C, 37°C and 42°C. Based on the phenotype they are grouped as wt, temperature sensitive (*ts*) or cold sensitive (*cs*) mutants.

Orientation test and sequencing of portal binding site I and Site II mutants

Plasmids were isolated from all functional and some null mutants from both libraries (DD330-331 & DY560-561). As the clones of the libraries were constructed by digesting with single enzyme (BamH1) on both the ends, ligated insert will have 50:50 chances of right: wrong orientation. To test the orientation, the plasmids were cut with BglII which will cut the insert and vector only once. If the clone is in right orientation, digestion with BglII will produce a large fragment of 7100bp and a smaller fragment of 425bp: otherwise, 5783bp and 1755bp fragments will be generated if the clone is in wrong orientation. Only the right orientation clones were sequenced.

Suppressors screening

The *ts* (DD-331-DQ) phage was prepared as described earlier. Serially diluted phage containing 10^6, 10^5, 10^4 were plated on *E.coli* P301 plates and incubated at 45°C for overnight. To compare the plaque morphology, ~ 10^2 wt phages were plated and incubated at the same temperature. Since DQ (*ts*) does not form plaques at 45°C, any plaques formed in those plates will be the suppressor phage. The plaques can be either true revertants or second site suppressors. In true revertant, the mutation is suppressed by

reverting back to the original wt sequence, hence it will have the plaque morphology similar to the wt phage. In case of second site suppressor the mutation is suppressed by 2nd site mutation somewhere in the same gene or different gene. The second site suppressor will have plaque morphology slightly different from the wt phages, usually of smaller size. In this study only the second site suppressors were screened. Isolated suppressor plaques were collected in 1ml of Pi-Mg buffer and plaque purified. Suppressor phenotype was further confirmed by spot testing at 3 different temperatures as described earlier. To prepare the template DNA, the plaques were collected in 100 ul of 1X PCR buffer and centrifuged at 3000 rpm for 5 minutes. 5 ul of supernatant was used in PCR amplification of gp17 and gp20 genes. The amplified DNA was gel purified and sequenced to determine the precise second site mutation.

Bioinformatics tools used

Multiple sequence alignments of terminase protein of T4 family phages were done using ClustalW program from http://www.ebi.ac.uk/Tools/clustalw2/index.html, with the default parameter. Analysis of crystal structure of gp17 of T4 phage (protein ID 3CPE, 3EZK) and Spp1 gp20 (protein ID 2JES) were done by using Pymol software from Delano 2002; http://pymol.sourceforge.net

Results

Characterization of gp17-gp20 interaction

The major components of the DNA translocation machinery are the terminase protein (gp17) and the portal protein (gp20). Genetic and biochemical studies indicated gp17-gp20 interaction in T4, λ, T3 and Spp1 (Lin *et al*., 1999; Yeo and Feiss, 1995; Morita *et al*., 1995; Oliveira *et al*., 2006). However, direct evidence for this interaction has not been established in any of the phages. In T4, overexpression of gp20 in *E.coli* yields inclusion bodies, thus making direct use of gp20 protein to study gp17-gp20 interaction has not been possible. Proheads containing gp20 attached were used to develop a prohead binding assay (Figure: 12). Moreover, prohead associated gp20 would be the natural substrate for packaging motor assembly. In this assay, the purified proheads were incubated with gp17 at room temperature for 30 minutes and the gp17-prohead complex was isolated by high speed centrifugation. The binding was assessed by SDS-PAGE. If gp17 interacts with gp20 on the prohead, after high speed centrifugation it would have associated with proheads and appeared as a gp17 band upon gel electrophoresis. The results clearly showed that gp17 binds to the proheads as it is associated with the proheads (Figure: 13, lane 2) compare to control (lane 3). The binding of gp17 to the proheads is specific since the control protein BSA did not show any binding (lane 4).

Prohead is a complex structure, composed of more than 1500 protein molecules encoded by 12 different genes with molecular weights varying from 10-70 kDa,

(Mesyanzhinov *et al*., 2004). In SDS-PAGE, the bands of full-length gp17 (70 kDa) and gp17-K577 (63 kDa), which is C-terminal 33 amino acid truncated gp17 (see Figure: 5 from Introduction) overlap with the prohead protein bands corresponding to alt (70 kDa) and gp20 (61 kDa) respectively (Figure: 13B). K577 band can be separated from the gp20 band by running the sample on 12% SDS-PAGE for 4 hours, However, running the SDS gel for such a long period of time causes smearing of bands, which makes further analysis difficult (see Figure: 13A, lane 1). To overcome this limitation, a gfp-gp17 fusion protein was constructed. The fusion of gfp to gp17 increases the molecular weight by 29 kDa which shifts the gp17 band to a higher position on the gel (104 kDa) where there are no background bands from the proheads.

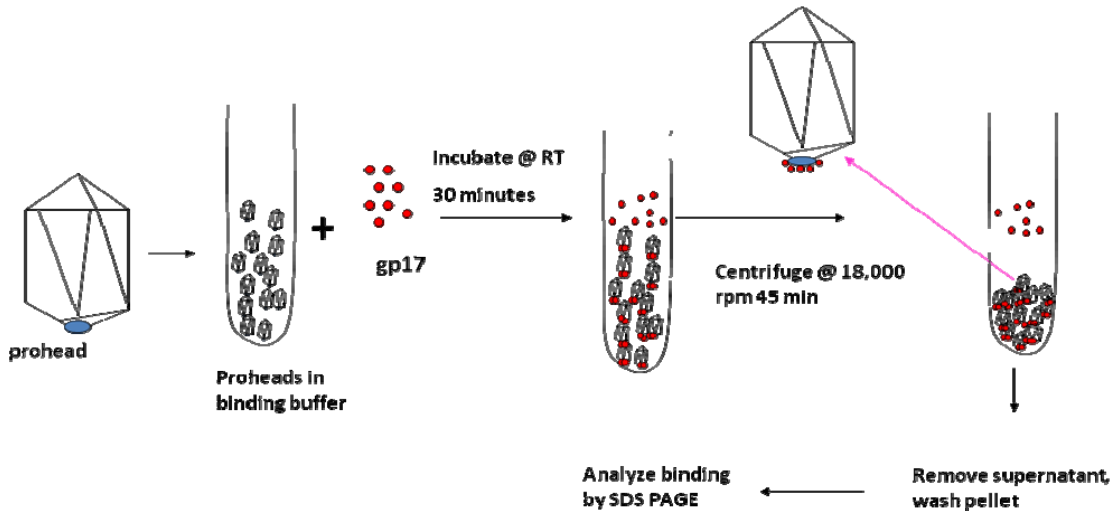


Figure 12: Schematic of prohead binding assay. Proheads were incubated at room temperature for 30 minutes and then centrifuged at 18000 rpm for 45 minutes. The bound gp17 will sediment along with proheads and unbound gp17 will be in the supernatant. Pellets were washed with binding buffer and binding was analyzed on SDS-PAGE.

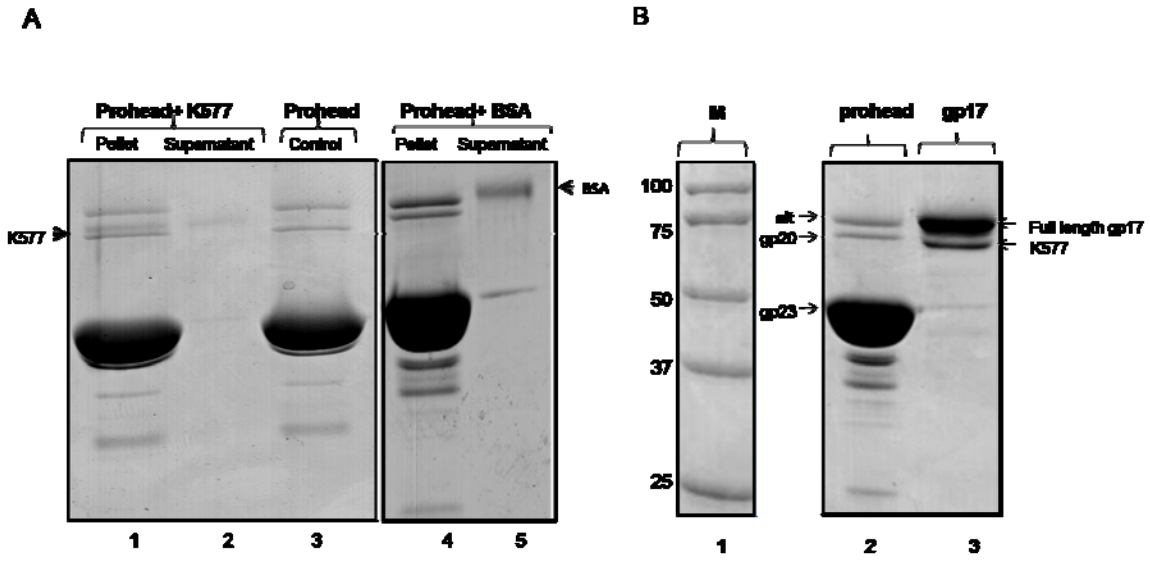


Figure 13: Specific binding of gp17 to proheads.
A: Proheads (30 nM, $2x10^{11}$ particles) were incubated either with K577 (lane1) or BSA (lane 4) at 50:1 ratio of protein to gp20 in binding buffer (50 mM Tris pH 7.5, 5 mM MgCl2, 100 mM NaCl) for 30 minutes at room temperature and binding experiment was performed as described in Materials and Methods. On SDS-PAGE, the K577 band overlaps with gp20 band from prohead to separate these two bands, samples were electrophoresed on 12% gel for 4hours (lane 1-3). Lanes 2 and 5 are the unbound K577 and BSA respectively in supernatant fraction.
B: SDS-PAGE showing overlapping of alt and gp20 bands from proheads (lane 2) with gp17-full length and gp17-K577 (lane 3) band respectively. M (lane 1) is the molecular weight markers

Construction and characterization of gfp-gp17 fusion protein

The gfp-gp17 fusion protein was constructed using the splicing by overlap extension PCR technique. gfp was fused to the N-terminus of full length gp17 with a seven amino acid linker in between (Figure: 14A) to add enough flexibility for gfp and gp17 to fold in a native conformation. The protein was overexpressed and purified by affinity and gel filtration chromatography as described in Materials and Methods (Figure: 14B). The gel filtration profile showed that the protein elutes as a monomer, similar to the wt gp17. The fusion protein was then tested for *in vitro* DNA packaging functional

assay. It showed comparable ATPase and DNA packaging activities (Figure: 14C) to that of wt gp17, suggesting that fusion of gfp to the N-terminus of gp17 doesn't affect the function.

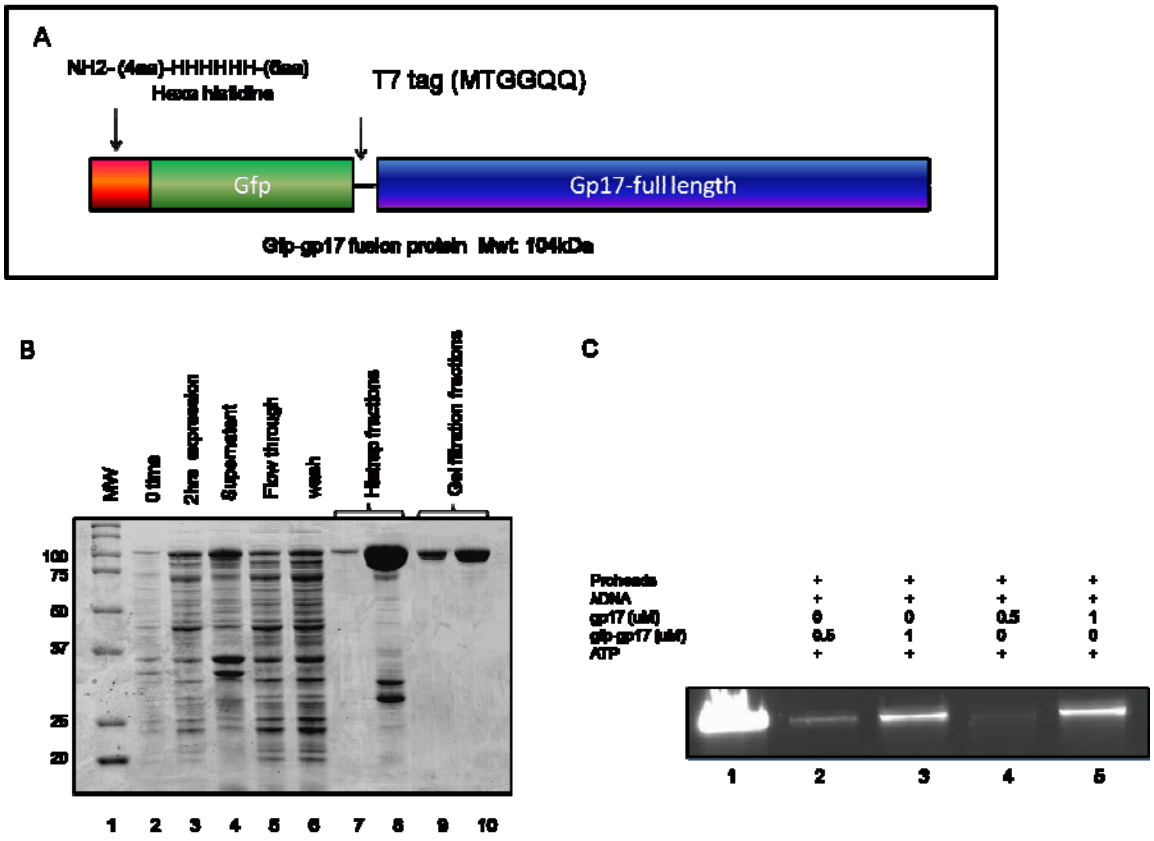


Figure 14: Construction and functional characterization of gfp-full length gp17.
A: Schematic of gfp-gp17 fusion construct. gfp vector is first constructed by cloning gfp at Nhe1 restriction site of pET-28b vector. Full length gp17 gene is then cloned to gfp-vector at BamH1 restriction site with T7 tag (MTGGQQ) as a linker between gfp and gp17). Histag for the protein purification is at the N-terminus of gfp.
B: Expression and purification of gfp-gp17. gfp-gp17 fusion protein was overexpressed in *E.coli* and purified by passing through histrap column followed by gel filtration column. Molecular weight marker is shown in lane 1. Before (0 h) and after (2 h) IPTG induction shown in lanes 2 and 3 respectively. Lysate (lane 4) was applied to the histrap column. Flow-through and wash from histrap column are shown in lanes 5 and 6 respectively. Histrap fractions (lane 7 and 8) were pooled together and applied to gel filtration column. The peak corresponding to gfp-gp17 (lane 9 and 10) was collected and used for functional assays.
C: Comparison of *in vitro* DNA packaging activity of gfp-gp17 fusion protein with gp17. DNA packaging was done by incubating 0.3 nM, $2x10^9$ prohead particles and 1 nM, 1x10

[10] DNA molecules (lane 1) with either 0.5 uM gfp-gp17 (lane 2) , 1 uM gfp-gp17 (lane 3) or 0.5 uM gp17 (lane 4), 1 uM gp17 (lane 5).

To test binding of gfp-gp17 to the proheads, proheads were incubated with an increasing ratio of gfp-gp17 to gp20 at room temperature for 30 minutes. The unbound protein was removed by washing with binding buffer. The binding sample was then analyzed by SDS-PAGE. As shown in Figure: 15A, binding of gfp-gp17 to the proheads can be visualized clearly by the appearance of a new band in lanes 3-5, when compared to the control lane 2. The binding of gfp-gp17 to the proheads increased with an increasing ratio of protein to the proheads. At a ratio of one molecule of gp20 to fifty molecules of gfp-gp17, no further increase in binding was observed (compare lanes 4 & 5). To evaluate further, the specificity of gfp-gp17 binding to the proheads, proheads were also incubated with gfp alone as a control. As shown in (Figure: 15B, lane 5) gfp itself does not to bind to the proheads. Binding assay was also done in the presence of ATP, DNA, or ATP and DNA to test their effects on gp17 binding to the proheads. As shown in Figure: 15C, neither ATP (lane 4) nor DNA (lane 6) or ATP and DNA (lane 8) had any effect on gfp-gp17 binding (compare to lane 2). However, due to low intensity of gfp-gp17 binding in this experiment, the result needs to be confirmed further.

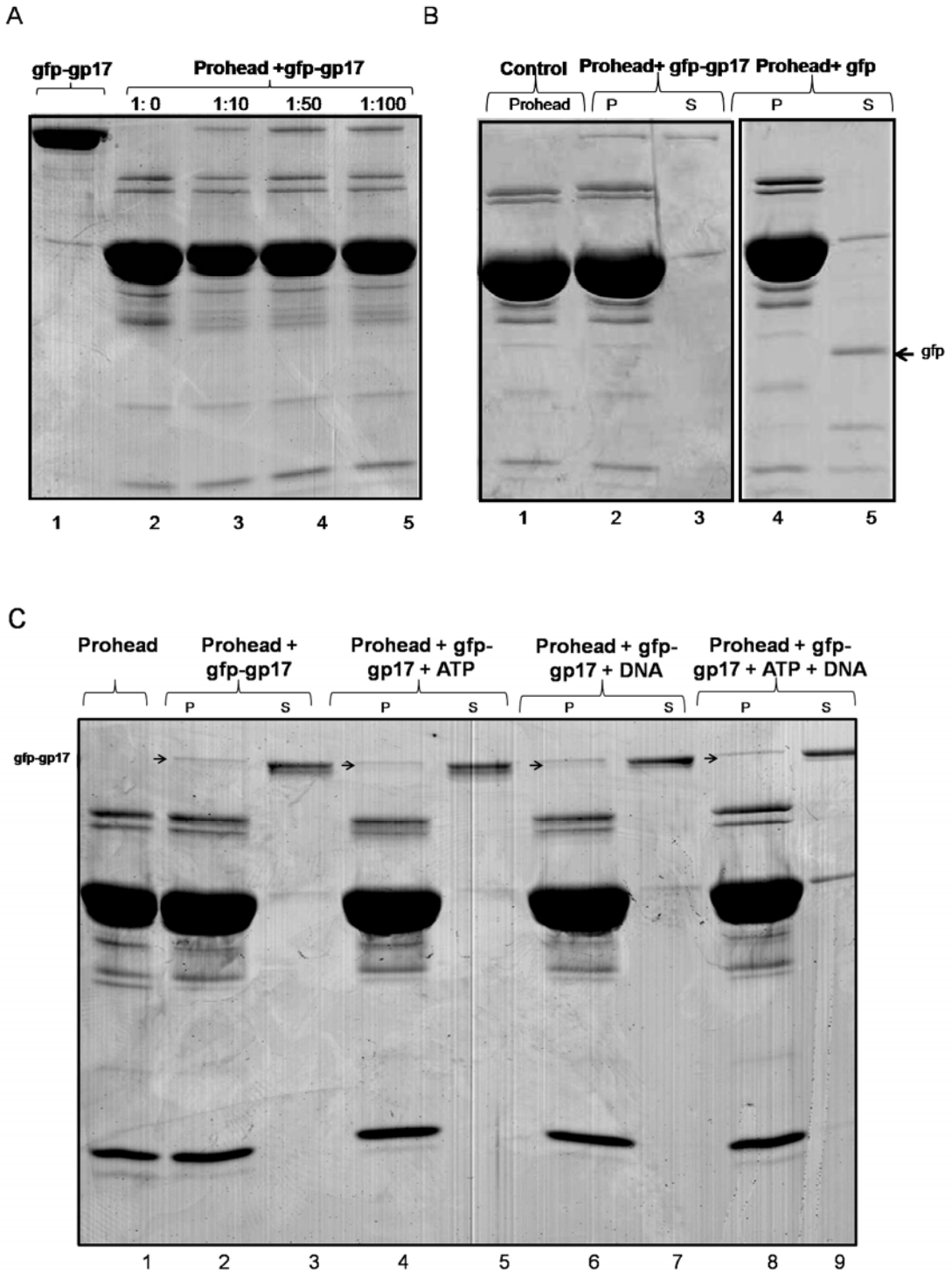


Figure 15: Binding of gfp-gp17 to the proheads. A: Binding experiment was done by incubating proheads (2x 10^{11}) particles with increasing ratio of gfp-gp17 (lane 2-5) to gp20 as described in Material and Methods. B: Specificity of binding: proheads (2x 10^{11}) were incubated with either gfp-gp17 (lane 2) or gfp (lane 4) at ratio of 1:50 of protein to

gp20 at room temperature for 30 minutes and binding was analyzed described in Materials and Methods. Unbound proteins are shown in supernatant fraction gfp-gp17 (lane 3) and gfp (lane 5).
C: Effect of ATP or DNA or ATP and DNA on gfp-gp17 binding. Binding experiment was done by incubating proheads (2x 10^{11}) with gfp-gp17 at 50: 1 ratio of protein to gp20 along with either 500 mM ATP (lane 4) or 100bp of DNA (2x 10^{13}) molecules (lane 6) or 500 mM ATP and DNA (2x 10^{13}) molecules (lane 8) at room temperature for 30 minutes. Supernatant (S) from each binding sample is loaded next to the pellet fractions (P). Control proheads (lane 1) and gfp-gp17 binding (lane 2).

Interaction of gp17 with the prohead occurs through the gp20 portal

To show that gfp-gp17 interacts with gp20 but not with other capsid proteins (major capsid protein gp23; minor capsid protein gp24). gp20-minus proheads were prepared as described in the Materials and Methods.

Binding assays were done by incubating gp20 minus proheads with gfp-gp17. No significant binding of gfp-gp17 to the gp20 minus proheads was observed (Figure: 16A, lane 2) when compared to binding of gfp-gp17 to the wt proheads (lane 1). To rule out the possibility that the loss of gp17 binding to the gp20 minus proheads was due to the SDS treatment, a PA-Hoc (capsid binding protein) was used as a control. No significant difference in binding of PA-Hoc was observed (compare lane 3& 4), suggesting that the loss of gp17 binding is due to release of gp20 from proheads, but not due to SDS treatment *per se*.

To further validate that gfp-gp17 interaction requires the exposed region of gp20 on the proheads, gfp-gp17 was also tested for binding to mature T4. Once the DNA packaging is completed, the terminase protein dissociates from the portal, the tail and tail fibers are attached to the portal. Thus, gp17 binding site is not be exposed in matured T4

phages; therefore, gfp-gp17 should not bind to the T4 phages. As expected, gfp-gp17 did not bind to the T4 phages (Figure: 16B, lane 3).

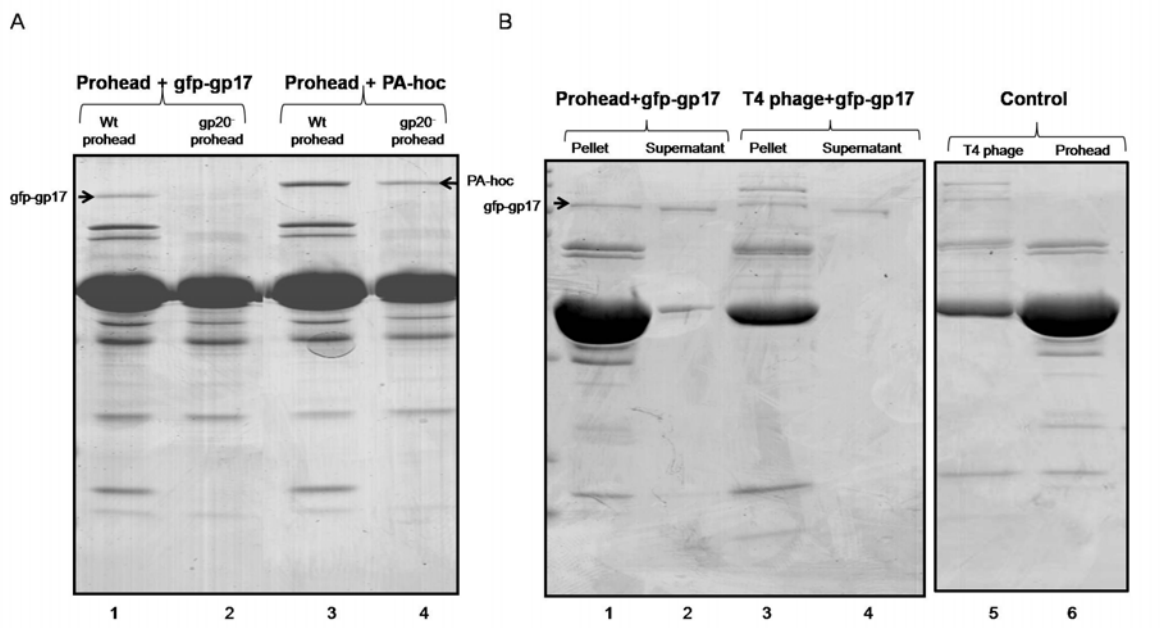


Figure 16: gfp-gp17 specifically interacts with gp20 on the prohead. A: Binding assay with gp20 minus proheads. gp20-minus proheads were prepared as described in Materials and Methods. The concentration of gp20-minus proheads were determined by quantifying the gp23 band. wt and gp20-minus proheads (2x 10^{11}) were incubated with gfp-gp17 (lane 1 and 2) or control protein PA-Hoc (lane 3 and 4) at 50: 1 ratio of protein to gp20, under standard prohead binding condition.
B: Binding assay with T4 phage. gfp-gp17 was incubated either with wt proheads (lane 1) or T4 phages (lane 3) at 50: 1 ratio of protein to gp20, under standard prohead binding conditions. Control: T4 phages (lane 5) and proheads (lane 6).

gp17 interacts with gp20 through the N-terminal ATPase

Earlier, it was shown that gp17 can be separated into N360 (1-360 residues) as the ATPase domain and C360 (360-577) as the nuclease domain. The separated domains exhibit their individual activities but, either individually or together cannot support DNA packaging, suggesting that communication between the two domains is critical for translocation of DNA into the prohead (Kanamaru *et al*, 2004).

To determine which domain interacts with gp20 on the proheads, prohead binding assay was done using N360 and C360 domains. As shown in (Figure: 17, lane 2 and 4) at 100 mM salt, the domains, N360 as well as C360 bound to the proheads. To rule out non-specific interaction between the proheads and the domains, the binding assay was repeated with more stringent conditions (300 mM salt). Binding of C360 was greatly reduced under the 300 mM salt condition (lane 5), whereas N360 binding was not affected (lane 3) suggesting that N360 binding is a specific binding. In addition, Cryo EM reconstruction results confirmed that, the N360 domain interacts with gp20 on the prohead, and that the C360 domain was away from the gp20 in gp17- prohead complex (Figure: 20 Sun *et.al*., 2008).

The reason for C360 binding at 100 mM salt is not clear. One of the possible explanations could be that, separation of full length gp17 into domains exposes charged residues on the surface of the C360 domain, hence, at low salt condition C360 binds non-specifically to the capsid.

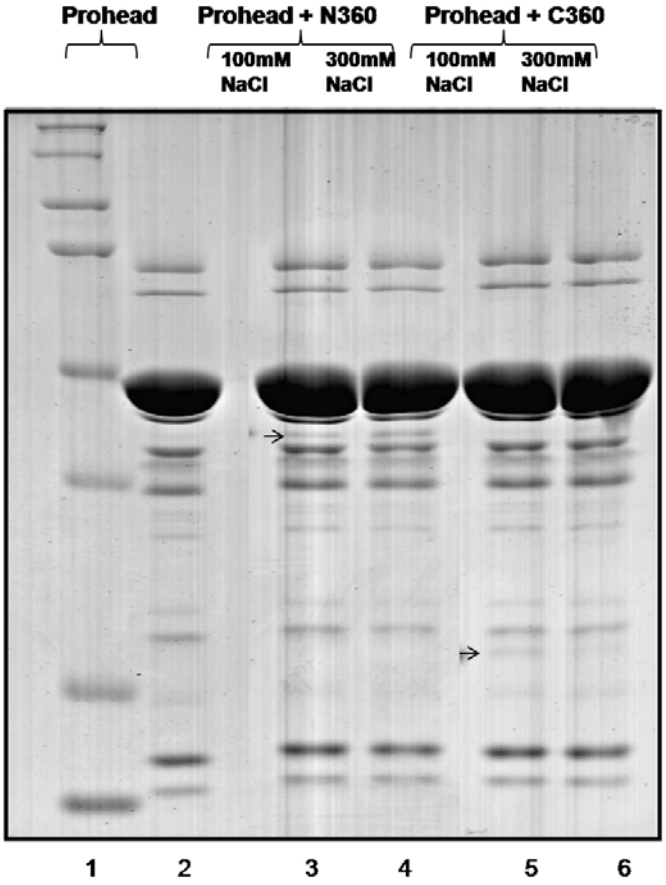


Figure 17: Prohead binding assay with N360 and C360 domain of gp17. Proheads ($2x10^{11}$) were incubated with either N360 or C360 at 50:1 ratio of protein to gp20 under standard prohead binding condition. Lane 1: Molecular weight marker. Lane 2: control prohead. Lane 3 and 4: proheads incubated with N360 in binding buffer containing 100 mM and 300 mM NaCl respectively. Lane 5 and 6: proheads incubated with C360 in the presence of 100 mM and 300 mM salt respectively. The arrow indicates the position of N360 and C360 bands in respective lanes.

Stoichiometry of gp17 packaging motor assembled on the prohead

Although the stoichiometry of portal protein is well conserved among the phages the stoichiometry of gp17 terminase in the packaging motor complex has not been established. In T3, biochemical studies suggested that there are six subunits of the

terminase per prohead (Miyo Morita *et.al*, 1991). In λ, four heterotrimers of two gpNu1 to one gpA has been proposed (Maluf *et al*, 2005). The stoichiometry of gp17 in the gp17-prohead complex was determined by a direct binding assay. The number of gp17 bound to the prohead was calculated from the bound gp17 density. A total of fourteen independent gfp-gp17–prohead binding assays were done. The calculated copy number was five in 50% of the experiments, six in 42% of experiments and four in 14% of the experiments (Figure: 18). These data suggested that there might be five subunits of gp17 per prohead. However, SDS-gel density scanning can not unambiguously resolve the difference between the copy numbers of five and six. To resolve this difference more direct structural analysis of gfp-gp17 bound to prohead was done. The Cryo-EM reconstruction of prohead bound gfp-gp17 at the portal was performed by imposing five-fold as well as six-fold symmetry. The five-fold symmetry showed discrete densities for five gp17 molecule whereas, the six-fold symmetry was smooth and featureless (Figure: 19, Sun *et al*, 2008). Thus, the Cryo-EM reconstruction results clearly indicated that there are five subunits of gfp-gp17 per prohead particle.

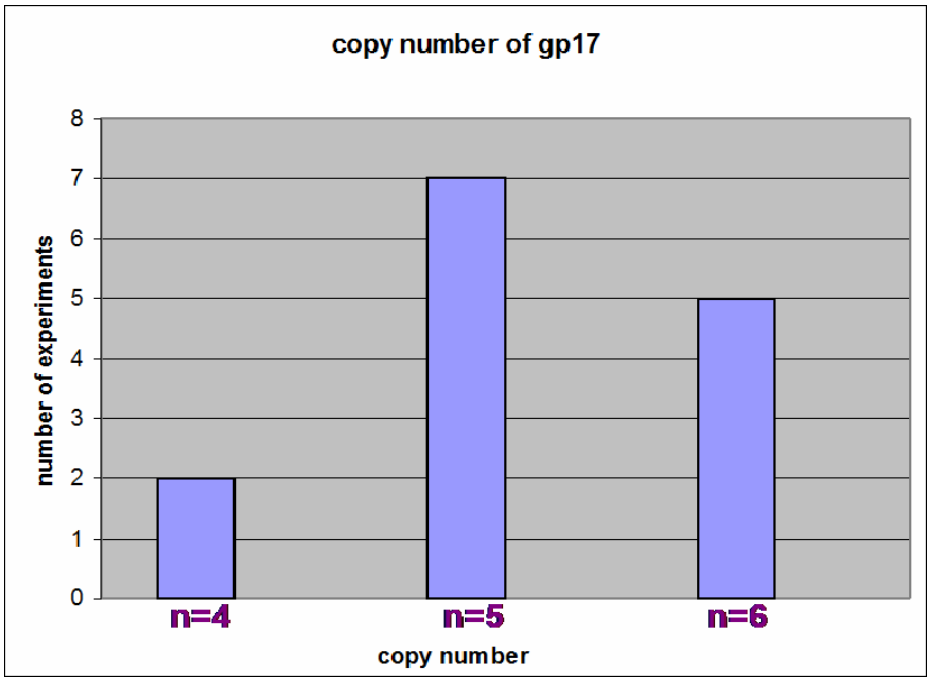


Figure 18: Stoichiometry of terminase in DNA packaging complex. A: The copy number of gfp-gp17 in prohead complex was calculated by scanning the density of gfp-gp17 band and normalizing it to the density of gp20 band whose copy number is known as 12 per prohead particle. Y axis is the number experiments done and X axis is the copy number of gp17 prohead particle. Total number of experiments n= 14.

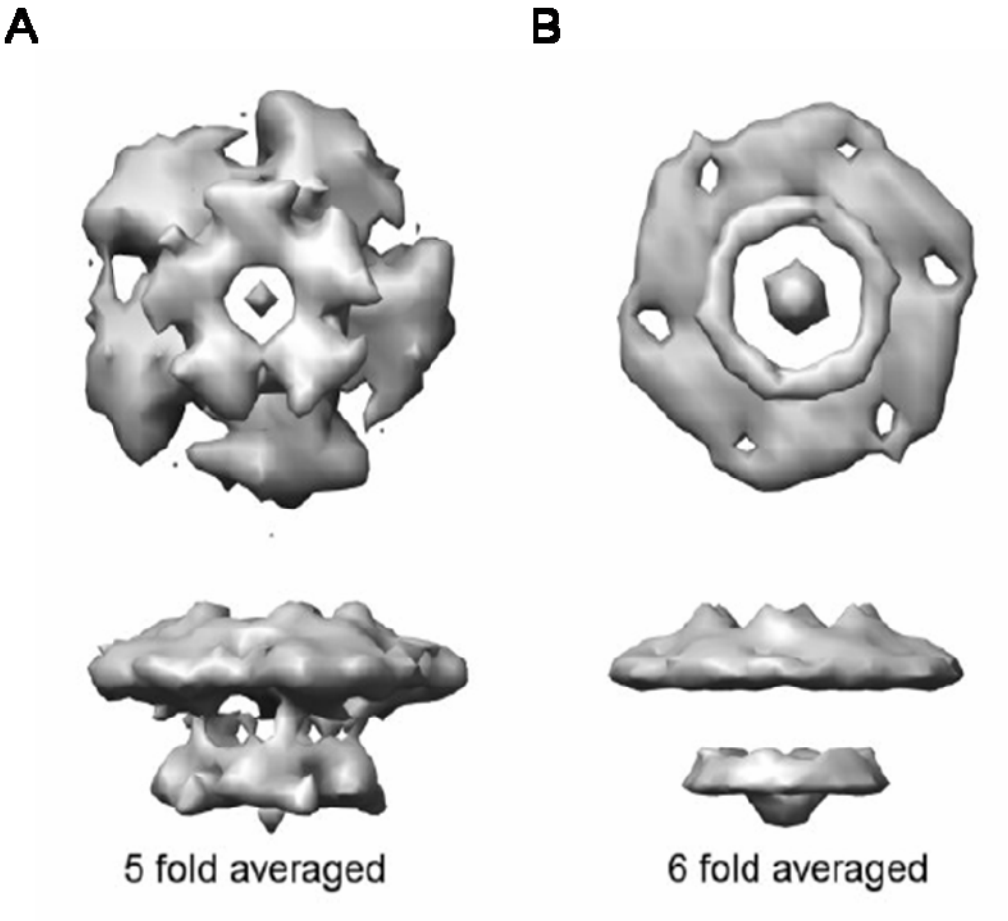


Figure 19: Cryo-EM reconstruction of gp17 bound to the prohead assuming 5 or 6 fold symmetry (Sun *et.al*, 2008).
A: Five fold averaged gp17 map shows discrete density for five subunits of N- and C-domains of gp17.
B: Six-fold averaged gp17 map shows smooth, featureless density for N- and C-domain of gp17.

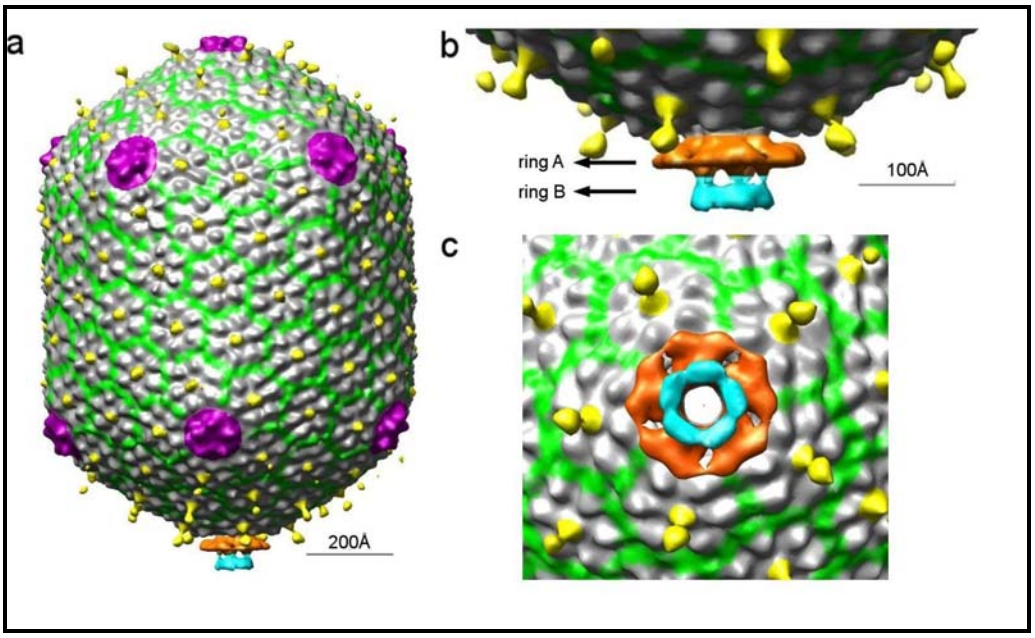


Figure 20: Cryo-EM reconstructions prohead assembled packaging motor.
a: Cryo-EM reconstructions showing the prohead assembled gp17 through gp20, situated at the special portal vertex of the capsid. b: Cryo-EM reconstruction at higher magnification showing the N-domain in orange (ring A) attached to the gp20 and C-domain in cyan (ring B) facing away from gp20. c: bottom view showing gp17 binding to the special vertex of the capsid where gp20 is present (Sun *et al*., 2008).

Mapping of gp20 interacting region/residues on gp17

Sequence analysis proposed two regions on gp17 (Figure: 21) as the portal interacting sites, portal binding site I: LYNDEDIFDD $_{323\text{-}331}$ and portal binding site II: FIDYADKDD $_{559\text{-}567}$ (Hsiao & Black, 1977; Lin *et al*., 1999). This was based on genetic evidence in which *cs/ts* mutations in gp20 were suppressed by second site mutation in gp17 near these binding sites. These sites contain acidic and hydrophobic residues similar to that reported for the putative binding sites in phage λ and T3 terminases.

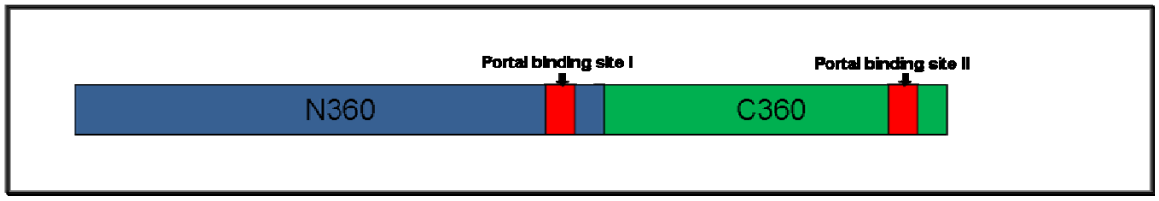


Figure 21: Predicted gp20 interacting regions on gp17. A: Schematic of gp17 protein showing the position of predicted gp20 interacting regions. The portal binding site I (highlighted in red) is located in the N360 domain of gp17, portal binding site II (red) is in the C360 domain of gp17.

Analysis of portal binding site I & site II

Bioinformatics

To assess the importance of the portal binding sites in T4 related phages, extensive sequence alignments of T4 family phage was carried out (Figure: 22). Although at present, 18 T4 family phage sequences are available in the database, at the time of constructing the portal binding site mutants only 9 T4 related phages were sequenced.

Sequence alignment showed that almost all phage terminases retained a similar stretch of amino acids at both the sites.

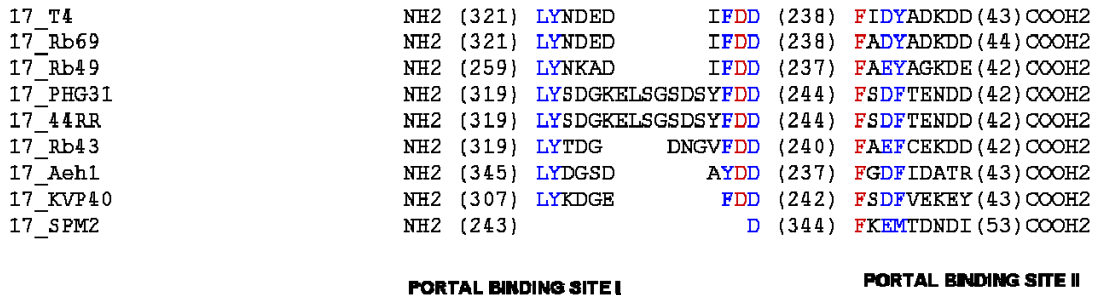


Figure 22: Sequence alignment of eight T4 family phage terminases by CLUSTAL W. The numbers in parenthesis represent the amino acids which are not shown in the alignment. The conserved and semi-conserved residues from the portal binding site I and II are represented in blue and red respectively. The conserved residues FDD 329-331 from portal binding site I and F558, DY560-561 from portal binding site II were selected for mutational analysis. All sequences were obtained from Genebank, the accession numbers are; T4: NP 17312, RB69: NP 861869, RB49: NP 891724, 44RR: NP 932508, RB43: YP 238880, Aeh1: NP 944105, KVP40: NP 899601, SPM2: YP 195134

Mutagenesis

Combinatorial mutagenesis approach was used to construct the double mutant library for DD330-331 and DY560-561 residues. Using this strategy, the importance of double or triple amino acids can be tested by substituting with all possible combinations. However, this approach requires an amber mutant phage with the amber mutation very close to the target site. Therefore, F329am and F558am mutant phages were first constructed. An amber stop codon at F329 or F558 position was introduced using the SOE-PCR strategy. The amber mutation was then transferred to the T4 genome by recombinational marker rescue. F329am and F558am phages were then recovered in *E.coli*-phenylalanine suppressor background. To check the importance of F329 and F558

residues, the F329am and F558am mutant phages were plated on 12 available *E.coli* suppressors. The F329am phage only tolerated relatively conservative substitutions; leucine, tyrosine and histidne (Figure: 23A). On the other hand, the F558am phage tolerated more substitutions; alanine, cystine, leucine, tyrosine and histidine (Figure: 23B).

Double mutant library was constructed for D330-D331 and D560-Y561 residues. Using SOE-PCR all possible codon combinations were introduced at DD330-331 and DY560-561 positions. Each template of gp17 mutant gene will have different combinations of amino acids in those positions. The amplified mutant gene was cloned into pET-15b under T7 promoter. The recombinant plasmids were then transferred to *E.coli* Bl-21 competent cells. The importance of each transformants carrying the mutated gene was then tested by transferring the mutant gene into the T4 genome by marker rescue. F329am phage was used for the marker rescue of D330-D331 mutant library and F558am phage was used for the marker rescue of D560-Y561 mutant library. Each transformant from library was scored either as a functional (lysis of *E.coli*) or Null (no lysis) mutant. In the case of DD330-331 library, a total of 780 clones were screened of which only 6 clones were functional (0.7%). A battery of functional and non functional clones was sequenced to determine which substitutions were tolerated. Interestingly all functional clones retained aspartic acid at at-least one of the positions. The second tolerated position contained relatively conserved substitutions such as glutamic acid, aspargine, glutamine, alanine and cysteine (Figure: 23A). On the other hand, D560-Y561 library screen showed that 47% of the clones are functional, indicating that many

substitutions at DY560-561 position are tolerated. Sequencing of the functional and null mutants showed that many non conserved substitutions such as proline and alanine or isoleucine and tryptophan are tolerated. This suggests that the conserved residues from the portal binding site I are important for function but those from the portal binding site II are not that important.

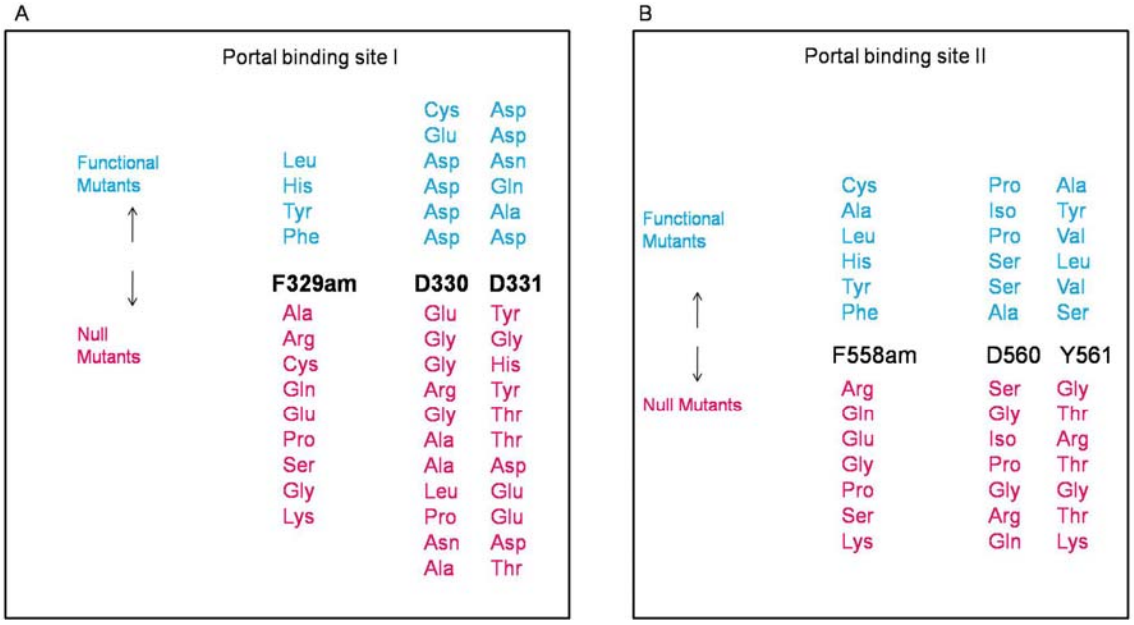


Figure 23: Mutagenesis of portal binding site I and II residues.
A: Mutational analysis of conserved residues F329, DD330-331 of portal binding site I.
B: Mutational analysis of portal binding site II F558, DY560-561 residues. Amber mutation was introduced to F329 and F558 residues. The wt sequences are shown in black color in the middle. Sequences in blue above the wt sequence are tolerated (functional clones) and the sequences in pink below the wt sequence are not tolerated (null-mutants). Only few null mutants' sequences are listed here.

Conditional lethality of portal binding site I mutants

Functional clones from the DD library were further tested for conditional lethality i.e. plaque forming ability at 3 different temperatures (25°C, 37°C, 42°C). Of the six functional clones, one clone showed a wt phenotype; means it formed plaques at all 3 temperatures, three clones showed *ts* phenotype (did not form plaques at 42°C) and two clones showed extreme *ts* phenotype (did not form plaques at 42°C and formed small plaques at 37°C). The sequencing data revealed that the wt phenotype turned out to be truly wild type showing the same DD sequence. Any functional substitutions at the DD residues, even the conservative substitution such as aspargine resulted in temperature sensitive phenotype. Substitutions at D330 resulted in more severe temperature sensitive phenotype producing small plaques even at 37°C (table 1). These results suggest that both the DD330-331 residues are very important for gp17 function; however the D330 residue is even more critical than the D331 residue.

Clones	37^{0}C	42^{0}C	25^{0}C	Phenotype	Sequence
129	+	+	+	Wt	DD
1033	+	-	+	*ts*	DQ
1036	+	-	+	*ts*	DA
648	+	-	+	*ts*	DN
43	+(small plaques)	-	+	*ts*	CD
990	+(small plaques)	-	+	*ts*	ED

Table1: Characterization of portal binding site I functional clones. Functional clones were marker rescued at three different temperatures 25°C, 37°C and 42°C. Based on the plaque morphology, clones were scored either as wt or *ts*. (+) indicates plaque formation. (-) no plaque formation.

Isolation of second site suppressors

According to our hypothesis if the portal binding site I (DD330-331) residues are involved in gp20 interaction, then the *ts* phenotype of these mutants can be suppressed by second site suppressor mutations in gp20 (inter-genic suppressors) or in gp17 (intra-genic suppressors). Second site suppressor mutants were screened for the DQ 330-331(*ts*) phage. Different dilutions of DQ 330-331 (*ts*) phage were plated at 45°C and screened for suppressor plaques. Only those plaques whose plaque morphology at 45°C was different from the wt phages (true revertants) were analyzed further. A total of 7 suppressors were selected and plaque-purified. The gp20 and gp17 genes were amplified and sequenced (table 2). All the suppressors had a second site mutation in the gp17 N360 domain at residue I337. This residue is adjacent to the *ts* mutation (D331) in a helix-loop-helix

motif. In four of the suppressor I337 was mutated to M and in one of the suppressor I337 was mutated to K. All retained the original D331Q mutation but none of the suppressors had any second site mutation in gp20 gene.

Mutants	Mutation	Phenotype
Temperature sensitive	gp17 D331Q	No plaques @ 45^{0}C
Suppressor	gp17D331Q+ I337M/I337K	Plaque formation @ 45^{0}C

Table 2: The second site mutation and phenotype of (D331Q) *ts* suppressor mutant phage D331Q temperature sensitive phages were prepared as described in Material and Methods. *E.coli* P301 plates were infected with about 10^5 phages and the plates were incubated at 45^{o}C and screened for suppressors. Sequencing of isolated suppressor showed a second site mutation in gp17.

Structural analysis identified helix-loop-helix region as gp20 interacting region

Recent structural analysis i.e. fitting of gp17 structure to the Cryo-EM identified the same helix-loop-helix (WQWSIQT INGSSLAQFRQEH $_{333\text{-}352}$) contact with gp20 on the proheads (Figure: 24). Helix- loop- helix region is just one amino acid downstream of predicted portal binding site I.

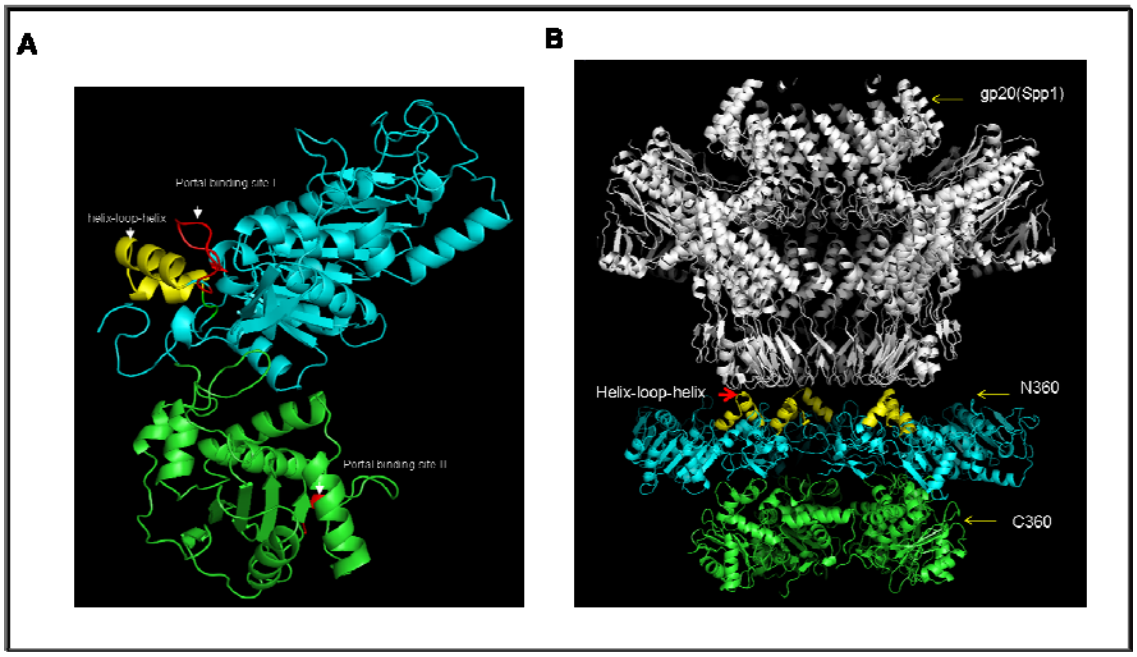


Figure 24A: X-ray structure of gp17 with portal binding sites and helix loop helix region are highlighted. The portal binding site I (red) in N360-domain (cyan), site II (red) in C360-domain (green) and helix loop helix region (yellow) in N360-domain
B: Structural alignment of Spp1 portal with T4 gp17 structure. Using PyMol structural analysis software, Spp1 portal (PDB accession number 2jes) was structurally aligned with T4 gp17 pentamer (PDB accession number 3ezk). The N360-domain of gp17 was then closely analyzed for region/regions which contact the gp20 structure. The helix-loop-helix region from sub-domain II of N360 was found at the interface of gp17 and gp20 structure which was selected as the gp20 interacting region (highlighted in yellow). Spp1 gp20 is represented in white, N360 is in Cyan and C360 is in green. The image is created using PyMol software.

Analysis of helix-loop-helix (HLH) motif

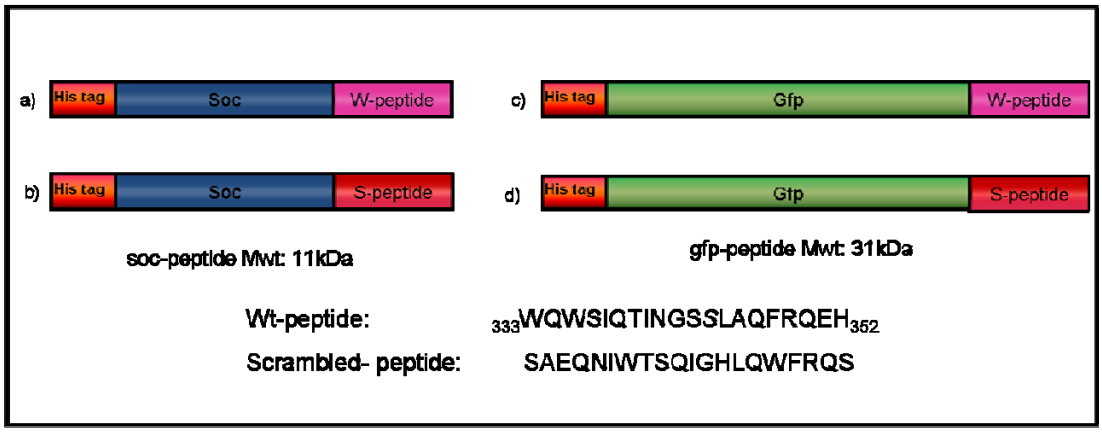
To test whether the HLH motif is a gp20 interacting region, a peptide corresponding to HLH motif was constructed. As this peptide consists of large percentage of hydrophobic residues, the synthetic peptide would highly be insoluble. To produce a soluble peptide, an alternate strategy was used. The peptide sequence was fused to a soluble protein, RB69-soc (9 kDa) and gfp (29 kDa). wt-peptide and the control scrambled peptide were fused to the C-terminus of soc or gfp (Figure. 25A). Even after fusing the peptides to highly soluble proteins, only ~10% of soc-wt peptide was in soluble form, whereas gfp-wt peptide was completely insoluble. gfp-wt peptide was purified from insoluble fraction by 8 M urea denaturation and renaturation. On the other hand, ~ 40-50% of scrambled peptide (soc-scrambled and gfp-scrambled) fusion proteins were in soluble form. The fusion peptides were purified by passing through the histrap affinity chromatography and Superdex 200gel filtration, as described in Materials and Methods (Figure: 25B). Peptides fused to the soc protein were tested in the, *in vitro* DNA packaging assay for their ability to inhibit DNA packaging by competing with full length gp17 for gp20 binding site on the proheads (Figure: 26). As shown in (Figure: 27, lane 4) soc-wt peptide at 6X higher ratio than the gp17 molecules inhibited 80% of DNA packaging, whereas the soc-scrambled peptides at the same concentration (lane 6) did not show any effect on the DNA packaging activity. In addition, a F348A mutant peptide behaved like the wt-peptide, showing 82% DNA packaging inhibition at 6 uM peptide concentration.

The peptides were next tested for their effect on gp16 stimulated ATPase activity of the gp17. As shown in (Figure: 28), none of the peptides, wt, scrambled, F348A mutant peptides at 6 uM (12 times higher molar ratio than gp17) concentration had any effect on gp16 stimulated ATPase activity of gp17. This suggests that the inhibitory effect of peptides on the DNA packaging is likely due to the inhibition of gp17 binding to gp20 on the proheads.

To further determine, whether the helix-loop-helix region is sufficient for gp20 binding, peptides were examined for prohead binding. The gfp–peptide construct was used for binding. The reasons for selecting gfp-peptide constructs for binding are, 1) from earlier binding results it is known that gfp itself doesn't bind to the proheads. 2) The molecular weight of the gfp-peptide is 31 kDa, on SDS-PAGE, at that position there is no background from the proheads. 3) soc-peptides cannot be used for prohead binding due to background binding of soc to the soc binding sites on the proheads.

Purified gfp-wt peptide, gfp-scrambled peptide and gfp, were incubated with proheads at ratio of 50 molecules of gfp-peptide/gfp to one molecule of gp20. As shown in (Figure: 29, lane 3) gfp-wt peptides bound to the proheads, whereas no binding was observed with the gfp-scrambled peptides or gfp (lanes 4 & 5) respectively. To test the specificity of peptide binding, gfp-wt peptides were also tested for binding with T4 phage. gfp-wt peptide did not show any significant binding to T4 phage, (Figure: 29, lane1). These results provide strong evidence that the gp17 HLH motif is the gp20 interacting region.

A



B

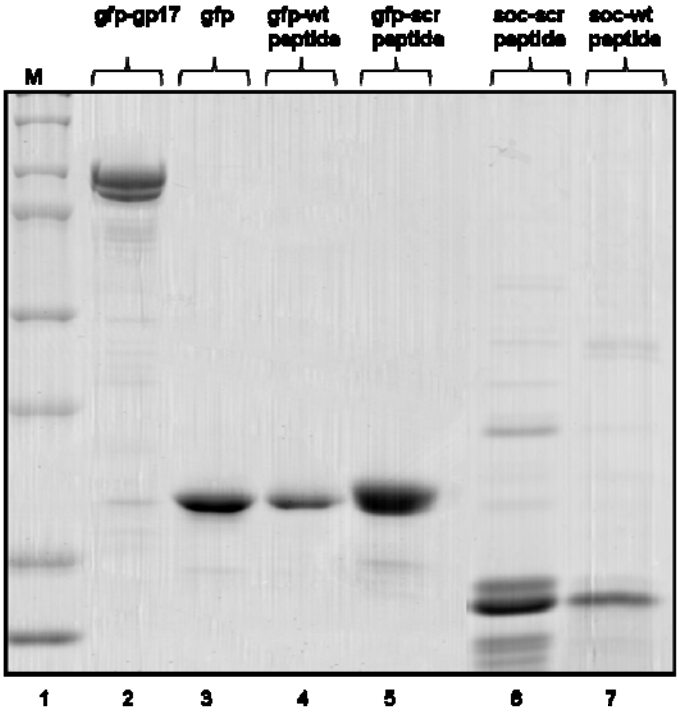


Figure 25: Construction and purification of helix-loop-helix peptide.
A: Constructs of helix-loop-helix peptide. wt sequence corresponding to helix-loop-helix region was fused inframe to the C-terminus of RB69-soc (a) and to the C-terminus of gfp protein (c). Scrambled peptide sequence fused to soc and gfp were used as a negative control, Scrambled sequence fused to RB69-soc (b) to gfp (d). The sequence corresponding to the wt and scrambled peptides are shown bellow.

B: SDS-PAGE showing the purity of fusion peptides. Lane 2: gfp-gp17. Lane 3: gfp. Lane 4: gfp-wt-peptide. Lane 5: gfp-scrambled peptide. Lane 6: soc-scrambled peptide. Lane 7: soc-wt peptide

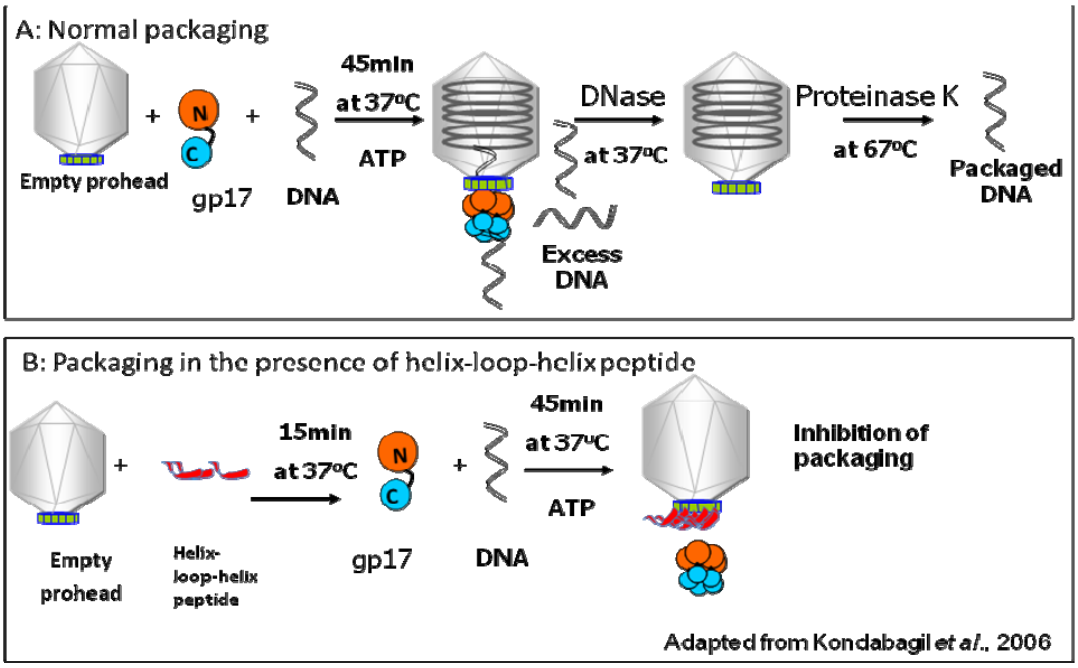
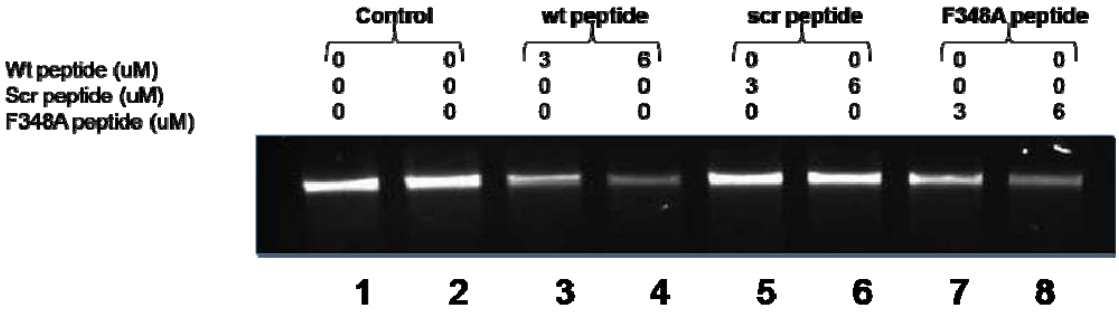


Figure 26: Schematic of *in vitro* DNA packaging inhibition assay.
A: In the defined packaging assay, proheads were incubated with DNA, gp17 and ATP at 37° C for 45 minutes. gp17 by using ATP energy translocates DNA into the capsid. Excess of DNA outside the capsid is degraded by DNase treatment. By proteinase K treatment the protected DNA will be released and quantified by agarose gel electrophoresis. B: DNA packaging in the presence of peptide, proheads were incubated with peptide at 37°C for 15 minutes then DNA, gp17 and ATP were added.

A



B

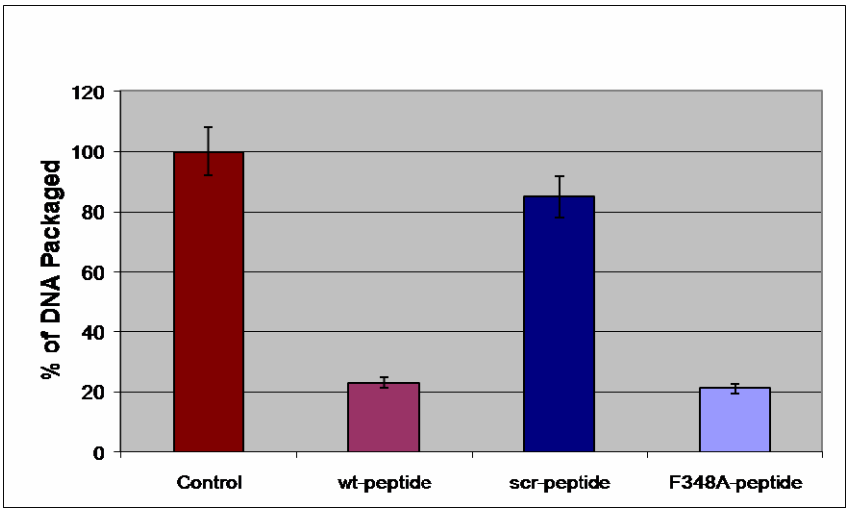


Figure 27: Inhibition of *in vitro* DNA packaging by helix-loop-helix peptide. Proheads ($4x10^9$particles) were first incubated with different peptides at indicated concentrations for 15 minutes and then rest of the packaging components were added. A: agarose gel showing DNA packaging in the presence of wt-peptide (lanes 3-4), scrambled peptide (lanes 5-6) and F348A mutant peptide (lanes 7-8) at the indicated concentrations shown above each lane. The amount of DNA packaged in the control (absence of any peptides) (lanes 1-2) is taken as 100%.

B: Histogram showing the percentage of DNA packaged in the presence of different peptides at 6 uM concentration. Y-axis is the percentage of DNA packaged. X-axis is the different peptides tested.

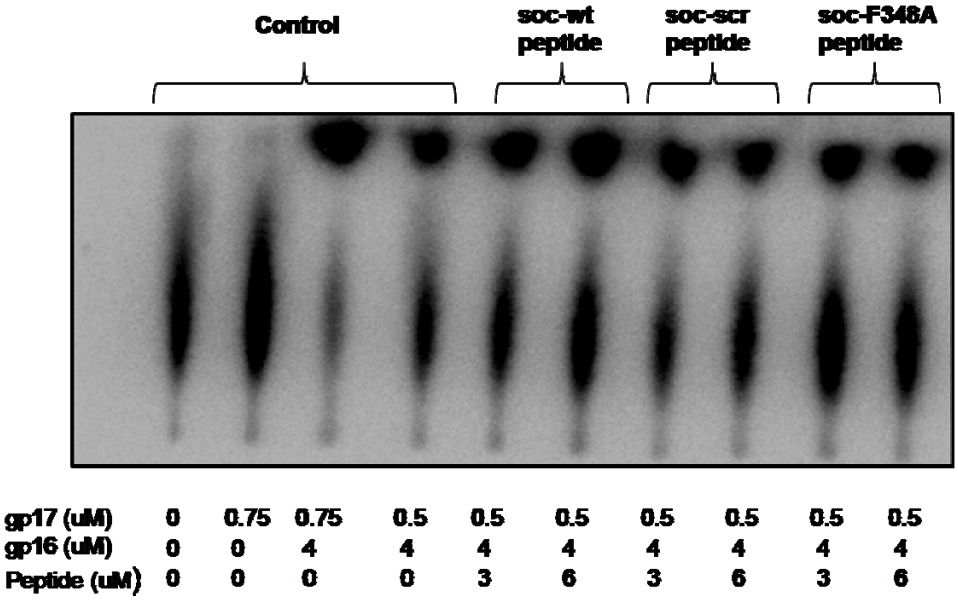


Figure 28: Effect of helix-loop-helix peptides on gp16 stimulated ATPase activity of gp17. Autoradiogram showing gp16 stimulated ATPase activity of gp17 in the presence of wt-peptides, scrambled peptides and F348A mutant peptides at 3 and 6uM concentrations. The concentrations of gp17 and gp16 used in the reaction mixture are indicated above.

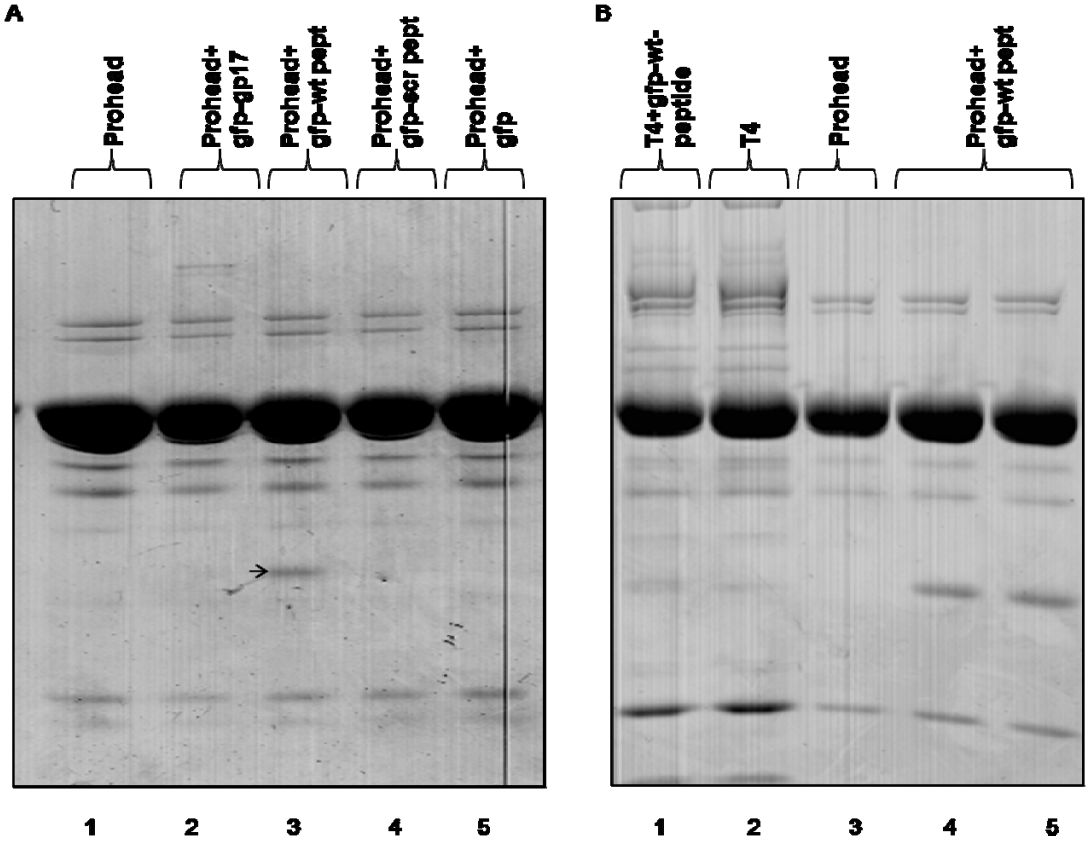


Figure 29: Specific binding of gfp-wt peptides to the prohead.
A: SDS-PAGE showing the binding of gfp-wt peptides (lane 3) and no binding of gfp-scrambled peptides (lane 4) or gfp (lane 5) to the proheads. Control proheads (lane 1), gfp-gp17 binding to proheads (lane 2) are shown.
B: SDS-PAGE showing binding of gfp-wt peptides to the proheads (lane 4 & 5) but not to the T4 phage (lane 1). Lane 2, T4 phage (control), lane 3, proheads (control).

gp20 interacting residues of the HLH motif

To identify the residues of helix-loop-helix region that are important for gp20 interaction, mutations were introduced to the HLH peptide. Sequence alignment and the secondary structure predictions showed that the secondary structure and sequence of helix-loop-helix region is highly conserved among T4 family phages (Figure: 30). Small variations in amino acid sequence were observed mainly in helix-1 region of HLH. We hypothesized that the conserved residues are important for maintaining the structure of the region whereas the non conserved residues provide the specificity to different T4-type phages and thus important for gp17-gp20 interaction. To test this hypothesis, three swap mutants were constructed. In case of swap mutants 1 & 2, non conserved residues from the helix-1 and the helix-2 were swapped (mutant 1 swapped with RB43 /RB16 sequence, mutant 2 swapped with phi-1 /RB49 sequence). In case of mutant 3 conserved residues from the loop region were swapped with SPM2 sequence (Figure: 31A). The F348A mutation was introduced to all the swap mutants, as this mutation was found to enhance the solubility of the peptides. Experimental design for the mutant constructs was similar to that of the wt peptide (Figure: 31B). soc- peptide fusions were purified (Figure. 31C) and tested for their effect on the DNA packaging activities. As shown in (Figure: 32A, lane 5 and 8), mutants 1 and 2 in which the helix residues were swapped, even at 9 uM concentration (9 times higher than gp17 molecules in the packaging reaction) didn't show any packaging inhibition, whereas the wt peptide (F348A is used as wt control) shows 80% inhibition at 6 uM concentration (lane 2). However, mutant 3 in which the loop residues were swapped showed 60% packaging inhibition at 7.5 uM concentration (lane

10) and 70% inhibition at 9 uM concentration (lane 12). In addition to this, GSS-AAA mutation was introduced to the full-length gp17 in gfp-gp17 background (Figure: 33A). The mutant protein was purified and tested for *in vitro* DNA packaging activities. The gfp-gp17GSS-AAA loop mutant showed no DNA packing activities. Even at 3 uM concentration it did not show any DNA packaging (Figure: 33 B). However, in consistent with the swap-peptide mutant 3 packaging inhibition result, gfp-gp17 GSS-AAA loop mutant also showed DNA packaging inhibition. At 2 uM protein concentration, it showed almost 95% packaging inhibition (Figure: 33C, lane 5). The DNA packaging inhibition by the loop mutant is likely due to the binding of mutant protein to the gp20 on the proheads and inhibition of wt-gp17 binding. To conform that, the mutation on loop residues doesn't affect the prohead binding, mutant protein was incubated with proheads and binding was analyzed on SDS-PAGE. As shown in the (Figure: 33D), the mutant protein showed clear binding to the proheads (lane 4) similar to gfp-gp17 wt (lane 2). These results are consistent with our prediction is that the helix residues in the HLH motif are important for function and variations in the helix sequence impart specificity and binding to the respective portal among the T4 related phages. The loop residues are less important and probably required for proper positioning of the helix residues.

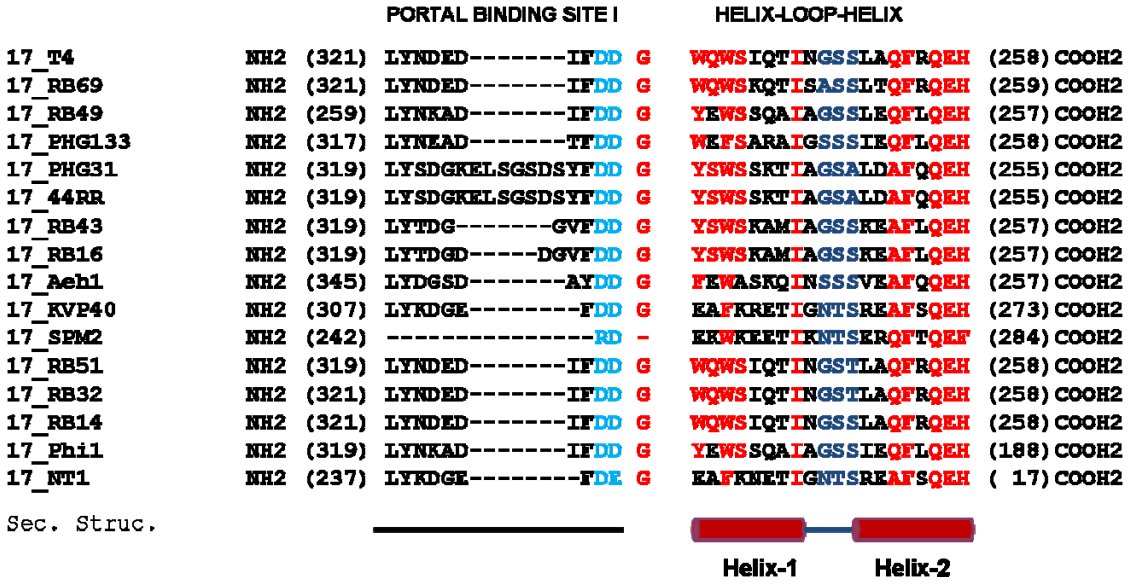


Figure 30: T4 family phage terminases were aligned with CLUSTALW.
The default alignment was then manually adjusted. Portal binding site I sequence is also included in the alignment. As shown in secondary structure, it is a long loop connected to helix-loop-helix structure. DD330-331 residues from portal binding site I which are critical (as observed from combinatorial mutagenesis) are highlighted in cyan.
Most of the residues from HLH region are highly conserved, which are colored in red (helix) and blue (loop). Non-conserved sequences are in black. Non-conserved sequences are mostly observed for helix-1 of the helix-loop-helix structure. Loop residues are colored in blue.
RB49 & phg133, phg31 & 44RR, RB43 & RB16, RB51 & RB32, T4 & RB14 phage terminases has identical sequence at HLH region.
All sequences were obtained from Gene bank, the accession numbers are; T4: P 17312, RB69: NP 861869, RB49: NP 891724, 44RR: NP 932508, RB43: YP 238880, Aeh1: NP 944105, KVP40: NP 899601, SPM2: YP 195134, RB51: YP 002854122, RB32: YP 803107, RB14: YP 002854500, Phi1: YP 001469498.

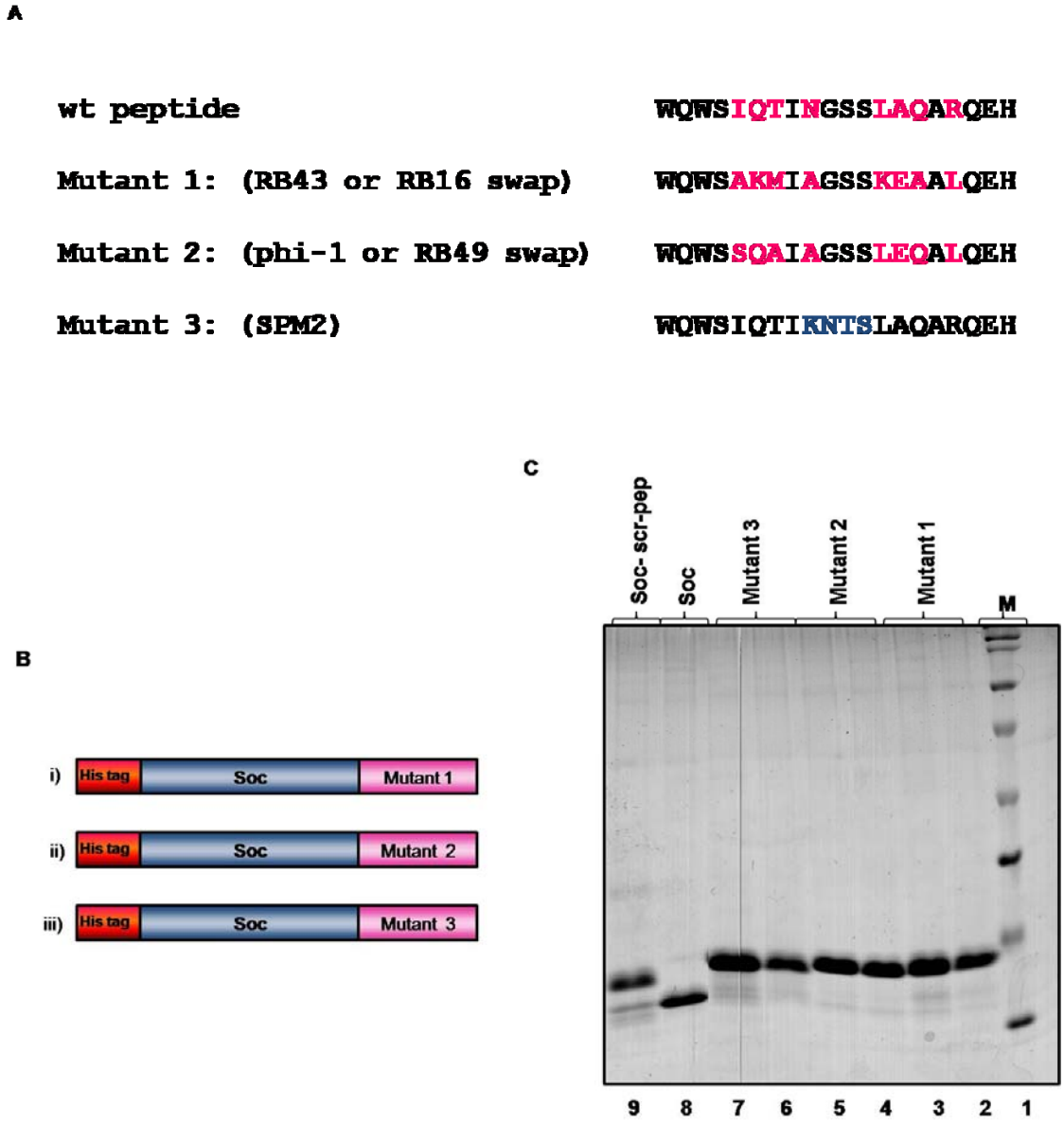


Figure 31: Construction and purification of HLH swap peptide mutants.
A: Sequences of swap mutants are shown. All the mutants include F348A mutation which enhanced the solubility of peptide. The amino acid sequences in pink are the non-conserved sequence. In case of mutant I & II, the non- conserved sequences from both the helices (colored pink) are swapped with RB-43/RB16 sequence and phi/RB49 sequence respectively. For mutant 3, conserved sequences from the loop region are swapped with SPM2 sequence (colored in blue).
B: Constructs of swap mutants. Mutant 1 (i), mutant 2 (ii), mutant 3 (iii) are fused in-frame with C terminus of Rb-69 soc. His-tag for the protein purification is added at the N-terminus of soc protein.

C: Purified swap mutant1 (lane 2 & 3), swap mutant 2 (lane 4 & 5), swap mutant 3 (lane 6 & 7). Molecular marker is shown in lane 1. soc and soc-scrambled peptide are used as control proteins (lane 8 & 9 respectively).

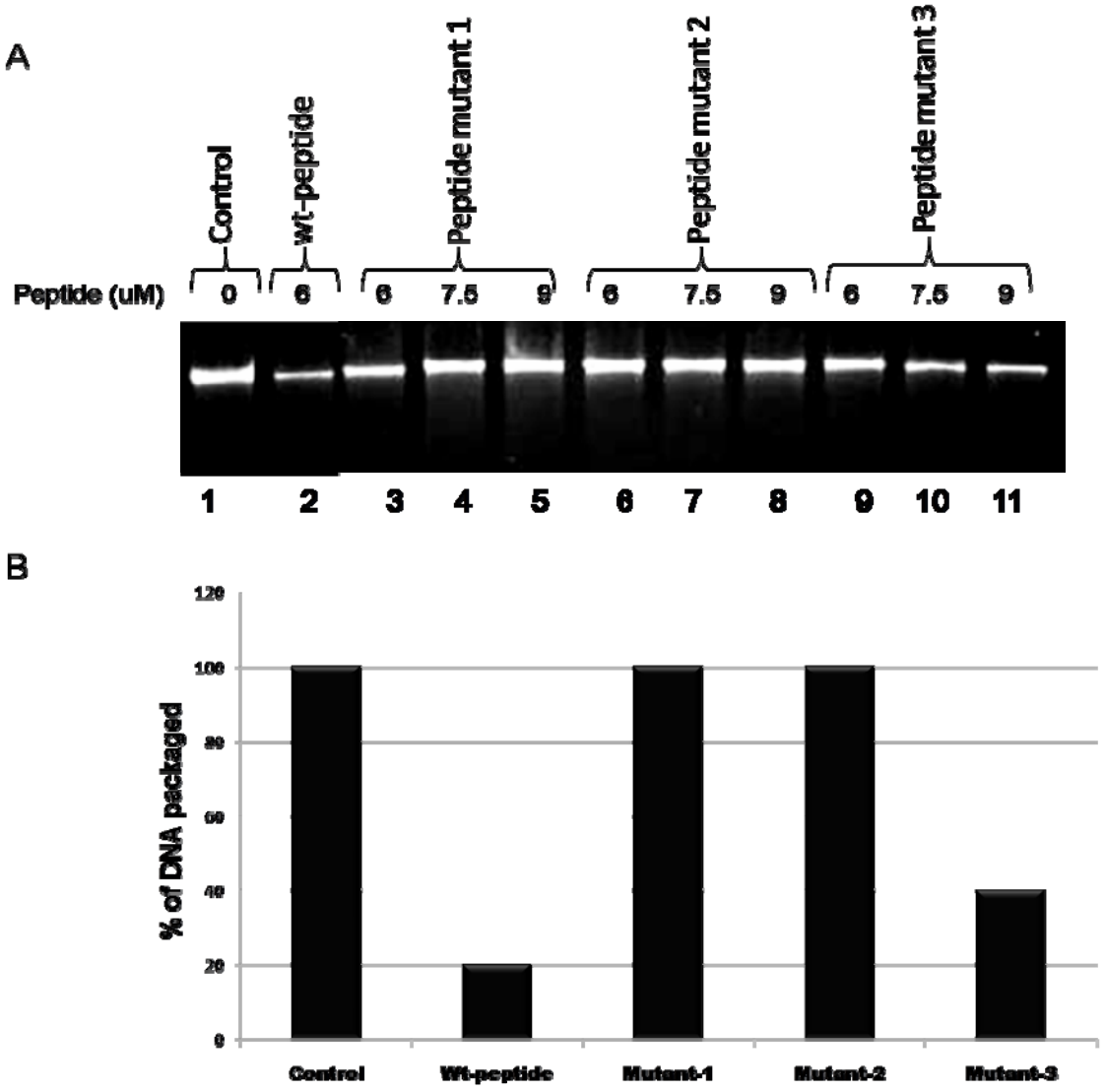


Figure 32: Effects of swap peptide mutants on DNA packaging activities. A: Proheads were incubated with increasing ratios of mutant peptides from (6 to 9) for 15 minutes and DNA packaging experiment was performed as described in Materials and Methods. Control: DNA packaging in the absence of any peptides (lane1). DNA packaging in the presence of wt (F348A) peptide (lane 2), mutant 1 (lanes 3-5), mutant 2 (lanes 6-8), and mutant 3 (lanes 9-11) at the concentrations/ratios of the peptides indicated above. The concentration of gfp-gp17 used in the DNA packaging assay is 1 uM.

B: Bar diagram of DNA packaging in the presence of swap peptide mutants at 7.5 uM concentration of peptides. Y-axis is the percentage of DNA packaged and X-axis is the different peptides tested.

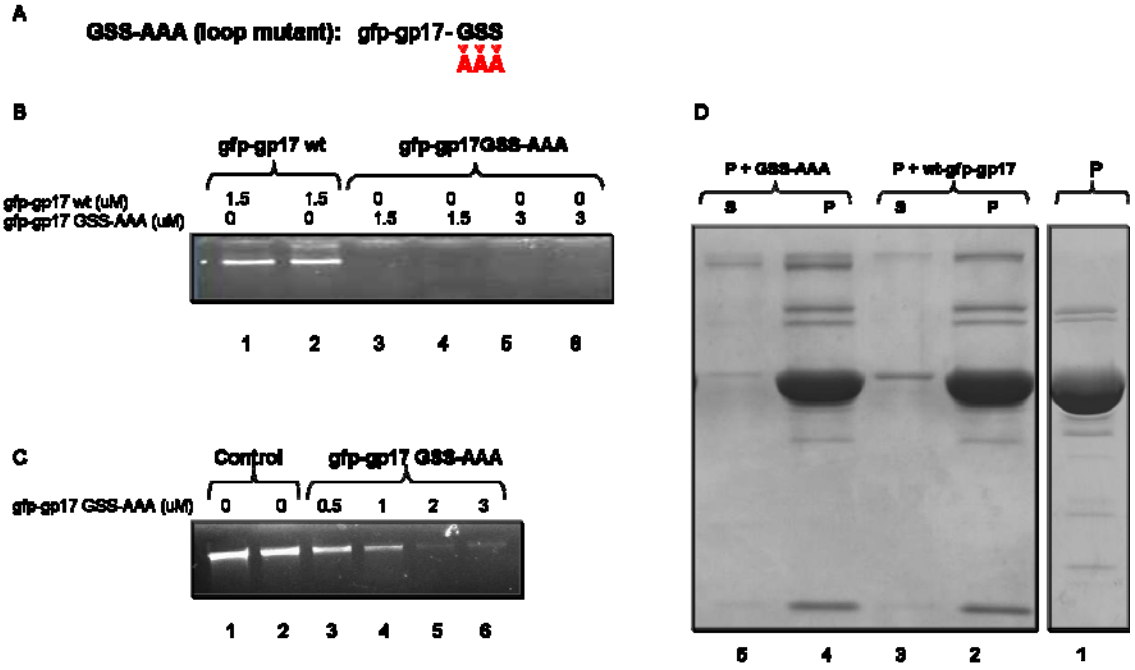


Figure 33: Characterization of gfp-gp17 GSS-AAA loop mutant. A: Construct of gfp-gp17 GSS-AAA loop mutant. GSS-AAA mutation was introduced into to the loop region in gfp-full length gp17 background.
B: Agarose gel showing the loss of *in vitro* DNA packaging activities of gfp-gp17 GSS-AAA loop mutant. Amount of DNA packaged by (control) gfp-gp17 wt at 1.5 uM concentration (lane 1 and 2), and no DNA packaging by gfp-gp17 GSS-AAA mutant at 1.5 uM (lanes 3 and 4) and 3 uM concentration (lanes 5 and 6).
C: Inhibition of *in vitro* DNA packaging activities by gfp-gp17 GSS-AAA loop mutant. Proheads were incubated with increasing concentration of mutant protein for 15 minutes and DNA packaging experiment was performed as described in Materials and Methods. Lane 1 and 2, control DNA packaging (in the absence of mutant protein), lanes 2-6, in the presence of increasing concentration (0.5 – 3 uM) of gfp-gp17GSS-AAA.
D: gfp-gp17 GSS-AAA loop mutant showing binding to the proheads. Prohead (control) (lane 1), prohead incubated with gfp-gp17 wt (lane 2), prohead incubated with gfp-gp17 GSS-AAA loop mutant (lane 4).

Discussion

DNA packaging is the most important step in the life cycle of a large dsDNA virus. Many tailed dsDNA bacteriophages e.g., λ, T3, T4, T7, Spp1 and some eukaryotic viruses such as herpes viruses exploit the terminase enzymes for this purpose (Catalano, 2005). The terminase holoenzymes from different phages characterized to date are found to be heteroligomers, encoded by two different gene products (Black, 1989; Catalano, 2000). In T4, these are called gp16 (small terminase) and gp17 (large terminase). The small terminase oligomerizes to form a ring of 8-10 subunits, stimulates the ATPase activity of the large terminase. Small terminase is required for recognizing the viral genome *in vivo,* however it is dispensable for *in vitro* DNA packaging (Rao & Black, 1988; Lin *et al*., 1997; Catalano *et al*., 1999; Kondabagil and Rao, 2006). Although the exact role of the small terminase in viral DNA packaging is still unclear, our recent data suggest that it is a regulator of the DNA packaging motor (Al-Zahrani *et al*., 2009). On the other hand, the large terminase posses multiple enzymatic activities required for DNA packaging. It has an N-terminal (N360) ATPase activity and a C-terminal (C360) nuclease activity (Bhattacharya and Rao, 1993; Baumann and Black, 2003; Rentas and Rao, 2003; Kanamaru *et al.*, 2004). The coordinated ATPase and the nuclease activities are coupled to DNA translocation into the capsid (Kubler and Rao 1998; Lin & Black 1998; Leffers and Rao, 2000; Mitchel and Rao, 2006; Draper and Rao, 2007).

DNA packaging is a complex multi-step process involving the assembly of motor onto the prohead, DNA translocation, and undocking of motor from the prohead. The intricate interaction between the terminase and the portal protein is thus critical for

successful DNA translocation. Indeed, in diverged phages like phi29, ATPase does not directly interact with portal instead the ATPase-portal interaction is mediated by a pRNA molecule (Morais *et al*., 2005). Five 170 bp RNA molecules appear to connect the pentameric ATPase to the dodecameric connector (Morais *et al*., 2005; Morais *et al*., 2008). This implicates that, although additional components are involved, the basic communication between the portal and the terminase as well as the symmetry mismatch between them is well preserved among diverged phages.

The X-ray structure of phi29 and SPP1 portals and Cryo-EM reconstructions of P22, T4 and e15 portals are available (Simpson *et al*., 2000; Lebedev *et al*., 2007; Leiman *et al*., 2004; Chang *et al*., 2006; Lander *et al*., 2006). In all these cases, the dodecameric ring structure of the portal is observed with wider end inside the capsid and narrower end protruding out of the capsid. This suggests that, despite very little sequence similarities, the overall structure and stoichiometrys of portal proteins are well conserved. In contrast to the well characterized portal protein, the oligomeric state of the terminase protein in packaging complex remains elusive. In fact, there is no direct evidence for portal-terminase interaction in any phage system. Direct structural-functional evidence is required to understand the interactions and role of symmetry mismatch between the portal and the motor and to characterize the detailed packaging mechanisms.

Direct evidence for gp17-gp20 interaction

Prohead binding data suggests that gp17-K577 or gfp-gp17 binds to the proheads. Binding of gp17 to the proheads is specific, as the control protein BSA or gfp did not show any binding. Binding of gfp-gp17 to the proheads increased with increasing ratio of

protein with maximum binding observed at protein: portal ratio of ~ 4:1 (Figure: 15A). As evident from gp20 minus prohead binding assay, gp17 specifically interacts with gp20, but not the other components of the proheads. SDS treatment releases the gp20 portal from the capsid without affecting the overall structure and functions of the capsid. This was clearly demonstrated by the binding of PA-Hoc to the hoc binding sites on the major capsid protein gp23.

In addition, gp17 did not bind to the T4 phage, because in mature phage neck proteins gp13 and gp14, tail and tail fibers are attached to gp20. Thus, gp20 binding site is inaccessible for terminase interaction. Furthermore, another important piece of evidence for gp17-gp20 interaction is the Cryo-EM reconstructions of proheads-gp17 complex. The Cryo-EM reconstructions clearly showed that the extra densities corresponding to the gp17 was observed only at the special vertex of the capsid where gp20 is located (Sun *et al*., 2008). Taken together, the biochemical and structural evidence clearly show for the first time that gp17 specifically interacts with gp20 on the proheads.

Stoichiometry of gp17 packaging motor

The number of gp17 subunits bound to the prohead was determined by scanning the density of gp17 band in the prohead-gp17 complex. In 14 independent experiments, 50% of the time, the calculated copy number was five and 35% of the time calculated copy number was six. From these results it is reasonable to suggest that five subunits of gp17 are bound per prohead particle. However, in view of the small copy number and potential experimental error in the technique used to determine the copy number, a copy

number six cannot be ruled out. In T3 phages biochemical studies suggested that six copies of terminase (gp19) are required for optimal DNA packaging (Fujisawa *et al*., 1991). The copy number six is more attractive than copy number five, as six subunits will have symmetrical interactions with the dodecameric protein and mechanistically it will be similar to the hexameric helicases. However, Cryo-EM of gp17 bound to the proheads consistent with the binding results, showed that there are only five subunits of gp17 bound at the special vertex of the capsid, similar to five pRNA molecules bind as a ring around the connector of the procapsid (Sun *et al*., 2008; Morais *et al*., 2008; Simpson *et al*., 2000). The symmetry mismatch between the portal (12-fold) and the terminase (5-fold) is very intriguing. Symmetry mismatch is also documented in other protein complexes such as F1-F0 ATPase (Junge *et al*., 1997), CipA-CipP, ATP dependent proteases (Beuron *et al*., 1998). Well-accepted idea about the symmetry mismatch is that it would serve as lubrication device to facilitate the relative rotations (Hendrix, 1978). Symmetry mismatch might also represent the quick release mechanism. It is easier for the gp17 motor to separate if the interactions between gp17 and gp20 takes are not as strong The symmetry mismatch could also allow synchronization of of different activities of the terminase protein during the DNA packaging and help in efficient assembly and disassembly of the motor during initiation and termination of DNA packaging.

gp20 interacting region resides in the N-terminal ATPase domain of gp17

Genetic and biochemical studies in T3 and λ suggested that the portal binding site is localized within last 15 amino acids of the large terminase protein. The proposed binding site in λ is $LYWEDD_{571\text{-}576}$ and in T3 is $LSGEDE_{636\text{-}641}$, both comprising of a

mixture of hydrophobic and acidic residues. In T4, however, the last 33 amino acids are not critical for functioning of the protein. gp17-K577 in which the C-terminal 33 amino acids are truncated retains all the activities including the ATPase and DNA packaging activities, of the full-length gp17 (Kannamaru *et al*., 2000). In addition, second site suppressors of gp20 *cs* mutations were mapped to the central S364N and C-terminal S583N regions of gp17 (Black *et al*., 1997). These data indicated that in T4, the portal-binding site is not restricted to the C-terminus of gp17 in contrast to what was observed for T3 and λ phages.

Initial prohead binding experiments showed that both N-domain (N360) and C-domain (C360) bind to the prohead. However under a stringent (300 mM salt) binding condition, only N360 bound to the proheads whereas C360 binding was drastically reduced. This result suggests that binding of N360 to the proheads is specific. Perhaps the most important piece of evidence for N-domain interaction with gp20 comes from Cryo-EM reconstruction studies. The Cryo-EM reconstructions of gp17-bound to proheads showed that the N360 domain was in close contact with the gp20 portal vertex of the proheads, whereas the C360 domain was away from the gp20 structure (Sun *et al*., 2008). These results clearly show that the gp20 interacting region is located in the N360 domain of gp17. Although, at present, it is not very clear the reason for C360 binding at low salt concentration (100 mM) and its inhibitory activities in the *in vitro* DNA packaging reaction. Two possible explanations could be that a) separation of full length gp17 into two domains exposes the charged residues on the surface of C-domain, which are responsible for non specific binding of C-domain to the capsid or b) at some point during

DNA translocation C-domain might be interacting with gp20 or capsid. If, the latter is true then Cryo-EM reconstructions of packaging complex at a different time point during the DNA packaging, should capture the C-domain near the capsid.

Mutational analysis of putative portal binding sites showed that the N-terminal portal binding site I is critical for gp17 function, whereas the C-terminal portal binding site II is not. Many substitutions were tolerated at DY560-561 residues of site II, whereas very few conserved substitutions were tolerated at DD330-331 residues of site I. Moreover, even the conserved substitutions such as N, A, and Q in the case of D331 or C and E in the case of D330 resulted in *ts* phenotype. Thus subtle changes such as increase in chain length by one C-C bond (D330E, D331Q) or changing the carboxyl group to amide group (D331N, D331Q) or sulphur group (D330C) leads to defective protein. From the structure it is clear that the DD330-331 residues are involved in interaction with the neighboring residues in the HLH motif. Indeed, the suppressor screening results showed that D331 residue can be suppressed by changing the neighboring I337 residue of the HLH motif. However, no second site suppressors in gp20 were isolated in our screening. The second site suppressor mutation was mapped to helix loop helix residue of gp17 gene. This implicated the importance of helix-loop-helix region in gp20 interaction. Structural analysis done by fitting the X-ray structures of T4 gp17 and Spp1 portal into the Cryo-EM density showed that the same HLH motif interacts with the portal. Indeed, the peptides corresponding to this region inhibited the DNA packaging activities of gp17 in a competitive manner. But a scrambled sequence of the same peptide didn't show any significant effect on the DNA packaging activities of gp17. In addition, a soc-F348A

mutant peptide also competitively inhibited the gp17 DNA packaging activities. From these data, it is reasonable to suggest that the HLH peptides inhibit DNA packaging by competing with gp17 for gp20 binding site on the proheads. Moreover, none of the peptides showed any effect on gp16 stimulated ATPase activity of gp17. Even at 6 uM concentration of peptide (twelve times higher molar ratio than gp17), the ATPase activity of gp17 was unaffected. Furthermore, when the peptides were incubated separately with proheads and T4 phage and analyzed for binding, gfp-wt peptides showed binding to the proheads but not to T4 phage. Neither the gfp-scrambled peptides nor gfp alone showed any binding to the proheads. Taken together, these results strongly suggest that the HLH peptide is sufficient for stable interaction with gp20.

Secondary structure predictions and the sequence alignment of T4 family phages suggested that the secondary structure as well as the sequence of HLH peptide is highly conserved with very little variations in the amino acid sequence. The variations were observed mainly in the helix-1 of the HLH peptide. These results led to hypothesis that the conserved residues from the helix-loop-helix region are important for maintaining the structure, whereas the variant residues are the gp20 interacting residues that determine the specificity of gp17-gp20 interaction. This hypothesis was tested by constructing three swap mutants. In swap mutant 1 & swap mutant 2, the non-conserved sequences from the T4 gp17 helices were swapped (see Figure: 30A) with that of T4-type phages. These peptides could not inhibit the DNA packaging suggesting that they lost the specificity and could not bind to T4 gp20. Even at 9 uM concentration (9 times of gp17 molecule) no packaging inhibition was observed. However swap mutant 3 in which the residues of the

T4 loop NGSS were swapped to that of SPM2 phage KNTS sequence (see Figure: 30 A) 60% DNA packaging inhibition was observed. Moreover, when the mutation was introduced to the conserved loop residues GSS-AAA in full length gp17 in gfp-gp17 background, the mutant protein had lost *in vitro* DNA packaging activities; however it retained the prohead binding activity and thus DNA packaging inhibition. At 2 uM concentration it completely inhibited DNA packaging activities of wt gp17, suggesting that mutant protein can compete with wt gp17 for gp20 binding site on prohead (Figure: 33D) and inhibit DNA packaging activities (Figure: 33C). Thus mutational analysis of HLH region concludes that mutation on conserved loop residues doesn't affect the gp20 interaction, whereas the mutation of non conserved helix residues affect the gp20 interaction. Therefore, it is likely that non conserved residues are gp20 interacting residues and conserved residues are not required for gp20 interaction, however they are important for maintaining the structure and overall functioning of the protein. In addition, the detailed structural analysis of helix-loop-helix region on solved gp17 structure indicates that only residues from helix-1 might contact with the gp20 residues. Therefore, the swap mutant 1 and swap mutant 2 which lost the DNA packaging inhibition must be due to loss of specific interaction between helix-1 residues and the gp20. If this is true that non conserved residues from helix-1 are involved in gp20 interaction, then the mutation of non conserved residues from helix-2 should not affect the gp20 interaction, and such mutant should not lose the DNA packaging inhibition activity. One way to test this would be to construct mutants by swapping the non-conserved residues from the helices one amino acid at a time and then testing for DNA packaging inhibition. However,

in the context of full length protein, even though only helix-1 residues are involved in specific interaction, it is likely that the helix-2 and loop region will also play a role in gp20 interaction.

Discovery of a communication relay center in the DNA packaging motor

HLH region is flanked by critical loops on both the sides, portal binding site I loop 323-331 on one side and hinge region loop 353-364 on other side. Portal binding site I loop connects the helix-loop-helix region to sub-domain I (ATP binding domain) and hinge region loop connects helix-loop-helix to C-domain (DNA binding domain). Therefore, the whole region must be presumably important for maintaining the structure and orientation of the helix-loop-helix as well as for transferring the signal to respective domains during the DNA translocation.

Symmetrical arrangement of interactions between the mismatched portal and motor

The data presented in this work first leads to the question of how HLH region accounts for symmetry mismatch between gp17 (five subunits) and gp20 (twelve subunits). Initially gp17 assembles into a pentameric motor through relatively weak interactions with gp20. In this arrangement each HLH will likely have different i.e., quasi-equivalent interactions, with the same protruding loops of gp20.

Molecular lever model for coupling of ATPase and DNA translocation through gp17-gp20 interactions

When DNA binds to C-domain of a gp17 subunit, it causes a conformational change in HLH which will insert into the gp20, results in the formation of a more extensive and unique interaction with gp20. In other words DNA binding locks that

motor subunit into a portal subunit causing that gp17 subunit to fire ATP hydrolysis. There is structural evidence for the rotation of sub-domain II by 6° in the tensed state of gp17. This then causes the movement of C-domain bound with DNA as well as movement of channel helix-5 connected to tunnel loop. This then causes the translocation of 2-bp into the capsid. The release of ADP and Pi products brings the sub-domain II and HLH back to the original position and the DNA is handed over to the next subunit for firing another gp17 ATPase.

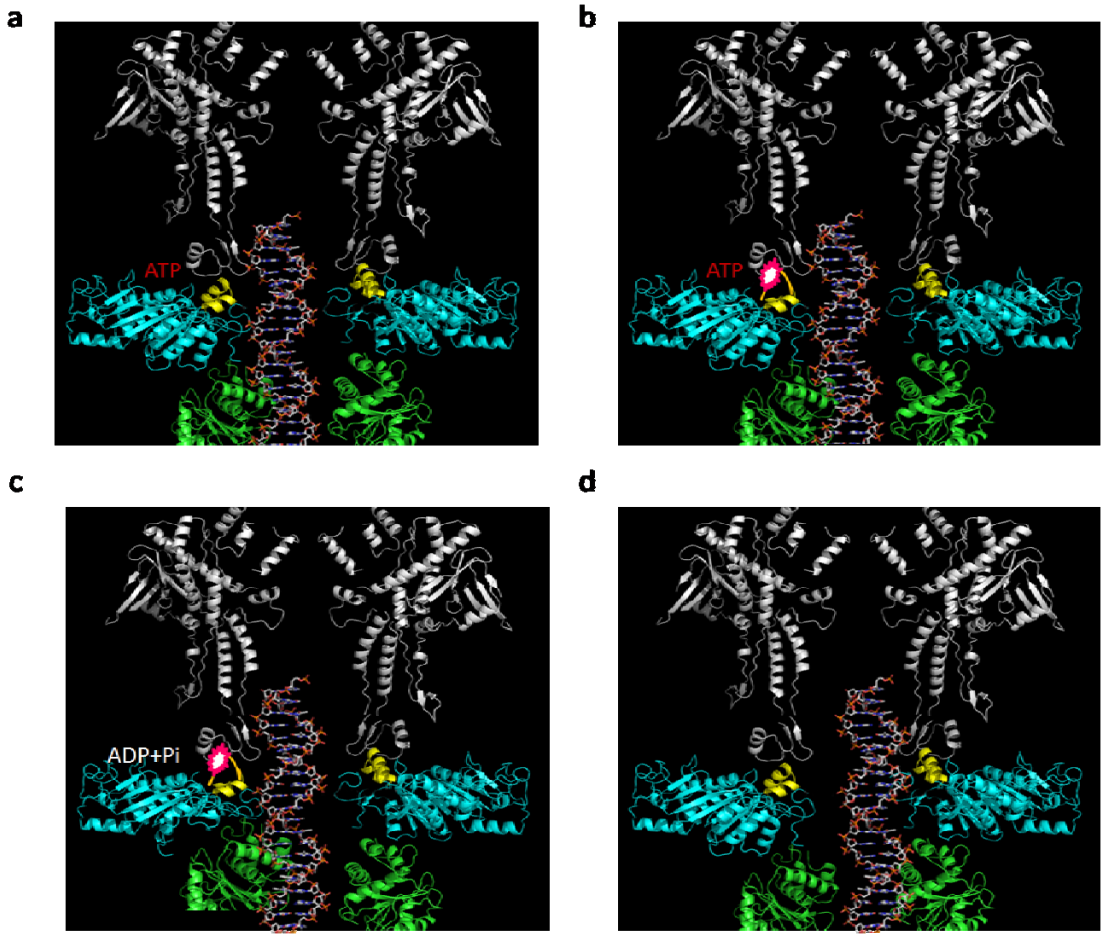


Figure 34: Symmetry mismatch model for DNA packaging. a: ATP binds to the N-domain of gp17 subunit whose C-domain bound with DNA. b: Binding of DNA causes conformational changes in helix-1 of helix-loop-helix region, results in formation of long mobile loop which inserts into gp20 and establish tight interaction, which will then signal the ATP hydrolysis. c: The signal from ATP hydrolysis from N-domain first transferred to gp20 through helix-loop-helix region possibly for rearrangement of tunnel loop for smooth passage of DNA. The signal is also transferred to the C-domain through the hinge region to move C-domain upwards to translocate the DNA into the capsid. d: Product release brings back the normal conformation of helix and C-domain back to original position. The DNA is handed over to adjacent gp17 subunit.

Although the model presented here is highly speculative, there are a number of evidences in T4 and other phages that are consistent with the model. In T3, terminase protease digestion sensitivity studies have shown that following ATP binding terminase

protein (gp19) undergoes conformational change which is required for terminase to form a functional complex with proheads (Morita *et al*., 1995). There are also studies on phi29 phage which suggests conformation change in terminase after binding to the proheads (Guo *et al*., 1987a). In Spp1, the biochemical and structural studies showed that the interaction of gp2 (terminase) with gp6 (portal protein) convert gp2 to a powerful packaging ATPase. In addition, mutations in the helices α3- α5- α6 residues of portal result in loss of gp6 dependent stimulation of gp2 ATPase activity and DNA translocation. Structural studies on Spp1 portal protein indicated differential positioning of tunnel loops of the subunits which is connected to a conformational change in helix α5 (Oliveira *et al*., 2006; Lebedev *et al*., 2007).

According to our electrostatic dependent mechanism, sub-domain I of gp17 produces energy by ATP hydrolysis and sub-domain II transfers this energy to the C-domain for DNA translocation (Sun *et al*., 2008). Thus, helix-loop-helix region being in sub-domain II is in ideal position to transfer the signal from the ATPase domain to gp20 on one hand and C-domain on the other. Thus translocation is synchronized or coupled through the domain movements communicated through this newly discovered communication relay center in the DNA packaging machine.

References

1. Alam, T.I., B. Draper, et al. (2008). “The headful packaging nuclease of bacteriophage T4.” Mol Microbiol **69**, 1180-1190 (2008).
2. Alam, T. I. and V. B. Rao (2008). "The ATPase domain of the large terminase protein, gp17, from bacteriophage T4 binds DNA: implications to the DNA packaging mechanism." J Mol Biol **376**(5): 1272-81.
3. Al-Zahrani, AS, K. Kondabagil, et al. (2009). “The small terminase, gp16, of bacteriophage T4 is a regulator of the DNA packaging motor.” Journal of biological chemistry **284** (36) 24490-500
4. Baumann, R. G., J. Mullaney, et al. (2006). "Portal fusion protein constraints on function in DNA packaging of bacteriophage T4." Mol Microbiol **61**(1): 16-32.
5. Beuron, F., A.C. Steven, et al. (1998). “At sixes and sevens: Characterization of the symmetry mismatch of the CIpAP chaperone-assisted protease. Journal of structural biology **123** (3): 248-259.
6. Bhattacharyya, S.P. & V.B. Rao, (1993). A novel terminase activity associated with the DNA packaging protein gp17 of bacteriophage T4. *Virology* **196**, 34-44.
7. Black, L. W. and D. J. Silverman (1978). "Model for DNA packaging into bacteriophage T4 heads." J Virology **28**(2): 643-55
8. Black, L.W., K. Ray, et al. (1989). “Portal control of viral prohead expansion and DNA packaging.” J Virology **391**(1): 44-50

9. Camacho, A. G., A. Gual, et al. (2003). "Bacillus subtilis bacteriophage SPP1 DNA packaging motor requires terminase and portal proteins." J Biol Chem **278**(26): 23251-9.

10. Chang, J., W. Chiu, et al. (2006). "Structure of epsilon 15 bacteriophage reveals genome organization and DNA packaging /injection apparatus." Nature **439**: 612-615.

11. Casjens, S., E. Wyckoff, et al. (1992). "Bacteriophage P22 portal protein is part of the gauge that regulates packing density of intravirion DNA." J Mol Biol **224**(4): 1055-74.

12. Crick, F. H., S. Brenner, et al. (1961). "General nature of the genetic code for proteins" Nature **192**: 1227-32

13. Cuervo, A., M. C. Vaney, et al. (2007). "Structural rearrangements between portal protein subunits are essential for viral DNA translocation." J Biol Chem **282**(26): 18907-13.

14. d'Herelle F. The bacteriophage. *Science News.* 1949:44–59.

15. Dhar, A. and M. Feiss (2005). "Bacteriophage lambda terminase: alterations of the high-affinity ATPase affect viral DNA packaging." J Mol Biol **347**(1): 71-80.

16. Draper, B. and V. B. Rao (2007). "An ATP hydrolysis sensor in the DNA packaging motor from bacteriophage T4 suggests an inchworm-type translocation mechanism." J Mol Biol **369**(1): 79-94.

17. Driedonks, R. A. and J. Caldentey (1983). "Gene 20 product of bacteriophage T4. II. Its structural organization in prehead and bacteriophage." J Mol Biol **166**(3): 341-60.

18. Fuller, D. N., D. M. Raymer, et al. (2007). "Single phage T4 DNA packaging motors exhibit large force generation, high velocity, and dynamic variability." Proc Natl Acad Sci U S A **104**(43): 168-73.

19. Goetzinger, K. R. and V. B. Rao (2003). "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**(1): 139-54.

20. Hendrix, R.W. (1978). "Symmetry mismatch and DNA packaging in large bacteriophages." Proc. Natl. Acad. Sci. USA **75**: 4779-783.

21. Hershey, A.D. and M. Chase (1952). "Independent functions of viral protein and nucleic acid in growth of bacteriophage".The journal of general physiology **36**(1): 39-56.

22. Horton, R.M, et al. (1990). "Gene splicing by overlap extension: tailor- made genes using the polymerase chain reaction." Biotechniques **5**: 528-35

23. Hsiao, C. L. and L. W. Black (1977). "DNA packaging and the pathway of bacteriophage T4 head assembly." Proc Natl Acad Sci U S A **74**(9): 3652-6.

24. Hsiao, C. L. and L. W. Black (1978). "Head morphogenesis of bacteriophage T4. III. The role of gene 20 in DNA packaging." Virology **91**(1): 26-38.

25. Hugel, T., C. Bustamante, et al. (2007). “Experimental Test of Connector Rotation during DNA Packaging into Bacteriophage φ29 Capsids.” PLoS Biology **5**(3):0558-0567

26. Iyer, L.M., K.S. Makarova, et al. (2004). “Comparative genomics of the FtsK-HerA superfamily of pumping ATPases: implications for the origins of chromosome segregation, cell division and viral capsid packaging”. Nucleic acid research **32**: 5260-279

27. Isidro, A., A. O. Henriques, et al. (2004). "The portal protein plays essential roles at different steps of the SPP1 DNA packaging process." Virology **322**(2): 253-63.

28. Junge, W., H. Lill, et al. (1997). “ATP synthase: an electrochemical transducer with rotary mechanics.” Trends Biochemical Sci **11**: 420-3

29. Kanamaru, S., K. Kondabagil, et al. (2004). "The functional domains of bacteriophage t4 terminase." J Biol Chem **279**(39): 40795-801.

30. Kellenberger E. (1990). “Form determination of the heads of bacteriophages.” European journal of biochemistry 190(2): 233-48.

31. Kuebler, D. and V. B. Rao (1998). "Functional analysis of the DNA-packaging/terminase protein gp17 from bacteriophage T4." J Mol Biol **281**(5): 803-14.

32. Kondabagil, K. R., Z. Zhang, et al. (2006). "The DNA translocating ATPase of bacteriophage T4 packaging motor." J Mol Biol **363**(4): 786-99.

33. Koti, J. S., M. C. Morais, et al. (2008). "DNA packaging motor assembly intermediate of bacteriophage phi29." J Mol Biol **381**(5): 1114-32.

34. Laemmli, U. K., N. Teaff, et al. (1974). "Maturation of the head of bacteriophage T4. III. DNA packaging into preformed heads." J Mol Biol **88**(4): 749-65.

35. Lane T and Eiserling F. (1992). “Genetic control of capsid length in bacteriophage T4. VII. A model of length regulation based on DNA size.” Journal of structural biology **104**(1-3):9-23

36. Lebedev, A. A., M. H. Krause, et al. (2007). "Structural framework for DNA translocation via the viral portal protein." Embo J **26**(7): 1984-94.

37. Lee, T. J. and P. Guo (2006). "Interaction of gp16 with pRNA and DNA for genome packaging by the motor of bacterial virus phi29." J Mol Biol **356**(3): 589-99.

38. Leffers, G. and V. B. Rao (2000). "Biochemical characterization of an ATPase activity associated with the large packaging subunit gp17 from bacteriophage T4." J Biol Chem **275**(47): 37127-36.

39. Lin, H. and L. W. Black (1998). "DNA requirements in vivo for phage T4 packaging." Virology **242**(1): 118-27.

40. Lin, H., V. B. Rao, et al. (1999). "Analysis of capsid portal protein and terminase functional domains: interaction sites required for DNA packaging in bacteriophage T4." J Mol Biol **289**(2): 249-60.

41. Maluf, N. K., Q. Yang, et al. (2005). "Self-association properties of the bacteriophage lambda terminase holoenzyme: implications for the DNA packaging motor." J Mol Biol **347**(3): 523-42.

42. Masey, T., J. Lowe, et al. (2006). “Double stranded DNA translocation: Structure and mechanism of Hexameric FtsK. Molecular cell **23**: 457-69

43. Matthews, CK, et al (1994). An overview of the T4 developmental program, in: Molecular biology of bacteriophage T4. Karam, JD., ed. ASM press, Washington, D.C.

44. Mesyanzhinov, V. V., M. M. Shneider, et al. (2004). “Molecular architecture of bacteriophage T4.” Biochemistry (Mosc) **69** (11): 1190-202.

45. Miyo Morita, et al. (1991) “Analysis of the fine structure of the prohead binding domain of the packaging protein of T3 using a hexapeptide, an analog of prohead binding site.” Virology **211** (2) 516-24

46. Mitchell, M. S., S. Matsuzaki, et al. (2002). "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**(18): 4009-21.

47. Mitchell, M. S. and V. B. Rao. (2006). "Functional analysis of the bacteriophage T4 DNA-packaging ATPase motor." J Biol Chem **281**(1): 518-27.

48. Moffitt, J. R., Y. R. Chemla, et al. (2009). "Intersubunit coordination in a homomeric ring ATPase." Nature **457**(7228): 446-50.

49. Moris, M.C., M. G. Rossmann, et al. (2005). “Conservation of capsid structure in tailed dsDNA bacteriophages: the pseudoatomic structure of phi29.” Molecular cell **18**(2): 149-59.

50. Moris, M.C., M. G. Rossmann, et al, (2008) “Defining molecular and domain boundaries in the bacteriophage phi29 DNA packaging motor”. Structure **16**(8): 1267-74

51. Morita, M., Tasaka, M. & H. Fujisawa (1995). “Analysis of the fine structure of the prohead binding domain of the packaging protein of bacteriophage T3 using a hexapeptide, an analog of a prohead binding site”. *Virology* **211**, 516-24.

52. Mosig, G. and F. A. Eiserling, (2004). T4 and related phages: Structure and development. In: The bacteriophages, 2nd edition. Calender, R. ed. Oxford university press, London

53. Oliveira, L., A. O. Henriques, et al. (2006). "Modulation of the viral ATPase activity by the portal protein correlates with DNA packaging efficiency." J Biol Chemistry **281**(31): 21914-23.

54. Leiman, Petr G., M. G. Rossmann, et al. (2004). “Three dimensional rearrangement of proteins in the tail of bacteriophage T4 on infection of its host.” Cell **118** (4) 419-29.

55. Rao, V. B. and M. Feiss (2008). "The bacteriophage DNA packaging motor." Annu Rev Genet **42**: 647-81.

56. Rao, V. B. and M. S. Mitchell (2001). "The N-terminal ATPase site in the large terminase protein gp17 is critically required for DNA packaging in bacteriophage T4." J Mol Biol **314**(3): 401-11.

57. Rishovd,S., Holzenburg, A., et al. (1998).“Bacteriophage P2 and P4 morphogenesis: structure and function of the connector. Virology **245**: 11–17

58. Simpson, A. A., Y. Tao, et al. (2000). "Structure of the bacteriophage phi29 DNA packaging motor." Nature **408**(6813): 745-50.

59. Stent G. S and S. Brenner (1961). "A genetic locus for the regulation of ribonucleic acid synthesis." Proc Natl Acad Sci U S A **47**: 2005-14

60. Studier F W., J.W. Doubendorff, et al. (1990). "Use of T7 RNA polymerase to direct expression of cloned genes." Methods Enzymology **185**: 60-89

61. Sun, S., K. Kondabagil, et al. (2007). "The structure of the ATPase that powers DNA packaging into bacteriophage T4 procapsids." Mol Cell **25**(6): 943-49.

62. Sun, S., K. Kondabagil, et al.,(2008). "The structure of the phage T4 DNA packaging motor suggests a mechanism dependent on electrostatic forces." Cell **135**(7): 1251-62.

63. Tavares, D, B.H. Lindqvist and S.D. Fuller (1993). "Image reconstruction from cryo-electron micrographs reveals the morphopoietic mechanism in the P2–P4 bacteriophage system, *EMBO J.* **11**: 839–46.

64. Yeo, A. and M. Feiss (1995). "Specific interaction of terminase, the DNA packaging enzyme of bacteriophage lambda, with the portal protein of the prohead." J Mol Biol **245**(2): 141-50.